

## SOLAR IRRADIATION FORECAST BY DEEP LEARNING ARCHITECTURES

by

**Omer DAGISTANLI<sup>a</sup>, Hasan ERBAY<sup>b</sup>,  
Ahmet Hasim YURTTAKAL<sup>c</sup>, and Hakan KOR<sup>d\*</sup>**

<sup>a</sup>Department of Computer Technology, Bogazliyan Vocational School,  
Yozgat Bozok University, Yozgat, Turkey

<sup>b</sup>Computer Engineering Department, Engineering Faculty,  
University of Turkish Aeronautical Association, Etimesgut Ankara, Turkey

<sup>c</sup>Computer Engineering Department, Engineering Faculty,  
Afyon Kocatepe University, Erenler Afyon, Turkey

<sup>d</sup>Department of Computer Engineering, Engineering Faculty,  
Hitit University, Corum, Turkey

Original scientific paper

<https://doi.org/10.2298/TSCI2204895D>

*Global solar irradiation data is a crucial component to measure solar energy potential when we plan, size, and design solar photovoltaic fields. Often, due to the absence of measuring equipment at meteorological stations, data for the place of interest are not available. However, solar irradiation can be estimated by ordinary meteorological data such as humidity, and air temperature. Herein we propose two different deep learning methods, one based on a deep neural network regression and the other based on multivariate long short term memory unit networks, to estimate solar irradiation at given locations. Validation criteria include mean absolute error, mean squared error, and coefficient of determination ( $R^2$  value). According to the simulation results, multivariate long short term memory unit networks performs slightly better than deep neural network. Even though both have very close  $R^2$  values, multivariate long short term memory's  $R^2$  values are more consistent. The same is true for mean squared error and mean absolute error.*

*Key words: solar irradiation, deep learning, estimation*

### Introduction

The term *energy personality* refers to the daily activities of individuals related to energy consumption. According to Opower [1], a Virginia-based company, there are five *energy personalities* that represent five separate lifestyles people tend to pursue, including Daytimers, Evening Peakers, Steady Eddies, Night Owl, and Twin Peaks. The company works with energy firms to help them better connect with their customers and potentially change their customers' behavior. Based on individuals' energy personalities, the company suggests some energy efficiency programs to save money and to reduce energy use. The adoption of sustainable energy technologies by people is seen as one of the biggest factors that will ensure sustainable development. Sunlight is one of the most important sustainable and renewable energy sources [2, 3]. Today, it is known that academic studies have been conducted on the effects of individuals' personality traits on career choice, academic success, social communication and energy

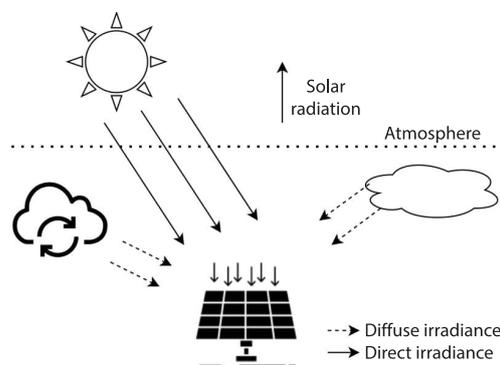
\* Corresponding author, e-mail: hakankor@hitit.edu.tr

usage habits [4, 5]. In addition, governments are investing in green technologies such as solar cells, smart lamps, and electric vehicles to achieve a cleaner and healthier environment. Moreover, worldwide energy consumption, especially electricity consumption, continues to increase faster than ever before. Electricity consumption occupies the leading position among others. On the other hand, global energy demand is expected to reach about 660 quadrillions BTU in 2050, up 15% increase compared to 2019. Among electricity demand is supplied primarily by wind, solar, natural gas-fired generation, and nuclear. Governments encourage companies to generate electricity from renewables including hydro power, wind, and solar power system. Solar irradiation values are crucial when we plan solar power systems. Moreover, architects utilize solar irradiance when they design cooling and passive heating systems for buildings. There are various components of solar irradiance such as global horizontal irradiance (GHI), diffused horizontal irradiance (DHI), and direct normal irradiance (DNI). The GHI value is of particular interest to photovoltaic installations. The GHI represents the total solar radiation incident on a horizontal surface. So, solar power firms use Solar GHI value to compute flat-panel photovoltaic output. The GHI value is determined:

$$GHI = DNI \times \cos \theta + DHI \quad (1)$$

where  $\theta$  is the solar zenith angle (SZA), which is the angle between the local zenith (that is just above the point on the ground) to the line of sight from that point to the Sun. It can be calculated by using:

$$\cos \theta = \sin \delta \sin \phi + \cos \delta \cos \phi \cos \omega \quad (2)$$



**Figure 1. Diffuse irradiance and direct irradiance**

led computerized decision systems to spread diverse fields, such as medicine [6], agriculture [7], and [8, 9]. Regarding the content of the current study, different approaches such as fuzzy model, artificial neural network, regression model, empiric model, and long short term memory (LSTM) unit network are used for solar irradiance predictions [10]. Using the LSTM to model hourly irradiance data not only captures the dependence between consecutive hours of the same day but long-term (*e.g.* seasonal) behavior can be learned. Once the LSTM network is trained and tuned, it can be used to estimate daily solar irradiance values using weather forecasts [11, 12]. Herein, two models are proposed to predict solar irradiance's components by utilizing ordinary meteorological data such as humidity, wind, cloudiness, and air temperature at certain locations. One model is based on deep neural network regression (DNN) while the other is based on multivariate long short term memory (M-LSTM) unit networks. The simulation results indicate that both DNN and M-LSTM produce strong results, but M-LSTM performs slightly better than

where  $\delta$  is the declination of the Sun,  $\omega$  – the hour angle, and  $\phi$  – the latitude (characterized as a positive value in the northern hemisphere). The latter is a measure of local time, that is, defined as the angle the Earth must rotate to bring the meridian of the observatory directly under the Sun. Solar declination is a function of the day of the year and it is independent of location. Its value ranges from  $23^{\circ} 27'$  on June 21 to  $23^{\circ} 27'$  on December 22. See fig. 1 for visual explanation.

On the other hand, the introduction of deep learning in the artificial intelligence field along with advances in computational power

DNN. For the test set, the  $R^2$  values for the triplet (Clearsky DHI, Clearsky DNI, Clearsky GHI) are (0.948, 0.987, and 0.999) relative to DNN and (0.987, 0.989, and 0.988) relative to M-LSTM. As can be seen, the  $R^2$  values of M-LSTM are more consistent compared to DNN ranging from 0.948-0.999. Similar observation is true for both mean absolute error, mean squared error.

### Related studies

Increasing demand for solar energy requires accurate solar resource estimation optimize energy management and grid operation. Observational models are widely used for estimating solar radiation over a short-term time period ranging from five minutes to six hours [13]. The simplest form of these models, which takes advantage of the time persistence of a variable, has often been accepted as a reference for evaluating more advanced models [14]. Machine learning models and statistical techniques are used to improve solar radiation prediction [15-17]. Most such models are based on statistical relationships between prediction and observation, with empirically determined weights that are difficult to interpret in a clear physical sense. Moreover, although different attempts have been made to extract DNI or DHI, observational models are generally limited to estimating GHI [18-20].

Solar resource assessments can be performed with a combination of ground-based and satellite-derived weather data. Optimally, correlating satellite and ground-based weather data with photovoltaic system performance requires knowledge of four different quantities: GHI, DNI, DHI, and plane of array (POA) radiation. High accuracy instrumentation measure all four quantities can be used in weather stations for utility scale projects [21].

Gustafson *et al.* [22] used two models to estimate DNI for solar concentration or tracking photovoltaic applications. The first of these are broadband luminescence models that predict DNI under open skies from atmospheric data. Second, the methods used to extract DNI from spherical GHI. While the second model has had multiple validation studies, the first has only been evaluated in a dozen or less local studies, mostly in the US [23, 24]. Finally, a global validation of the REST2 direct radiation model is recommended against both GHI and DNI at more than 100 locations worldwide.

In their study, Liu *et al.* [25] studied four new physics-informed models to predict GHI, DNI, and DHI in a similar way. Ten years of measurements (1998-2014) at the US Department of Energy's Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) Central Facility site were used to evaluate the performance of new models. Simple and smart continuity models improve forecast accuracy with lead times from 1.25-6 hours. It has been stated that the best model for estimating different radiative components is *The relationship between solar radiation and cloud properties*.

Accurate estimation of solar energy is of great importance for photovoltaic based power plants. Srivastava and Lessmann [see in 26] use LSTM as the estimation method in their study. Data were collected by remote sensing in 21 different locations, 16 in Europe and five in the USA. The results showed that the LSTM was more successful than other models with a 52.2% success rate.

Pazikadin *et al.*, [27] also estimate solar energy production in their study. They apply artificial neural network (ANN) in this study. His publications from February 1, 2014 to February 1, 2019 are examined. Selected articles were obtained from the databases of Direct Science, IEEE Xplore, Google Scholar, MDPI, and Scopus. The study showed that the use of ANN has increased in publications on solar power generation, and again, the use of hybrid ANN has revealed the argument that it is more successful than individual models.

In a study by Zang *et al.*, [28] they also discussed solar radiation estimation with a spatio-temporal correlation model based on deep learning. First of all, convolutional neural network (CNN) was applied to the meteorological parameters related to the targeted region and its neighbors in the study. Then, LSTM was applied to the past solar radiation series data. The LSTM is for the extraction of temporal features. The space-time relationship is combined to predict spherical horizontal radiation. The results determined that the proposed hybrid model has advantages over other models. Benali *et al.* [29] also compare the three components of solar radiation with memorability, ANN, and random forest algorithm in estimating the radial horizontal, ray normal and bright horizontal values. The aim is to estimate hourly solar irradiance for the time periods  $h + 1$  to  $h + 6$ . The random forest (RF) method produces the most successful result. Compared to ANN and smart persistence (SP), the RF usage prediction horizon increases. However, Verbois *et al.* [30] propose a new method for estimating solar radiation. This method uses principal component analysis to combine a High Resolution medium scale numerical weather prediction (NWP) model with a quantile gradient boosting algorithm (QGB). For this process, the Lasso (lasso) model and the analog ensemble (AnEn) model are the methods used as a comparison. The study determined that the newly proposed model, the QGB model, was at least as successful as the Lasso model and less biased. It has also been shown in the study that QGB is at least as competitive as AnEn.

On the other hand, investments in renewable energy facilities significantly reduce GHG emissions. These investments are also influenced by individual psychological factors. In their study, Basic-Sontic *et al.* [31] try to determine whether the household's investment in renewable energy facilities is related to the big five personality traits with the data obtained from the German Socio-Economic Panel ( $N = 3.468$ ). By using a representative sample in Germany, it has been tried to determine how the personality traits that reduce the environmental impact have a relationship with the houses with RES. In the study, openness and neuroticism were found to have a weak positive link with renewable energy facilities through the environmental concern channel, while extraversion had a weak negative link.

In addition, improvements in forecasting could limit the impact of fluctuations in solar power generation, especially on cloudy days. Achleitner *et al.* [32] propose the detection system SIPS, which uses wireless sensor networks (WSN) to detect solar radiation in solar power generation prediction. The study evaluates the findings of a hierarchical WSN system consisting of 19 TelosB nodes with solar radiation sensors and five MicaZ nodes with GPS cards placed near a on 1 MW solar array. Using SIPS data, the success of different forecasting methods has been tested and the existence of spatio-temporal cross-correlation between sensor node readings and solar array output power has been determined.

## Materials and methods

### *Solar irradiance*

Solar irradiation represents the amount of electromagnetic radiation received from the sun per unit area, namely, the amount of the Sun's power detected by a measuring instrument. When the data is measured over time, the information is called solar irradiation. The solar irradiance data varies depending on how far the object is from the Sun, the angle of the Sun, and the solar cycle, that is, the change in the Sun's appearance and activity every 11 years. Solar irradiance, in addition planning solar power systems and designing passive heating and cooling systems for buildings, plays a key role in climate modelling and weather forecasting. Scientists combine irradiation data with other quantitative measurements to answer the relationship between them and develop predictive models. Below we describe the main quantitative measurements in solar irradiance.

*Global horizontal irradiance.* The GHI is the total amount of shortwave radiation received from above by a surface which is horizontal to the ground. The value is of particular interest to photovoltaic installations. It is the most important parameter for calculation of photovoltaic electricity yield.

*Direct normal irradiance.* The DNI is the amount of light that is coming perpendicular to surface. The surface here represents ground or something parallel to ground. This type of irradiance belongs to rays that come in a straight line from the direction of the Sun at its current position in the sky. This quantity is of particular interest to concentrating solar thermal installations and installations that track the position of the Sun. Solar collectors, panels maximize this DHI by means of tilting or rotating with angle of Sun.

*Diffused horizontal irradiance:* The DHI is the solar radiation that does not arrive on a direct path from the Sun, but has been scattered by clouds and particles in the atmosphere and comes equally from all directions. As a footnote, here is the correct place to mention that solar energy companies utilize Solar GHI to compute flat-panel photovoltaic output. Equation one describes GHI in terms of DNI and DHI.

### Dataset

The dataset used in the experiment was first presented in a challenge organized by an Indian multinational company providing information technology, consultancy, and business process services. It is open access under a creative commons (CCO) license on the Kaggle platform, a community of data scientists and machine learning practitioners [33]. The dataset consists of 175296 rows and 18 columns. The dataset characterization is given in tab. 1.

**Table 1. The dataset characterization**

Features	Values
Year	2009-2018
Month	1-12
Day	1-31
Hour	0-23
Minute	0-30
Temperature	-9-40 °C
Clearsky DHI	0-565 w/m <sup>2</sup>
Clearsky DNI	0-1040 w/m <sup>2</sup>
Clearsky GHI	0-1063 w/m <sup>2</sup>
Cloudtype	0 – Clear, 1 – Probably clear, 2 – Fog, 3 – Water, 4 – Super-cooled water, 5 – Mixed, 6 – Opaque ice, 7 – Cirrus, 8 – Over lapping, 9 – Overshooting, 10 – Unknown, 11 – Dust, 12 – Smoke, 15 – N/A
Dew point	-17-26.9
Fill flag	0 – N/A, 1 – Missing image, 2 – Low irradiance, 3 – Exceeds clearsky, 4 – Missing cloud properties, 5 –Rayleigh violation
Relative humidity	7.19-100%
SZA degree	8.55-171.55°
Pressure	970-1024 mbar
Precipitable water	0.122-7.70 cm
Wind direction degrees	0-360°
Wind speed	0-12.9 m/s

**Models**

Within the scope of the study, we developed two models based on DNN and M-LSTM architectures. This subsection briefly describes the structure of these base architectures.

*Deep neural network regression*

The purpose of DNN is to calculate the convergence of a function  $f(x)$ . For an example regression model, input  $x$  is passed to the value  $y$  using  $y = f(x)$ . The feedforward network is defined by the formula  $y = f(x, w)$ . Learns the weight parameter  $w$  which provides the best approximation. The output of the model has no feedback links fed into it. The model consists of functions linked like a chain,  $f(x) = z_3\{z_2[z_1(x)]\}$ . In this case,  $z_1$  is the input layer of the network,  $z_2$  is the hidden layer, and  $z_3$  is the output layer of the network. The number of hidden layers gives the depth of the model. This is where the concept of depth in a DNN comes from [34]. The DNN structure employed in the current study is given in the fig. 2.

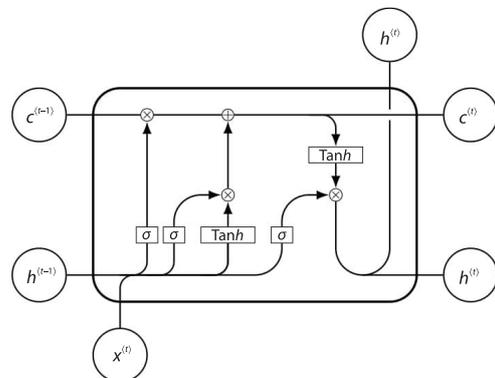
Layer (type)	Output Shape	Param #
dense (Dense)	(None, 256)	9984
dense_1 (Dense)	(None, 128)	32896
dense_2 (Dense)	(None, 64)	8256
dense_3 (Dense)	(None, 32)	2080
dense_4 (Dense)	(None, 1)	33

-----  
 Total params: 53,249  
 Trainable params: 53,249  
 Non-trainable params: 0

**Figure 2. The DNN**

*Multivariate long short term memory unit networks*

The traditional LSTM networks are examples of recurrent neural networks and are capable of learning order dependence in sequence prediction problems, making them favorable in time series problems. Hochreiter and Schmidhuber [11] introduced LSTM as a solution the vanishing gradient problem by letting LSTM units allow gradients to flow unchanged. The LSTM addresses the long-term dependency problem. The core ingredients of an LSTM network are a sequence input layer and an LSTM layer. The sequence input layer inputs sequence data into the network, but the LSTM layer learns long-term dependencies between the steps of sequence data. The LSTM layers are equipped with gating mechanisms controlling the memorizing process, where information can be stored, written, or read via gates that open and close. The LSTM layers contain three gates such as forget gate, input gate, and output gate.



**Figure 3. The LSTM**

The LSTM layers contain three gates such as forget gate, input gate, and output gate.

Figure 3 illustrates the components of an LSTM layer.

The equations that govern the content of the gates are:

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}ct - 1 + b_i) \quad (3)$$

$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}ct - 1 + b_f) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \mathcal{G}(W_{cx}x_t + W_{cm}m_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}ct + b_o) \quad (6)$$

$$m_t = o_t \odot \mathcal{h}_{(c_t)} \quad (7)$$

Note that LSTM layers include neurons with memories that store the past data. The stored data is shaped during the training process [11]. The LSTM provide successful solutions, especially for time series problems. Also, as in this study, besides the time data, there are additional parameters such as temperature, pressure, and humidity employed because these parameters also affect the Irradiation variable. When additional parameters, the method is called M-LSTM. Recall that in time series problems, if the present time is  $t$ , then the future time is represented by  $t + 1$ . Meaning, all the variables in the dataset are transformed into a supervised learning problem by shifting 1 Step forward to create multivariate input and output sequences.

### Performance metrics

We used three metrics to measure performance of the models such that mean absolute error (MAE), mean squared error (MSE), and  $R^2$  value. The MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. In other words, MAE is the average of the absolute differences between the predicted value and the corresponding observation over the dataset. Mathematically speaking, let  $y = (y_1, y_2 \dots y_n)$  be observed values and  $\hat{y} = (\hat{y}_1, \hat{y}_2 \dots \hat{y}_n)$  be predicted values at  $x = (x_1, x_2 \dots x_n)$ . Then MAE is given:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (8)$$

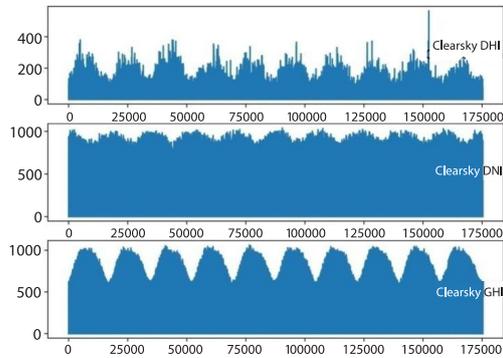
where MSE measures the amount of error in prediction models. It assesses the average squared difference between the observed values and predicted values, mathematically, it is given:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (9)$$

The coefficient of determination,  $R^2$  value, summarizes the proportion of variance in the dependent variable associated with the predicted (independent) variables, with larger  $R^2$  values indicating that more of the variation is explained by the model, to a maximum of 1:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (10)$$

where  $\bar{y}$  is the mean of the observed values.

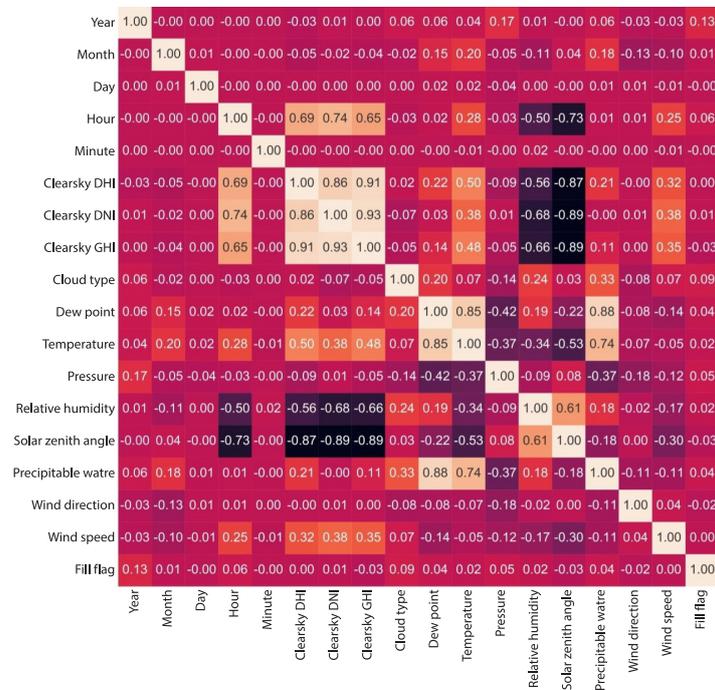


**Figure 4. The DHI, DNI, and GHI values variations**

**Result and discussion**

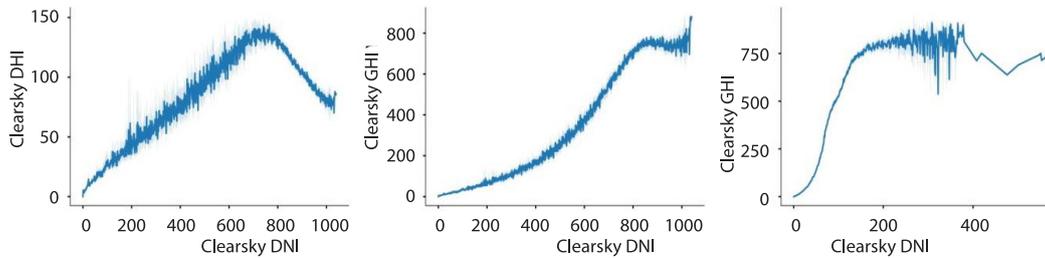
*Exploratory data analysis:* The data set was analyzed before the models were built. The main goal was to better understand the patterns in the data and find interesting relationships between the variables [35]. The variation of DHI, DNI, and GHI values over time is given in fig. 4. One can observe from the figure that the intensity of solar radiation increases in summer. On the other hand, the relationship of the variables with each other affects the regression result. In the correlation matrix given in the fig. 5, the values are positioned between  $-1$  and  $+1$ .

Those values close to  $-1$  are interpreted as negative correlations, and those close to  $+1$  are interpreted as positive correlations [36]. According to the figure, there is high negative correlation between SZA and Clearsky GHI-Clearsky DNI, while there is a high positive correlation between precipitable water and dew point. There is also a high positive correlation between Clearsky DNI, Clearsky DHI, and Clearsky GHI. Trend analysis is a statistical technique that helps to interpret the trend of long-term data in the current state and future situation, and helps to find the line or curve suitable for the series in time series showing a general trend in a particular direction. Output parameters were visualized with line plots in order to analyze the trend and observe the sequentiality of the observations. Trend analysis graphs of target variables are given in the fig. 6. According to the graph, when Clearsky DNI reaches



**Figure 5. Correlation matrix**

800 w/m<sup>2</sup>, Clearsky DHI and Clearsky GHI also reach their maximum values. The Clearsky GHI value increased until the Clearsky DHI reached approximately 300 w/m<sup>2</sup>.



**Figure 6. Trend analysis; (a) Clearsky DNI vs. DHI, (b) Clearsky DNI vs. GHI, and (c) Clearsky DHI vs. GHI**

*Regression performance:* The study was developed in the Python environment. Before the training phase, 80% of the dataset is reserved for training (140236 samples) and 20% for testing (35060 samples). In addition the time data, the dataset has input parameters such as temperature and pressure that can affect the output parameters. The problem is multivariate time series estimation in terms of its structure. The dataset was analyzed using DNN and M-LSTM. The number of learnable parameters of the proposed DNN model was 53249. The main hyperparameters of the model are presented at tab. 2. The success of the DNN model was evaluated according to MAE, MSE and  $R^2$  values. Performance metric values are given in tab. 3. According to the results obtained, the estimation of Clearsky GHI values for the test set were highly predicted with  $R^2$  values of 0.948 for DHI, 0.987 for DNI, and 0.999. Moreover, MSE values for the test set were 0.001 for DHI, 0.002 for DNI, and 0.000 for GHI.

**Table 2. Hyperparameters**

Hyperparameters	Values
Loss function	Mean squared error
Optimizer	Adam
Learning rate	0.001
$\beta_1$	0.9
$\beta_2$	0.999
$\epsilon$	$e^{-7}$
Epoch	200
Batch	256
$L_1$ regularization	$e^{-5}$
$L_2$ regularization	$e^{-4}$
Early stopping	25

On the other hand, the M-LSTM model was also trained with the same dataset. The number of learnable parameters of the proposed M-LSTM model was 73.851. During the simulation experiments, it was observed that the performance decreased as the depth increased in the M-LSTM model. Therefore, getting the optimum number of layers was also a challenge. The proposed M-LSTM consists of two LSTM Layers and a Dense Layer. The M-LSTM structure used in the present study is given in fig. 7.

**Table 3. Comparison of experimental results of DNN**

Train performance metrics	Clearsky DHI	Clearsky DNI	Clearsky GHI
<i>MSE</i>	0.001	0.002	0.000
<i>R</i> <sup>2</sup>	0.961	0.989	0.999
<i>MAE</i>	0.011	0.019	0.004
Test performance metrics	Clearsky DHI	Clearsky DNI	Clearsky GHI
<i>MSE</i>	0.001	0.002	0.000
<i>R</i> <sup>2</sup>	0.948	0.987	0.999
<i>MAE</i>	0.012	0.021	0.004

```

Layer (type)                Output Shape                Param #
-----
lstm (LSTM)                  (None, 3, 100)             43600
lstm_1 (LSTM)                 (None, 50)                  30200
dense (Dense)                 (None, 1)                   51
-----
Total params: 73,851
Trainable params: 73,851
Non-trainable params: 0

```

**Figure 7. The architecture of the M-LSTM model**

During fine tuning process of M-LSTM model, it was observed that the error rate increased when the layers were increased. Thus, too deep network architecture was not preferred. While the epoch number was determined as 200, EarlyStopping patience value was set to 10. The M-LSTM model's results obtained were given in tab. 4.

**Table 4. Comparison of experimental results of M-LSTM**

Train Performance Metrics	Clearsky DHI	Clearsky DNI	Clearsky GHI
<i>MSE</i>	0.000	0.000	0.000
<i>R</i> <sup>2</sup>	0.986	0.989	0.986
<i>MAE</i>	0.017	0.015	0.017
Test performance metrics	Clearsky DHI	Clearsky DNI	Clearsky GHI
<i>MSE</i>	0.000	0.000	0.000
<i>R</i> <sup>2</sup>	0.987	0.989	0.988
<i>MAE</i>	0.016	0.015	0.016

According to the table, the estimations to Clearsky DHI, Clearsky DNI, and Clearsky GHI values were close to corresponding observed values because both MSE and MAE values were almost zero, while *R*<sup>2</sup> values of approximately 0.987 were reached. Tables 3 and 4 yield the M-LSTM model performed slightly better than the DNN model in estimating Clearsky DNI and Clearsky DHI.

## Conclusion

In this study, two models based on DNN and M-LSTM architectures are developed to predict solar radiation. These two models were compared with each other. Our dataset is open

access under the Creative Commons license on the Kaggle platform. It consists of 175296 rows and 18 columns. First, the data set was analyzed and then the correlation of each variable in the data was calculated. Figure 5 shows the correlation matrix. According to the correlation values, there is a negative high correlation between SZA and clearsky DHI, clearsky DNI, and clearsky GHI. There is a high positive correlation between dew point and precipitable water and again between dew point and temperature. Also clearsky DHI, clearsky DNI, and clear sky GHI have high positive correlation with each other. Trend analysis, where we can analyze the long-term status of the data, is also shown in fig. 5. We used three metrics to measure performance of the models such that MAE, MSE, and  $R^2$  value. In this scope, our M-LSTM model was more successful than the DNN model, except for one metric. According to these results, Clearsky GHI value is estimated with the DNN model with an  $R^2$  value of approximately 99%.

### Acknowledgment

No funds, grants, or other support was received. The authors declare that they have no conflict of interest.

### References

- [1] \*\*\*, Opower, 2022. <https://www.oracle.com/industries/utilities/opower-energy-efficiency/>
- [2] Dincer, I., Renewable Energy and Sustainable Development: A Crucial Review, *Renewable and Sustainable Energy Reviews*, 4 (2000), 2, pp. 157-175
- [3] Panwar, N., et al., Role of Renewable Energy Sources in Environmental Protection: A Review, *Renewable and Sustainable Energy Reviews*, 15 (2011), 3, pp. 1513-1524
- [4] Basic-Sontic, A., et al., The Role of Personality Traits in Green Decision-Making, *Journal of Economic Psychology*, 62 (2017), Oct., pp 313-328
- [5] Gu, Y., Zhang, X., A Solar Photovoltaic/Thermal (pv/t) Concentrator for Building Application in Sweden Using Monte-Carlo Method, in: *Data-driven Analytics for Sustainable Buildings and Cities*, Springer, Heilberberg, Germany, 2021, pp. 141-161
- [6] Varcin, F., et al., End-to-end Computerized Diagnosis of Spondylolisthesis Using Only Lumbar x-Ray, *Journal of Digital Imaging*, 34 (2021), 1, pp. 85-95
- [7] Hayit, T., et al., Determination of the Severity Level of Yellow Rust Disease in Wheat by Using Convolutional Neural Networks, *Journal of Plant Pathology*, 103 (2021), 3, pp. 923-934
- [8] Alterkavi, S., Erbay, H., Design and Analysis of a Novel Authorship Verification Framework for Hijacked Social Media Accounts Compromised by a Human, *Security and Communication Networks*, 2021 (2021), ID8869681
- [9] Alterkavi, S., Erbay, H., Novel Authorship Verification Model for Social Media Accounts Compromised by a Human, *Multimedia Tools and Applications*, 80 (2021), 9, pp. 13575-13591
- [10] Akarслан, E., et al., Novel Short Term Solar Irradiance Forecasting Models, *Renewable Energy*, 123 (2018), Aug., pp. 58-66
- [11] Hochreiter, S., Schmidhuber, J., Long Short-Term Memory, *Neural Computation*, 9 (1997), 8, pp. 1735-1780
- [12] Shakya, A., et al., Solar Irradiance Forecasting in Remote Microgrids Using Markov Switching Model, *IEEE Transactions on Sustainable Energy*, 8 (2016), 3, pp. 895-905
- [13] Reikard, G., Predicting Solar Radiation at High Resolutions: A Comparison of Time Series Forecasts, *Solar Energy*, 83 (2009), 3, pp. 342-349
- [14] Diagne, M., et al., Review of Solar Irradiance Forecasting Methods and a Proposition for Small-Scale Insular Grids, *Renewable and Sustainable Energy Reviews*, 27 (2013), Nov., pp. 65-76
- [15] Voyant, C., et al., Machine Learning Methods for Solar Radiation Forecasting: A Review, *Renewable Energy*, 105 (2017), May, pp. 569-582
- [16] Yang, D., et al., Forecasting of Global Horizontal Irradiance by Exponential Smoothing, Using Decompositions, *Energy*, 81 (2015), Mar., pp. 111-119
- [17] Yang, D., et al., Reconciling Solar Forecasts: Geographical Hierarchy, *Solar Energy*, 146 (2017), Apr., pp. 276-286

- [18] Bailek, N., *et al.*, A New Empirical Model for Forecasting the Diffuse Solar Radiation over Sahara in the Algerian Big South, *Renewable Energy*, 117 (2018), 3, pp. 530-537
- [19] Chu, Y., *et al.*, Real-Time Prediction Intervals for Intra-Hour DNI Forecasts, *Renewable Energy*, 83 (2015), C, pp. 234-244
- [20] Law, E. V., *et al.*, Direct Normal Irradiance Forecasting and Its Application Concentrated Solar Thermal Output Forecasting – A Review, *Solar Energy*, 108 (2014), Oct., pp. 287-307
- [21] Gostein, M., *et al.*, Evaluating a Model to Estimate GHI, DNI, & DHI from POA Irradiance, *Proceedings*, 2016 IEEE 43<sup>rd</sup> Photovoltaic Specialists Conference (PVSC), Portland, Org., USA, 2016, pp. 0943-0946
- [22] Gustafson, W. T., *et al.*, Global Validation of Rest2 Incorporated into an Operational DNI and GHI Irradiance Model, *Proceedings*, 2016 IEEE 43<sup>rd</sup> Photovoltaic Specialists Conference (PVSC), Portland, Ore., USA, 2016, pp. 0947-0952
- [23] Gueymard, C. A., Rest2: High-Performance Solar Radiation Model for Cloudlessky Irradiance, Illuminance, and Photosynthetically Active Radiation-Validation with a Benchmark Dataset, *Solar Energy*, 82 (2008), 3, pp. 272-285
- [24] Sengupta, M., *et al.*, Physics-Based Goes Satellite Product for Use in NREL's National Solar Radiation Database, Technical Report, National Renewable Energy Lab. (NREL), Golden, Col., USA, 2014
- [25] Liu, W., *et al.*, Use of Physics to Improve Solar Forecast: Physics-Informed Persistence Models for Simultaneously Forecasting GHI, DNI, and DHI, *Solar Energy*, 215 (2021), Feb., pp. 252-265
- [26] Luo, J., *et al.*, Progress in Perovskite Solar Cells Based on ZnO Nanostructures, *Solar Energy*, 163 (2018), Mar., pp. 289-306
- [27] Pazikadin, A. R., *et al.*, Solar Irradiance Measurement Instrumentation and Power Solar Generation Forecasting Based on Artificial Neural Networks (ANN): A Review of Five Years Research Trend, *Science of The Total Environment*, 715 (2020), 136848
- [28] Zang, H., *et al.*, Short-Term Global Horizontal Irradiance Forecasting Based on a Hybrid CNN-LSTM Model with Spatiotemporal Correlations, *Renewable Energy*, 160 (2020), Nov., pp. 26-41
- [29] Benali, L., *et al.*, Solar Radiation Forecasting Using Artificial Neural Network and Random Forest Methods, Application Normal Beam, Horizontal Diffuse and Global Components, *Renewable Energy*, 132 (2019), Mar., pp. 871-884
- [30] Verbois, H., *et al.*, Probabilistic Forecasting of Day-Ahead Solar Irradiance Using Quantile Gradient Boosting, *Solar Energy*, 173 (2018), Oct., pp. 313-327
- [31] Busic-Sontic, A., *et al.*, Personality Trait Effects on Green Household Installations, *Collabra: Psychology*, 4 (2018), 1, 8
- [32] Achleitner, S., *et al.*, Solar Irradiance Prediction System, *Proceedings*, 13<sup>th</sup> PSN14 International Symposium on Information Processing in Sensor Networks, IEEE, Berlin, Germany, 2014, pp 225-236
- [33] \*\*\*, Wipro's sustainability machine learning challenge, 2021, URL <https://www.kaggle.com/datasets/vickeytomar/wipros-sustainability-machine-learning-challenge>, 2022
- [34] Goodfellow, I., *et al.*, *Deep Learning*, MIT Press, Cambridge, Mass., USA, 2016
- [35] Komorowski, M., *et al.*, Exploratory Data Analysis, in: *Secondary Analysis of Electronic Health Records*, Springer, New York, USA, 2016, Chapter 15, pp. 185-203
- [36] Dowdy, S., *et al.*, *Statistics for Research*, John Wiley & Sons, New York, USA, 2011