

# Solar Radiation Prediction Using LSTM

Mr. Koushik De

Dept. of Computer science and Engineering  
Netaji Subhash Engineering College  
Kolkata, India

Dr. Chandra Das

Dept. of Computer science and Engineering  
Netaji Subhash Engineering College  
Kolkata, India

Dr. Shilpi Bose

Dept. of Computer science and Engineering  
Netaji Subhash Engineering College  
Kolkata, India

**Abstract**— The demand for energy generation from solar energy resource has been exponentially increasing in recent years. It is integral for a grid operator to maintain the balance between the demand and supply of the grid. Solar radiation forecasting paves the way for proper planning, reserve management, and elude penalty since solar energy is sporadic in nature. The two primary categories of methods for predicting the world's solar radiation were scientific estimates & machine learning models. The purpose of this work is to provide a summary of solar radiation prediction in this context using machine learning algorithms. It will be shown that, despite the fact that several studies describe methodology such neural network models or support vector regression, different methods (such as regression model, random forest, XGBoost, etc.) tend to be used in this prediction. In this study, firstly a comparison has been done among different regression algorithms. A new deep neural network based model using RNN and LSTM has been proposed providing more accurate forecast.

**Keywords**—Solar radiation prediction, regression, LSTM, mean square error, mean absolute error

## I. INTRODUCTION

Now a day Solar Radiation carries a huge impact on the globe. We human make pollution over environment, for that global warming comes. **Global warming** causes a barrier on solar radiation, it creates a layer at the atmosphere of our mother earth so that the solar ray cannot radiate back to Sun, and for this our mother earth gets warmer day by day.

As a result, the amount of Solar Radiation gets decrease day by day [1] and for this the associated sectors such as animal husbandry, agricultural firms etc. are performing worsen. Without solar radiation, there would be no life on Earth; moreover, it currently allows us to produce photovoltaic energy, which is essential in the fight against climate change.

However, it can also be harmful to human health; for example, due to its effects on our skin, and these effects have become more dangerous in recent years due to the greenhouse effect, which also influences the rising temperatures on our planet.

Solar radiation is measured on a horizontal surface by means of a radiation sensor or pyranometer, which is placed in a south-facing, shadow-free location. Data are collected in units

of power, watts per square meter ( $\text{W/m}^2$ ), at all weather stations and tend to be taken at ten-minute or 24-hour intervals to establish averages. In the case where it is desired to convert solar radiation from power units to energy units, the data in  $\text{W/m}^2$  must be multiplied by the number of seconds comprising ten minutes (600) or 24 hours (86,400) and the result will be provided in joules per square meter ( $\text{J/m}^2$ ).

Deep learning is a branch of machine learning which is based on artificial neural networks. It is capable of learning complex patterns and relationships within data. In deep learning, we don't need to explicitly program everything. It has become increasingly popular in recent years due to the advances in processing power and the availability of large datasets. Because it is based on artificial neural networks (ANNs) also known as deep neural networks (DNNs). These neural networks are inspired by the structure and function of the human brain's biological neurons, and they are designed to learn from large amounts of data.

The key characteristic of Deep Learning is the use of deep neural networks, which have multiple layers of interconnected nodes. These networks can learn complex representations of data by discovering hierarchical patterns and features in the data. Deep Learning algorithms can automatically learn and improve from data without the need for manual feature engineering.

Deep Learning has achieved significant success in various fields, including image recognition, natural language processing, speech recognition, and recommendation systems. Some of the popular Deep Learning architectures include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Deep Belief Networks (DBNs).

## II. EXISTING METHOD

A. Khosravi et al. [1] proposes machine learning algorithms to predict the hourly solar irradiance. Forecasting models were developed based two types of the input data. The first one uses local time, temperature, pressure, wind speed, and relative humidity as input variables of the models (N1); the second one

is the time-series prediction of solar irradiance (N2) (forecasting models only use from past time-series solar radiation values to estimate the future values). For this purpose, multilayer feed-forward neural network (MLFFNN), radial basis function neural network (RBFNN), support vector regression (SVR), fuzzy inference system (FIS) and adaptive neuro-fuzzy inference system (ANFIS) are developed. The results demonstrated that for the N1, SVR and MLFFNN models have the maximum performance to predict the solar irradiance with  $R = 0.9999$  and  $0.9795$ , respectively. For the N2, SVR, MLFFNN and ANFIS models have reported the correlation coefficient more than  $0.95$  for the testing dataset.

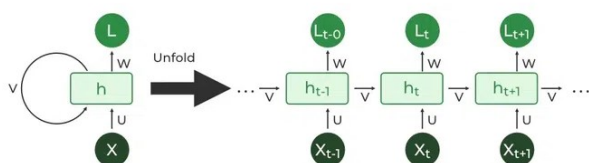
Ümit Ağbulut et al. [2] predict daily global solar radiation data of four provinces (Kırklareli, Tokat, Nevşehir and Karaman) which have different solar radiation distribution in Turkey. In the study, four different machine learning algorithms (support vector machine (SVM), artificial neural network (ANN), kernel and nearest-neighbor (k-NN)[3], and deep learning (DL)) are used. In the training of these algorithms, daily minimum and maximum ambient temperature, cloud cover, daily extraterrestrial solar radiation, day length and solar radiation of these provinces are used. To decide on the success of these algorithms, seven different statistical metrics ( $R^2$ , RMSE, rRMSE, MBE, MABE, t-stat, and MAPE) are discussed in the study. The results shows that  $R^2$ [4], MABE, and RMSE values of all algorithms are ranging from  $0.855$  to  $0.936$ , from  $1.870$  to  $2.328$  MJ/m<sup>2</sup>, from  $2.273$  to  $2.820$  MJ/m<sup>2</sup>, respectively. At all cases, k-NN exhibited the worst result in terms of  $R^2$ , RMSE, and MABE metrics.

Rifat Hacıoğlu[5] shows that machine learning methods which linear regression and Gaussian process regression[6] models are used to estimate the solar radiation on daily data set taken from the wind central in Zonguldak province in Turkey. The measured wind speed, temperature, pressure, humidity parameters together with solar radiation are used for the prediction process. In the prediction process, number of delay steps from 3 to 12 for these parameters are applied to the developed models. In order to determine the performance of the obtained model, the model is evaluated in terms of statistical error criteria such as MAE, MSE and RMSE[7].

### III. PROPOSED METHOD

#### A. Recurrent Neural Network

Recurrent Neural Network (RNN) is a type of Neural Network where the output from the previous step is fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other. The main and most important feature of RNN is its Hidden state, which remembers some information about a sequence.

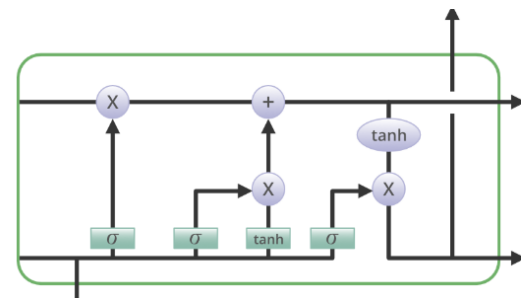


The state is also referred to as Memory State since it remembers the previous input to the network. It uses the same parameters for each input as it performs the same task on all

the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks. The fundamental processing unit in a Recurrent Neural Network (RNN) is a Recurrent Unit, which is not explicitly called a “Recurrent Neuron”. This unit has the unique ability to maintain a hidden state, allowing the network to capture sequential dependencies by remembering previous inputs while processing. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) versions improve the RNN’s ability to handle long-term dependencies.

#### Long Short-Term Memory (LSTM)

Long Short-Term Memory works on the read-write-and-forget principle where given the input information network reads and writes the most useful information from the data and it forgets about the information which is not important in predicting the output.



For doing these three new gates are introduced in the RNN. In this way, only the selected information is passed through the network.

#### B. Equations

The formula for calculating the current state:

$$h_t = f(h_{t-1}, x_t) \text{ where,}$$

$h_t$  -> current state

$h_{t-1}$  -> previous state

$x_t$  -> input state

Formula for applying Activation function(tanh)

$$h_t = \tanh(W_{hh}h_{t-1} + W_{hx}x_t) \text{ where,}$$

$W_{hh}$  -> weight at recurrent neuron

$W_{hx}$  -> weight at input neuron

The formula for calculating output:

$$y_t = W_{hy}h_t$$

$y_t$  -> output

$W_{hy}$  -> weight at output layer

These parameters are updated using Backpropagation. However, since RNN works on sequential data here we use an updated backpropagation which is known as Backpropagation through time.

#### IV. IMPLEMENTATION DETAILS

We have implemented above mentioned models using Python programming language, it is an open source language along with some libraries which are also open source and ready to download and use materials.

##### A. Dataset

Dataset which has been used in this project is "SolarPrediction". It is in comma separated value format (\*.csv).

Dataset has been derived from a thermal power station situated between Kuntighat and Jirat Railway Station under Indian Railway, Eastern Zone, Howrah Division. As both the stations are in Hooghly District that's why all the data comes from Hooghly District only.

**Dimension of Dataset:** 32686 rows  $\times$  11 columns

**Parameters of Dataset:** UNIXTime

Date  
Time  
Radiation  
Temperature  
Pressure  
Humidity  
WindDirection  
Speed  
TimeSunRise  
TimeSunSet

##### B. Preprocessing of Dataset

Unix time is currently defined as the number of non-leap seconds which have passed since 00:00:00 UTC on Thursday, 1 January 1970, which is referred to as the Unix epoch. Unix time is typically encoded as a signed integer. This UNIXTime column has been removed from the dataset as this is un useful for prediction of solar-radiation.

##### C. Figures and Tables

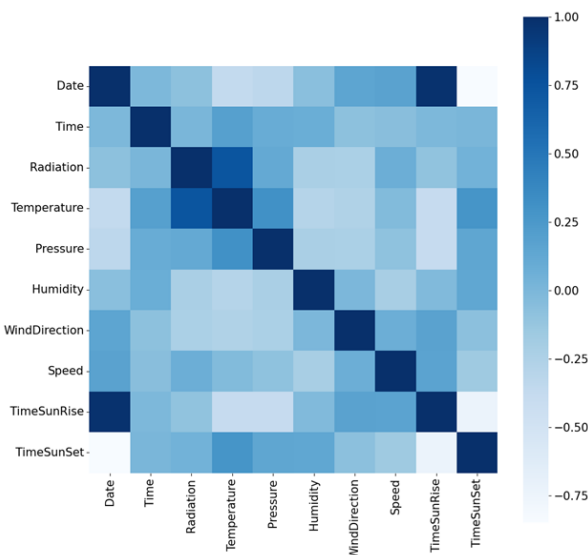


Fig 1: Correlation Heatmap

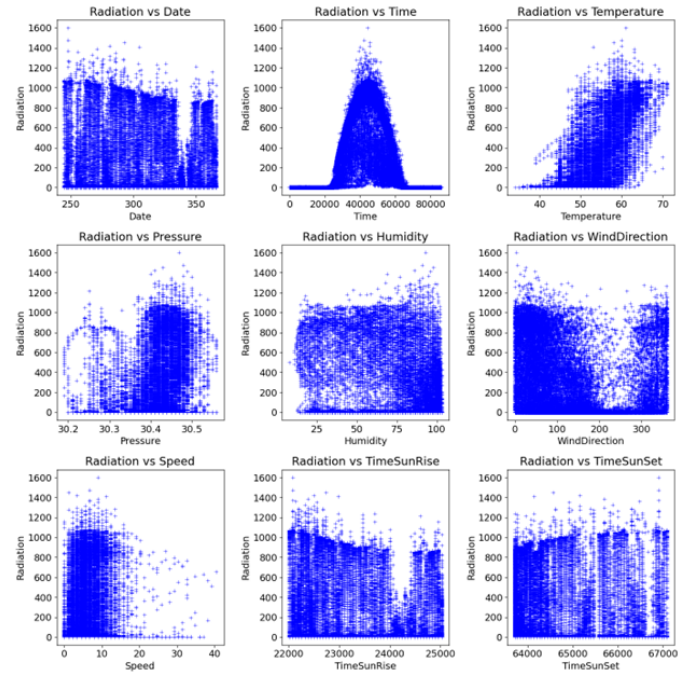


Fig 2: Scatterplot with relation with radiation

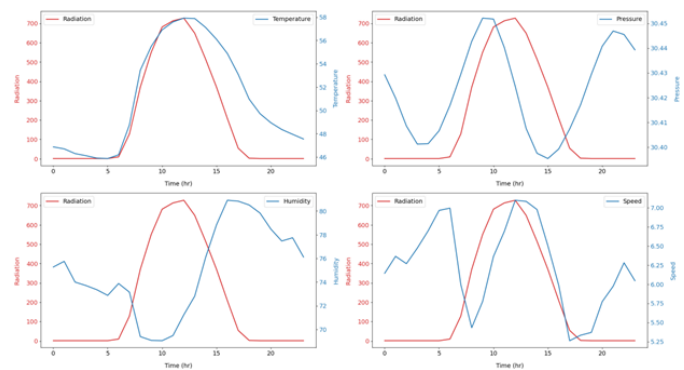


Fig 3: Hourly Average Plots wrt Radiation

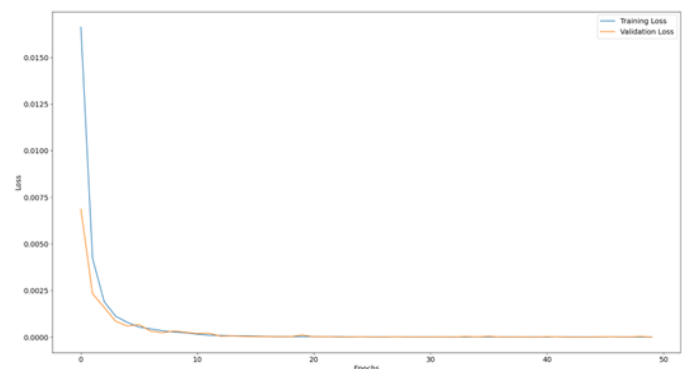


Fig 4: Training Loss vs. Validation Loss

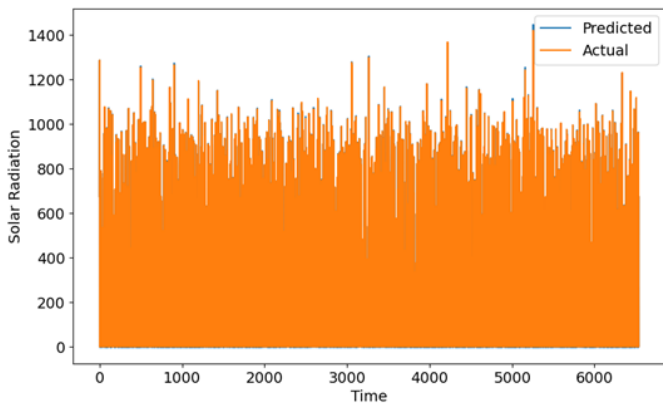


Fig 5: Actual Values vs. Predicted Values

TABLE I. R2 SCORE OF POLYNOMIAL REGRESSION IN 5-FOLD AND 10-FOLD VALIDATION TESTING

POLYNOMIAL REGRESSION		
Test - size	5-fold	10-fold
10	0.6585304957637894	0.6303591931929204
20	0.650480046005116	0.8365516149304206
30	0.8858606383258486	0.8650821891524767
40	0.730219214724724	0.8482285588424687
50	0.23256142153424542	0.6378232370584094

TABLE II. VALIDATION LOSS SUMMARY FOR LSTM

Epoch-size	Validation Loss
10	2.5524e-04
20	1.1212e-04
30	6.5811e-06
40	2.7272e-06
50	3.9534e-06

## ACKNOWLEDGMENT

This entire work has been done by me under supervision of my project guide Prof. (Dr.) Shilpi Bose Sir, I would like pay warm gratitude and a ton of thanks to him.

## REFERENCES

Till now no further changes need to do with the newly developed model, but I want to do a comparative study by fitting my dataset with the existing deep neural networks and check how much accurate result it gives. I have used the below mentioned papers as my reference papers.

- [1] Khosravi, A., et al. "Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms." *Journal of Cleaner Production* 176 (2018): 63-75.
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [3] Ağbulut, Ümit, Ali Etem Gürel, and Yunus Biçen. "Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison." *Renewable and Sustainable Energy Reviews* 135 (2021): 110114.
- [4] K. Elissa, "Title of paper if known," unpublished.
- [5] Hassan, Md Ziaul, et al. "Forecasting Day-ahead solar radiation using machine learning approach." 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE). IEEE, 2017.
- [6] Hedar, Abdel-Rahman, et al. "Hybrid machine learning for solar radiation prediction in reduced feature spaces." *Energies* 14.23 (2021): 7970.
- [7] M Hacıoğlu, Rifat. "Prediction of solar radiation based on machine learning methods." *The journal of cognitive systems* 2.1 (2017): 16-20.
- [8] Voyant, Cyril, et al. "Machine learning methods for solar radiation forecasting: A review." *Renewable energy* 105 (2017): 569-582.
- [9] Solano, Edna S., Payman Dehghanian, and Carolina M. Affonso. "Solar radiation forecasting using machine learning and ensemble feature selection." *Energies* 15.19 (2022): 7049.