



Article Deep-Learning-Based Adaptive Model for Solar Forecasting Using Clustering

Sourav Malakar ^{1,*,†}, Saptarsi Goswami ^{2,†}, Bhaswati Ganguli ³, Amlan Chakrabarti ¹, Sugata Sen Roy ³, K. Boopathi ⁴ and A. G. Rangaraj ⁴

- ¹ A.K. Choudhury School of Information Technology, University of Calcutta, Kolkata 700073, India; achakra12@yahoo.com
- ² Bangabasi Morning College, University of Calcutta, Kolkata 700073, India; sgakc@caluniv.ac.in
- ³ Department of Statistics, University of Calcutta, Kolkata 700073, India; bgstat@gmail.com (B.G.); sugatasr@gmail.com (S.S.R.)
- ⁴ National Institute of Wind Energy (NIWE), The Ministry of New and Renewable Energy, Government of India, New Delhi 110003, India; boopathi.niwe@nic.in (K.B.); rangaraj.niwe@nic.in (A.G.R.)
- * Correspondence: sourav.xaviers@gmail.com
- † These authors contributed equally to this work.

Abstract: Accurate short-term solar forecasting is challenging due to weather uncertainties associated with cloud movements. Typically, a solar station comprises a single prediction model irrespective of time and cloud condition, which often results in suboptimal performance. In the proposed model, different categories of cloud movement are discovered using K-medoid clustering. To ensure broader variation in cloud movements, neighboring stations were also used that were selected using a dynamic time warping (DTW)-based similarity score. Next, cluster-specific models were constructed. At the prediction time, the current weather condition is first matched with the different weather groups found through clustering, and a cluster-specific model is subsequently chosen. As a result, multiple models are dynamically used for a particular day and solar station, which improves performance over a single site-specific model. The proposed model achieved 19.74% and 59% less normalized root mean square error (NRMSE) and mean rank compared to the benchmarks, respectively, and was validated for nine solar stations across two regions and three climatic zones of India.

Keywords: clearness index forecasting; cloud cover; clustering; DTW

1. Introduction

Solar power is one of the viable alternatives to fossil-fuel-generated power, which causes serious environmental damage [1]. In terms of total energy consumption, India is ranked third after China and the United States [2], and has a target of producing 57% of total electricity capacity from renewable sources by 2027 [3]. In this paper, we developed a novel method for the short-term (some hours ahead) [4] forecasting of the clearness index (Kt) (defined as the ratio of global horizontal irradiance (GHI) to extraterrestrial irradiance) [5–8] while accounting for unpredictable weather conditions, focusing on variability in cloud cover [9–12]. Cloud variability leads to highly localized solar prediction, as a single model is unable to provide accurate forecasts under different weather conditions [13,14].

Long short-term memory (LSTM) [15] is one of the most popular deep-learning algorithms, mainly used to handle sequential data, and it can preserve knowledge by passing through the subsequent time steps of a time series [16]. In [17], the authors developed a site-specific univariate LSTM for the hourly forecasting of photovoltaic power output. In [18], the authors compared the performance of several alternative models for forecasting clear-sky GHI. These included gated recurrent units (GRUs), LSTM, recurrent neural networks (RNNs), feed-forward neural networks (FFNNs), and support vector regression (SVR). GRU and LSTM outperformed the other models in terms of root mean square error



Citation: Malakar, S.; Goswami, S.; Ganguli, B.; Chakrabarti, A.; Roy, S.S.; Boopathi, K.; Rangaraj, A.G. Deep-Learning-Based Adaptive Model for Solar Forecasting Using Clustering. *Energies* **2022**, *15*, 3568. https://doi.org/10.3390/en15103568

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 24 March 2022 Accepted: 19 April 2022 Published: 13 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). (RMSE). In [19], the authors proposed an hour-ahead solar power forecasting model based on RNN-LSTM for three different solar plants. In [9], LSTM and GRU dominated over artificial neural networks (ANNs), FFNNs, SVR, random forest regressor (RFR), and multilayer perceptron (MLP) in solar forecasting. The above discussion suggests that the authors used a single model to forecast solar irradiation for a particular day and did not consider cloud cover at the time of forecasting. In [20], the authors designed a forecasting model for one-day ahead hourly prediction using LSTM. The authors reported that the algorithm performed effectively under fully or partially cloudy conditions. In [21], the authors proposed

formed effectively under fully or partially cloudy conditions. In [21], the authors proposed a one-hour-ahead hybrid solar forecasting model using traditional machine-learning models such as random forest (RF), gradient boosting (GB), support vector machines (SVMs), and ANNs. The RF model showed the best forecasting accuracy for the spring and autumn seasons, while the SVR model performed best for the winter and summer seasons. In [22], the authors evaluated 68 machine-learning models for 3 sky conditions, 7 locations, and 5 climate zones in the continental United States. No universal model exists, and specific models for each sky and climate condition are recommended. Hence, it is well-established that a single site-specific forecasting model is unable to produce consistent forecasting performance in all cloud conditions and seasons [20–22].

Typically, a site-specific model is built for solar-energy prediction, and multiple models are built for different seasons. However, even within the same day, there can be fluctuations due to variability in cloud cover [23]. So, a single model gives a very high error in terms of NRMSE. The error is further pronounced for time windows with high cloud variability. The authors in [22], found that a specific forecasting model showed very high error in NRMSE in overcast cloud conditions in comparison with clear-sky conditions on a particular day. They also stated that forecasting performance significantly changed with the change in cloud conditions on a particular day. The authors in [24], stated that LSTM outperformed other predictive models in short-term solar forecasting. Nevertheless, its ability to predict cloudy days with low solar irradiance is significantly reduced. This serves as a motivation to implement an adaptive model. Table 1 summarizes the forecasting error of LSTM for nine solar stations across three climatic zones of India. Solar stations are described in Section 3.1. Figure 1 depicts the deviation of NRMSE in high- and lowvariability cloud-cover conditions in comparison with overall NRMSE. In high cloud-cover variability, forecasting error was significantly higher compared to overall NRMSE. This signifies that if cloud variability ever increases too much, site-specific LSTM cannot handle such a situation very well. Another motivation is provided by the parity plot in Figure 2, which shows forecast and actual clearness indices for three solar stations separately for high and low cloud-cover variability conditions. Forecasts were more accurate under low-variability cloud cover conditions than those for high-variability cloud conditions.

In this paper, we propose a novel short-term (2 h ahead) solar forecasting approach [25] that uses clustering on the basis of cloud parameters as a preprocessing step, and subsequently uses LSTM that is cluster-specific for the forecasting clearness index. Specifically,

- For each forecasting site, the nearest three neighboring stations were selected on the basis of DTW similarity scores [26].
- A global dataset was created by combining some derived features of cloud cover and clearness indices of each station and those of its neighbors. The derived features were obtained following [23].
- The entire day was divided into time windows. For clustering, the K-medoids [27] algorithm was applied on those time windows.
- A separate LSTM model was trained for each cluster that represented different cloud conditions.

The major contributions of the paper are as follows:

- An adaptive forecasting model (CB-LSTM) is proposed that can apply multiple models for a site on the basis of existing weather conditions.
- A global dataset was created on the basis of derived cloud-related information that is used to cluster a day into different weather types.

- The proposed model showed promising forecasting performance compared to benchmark models such as convolutional neural network (CNN)-LSTM and nonclusteringbased site-specific LSTMs. The model achieved less forecasting error for solar stations having significant solar variability.
- Performance (measured in terms of forecasting accuracy) was validated for nine solar stations from three climatic zones in India. To our knowledge, this is the first time that such an approach was applied to data from the Indian subcontinent.

The rest of the paper is organized as follows. Section 2 provides the background on various deep-learning model architectures. Section 3 presents the proposed method, and Section 4 presents details on its forecasting performance. The paper is concluded with a discussion in Section 5.

Table 1. Forecasting performance (NRMSE in %) of LSTM for high- and low-cloud-variability cloud cover.

Stations	High Cloud-Cover Variability	Low Cloud-Cover Variability	Overall NRMSE		
Bhainsdehi	44.72	17.71	29.85		
Begamganj	39.03	20.80	27.98		
Dindori	41.16	14.90	25.73		
	Hot and dry	y climate zone			
Tiruchirappalli	59.06	44.02	47.57		
Idukki	87.99	47.70	61.13		
Madurai	47.89	44.29	45.65		
Warm and humid climate zone					
Khaga	34.23	17.42	23.63		
Vaibhavwadi	92.22	38.22	57.31		
Osmanabad	71.72	33.58	47.83		



Figure 1. Deviation of NRMSE (%) in high and low cloud-cover variability conditions compared to overall NRMSE (%).



Figure 2. Parity plot showing forecast and actual clearness indices for three solar stations for high and low cloud-cover variability conditions. (**a**) Bhainsdehi (high cloud-cover variability); (**b**) Bhainsdehi (low cloud-cover variability); (**c**) Osmanabad (high cloud-cover variability); (**d**) Osmanabad (low cloud-cover variability); (**e**) Khaga (high cloud-cover variability); (**f**) Khaga (low cloud-cover variability).

2. Deep-Learning Models

ANN in the learning phase is unable to utilize information learned from the past time steps while processing the current time step, which is the major drawback of traditional neural networks. An RNN can solve this problem and is one of the deep-learning models designed to handle sequential data. To preserve information, it recursively transfers learning from previous time steps of the network to the current time step. However, it is susceptible to the vanishing gradient problem. As a result, it is unable to remember long-term dependencies.

LSTMs are a special type of RNN that is especially designed to learn both long- and short-term dependencies [15]. Compared to a traditional neural network, LSTM units encompass a 'memory cell' that can retain and maintain information for long periods of time [28]. Figure 3 is a schematic diagram of an LSTM cell. A set of gates are used to customize the hidden states. Three different gates are used, representing input, forget, and output. The functionality of each gate is summarized as follows.

- Forget gate [29] $(f_t) = \sigma(w_f[h_{t-1}, x_t] + b_f)$: On the basis of certain conditions such as x_t , h_{t-1} , and a sigmoid layer, a forget gate produces either 0 or 1. If 1, memory information is preserved; otherwise, it is discarded.
- Input gate [29] (i_t) = σ(w_i[h_{t-1}, x_t] + b_i): helps in deciding which values from the input are used for the current memory state.
- **Cell state** [29] $(c_t) = tanh(w_c[h_{t-1}, x_t] + b_c)$: new cell state c_t is the summation of $c_{t-1} * f_t$, and $\hat{c}_t * i_t$. $c_{t-1} * f_t$ decides the fraction of the old cell state that is discarded, and the amount of new information that is added is decided through $\hat{c}_t * i_t$.
- **Output gate** [29] $(o_t) = \sigma(w_o[h_{t-1}, x_t] + b_o)$: decides what to output on the basis of the current input and previous hidden state.
- **Hidden state** [29] $(h_t) = o_t * tanh(c_t)$: current hidden state is computed by multiplying the output gate by the current cell state using the tanh function.



Figure 3. Schematic diagram of an LSTM cell.

Here, w_f , w_i , w_c and w_o are weight matrices. b_f , b_i , b_c and b_o are biases for individual gates. σ indicates a sigmoid activation function. * stands for element-wise multiplication, and + implies element-wise addition.

3. Materials and Methods

3.1. Dataset Description and Preprocessing

We merged two sources of data for this analysis. The first was obtained from Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) [30] satellite. This provides information on PM2.5, surface wind speed (m/s), surface air temperature (k), total cloud area fraction, dew point temperature at 2 m (k), 2 m eastward wind (m/s), and 2 m northward wind (m/s). The second set of data was extracted from from the National Renewable Energy Laboratory (NREL) [31] for 2013. This contains information on DHI, DNI, GHI, clear-sky DHI, clear-sky DNI, clear -sky GHI, and solar zenith angle. The clearness index at time t (denoted by Kt) was calculated on the basis of GHI values. These two datasets were merged on the basis of latitude and longitude. For each location (unique combination of latitude and longitude), a 10 km radius was used for the merge. Table 2 describes the nine solar stations studied in this paper.

Table 2. Dataset description.

Latitude and Longitude	Region	Climatic Zone	Location
23.50 and 78.75	Inland	Composite	Begamganj (Madhya Pradesh)
22.50 and 81.25	Inland	Composite	Dindori (Madhya Pradesh)
25.50 and 81.25	Inland	Composite	Khaga (Uttar Pradesh)
21.50 and 77.50	Inland	Hot and dry	Bhainsdehi (Madhya Pradesh)
16.50 and 73.75	Coastal	Hot and dry	Vaibhavwadi (Maharashtra)
18.00 and 76.25	Coastal	Hot and dry	Osmanabad (Maharashtra)
10.00 and 78.125	Inland	Warm and humid	Madurai (Tamil Nadu)
11.00 and 78.75	Coastal	Warm and humid	Tiruchirappalli (Tamil Nadu)
10.00 and 76.875	Coastal	Warm and humid	Idukki (Kerala)

After collecting solar data (time series) from the PV module, they were stored in a database, and a series of standard preprocessing steps were applied.

- Night hours (8 p.m. to 7 a.m.) are removed. The resolution of the collected data was 15 min.
- Next, a standard sliding-window approach was applied to the time-series data to convert them into a suitable representation (supervised) for deep-learning models. Figure 4 depicts the generic approach of a sliding window with input and output window sizes *n* and *m*. The input window covered *n* past observations, such as {*X*₁, *X*₂, *X*₃, ..., *X_n*}, and used to predict the next m observations as {*X*_{n+1}, ..., *X*_{n+m}}. After that, the input window is shifted one position to the right as {*X*₂, *X*₃, *X*₄, ..., *X_{n+1}*}, and {*X_{n+2}, ..., X_{n+m+1}*} are the new input and output sequences. This process continues until no data points of the time series are left.



Figure 4. Sliding-window approach.

3.2. Locally and Remotely Derived Variables for Clustering

Variables that were obtained on the basis of the meteorological parameters of the forecast site are called locally derived variables, while variables acquired depending upon the neighboring forecasting sites are called remotely derived variables. Derived variables were obtained following [23].

Neighboring solar stations relative to the forecast site were selected as follows:

- A radius of 120 km was used to select the solar stations around the forecasting site.
- DTW [32] was used as the similarity measure of clearness index (Kt).

• Three solar stations with the best similarity score were selected and used for clustering. Table 3 lists all locally and remotely derived variables.

Table 3. Variables used for K-medioid clustering.

Predictors	Equation	Description					
	Locally derived variables						
kt_{trend}	KtPrev15 – KtPrev30	It was calculated to capture the most recent trend					
Kt temporal variability (Stdev 1-h)	$\sigma(\text{KtPrev15},\text{KtPrev30},\text{KtPrev45},\text{KtPrev60})$	Computed by taking the standard deviation of the four subsequent observations in an hour					
Kt Slope (1-h)	$\beta_0 + \beta_1 KtPrev15 + \beta_2 KtPrev30 + \beta_3 KtPrev45 + \beta_4 KtPrev60$	It was calculated by fitting a linear equation through four consecutive observations of the clearness index.					
	Remotely derived variables						
KtPrev15 nearby mean	$\mu(\text{KtPrev15}_{Source1},\text{KtPrev15}_{Source2},\text{KtPrev15}_{Source3})$	It was computed by taking the mean of the clearness index of three neighboring sites.					
KtPrev15 nearby std	$\sigma(\text{KtPrev15}_{Source1},\text{KtPrev15}_{Source2},\text{KtPrev15}_{Source3})$	It was computed by taking the standard deviation of the clearness index of three neighboring sites.					
Cloud-cover variability (Stdev)	$\sigma(\text{CV}_{Source1}, \text{CV}_{Source2}, \text{CV}_{Source3})$	It is the standard deviation of the cloud cover of neighboring solar stations.					
Cloud Cover Squired	$(\mu(\text{CV}_{\text{Source1}},\text{CV}_{\text{Source2}},\text{CV}_{\text{Source3}}))^2$	It is the squared value of the mean of cloud covers of three neighboring sites.					

3.3. Multivariate LSTM (M-LSTM)

Figure 5b describes the architecture of the multivariate LSTM model, summarized as follows:

- It is a site-specific model where predictors are directly used from the forecasting site. The model uses additional information on predictors such as temperature, dew point, wind speed, and cloud cover.
- A stateful LSTM was used for maintaining inter- and intrabatch dependency. The input layer of an LSTM consists of 5 input features of 16 time steps (4 h) each. Two hidden layers and a tanh activation were used.

3.4. Spatiotemporal LSTM (ST-LSTM)

This implementation led to a spatiotemporal model, as information from the forecast site is used together with information from neighboring sites.

Figure 5a shows the architecture in detail and can be summarized as follows:

- A spatiotemporal dataset was created by combining information on meteorological parameters, including dew point, temperature, wind speed, and cloud cover from the three neighboring sites and from the forecast site.
- Stateful LSTM was used. The input layer consisted of 20 input features of 16 time steps each. Two hidden layers were used with a dropout rate of 20%. As hidden-layer activation, tanh was used.
- A spatiotemporal dataset was used to train the LSTM and forecast clearness index for three different times of the day.



(a) Spatio-temporal Forecasting

Figure 5. Architecture of—(a) spatiotemporal and (b) multivariate forecasting models.

3.5. Clustering-Based ANN (CB-ANN) and LSTM (CB-LSTM)

CB-ANN was designed following [23]. CB-LSTM is a global forecasting model, designed using K-medoid clustering followed by LSTM. Meteorological parameters collected from the neighboring sites together with clearness index values from the forecast site were not directly used as predictors. As mentioned in Section 3.2, derived features were extracted for Kt from the forecast site and cloud cover information from neighboring sites.

Figure 6 shows a cluster-specific feature identification to understand the important features of a cluster. For each forecasting site, a spatiotemporal dataset was created that was split into training (80%) and test sets (20%). A global dataset was created by combining the training sets of all the forecast sites, and normalized using the min–max normalizer [33]. Next, the optimal number of clusters was determined on the basis of the elbow–silhouette [34] method. The K-medoids algorithm is used to cluster the time windows in the dataset. As the input attributes were related to cloud formation, the clusters intuitively represent different cloud types. As a result, the dataset was split into k clusters, where each cluster center was represented by a medoid.



Figure 6. Cluster-specific best feature identification strategy.

The proposed approach is presented in Figure 7 and is described as follows:

- For each cluster, separate stateless LSTM was built. The network was implemented using the keras package of version 2.7.0 in Python. The input layer of the LSTM consisted of 13 input features of 1 time step each. In the hidden layers, hyperbolic tangent activation function (tanh) was used. After each hidden layer, a batch-normalization [35] layer was used to transform inputs into a mean of 0 and a standard deviation of 1. We used a dropout and L2 regularization [36] to protect the network from overfitting. The network weights were initialized using the Xavier uniform initializer [37]. The output layer consisted of 8 nodes with linear activation and was used to forecast Kt for the next two hours. Figure 8 shows the network configuration of CB-LSTM together with input and output features. For obtaining the best performance of CB-LSTM, specific hyperparameters such as number of layers, number of nodes in each layer, batch size, number of epochs, dropout, and learning rate were optimized. The tree-structured Parzen estimator (TPE) [38] algorithm was used for optimization. Table 4 presents the hyperparameter settings. Table 5 shows the optimal hyperparameter settings for the proposed approach (CB-LSTM) in Section 4.
- Overall forecasting accuracy was computed for each site using the weighted average of the generated accuracy by the different LSTM models. Forecasting accuracy was separately computed for three times of the day.

	-
Model	Hyperparameter

Table 4. Hyperparameters to optimize.

Model	Hyperparameter	Value
	Number of layers	1, 2, 3
	Nodes in layers	25, 50, 75, 100
CB-LSTM	Learning rate	0.1, 0.01, 0.001
	Batch size	1, 10, 20, 50, 100
	Epoch	25, 50, 100, 150, 200
	Dropout	0.05, 0.1, 0.2







Deep Neural Network

Figure 8. Network configuration of CB-LSTM.

Hyperparameters	Broken	Clear/Sunny	Broken/Overcast
Number of hidden layers	2	2	2
Nodes in hidden layer one	100	25	100
Nodes in hidden layer two	100	25	100
Learning rate	0.001	0.001	0.001
Batch size	1	1	1
Epochs	200	200	100
Dropout	0.2	0.2	0.2
Model parameters	127,608	9408	127,608

Table 5. Optimal hyperparameter settings for CB-LSTM in different cloud conditions.

3.6. Evaluation

Forecasting performance was evaluated using three performance evaluation metrics, namely, mean absolute error (MAE) [39], RMSE [40], and NRMSE [39]. They are defined as follows:

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

RMSE =
$$\sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}}$$
 (2)

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}}}{\mu}$$
(3)

NRMSE(%) =
$$\frac{\sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}}}{\mu} * 100\%$$
 (4)

4. Result and Discussion

Figure 9 shows the average value and standard deviation of the clearness index for all forecasting sites. Greater variability was observed for Idukki, Vaibhavwadi, Tiruchirappalli, and Osmanabad. Khaga had the lowest variability.

Figure 10 shows cluster-specific features, and K-medoids led to three clusters. Features were selected on the basis of the reduction in impurity scores. For Cluster 0, the mean decrease in impurity was highest for features Kt slope (1-h) and KtPrev15. For Cluster 1, cloud cover squared and KtPrev15-nearby-mean were the most important features. For Cluster 2, cloud cover squared and Kt Slope (1-h) were the most important features. To understand the cloud type of each cluster, we calculated the percentage of cloud-cover information falling in each cluster, and this is illustrated in Figure 11. For Cluster 0, the majority of observations belonged to the broken-cloud type. In Cluster 1, the majority of observations belonged to the clear/sunny-sky type. For Cluster 2, the total numbers of observations in the broken and bvercast cloud types were relatively similar.

Table 5 shows the optimal hyperparameters of the proposed approach (CB-LSTM) for three different cloud conditions of broken, clear/sunny, and broken/overcast. Complex cloud conditions (broken or overcast) require more hidden nodes and parameters to produce good forecasting. Nevertheless, model complexity is less in clear/sunny sky conditions.

Table 6 provides information on the climatic zone-specific forecasting performance of ST-LSTM and a comparison with M-LSTM. For the composite climatic zone, ST-LSTM achieved 5.96%, 3.71%, and 8.80% less RMSE, NRMSE, and MAE, respectively, than the M-ULSTM did. For the hot and dry climatic zone, the corresponding values were 1.58%,



1.44%, and 0.25% respectively. The biggest gain was in the warm and humid climatic zone, with corresponding percentages at 8.65%, 8.34%, and 11.55% respectively.

Figure 9. Average value and variability of clearness index of forecasting sites.



Figure 10. Cluster-specific best features in terms of mean decrease in impurity.



Figure 11. Understanding cloud patterns via cluster-specific distribution of cloud type.

		M-LSTM			ST-LSTM	
Forecasting Sites	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE
		Con	nposite			
Bhainsdehi	0.1429	0.2550	0.0870	0.1408	0.2515	0.0812
Begamganj	0.2211	0.3568	0.0952	0.2090	0.3410	0.0960
Dindori	0.2430	0.3541	0.1258	0.2164	0.3352	0.1358
		Hota	and dry			
Tiruchirappalli	0.2859	0.5426	0.1894	0.2854	0.5419	0.1888
Idukki	0.3087	0.6243	0.1957	0.2969	0.6037	0.1837
Madurai	0.2725	0.5497	0.1746	0.2704	0.5447	0.1845
Warm and humid						
Khaga	0.2253	0.3888	0.1049	0.1890	0.3300	0.1113
Vaibhavwadi	0.2911	0.4708	0.1821	0.2726	0.4423	0.1831
Osmanabad	0.3001	0.4541	0.2188	0.2896	0.4408	0.1941

Table 6. Forecasting performance of spatial LSTM compared to univariate and multivariate LSTMs.

Table 7 demonstrates the superiority of CB-LSTM over CB-ANN and ST-LSTM in terms of RMSE, NRMSE, and MAE. For the composite climatic zone, CB-LSTM outperformed CB-ANN by 27.16%, 29.49%, and 38.86% in terms of RMSE, NRMSE, and MAE, respectively. For the hot and dry climatic zone, percentages were -5.28%, 8.03%, and 8.85%, respectively. For the warm and humid climatic zone, CB-LSTM achieved 9.80%, 22.04%, and 19.94% less RMSE, NRMSE, and MAE, respectively, as compared to CB-ANN. ST-LSTM was dominated by CB-LSTM in the composite climatic zone by 33.77%, 28.49%, and 19.64% in terms of RMSE, NRMSE, and MAE, respectively. In the hot and dry climatic zone, CB-LSTM led to reductions of 35.37%, 35.26%, and 34.74% in RMSE, NRMSE, and MAE, respectively, as compared to ST-LSTM. For the warm and humid climatic zone, CB-LSTM led to corresponding reductions of 27.65%, 17.78%, and 25.34%, respectively.

The largest gain was observed in the composite climatic zone compared to CB-ANN in terms of RMSE, NRMSE, and MAE. On the other hand, compared to CB-LSTM, the greatest gain was seen in the hot and dry climatic zone. Thus, CB-LSTM led to less forecasting error than that of the M-LSTM and ST-LSTM. CB-LSTM dominated both M-LSTM and ST-LSTM at each of the three different times of day in terms of NRMSE.

		CB-LSTM			CB-ANN			ST-LSTM	
Forecasting Sites	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE
				Composit	e				
Bhainsdehi	0.1096	0.1936	0.0677	0.1569	0.2762	0.1243	0.1408	0.2515	0.0812
Begamganj	0.1148	0.2016	0.0751	0.1664	0.3118	0.1384	0.2090	0.3410	0.0960
Dindori	0.1426	0.2530	0.1079	0.1790	0.3440	0.1445	0.2164	0.3352	0.1358
			I	Hot and di	ry				
Tiruchirappalli	0.1605	0.2903	0.1094	0.1617	0.3414	0.1173	0.2854	0.5419	0.1888
Idukki	0.2118	0.4934	0.1339	0.2001	0.5288	0.1376	0.2969	0.6037	0.1837
Madurai	0.1793	0.3208	0.1193	0.1619	0.3288	0.1447	0.2704	0.5447	0.1845
			Wai	rm and hu	ımid				
Khaga	0.1483	0.2641	0.1139	0.1623	0.3195	0.1258	0.1890	0.3300	0.1113
Vaibhavwadi	0.2009	0.3599	0.1200	0.2109	0.4980	0.1475	0.2726	0.4423	0.1831
Osmanabad	0.1879	0.3757	0.1089	0.2238	0.4760	0.1595	0.2896	0.4408	0.1941

Table 7. Forecasting performance of CB-LSTM compared to multivariate and spatiotemporal LSTM.

Table 8 illustrates the climatic-zone-specific forecasting superiority of CB-LSTM compared to three benchmark models [21,23,41]. In the hot and dry climatic zone, CB-LSTM achieved maximal gain with 8.86% and 26.81% lower NRMSE compared to [21,41]. On the other hand, in the composite climatic zone, the best NRMSE was 30.56% compared to [23].

Forecasting Sites	CB-LSTM	CB-LSTM [21]		[23]		
	Con	nposite				
Bhainsdehi	19.36%	22.75%	26.09%	27.62%		
Begamganj	20.16%	20.74%	26.39%	31.18%		
Dindori	25.30%	25.09%	27.93%	34.40%		
	Hot	and dry				
Tiruchirappalli	29.03%	38.06%	47.62%	34.14%		
Idukki	49.34%	49.82%	54.04%	52.88%		
Madurai	32.08%	34.14%	47.67%	32.88%		
Warm and humid						
Khaga	26.41%	20.90%	24.02%	31.95%		
Vaibhavwadi	35.99%	45.70%	46.61%	49.80%		
Osmanabad	37.57%	48.89%	41.98%	47.60%		

Table 8. Forecasting performance of CB-LSTM compared with benchmark models in terms of NRMSE (%).

Figure 12a shows climatic-zone-specific variability in predictions of CB-LSTM in terms of NRMSE. clustering-based ANN [23] for the composite and hot and dry climatic zone, and RF-SVR [21] for the warm and humid climatic zone were the worst-performing models. For all climatic zones, CB-LSTM achieves the least prediction error.

Figure 12b shows region-specific variability in predictions of CB-LSTM in terms of NRMSE. CB-LSTM had the least prediction error in both inland and coastal regions.



Figure 12. (a) Climatic-zone-specific variability in predictions; (b) region-specific variability in predictions. The symbol "†" indicates an outlier.

Table 9 shows a comparison of the overall forecasting performance of CB-LSTM to that of three benchmark models in terms of NRMSE and mean rank. CB-LSTM showed the lowest overall NRMSE and mean rank.

Table 9. Overall forecasting performance of CB-LSTM compared to that of benchmarks.

Models	Overall NRMSE	Mean Rank
[41]	0.3803	3.11
[21]	0.3378	2.11
[23]	0.3805	3.33
CB-LSTM	0.3058	1.44

5. Conclusions

CB-LSTM achieved better forecasting performance than that of M-LSTM and ST-LSTM for all climatic zones and regions. In terms of RMSE, MAE, and NRMSE, CB-LSTM dominated M-LSTM and ST-LSTM by 32.07%, 26.50%, 30.59%, 32.26%, 26.57%, and 27.18%.

CB-LSTM also outperformed three benchmark models [21,23,41] by 10.46%, 24.42%, and 24.36% in terms of overall NRMSE. CB-LSTM achieved the best mean rank compared to all the benchmark models. This holds for all the climatic zones and regions compared to the three benchmark models.

Thus, the performance of CB-LSTM was robust under differing conditions. We also obtained insights into the common nature of cloud patterns in India, as the clustering algorithm indicated relevant features about cloud patterns that could lead to improved forecasts.

The proposed model helps grid operators in better distributing power across the national grid. A future goal is to validate our model for more locations across other climatic zones, seasons, and topographical regions of India.

Author Contributions: Conceptualization, S.M. and S.G.; methodology, S.M.; software, S.M.; validation, S.M., S.G., B.G., A.C., S.S.R., K.B. and A.G.R.; formal analysis, S.M.; investigation, S.M.; resources, S.M.; data curation, S.M., K.B. and A.G.R.; writing—original draft preparation, S.M.; writing—review and editing, S.M., S.G., B.G. and A.C.; visualization, S.M.; supervision, S.G., B.G. and A.C.; project administration, S.G., B.G. and A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: This article resulted from Indo-USA collaborative project LISA 2020 between the University of Calcutta, India, and the University of Colorado, USA, and the research was supported by the National Institute of Wind Energy (NIWE) and the Technical Education Quality Improvement Programme (TEQIP).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Elizabeth Michael, N.; Mishra, M.; Hasan, S.; Al-Durra, A. Short-Term Solar Power Predicting Model Based on Multi-Step CNN Stacked LSTM Technique. *Energies* 2022, 15, 2150. [CrossRef]
- 2. Dudley, B. BP statistical review of world energy. In *BP Statistical Review*; BP p.l.c.: London, UK, 2018; Volume 6, p. 00116.
- 3. Safi, M. India plans nearly 60% of electricity capacity from non-fossil fuels by 2027. The Guardian, 21 December 2016; p. 22.
- Gensler, A.; Henze, J.; Sick, B.; Raabe, N. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 002858–002865.
- 5. Akarslan, E.; Hocaoglu, F.O. A novel adaptive approach for hourly solar radiation forecasting. *Renew. Energy* **2016**, *87*, 628–633. [CrossRef]
- Kumar, R.; Umanand, L. Estimation of global radiation using clearness index model for sizing photovoltaic system. *Renew.* Energy 2005, 30, 2221–2233. [CrossRef]
- 7. Liu, B.; Jordan, R. Daily insolation on surfaces tilted towards equator. ASHRAE J. 1961, 10, 5047843.
- 8. Perez, R.; Ineichen, P.; Seals, R.; Zelenka, A. Making full use of the clearness index for parameterizing hourly insolation conditions. *Sol. Energy* **1990**, *45*, 111–114. [CrossRef]
- 9. Rajagukguk, R.A.; Ramadhan, R.A.; Lee, H.J. A Review on Deep Learning Models for Forecasting Time Series Data of Solar Irradiance and Photovoltaic Power. *Energies* 2020, *13*, 6623. [CrossRef]
- Inman, R.H.; Pedro, H.T.; Coimbra, C.F. Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* 2013, 39, 535–576. [CrossRef]
- 11. Yang, D.; Jirutitijaroen, P.; Walsh, W.M. Hourly solar irradiance time series forecasting using cloud cover index. *Sol. Energy* **2012**, *86*, 3531–3543. [CrossRef]

- Sanjari, M.J.; Gooi, H. Probabilistic forecast of PV power generation based on higher order Markov chain. *IEEE Trans. Power Syst.* 2016, 32, 2942–2952. [CrossRef]
- Feng, C.; Cui, M.; Hodge, B.M.; Lu, S.; Hamann, H.F.; Zhang, J. Unsupervised clustering-based short-term solar forecasting. *IEEE Trans. Sustain. Energy* 2018, 10, 2174–2185. [CrossRef]
- 14. Bandara, K.; Bergmeir, C.; Smyl, S. Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Syst. Appl.* **2020**, *140*, 112896. [CrossRef]
- 15. Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780. [CrossRef] [PubMed]
- 16. Wang, H.; Li, M.; Yue, X. IncLSTM: Incremental Ensemble LSTM Model towards Time Series Data. *Comput. Electr. Eng.* **2021**, 92, 107156. [CrossRef]
- 17. Abdel-Nasser, M.; Mahmoud, K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Comput. Appl.* **2019**, *31*, 2727–2740. [CrossRef]
- Aslam, M.; Lee, J.M.; Kim, H.S.; Lee, S.J.; Hong, S. Deep learning models for long-term solar radiation forecasting considering microgrid installation: A comparative study. *Energies* 2020, 13, 147. [CrossRef]
- Akhter, M.N.; Mekhilef, S.; Mokhlis, H.; Almohaimeed, Z.M.; Muhammad, M.A.; Khairuddin, A.S.M.; Akram, R.; Hussain, M.M. An Hour-Ahead PV Power Forecasting Method Based on an RNN-LSTM Model for Three Different PV Plants. *Energies* 2022, 15, 2243. [CrossRef]
- Qing, X.; Niu, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* 2018, 148, 461–468. [CrossRef]
- Ali-Ou-Salah, H.; Oukarfi, B.; Bahani, K.; Moujabbir, M. A New Hybrid Model for Hourly Solar Radiation Forecasting Using Daily Classification Technique and Machine Learning Algorithms. *Math. Probl. Eng.* 2021, 2021, 6692626. [CrossRef]
- 22. Yagli, G.M.; Yang, D.; Srinivasan, D. Automatic hourly solar forecasting using machine learning models. *Renew. Sustain. Energy Rev.* 2019, 105, 487–498. [CrossRef]
- McCandless, T.; Haupt, S.; Young, G. A regime-dependent artificial neural network technique for short-range solar irradiance forecasting. *Renew. Energy* 2016, 89, 351–359. [CrossRef]
- Yu, Y.; Cao, J.; Zhu, J. An LSTM short-term solar irradiance forecasting under complicated weather conditions. *IEEE Access* 2019, 7, 145651–145666. [CrossRef]
- Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* 2017, 105, 569–582. [CrossRef]
- Berndt, D.J.; Clifford, J. Using dynamic time warping to find patterns in time series. In Proceedings of the KDD Workshop, Seattle, WA, USA, 31 July 1994; Volume 10, pp. 359–370.
- 27. Park, H.S.; Jun, C.H. A simple and fast algorithm for K-medoids clustering. Expert Syst. Appl. 2009, 36, 3336–3341. [CrossRef]
- Sharma, N.; Mangla, M.; Yadav, S.; Goyal, N.; Singh, A.; Verma, S.; Saber, T. A sequential ensemble model for photovoltaic power forecasting. *Comput. Electr. Eng.* 2021, 96, 107484. [CrossRef]
- 29. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2222–2232. [CrossRef]
- Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J. Clim. 2017, 30, 5419–5454. [CrossRef]
- Marion, B.; Kroposki, B.; Emery, K.; Del Cueto, J.; Myers, D.; Osterwald, C. Validation of a Photovoltaic Module Energy Ratings Procedure at NREL; Technical Report; National Renewable Energy Lab.: Golden, CO, USA, 1999. [CrossRef]
- 32. Müller, M. Dynamic time warping. In *Information Retrieval for Music and Motion*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 69–84. [CrossRef]
- 33. Patro, S.; Sahu, K.K. Normalization: A preprocessing stage. arXiv 2015, arXiv:1503.0646.
- Krishna, T.S.; Babu, A.Y.; Kumar, R.K. Determination of optimal clusters for a Non-hierarchical clustering paradigm K-Means algorithm. In *Proceedings of International Conference on Computational Intelligence and Data Engineering*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 301–316. [CrossRef]
- 35. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 448–456.
- 36. Cortes, C.; Mohri, M.; Rostamizadeh, A. L2 regularization for learning kernels. arXiv 2012, arXiv:1205.2653.
- Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
- Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for hyper-parameter optimization. In Proceedings of the 25th Annual Conference on Neural Information Processing Systems (NIPS 2011), Granada, Spain, 12–17 December 2011, Volume 24.
- Shcherbakov, M.V.; Brebels, A.; Shcherbakova, N.L.; Tyukov, A.P.; Janovsky, T.A.; Kamaev, V.A. A survey of forecast error measures. World Appl. Sci. J. 2013, 24, 171–176. [CrossRef]
- 40. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* 2014, 7, 1247–1250. [CrossRef]
- 41. Zang, H.; Liu, L.; Sun, L.; Cheng, L.; Wei, Z.; Sun, G. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renew. Energy* **2020**, *160*, 26–41. [CrossRef]