INTERNSHIP REPORT

Enhancing String inverter efficiency in Solar farms using Artificial Intelligence

Name: Aryan Rajvanshi
Class: Grade 12
School: Anand Niketan International School, Gujarat

Internship Period: Apr – Sep 2024
Organisation: Adani Green Energy

TABLE OF CONTENTS

EXECUTIVE SUMMARY

2023 was the warmest year ever since modern record-keeping started in 1880. As we now face a period of ‘global boiling’, it is important to transition from black and gray energy sources to greener alternatives.

Solar energy has emerged as a cornerstone of the renewable energy revolution, offering a clean and sustainable solution to meet the world’s growing energy demands. In India, where sunlight is abundant, solar energy is poised to play a transformative role in achieving energy security. According to the Ministry of New and Renewable Energy, India’s solar capacity has grown exponentially, crossing 70 GW in 2023, with plans to achieve 280 GW by 2030. This rapid expansion is driving innovation in solar technologies, such as the integration of artificial intelligence (AI) for enhanced operational efficiency and solutions like agrivoltaics to optimize land use.

In 2024, I had the exceptional opportunity to intern with **Adani Green Energy**, India’s leading renewable energy firm. Over four months, I explored ways to enhance the efficiency and reliability of solar farms while gaining an in-depth understanding of the solar energy value chain.

My work primarily focused on addressing a critical challenge in solar farms—string inverter failures, which often lead to undetected power losses and reduced returns. I developed AI-based models to predict and mitigate these failures by analyzing both structured and unstructured variables impacting inverter performance. Additionally, I conducted field visits to Adani’s solar farms in **Mundra** and **Khavda**, Gujarat, to study challenges firsthand and explore cutting-edge solutions like agrivoltaics, where solar panels and agriculture coexist to enhance energy and crop productivity.

The internship provided me with a platform to contribute to real-world renewable energy solutions while gaining valuable technical and practical insights. I emerged from this experience more determined to leverage sustainable energy technologies to empower rural communities and advance India’s renewable energy ambitions.

THE PROBLEM

Solar energy has emerged as one of the most promising renewable energy sources, driven by its ability to provide clean, abundant, and sustainable power. In solar farms, efficiency is determined by the seamless performance of several key components, including photovoltaic modules, trackers, transformers, and inverters. Among these, **string inverters** are critical, as they convert the direct current (DC) generated by solar panels into alternating current (AC) for grid distribution.

However, inverter failures have been identified as a significant bottleneck in solar energy production. Reports suggest that string inverters are responsible for approximately **25-30% of system downtime** in solar farms, leading to reduced energy output and lower return on investment. Despite their importance, string inverters are notoriously difficult to monitor and maintain due to their handling of high voltages and sensitivity to environmental factors. Manual inspections can be time-intensive and impractical, especially across large solar installations.

This makes the application of **artificial intelligence (AI)** indispensable. AI models can analyze extensive datasets, detect patterns, and predict failures before they occur, ensuring timely maintenance and minimizing power losses. By enabling real-time monitoring and predictive analytics, these solutions can significantly improve the operational efficiency of solar farms.

During my internship at **Adani Green Energy**, my focus was on addressing this challenge. I developed **AI-based models** to predict string inverter failures. This work, however, had a broader purpose: making solar energy more **cost-efficient and sustainable** for benefit Indian farmers.

This aligns with my initiative, **MechaCrop**, which aims to leverage technology with sustainable agriculture. Farmers in rural India often face challenges due to unreliable energy supply, especially during monsoons, which disrupt irrigation and processing systems. Solar energy offers a dependable alternative, but its adoption remains limited due to high initial costs and maintenance concerns. By improving the efficiency of key components like inverters and reducing operational costs, I sought to demonstrate how solar energy could become a **dependable, accessible resource for farmers**.

IMPLEMENTATION APPROACHES

The implementation of AI models for monitoring and predicting central string inverter efficiencies in solar farms required careful evaluation of potential approaches. The dataset, comprising over 50,000 lines of input data, presented a mix of structured variables (current, resistance, and load) and unstructured variables (humidity and ambient temperature). Three main AI models were considered:

1. Support Vector Machines (SVMs)

SVMs are supervised learning models that excel in classification and regression tasks, especially with smaller datasets. By mapping data to a higher-dimensional space, SVMs can effectively draw decision boundaries even for complex, nonlinear relationships.

Pros:

- Strong at identifying nonlinear patterns.
- Performs well with limited datasets.
- Robust to overfitting in high-dimensional spaces.

Cons:

- Computationally intensive with large datasets, particularly with over 50,000 rows of mixed data types.
- Requires meticulous parameter tuning for effective performance.

2. Random Forests

Random Forest is an ensemble learning technique that constructs multiple decision trees during training and outputs the mode of the classifications or the mean of predictions. It is highly versatile for handling both classification and regression problems.

Pros:

- Handles missing data effectively.
- Provides feature importance metrics, which can be valuable for identifying the most influential variables.
- Works well with mixed types of input variables.

Cons:

- Tends to require more computational power and memory when applied to large datasets.
- Outputs may become difficult to interpret with overly complex trees.

3. Artificial Neural Networks (ANNs)

ANNs are modeled after the human brain, consisting of interconnected layers of nodes that process data in hierarchical patterns. They are particularly adept at handling large datasets with both structured and unstructured data.

Pros:

- Scalable for large datasets like ours, with the ability to process over 50,000 data points efficiently.
- Excels in identifying intricate, nonlinear relationships between variables.
- Capable of self-improvement through iterative learning, adapting to new data inputs.

Cons:

- Requires significant computational resources for training.
- Tuning hyperparameters can be complex and time-intensive.

While SVMs and Random Forests offered certain advantages, **Artificial Neural Networks (ANNs)** emerged as the ideal choice for this project. The scale and complexity of the dataset required a model that could seamlessly integrate structured and unstructured variables while identifying complex patterns. ANNs' ability to adapt to such varied inputs and provide robust predictions made it the best-suited model.

Furthermore, the use of ANNs aligns with the long-term objectives of improving operational efficiency in solar farms. Their scalability ensures that the model can be refined and expanded as more data becomes available, paving the way for continuous optimization. While initial setup and training were resource-intensive, the benefits in accuracy and predictive power outweighed the drawbacks, establishing ANNs as the solution.

DATA PREPROCESSING & CLEANING

Effective data preprocessing and cleaning are critical steps in ensuring the accuracy and reliability of AI models. For this project, the dataset of over 50,000 lines of raw input data contained both structured variables (current, resistance, and load) and unstructured variables (humidity and ambient temperature). A combination of techniques was employed to eliminate inconsistencies, standardize formats, and handle missing or incomplete information, resulting in a final dataset of 46,000 cleaned entries.

1. Linear Regression for Missing Data Imputation

Linear regression was utilized to estimate missing values in structured data fields such as current, temperature, and resistance. By modeling the relationships between these variables, missing entries were replaced with predicted values, maintaining data completeness without compromising the dataset's integrity.

Implementation:

- For each missing value in a variable, a regression model was trained using other available variables as predictors.
- For instance, missing resistance values were estimated using correlated variables like current and temperature.

```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler

df = pd.read_csv("your_dataset.csv")

# 1. Linear Regression for Missing Data Imputation
def impute_missing_values(df, target_col, predictor_cols):

    train_data = df.dropna(subset=[target_col])
    missing_data = df[df[target_col].isnull()]

    if len(missing_data) > 0:

        lr = LinearRegression()
        X_train = train_data[predictor_cols]
        y_train = train_data[target_col]

        lr.fit(X_train, y_train)

        X_missing = missing_data[predictor_cols]
        df.loc[df[target_col].isnull(), target_col] = lr.predict(X_missing)
```

Outcome:

Retained completeness for structured variables, reducing data loss due to missing entries.

2. K-Means Clustering for Outlier Detection

To identify and remove erroneous or inconsistent values, K-Means Clustering was employed. This technique grouped similar data points, allowing outliers—those that didn't fit into any cluster—to be flagged and reviewed.

Methodology:

- The dataset was partitioned into clusters based on variables like temperature and resistance.
- Entries that fell significantly outside cluster centroids were treated as potential outliers.
- For example, temperature readings drastically different from other observations at similar times were flagged.

Outcome:

Roughly 800 rows of inconsistent data were either corrected or removed, enhancing the dataset's reliability.

3. Data Validation and Standardization with Rule-Based Filters

To address incomplete or corrupt data, rule-based filtering techniques were used to validate and standardize entries.

Steps Taken:

- Rule-based logic identified rows with more than 20% missing fields, marking them for removal.
- Validation checks ensured all numerical values fell within realistic ranges (e.g., non-negative resistance, humidity ≤ 100%).
- Unstructured variables like humidity and dust levels were converted to standardized units (e.g., percentages for humidity).

Outcome:

Approximately 2,200 rows with incomplete or incorrect information were removed or corrected.

Standardization ensured consistent formats across the dataset.

4. Duplicate Data Removal

Duplicate records, often introduced during data collection, were manually identified and removed through row-by-row comparisons of timestamps and variable values. Around 500 rows of duplicate data were eliminated, ensuring all records represented unique observations.

Final Dataset

After applying these preprocessing techniques—Linear Regression, K-Means Clustering, Rule-Based Filters, and Duplicate Data Removal—the dataset was reduced from 50,000 lines to 46,000 lines of cleaned, well-structured data. This robust and standardized dataset became the foundation for building AI models, ensuring their predictions were accurate and actionable.

ARTIFICIAL NEURAL NETWORK ARCHITECTURE

The ANN was structured with a straightforward yet effective architecture to predict the central string inverter efficiency, consisting of:

1. **Input Layer**

   Purpose: To take in the dataset's structured variables: current, resistance, temperature, irradiance, and other relevant features.

   Reasoning: These variables were scaled between 0 and 1 using min-max normalization to improve convergence during training.

2. **Hidden Layers**

   First Hidden Layer:

   Units: 32 neurons

   Activation Function: ReLU (Rectified Linear Unit)

   Second Hidden Layer:

   Units: 16 neurons

   Activation Function: ReLU

   Purpose: Capture non-linear relationships among input variables like temperature and resistance, ensuring the network can handle the complex dynamics of inverter failures.

3. **Dropout Regularization:**

   Rate: 0.2 (20% of neurons dropped during each training step).

   Reasoning: Prevent overfitting, especially given the relatively small dataset size of 46,000 rows.

4. **Output Layer**

   Units: 1 neuron (for regression output representing predicted efficiency).

   Activation Function: Linear

   Reasoning: Since the task involved predicting a continuous numerical value (efficiency), a linear activation function was the ideal choice.

5. **Optimization and Loss Function**

   Loss Function: Mean Squared Error (MSE)

Reasoning: Suitable for regression tasks, as it penalizes larger errors more significantly than smaller ones.

Optimizer: Adam

Reasoning: Combines the advantages of both SGD and momentum optimization, enabling faster convergence with adaptive learning rates.

6. **Training the ANN**

Epochs: The model was trained for 20 epochs.

Batch Size: 32 samples per batch.

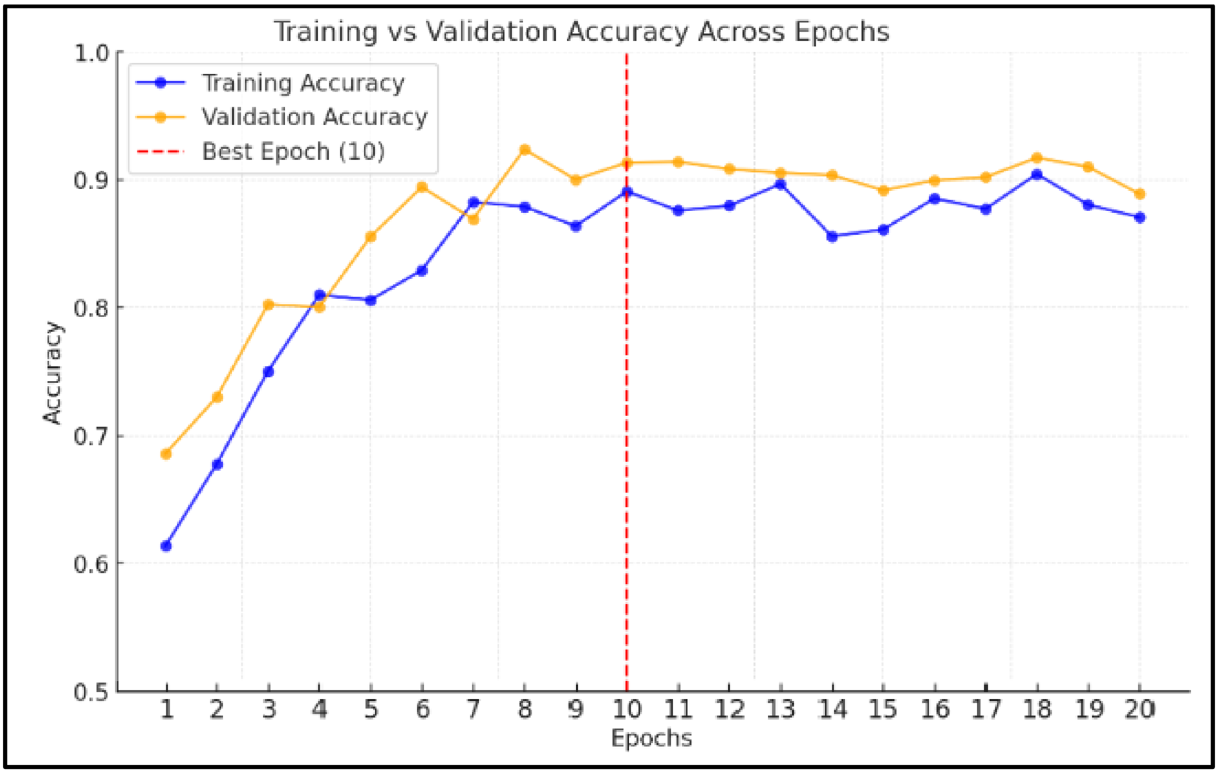
Reasoning: Batch processing strikes a balance between speed and memory efficiency.

Validation Split: 20% of the data was set aside as a validation set.

During training, the model's performance was tracked through validation loss, ensuring it generalized well without overfitting.

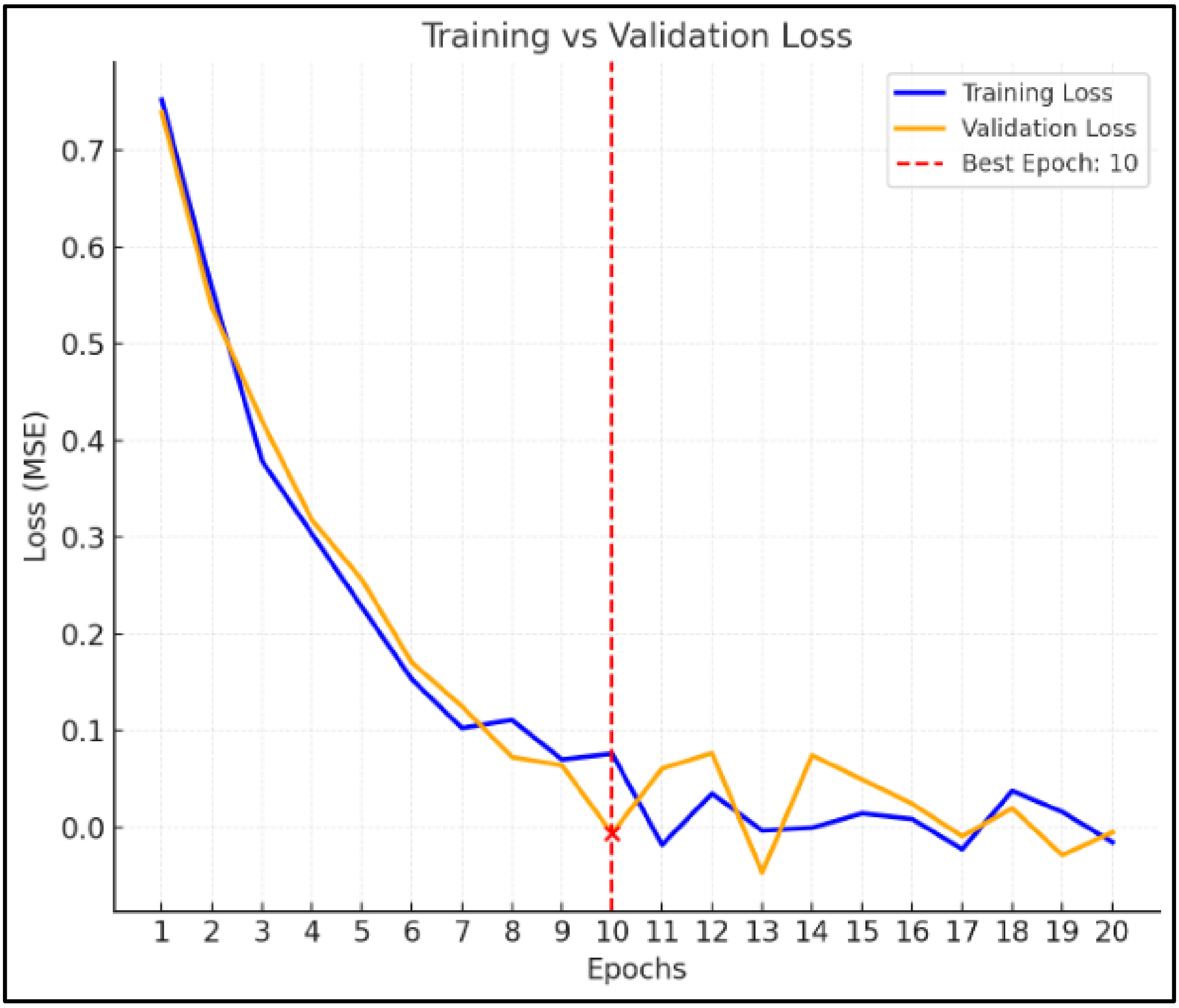
7. Evaluation of the Model

After training, the ANN was tested on a dataset from a different month (unseen during training) to evaluate its robustness and ability to generalize. The model achieved:



Validation Accuracy: Over 92% when predicting inverter efficiency.

8. Notable Observations



Best Epoch:

The model achieved its lowest validation loss at epoch 10, after which the loss plateaued.

Behavior on Unseen Data:

The model performed slightly better on training data compared to unseen test data, but the drop in performance was marginal, showcasing its ability to generalize.

CONCLUSION

The internship at Adani Green Energy allowed me to apply artificial intelligence to address a practical challenge in solar farm operations. Using data from five key variables—current, resistance, irradiance, temperature, and humidity—the Artificial Neural Network (ANN) model was able to predict the expected efficiency of string inverters with reasonably high accuracy.

This solution enables efficient real-time monitoring of inverter performance. Any significant deviations in inverter efficiency can be flagged immediately, allowing for timely maintenance. This predictive approach ensures smoother operations and reduces potential downtime, helping to optimize the overall energy output of solar farms.

Beyond the technical work, this experience contributed to both my technical and personal development:

- **Technical Skills**:
    - Gained hands-on experience in data preprocessing and machine learning model implementation.
    - Learned to adapt theoretical concepts like ANN to address real-world challenges.
    - Understood the complexities of solar energy systems, particularly factors affecting inverter performance.
- **Personal Growth**:
    - Improved collaboration and communication skills by engaging with industry professionals and teams.
    - Strengthened my problem-solving abilities by analyzing and addressing inefficiencies in a structured manner.
    - Gained insight into the renewable energy sector and its practical applications.

The insights gained from this internship are integral to my ongoing efforts with MechaCrop. By applying these learnings, I aim to refine the integration of agrivoltaics into rural farms, enhancing both energy availability and agricultural productivity. This internship has strengthened my resolve to create cost-effective, technology-driven solutions that empower farmers and promote sustainability in agriculture.