Prediction and Monitoring of Solar Radiation Using Artificial Neural Networks for Renewable Energy Optimization

Faisal Alajmi[1]

Eng.[1]

[1]Specialist Trainer in Public Authority for Applied Education and Training, University of Sunderland, United Kingdom

Publication Date: 2025/11/04

Abstract: Accurate prediction of solar radiation is essential in the maximization of the output and planning of the renewable energy systems. The current paper proposes and tests the experimental daily global horizontal irradiance (GHI) forecasting of an Artificial Neural Network (ANN) model with a number of meteorological variables taken as input variables in the NASA POWER database. For the input features, ANN was trained with five input parameters: air temperature, relative humidity, wind speed, surface pressure, thermal range, and it utilized architecture with two hidden layers (128-64 neurons). Mean Absolute Error (MAE) calculation, Root Mean Square Error (RMSE) calculation and the coefficient of determination (R2) were used to assess the performance of models. Predictive capacity was high as indicated by a low MAE of 0.754 MJ/ m2/day, RMSE of 0.943 MJ/ m2/day, and R2 of 0.725 that interprets data to mean the model explains about 73% of GHI variation. Model stability and aids of monthly boxplots, visual diagnostics, and residual analysis were all in agreement in terms of the accuracy and stability of the model. The method is mathematically lean, interpretable, and is appropriate in data-scarce environments. This ANN model provided a viable and scalable solution to solar energy forecasting and helps in making decisions on planning and integrating photovoltaic systems into the grid.

Keywords: *Solar Radiation Forecasting; Artificial Neural Networks (ANN); NASA POWER Data; Renewable Energy; Meteorological Variables; Machine Learning; GHI Prediction; Time-Series Modelling; Feature Importance; PV System Planning.*

How to Cite: Faisal Alajmi (2025) Prediction and Monitoring of Solar Radiation Using Artificial Neural Networks for Renewable Energy Optimization. *International Journal of Innovative Science and Research Technology,* 10(9), 3152-3166 https://doi.org/10.38124/ijisrt/25sep1052

I. INTRODUCTION

The growing trend of turning to renewable sources of energy is also putting a new strain by introducing new economic and technical challenges of the field of photovoltaic (PV) systems and their integration into power systems [1]. These difficulties are based on a number of factors, like intermittency of solar power, seasonal imbalances between energy delivery and demand and the fact that the cost of storing solar energy is high [2]. In order to control these, these solar power facilities tend to use other generation (ancillary generators) to stabilise production during the times when it is volatile thus increasing both capital and operational costs [3].In order to maintain grid stability, utility and network operators must consider what would be happening to these additional generators by developing short-term and long-term plans that consider their effects [2]. Precise PV or solar radiation predictions are critical to the system operators and utilities, since this allows them to address high levels of PV penetration, without losing their reliability [4]. This means that the importance of solar radiation forecasting in increasing grid integration of solar energy lies in the fact that prior information regarding the power generation capacity of solar panels is given, and thus informed strategic planning.

Because of these problems, there are many studies that explored the potential use of machine learning methods in creating suitable predictions to achieve reliable data about solar radiation “for energy conversion and system-level planning [2] [3] [5]. The time series which represents solar radiation has random and unpredictable patterns. Thus, such a behaviour has to be modelled accurately and translated to a larger scale through proper mathematical modelling. Using a well-defined predictive model, one may express the radiation values mathematically as a conditional expectation by using datasets that analysed historical statistics [5].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1052

The result is that the forecasting of solar radiation has emerged as the major artificial intelligence application in the current power system. Numerical weather prediction (NWP) models, statistical methods of the machine learning type, and climatological methods of solar irradiance estimating and energy generation are used depending on the particular use case [2].

The predictive models commonly rely on the readily available meteorological elements that include temperature, humidity, and wind speed [4] [2]. Empirical methods have been formulated by creating linkages of the climatic variables which in most cases use both linear and nonlinear regressions which involve the use of geographic and weather-based data. These models come quite in handy because direct solar radiation measurement instruments are particularly costly and require maintenance. These models make it possible to estimate global solar radiation (GSR) in different locations because the inputs, in known conditions, such as temperature, humidity, sunshine duration, and solar brightness, as well as geographic coordinates like longitude and latitude are well known [6].

However, one of the limitations that empirical models have is the fact that their accuracy is lower in respect to applying them to inherently nonlinear systems. In order to address these shortcomings, a huge amount of research has focused on the use of machine learning methods, especially the artificial neural networks (ANNs) technique in environmental and renewable energy studies. ANNs represent one of the most popular ML tools, which is the reason they are widely used due to the possibility to model nonlinear relationships and accommodate noisy or insufficient data [7].

ANNs have been studied extensively both in scientific and industrial circles (in particular, in estimation and forecasting of solar radiation). The superiority of ANN-based models has been evident in numerous comparative studies, where it may be taken as a point of reference in the comparison between various AI methods [8]. Their architecture, particularly feedforward ANNs of single hidden layer, have been known to be universal approximator to continuous functions [9]. Their computational efficiency and ability to identify complex patterns in environmental data, have made these models popular in weather and solar forecasts [10].

It is against this background that the current research takes the ANN-driven model to forecast (predict) daily global horizontal irradiance (GHI) based on the application of meteorological characteristics, including air temperature, humidity, wind speed, and subsequent indicators. The model architecture is optimised through automated hyperparameter tuning, and feature selection is enhanced using filter-based techniques. In implementing this hybrid model, the study will enhance accuracy in prediction and cut computational cost therefore increasing prediction efficiency of solar-based energy forecasting and planning investments [7].

II. LITERATURE REVIEW

Solar radiation needs to be anticipated with the highest degree of precision so that the photovoltaic systems can be integrated into the current power networks in the most optimised method [4] [10]. In order to accomplish that, different studies have been conducted and analysed many factors or influencing factors that impact the forecasting accuracy like the prediction horizon, weather classification, as well as the model performance measures. This chapter unravels the essential elements that define the progress and trustworthiness of forecasting solar irradiance strategies in an orderly manner. In Section 2.1, the types of technical and operational factors that influence the accuracy of forecasting have been brought up, such as the importance of time horizons, meteorology, and measures of evaluation. The second section (2.2) summarises briefly on the concept of Artificial Neural Networks (ANNs) that have gained popularity in modelling of solar radiation because of their efficiency in approximating nonlinear non-trivial dynamics within observations of environmental related matters. Collectively, these sections present a framework of methods on how to design ANN-based solar forecasting models and that can be used in the research and practical implementations.

➢ *Influencing Elements in Solar Power Forecasting Accuracy*

There are a variety of factors that influence the accuracy and reliability of the forecasting of solar irradiance, including prediction horizon length, weather conditions classification, performance measurement metrics selection and optimisation and input features selection and optimisation [11]. Each of those aspects is important for producing a quality forecast, which is addressed in more details in the following subsections.

- *Forecasting Horizon*

The commonly used typologies of solar energy forecasting models are based on the time span of predictions. The ability of operators of power generators to forecast energy future generation and energy consumption is crucial. The success of solar irradiance forecasting greatly depends on the choice of the forecasting horizon that is essential to optimum implementation of solar energy applications such as photovoltaic (PV) plants [11]. Grid management especially requires forecasts over various intervals to ensure its stability as well as the assignment of spinning reserves, and the allocation of unit commitments effectively [12]. This may cause forecasting horizons to be categorised into mainly three (1) short-range, (2) medium-range, and (3) long-range. There are also pieces of research, adding a fourth type of forecasting: the use of ultra-short-term forecasting to focus on immediate operational requirements [9].

Currently, forecasting horizon classification of solar energy forecasting has not yet agreed to any classification (system). Nevertheless, the categories commonly distinguished could be described as follows [11]:

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1052

Nowcasting is forecasting in the extremely short term between one minute and several minutes ahead, which is also referred to as intra-hour forecasting. The forecasting acts in especially dynamic fields like pricing in electricity markets, bidding, real-time operation of electricity systems and power system stability plans with peak loads.

Short-term forecasting includes a couple of hours, a day or a week in future. This horizon is essential to certain operations, including the optimal unit commitment tasks, the rotating reserve management tasks, and the task of analysing energy trading contracts between businesses. It enhances the installation of energy management systems in combination with photovoltaic (PV) systems, and makes the whole grid more reliable as well.

Medium-term forecasting takes a period of one month to one year. It is particularly helpful in the planning of maintenance schedules of solar power infrastructure such as the transformers and the auxiliary systems with the aim of reducing energy loss during use.

Long-range forecasting goes up to ten years and can be as short as one year. This type is best applied to strategic planning purposes like the choice of suitable site of PV power plants, development of solar energy infrastructure on the large scale, and managing the transmission and distribution networks. The long-term models are restricted by a fact that they cannot predict the future weather change with high accuracy, but they could be useful in a long-term program in scheduling and forecasting prices, as well as planning site development.

Forecasting horizon has a significant influence on the prediction model accuracy, even when the forecasting inputs are held constant [13]. The accuracy of solar irradiance predictions depends on numerous factors that are interdependent and include such aspects as ultra-short periods (down to a few seconds or a few minutes), amount of cloud coverage, and the intensity of solar radiation reaching the planet [14]. Thus, to enhance the prediction of irradiance, one should include cloud types and predictions. However, most of the current studies neglect cloud conditions during the modelling procedures which eventually compromises the value of forecast reliability [11].

- *Weather Classification*

The quality of production of solar energy largely depends on the amount of incoming solar radiation and this is considered as a baseline figure in the estimation of the performance of photovoltaic (PV) systems. The existence and fluctuation of this radiation depend on much atmospheric conditions such as the nature of cloud cover, temperature of the surrounding environment, existence of moisture, wind conditions, barometric pressure, and particles quantity. With this reliance, changes in meteorological environments have immense influences on the predictability of any solar irradiance models [11]. This makes the categorisation of the weather conditions vital in guaranteeing the accuracy as well as the robustness of the forecasting algorithms. Several studies have emphasised that weather pattern classification is a very essential module in the preparation of first level data to tackle the short-term solar radiation forecasting operations [15, 16] .

A major problem in weather classification that can be used in forecasting lies in the fact that information that is fed to predictive models is scarce [17]. To solve this, weather types may be classified in bigger categories like shrinking dozens of classifications to approximately 10 abridged categories. Such a method is beneficial when creating enough patterns to train the model. The other approach in the management of data scarcity is the simplification of weather classification schemes to three or four major categories. Such a reduction is used to make model training more robust and to better generalise [11] [18]. The popular method that has been used with surface meteorological variables and solar irradiance measurements is the K-means clustering to study the variability related to clouds. Such an approach helps in grouping the similar weather conditions in terms of their statistical features. Solar irradiance forecasting requires operations of Artificial Neural Networks (ANN) especially in regimes that experience piecewise-smooth weather patterns. Such networks are flexible depending on the different conditions in different atmospheres and have been more accurate in predictions.

Clustering meteorological inputs of different weather classes is also performed commonly by means of Self-Organisation of Maps (SOMs) [11]. The classes can be included in forecasting systems when carrying out weather specific predictions to improve the system performance. An even better framework by categorising weather is based on a four-class first-order model of typical ground-level profiles of solar radiation. In other applications weather conditions can be coarsely classified as dry and wet regions which can then be clustered by a clustering algorithm thus defining regions with similar behaviour of weather.

The most recent new forecasting methods are based on generative adversarial networks (GANs) together with CNN1D, CNN2D, multilayer perceptrons (MLPs) and support vector machines (SVMs) and k-nearest neighbours (KNNs). Translation: these models work with the input data divided according to weather type, and the performance is often assessed with the help of confusion matrices, measures of the accuracy of identifying different classes of weather [11].

Out of these, CNN2D has proved to be highly effective as it is capable of handling non-linear relationships both between the inputs and outputs. The models based on GANs, including the ones combined with the CNN structures, have potential in classifications accuracy enhancement and overcoming the data imbalance, especially in the case of short-term forecasting within the scope of machine learning [11]. In summation, proper identification of weather condition greatly contributes to the success of the system of forecasting solar irradiance and photovoltaic power. The catagorisation of weather becomes of paramount importance, as it increases the accuracy and trustable model development.

- *Model Performance Metrics*

Evaluation metrics are quite crucial at each point of developing a deep learning model. Because of the lack of these measures, it is hard to quantify the performance of a model in comparison to the performance of others [19]. The effectiveness of the model may depend on a number of factors such as the interval that the model will be used to predict, the conditions on which the training has been done and localised weather conditions. In general, the model's performance is checked by comparing predicted solar irradiance values with the actual ones. Such metrics are used to assess the accuracy of the predictions and make adjustments that can enhance precision. Moreover, the lower the metrics the better as it is an indication that the forecast is near reality. There are certain commonly used statistical tools that are being used to achieve this purpose as presented below [11] [19] [5]:

✓ *Mean Absolute Error (MAE):*

The MAEs quantify the mean value of the differences between the measured values and the computed values of the GHI. It does not put more weight on more significant errors but normalises all of them by giving them absolute values before averaging. This will make sure that no error is given too little or too much weight as compared to the other errors. The equation is written in Eq. (1) [11]:

MAE=1/N ∑_(i=1)^n▒Ga_i-Gp_i | ………,……….……..(1)

Here, Gai, and Gpi are the observed and predicted GHI at instance i and n is the total number of observations.

✓ *Mean bias error (MBE):*

This metric examines the mean error of the predictions of the model. Unlike MAE, it takes direction of the error into account by taking underestimations and overestimations to cancel each other. Although MBE cannot necessarily be used when judging solely performance, it can be used to determine whether the model either overpredicts or underpredicts values. This is illustrated by the following equation in Eq. (2) [11]:

MBE=1/N ∑_(i=1)^N▒(Gp_i-Ga_i)| …………………..…..(2)

✓ *Mean square error (MSE):*

In MSE, the mean of the squared differences between predicted and real solar irradiance values are calculated. It amplifies both predictor and actual values since it squares them, thus the difference between a predicted and an actual value has a stronger effect. This predisposes it against large errors as compared to MAE. The expression is provided in the following Eq. (3) [11] [7]:

MSE=1/N ∑_i^n▒Gp_i-Ga_i | ……………..…………..(3)

✓ *Root Mean square error (RMSE):*

Square root of mean of the squared differences between the estimated and observed values of solar irradiance is calculated by RMSE. It is also one of the most reliable measures of the model accuracy as it places greater weight on larger errors that are of great assistance in identifying the outliers. The calculation of this measure is done by Eq. (4) [11] [7]:

R=√(1/N ∑_i^n▒(Gp_i-Ga_i)^2)……………..………..(4)

✓ *Mean absolute percentage error (MAPE):*

MAPE is a form of average of any absolute error in the form of a percentage of actual values. One calculates it by dividing the absolute error of each prediction by the correct value associated with that specific prediction giving a normalised measure of error. According to Eq. (5), this method can be characterised as follows [11] [7]:

MABE=1/N ∑_(i=1)^N▒(Gp_i-Ga_i)/(Ga_i)|
……………..(5)

✓ *Normalised RMSE (nRMSE):*

To compare model error with the mean actual, the normalised RMSE is computed when large datasets are involved. it normalises the RMSE in respect to the observations so as to facilitate comparison of performance between datasets of varying size. It is expressed in Eq. (6) [11] [6]:

nRMSE=1N inGpi-Gai2(Ga_i)¯ ……,….………...……..(6)

✓ *Correlation coefficient (R):*

It is used to record how well the values of the output of the forecast of solar irradiance values linearly correlate with the actual ones. When the value is near 1 then there is a high positive correlation between the predicted and the actual data. The Eq. (7) was not specified exactly, but usually it has this form [11] [20]:

R=(∑_i^n▒〖(Ga_i-(Ga_i)¯)(Gp_i-(Gp_i)¯)〗
)/√(∑_i^n▒〖(Ga_i-(Ga_i)¯)∑_i^n▒(Gp_i-(Gp_i)¯)^2 〗)
……………..(7)

Here, (Gp_i)¯and (Ga_i)¯are the forecasted average along with the actual GHI, correspondingly.

Forecast skill score (FS):

The FS measure quantifies the quality of a forecasting model in relation to a simpler reference model known as a persistence or benchmark model. The determination of it occurs under the Eq. (8) [11] [20]:

FS=1-(RMSE_model)/(RMSE_(simple model)) …..........(8)

RMSE model in this formula is the forecasting model being tested and RMSE simple model is the error made within the simple reference model. When 1 is obtained then it is an ideal forecast. A result of 0 is an indication that the model is as good as the benchmark. When the FS score is negative, this indicates that the tested model would be worse than the baseline method.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1052

- *Theoretical Background: Artificial Neural Network (ANN)*

Artificial Neural Networks (ANNs) are computer systems used to model the complex relationships in nonlinear and inconsistent data sets that often occur in a nonlinear fashion. These programs are easy to use and they can be implemented without necessarily having an extensive knowledge in higher mathematics. Unlike conventional methods, ANNs are more effective in the process of mapping the relationship between inputs characteristics and the desired outputs, and take significantly less computational resources [21] [6];

ANNs are non-linear and are characterised by parallel processing as non-compliant of any strict rules [21]. The ANNs have, at their lowest level, processing units known as neurons, which are organised into layers of interconnections. It is usually a layered structure that comprises the following layers as the input layer (where the raw data is given), one or more layers of hidden layers (where computations are made internally), and the output layer (where the final product is delivered) [21] [22].

The data is extracted by neurons through the adding of a bias and the implementing of a weight on the input. Such a weighted input is transferred to an activation function in the hidden layer. This is transformed signal which is sent into another layer. Purelin is a popular activation function of the output layer and results in direct proportionality between the input and the output [21] [22]:

Inputs into neurons consist of a number of variables (e.g., x 1, x 2... x n) and the output of the neurons is denoted by Y. We tend to express the relationship between the inputs and the output in a mathematical form, according to what Eq. (9) entails [6]:

$$Y=f(v(w,x)) \ldots\ldots\ldots\ldots\ldots (9)$$

$$v(w,x)=\sum_{k=1}^{n} w_k x_k + b = W'+B \quad (10)$$

The weights of each input would be represented by the function v in an artificial neural network and the bias term is the net summation of the weights (i.e., W = w1, w2, wn) of all the input connections. The weight vector and the associated value of input are a vector by (X = x1, x2,.,xt n). The bias, which is shown by b, is added to the general expression as is illustrated in Eq. (10).

f is the activation function that takes the signal given to it by the previous layer to give off the final signal. This particular two-level activation function to be implemented is dependent on the type (hidden or out) layer of which the neuron is a part but also on task at hand.

There are a number of activation functions commonly mentioned in the literature when applying them in the context of a neural network. These are logistic sigmoid, hyperbolic tangent sigmoid, Gaussian radial basis function, linear, unipolar and bipolar step functions together with their linear versions [6]. Both of them have a very different role in creating a response of the network.

With the feedforward neural network structure, different model structures were tried with models having a different number of layers. The outcomes revealed that adding layers to the depth that is more than two layers did not provide an improved performance, just bringing in the layers of complexity. Thus, the two-layer structure was the discussed system as the most efficient one [21] [22].

As mentioned before, Artificial Neural Networks (ANNs) are a branch of artificial intelligence (AI) and their tools of complex tasks. They are especially useful in solving nonlinear problems, e.g. approximation to functions, pattern recognition, data classification and clustering, optimisation and simulation [11]. ANNs can also be referred to as black-box models, since they are used to simply learn nonlinear patterns between the inputs and outputs without actually articulating the connections. The standard ANN has three basic layers: the input layer (where the information settles), one or more hidden layers (where the processing is performed) and the output layer (where the final result is given out). A combination of weighted connections, a bias value, an activation function and summation nodes are some of the key factors [21] [22].

The ANNs operate in two main phases which include: training (or learning) and testing (or generalisation). The model uses an error function to change its weights and biases during the training to lower the gap between the anticipated and empirical outputs. The algorithm that drives the process is known as a learning algorithm and it goes through multiple epochs, with each epoch consisting of one pass through the training dataset with the network once [7].

The four major types of learning in ANN are generally supervised, unsupervised, reinforcement, and evolutionary. In supervised learning, the network is compared with what it is supposed to be giving. The error is processed as a way of updating the weights and the biases in order to make better forecasts in the future. This method is used as a closed-loop feedback system with the error being the signal that gives the corrections [2] [4].

One of the most important metrics of error is the so-called Mean Squared Error (MSE) that is computed at every training epoch. The process of training the network will go on until this error value is made as small as possible, which means that the network has mastered the rule matter. Solar radiation modelling, prediction, and forecasts are among several activities in solar radiation that the techniques provided by ANN are specifically useful [23].

These recent occurrences in the deployment of Artificial Neural Networks (ANNs) have dotted the usefulness of the networks in modelling and forecasting of solar energy. In another research by Ganotra et al., a multilayer perceptron ANN model was also developed to forecast solar power generation based on a set of historical records on solar energy production, weather conditions and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1052

environmental parameters. Since the ANN technique had shown commendable levels of accuracy as compared to conventional statistical and machine learning techniques, MAE and RMSE have shown that ANN is much more accurate. This was the confirmation of the model to reflect nonlinear impacts due to the complication between multiple inputs [24].

ANN models were also used by Lim and Chaiwiwatworakul to predict the solar radiations in Bangkok, Thailand on one-hour-ahead basis over vertical planes bearing to various cardinal directions. The model they used, based on the use of the meteorological data of 2019-2022, showed good results in all orientations and had a high level of the forecasting accuracy without using other transposition models. Prediction of directional solar irradiance using ANN: This research shows the effectiveness of ANN methods in predicting directional solar irradiance in the real-world condition [24].

In one of their recent articles, Abdlrazg and Atetalla (2024) investigated solar radiation forecast in various places in Libya through two ANN models Backpropagation Neural Networks (BPNNs) and Radial Basis Function Networks (RBFNs). The RBFN model was more precise and faster in execution as it got the regression score of 93.15 percent and Mean Squared Error (MSE) of 0.0090. These results reinforce the overall effectiveness of ANN models in the forecasting of regional solar radiation, particularly when adapted to a particular geographical data [26].

The new studies unconditionally present ANNs as models with potent capabilities in solar energy forecasting since they produce high accuracy, flexibility across different climatic and geographic conditions, and adaptability. In the part below, the methodology of the study will be described.

III. METHODOLOGY

In this section, the strategy used to estimate daily global horizontal solar radiation by use of a neural network-based prediction framework is described. The workflow contains, the raw data measurement collection, the data cleaning and data structure, the creation of input attributes, the data statistical filtering method to detect which predictors are relevant, the model training, and the evaluation of its prediction capabilities. It was all worked in Python with the utilisation of its data science libraries, including pandas, scikit-learn, matplotlib, and seaborn. The following repeating bending process diagram shows the adopted methodology steps:

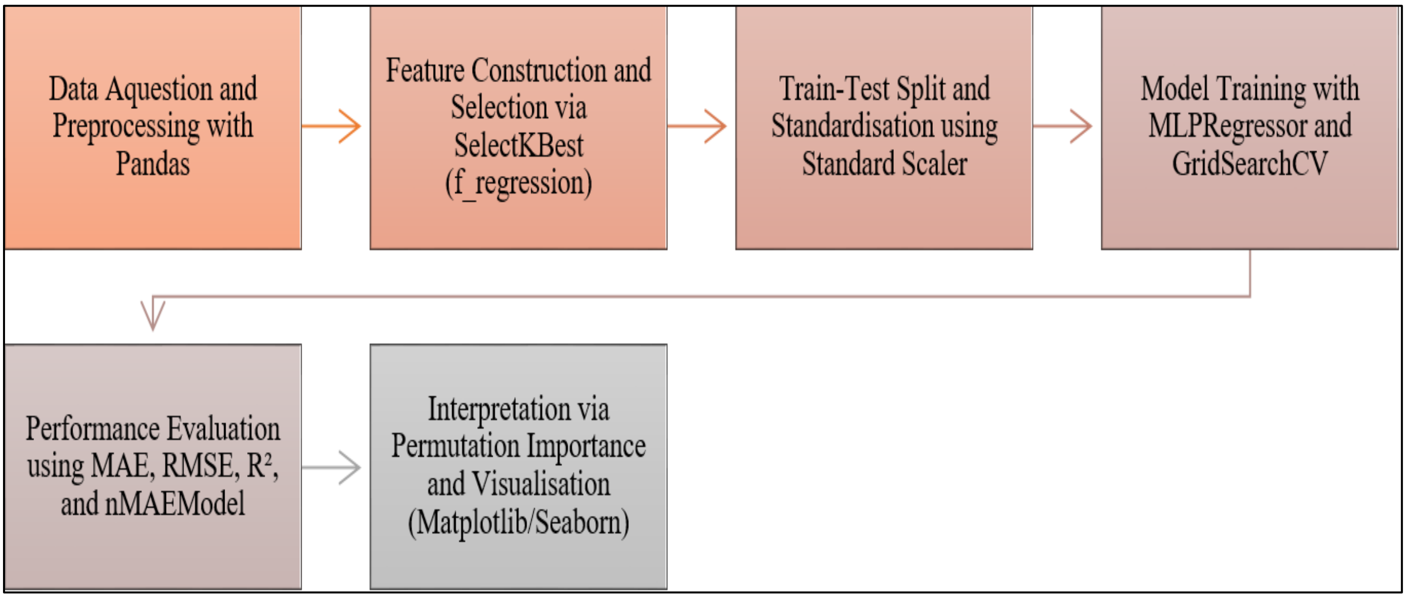

Fig 1The Adopted Methodology Steps

➢ *Data Acquisition and Preprocessing*

The data sets used in this research were downloaded to a web-based power platform of the NASA agency that is known as POWER (Prediction of Worldwide Energy Resource). The data includes a chronicle of meteorological and solar parameters that are peculiar to a geographical point located at 29.36 N and latitude and 47.97 E longitude. The file name of CSV file is POWER_Point_Daily_20200101_20250701_029d36N_047d97E_LST.csv. The data is observed between the date of [January 2020-July 2025] at a particular point and its variables are in the form of the air temperature, relative moisture content, wind strength, and global horizontal irradiance (GHI). Python (pandas' library) was used to create column labels assigned to variables of interest after the raw data was imported. The date elements (Year, Month, Day) were transformed into a single datetime object, and the records with missing values represented systematically in order to make the dataset we have integrity as well.

➢ *Feature Engineering and Selection*

In addition, novel input attributes are derived in order to enhance the model in the ability of realising meaningful environmental interactions. These were daily thermal variation (calculated as the difference between the maximum and minimum temperature of the day) and a combined

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1052

dimension that was between humidity and the velocity of the wind (both of which were thought to have an impact on the means by which solar energy is transported into the atmosphere). In order to obtain the most relevant predictors, statistical filtering method was adopted with SelectKBest method and f regression scoring criterion, that ranks inputs depending on their linear relationship to the output variable. Further modelling was based on selecting five most informative features, i.e. T2M, T2M_MAX, T2M_MIN, TEMP_RANGE and RH2M.

This method simplified not only the computing burden, but also reduced the chances of overfitting due to removal of less important inputs. The following table defines every single term adopted in the work, in the background of solar radiation and weather data:

Table 1: Definition Term Adopted in The Work of Solar Radiation and Weather Data.

Feature Name	Full Term	Definition	Units	Source/Standard
T2M	Air Temperature at 2 meters	The ambient air temperature measured or modeled at a height of 2 meters above ground level.	°C (Celsius)	NASA POWER / meteorological stations
T2M_MAX	Maximum Air Temperature	The highest recorded or estimated air temperature at 2 meters above ground level for a specific day.	°C (Celsius)	NASA POWER / daily summary
T2M_MIN	Minimum Air Temperature	The lowest recorded or estimated air temperature at 2 meters above ground level for a specific day.	°C (Celsius)	NASA POWER / daily summary
TEMP_RANGE	Temperature Range	The calculated difference between the daily maximum and minimum air temperatures (T2M_MAX – T2M_MIN).	°C (Celsius)	Derived feature
RH2M	Relative Humidity at 2 meters	The ratio of actual water vapor content to the maximum possible at that temperature, measured or modeled at 2 meters above ground level.	% (percent)	NASA POWER / meteorological stations

➢ *Data Partitioning and Normalisation*

In this study data set was split into training and testing sets in an 80-20 ratio by using train test split. In order to keep input contributions balanced in spite of having a wide variety of input to offer at different points in the network, normalisation was performed on all data that came up using the StandardScaler trained on the training set and used whenever converting the subsets. This process was used to normalise all the input values as having a mean of zero and standard deviation of one. This type of standardisation is extremely important to have consistent gradient descent when training neural networks and will make the networks converge faster and achieve better stability.

➢ *Model Selection/Development and Hyperparameter Optimisation*

In order to determine the best structure of artificial neural network (ANN) that could be used in predicting the solar radiation, a systematic model exploration process was done through a grid search with cross-validation. Other tests on the Multi-Layer Perceptron (MLP) regressor were conducted including different combinations of hidden layer architectures, activation functions (ReLU and tanh), and initial learning rates (0.001 and 0.01). This search was aimed at finding the hyperparameter set that resulted in the minimum mean absolute error (MAE) on the validation subsets. The selection of an effective yet moderately complex ANN structure without overfitting or underfitting risk in this methodology was achieved. The last model was selected due to its stable cross-validation and further tested on another test data-set in order to validate.

Model Development is based on a multilayer perceptron (MLP) neural network which was implemented by MLPRegressor module. To optimise the model, the hyperparameter search was conducted with GridSearchCV in order to maximise the model. This search was performed to examine a range of network architectures, namely, the number of different sized hidden layers, activation functions (ReLU and tanh), and initial learning rates [1] [2].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1052

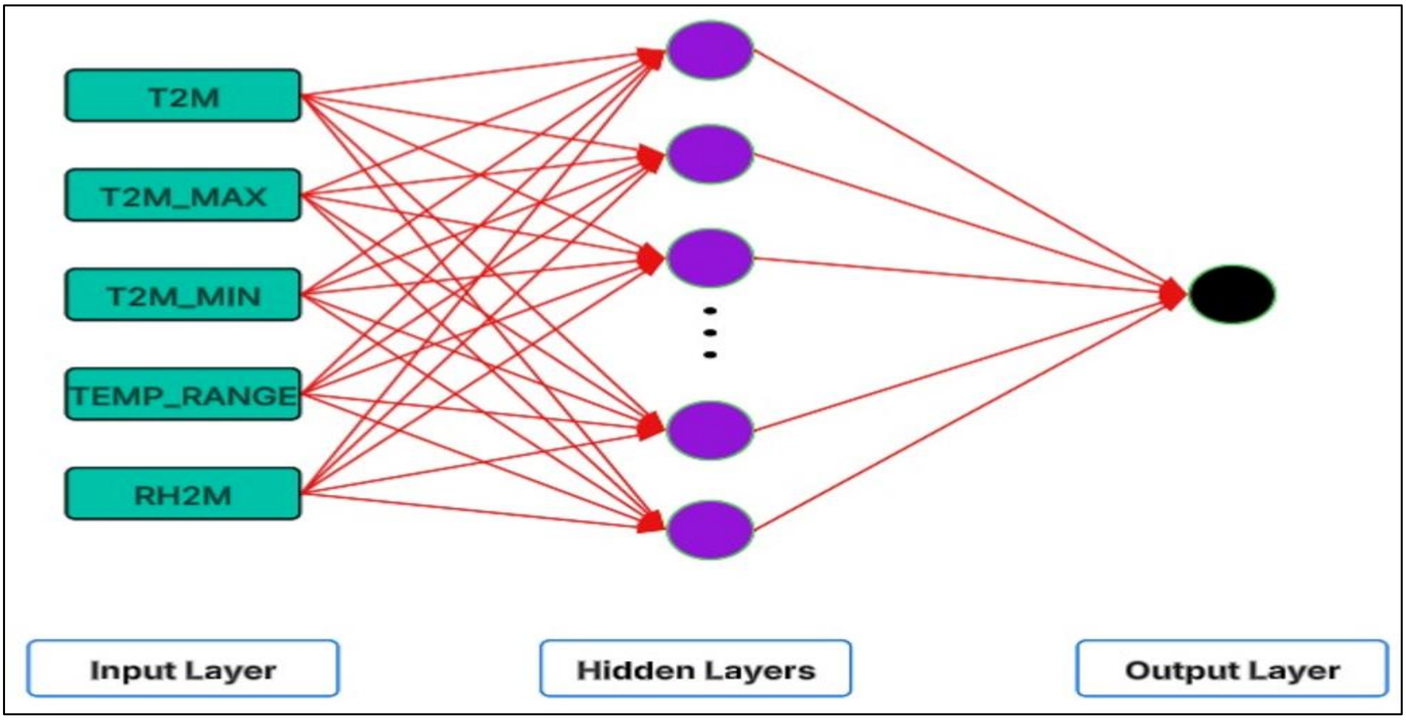

Fig 2 ANN Structure for GHI Prediction.

The model was evaluated using 3-fold cross-validation and the negative mean absolute error was used to optimise its performance. he following hyperparameters were optimised using grid search with 3-fold cross-validation via GridSearchCV:

Hidden layer configurations: (64,), (128,), (64, 32), (128, 64), (128, 64, 32), (128, 64, 32, 16) Activation functions: relu, tanh Initial learning rates: 0.001, 0.01

The best model presented consisted of two soft layers with 128 and 64 neurons and the activation function was ReLU and learning rate 0.001. Early stopping mechanism was also introduced into the training with the aim of preventing overfitting.

➢ *Model Evaluation and Prediction*

The concluded neural model was re-trained with the whole training dataset containing the optimised set of inputs variables, and thereafter used to predict on the test data. Four key performance indicators were used to measure the effectiveness of the predictions, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination (R2), and Normalised MAE (nMAE) expressed in equations (1)-(3) [3].

$$MAE = (1/n)^* \Sigma |yi - \hat{y}i| \quad \text{for } i = 1 \text{ to } n \ldots\ldots(1)$$

$$RMSE = \sqrt{((1/n)^* \Sigma (yi - \hat{y}i)^2)} \quad \text{for } i = 1 \text{ to } n \ldots\ldots(2)$$

$$R^2 = 1 - [\, [\![\Sigma(yi - \hat{y}i)]\!]^2/(\Sigma (yi - \bar{y})^2)] \ldots\ldots(3)$$

These assessment measures were applied in determining the accuracy of the forecasts as well as the relative accuracy of the model. An exact prediction was also made on August 14, 2022, and it was observed that the computed predicted value backward is very close to the actual value proving that the model can also be used in real-life situations.

➢ *Model Interpretation and Visualisation*

In order to enhance the transparency of the model decision process, feature importance analysis has been done using permutation to quantify the extent to which each input variable affected the results of the prediction. Different visual measures to evaluate the models were used such as:

- Scatter plots of actual and forecast GHI.
- A histogram of the residual errors and plots of residuals versus the observed GHI.
- Also, a time-series plot was employed to see the extent to which actual values were closely approximated by the predictions over the test period.
- Boxplots on a monthly basis were also done to determine seasonal variations and distributions of GHI. All these visualisations confirmed the accuracy of the model and proved that it is stable in dissimilar time frames.

IV. RESULTS

The results given in this section is a clear investigation of the outcomes generated in the ANN based model of predicting daily Global Horizontal Irradiance (GHI). Results are categorised in terms of such dimensions as numerical performance indicators, visual performance diagnostics, and feature relevance. Combined, these performance metrics are utilised to confirm the accuracy, generalisation ability, and practical applicability of the model to make predictions on solar radiation when the atmospheric circumstances are not

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1052

conversant. The last model was trained on five features being chosen: Air Temperature at 2 meters (T2M), Maximum Air Temperature (T2M_MAX), Minimum Air Temperature (T2M_MIN), Relative Humidity at 2 meters (RH2M) and Temperature Range (TEMP_RANGE) which was derived. These predictors have been chosen by statistical sifting (SelectKBest with F-regression) in order to choose the most useful predictors and yet make the model simpler. The scatter plot, in Figure 3 shows a plot of the observed and predicted values of GHI on the test dataset.

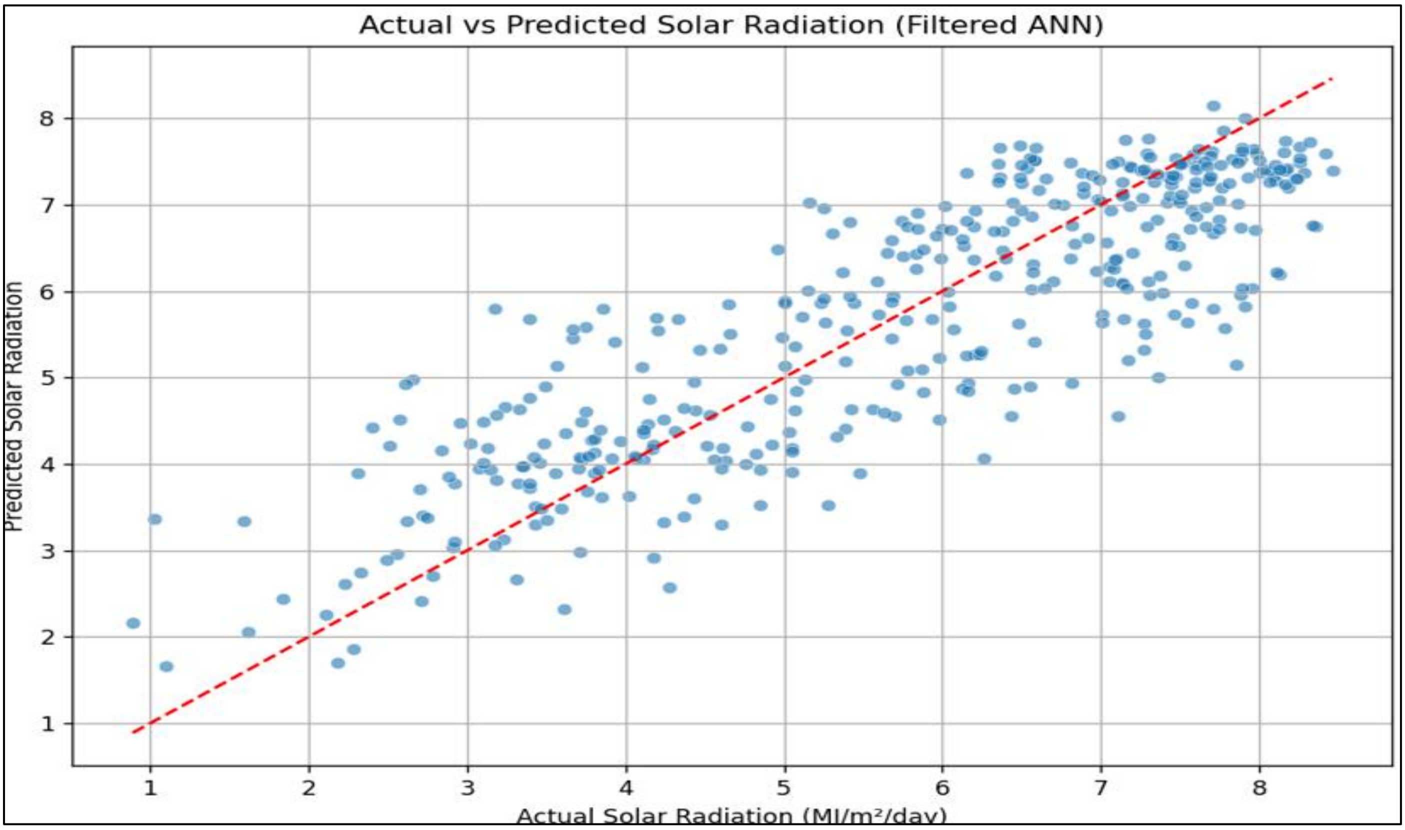

Fig 3 Scatter of Actual and Predicted Values of GHI with Trained ANN Model.

The numbers on the x-axis are the days that they are predicting, and the red dash line is ideal where the actual and predicted values are 1:1 pairing. A high concentration of points along this line will connote that the ANN model had closely approximated the values of the true GHI at diverse conditions. The separation of the line is small which implies low variance of the prediction error. This number gives instant and intuitive information to the accuracy of models.

Fig 4 shows the Actual versus Predicted Solar Radiation 2000-2008. This line graph shows the comparison of predicted and measured data of GHI on part of the test period. The high similarity between the real and the fitted adornments shows that the model is able to capture the short-term variation as well as the long-term seasonal patterns. This matching is essential in energy planning activities where either under or overestimation of sunlight may attract economic and performance costs of photovoltaic (PV) devices.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1052

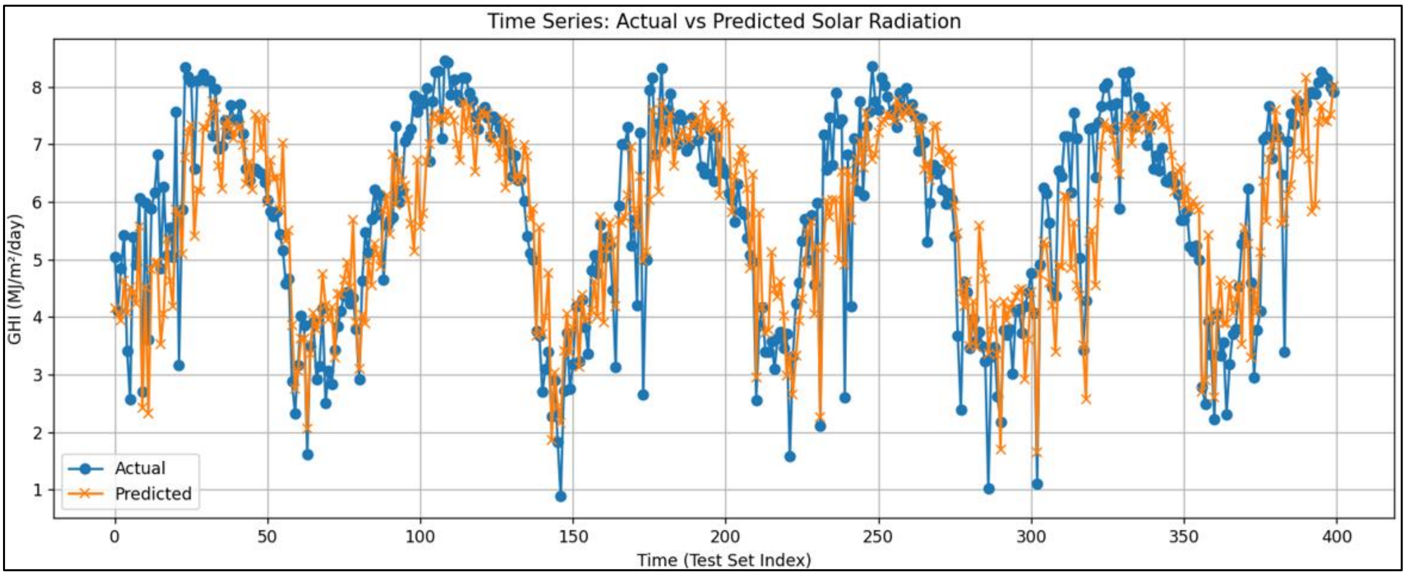

Fig 4 Actual- Predicted Plot of Time Series of GHI Against The Test Data

Fig 5 shows a histogram of the actual versus predicted GHI (obtained as predicted GHI-actual GHI). The fact that the bell-shaped curve is centred on zero proves that the errors of the model are symmetrical. This is a significant diagnostic, because it means that systematic bias is not a problem in the model. The bulk of errors is within a narrow range (approximately ±1 MJ/m²/day) which indicates the high stability of the testing data.

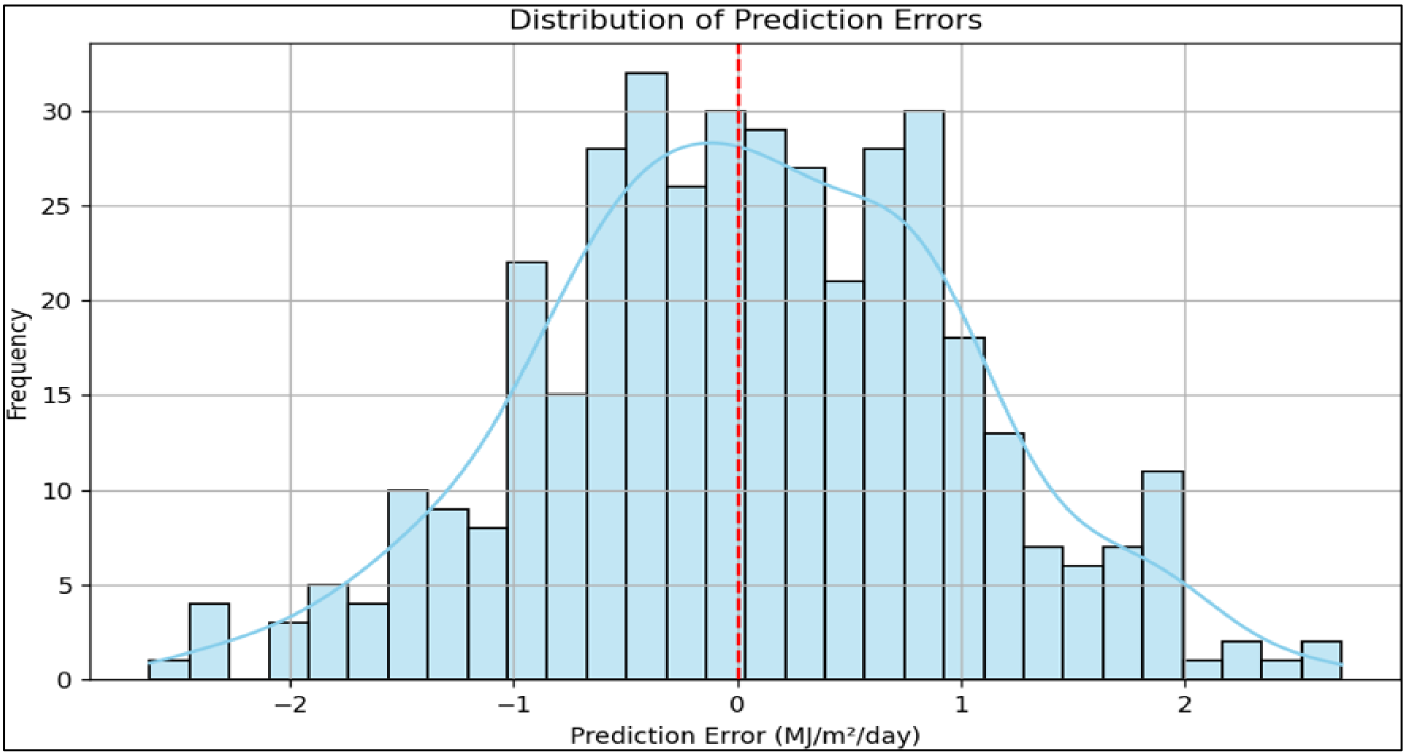

Fig 5 Histogram of Prediction Errors of ANN-Based GHI Forecast.

The residual plot depicted in Figure 6 shows error in prediction against the actual GHI values. Ideally, the residuals ought to display no perceptible pattern and this has remained the case as the residuals are scattered randomly about zero with the levels of actual solar radiation. This will demonstrate that the model exhibits the same accuracy in the decreased radiation value in winter months as well as maximum radiation value in summer.

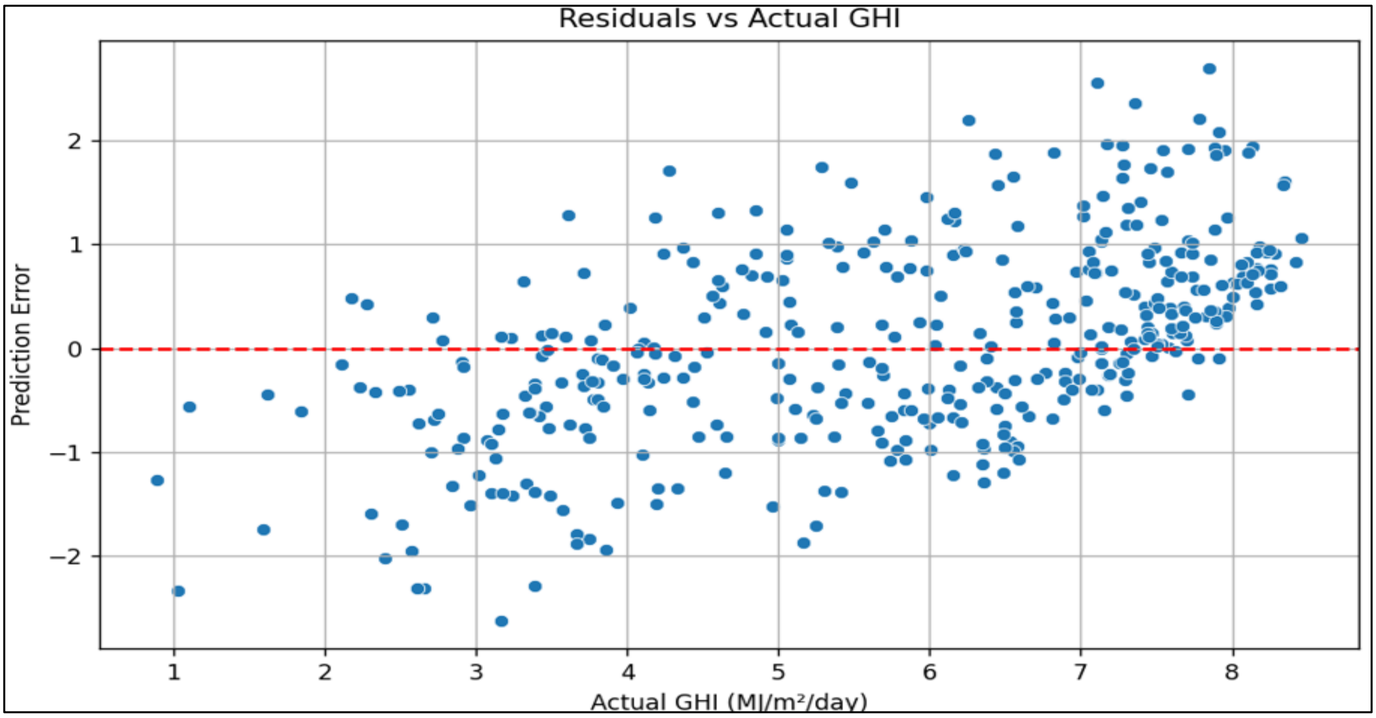

Fig 6 Residuals vs. Actual GHI.

Fig 7 represents the relative significance of all the input features of the trained ANN model through the permutation importance. Out of the five predictors, Air Temperature (T2M) and Relative Humidity (RH2M) exhibited the highest levels of influential variables. The feature TEMP_RANGE was also generated, which also helped to some degree indicating that the model is sensitive to the daily variation in temperature. The maximum temperature data and the minimum temperature data made a significant impact individually, whereas the other two values (mean and median) did not create as big of an impact as the maximum and minimum values did.

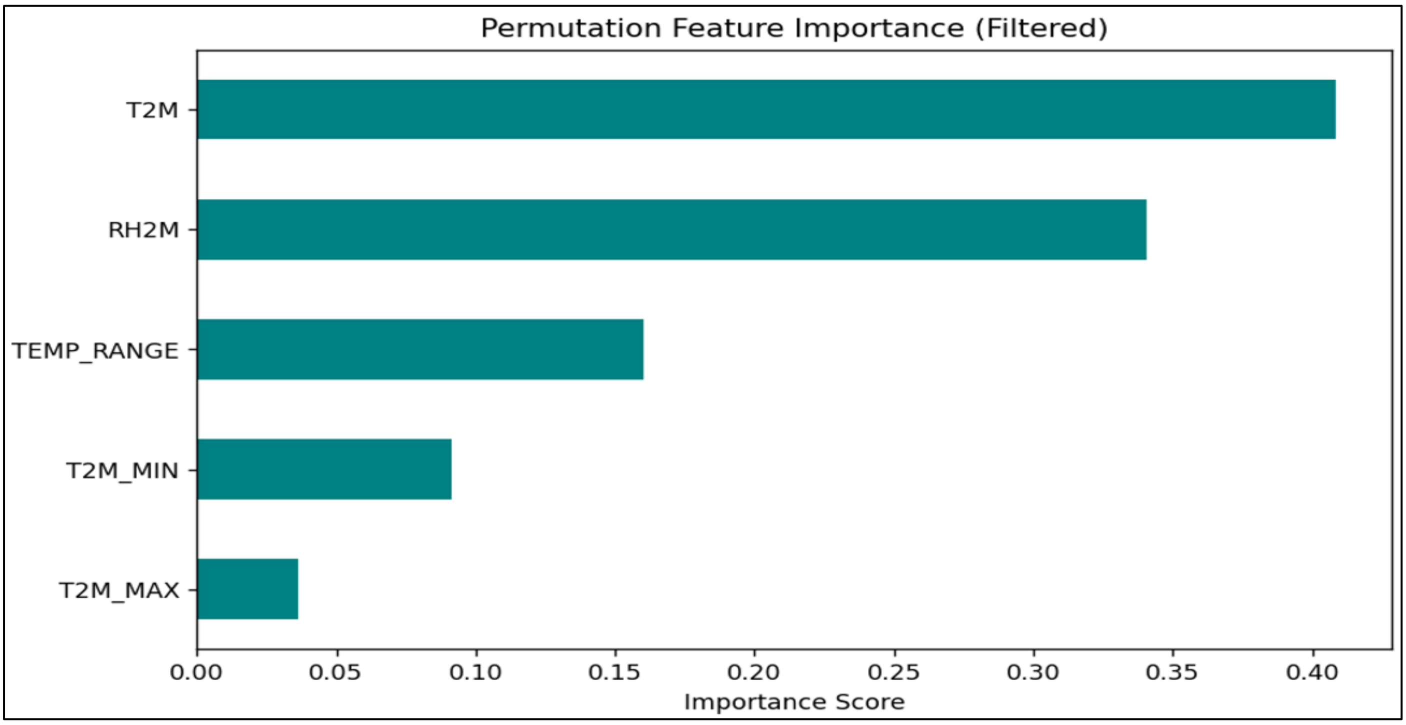

Fig 7 Permutation Feature Importance (Filtered).

Fig 8 shows boxplots of monthly GHI values to reflect the seasonal change in the solar radiation. It is also clear that distribution of radiation is most in the months of May to July so as to correspond with summer months of that region being studied. The radiation is found to be lower in the winter months (December to February) in the ranges as expected.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1052

The boxplots further show relatively consistent median value and interquartile range throughout summer season, which means that the irradiance in the peak generation times will be rather stable.

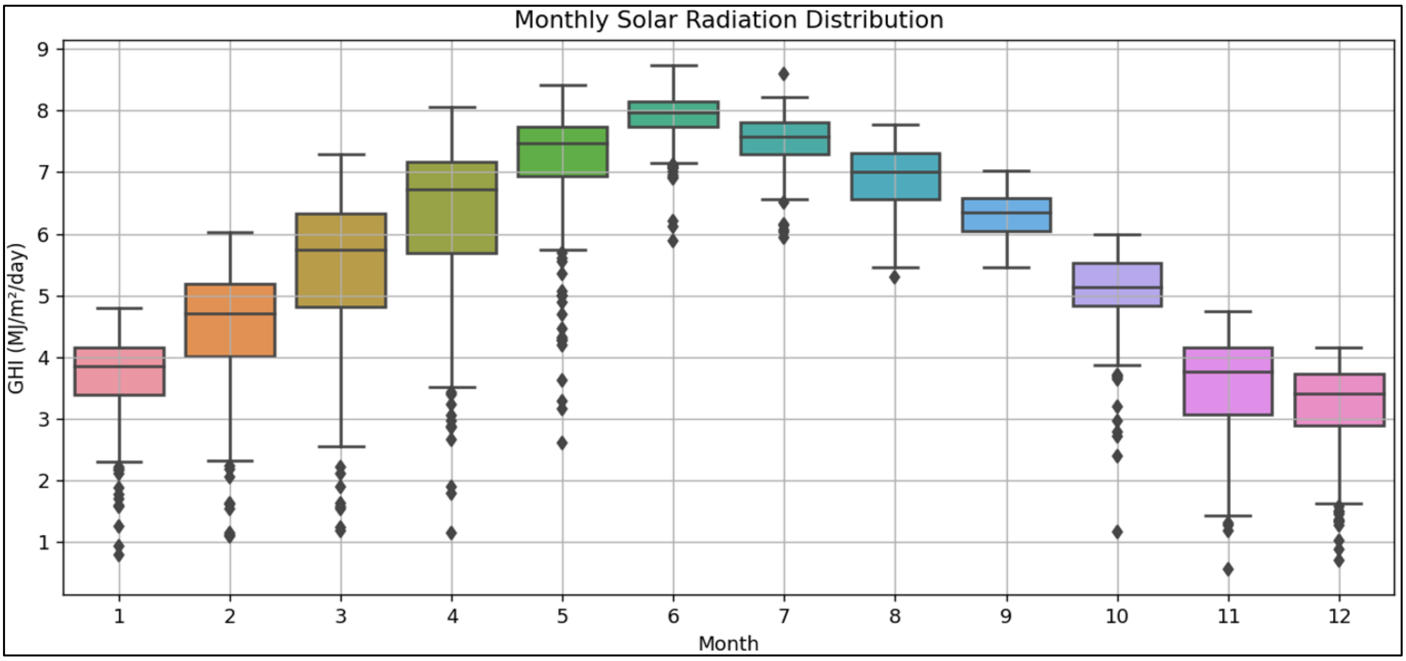

Fig 8 Monthly Solar Radiation Distribution

The summary of optimal ANN model architecture, chosen using the grid search optimisation is presented in Figure 9. Its architecture is made up of two concealed layers that contain 128 and 64 neurons, the activation function is the ReLU function, and the preliminary learning rate is 0.001. Overfitting was avoided by applying the early stopping mechanism. The test set key assessment criteria are to include:

MAE: 0.754 MJ/m2 / day
RMSE: 0.943 MJ/m2 / day
R2: 0.725

Normalised MAE: 13.05%

August 14, 2022 prediction provided a close prediction of 6.964 MJ/ m2/day compared to an actual value (6.93 MJ/ m2/day), which, therefore, makes this assumption useful in predicting the level of GHI in real-time.

```
✅ Selected Features: ['T2M', 'T2M_MAX', 'T2M_MIN', 'TEMP_RANGE', 'RH2M']
Fitting 3 folds for each of 20 candidates, totalling 60 fits

📊 Best Model Hyperparameters: {'activation': 'relu', 'hidden_layer_sizes': (128, 64), 'learning_rate_init': 0.001}

📊 Model Evaluation:
MAE      = 0.754 MJ/m²/day
RMSE     = 0.943 MJ/m²/day
R²       = 0.725
nMAE     = 13.05%

📅 Prediction for 2022-08-14:
🔍 Predicted GHI: 6.964 MJ/m²/day
🎯 Actual GHI   : 6.93 MJ/m²/day
```
Fig 9 Model Summary and August 14, 2022 Prediction.

Fig 10 represents Sample Prediction of January 1, 2019. In order to extend evidence of the power of the model, one day forecast with another year (January 1, 2019) is presented. The mean solar radiation GHI was estimated to be 2.499 MJ/ m2/day against 2.875 MJ/ m2/day that was considered to be the observed value. The absolute error value 0.376 is not high especially in a winter day when irradiance is normally higher and more variable.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1052

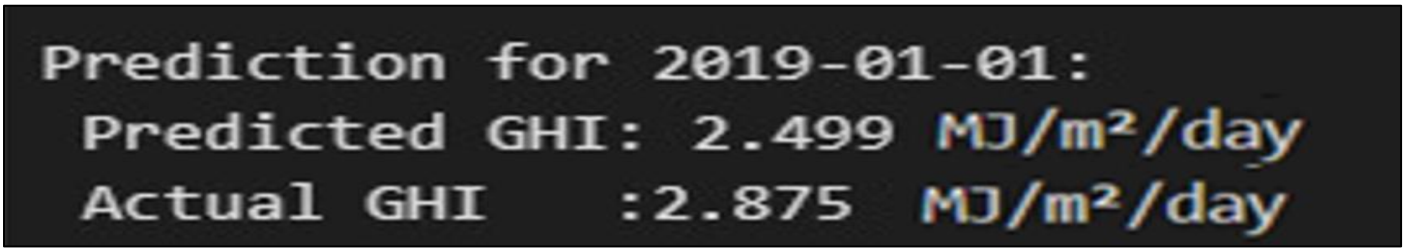

Fig 10 Model Summary and January 01, 2019 Prediction.

Fig 11 represents another Sample Prediction of July 26, 2019. In order to extend evidence of the power of the model, one day forecast with another year (January 1, 2019) is presented. The mean solar radiation GHI was estimated to be 7.8853MJ/ m2/day against 7.456 MJ/ m2/day that was considered to be the observed value. The absolute error value 0.429 is not high especially in a summer day when irradiance is normally higher and more variable.

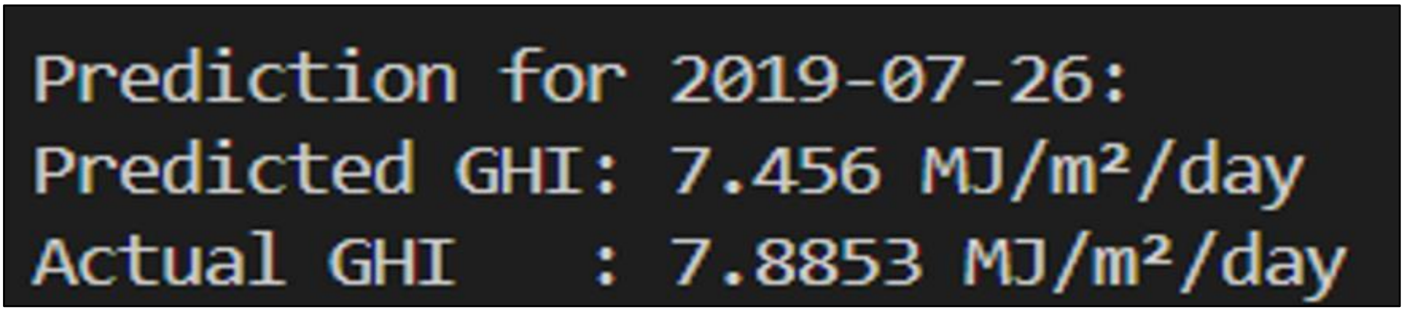

Fig 11 Model Summary and July 26, 2019 Prediction

V. DISCUSSION

This section discusses and interprets the results obtained from the ANN model in the previous chapter in terms of its effectiveness, reliability and practical implications towards solar radiation forecasting. This model uses a subset of selected variables of meteorology in an effort to obtain the maximum degree of accuracy in the estimate of daily solar radiation. The metrics of the performance and the visual diagnostics give a strong confirmation of the merits and weaknesses of the model.

A high predictive performance was achieved by the ANN with an MAE = 0.754 MJ/ m2/day and RMSE = 0.943 MJ/ m2/day also having a coefficient of determination (R2) of 0.725. These values show that the model accounts to about 73 percent of the variation in the values of GHI. This type of performance is commendable, particularly since only five input features were involved, and no external imagery of satellite sources was assimilated and no ground-based sensor feeds were also included.

After examining both the scatter and time-series plots, it can be seen that the model is competent at following actual values, both on a macro basis as well as a day-to-day basis. This is critical in real-life applications like energy management in real-time, planning grid integration, and battery energy storage system forecasting.

The residual analysis and the plots of error distribution serve to support the accuracy of the ANN further. A centred and symmetrical distribution of the model errors is an indication of absence of directional bias in the model. Moreover, the even spread of residuals across the full GHI range confirms stable performance under varying conditions.

The model interpretability is another important requirement. The research has identified how each of the environmental variables contributes to predicting solar irradiance by employing the permutation importance. The extreme applicability of air temperature (T2M) and relative humidity (RH2M) is aligned to atmospheric physics in which the two are important items determining the transmittance of solar radiation. The introduction of additive features such as TEMP_RANGE was introduced to reflect subtle thermal variations that influence solar irradiance, and it probably aided performance.

Also, the validity of the generalisation of the model over time is demonstrated as it managed to predict both the days of high irradiance (e.g., August 2022) and days of low irradiance (e.g., January 2019). This establishes its stability and indicates that it can be trusted to perform effectively in the form of operational forecasting.

The boxplots at the monthly level indicate that the model will adequately use seasonal variation which is a characteristic not well represented or used in data-driven models. This is critical to modelling aspects such as PV system design where an annual solar yield estimates are the driving force to the investment decisions.

Computationally, the chosen ANN structure (128-64 hidden neurons) has an appropriate balance of more dense layers against speed. The early stopping and a simple structure will guarantee that training is neither underfitting, nor excessively complex, that is, that overfitting on training data or sluggish convergence does not occur.

Notably, the model also works across locally available NASA POWER datasets hence it is particularly well-suited to the application in resource-poor areas. The world has many regions that lack the high-quality solar monitoring facilities and this model creates that gap to provide accurate provision by radiation using the open access data on meteorological conditions.

Lastly, a combination of other contextual factors can further be incorporated in this model in future works, where there is an inclusion of factors like solar zenith angle, cloud cover data or satellite-retrieved indices of irradiance. In addition, integrating methods or hybrid models in which ANN would be combined with locally optimal decision trees or SVMs may be used to achieve even better results.

VI. CONCLUSION

In this study, the relevance of ANNs has been established in predicting the actual amount of solar radiations in a day based on meteorological conditions. Through a two-hidden-layer ANN model that was trained on NASA POWER data the study managed to fulfil its accurate predictive potential without complex input satellite data and thick net of ground sensors. The realised measures; 0.754 MJ/m2/day, 0.943 MJ/m2/day and 0.725 highlight the effectiveness of the model in predicting the variability of the global horizontal irradiance (GHI) among seasons and atmospheric variations.

The realisation of generalisation ability of the model in time and its permanent performance in the period of high and low solar irradiance level prove its application in operational forecasting. The stability and the absence of bias, however, are emphasised by the residual analysis and monthly performance plots, which are critical aspects in solar energy planning to occur especially in resource constrained regions. The paper also ascertained the fact that a fairly complicated ANN design, through balanced use of more rates of regularisation, such as early stopping, can produce excellent results without undergoing the danger of becoming over-fit.

The results confirm that ANNs have the potential of being a great replacement to conventional empirical models and especially in solar energy applications. Further research can be based on hybrid ANN-decision tree or ANN-support machine models, and it might be suggested that more specific parameters including satellite-based cloud indices or zenith solar angle may increase the sensitivity. On the whole, this study is a part of emerging knowledge on the application of AI to forecasting and planning in renewable energy.

RESEARCH CONTRIBUTION

The study provides a number of novel contributions in the context of solar radiation forecasting with the application of Artificial Neural Networks (ANNs) and specifically in terms of data efficiency, the robustness of the models and the overall methodological advancement:

➢ *Embedding of Advanced Features Selection Methods:*

The major methodological innovation of the present study is associated with the fact that it introduced a hybrid feature selection pipeline that included filter, wrapper, and embedded feature selection threads. First, it filtered out less informative meteorological variables with the help of the filter approach (SelectKBest with f_regression). This was followed by the recursive feature elimination (RFE) use of the MLPRegressor as estimator in refining the subset. Also, an embedded approach was examined: the MLP removed in favour of the Lasso regression, and, as a result, it was possible to select coefficients and make it easier to interpret the model. This procedure in steps guarantees an initial identification of the most salient qualities of interest before the last stage eliminates redundancies thus having a simple model and its predictive improvement.

➢ *Interpretability of Models through Permutation Importance and Lasso Weights:*

In addition to accuracy, the analysis focuses on model interpretability by use of the permutation importance analysis and weights analysis on Lasso. The tools give an idea of feature relevance so domain experts can get to grips with and be confident with the predictive process of the ANN.

➢ *Efficient Computationally and Scale:*

Its ANN architecture (128-64) hidden neurons with its early stopping when training) has a balance between the speed and accuracy. Its minimal computational overheads can be used in deployable software at real-time or embedded systems to provide solar energy prediction.

In conclusion, the paper adds value to the current interpretable, efficient, and replicable ANN-based architecture approach in predicting solar radiation with the use of a new, hybrid feature selection framework. This leads not only to an increased performance but also in setting out a roadmap towards future studies aimed at integrating machine learning in renewable energy forecasting based on data-efficient pipelines.

REFERENCES

[1]. A. Dindar, S. Ourang and E. Ghadikola, "Development of a communication-assisted adaptive overcurrent protection scheme in smart distribution networks in presence of wind and solar generation.," 2022.
[2]. R. Nematirad and A. Pahwa, "Solar radiation forecasting using artificial neural networks considering feature selection," In 2022 IEEE Kansas Power and Energy Conference (KPEC), pp. 1-4, 2022.
[3]. S. Soleymani and A. Talebi, "Forecasting solar irradiance with geographical considerations: integrating feature selection and learning algorithms," Asian Journal of Social Science, vol. 8, no. 5, 2025.
[4]. M. Hossain, A. Wadi Al-Fatlawi, L. Kumar, Y. Fang and M. Assad, "Solar PV high-penetration scenario: an overview of the global PV power status and future growth," Energy Systems, pp. 1-57, 2024.

[5]. S. Rahman, S. Rahman and A. Haque, "Prediction of solar radiation using artificial neural network," In Journal of physics: conference series. IOP Publishing, vol. 1767, no. 1, p. 012041, 2021.
[6]. T. Woldegiyorgis, N. Benti, M. Chaka, A. Semie and A. Jemberie, "Estimating solar radiation using artificial neural networks: A case study of Fiche, Oromia, Ethiopia," Cogent engineering, vol. 10, no. 1, p. 2220489, 2023.
[7]. M. Ali, A. Elsayed, I. Elkabani, M. Akrami, M. Youssef and G. Hassan, "Optimizing artificial neural networks for the accurate prediction of global solar radiation: A performance comparison with conventional methods," energies, vol. 16, no. 17, p. 61, 2023.
[8]. P. Malik, A. Gehlot, R. Singh, L. Gupta and A. Thakur, "A review on ANN based model for solar radiation and wind speed prediction with real-time data," Archives of Computational Methods in Engineering, vol. 29, no. 5, pp. 3183-3201, 2022.
[9]. F. Nawab, A. Abd Hamid, A. Ibrahim, K. Sopian, A. Fazlizan and M. Fauzan, "Solar irradiation prediction using empirical and artificial intelligence methods: A comparative review," Heliyon, vol. 9, no. 6, 2023.
[10]. S. Al-Dahidi, M. Madhiarasan, L. Al-Ghussain, A. Abubaker, A. Ahmad, M. Alrbai, M. Aghaei, H. Alahmer, A. Alahmer, P. Baraldi and E. Zio, "Forecasting solar photovoltaic power production: A comprehensive review and innovative data-driven modeling framework," Energies, vol. 17, no. 16, p. 4145, 2024.
[11]. A. Assaf, H. Haron, H. Abdull Hamed, F. Ghaleb, S. Qasem and A. Albarrak, "A review on neural network based models for short term solar irradiance forecasting," Applied Sciences, vol. 13, no. 14, p. 8332, 2023.
[12]. K. Iheanetu, "Solar photovoltaic power forecasting: A review," Sustainability, vol. 14, no. 24, p. 17005, 2022.
[13]. Ramirez-Vergara, L. Bosman, W. Leon-Salas and E. Wollega, "Ambient temperature and solar irradiance forecasting prediction horizon sensitivity analysis," Machine Learning with Applications, vol. 6, p. 100128, 2021.
[14]. D. Fomenko, M. Kostoev and A. Vaskov, "Review of Ultra-Short-Term Solar Radiation Prediction Methods," In 2023 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). IEEE, pp. 212-216, 2023.
[15]. H. Nguyen, Q. Tran, C. Ngo, D. Nguyen and V. Tran, "Solar energy prediction through machine learning models: A comparative analysis of regressor algorithms," PloS one, vol. 20, no. 1, p. e0315955, 2025.
[16]. I. Bazionis, M. Kousounadis-Knousen, P. Georgilakis, E. Shirazi, D. Soudris and F. Catthoor, "A taxonomy of short-term solar power forecasting: Classifications focused on climatic conditions and input data," IET renewable power generation, 2023.
[17]. R. Meenal, D. Binu, K. Ramya, P. Michael, K. Vinoth Kumar, E. Rajasekaran and B. Sangeetha, "Weather forecasting for renewable energy system: a review," Archives of Computational Methods in Engineering, vol. 29, no. 5, pp. 2875-2891, 2022.
[18]. M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. Refaat and F. Oueslati, "Enhanced random forest model for robust short-term photovoltaic power forecasting using weather measurements," Energies, vol. 14, no. 13, p. 3992, 2021.
[19]. V. Prema, M. Bhaskar, D. Almakhles, N. Gowtham and K. Rao, "Critical review of data, models and performance metrics for wind and solar power forecast," IEEe Access, vol. 10, pp. 667-688, 2021.
[20]. D. Markovics and M. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," Renewable and Sustainable Energy Reviews, vol. 161, p. 112364, 2022.
[21]. M. Jaber, A. Abd Hamid, K. Sopian, A. Fazlizan and A. Ibrahim, "Prediction model for the performance of different PV modules using artificial neural networks," Applied Sciences, vol. 12, no. 7, p. 3349, 2022.
[22]. A. Kurniawan and A. Harumwidiah, " An evaluation of the artificial neural network based on the estimation of daily average global solar radiation in the city of Surabaya," Indonesian Journal of Electrical Engineering and Computer Science, vol. 22, no. 3, pp. 1245-1250, 2021.
[23]. N. El-Amarty, M. Marzouq, H. El Fadili, S. Bennani and A. Ruano, "A comprehensive review of solar irradiation estimation and forecasting using artificial neural networks: data, models and trends," Environmental Science and Pollution Research, vol. 30, no. 3, pp. 5407-5439, 2023.
[24]. M. Ganotra, N. Gupta, P. Singh, V. Kharpuse, V. Jha and P. Rao, "ANN Technique for the Prediction of Solar Power Generation," In 2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS), pp. 1-6, 2024.
[25]. V. Lim and P. Chaiwiwatworakul, "Artificial Neural Network (ANN) Modeling to Forecast the Solar Radiation on Vertical Planes in Thailand," In 2024 9th International Conference on Business and Industrial Research (ICBIR). IEEE, pp. 1048-1053, 2024.
[26]. B. Abdlrazg and H. Atetalla, "Prediction of solar radiation potential in Libya using artificial neural networks.," International Science and Technology Journal, 2024.
[27]. Z. Kadhim, H. Abdullah and K. Ghathwan, "Artificial Neural Network Hyperparameters Optimization: A Survey," Int. J. Online Biomed, vol. 18, no. 15, pp. 59-87, 2022.
[28]. L. Japa, M. Serqueira, I. MendonçA, M. Aritsugi, E. Bezerra and P. González, "A population-based hybrid approach for hyperparameter optimization of neural networks," IEEE Access, vol. 11, pp. 50752-50768, 2023.