Report for Optimizing Renewable Energy Management through Solar Power Forecasting

MSc Research Project
MSCDAD_C

Karthikram Chockalingam Swaminathan
x23127651@student.ncirl.ie

School of Computing
National College of Ireland

Supervisor: Abid Yaqoob (Abid.Yaqoob@ncirl.ie)

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Karthikram Chockalingam Swaminathan …………………………………………………………………………………………..
Student ID:	X23127651 …………………………………………………………………………………………..……
Programme:	MSCDAD_C ……………………………………………………………… **Year:** Sep(2023 – 2024) …………………………..
Module:	Research Project ……………………………………………………………………………………….………
Lecturer:	Abid Yaqoob …………………………………………………………………………………………..………
Submission Due Date:	16-09-2024 ……………………………………………………………………………………….………
Project Title:	Optimizing renewable Energy management - Solar power Forecasting ……………………………………………………………………………………….………
Word Count:	7497 ……………………………………… **Page Count:** 21 …………………………………….……….………

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: Karthikram CS ……………………………………………………………………………………………………

Date: 16-09-2024 ……………………………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
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.	□

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):	

Optimizing Renewable Energy Management through Solar Power Forecasting

Karthikram Chockalingam Swaminathan
x23127651@student.ncirl.ie

Abstract

With the shift towards renewable energy has increased exponentially, improving the grid stability, encouraging the transition navigating towards renewable energy sources and an efficiently predicting model for solar power are important for enhancing the management of renewable energy. This research study aims to use deep learning base approach specifically, Transformers for learning the complex patterns and trends in solar time series data. We also compared transformers based approach with various machine learning approaches like the Decision Tree Regression model, KNN (K-Nearest Neighbors) Regression model, Gradient Boosting Regres- sion model. Transformer model has shown exceptional performance compared to other models, where the values of R^2, MSE, and MAE are 0.99999, 0.97, and 0.56 respectively while other machine learning approaches also performed better. Transformer model able to make better predictions because it able to learn the long range dependencies of the solar data as it is time series data. The transformer model has shown significant enhancements in terms of identifying dependencies and complex patterns present in the data, this helps in improved predicting accuracy. This research study aims is understand and show the capabilities of enhanced deep learning and machine learning models for improving the predicting accuracy of solar power. The knowledge is acquired to contribute enhancements in grid integration and energy storage management. Future works can concentrate on working with real-time data to improve prediction accuracy and develop more integrated models. This study contributes to reliable and sustainable energy systems.

1 Introduction

Solar power is the most important renewable energy source because of its abundance in nature and sustainability. Globally renewable energy utilization has increased exponentially and there is a great requirement for efficient energy management systems. Predicting solar power accurately has included several attributes like historical data, panel characteristics, and weather conditions for the future output of solar panels. But, solar energy is not a continuous form of energy source, its variable nature has shown remarkable limitations for combining into an energy grid (Marquez et al.; 2020). A predicting model is required to mitigate the current limitations and enhance the prediction accuracy for solar power, the assistance of an efficient predicting model leads to advanced grid stability and mitigates the dependency on fossil fuels(Antonanzas et al.; 2016; Ghimire; 2018).

1.1 Motivation

This research goal is to mitigate the dependency on fossil fuels and navigate toward the utilization of renewable energies. One of the most promising renewable energy sources is solar power because of its abundance and sustainable nature. There is a continuous energy expansion and an efficient energy management model is required for predicting the solar power output and integrating solar power with current energy grids. According to IEA, the capacity of solar PV has enhanced by 22% in 2020 and expanded to 707GW in 2021.

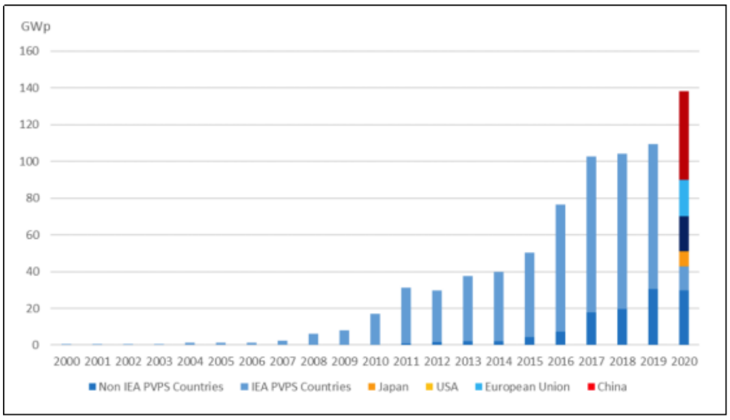

Figure 1: Evaluation of Annual PV Installations (Lappalainen et al.; 2017)

As solar power is not a continuous source of energy, it is important to predict solar power accurately, the main reason is weighing the supply and the demand in the market, if the supply is too low compared to the predicted output, there will be chances of power outages(Bacher et al.; 2009). enhanced solar power forecasting assists in economic planning and mitigates operational costs(Voyant et al.; 2017). With the better-predicting models, it will be easy to organize the energy storage system which is important for reducing the recurrence of solar power(Lappalainen et al.; 2017).

FOR ANNUAL INSTALLED CAPACITY			FOR CUMULATIVE CAPACITY		
1	China	48,2 GW	1	China	253,4 GW
(2)	*European Union*	19,6 GW	(2)	*European Union*	*151,3 GW*
2	United States	19,2 GW	2	United States	93,2 GW
3	Vietnam	11,1 GW	3	Japan	71,4 GW
4	Japan	8,2 GW	4	Germany	53,9 GW
5	Germany	4,9 GW	5	India	47,4 GW
6	India	4,4 GW	6	Italy	21,7 GW
7	Australia	4,1 GW	7	Australia	20,2 GW
8	Korea	4,1 GW	8	Vietnam	16,4 GW
9	Brazil	3,1 GW	9	Korea	15,9 GW
10	Netherlands	3 GW	10	UK	13,5 GW

Figure 2: Top 10 Countries for Installations and Total Installed Capacity in 2020(Marquez et al.; 2020)

Transformer models have recently been developed in machine learning to assist in recognizing dependencies and complex patterns, and advancements in deep learning have paved a new path for enhancing the reliability and accuracy of solar power predictions.

However, the classical statistical models can not identify the complex patterns and long-term dependencies present in the solar power data(Antonanzas et al.; 2016). Machine learning and deep learning models acknowledge large datasets and recognize dependencies and complex patterns which helps in improving the prediction accuracy.

1.2 Research Questions

This study aims to concentrate on advanced machine learning and deep learning tech- niques for developing a novel framework for accurately predicting solar power which assists in enhancing renewable energy management. This research navigates toward finding the answers to these important questions:

- To improve grid integration and renewable energy management, how predicting solar power techniques can be used?
- Exploring various machine learning models and identifying which model provides the highest accuracy for predicting solar power?
- Define the important attributes that affect the solar power output, and how these attributes can be managed efficiently for enhancing solar power output.
- Exploring and evaluating machine learning and deep learning models and comparing the results between the advanced models and classical models, to define the best model for accurately predicting the solar power?
- How efficiently these models will work on real-time data to enhance solar power prediction for energy storage management and grid stability?

1.3 Objectives

In this research, several machine learning and deep learning models are explored for accur- ately predicting solar power, few models are KNN, Random Forest, Transformer model, and Decision Trees. This research goal is to focus on developing a novel framework for enhancing prediction models for solar power and renewable energy management. By training and testing the machine learning models with the same dataset, we can identify the most efficient method for accurately predicting solar power.

The dataset is used to train the predicting models and the comprehensive dataset is used in this study, the attributes influencing solar power prediction are solar irradiation levels, ambient and module temperatures, and DC (direct current) and AC (alternating current) power. These different attributes assist in the complete analysis of features af- fecting solar power output.

The objectives of this research are:

- For predicting solar power different machine learning models are developed and evaluated.
- To recognize the model which provides the highest prediction accuracy and reliab- ility.

- To offer insights and guidance for enhancing the integration of solar power into the energy grid.

This research aim is to contribute globally to improving renewable energy management systems and mitigate the risk of grid stability. Accurately predicting solar power will help in balancing energy distribution and storage. The assistance of better-predicting models will offer more support towards reliable and sustainable energy systems.

1.4 Significance

The enhancements in machine learning and deep learning models have shown significant enhancements in predicting strategies and exploring their capability for accurately predicting solar power and enhancing renewable energy management. As solar power is not a continuous source of energy, predicting accurately will assist grid operators in efficiently managing the uncertainty as well as the variability of solar power, overall it helps in enhancing the sustainability and reliability of the energy grid. The main goal of this research is to find a novel prediction model that can accurately predict solar power and contribute to a reliable and sustainable use of renewable energy resources.

In conclusion, this study aims for a reliable energy solution globally and the advanced machine learning and deep learning models assist in accurately predicting solar power. With the help of these advanced methodologies, this study acquires enhanced predictions for solar power and offers valuable insights and real-time solutions for improving solar energy management, contributing towards a sustainable and greener feature.

2 Related Work

Solar power is one of the most reliable sources of renewable energy because of its abundance and sustainable nature. The dependency on solar power is increasing exponentially and installing solar PV (photovoltaic) has increased globally. This research aims to enhance the integration of solar power into energy grids, and enhancing solar power predictions has evolved remarkably, with the integration of advanced machine learning models and the classical models. The literature review aims to offer extensive knowledge about solar power management, the current situation of predicting solar power, advanced machine learning algorithms influencing the enhancements in predicting solar power, challenges posed by advanced methodologies, and the scope for future work.

2.1 Traditional Methods for Solar Power Forecasting

Traditional models used for predicting solar power mostly depend on historical data- sets, these classical methodologies include statistical as well as physical models. ARIMA and linear regression models are used in predicting solar power, autoregressive integrated moving averages model is simple and can be used for accurately predicting solar energy, but these models are good at predicting short-term dependencies(Ghimire; 2018) . The performance of ARIMA and linear regression models is good at capturing short-term dependencies, they are not capable of identifying complicated relations among the attrib- utes and non-linear data (Bacher et al.; 2009) . When it comes to physical models, they utilize meteorological data, these models depend on the solar radiation, temperature, and color cover data for predicting the solar power (Lorenz et al.; 2011).

2.2 Machine Learning Approaches for Solar Power Forecasting

The traditional approaches are capable of predicting solar power for short range like on a fixed data, to overcome these challenges machine learning techniques are used for temporal. Advanced machine-learning models can learn from large datasets to undertand and learn non-linear relationships present within in temporal solar power data (Voyant et al.; 2017). SVM (Support Vector Machines) can effectively predict solar power data, but these models have to be trained rigorously and the attributes have to be fine-tuned for predicting the solar power accurately. SVM models can handle non-linear data effectively and allow the model to manage the relation between the input attributes as well as the solar power output (Chen et al.; 2011). KNN (K-Nearest Neighbors) model depends on the output results of the nearest neighbors, it is an efficient model for pre- dicting solar power, but the prediction accuracy depends on the number of neighbors in the feature space and this can influence the dimensionality (Voyant et al.; 2017). ANN (neural Networks) models are known for identifying non-linear relationships and complex patterns available in the dataset, but training the ANN model is more resource-intensive and time-consuming. ANN models are more advanced in comparison to traditional statistical models (Yadav and Chandel; 2014). GBM (Gradient Boosting Machines) is an ensemble model, that integrates various weak learners and the average result of these models is considered output, this model has shown significant enhancements for accurately predicting solar power, identifying complex patterns, and managing large datasets. However, these models are expensive and need a delicate tuning process (Huang et al.; 2013). Random Forests, integrates multiple decision trees, and these trees are assigned to predict the solar power accurately. Decision trees are easy to elucidate, the model's prediction depends on the attribute values and the forecasting depends on the leaf node's average output, but these models consider multiple feature values which creates the limitations of overfitting with large and complex data. To overcome the limitations posed by decision trees, the Random Forest model uses the average output of multiple decision trees, which improves the model's accuracy (Al-Fetyani and Tawalbeh; 2017).

2.3 Deep Learning Approaches for Solar Power Forecasting

Deep learning models are known for accurately predicting the results, they are capable of automatically acknowledging complex patterns and long-term dependencies present in raw data. Deep learning models shown significant results in various fields, with these trends even solar power for predicting is being done accurately, and these models can im- prove the efficiency in identifying complex feature patterns (Marquez et al.; 2020). Trans- formers are mainly utilized in the NLP field with significant improvement of performance on time series data. With these advantages transformers are utilized for predicting solar power accurately,because of its long term dependencies on the historical data as well as it is time series data. Further, hybrid models are used to enhance prediction accuracy in various fields, these models integrate various machine-learning techniques. For instance, by combining both classical and machine learning models, we can use the advantage of identifying complex patterns, and short-term and long-term dependencies, resulting in enhancing the accuracy of the model's prediction and robustness. Overall deep learning models outperform machine learning and traditional models (Nespoli et al.; 2019).

Table 1: Comparison between Traditional and Modern Methods for Solar Power Forecasting

Feature	**Traditional Methods**	**Modern Methods (Machine Learning & Deep Learning)**
Approach	Statistical and physical models	Machine learning and deep learning models
Examples	ARIMA, Linear Regression, Physical Models	SVM, ANN, GBM, KNN, Decision Trees, Random Forests, LSTM, Transformer
Data Requirements	Historical time series data, meteorological data	Large datasets including historical, meteorological, and additional features
Accuracy	Reasonable for short-term, less accurate for long-term	High accuracy for both short-term and long-term forecasts
Complexity	Lower, simpler to implement	Higher, complex algorithms and model structures
Computational Cost	Low to moderate	High, especially for deep learning models
Ability to Capture Non-linearity	Limited	High
Scalability	Moderate	High
Interpretability	High	Often lower, especially with deep learning models
Adaptability	Limited	High, can adapt to new patterns and data

2.4 Comparative Analysis

As solar energy is the most important source of renewable energy, it is important to predict solar power and improve the storage management capabilities. To acknowledge the predicting techniques of the current state, it is important to evaluate key research studies:

Table 2: Summary of Studies on Forecasting Methods and Their Performance

Study	Method	Results	R^2 Score
(Huang et al.; 2013)	GBM	Excellent performance, handles large data-sets well	0.85
(Voyant et al.; 2017)	KNN, Ensemble Methods	KNN provides reasonable accuracy, ensemble methods outperform individual models	0.85
(Al-Fetyani and Tawal-beh; 2017)	Decision Trees, Random Forests	Random Forests mitigate overfitting, better accuracy than Decision Trees	0.88
(Marquez et al.; 2020)	LSTM	Superior performance in capturing temporal dependencies	0.92
(Lim et al.; 2021)	Transformer Models	State-of-the-art performance, surpassing RNNs and LSTMs	0.97
(Nespoli et al.; 2019)	Hybrid Models	Enhanced accuracy by integrating multiple methods	0.93

The above table explains the important research studies that used different predicting models with various technologies like traditional models, machine learning, and deep learning models and includes the results of every prediction model, their significance in predicting solar power, and the areas where further enhancement is required.

2.5 Challenges and Future Directions

Several enhancements were made to predict solar power accurately, still, there are many limitations because solar power is not a continuous source of energy, there will be changes depending on the cloud coverage, and weather conditions, and there is a requirement for great quality meteorological data, one of the major challenges is the computational complexity of enhanced machine learning and deep learning models. Further, this process requires a novel framework for combining the predictions into grid management systems, and real-time data is needed for processing. The future research scope is to concentrate on enhancing the efficiency of machine learning models and exploring hybrid approaches. There is a great possibility for enhancing the edge computing devices and integrating the satellite and sensor data will offer more accurate predictions of solar power (Antonanzas et al.; 2016).

The literature review explains various models for predicting solar power from basic traditional models to advanced deep learning models and their enhancements for pre- dicting solar power. Traditional models can capture short-term dependencies and require large amounts of meteorological data for predicting accurately, machine learning models showed some promising results compared to traditional models, and transformers have evolved and are known for identifying long-term dependencies and complex patterns avail- able in the data. As solar energy is evolving as the most important source of renewable energy, there is a need to develop a novel approach for enhancing energy management and grid stability.

3 Methodology

Solar energy is not a continuous form of energy but it is abundant and sustainable, the methodology for this study mainly concentrates on a novel approach for predicting solar power accurately through Transformer based models is developed and evaluated on various factors influencing solar power prediction. This methodology outlines the processes involved in collecting data, preprocessing, selecting the models for predicting solar power, implementing, and comparison, this section offers an extensive approach to the objectives.

For this research study, the dataset is collected and can be accessed from Kaggle, the same dataset is used for preprocessing, training, and testing various machine learning and deep learning models. The solar power dataset offers complete data on solar power gener- ation, which is very important for developing models for predicting solar power accurately.

Characteristics of the Dataset:

Table 3: Feature Characteristics of Dataset1

Feature Name	**Description**	**Data Type**
S.NO	Unique identifier for each data entry	Int64
SOURCE_KEY	Identifier for the specific solar panel or source	Object
DC_POWER	Power output from the solar panel in direct current (DC), measured in watts	Float64
AC_POWER	Power output converted to alternating current (AC), measured in watts	Float64
DAILY_YIELD	Total energy produced that day, measured in kilowatt-hours (kWh)	Float64
TOTAL_YIELD	Total energy produced by the solar panel since its installation, measured in kilowatt-hours (kWh)	Float64
DATE_TIME	Timestamp recording the date and time of the data entry	Object
AMBIENT_TEMP	Temperature surrounding the solar panel, meas- ured in degrees Celsius	Float64
MODULE_TEMP	Temperature of the solar panel's module, meas- ured in degrees Celsius	Float64
IRRADIATION	Solar radiation falling on the panel's surface, meas- ured in watts per square meter (W/m^2)	Float64
DATE	Date of the data entry	Object
TIME	Time of the data entry	Object
DAY	Day of the month (ranging from 1 to 31)	Int64
MONTH	Month of the year (ranging from 1 to 12)	Int64
WEEK	Week number of the year	Int64
HOURS	Hour of the day (ranging from 0 to 23)	Int64
MINUTES	Minute of the hour (ranging from 0 to 59)	Int64
TOTAL MINUTES	Total elapsed minutes since data collection began	Int64
DATE_STRING	Date represented as a string	Object
SOURCE_NUMBER	Numerical version of the source key	Int64

The data set collected from Kaggle is solar power generation data, to enhance this dataset, we integrated it with other datasets with relevant attributes influencing the solar power prediction and more robust analysis. The merged dataset will be extensive and this process includes standardizing attributes to guarantee consistency across the combined dataset and aligning the dataset depending on time stamps. By combining more datasets, we acquired a high-quality dataset, the main aim of this research is to improve storage management, develop a novel approach for predicting solar power accurately, and concentrate on enhancing renewable energy resources.

3.1 Data Preprocessing

This is one of the most important steps in methodology because this step guarantees the quality and sustainability of the dataset which is further used for training the models (Garc´ıa et al.; 2016). The steps involved in preprocessing are:

- Data Cleaning: Checking for inconsistencies present in the dataset, if any duplicate values are present, in this stage duplicate data is removed, and missing values are handled.
- Feature Engineering: New features are created depending on the current models to provide more information regarding the date and time components like minutes, hours, weeks, and months.
- Normalization: This step guarantees that the model training process is not based on the feature scales, so the standard deviation value is set to 1 and the mean is 0.
- Splitting Data: Data used for Training will be 80% where as 20% used for testing.

3.2 Baselines

Accurately predicting solar power, various ML model have been trained on solar data as baselines to compare with transformer based model.

3.2.1 Gradient Boosting Regressor (GBM)

GBM is an ensemble model (Huang et al.; 2013), that integrates various weak learners and the average result of these models is considered output, this model has shown significant enhancements for accurately predicting solar power, identifying complex patterns, and managing large datasets.

The core idea is to sequentially train a series of models such that each new model corrects errors made by its old learners as shown in figure 3. This is achieved through an iterative process where each subsequent tree is trained to predict the residuals or errors of the combined predictions of all previous trees. The model's predictions are keep on updated by adding these new trees' predictions until the model gets better performance.

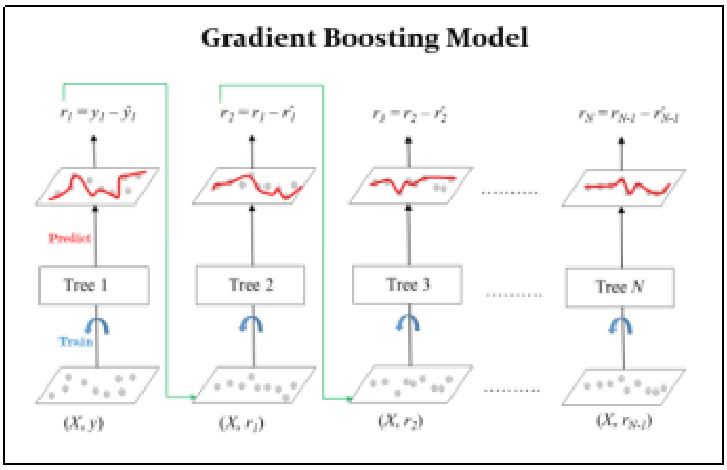

Figure 3: Architecture of Gradient Boosting Model (Natekin and Knoll; 2013)

3.2.2 K-Nearest Neighbors (KNN) Regressor

KNN is a Non -parametric algorithm used for regression tasks (Voyant et al.; 2017). this method uses the concept of nearest neighbour to classify the points as shown in figure 4. To determine the nearest neighbors, KNN uses various distance metrics like euclidean or manhattan distance to measure closeness between any two data points. Once all nearest neighbors for a specific point selected, their target values are averaged or mean to make the final prediction. But limitations of KNN it is very costly in terms of computational and modelling resources and often suffer from curse of dimensionality.

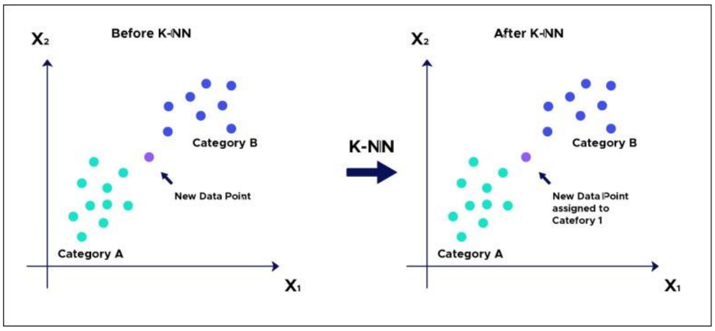

Figure 4: Architecture of K-Nearest Neighbors Model (Guo et al.; 2003)

3.2.3 Decision Tree Regressor (DT)

DT can be used for regression as well as classification (Al-Fetyani and Tawalbeh; 2017). This approach used for regression tasks by splitting the data into very small subsets based on the attributes shown in figure 5. It consists of tree structure with nodes are known for conditions for split and branches that lead to subsequent nodes or leaf nodes through recursive partitioning. In the last year of tree, leaf nodes contains a prediction by the averaging of the target values of the data points that fall into that leaf. Algorithm selects the feature and threshold that best separates using entropy or Gini index based approaches.

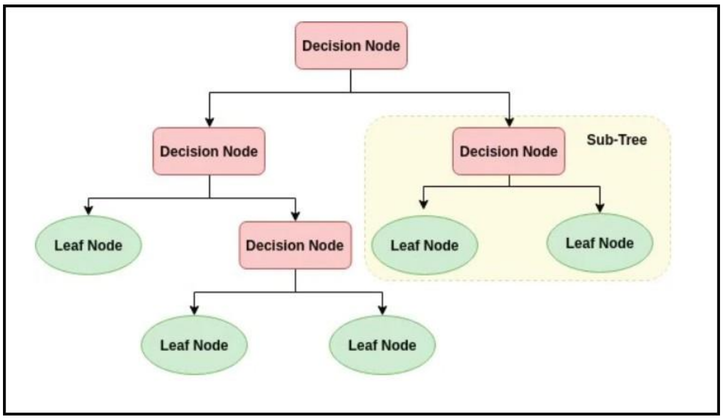

Figure 5: Architecture of Decision Tree Model(Ali et al.; 2012)

3.2.4 Random Forest Regressor (RF)

RF is an ensemble learning method that uses the concept of averaging over multiple decision trees (Al-Fetyani and Tawalbeh; 2017). It builds multiple DT's such that all individual trees are trained on random subset of features. Each DT tree is expanded or splitted full depth, with decisions made at each node based on a random subset of features rather than the entire feature set as shown in figure 6. This degree of entropy in the individual trees helps to ensure that the trees are diverse to avoid overfitting c. The final prediction is made by averaging the predictions from all the individual trees in the forest. Random Forests are known for their robustness and accuracy, and they generally require less tuning than other complex models, making them a popular choice for many regression tasks.

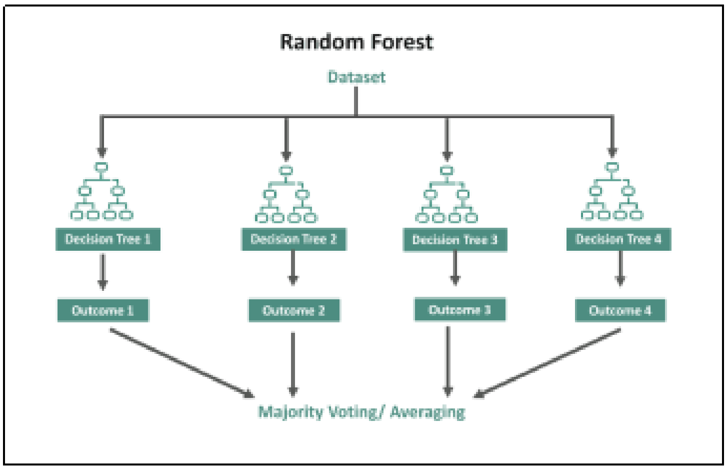

Figure 6: Architecture of Random Forest Model (Ali et al.; 2012)

3.3 Transformer Model

Transformers (Marquez et al.; 2020) are mainly utilized in the natural language processing field and have shown significant results, so they are used in several time series prediction models. Now transformers are utilized for predicting solar power accurately, because of their capability to identify complex patterns and long-term dependencies in the dataset. Mainly transformers utilize a self-attention mechanism for identifying long-term dependencies in the data despite their distance. The mechanism used in traditional models like Recurrent Neural Networks is sequential whereas in transformers the input data is processed simultaneously, this remarkably improves the predicting accuracy by capturing complex patterns and dependencies. For predicting the time series data, the transformer model is very helpful for learning and understanding the patterns and long-term dependencies. The transformer models are known for training and learning from large volumes of data because of their processing abilities on temporal data.

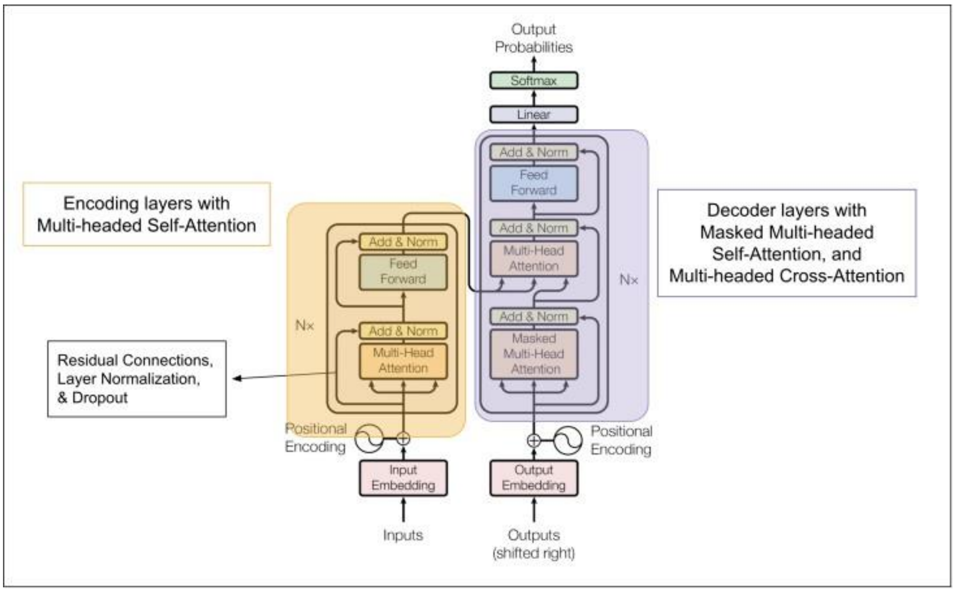

Figure 7: Architecture of Transformer Model (Dosovitskiy et al.; 2020)

This methodology highlights the enhancements of different traditional, machine learning, and advanced deep learning models for accurately predicting solar power. With the help of advanced machine learning and deep learning models as well as date preprocessing, this research goal is to enhance the grid integration and effectively predict solar power. The valuable insights from this methodology help in real-time implications for enhancing the sustainability, efficiency, and reliability of solar power systems, finally contributing towards globally sustainable and greener renewable energy sources.

4 Design Specification

This study on predicting solar power for enhancing the management of renewable energy framework is a two-tiered architecture: divided into the Business logic tier and the Presentation tier. This design guarantees reliable data processing, various models ro- bust training and evaluation, and overall an effective model for efficiently predicting solar power for real-time applications.

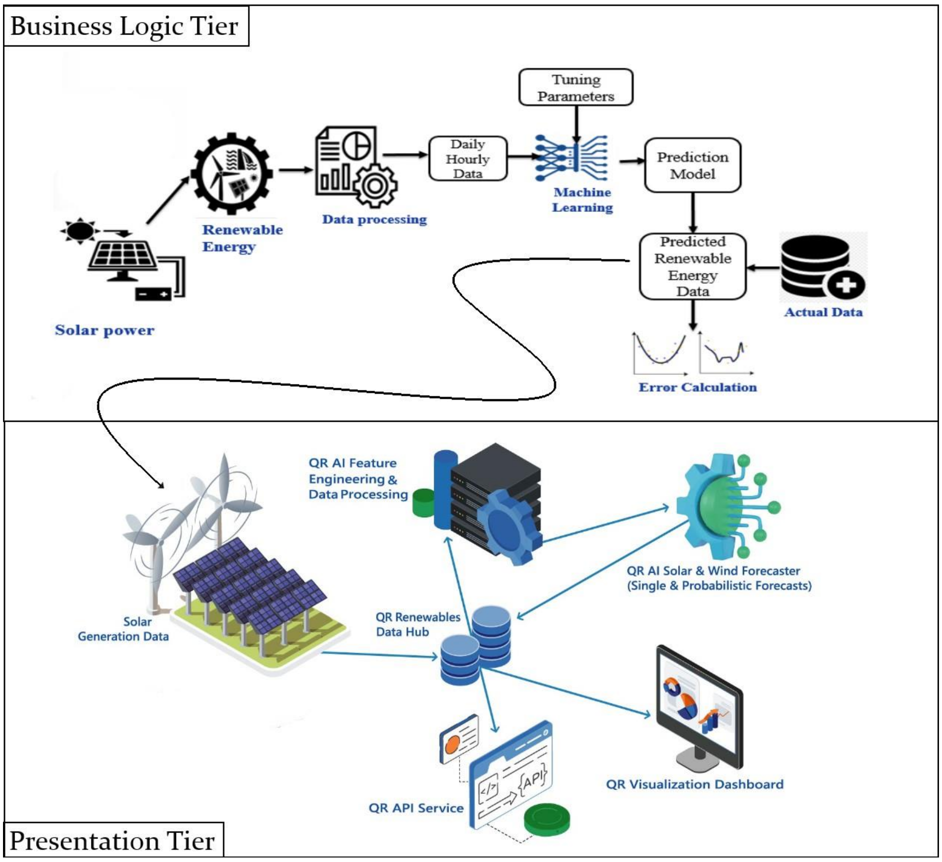

Figure 8: Design Specification for Solar Power Forecasting

The framework is divided into the two primary tiers, The first is the Business logic tier which accounts for different machine learning models' predictions depending on their previous solar power data. Most of the important tasks are implemented in this tier, like collecting datasets, merging multiple datasets to form a comprehensive data- set, data preprocessing, and feature engineering. Normalization of data guarantees that the model training process is not biased and the dataset is split into training and test- ing sets. The models developed and trained in this business logic tier are GBR, KNN, Random Forest Regressor, Decision Tree Regression model, and transformer model. The trained models are evaluated with performance metrics like R^2 (R-squared), MAE (Mean Absolute Error), and MSE (Mean Squared Error), these metrics are used to define the model's strengths and limitations. To understand data distribution and the analysis of the model's performance different visualization techniques are assigned. On the other hand, the Presentation tier concentrates on the enhancements in the predictions and per- formance metrics with their applications in real-world problems. It guarantees that the predicted output and the performance metrics are communicated effectively. The pre- dicted output is viewed along with the expected output in the form of graphs and charts, it will be easy to understand the difference between the actual and predicted values. The model's performance can be outlined, the main aim is accurately predicting solar power which helps in enhancing resource management and assists in great decision-making for energy demand, distribution, consumption, and operational efficiency. This tier utilizes the advantages of advanced machine learning and deep learning models to predict solar power accurately, this will navigate toward sustainable goals and mitigate the dependency on fossil fuels. This framework helps in predicting solar power efficiently and the advanced machine learning models contribute towards sustainable and efficient energy infrastructure.

5 Implementation

This implementation process begins by collecting and merging solar power datasets, this extensive dataset helps the models for predicting solar power accurately, Datasets are merged because of the relevant attributes present in different datasets and these attributes influence the solar power prediction. There are several attributes in the dataset, but for this research, few attributes are considered and they are as follows: daily and total energy yields (DAILY YIELD, TOTAL YIELD), ambient and module temperatures (AMBIENT TEMPERATURE, MODULE TEMPERATURE), solar irradiation levels (IRRADIATION), direct current power (DC POWER), and alternating current power (AC POWER). These are a few important factors for predicting solar power ac- curately and help in recognizing the complex patterns and dependencies present in the dataset. To understand the data, the dataset is loaded into pandas DataFrame, evaluat- ing its structure as well as acknowledging the relation between various attributes. Once this step is clear with the dataset, then the successive preprocessing activities can be guided.

To make model has to be effective, then the two important steps are data cleaning and data preprocessing. When the data is collected, the raw data will have inconsistencies, all these inconsistencies, missing data, and duplicate data have to be removed to guarantee the integrity of the model. In the feature engineering step, new features are derived from the current feature set. For example, from the solar power generation dataset, features like time stamp and date are extracted and these features help identify complex patterns and dependencies in the dataset, finally improving the accuracy of prediction models. All the features undergo normalisation to make values between [0,1] for strandadising. This Step is very crucial for training because without normalising some features will get high priority with high values leads to inappropriate predictions.s

Evaluating the models was a critical step to assess their accuracy and reliability in pre- dicting solar power output. Performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) were used to provide a comprehensive assessment of each model's performance. These metrics allowed for a detailed compar- ison of the models, highlighting their strengths and weaknesses. The evaluation involved comparing the predicted values from the models with the actual values from the testing dataset. This comparison provided valuable insights into how well each model performed in real-world scenarios, guiding the selection of the best-performing model based on its accuracy and reliability.

Overall, This research study implementation has included a few steps starting from collecting the dataset, cleaning the data is one of the most important parts for predicting models, the inconsistent and duplicate values have to be removed from the dataset, and then the dataset is split into training and testing dataset, in this phase several machine learning and deep learning models are trained and tested, the evaluated results explain

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}|$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y})^2$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y})^2}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

Where,
$\hat{y}$ $-$ *predicted value of y*
$\bar{y}$ $-$ *mean value of y*

Figure 9: Various Regression metrics used for effectiveness models Marquez et al. (2020)

the model's advantages and limitations for performing the accuracy of predicting solar power. The outcomes of this research assist in optimizing solar energy management and increasing grid stability and integration of solar power into energy grids. When compared to the other advanced machine learning and deep learning models, transformers have shown promising results because of their capability to capture long-term dependencies and complex patterns in the dataset. Addressing all the limitations currently present in predicting solar power improves and creates a way for more reliable and accurate solar power forecasting.

6 Evaluation of Implementation Results

The evaluation of implementation results for this research study focused on assessing the performance of various models in accurately forecasting solar power output can be done using various metrics shown in Figure 9

6.1 Gradient Boosting Regressor

GBR is an ensemble model that integrates the different weak learners and the average results of these models are considered as output, In this section the actual values from the solar power generation dataset are compared to GBR predicted values, the errors, and the percentage errors are shown in the following table:

There are evaluation metrics that define the prediction capabilities of the models, GBR has achieved an MSE, MAE, and R^2 values of 114.62, 8.54, and 0.9992 respectively. These outcomes indicate the great quality of accuracy and explain that the model has effectively identified the complex patterns and dependencies in the dataset, but a few points showed more percentage error and suggest the requirement for data preprocessing.

Table 4: Comparison of Actual and Predicted Values with Error and Percentage Error

Actual	Predicted	Error	Percentage Error
845.71	835.43	10.29	1.22
190.29	201.09	-10.80	5.68
389.30	393.80	-4.50	1.16
26.60	42.90	-16.30	61.26
466.20	471.21	-5.01	1.07

6.2 K-Nearest Neighbors Regressor

KNN regression model, has shown some great results for accurate prediction of the model, but they mostly face challenges with overfitting the model, and the following is the evaluation table of the RNN model:

Table 5: Comparison of Actual and Predicted Values with Error and Percentage Error

Actual	Predicted	Error	Percentage Error
845.71	797.55	48.16	5.70
190.29	193.43	-3.14	1.65
389.30	378.14	11.16	2.87
26.60	22.93	3.67	13.79
466.20	464.52	1.68	0.36

The KNN model values of MAE, MSE, and R^2 are 16.78, 594.54, and 0.9956 respectively. When compared to other machine learning models the acquired accuracy is relatively low, they have shown some great errors specifically for certain data points. This outlines that the number of neighbors decides the prediction accuracy, and the challenges for identifying complex patterns increase if there are more neighbors present.

6.3 Decision Tree Regressor

The following table demonstrates the performance of Decision Tree Regression model: Table

6: Comparison of Actual and Predicted Values with Error and Percentage Error

Actual	Predicted	Error	Percentage Error
845.71	847.04	-1.33	0.16
190.29	197.60	-7.32	3.84
389.30	394.50	-5.20	1.34
26.60	14.41	12.19	45.82
466.20	466.32	-0.12	0.03

The evaluation parameter values of the decision tree regression model have achieved MSE, MAE, and R^2 values of 41.17, 5.35, and 0.9997 respectively. The results of the decision tree model explain that the dataset is divided very efficiently depending on the feature values. Few data points have deviated completely from the actual values, which suggests that there is a need for enhancements in that particular area.

6.4 Random Forest Regressor

The Random Forest Regression Model works by integrating multiple decision trees and their average output is considered as the result. The evaluation performance of RF regression model is shown in the following table:

Table 7: Comparison of Actual and Predicted Values with Error and Percentage Error

Actual	Predicted	Error	Percentage Error
845.71	846.92	-1.20	0.14
190.29	197.27	-6.98	3.67
389.30	393.06	-3.76	0.97
26.60	25.20	1.40	5.27
466.20	467.04	-0.84	0.18

From the above table, the performance of the RF regression model is achieved as the MSE, MAE, and R^2 values of 14.60, 3.01, and 0.9999 respectively. As the Random Forest Regression model utilizes the multiple decision trees, they have overcome the limitation of overfitting and enhanced its performance for accurately predicting solar power.

6.5 Transformer Model

The transformer models are known for capturing complex patterns and dependencies in natural language and they have shown similar results in various fields, here the trans- former model's performance is shown in the following table:

Table 8: Comparison of Actual and Predicted Values with Error and Percentage Error

Actual	Predicted	Error	Percentage Error
845.71	845.84	-0.12	0.01
190.29	190.08	0.21	0.11
389.30	389.19	0.11	0.03
26.60	26.72	-0.12	0.47
466.20	465.83	0.37	0.08

The transformer model values of MSE, MAE, and R^2 are 0.97, 0.56, and 0.99999 respectively. When compared to other machine learning and deep learning models, transformer models have shown exceptional results by identifying the complex patterns and long-term dependencies present in the huge volumes of the dataset.

6.6 Summary of Model Performance

The table below summarizes the performance metrics for all the models:

The evaluation results shows the effectiveness of transformer model over machine learning models shows transformer Models superiority in capturing the patterns within the data. The Random Forest Regressor and Decision Tree Regressor also showed high accuracy but were slightly less precise than the Transformer Model. The Gradient Boost- ing Regressor and KNN Regressor, while still effective, exhibited higher errors compared to the other models.

Table 9: Performance Metrics for Different Models

Model	MSE	MAE	R^2
Gradient Boosting Regressor	114.62	8.54	0.9992
K-Nearest Neighbors Regressor	594.54	16.78	0.9956
Decision Tree Regressor	41.17	5.35	0.9997
Random Forest Regressor	14.60	3.01	0.9999
Transformer Model (Deep Learning)	0.97	0.56	0.99999

Performance metrics are used to outline the enhancements made by various machine learning and deep learning models, their accurate predicting capability of solar power, integrating solar power into energy grids, and enhancing solar power management systems. Out of all the other predicting models transformers have shown exceptional results for predicting solar power accurately, by identifying the complex patterns and dependencies. From the knowledge acquired these models contribute to more robust and sustainable energy management systems, further hybrid models can be explored and the predicting model's accuracy can be enhanced.

7 Discussion of Results

Various machine learning and deep learning models are trained and tested for their performance in accurately predicting solar power and for enhancing renewable energy management. The various machine learning algorithms are evaluated for their performance, and the models evaluated are GBM (Gradient Boosting Regressor), KNN (K-Nearest Neighbors) Regression model, Random Forest Regressor, Decision Tree Regressor, and Transformer model, the prediction accuracy and reliability of these models are shown below.

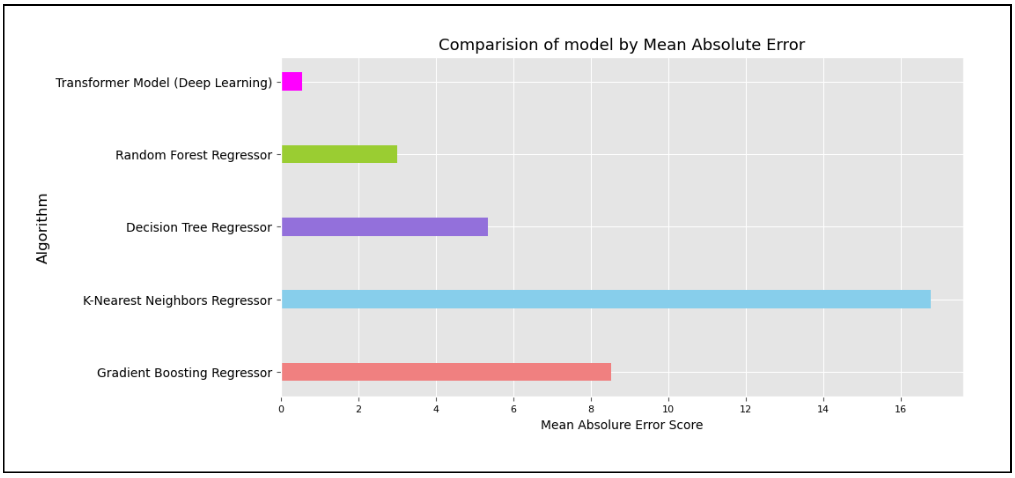

Figure 10: Performance Comparision of model by Mean Absolute Error

From results reported in figure 10 we can observe that transformer model able to give minimal error compared to any other machine learning model because Transformer models are known for capturing complex dependencies and complex in the dataset, they have shown great results in natural languages and also in various other fields. Now in this research, we explored the capabilities of Transformer models and they have shown significant results with the lowest MSE, MAE, and R^2 values of 0.97, 0.56, and 0.99999. These models are capable of handling large volumes of data effectively, and this makes them well suitable for improving solar power management and energy grid integration.

The Random Forest Regression Model works by integrating multiple decision trees and their average output is considered the result. the performance of the Random Forest regression model is achieved as the MSE, MAE, and R^2 values of 14.60, 3.01, and 0.9999 respectively. As the Random Forest Regression model utilizes the multiple decision trees, they have overcome the limitation of overfitting and enhanced its performance for accur- ately predicting solar power. The performance of the Random Forest Regressor is slightly low compared to the transformer model, but these models can still be used as ensemble models for predicting with great accuracy.

The evaluation parameter values of the decision tree regression model have achieved MSE, MAE, and R^2 values of 41.17, 5.35, and 0.9997 respectively. The results of the decision tree model explain that the dataset is divided very efficiently depending on the feature values. Few data points have deviated completely from the actual values, which suggests that there is a need for enhancements in that particular area.

GBR is an ensemble model that integrates the different weak learners and the aver- age results of these models are considered as output. There are evaluation metrics that define the prediction capabilities of the models, GBR has achieved an MSE, MAE, and R^2 values of 114.62, 8.54, and 0.9992 respectively. These outcomes indicate the great quality of accuracy and explain that the model has effectively identified the complex pat- terns and dependencies in the dataset, but a few points showed more percentage error and suggest the requirement for enhancement in data preprocessing.

The KNN regression model has shown some great results for accurate prediction of the model, but it mostly faces challenges with overfitting the model. The KNN model values of MAE, MSE, and R^2 are 16.78, 594.54, and 0.9956 respectively. When compared to other machine learning models the acquired accuracy is relatively low, they have shown some great errors specifically for certain data points. This outlines that the number of neighbors decides the prediction accuracy and the challenges for identifying complex pat- terns.

Finally, the results of various machine learning and deep learning models are highlighted, especially advanced deep learning models like Transformer models have shown excep- tional accuracy in predicting solar power. These methodologies promise an efficient and reliable management system for solar energy, integration of solar power into energy grids, and contributing towards a sustainable and reliable energy system.

8 Conclusion

This research study aimed to optimize renewable energy management through accur- ate solar power forecasting by using deep learning base model specifically using trans- former.We have also compared with various baselines models Gradient Boosting Re- gressor, K-Nearest Neighbors (KNN) Regressor, Decision Tree Regressor, Random Forest Regressor, and Transformer Model. The Transformer Model emerged as the most accur- ate and reliable model, demonstrating exceptional performance with minimal errors and an R^2 value of 0.99999. Its ability to capture complex patterns and dependencies within the data underscores its potential for high-accuracy forecasting tasks. The Random Forest Regressor and Decision Tree Regressor also showed strong performance, while the Gradi- ent Boosting Regressor and KNN Regressor exhibited higher errors and lower accuracy. These findings highlight the potential of advanced machine learning and deep learning techniques in optimizing renewable energy management. Accurate solar power forecast- ing can significantly enhance grid integration, improve energy storage management, and support economic planning. The insights gained from this research are expected to con- tribute to the development of more reliable and efficient solar power systems, ultimately supporting the global transition towards a sustainable and resilient energy system.

The future scope is to concentrate on exploring hybrid models that integrate the ad- vantages of different algorithms, fine-tuning the existing models, and with the help of real-time data, the prediction accuracy of the models can be enhanced. Further, by improving the interpretability of deep learning models, it can be applied to real-time scenarios and there will be chances for enhancing the accuracy of predicting models. Overall, by solar power prediction enhancements, we can contribute towards achieving sustainability goals.

References

Agency, I. E. (2021). Renewables 2021: Analysis and forecast to 2026.
URL: *https://www.iea.org/reports/renewables-2021*

Al-Fetyani, S. and Tawalbeh, L. (2017). Comparative study of different regression techniques for solar power forecasting, *Renewable Energy* **123**: 284–292.

Ali, J., Khan, R., Ahmad, N. and Maqsood, I. (2012). Random forests and decision trees, *International Journal of Computer Science Issues (IJCSI)* **9**(5): 272.

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., de Pison, F. M. and Antonanzas- Torres, F. (2016). Review of photovoltaic power forecasting, *Solar Energy* **136**: 78–111.

Bacher, P., Madsen, H. and Nielsen, H. A. (2009). Online short-term solar power forecasting, *Solar Energy* **83**(10): 1772–1783.

Chen, C., Duan, S., Cai, T. and Liu, B. (2011). Online 24-h solar power forecast- ing based on weather type classification using artificial neural network, *Solar Energy* **85**(11): 2856–2870.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale, *arXiv preprint arXiv:2010.11929* .

Garc´ıa, S., Ram´ırez-Gallego, S., Luengo, J., Ben´ıtez, J. M. and Herrera, F. (2016). Big data preprocessing: methods and prospects, *Big data analytics* **1**: 1–22.

Ghimire, S. (2018). Solar power forecasting using arima model, *Journal of Engineering Research and Development* **2**(3): 45–51.

Guo, G., Wang, H., Bell, D., Bi, Y. and Greer, K. (2003). Knn model-based approach in classification, *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, Springer, pp. 986–996.

Huang, C. J., Kuo, P. H. and Shih, J. Y. (2013). A comparison of prediction accuracy of three artificial intelligence methods applied to hourly solar radiation data in taiwan, *Solar Energy* **93**: 11–19.

Lappalainen, K., Repo, S. and Järventausta, P. (2017). Solar power forecast uncertainty and energy storage control, *Electric Power Systems Research* **146**: 29–39.

Lim, B., Zohren, S. and Roberts, S. (2021). Time-series forecasting with deep learning: A survey, *Philosophical Transactions of the Royal Society A* **379**(2194): 20200209.

Lorenz, E., Schel, H. and Heinemann, D. (2011). Short term forecasting of solar radiation based on satellite data, *Energy Procedia* **9**: 199–206.

Marquez, R., Coimbra, C. F., Kleissl, J. and Navarro, A. (2020). Neural network solar forecasting: Time series analysis, *Renewable Energy* **146**: 1781–1788.

Natekin, A. and Knoll, A. (2013). Gradient boosting machines, a tutorial, *Frontiers in neurorobotics* **7**: 21.

Nespoli, L., De Paris, A., Grimaccia, F., Leva, S. and Mussetta, M. (2019). An innovative hybrid forecasting system for pv power generation, *Solar Energy* **174**: 886–896.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M. L., Paoli, C., Motte, F. and Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review, *Renewable Energy* **105**: 569–582.

Yadav, A. and Chandel, S. S. (2014). Solar radiation prediction using artificial neural network techniques: A review, *Renewable and Sustainable Energy Reviews* **33**: 772–781.