International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 499

HOURLY SOLAR IRRADIANCE PREDICTION FROM SATELLITE DATA USING LSTM

Pratima Kumari[1*], Durga Toshniwal[1]

[1]Indian Institute of Technology Roorkee, Uttarakhand, India

*Corresponding Author e-mail id: pratimatomar93@gmail.com

ABSTRACT

Solar irradiance prediction is an emerging area of research for various applications in renewable energy domain. So far, numerous physical models, statistical models and machine learning based techniques have been utilized to accomplish solar irradiance prediction. However, existing models are not good at learning long-term historical dependencies, lead to compromise in modeling non-linear solar irradiance patterns. In this paper, a novel prediction model (i.e. Long Short Term Memory, LSTM) from deep neural network family is used to predict hourly solar irradiance with enhanced prediction accuracy by considering long-term historical data dependencies. To provide an extensive and strong assessment of proposed model, present study employs National Solar Radiation Database (NSRDB) data for evaluating prediction accuracy at 7 locations of India having different climatic conditions. The proposed model is compared with Feed Forward Neural Network (FFNN), Extreme Gradient Boost (XGBoost) and Persistence model at broader coverage of geographical regions. Empirical outcomes suggest that proposed LSTM model outperforms different models with an average forecast skill of 50.72% over persistence model.

Keywords: renewable energy, deep learning, long short term memory, clearness index and climatic condition

1. INTRODUCTION

The International Energy Agency (IEA) is continuously motivating the research in the field of non-conventional energy related technologies for a clean and sustainable future. Solar energy is ample but less utilized resource, have higher potential to meet the vast energy demands [1]. The efficient penetration of solar energy resources into energy market is very challenging due to highly uncertain and irregular nature of solar resource in both time and space. Therefore, accurate prediction of energy generation is always a point of top priority in energy industry.

The solar irradiance prediction models can be broadly divided into three categories: (i) physical (ii) statistical and (iii) machine learning methods. Despite the pros and cons of each forecasting method, machine learning methods, a sub branch of artificial intelligence (AI) techniques have gained enormous attention due to their prediction accuracy and reliability [2]. In this work, machine learning methods are presented for solar irradiance forecasting. Moreover, one-fits-all approach is not applicable for solar irradiance forecasting. It can't be assured that a model which has been performed successfully at one location is as persuasive at another location. It has been observed that most of the studies in literature follow one-fits-all approach. To illustrate the state-of-the-art in solar irradiance forecasting, Table 1 presents some previous studies along with several climatic conditions, employed data, and utilized prediction techniques.

In reference of prediction models, Table 1 unfolds multiple conclusions. First, it reveals that FFNN dominates in solar irradiance prediction. It also suggests that a large number of literature studies on solar irradiance prediction have been performed but potential of deep learning models is still unexplored [19]. Second, very few studies include multiple geographical locations with different climate types as previous studies have generally employed measured ground-based solar irradiance readings.

In present work, we conquered these shortcomings by utilizing satellite-based recorded irradiance data which is available for many geo-locations. Using satellite-based irradiance data, we add some bias towards climate and location specific irradiance modeling. This enables us to evaluate the performance

Table 1
Previous study on solar irradiance forecast

Geographical span		Models		Reference
Locations covered	Climates covered	Proposed	Benchmark	
1	1	DRWNN	AP, FFNN	[3]
1	1	RBFNN	None	[4]
1	1	FFNN	None	[5]
7	2	Ensemble	SVR, GBR, RFR	[6]
4	2	SVM	Persistence	[7]
1	1	FFNN	None	[8]
83	1	FFNN	None	[9]
1	1	FFNN	NP, MC, K-NN, AR	[10]
1	1	LSTM	Persistence, LR	[11]
1	1	RNN, RBFNN	FFNN	[12]
7	4	LSTM	Persistence, FFNN, XGBoost	This study

of prediction models at multiple locations having different climatic conditions. Additionally, a data that includes all seasons throughout the year is used for the validation of the model.

2. RESEARCH CONTRIBUTIONS

The existing methods have several limitations such as physical methods are highly complexed and take large computational time. Next, statistical methods are simpler but they fail in capturing the non-linearity of data. Moreover, their incapability of parameter optimization is also an issue that need to be resolved. In contrast to statistical methods, AI based methods have been found highly effective in handling the non-linearity of data but they shortfall in handling the historical dependencies in time-series data. Hence, the present work may overcome the mentioned shortcomings and contribute as follows:

1. It proposes an optimized LSTM based deep learning model for hourly solar irradiance prediction to resolve the preexisting challenges.
2. It would be able to handle non-linear complexities present in solar irradiance data.
3. It would be capable to model both the long-term and short-term dependencies in data.
4. It quantifies the performance of prediction models on solar radiation data of multiple locations across different geographical regions.

3. PRILIMINARIES

3.1 Feed Forward Neural Network (FFNN)

FFNN is the first type of ANN, fully inspired by the architecture of a human brain nervous system. FFNN is widely popular to solve prediction and classification problems due to their self-learning and generalization capability. While FFNN can efficiently handle complex non-linear mappings of input-output, they fail to learn the historical dependencies in data. In this way, RNN emerged as an advancement over FFNN which inherently handles the dependencies in data. However, RNN can excel at handling short term dependencies, but it is ineffective to learn long term dependencies due to vanishing gradient problem.

3.2 Long Short Term Memory (LSTM)

Hochreiter and Schmidhuber [13] recommended that LSTM networks are successful in handling the vanishing gradient problem of RNN by incorporating gates and memory cell in the hidden layer nodes which regulates the information flow in the network. The LSTM model is capable to learn the long-term dependencies and temporal and spatial patterns of solar irradiance data, which helps to exploit the contextual information. Therefore, these unique capabilities are main motivation behind this study to employ LSTM network for hourly solar irradiance prediction.

3.3 Extreme Gradient Boosting (XGBoost)

Chen et al. [14] has recently proposed an XGBoost model which combines several regression trees to develop a powerful model. It works on the principle of boosting i.e. it combines models with low variance and high bias to lower the bias while maintaining the low variance. XGBoost reduces the over-fitting of model by adding a regularization term to the loss function.

3.4 Persistence model

Persistence model is the simplest and naive model of forecast which sets the irradiance value at previous hour y_{t-1} to be an hour ahead solar irradiance, y_t. Thus persistence model does not require training and parameter setting and generally used as a baseline model.

4. PROPOSED METHODOLOGY

4.1 Data description

The National Solar Radiation Database (NSRDB) SUNY Semi-Empirical Model data developed by National

Renewable Energy Laboratory is based on satellite-to-irradiance models. SUNY produces solar irradiance data for South Asia from 2000-2014 with the spatial resolution of 10 km x 10 km at a temporal resolution of 1 h [15].

Usually, energy industries does not share solar energy data for open access which leads to an alternate way to forecast GHI. The GHI data is recorded using pyranometer installed on a planar surface. However, the availability of such ground-recorded data is limited due to high maintenance and installation cost of the device (pyranometer). Addressing these issues related to recorded GHI data, several researchers advise modeling GHI using satellite-recorded data [16]. Satellite-based GHI data is often less accurate in comparison to pyranometer based readings. Hence, prediction models based on satellite-based data are relatively less validated. Though, it has also been reported that quality of satellite-based irradiance data has improved significantly and have shown comparable forecasting error benchmarked against satellite and ground-based data [17]. However, this study aims to provide a relative analysis of prediction models using satellite-based GHI data under similar settings with same input-output variables. The major advantage of satellite-based data is its availability for various geographical locations which facilitates the validation of forecasting models at multiple locations.

In this work, data of 7 different locations, covering 4 different climatic zones across India is used. To compile, Table 2 presents the geographical information of selected locations along with their climatic zones. The presented climate information is based on Köppen-Geiger climate classification system [18].

Table 2

Previous study on solar irradiance forecast

Location	Lat	Long	Climate
Dehradun	30.3165	78.0322	Hot-summer Mediterranean **(Csa)**
Dharamshala	32.2190	76.3234	Humid-subtropical **(Cwa)**
Gandhinagar	23.2156	72.6369	Semi-arid **(BSh)**
Guwahati	26.1445	91.7362	Humid-subtropical **(Cwa)**
Jaipur	26.9124	75.7873	Semi-arid **(BSh)**
Pune	18.5204	73.8567	Semi-arid **(BSh)**
Thiruvananthapuram	8.5241	76.9366	Tropical wet **(Aw)**

Five years (2010-2014) SUNY data at aforementioned locations were downloaded. The three years (2010-2012) and two years (2013-2014) data is used for training and testing of proposed model, respectively. Since, GHI readings are reduced to zero at night due to unavailability of sunlight therefore, data instances between 6:30 - 17:30 hours are picked up for training and testing.

4.2 Feature selection

4.2.1 Parameter selection from historical states of GHI data

The autocorrelation coefficient of the solar irradiance time-series (y_t) is utilized to determine the embedded input dimension as follows:

$$r_k = \frac{\sum_{i=1}^{N-k}(y_t-\bar{y})(y_{t-k}-\bar{y})}{\sum_{i=1}^{N}(y_t-\bar{y})^2} \tag{1}$$

where, r_k $(k = 1,2, \ldots .80)$ is the autocorrelation coefficient with lag k. s_y, μ_y and N are the standard deviation, mean and number of sample in GHI time-series, respectively. Autocorrelation results show that lagged hourly GHI value of previous three days can be used as input feature vector (12*3 = 36 features).

4.2.2 Parameter selection from meteorological variables

All meteorological variables present in dataset are not relevant. Therefore, Pearson correlation of each meteorological variable with solar irradiance is calculated. The correlation values suggest that temperature, relative humidity, pressure, wind speed and dew point are relevant variables for GHI forecasting as shown in Figure 1. The historical values of GHI and meteorological variables collectively consist the final input feature.

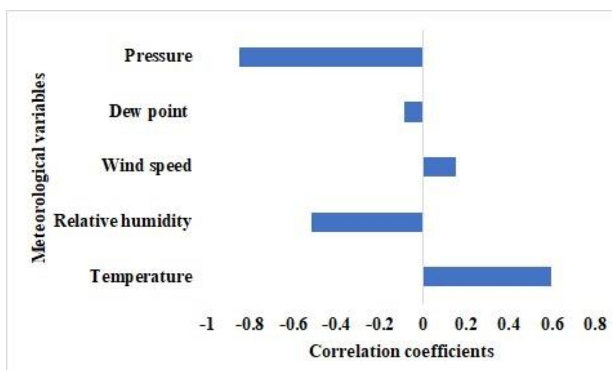


Fig 1 Correlation of meteorological variables with solar irradiance.

4.3 Performance evaluation

The performance of the proposed model is evaluated using normalized root mean square error (nRMSE) and forecast skill which is formulated by Eqs, (2) and (3):

$$nRMSE(\%) = \frac{\sqrt{\frac{1}{N}(\hat{y}_i - y_i)^2}}{\frac{1}{N}\sum_{i=1}^{N} y_i} * 100 \quad (2)$$

$$Forecast\ skill(\%) = 1 - \frac{nRMSE}{nRMSE_{persistence}} \quad (3)$$

Where, $\hat{y}_i$ and y_i is the predicted and actual value of GHI respectively, N is the total number of observations. Forecast skill provides the relative improvement of prediction models over persistence model.

4.4 Model tuning and parameter setting

GHI predictions of persistence model for Thiruvananthapuram exhibits high irregularities which suggest that this location is challenging to forecast. Therefore, Prediction models are fine-tuned on data of Thiruvananthapuram. Grid-search has been used to determine the hyper-parameters for models. Table 4 shows the considered parameter settings. The parameter values for which prediction model provides the best performance are fixed and then same values are reiterated for other location. To give adaption freedom to the prediction models, all but one hyper-parameter are fixed and rest are tuned for each location. For LSTM and FFNN the free hyper-parameter is the number of iterations whereas for XGBoost model, number of regression trees represents the same. FFNN, LSTM and XGBoost models are implemented using python based keras [19] and xGBoost library, respectively.

Table 3
Hyper-parameters used for grid-search.

Algorithm	Hyper-parameter	Values
FFNN	Neurons in 1st hidden layer	c(10,20,20,40,50,60,70,80,90, 100,110,120,130,140,150)
	Neurons in 2nd hidden layer	c(10,20,20,40,50,60,70,80,90, 100,110,120,130,140,150)
LSTM	Neurons in 1st hidden layer	c(10,20,20,40,50,60,70,80,90, 100,110,120,130,140,150)
	Neurons in 2nd hidden layer	c(10,20,20,40,50,60,70,80,90, 100,110,120,130,140,150)
XGBoost	Number of regression trees	k(10,20,30,40,50,60,70,80,90, 100)
	Maximum depth of tree	d(2,3,4,5,6,7,8)

5. RESULTS AND DISCUSSION

The simulation of considered models is performed using historical and meteorological data to forecast future solar irradiance. Table 4 shows the prediction results in terms of nRMSE (%) for selected locations. At majority of locations LSTM provides high accuracy in prediction results, suggests that LSTM can be accepted as an apt method for solar irradiance forecasting. Especially, LSTM exhibits superior results than XGBoost,

Table 4
nRMSE(%) of predicted GHI for selected locations.

Location	FFNN	LSTM	XGBoost	Persistence
Dehradun	30.93	**27.29**	29.76	51.59
Dharamshala	33.87	31.98	**30.78**	50.38
Gandhinagar	21.34	**17.86**	20.61	48.73
Guwahati	33.08	30.36	**30.06**	52.39
Jaipur	22.02	**19.03**	21.83	48.47
Pune	26.59	**22.79**	25.93	51.49
Thiruvananthapuram	30.79	**26.53**	29.61	53.06

FFNN and persistence model. The excellence of LSTM model as reflected in Table 4 may generalize over other models (less accurate than XGBoost and FFNN). Although, LSTM model has shown high accuracy in GHI prediction at most of locations, Table 4 suggests that our dataset contains two locations (Dharamshala and Guwahati), where LSTM prediction accuracy is remarkably lower than benchmark models. In addition, the forecast skill of FFNN, LSTM and XGBoost over persistence model is calculated which also indicates favorable results for LSTM at all locations except Dharamshala and Guwahati, as shown in Figure 2. Figure 2 suggests that FFNN also shows a sudden decline of performance at these two locations.

To delve into the cause of this particular behavior, we further analyzed the monthly nRMSE across these two locations as shown in Figure 3(a) and 3(b). Figure 3(a) indicates that the overall reduction in performance of LSTM for Dharamshala is subjected to high prediction error in June, July, August and September. Similarly, in Figure 3(b) the prediction error is high in April, May, June, July, August and September which reduces the overall performance of LSTM for Guwahati. While, in remaining months LSTM has performed better than benchmark models for both locations. These results confirms that less prediction accuracy of LSTM in aforementioned months is explicitly subjected to the characteristics of the locations (Dharamshala and Guwahati). Moreover, LSTM is capable of building efficient non-linear mapping function between input features and target variable which makes it difficult to elucidate the origin of reduction in prediction accuracy. Although, penetrating the mapping function is very complicated, an analysis of data may yield some fruitful explanation for the increased prediction error in specific months. To investigate the cause of increased error, we examined the monthly clearness index (K_t) for all locations.

As all models have shown highly accurate GHI forecast in Gandhinagar thus, monthly trend of clearness

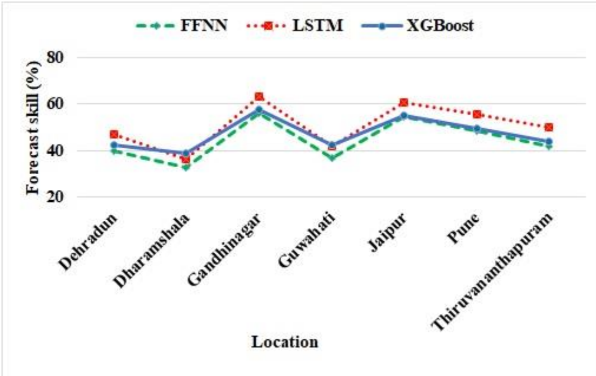
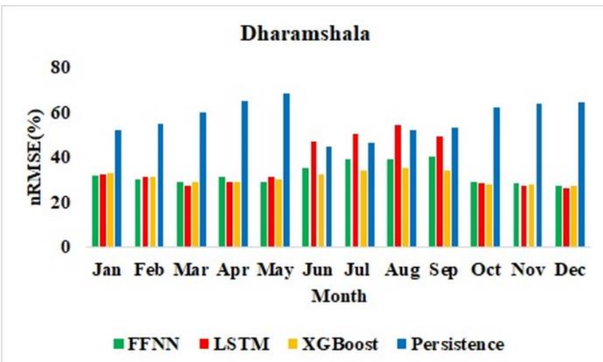
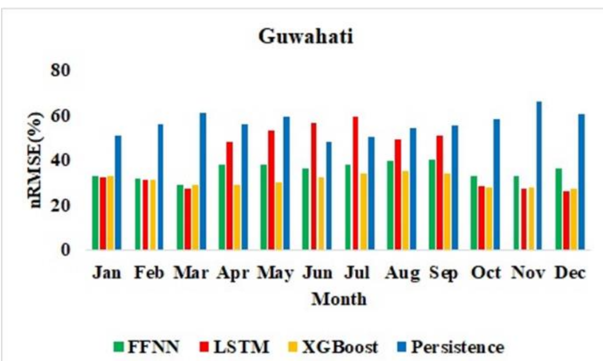


Fig 2 Forecast skill of FFNN, LSTM and XGBoost over persistence model.



(a)



(b)

Fig 3 Monthly nRMSE(%) for worst performing locations.

index for Dharamshala and Guwahati is compared with Gandhinagar as shown in Figure 4. It indicates that K_t in Dharamshala is low throughout the year and depicts a remarkable fall in months of June, July, August and September. Likewise, Guwahati has very low K_t throughout the year, especially in April, May, June, July, August and September which hits the lowest value (0.21) in July. The low K_t in aforementioned months specifies the less number of clear sky hours for Dharamshala and Guwahati which is in good accordance with their rainy climatic conditions. Furthermore, Table 4 suggests that

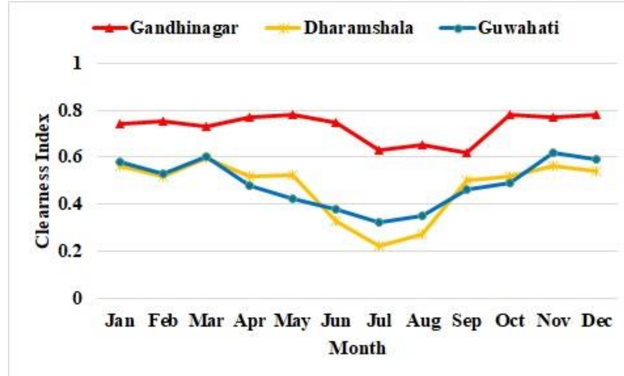


Fig 4 Comparison of monthly average clearness index value for Dharamshala, Gandhinagar and Guwahati.

LSTM has shown 17.86% nRMSE in Gandhinagar which dropped to 31.98% and 30.36% in Dharamshala and Guwahati, respectively. On the other hand, XGBoost has shown more accurate GHI forecast for these two locations which clearly indicates that XGBoost is a top performing model for such climates. Yagli et al. [20] has arrived at similar conclusion in their study of hourly GHI forecast.

The empirical results indicate that LSTM model is more accurate than FFNN, RNN, XGBoost and persistence on average. Occasionally, LSTM shows low performance than XGBoost (e.g. June, July, August, and September for Dharamshala and Guwahati) but never less than FFNN. The follow-up examination of two exceptional locations, where the benchmark model outperformed the LSTM has proved that distribution of clearness index shows significant irregularities in input features at these locations. These irregularities affect every prediction model. However, the LSTM model has upper-hand in distilling more informative patterns in data than benchmarks. Therefore, random distortion in the input features is likely to affect LSTM more firmly than benchmark models. This could illustrate the relatively less accurate performance of LSTM for Dharamshala and Guwahati.

CONCLUSIONS

This work introduces a deep learning based LSTM model as an influential approach for time series forecasting and evaluates its potential to predict hourly solar irradiance. LSTM outperforms the challenging benchmarks (FFNN and XGBoost) which evidently advise the use of LSTM to predict GHI. The proposed framework uses historical GHI and meteorological data as input feature. The feasibility and adaptability of models are investigated for 7 locations across India covering 4 different climates. Hence, it overcomes the limitations of prior works which presents location specific comparison of machine learning models and needs external validation at other locations with different spatial and climatic conditions. LSTM failed to perform well at certain locations (with low clearness index) at which XGBoost has shown praisable prediction accuracy. Hence, XGBoost model is recommended for those locations where clearness index is low throughout the year. The proposed framework suggests that climatic conditions could be a factor in deciding the appropriate model. Therefore, one could foresee an ensemble approach that automatically switch from one model to another according to the change in climatic condition for better GHI prediction.

ACKNOWLEDGEMENT

Authors would like to thank Ministry of Human Resource Development (MHRD), Govt. of India.

REFERENCE

[1] Perez R, Zweibel K, Hoff TE. Solar power generation in the US: Too expensive, or a bargain? Energy Policy 2011. doi:10.1016/j.enpol.2011.08.052.

[2] Urraca R, Antonanzas J, Alia-Martinez M, Martinez-De-Pison FJ, Antonanzas-Torres F. Smart baseline models for solar irradiation forecasting. Energy Convers Manag 2016;108:539–48. doi:10.1016/j.enconman.2015.11.033.

[3] Cao J, Lin X. Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks. Energy Convers Manag 2008. doi:10.1016/j.enconman.2007.12.030.

[4] Chen C, Duan S, Cai T, Liu B. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Sol Energy 2011;85:2856–70. doi:10.1016/j.solener.2011.08.027.

[5] Ding M, Wang L, Bi R. An ANN-based approach for forecasting the power output of photovoltaic system. Procedia Environ. Sci., 2011. doi:10.1016/j.proenv.2011.12.196.

[6] Gala Y, Fernández Á, Díaz J, Dorronsoro JR. Hybrid machine learning forecasting of solar radiation values. Neurocomputing 2016. doi:10.1016/j.neucom.2015.02.078.

[7] Da Silva Fonseca Junior JG, Oozeki T, Ohtake H, Shimose K ichi, Takashima T, Ogimoto K. Regional forecasts and smoothing effect of photovoltaic power generation in Japan: An approach with principal component analysis. Renew Energy 2014. doi:10.1016/j.renene.2014.02.018.

[8] Leva S, Dolara A, Grimaccia F, Mussetta M, Ogliari E. Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. Math Comput Simul 2017. doi:10.1016/j.matcom.2015.05.010.

[9] Linares-Rodríguez A, Ruiz-Arias JA, Pozo-Vázquez D, Tovar-Pescador J. Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks. Energy 2011. doi:10.1016/j.energy.2011.06.044.

[10] Paoli C, Voyant C, Muselli M, Nivet ML. Forecasting of preprocessed daily solar radiation time series using neural networks. Sol Energy 2010. doi:10.1016/j.solener.2010.08.011.

[11] Qing X, Niu Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy 2018;148:461–8. doi:10.1016/j.energy.2018.01.177.

[12] Yona A, Senjyu T, Saber AY, Funabashi T, Sekine H, Kim CH. Application of neural network to one-day-ahead 24 hours generating power forecasting for photovoltaic system. 2007 Int. Conf. Intell. Syst. Appl. to Power Syst. ISAP, 2007. doi:10.1109/ISAP.2007.4441657.

[13] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput 1997. doi:10.1162/neco.1997.9.8.1735.

[14] Chen T, Guestrin C. XGBoost. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD '16, 2016. doi:10.1145/2939672.2939785.

[15] Sengupta M, Xie Y, Lopez A, Habte A, Maclaurin G, Shelby J. The National Solar Radiation Data Base (NSRDB). Renew Sustain Energy Rev 2018. doi:10.1016/j.rser.2018.03.003.

[16] Lorenzo AT, Holmgren WF, Cronin AD. Irradiance forecasts based on an irradiance monitoring network, cloud motion, and spatial averaging. Sol Energy 2015. doi:10.1016/j.solener.2015.10.038.

[17] Perez R, Schlemmer J, Hemker K, Kivalov S, Kankiewicz A, Dise J. Solar energy forecast validation for extended areas & economic impact of forecast accuracy. 2017 IEEE 44th Photovolt. Spec. Conf. PVSC 2017, 2017. doi:10.1109/PVSC.2017.8366214.

[18] Peel MC, Finlayson BL, Mcmahon TA. Updated world map of the Köppen-Geiger climate classification Hydrology and Earth System Sciences Updated world map of the Köppen-Geiger climate classification. 2007.

[19] Chollet François. Keras: The Python Deep Learning library. KerasIo 2015. doi:10.1086/316861.

[20] Yagli GM, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev 2019. doi:10.1016/j.rser.2019.02.006.