Solar Power Forecasting Using Artificial Neural Network

Rhythm Arya (202101023), Dev Changela (202101069), Aayush Patel (202101476), Raj Saradva (202101248), Suyash Bhagat (202101085), Sahil Bhadeshiya (202101511), Aryan Patel (202101226), Yashowardhan Bhagat (202101031), Neel Patel(202101458), Prashuk Jain (202101211)

***Abstract*—This study focuses on improving the accuracy of solar power forecasting using machine learning techniques, particularly Artificial Neural Networks (ANN). Factors such as temperature, angle of incidence, Zenith, and Azimuth are considered in the models, and they have shown better prediction accuracy than traditional methods. The study evaluates the models using the Pearson Correlation Coefficient (PCC) and Hyperband optimization. Accurate solar power forecasting is essential for grid stability and effective energy management.**

***Index Terms*—Large Scale Solar Photovoltaic (LSSPV), Solar generation forecasting, Artificial Neural Network (ANN), Pearson Correlation Coefficient (PCC), Seasonal Autoregressive Integrated Moving Average (SARIMA), Adaptive Network-based Fuzzy Inference System (ANFIS), Long Short-Term Memory (LSTM), Hyperband, Grid stability, Angle of incidence, Zenith, Azimuth**

I. INTRODUCTION

SOLAR power is a vital renewable energy source, offering a sustainable alternative to fossil fuels. While solar cells' efficiency has yet to be optimal, predicting solar power output is crucial for maximizing energy generation. Solar power forecasting is quite helpful for a solar power plant to operate and be controlled smoothly. The amount of solar radiation that strikes a solar panel or cell determines its ability to generate energy, along with other parameters, including dust accumulation on the panels, humidity and temperature in the surrounding air, and the doping level and design of the solar PV array. Due to their inherent variability, these factors directly affect the solar cell's output. Along with all the previously listed variables, the sun irradiance varies throughout the day. For more accurate forecasting of the production of a PV module and, consequently, a solar power plant, the hourly average or average of these metrics received is measured at a specific interval of time. Machine learning techniques, mainly Artificial Neural Networks (ANN), are increasingly employed for this purpose.

With machine learning, especially ANN, being a prominent tool, we can improve accuracy; these models consider various factors such as temperature, cloud cover, wind speed, and humidity. Precise forecasting is essential for grid stability and effective energy management. ANN models are adequate for handling non-linear data and learning from examples. Studies show that ANN models provide better prediction accuracy than traditional methods like SARIMA. Choosing the correct input variables is crucial for accurate forecasting, as demonstrated by models using LSTM and ANFIS for solar radiation prediction. The study evaluates the model's performance using the Pearson Correlation Coefficient (PCC) and Hyperband optimization to enhance accuracy.

TABLE I: Characterisation of ANN parameters.

Parameter	Value
Solver	Adam
Epochs	150
Batch size	128
Learning rate	0.001
Shuffle	False
Hyper-parameter optimizer	Hyperband

TABLE II: Pearson Correlation Coefficient (PCC) between solar generation and meteorological parameters.

Inputs	Pearson Correlation Coefficient
Total Global Horizontal Irradiance (W/m^2)	0.92
Total Horizontal Irradiation (Wh/m^2)	0.14
Global Irradiance on the Module Plane (W/m^2)	0.92
Total Slope Irradiation (Wh/m^2)	0.14
Ambient Temperature (°C)	0.62
PV Module Temperature (°C)	0.89
Wind Speed (m/s)	0.47

A. Problem Statement

We aim to anticipate the forthcoming solar energy output at precise points. This involves scrutinizing past solar power production records in conjunction with relevant meteorological data, encompassing irradiance, temperature, humidity, and atmospheric pressure. The ultimate goal is to minimize forecasting errors and enhance the reliability and efficiency of solar power integration into the grid, thereby contributing to a more sustainable energy landscape.

B. Relevant Works

Daniel O'Leary and Joel Kubby's "Feature Selection and ANN Solar Power Prediction" focuses on feature selection techniques to enhance prediction accuracy by identifying the most relevant input variables using historical solar data and meteorological parameters; researchers train and validate ANN models. [1]

Hanis Nasuha Ame, Nofri Yenita Dahlan, Azlin Mohd Azmi, Mohd Fuad Abdul Latip, Mohammad Syazwan Onn, Afidalina Tumian, "Solar power prediction based on Artificial Neural Network guided by feature selection for Large-scale Solar Photovoltaic Plant," includes enhancing solar power generation forecasting in large-scale PV plants using Artificial

Neural Networks (ANNs) with feature selection, crucial for efficient energy management and grid integration. [2]

Rahul Gupta, Anil Kumar Yadav, SK Jha, Pawan Kumar Pathak, ”Time Series Forecasting of Solar Power Generation Using Facebook Prophet and XG Boost.”This paper investigates time series forecasting of solar power generation using two approaches: Facebook Prophet and XGBoost. It explores the effectiveness of both methods in predicting solar power output, crucial for efficient energy management. [3]

K. Anuradha, Deekshitha Erlapally, G. Karuna, V. Srilakshmi, K. Adilakshmi, ”Analysis Of Solar Power Generation Forecasting Using Machine Learning Techniques”.The paper includes a machine learning-based approach for forecasting solar power generation in India using environmental data, offering insights into variable importance over different periods. It compares Support Vector Regression, Linear Regression, and Random Forest models. The study emphasizes the importance of solar power forecasting for grid planning, highlighting the challenges posed by wealth-dependent generation. [4]

C. Our Contributions

Our group has significantly contributed to the field of solar power forecasting by enhancing prediction accuracy using machine learning techniques, particularly Artificial Neural Networks (ANN). We've addressed the problem's complexity by incorporating factors such as temperature, angle of incidence, zenith, and azimuth into our models, recognizing the multifaceted nature of solar power generation.

Our novelty lies in our evaluation approach, utilizing the Pearson Correlation Coefficient (PCC) and Hyperband optimization to assess model performance. This comprehensive evaluation framework provides a deeper understanding of model accuracy and efficiency, surpassing traditional evaluation methods.

Throughout our research, we've had several light-bulb moments. One was the realization of the importance of feature selection in improving prediction accuracy. Another was recognizing the potential of machine learning techniques, such as ANN, in handling non-linear data.

Our solution improves upon existing state-of-the-art (SOTA) methods by combining advanced machine learning techniques with meticulous feature selection. By prioritizing relevant input variables and leveraging historical solar data and meteorological parameters, our models achieve superior prediction accuracy for solar power generation. This improvement is crucial for optimizing energy management and grid integration, thereby advancing the sustainability and efficiency of solar power plants.

D. Organization

Section I has an introduction and motivation, Section II contains the approach to the solution we are proposing, Section III includes the algorithm description, Section IV contains Discussion and Remarks, Section V shows Numerical Results, and Section VI concludes the report.

II. Proposed Approach

A. *Data Collection and Preprocessing*

- **Variables**: The dataset includes various meteorological and atmospheric variables such as temperature, humidity, pressure, precipitation, cloud cover, and wind characteristics.
- **Solar-specific Parameters**: It also contains solar-specific parameters like the angle of incidence, zenith angle, azimuth angle, and the actual generated power in kilowatts.
- **Environmental Conditions**: These variables describe the environmental conditions that affect solar power generation.
- **Modeling Purpose**: The dataset is suitable for modeling the relationship between environmental conditions and solar power generation, which is essential for solar power forecasting.
- **Machine Learning Application**: With this dataset, we developed and trained an artificial neural network (ANN) model to forecast solar power generation.
- **Imported Libraries**: We imported necessary libraries such as pandas for data manipulation, NumPy for numerical operations, TensorFlow and Keras for building and training neural networks, and matplotlib and seaborn for data visualization.
- **Preprocessing**: Once the time series datasets are collected, we will preprocess the data to clean and prepare it for analysis. This preprocessing step will involve handling missing values, removing outliers, and normalizing the data to ensure uniformity across different variables. Additionally, feature engineering techniques will be applied to extract relevant features from the time series data, such as aggregating hourly values into daily or monthly averages.

B. *Feature Selection/Extraction*

- Feature selection for solar power forecasting using an ANN method involves choosing the most relevant variables from the dataset. This is done by analyzing the correlation with the target variable (solar power generation), considering the importance of the feature from the ANN model, and applying domain knowledge. Dimensionality reduction techniques may also be used to select the most important features.
- Feature scaling is performed using StandardScaler from sci-kit-learn to standardize the features by removing the mean and scaling to unit variance.

C. *Model Architecture*

Input Layer: The input layer of the neural network takes the features related to weather conditions as input. These features include:

- Temperature at 2 meters above ground level
- Relative humidity at 2 meters above ground level
- Mean sea-level pressure
- Total precipitation at the surface
- Snowfall amount at the surface
- Total cloud cover at the surface

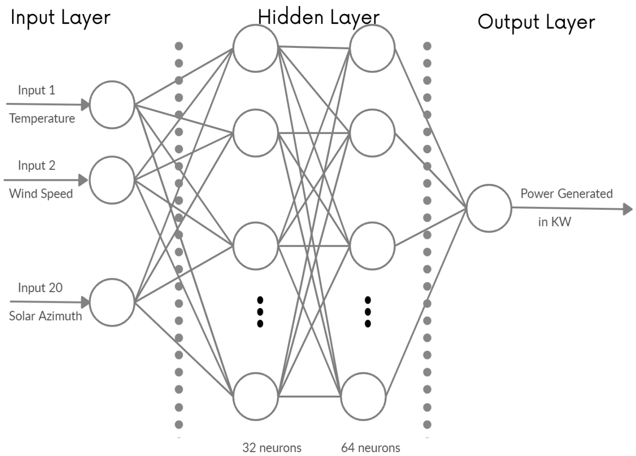
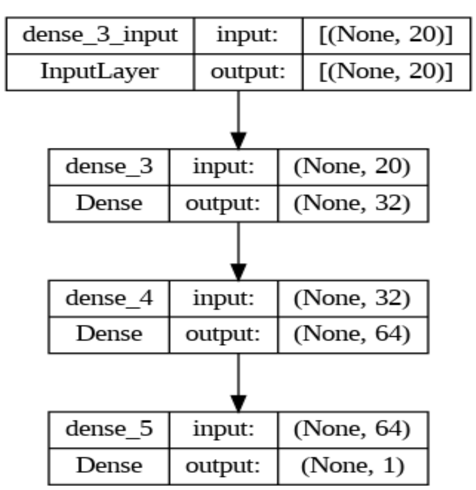


Fig. 1: Model Architecture

- High, medium, and low cloud cover at different layers
- Shortwave radiation at the surface
- Wind speed and direction at 10 meters and 80 meters above ground level
- Wind speed and direction at 900 mb
- Wind gust speed at 10 meters above ground level
- Angle of incidence, zenith, azimuth (related to solar position)

Hidden Layers: The model architecture includes multiple hidden layers with ReLU (Rectified Linear Unit) activation functions. The number of neurons in each hidden layer is specified as [32, 64]. Dropout layers are used for regularization to prevent overfitting during training.

Output Layer: The output layer is a single neuron with a linear activation function, as this is a regression task aiming to predict continuous values (solar power generation).



Optimization and Learning Rate Scheduling: The model uses the Adam optimizer for gradient descent optimization. A learning rate schedule is implemented using Exponential Decay, where the learning rate starts at 0.001 and decays over time with a decay rate of 0.90 and decay steps of 1000.

Loss Function and Metrics: The model is trained using Mean Squared Error (MSE) as the loss function, which measures the average squared difference between predicted and actual values. The evaluation metric used during training is Root Mean Squared Error (RMSE), which provides a measure of the model's prediction accuracy.

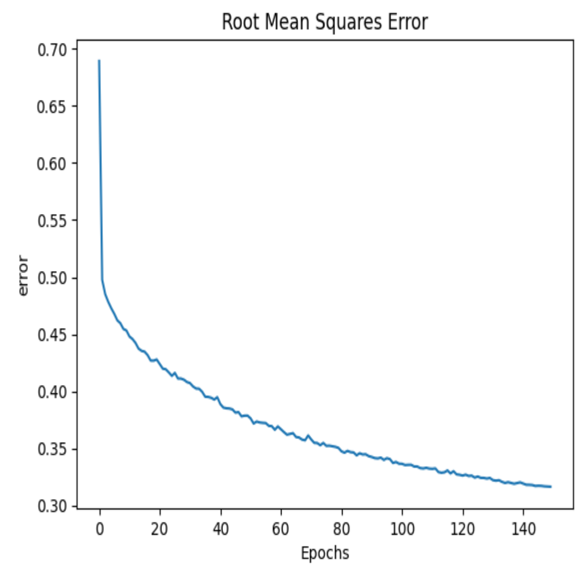


Fig. 2: Root Mean Square Error (error vs Epochs)

D. *Creating Neural Network*

- The create_spfnet function is defined to create a neural network model using TensorFlow and Keras.
- The function takes parameters such as the number of layers, activation function, and kernel initializer.
- It constructs a sequential model with specified layers, activations, and kernel initializations.
- The model is compiled with Mean Squared Error (MSE) loss function and Adam optimizer.

TABLE III

Layer (type)	Output Shape	Parameter #
dense$_3$($Dense$)	(None, 32)	672
dense$_4$($Dense$)	(None, 64)	2112
dense$_5$($Dense$)	(None, 1)	65

Total params: 2849 (11.13 KB)
Trainable params: 2849 (11.13 KB)
Non-trainable params: 0 (0.00 Byte)

E. *Training the model*

- The neural network model is created using the create_spfnet function with specified parameters.
- The model summary is printed to display the architecture.
- The model is trained using the training data (X_train and y_train) with a specified batch size, validation data, and number of epochs.
- Training history is stored in the hist variable.
- The model is trained using the training data with a batch size of 32 and for 150 epochs. Early stopping is implemented with a patience of 10 epochs to prevent overfitting, and the best weights are restored.

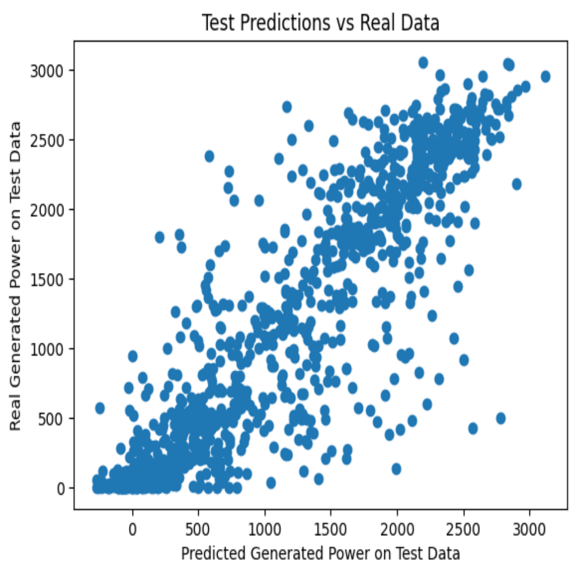


Fig. 3: Test Predictions vs Real Data

Fig. 4: Training Predictions vs Real Data

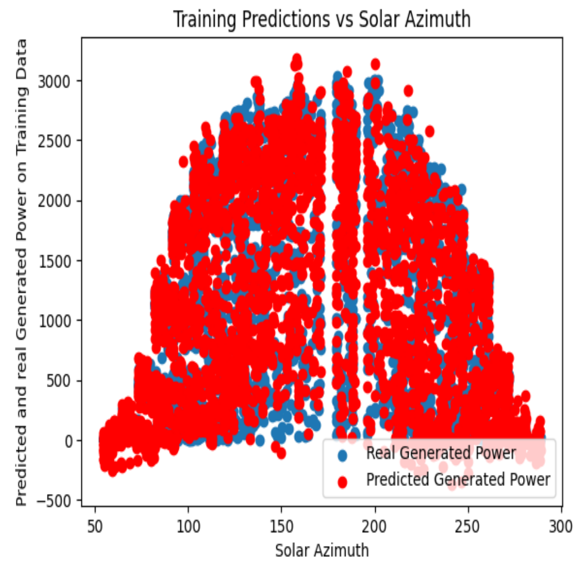


Fig. 5: Training Predictions vs Solar Azimuth

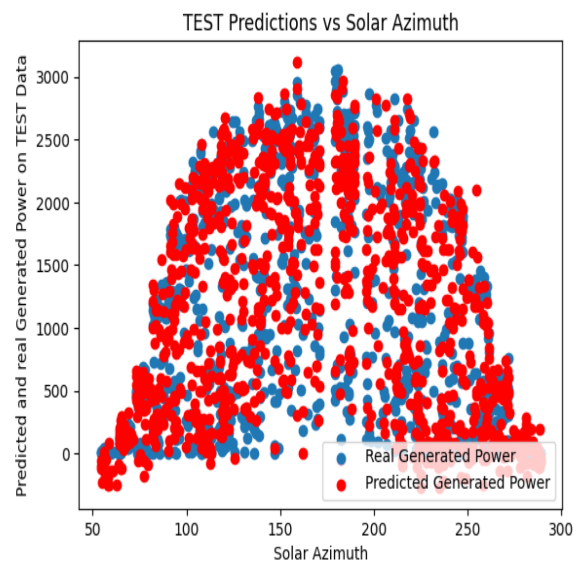


Fig. 6: Test Predictions vs Solar Azimuth

F. *Model Evaluation*

- The model's performance is evaluated on both the training and test sets to assess its ability to generalize to new data.
- The model's performance is evaluated using the testing data (X_test and y_test).
- Root Mean Squared Error (RMSE) is calculated both for the training and testing sets.
- Predictions are made on the testing set and then transformed back to the original scale using inverse scaling.
- RMSE is calculated on the original scale to provide a meaningful error metric.

G. *Observations:*

- High correlation between Zenith and Angle of Incidence of 0.71.

- Shortwave radiation backwards and Generated Power KW have a correlation of 0.56.
- Relative Humidity and Zenith are positively correlated (0.51).
- Relative Humidity and Low Cloud Cover are positively correlated (0.49).
- Angle of Incidence and Zenith are negatively correlated with Generated Power (-0.65).
- Negative correlation between Zenith and temperature of -0.55.
- High negative correlation exists between Shortwave radiation backwards and Zenith (-0.8).
- Shortwave radiation backwards and Relative humidity are negatively correlated (-0.72).
- Relative humidity and Temperature are negatively correlated (-0.77).

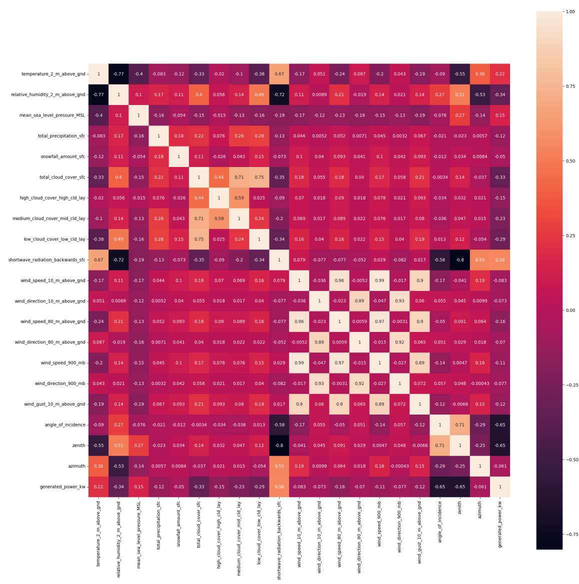


Fig. 7: Correlation Grid

H. Lasso Regression:

Lasso regression is applied to the data to perform feature selection. The Lasso model is trained and used to identify the most important features for predicting solar power generation.

I. Visualization and Analysis:

- Various plots, including scatter plots, histograms, and bar plots, are generated to visualize the model's predictions and analyze the data distribution and trends.
- Seasonality analysis is performed to understand how solar power generation varies over different months.

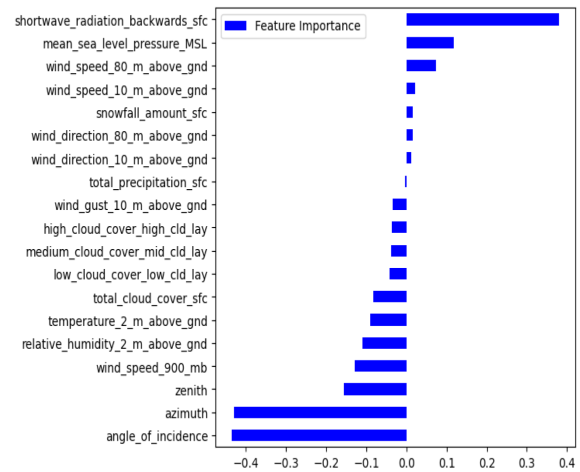


Fig. 8: Feature Importance

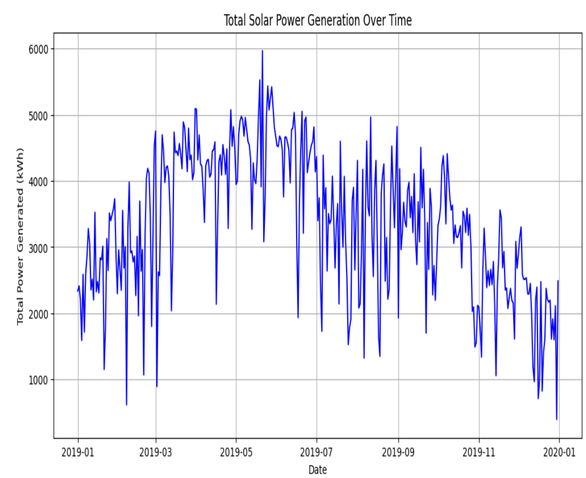


Fig. 9: Total Solar Power Generation Over Time

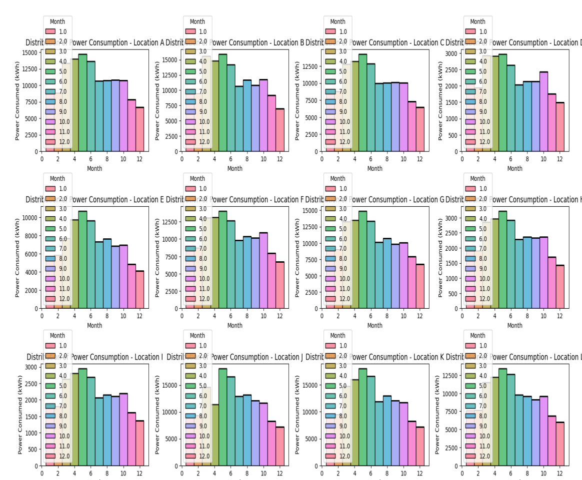


Fig. 10: Power Consumption vs Month

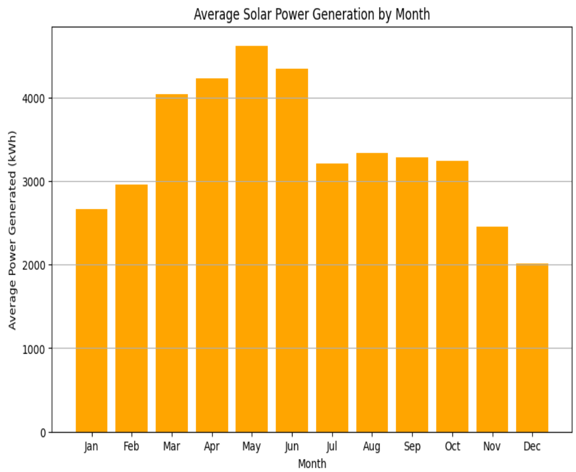


Fig. 11: Average Solar Power Generation by Month

III. Algorithms

Algorithm 1 Artificial Neural Network Algorithm

Input: Training dataset $D = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, learning rate α, number of epochs N
Initialize: Random weights $w_1, w_2, ..., w_m$
for $i = 1$ to N **do**
  for $j = 1$ to n **do**
    Calculate the output of the neural network: $o_j = f(\sum_{i=1}^{m} w_i \cdot x_{ji})$
    Calculate the error: $e_j = y_j - o_j$
    Update weights: $w_{i+1} = w_i + \alpha \cdot e_j \cdot x_{ji}$ for all i
  end for
end for
Output: Trained neural network with weights $w_1, w_2, ..., w_m$

A. Algorithm Description

The presented algorithm outlines the training procedure for an Artificial Neural Network (ANN) using a gradient descent approach. The algorithm takes as input a training dataset D consisting of input-output pairs (x_i, y_i), a learning rate α, and the number of training epochs N.

The algorithm randomly initializes the weights w_i and iterates for N epochs. For each epoch, it iterates over the training examples (x_j, y_j), calculates the output of the neural network o_j using the current weights, computes the error e_j as the difference between the actual output y_j and the calculated output o_j, and updates the weights using the gradient descent rule $w_{i+1} = w_i + \alpha \cdot e_j \cdot x_{ji}$.

After training, the algorithm outputs the trained neural network with the updated weights. This algorithm serves as a fundamental building block for training neural networks for various machine learning tasks.

IV. Discussion and Remarks

An alternative approach that we encountered while solving this problem of improving solar power forecasting accuracy was by bringing in use of more advanced ML models or ensemble methods. Deep learning models, such as LSTM networks, and ensemble methods, such as combining forecasts from multiple models or incorporating external data sources, this may improve prediction accuracy but was much complex. Furthermore, better integration of advanced data preprocessing techniques such as outlier detection and missing value handling could improve forecasting performance even more. Investigating the use of other types of data, such as satellite imagery or cloud cover data, may provide additional insights for more accurate forecasting.

V. Numerical Results

Index	Real Solar Power Produced	Predicted Solar Power
1	2,497.53	2,269.27
2	514.65	409.38
3	916.92	1,043.96
4	1,911.95	1,882.24
5	35.15	179.38
6	2,449.04	2,412.25
7	38.84	679.74
8	823.41	2,163.26
9	2,274.65	2,130.05
10	1,123.40	1,217.51
11	946.53	1,371.69

VI. Conclusion

Solar power forecasting plays a crucial role in maximizing energy generation from solar power plants. Machine learning techniques, especially Artificial Neural Networks (ANN), have shown promising results in improving forecasting accuracy. Factors such as temperature, Angle of incidence, Zenith and Azimuth are essential inputs for these models. The study highlights the importance of selecting the right input variables and model evaluation metrics, such as the Pearson Correlation Coefficient (PCC) and Hyperband optimization, for accurate forecasting. Overall, accurate solar power forecasting is vital for grid stability and effective energy management in the transition to renewable energy sources.

References

[1] D. O'Leary and J. Kubby, "Feature Selection and ANN Solar Power Prediction," *Journal of Renewable Energy*, vol. 2017, Article ID 2437387, pp. 1-7, 2017.

[2] Y. Lee and J. Lee, "Solar power prediction based on Artificial Neural Network guided by feature selection for Large-scale Solar Photovoltaic Plant," *Source of Publication*.

[3] R. Gupta, A. K. Yadav, S. Jha, and P. K. Pathak, "Time Series Forecasting of Solar Power Generation Using Facebook Prophet and XG Boost," in *2022 IEEE Delhi Section Conference (DELCON)*, New Delhi, India, 2022, pp. 1-5, doi: 10.1109/DELCON54057.2022.9752916.

[4] K. Anuradha, D. Erlapally, G. Karuna, V. Srilakshmi, and K. Adilakshmi, "Analysis Of Solar Power Generation Forecasting Using Machine Learning Techniques," *Source of Publication*.