Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 1, No.3, January : 2025

A REGRESSION LEARNING BASED APPROACH FOR FORECASTING SOLAR IRRADIATION FOR RENEWABLE ENERGY APPLICATIONS IN POWER SYSTEMS

Sweta Shalini, Research Scholar, Department of Electrical and Electronics Engineering, TIT& Science, Bhopal.
Saurabh Gupta, Associate Professor, Department of Electrical and Electronics Engineering, TIT& Science, Bhopal.
Vikas Gupta, Professor, Department of Electronics & Communication, TIT, Bhopal.

Abstract: Solar energy and wind power have emerged as two of the most advanced forms of renewable energy in the last several years. The abundance of locations with high irradiation and wind possibilities has led to a recent uptick in funding for the research and development of these two technologies. Solar radiation isn't constant, therefore we need a way to generate energy that's more adaptable and dependable. Energy system operators rely on traditional technologies to meet the grid's demands. In order to obtain a higher penetration of renewable generators, it is necessary to build accurate forecasters in order to lower the unreliability aspects. It is not a new task to predict the power output of renewable energies. Various approaches are presented in the literature. Forecasting relies on extrapolating future values from existing historical datasets. Solar irradiation prediction using data-driven models is one application that has been attracting a lot of attention. Since sun irradiation varies greatly and can even go to zero at night, it might be challenging for machine learning algorithms to track its trend. More issues arise as a result of this break. Therefore, a two-pronged strategy has been employed to forecast solar irradiation: first, by processing the data with the wavelet transform for all the pertinent parameters; second, by training a neural network with the processed data. We will be measuring performance with the mean absolute percentage error. The proposed approach attains an MAPE of 1.37%.
Keywords—Power Systems, Solar Irradiation Forecasting, Movement, Wavelet Transform, Multi-level Decomposition,Gradient Boost, Gradinet Boosting, Accuracy, Regression**.**

1. Introduction

One of the most popular renewable energy sources in the modern period is solar power, which is convenient and adaptable. Solar renewable electricity has the potential to meet the world's expanding energy needs. Even while there are numerous benefits to using solar power, there is a downside as well. The potential quantity of solar energy remains an open question. This is because, in contrast to other renewable energy sources such as wind and tidal power, the amount of solar energy is subject to numerous unpredictable changes. Consequently, accurately estimating and harvesting solar power becomes challenging due to the intermittent and variable nature of solar energy. This makes it difficult to quantify the amount of energy that can be harvested from solar panels at any one moment. The same can be explained by a number of natural phenomena. Harnessing power from the sun becomes pointless and unproductive when the appropriate solar energy measure cannot be accurately established. The system unit's stability, power quality, and longevity could all take a hit as a result. It might also lead to a number of other power-related issues. Recurring variations in sun irradiation power could lead to inaccurate and untrustworthy results. [5] Solar irradiation prediction could be an answer to the dilemma I just described. Predicting future sun irradiation can help achieve an accurate reading of solar energy. It has the potential to demonstrate tremendous benefits associated with electricity generating if put into place from the start. In this respect, the entire grid configuration and its functional units will be of service. The average daily sun irradiation is the primary determinant of the solar energy output. The problem of variable solar irradiation power can be effectively addressed in this way.

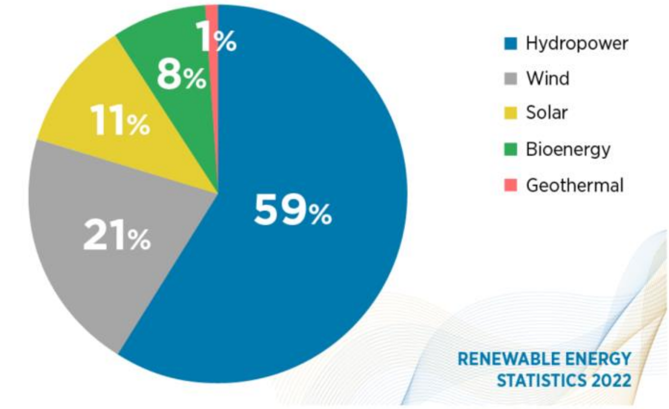

Fig.1 Global Solar Energy Share

A great deal of study has been conducted on the topic of solar irradiation predictions. The methodology utilized to classify solar forecasting methods allows for easy categorization. There are two main types: physical and statistical. Traditional methods mostly use patterns in numerical data sets derived from weather forecasts. The input data is used for the purpose of processing and evaluation after being supplemented with appropriate information in such a scenario. The area of Artificial Intelligence (AI) has been the subject of extensive study and practical application. eight to nine The application of AI in this field is highly advantageous due to the complexity of the task involved. With the help of artificial neural networks, non-linear connections and linkages between the input and output ends can be processed more efficiently. Their processing speed allows them to quickly determine the outcome of a huge data collection. Therefore, our proposed study makes use of the idea of Artificial Intelligence to establish an approach that is both accurate and versatile.

2. The Solar Energy Sector

On a global level, the population has been rapidly increasing. With the global technology revolution and the beginning of digitalization, there have been additional needs for high energy. Over the past few decades, nonrenewable energy sources have been used regularly and are almost to the point of exhaustion. Additionally, they have been involved in atmospheric imbalance and global warming. As a result, renewable energy sources are currently the most viable option for supplying energy needs in a way that is environmentally friendly. As is well known, the sun, environment, and other factors are the main considerations taken into account when forecasting solar power. The main producer of solar power is the sun.

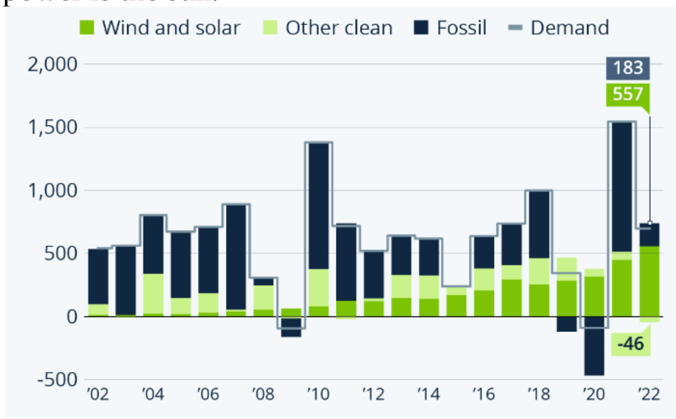

Fig.2 Global Energy Generation Statistics

This type of energy production prediction is essential for accurate solar power estimation purposes, efficient solar power plant management, and a well-distributed solar electricity grid and exports. However, a precise and effective forecasting technique is necessary since solar irradiation has both linear and nonlinear time series aspects. For effective prediction results, a method combining a

combination of neural networks with a holistic approach has been recommended in this study. This comprehensive strategy uses a two-stage feature selection filter that could eliminate undesirable characteristics from the input. It has been observed that the hybrid network executes better than the previous approaches.
The use of power has grown substantially in the past few decades, based on statistics and facts. The demand for power supplies has been far exceeding the amount of power provided globally. This indicates that the power demand has in addition to the capacity of the initial power generation. [10]The increasing rate of consumption globally could possibly be the cause of these significantly increasing electricity requirements. Some non-renewable energy sources have been investigated as potential solutions to these energy and power needs. Consequently, research efforts have been spurred to explore the vast potential of solar and wind energy resources. Therefore, solar power lays the path for supplying the globe's future energy needs.

3. Methodology
The suggested methodology aims to utilize data pre-processing techniques to eliminate noise from the data and provide the machine learning model with an external input in the form of a recent moving average. This will aid in analysing recent patterns. The data processing has been suggested using a dual strategy of noise filtering and dimensional reduction. In order to get the desired outcome, the wavelet transform and principal component analysis will be utilized and optimized for the data model. This approach facilitates the comprehension of how the data patterns persist over a significant period of time, while simultaneously eliminating the disruptive elements of the data. The applied machine learning model is better trained as a result of this creative method. The thorough statistical analysis offers insightful information about how accurate the data used to train the machine learning model was. The learning model is then proposed to use a moving average as an external input. The machine learning system would be able to recognize current patterns in the data and take into account the larger past data by using a moving average. The discrete wavelet transform coefficients from the decomposition and the moving average exogenous input would be included in the final data vector. Then, by adding all the outputs from the several learning models, gradient boosting is proposed as a way to determine the final prediction output. The mathematical modeling of the proposed method is described in detail in the accompanying section.

3.1 Data Processing
A technique for data optimization and dimension reduction in time series forecasting challenges is Principal Component Analysis (PCA). It is beneficial to distinguish between the variable and predictable components of the data in the context of this investigation. The coherent and predictable data samples are effectively separated from the noisy data samples by this separation. Applying Principal Component Analysis (PCA) to the raw data enables the reduction of dimensions, resulting in a more organized and concise data vector with fewer dimensions. This enables the reduction of disturbances in the data by truncating dimensions that have relatively lower coherence. The target vector can be represented by its data representation.

$$T = T(t), T(t-k) \dots \dots \dots T(t-nk) \qquad (1)$$

Here,
T is the composite target vector
t is the time variable
k is the delay or lag variable
n is the number of lags
Multiple governing parameters or characteristics, often nonlinear and non-stationary, influence the target vector. A lower-dimensional representation of the data vector can be obtained by using Principal Component Analysis (PCA), which also removes the uncorrelated data variables that are present in the data's noise spectrum. The data that has been processed through a filter can be represented as:

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 1, No.3, January : 2025

$$T_r = P_r^{n,m} S_r^{n,m} \Delta_r \qquad (2)$$

Here,
T_r is the truncated and dimensionally reduced data vector
$P_r^{n,m}$ is the principal components
$S_r^{n,m}$ is the reduced singular matrix
Δ_r is the PCA matrix
Subsequently, the PCA reduced data will be decomposed by the discrete wavelet transform. The discrete wavelet transform is a discretized variation of the continuous wavelet transform that is employed for the purpose of filtering time series data. The Discrete Wavelet Transform (DWT) functions as a multi-level filter, where scaling and shifting operations of even length serve as high and low pass filtering processes. In contrast to Fourier-based approaches, which are applicable to smooth signals and data patterns because of the limits imposed by Dirichlet's conditions. Nevertheless, the wavelet transform proves to be valuable in the analysis of signals and data that exhibit non-smoothness, non-stationarity, and sudden changes, mostly because of the presence of non-smooth kernel functions. The scaling operation adheres to the following conditions:

$$\sum_{n=0}^{L-1} k_i^2 = 1 \qquad (3)$$

$$\sum_{l=0}^{L-1} k_i^2 + 2n = \sum_{l=-\infty}^{\infty} k_i^2 + 2n = 0 \qquad (4)$$

Here,
$k_i: l = 0 \ldots\ldots L-1$ represents the kernel of the transform
n represents the set of non-zero integers.
An alternate perspective on the DWT is as a recursive pyramidal filter that decomposes the function into its component parts, allowing for the estimation and detailed calculation of coefficient values. The intrinsic patterns in the data, excluding the noisy component, would be contained in the approximate coefficient values, whereas the detailed values would include the noisy disturbances. In order to remove the noisy data, it is possible to apply the DWT iteratively while keeping the approximation coefficient values and rejecting the detailed ones. Although there might be some data loss during the procedure, the benefits would exceed the negative impacts of the data disturbances. It boils down to this, mathematically:

$$X(t) \xrightarrow{S,W} Co-efficients\ (Approx, Detailed) \qquad (5)$$

Here,
Approx. represents the approximate co-efficient values
Detailed represents the detailed co-efficient values
S and W are the scaling and wavelet filters of the DWT
X is the time domain samples of the data

3.2 Training

In order to get at the training data, the raw data is first processed and organized using PCA and DWT. This study makes use of the following data features: date, closing price of the previous day, opening price of the present day, volume (swing), and daily high and low prices. Back propagation-gradient descent is the training algorithm used in this case. As is customary, the neural network is trained using 70% of the data and tested using 30%. For temporal prediction issues, the back propagation method is selected because it produces generally decent results. This method involves training the network with the data vector and then calculating the objective function or error. The following limitations inform the iterative or recursive training process:
a) After a certain number of successive iterations, the values become stationary and the objective function is minimized.
b) We've reached the maximum number of iterations defined, but the objective function hasn't achieved stationarity.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 1, No.3, January : 2025

When you meet one of these two requirements, the instruction will end. The back propagation mostly makes use of the chain rule provided by:

$$w_{k+1} = w_k - \alpha \frac{\partial e}{\partial w} \quad (6)$$

Here,
w_{k+1} is the weight of the next iteration
w_k is the weight of the present iteration
e is the error
α is the learning rate

$$\frac{\partial e}{\partial w} = \frac{\partial e}{\partial y} \cdot \frac{\partial y}{\partial w} \quad (7)$$

The chain rule in relation (7) can be used for computing the error gradient. Summarizing the back propagation algorithm employed, the following steps are to be employed:
The performance evaluation of the proposed model is done based on the evaluation of the following parameters:
1) Mean Absolute Percentage Error (MAPE)
2) Mean square error (MSE)

$$MAPE = \frac{100}{N} \sum_{t=1}^{N} \frac{|V_t - \hat{V}_t|}{V_t} \quad (8)$$

$$MSE = \frac{1}{N} \sum_{t=1}^{N} {e_t}^2 \quad (9)$$

Here,
N is the number of predicted samples
V is the predicted value
$\hat{V}_t$ is the actual value
e is the error value
The next section discusses the obtained results.

4. Experimental Results

Time steps of one hour are used to collect the minded data. The weather station located at the "Solar Radiation Lab (SRL), University of Texas" in the USA gathers data. At 1-hour intervals, a total of 4,230 samples of the aforementioned metrics are at your disposal. The data processing and prediction methodologies are explained subsequently using the obtained results.

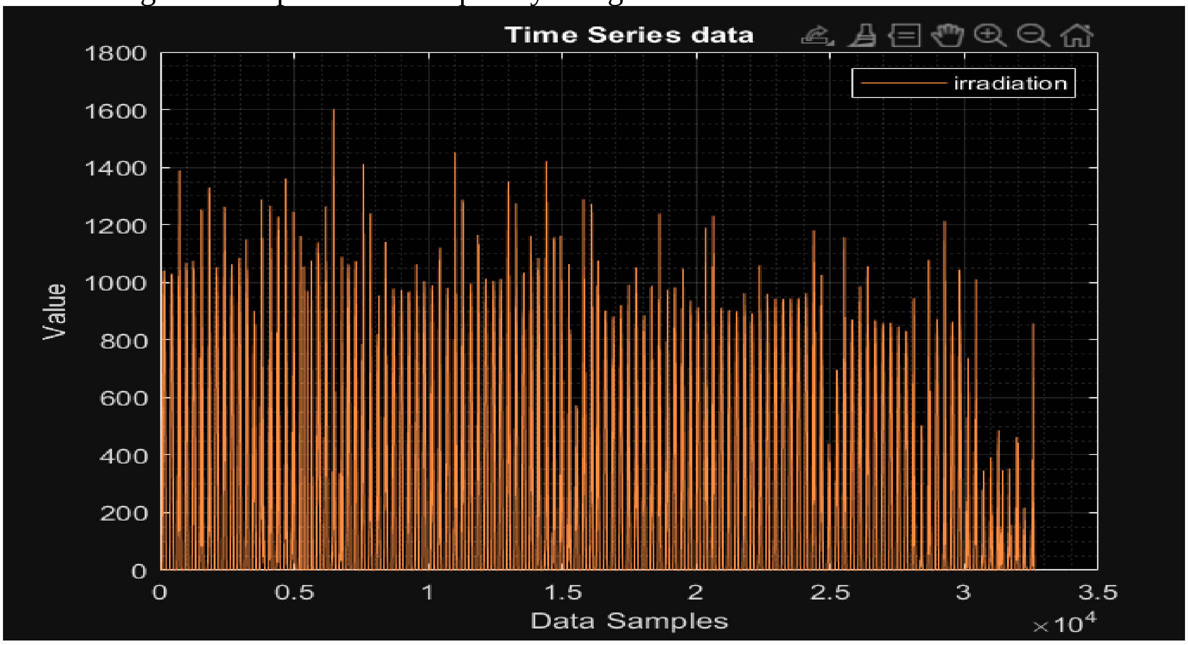

Fig.3 Raw Time Series Data

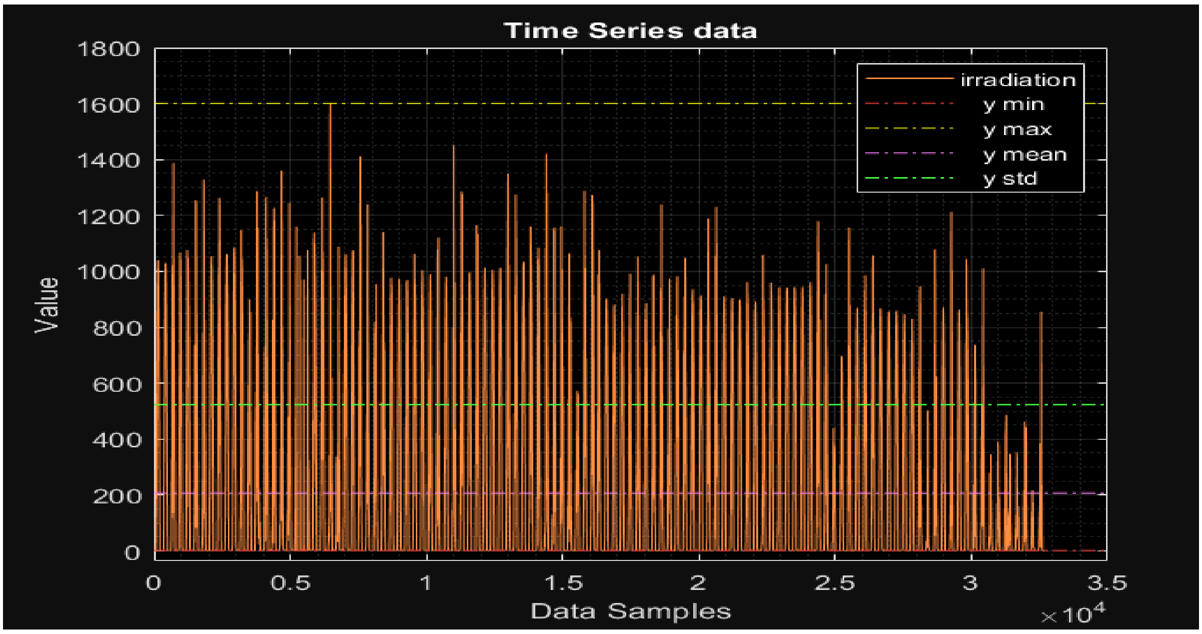

Fig.4 Statistical Measures of Data

Figure 3 depicts the raw data while figure 4 depicts the statistical features of the time series data.

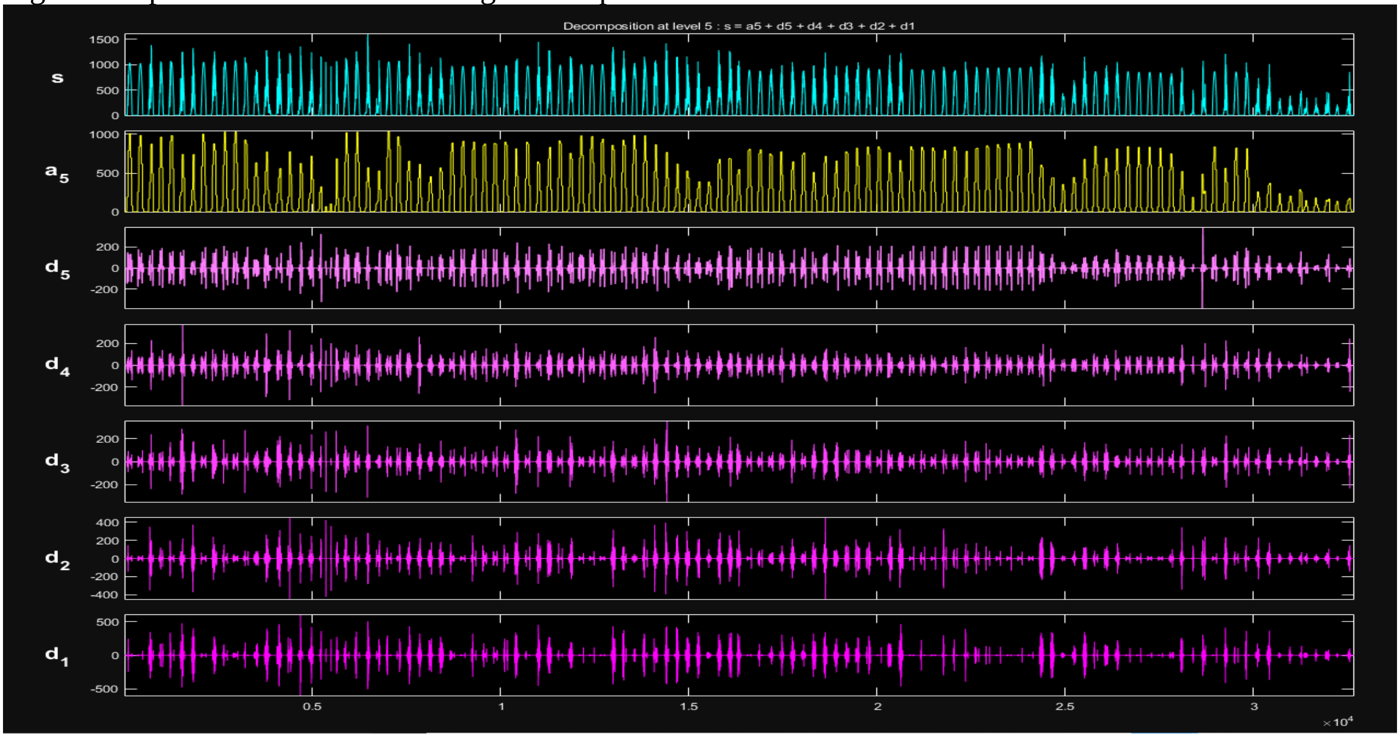

Fig.5 Wavelet Decomposition of the data at level 3 using Haar family

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 1, No.3, January : 2025

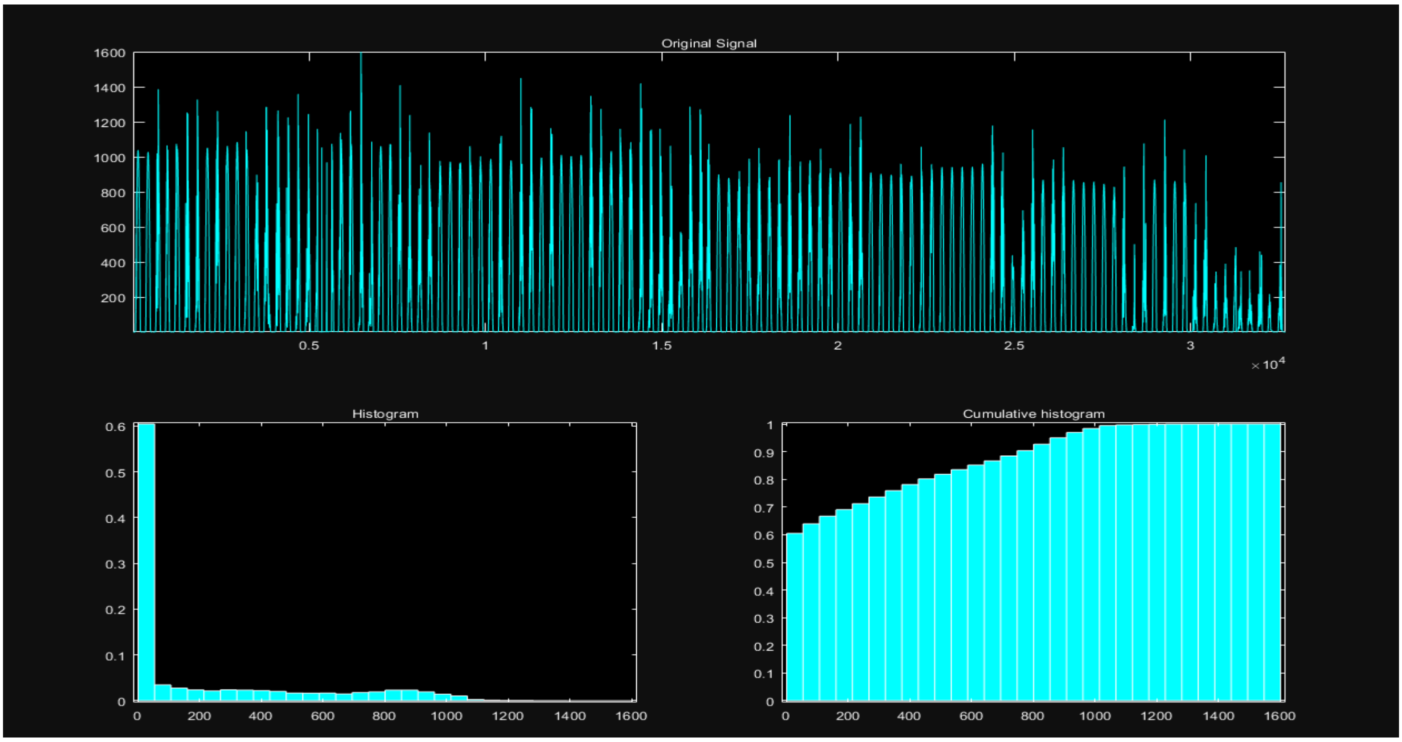

Fig.6 Histogram Analysis of raw data

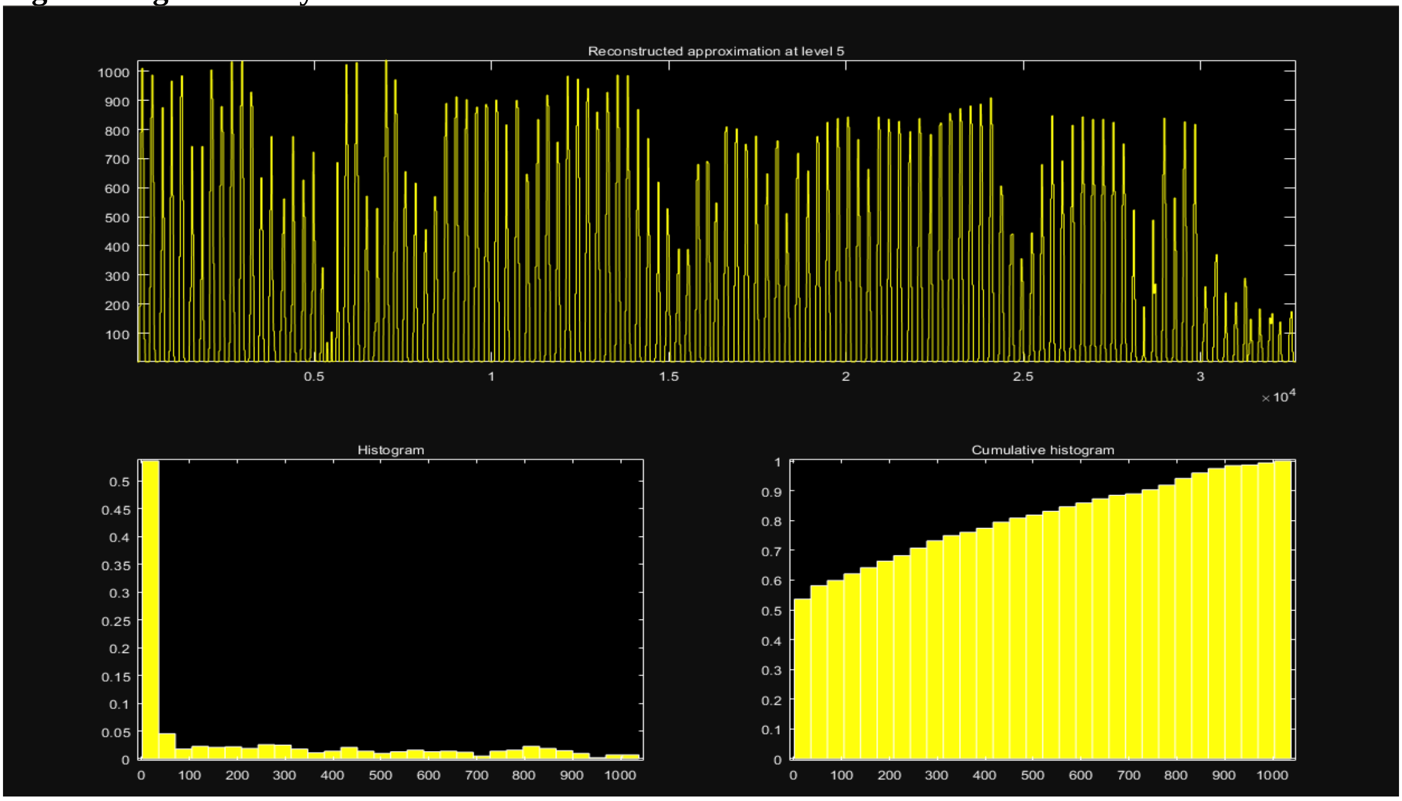

Fig.7 Histogram Analysis of Approximations

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 1, No.3, January : 2025

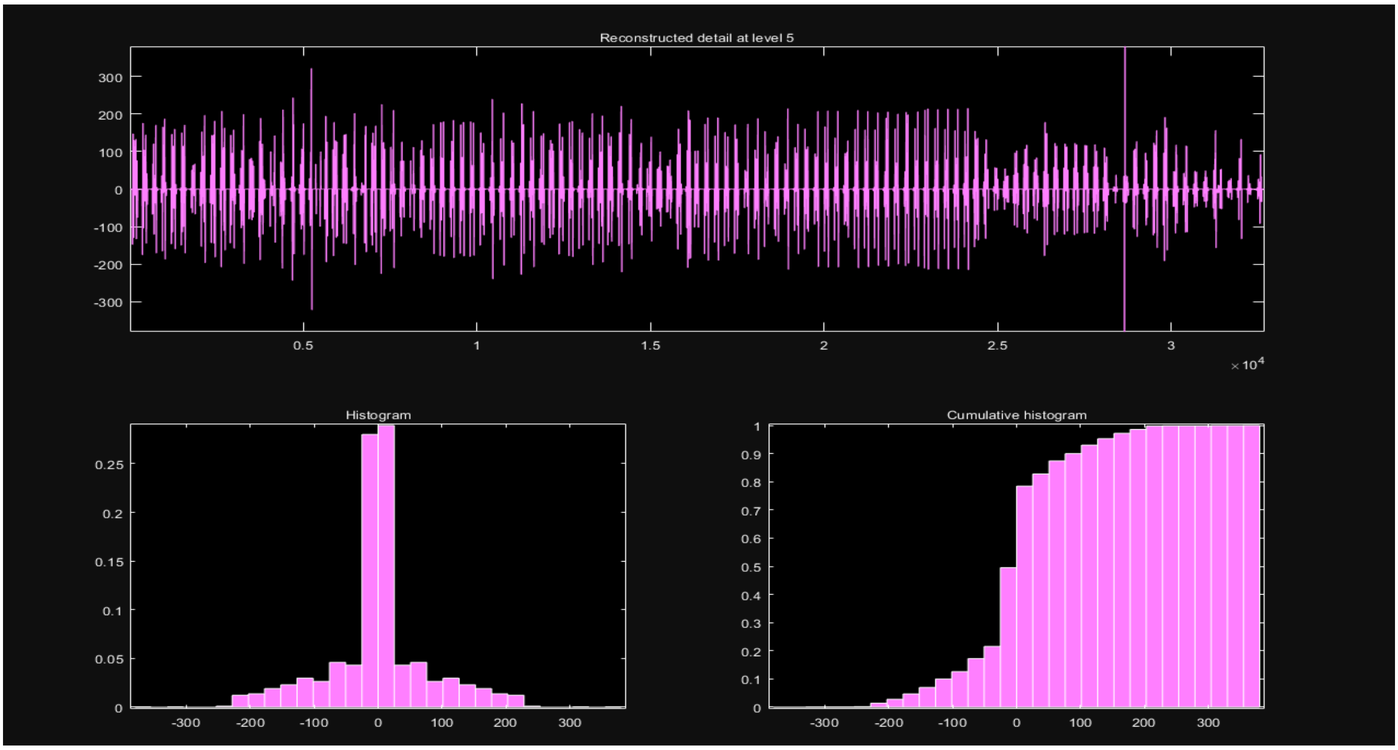

Fig.8 Histogram Analysis of Details

The data is broken down into precise coefficient values and shown in Figures 6, 7, and 8 for statistical analysis and histogram purposes. When comparing the original data or approximate coefficient values to the detailed ones, a noticeable difference is visible in the normal and cumulative histograms. The histogram of the original data and the approximate coefficient values are very similar, as shown in Figures 5, 6, and 7, but the detailed coefficient values differ significantly in size, distribution, and polarity. It is evident from this that the detailed coefficient values are far more affected by the noise and disturbance in the noise floor than the approximation coefficient values.

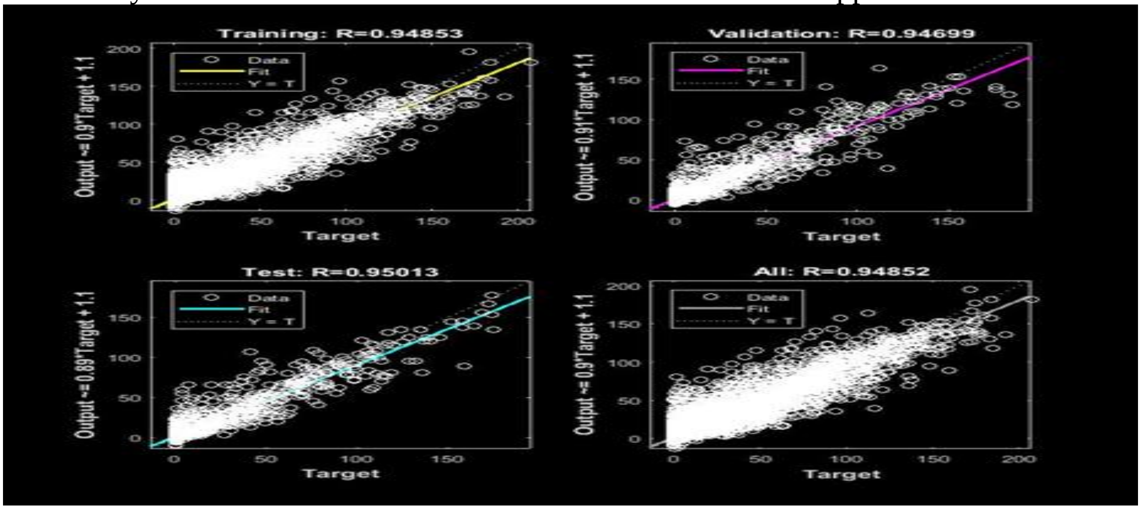

Fig.9 Regression of the Proposed Model

The **R (regression)** values have been depicted in figure 8 for the training, testing, validation and average overall cases. It can be observed that the proposed system attains an average regression of 0.94852. A high value of regression indicates the closeness in the forecasted and actual values.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 1, No.3, January : 2025

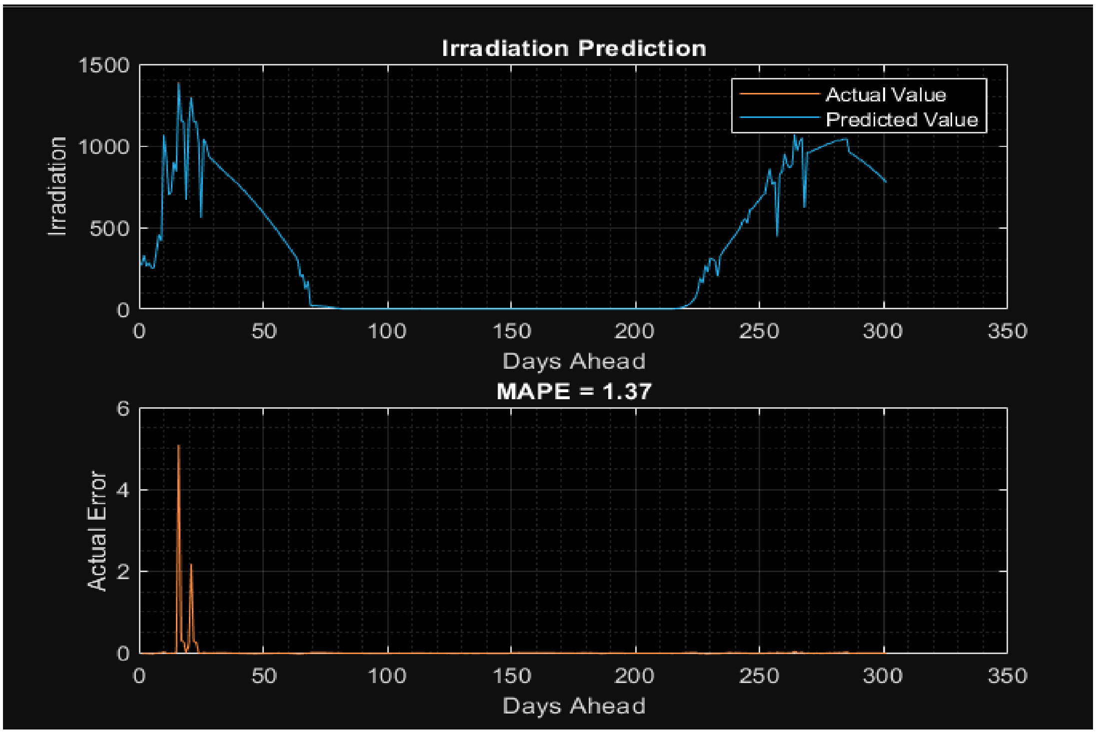

Fig.10 Forecasting MAPE for Model

All three datasets utilized in the study have undergone PCA and DWT decomposition. By comparing the actual and forecasted share prices, we have arrived at the ultimate outcome. The suggested method is correct, as shown by the comparison of the predicted and actual numbers. This study employed the system's mean absolute percentage error (MAPE) to calculate the prediction accuracy. Predicted and real irradiation values are very comparable, according to the regression analysis. The proposed work achieves a MAPE of only 1.37 percent.

Table 1
Summary of Results

S. No.	Parameter	Value
1.	Data	SRL, Texas
2.	Training Regression	0.94853
3.	Testing Regression	0.95013
4.	Validation Regression	0.94699
5.	Overall Regression	0.94852
6.	MAPE	1.37
7.	Accuracy	98.63%

The summary of results is presented in table 1.

5. Conclusion

It can be concluded that we must immediately switch from dirty fossil fuels to green, renewable energy. The sun's rays are the most plentiful and underappreciated energy source currently in existence.

Switching to solar power would necessitate the construction of massive solar farms, which in turn would necessitate substantial investments, making the transition capital demanding. Profitability of the projects depends on predicting, from past data, where solar irradiation will be strong in the near future.With the goal of attaining high forecasting accuracy, this work introduces a data-driven deep neural network model for solar irradiation forecasting.Predicting sun irradiation from hourly data using neural networks is the goal of the proposed approach. Predicting solar irradiation can be difficult due to the large fluctuation in the parameters that influence it, as mentioned above. Since solar irradiation varies greatly and can even go to zero at night, it can be challenging for any neural network system to track its pattern. More issues arise as a result of this break. Therefore, in order to forecast solar irradiation, a two-pronged strategy has been employed, which is incorporates DWT, PCA and deep nets. It is found that the proposed work attains an MAPE of just 1.37%.

Conflict of Interest:
The authors declare no conflicts of interest.

References
[1] G Etxegarai, A López, N Aginako, “An analysis of different deep learning neural networks for intra-hour solar irradiation forecasting to compute solar photovoltaic generators' energy production”, Energy for Sustainable Development, Elsevier, 2022, vol.68, pp.1-17.
[2] Chao Sun, Fengchun Sun, and Scott J. Moura, “Nonlinear Predictive Energy Management of Residential Buildings with Photovoltaics & Batteries”, 2016 ELSEVIER.
[3] Renno C, Petito F, Gatto A. Artificial neural network models for predicting the solar radiation as input of a concentrating photovoltaic system. Journal of Energy Convers Manag. vol. pp.999–1012.
106, Dec 2015.
[4] R. Amaro e Silva, M. C. Brito, “Impact of network layout and time resolution on spatio-temporal solar forecasting”, ELSEVIER 2019.
[5] R. Blaga, A. Sabadus, N. Stefu, C. Dughir, M. Paulescu and V. Badescu, "A current perspective on the accuracy of incoming solar energy forecasting", *Progr. Energy Combustion Sci.*, vol. 70, pp. 119-144, Jan. 2019.
[6] Hanmin Sheng, Jian Xiao, Yuhua Cheng, Qiang Ni, Song Wang, “Short-Term Solar Power Forecasting Based on Weighted Gaussian Process Regression”, 2018 IEEE.
[7] Y. Wen, *et al.* “Performance evaluation of probabilistic methods based on bootstrap and quantile
regression to quantify PV power point forecast uncertainty” IEEE Transact Neural Networks Learn
Syst, vol.31, no. 4, pp. 1134-1144, 2020
[8] T. Carriere, C. Vernay, S. Pitaval, G. Kariniotakis "A novel approach for seamless probabilistic photovoltaic power forecasting covering multiple time frames, "IEEE Trans Smart Grid, vol.11, no.3, pp. 2281-2292, 2020.
[9] L. Visser, T. AlSkaif, W. van Sark "Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution," Renew Energy, vol.183, pp. 267-282, 2022
[10] Wang, F. *et al.* Generative adversarial networks and convolutional neural networks based
weather classi fi cation model for day ahead short-term photovoltaic power forecasting. *Energy Convers. Manag.* Vol.**181**, pp.443–462
[11] Shibani Ghosh, Saifur Rahman, Manisa Pipattanasomporn, “Distribution Voltage Regulation Through Active Power Curtailment With PV Inverters and Solar Generation Forecasts”, 2017 IEEE.
[12] Alrashidi, M., Rahman, S. Short-term photovoltaic power production forecasting based on novel hybrid data-driven models. *Journal of Big Data.* vol.**10**, no. 26, pp.1-25,Mar. 2023.

[13] L. Saad Saoud, F. Rahmoune, V. Tourtchine, K. Baddari “Fully Complex Valued Wavelet Neural Network for Forecasting the Global Solar Irradiation.” Springer 2016.
[14] J. Cifuentes-Faura, "European union policies and their role in combating climate change over the years", *Air Qual. Atmos. Health*, vol. 15, no. 8, pp. 1333-1340, Aug. 2022
[15] Andr´e Gensler, Janosch Henze, Bernhard Sick, Nils Raabe, “Deep Learning for Solar Power Forecasting – An Approach Using Autoencoder and LSTM Neural Networks”, 2016 IEEE.
[16] U. H. Ramadhani, M. Shepero, J. Munkhammar, J. Widén and N. Etherden, "Review of probabilistic load flow approaches for power distribution systems with photovoltaic generation and electric vehicle charging", *Int. J. Electr. Power Energy Syst.*, vol. 120, Sep. 2020.
[17] Milad Fathi, Jafar Amiri Parian, “Intelligent MPPT for photovoltaic panels using a novel fuzzy logic and artificial neural networks based on evolutionary algorithms” Energy Reports, Vol. 7, November 2021, Pages 1338-1348.