Available online at www.sciencedirect.com

SciVerse ScienceDirect

A Deep Learning based Hybrid Method for Hourly Solar Radiation Forecasting

Chun Sing Lai [a,b], Cankun Zhong [c], Keda Pan [a],

Wing W. Y. Ng [c], Loi Lei Lai [a,*]

[a] Department of Electrical Engineering, School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[b] Brunel Institute of Power Systems, Department of Electronic and Electrical Engineering, Brunel University London, London, UB8 3PH, UK
[c] Guangdong Provincial Key Laboratory of Computational Intelligence and Cyberspace Information, School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

Email addresses: chunsing.lai@brunel.ac.uk (C. S. Lai); curran.z@qq.com (C. Zhong); 1111904017@mail2.gdut.edu.cn (K. Pan); wingng@ieee.org (W. W. Y. Ng); l.l.lai@ieee.org (L. L. Lai)

* Corresponding author

Abstract

Solar radiation forecasting is a key technology to improve the control and scheduling performance of photovoltaic power plants. In this paper, a deep learning based hybrid method for 1-hour ahead Global Horizontal Irradiance (GHI) forecasting is proposed. Specifically, a deep learning based clustering method, deep time-series clustering, is adopted to group the GHI time series data into multiple clusters to better identify its irregular patterns and thus providing a better clustering performance. Then, the Feature Attention Deep Forecasting (FADF) deep neural network is built for each cluster to generate the GHI forecasts. The developed FADF dynamically allocates different importance to different features and utilizes the weighted features to forecast the next hour GHI. The solar forecasting performance of the proposed method is evaluated with the National Solar Radiation Database. Simulation results show that the proposed method yields the most accurate solar forecasting among the smart persistence and state-of-the-art models. The proposed method reduces the root mean square error as compared to the smart persistence by 11.88% and 12.65% for the Itupiranga and Ocala dataset, respectively.

Keywords: Solar forecasting, deep learning, clustering, feature attention

Article Highlights:

1. Proposed a deep learning clustering method for solar irradiance feature learning.
2. Hourly solar forecasting with Feature Attention based Deep Forecasting (FADF).
3. RMSE reduced by 11.88%-12.65% as compared to smart persistence.

1. Introduction

Smart grids and cities concern with anticipating situations, with an efficient forecast of weather conditions (e.g., solar energy forecasting) and the possibility of making decisions in almost real-time. Solar forecasting is crucial in managing power network operations and solar photovoltaic applications (Huang et al., 2018; Lai et al., 2017a; Wang et al., 2018, 2019). High penetration of intermittent renewable sources and the lack of anticipatory capabilities will result in uneconomic choices (such as the severe curtailment of generation) or lack of resilience when facing faults and disturbances. The accurate forecasting of Global Horizontal Irradiance (GHI) is beneficial to quality future power productions.

In previous studies, many data-driven approaches are proposed for short-term solar forecasting. These approaches can be mainly divided into three parts: the physical methods, the statistical methods, and the machine learning methods. The physical methods consist of a set of mathematical equations describing the physical state and dynamic motion of the atmosphere (Zhang et al., 2018), and its forecasting performance is highly affected by the sharp changes in meteorological variables. The statistical methods utilizing statistical analysis of the different input features for solar forecasting include the auto-regressive integrated moving average, the exponential smoothing, and the Markov Chain model (Shakya et al., 2017), among others. Recently, machine learning based approaches have been widely used in modeling, design and prediction in solar energy systems. The combination of two or more machine learning methods is also used to provide a more accurate solar forecasting result known as the hybrid model.

From the literature, a collection of classical machine learning methods has been proposed and applied in the solar energy system. For example, an Artificial Neural Network (Fermín et al., 2018) model was developed to predict the levelized electricity cost of two parabolic trough solar thermal power plants coupled with a fuel backup system and thermal energy storage (Boukelia et al., 2017). The Support Vector Machine was adopted (Ma et al., 2017) to upgrade the estimation accuracy of solar irradiance levels from photovoltaic electrical parameters. Also, the potential of the Random Forest method for estimating solar radiation using air pollution index was assessed (Sun et al., 2016).

With the development and improvement of deep learning algorithms which are one of the advanced machine learning methods, there are additional deeper learning methods being applied to renewable energy challenges. Convolutional Neural Network (CNN) was used (Sun et al., 2018) to forecast the solar Photovoltaic (PV) output using the contemporaneous images of the sky. The meteorological features were utilized (Qing et al., 2018) as the input for a Long Short-Term Memory neural network (LSTM) for day ahead hourly solar radiance prediction. Furthermore, studies (Ghimire et al., 2019; Yan et al., 2020) applied CNN to robustly extract features from predictive variables while the LSTM or GRU was utilized to absorb the features for solar radiation forecasting. The study (Feng and Zhang, 2020) adopted the CNN to forecast the solar PV output using the contemporaneous images of the sky, while the study (Zhen et al., 2020) firstly assigns the sky image to the corresponding class using the deep clustering method and then utilizes a corresponding hybrid deep learning method for PV power forecasting.

Due to the limitation of a stand-alone machine learning method, the hybrid model combining multiple machine learning methods was conducted to improve the solar forecasting accuracy. The Adaptive Neuro-Fuzzy Interface Systems and the Wavelet Neural Network are among the early generation of hybrid models (Faizollahzadeh et al., 2018; Fotovatikhah et al., 2018). The study (David et al., 2016) evaluated performances of an ensemble of Autoregressive Moving Average (ARMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models to establish solar irradiance probabilistic forecasts. Still based on the ensemble learning, two advance base models, namely extreme gradient boosting forest and deep neural networks (XGBDNN), were proposed for hourly global horizontal irradiance forecast (Kumari et al., 2021). A two-stage method Coral Reefs Optimization Extreme Learning Machine (CRO-ELM) was applied (Salcedo-Sanz el al., 2018a) to select useful features and use selected features for solar forecasting. Since solar features are highly influenced by weather conditions, the combination of clustering and regression was conducted. A novel clustering method TB_K-means was proposed (Azimi et al., 2016) to partition the solar data into several clusters where the multiple layer perceptron was developed for each cluster to forecast hourly solar radiation. The similar hybrid forecasting method can also be found in other studies (Li et al., 2017; Feng et al., 2018; Fu et al., 2019), where the traditional clustering methods were utilized to categorize the data based on the original solar radiation sequence or the pre-extracted features. The hierarchical clustering technique was utilized in a different way (Sun et al., 2018) where the K-means clustering was used to cluster the forecasting results of each sub-component generated from the Ensemble Empirical Mode Decomposition method, and a least square support vector regression was applied to ensemble the sub-component forecasts of each cluster. Similar study (Theocharides et al., 2020) was conducted where the K-means was applied to categorize the forecasted daily GHI and for each cluster, coefficients were obtained by linear regression to correct the forecasted outputs of the machine learning model.

Most of the existing hybrid methods utilizing the clustering techniques for solar forecasting usually adopt traditional clustering methods (Azimi et al., 2016; Li et al., 2017; Feng et al., 2018; Fu et al., 2019) which may result in sub-optimal clustering outcomes because the feature extraction and clustering are conducted in two separate independent stages rather than jointly considered. Besides, the GHI features (e.g., historical GHI, clear-sky GHI) and the meteorological features (e.g., temperature, wind speed) are often used to forecast the future solar irradiance (Jeno and Kim, 2020; Pan et al., 2020; Wu et al., 2020). The features are treated equally to forecast solar radiation in some previous studies. Intuitively, the historical solar irradiance should play a more important role than other features for solar forecasting. Some studies consider possible redundant features that may hurt the solar forecasting performance and thus utilize feature selection methods to select an optimal feature subset from the original feature set for improving solar forecasting performance (Salcedo-Sanz et al., 2018b; Niu et al., 2020; Qadir et al., 2021). In this work, instead of utilizing the feature selection, a deep learning based feature weighting method is proposed alternatively to automatically enhance more useful features and restrain less important features for solar forecasting. Because the feature weighting can be observed as a generalization of feature selection where the feature weights are not limited to 0 or 1.

The proposed deep learning based hybrid method in this work consists of the Deep Time-series Clustering (DTC) and the Feature Attention based Deep Forecasting (FADF). The DTC groups solar time series with similar patterns into the same clusters using high-level useful features extracted by the Gated Recurrent Unit neural network (GRU) (Chung et al., 2014), where the feature learning and clustering are learned jointly. Each cluster has a corresponding hourly solar predictor (i.e., the FADF) trained with the data in the corresponding cluster. The FADF of each cluster utilizes a Feature Attention Sub-network to determine the feature importance (Song et al., 2018) and sends the weighted input to the main GRU network for hourly solar forecasting. The major contributions of this work are as follows:

1) A deep learning based clustering method (i.e., Deep Time-series Clustering, DTC) is proposed to group the solar irradiance (i.e., GHI) data with similar patterns into the same clusters. It optimizes the feature learning and clustering assignment simultaneously. DTC treats the feature learning and clustering assignment in two separate stages.
2) A Feature Attention based Deep Forecasting method (FADF) is proposed for hourly solar radiation forecasting of each cluster grouped by the DTC. The FADF utilizes a feature attention mechanism to dynamically allocate different importance to different features at each time step for 1-hour ahead

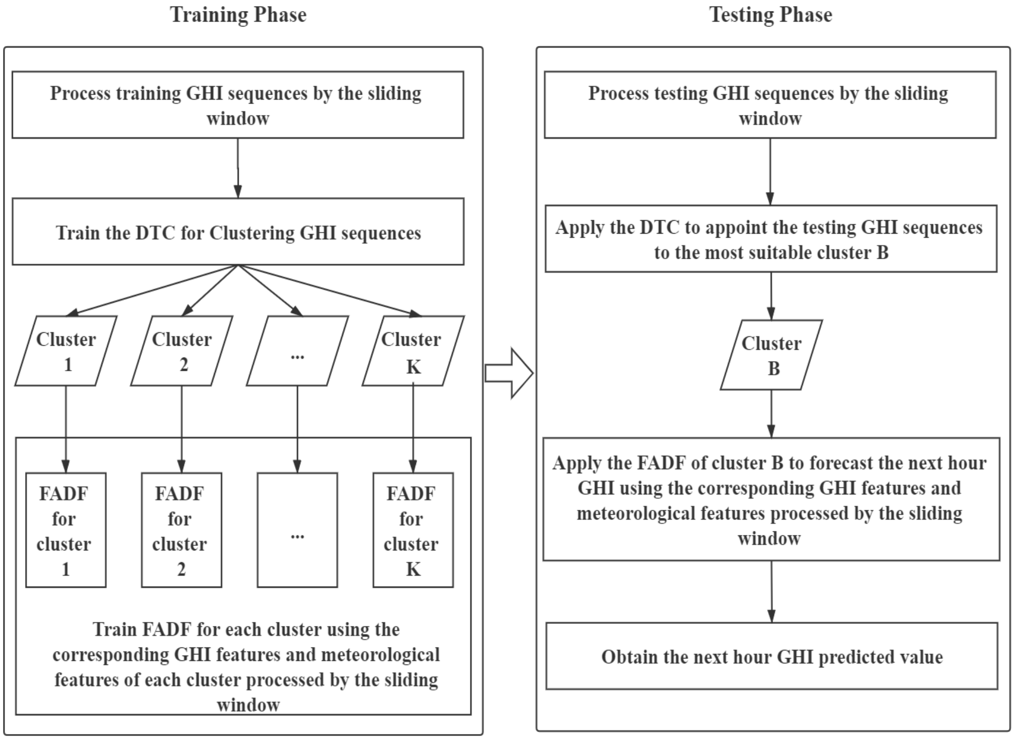
GHI forecasting.

3) Extensive simulations are carried out to confirm the superiority of the proposed method. Simulation results on the National Solar Radiation Database show that the proposed method outperforms existing methods for solar forecasting in most cases.

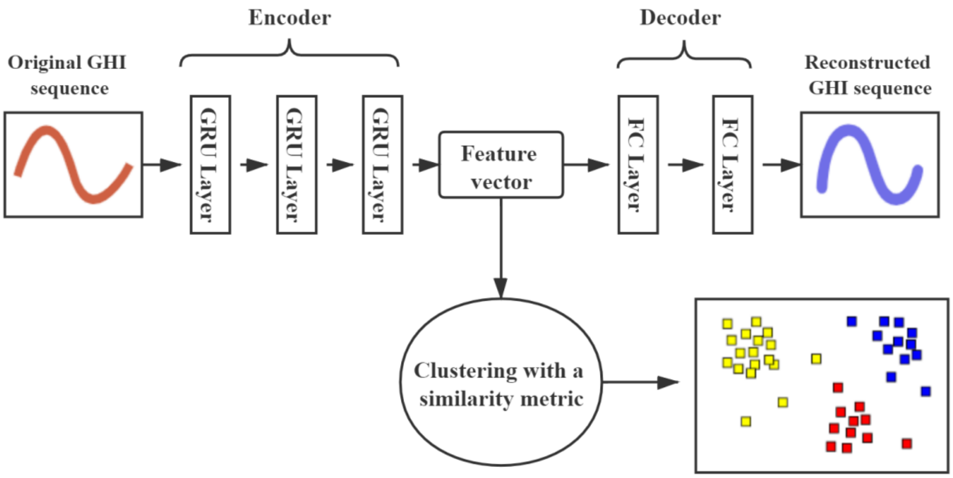
The remainder of this paper is organized as follows. Section 2 describes the proposed method for 1-hour ahead GHI forecasting. Section 3 gives the simulation results and detailed discussion. Finally, a conclusion is given in Section 4.

2. The Hybrid Method for One-hour Ahead GHI Forecasting

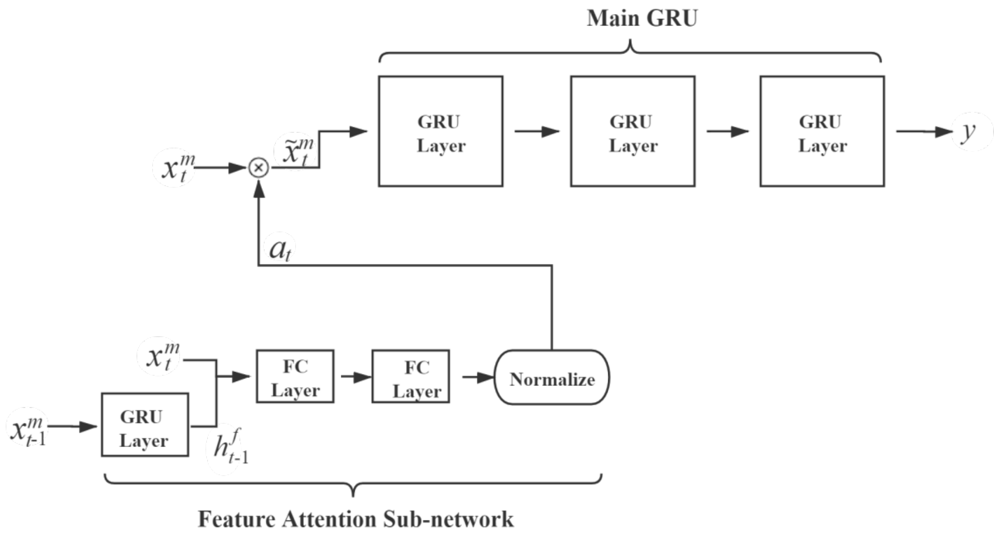
The training and testing phase of the proposed deep learning based hybrid method for 1-hour ahead GHI forecasting is shown in Fig. 1. The historical GHI time series of finite length (specified by the window size) is sent to the DTC to get its clustering label. Then, the FADF of the corresponding cluster assigns the feature importance of the GHI features (including historical GHI, clear-sky GHI, clear-sky index and solar zenith angle) and the meteorological features (temperature, relative humidity, wind speed, wind direction, and pressure) and utilizes the weighted features to forecast the next hour GHI.



(a) The training and testing phase of the proposed method.



(b) The overview of DTC.



(c) The overview of FADF.

Fig. 1. The proposed deep learning based hybrid method for 1-hour ahead GHI forecasting.

In the following section, we first introduce DTC for grouping the historical GHI time series in Section 2.1. Then, FADF is proposed for hourly GHI forecasting and details are given in Section 2.2.

2.1. Deep Time-series Clustering (DTC)

The DTC utilizes an encoder-decoder framework to learn the latent representation Z of the original data X^s (i.e., GHI time sequences) for clustering, where the encoder is a GRU model mapping the original GHI time sequences into the latent representation and the decoder is a Multiple Layer Perceptron (MLP) model reconstructing the latent representation into the original ones. It should be noted that the capacity of the decoder is lower than that of the encoder in this work such that the encoder is forced to learn the latent representation effectively instead of highly relying on the decoder.

The clustering is carried out on the latent representation space. After the initial clustering centers in the latent representation space are given, the clustering centers and the latent representation of the original data are updated by applying the clustering loss and reconstruction loss jointly. The clustering loss is responsible for forcing the data in the latent representation space to move closer to the corresponding clustering center while the reconstruction loss is utilized to preserve intrinsic structure in data and avoid the distortion of the latent representation space (Guo et al., 2020).

1) Gated Recurrent Unit Neural Network (GRU)

The GRU is invented to capture the long-term dependency of the time sequence data. The GRU consists of three basic components: the reset gate *r*, the update gate *p*, and the hidden state *h*. The hidden state acts as a memory storing useful information extracted by the reset gate and update gate. Specifically, at time step *t*, the reset gate takes the current time step input x_t^s (i.e., the GHI value at time step *t*) to determine how much information stored in the previous hidden state $\widetilde{h}_{t-1}^s$ should be ignored and to obtain a new candidate hidden state $\widetilde{h}_t^s$ of current time step while the update gate is used to decide how much memories should be updated by the candidate hidden state. This is implemented with the following equations:

$$r_t = \sigma(W_r \widetilde{h}_{t-1}^s + U_r x_t^s + b_r) \tag{1}$$

$$\widetilde{h}_t^s = \tanh(W_{\widetilde{h}}(r_t * h_{t-1}^s) + U_{\widetilde{h}} x_t^s + b_{\widetilde{h}}) \tag{2}$$

$$p_t = \sigma(W_p \widetilde{h}_{t-1}^s + U_p x_t^s + b_p) \tag{3}$$

$$h_t^s = (1 - p_t) * \widetilde{h}_t^s + p_t * \widetilde{h}_{t-1}^s \tag{4}$$

$$y_t = W_y h_t^s + b_y \tag{5}$$

where the $\sigma(\cdot)$ is the sigmoid function and the $\tanh(\cdot)$ is the hyperbolic tangent function. W_j, U_j, and b_j ($j \in \{r, p, \widetilde{h}, y\}$) are all trainable parameters. * denotes the element-wise multiplication operation.

2) Training of DTC

The DTC is first trained with the reconstruction loss only to provide a good initialization. Then, the K-means clustering (Lai et al., 2017b) is performed on the latent representation space to derive the initial cluster centers. Finally, the DTC model is fine-tuned with the clustering loss and reconstruction loss jointly.

The reconstruction loss is shown as below:

$$L_{\text{reconstruction}} = \left\| X - g_{dec}(f_{enc}(X^s)) \right\|_2^2 \tag{6}$$

where *X* is the original data, the $f_{enc}(\cdot)$ is the encoder (i.e., the GRU model), and the $g_{dec}(\cdot)$ is the decoder (i.e., the MLP model).

The clustering loss (Xie et al., 2016) consists of the Student's *t*-distribution *Q* measuring the similarity of the latent representation data point z_i and the cluster center μ_j, and the powerful auxiliary target distribution P to refine the clusters. This is shown in the following equations:

$$L_{\text{clustering}} = KL(P \| Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{7}$$

$$q_{ij} = \frac{(1 + \|z_i - \mu_j\|^2)^{-1}}{\sum_j (1 + \|z_i - \mu_j\|^2)^{-1}} \tag{8}$$

$$p_{ij} = \frac{q_{ij}^2 / \sum_i q_{ij}}{\sum_j (q_{ij}^2 / \sum_i q_{ij})} \tag{9}$$

where $KL(P\|Q)$ is the Kullback-Leibler divergence formulating the divergence between Student's t-distribution Q and the auxiliary target distribution P. Note that the distribution Q serves as the soft assignment of the clustering label. The auxiliary target distribution P provides the supervision information by using high confidential samples as supervision and then makes samples in each cluster distribute more densely.

The parameters (parameters of the encoder, parameters of the decoder, and the cluster center) of the DTC model are updated using the Adaptive Moment Estimation algorithm. Assume that the parameter of the encoder is W_{enc} and the parameter of the decoder is W_{dec}. The parameters are updated using the following equations:

$$\mu_j = \mu_j - \frac{\lambda}{n} \sum_{i=1}^{n} \frac{\partial L_{\text{clustering}}}{\partial \mu_j} \tag{10}$$

$$W_{\text{dec}} = W_{\text{dec}} - \frac{\lambda}{n} \sum_{i=1}^{n} \frac{\partial L_{\text{reconstruction}}}{\partial W_{\text{dec}}} \tag{11}$$

$$W_{\text{enc}} = W_{\text{enc}} - \frac{\lambda}{n} \sum_{i=1}^{n} \left(\frac{\partial L_{\text{reconstruction}}}{\partial W_{\text{enc}}} + \gamma \frac{\partial L_{\text{clustering}}}{\partial W_{\text{enc}}} \right) \tag{12}$$

where λ, n and γ denote the learning rate, the number of samples, and the coefficient balancing the clustering loss and the reconstruction loss respectively.

2.2. Feature Attention based Deep Forecasting (FADF)

The proposed FADF utilizes a Feature Attention Sub-network explicitly to weight the input feature and then send the weighted feature into the main GRU model to forecast the next hour GHI. In this work, the Feature Attention Sub-network is the GRU model and takes the previous hidden state h^f_{t-1} and the input of current time step x_t^m (including the GHI features and the meteorological features at time step t) as input to compute the normalized weight α_t of the input feature at the current time step.

1) Feature Attention Sub-network

The Feature Attention Sub-network automatically explores the different degrees of importance of features at each time step by jointly taking the previous hidden state h^f_{t-1} and the input of current time step x_t^m into account. The previous hidden state h^f_{t-1} stores the information of the past inputs benefiting from the merit of GRU while the input of current time step x_t^m serves as the indispensable element in determining their importance. With a soft attention mechanism, the weighted input of current time step $\widetilde{x}_t^m$ is as follows:

$$e_t = U_f \tanh(W_{xf} \cdot x_t^m + W_{hf} \cdot h_{t-1}^f + b_f) \tag{13}$$

$$a_t = \mathrm{softmax}(e_t) \tag{14}$$

$$\widetilde{x}_t^m = x_t^m * a_t \tag{15}$$

where $\mathrm{softmax}(\cdot)$ is the exponential normalization function. W_{xf}, Wh_f, and b_f are trainable parameters.

The Feature Attention Sub-network can also be viewed as a gating mechanism which adaptively controls the amount of information of each feature to flow to the main GRU model to forecast the next hour GHI value.

2) Training of FADF

In this work, the proposed deep learning model FADF is trained with the huber loss instead of the commonly used Mean Squared Error (MSE) loss as reported in previous study (Mosavi et al., 2019).

One of the major drawbacks of the MSE loss is that it is too sensitive to the outliers but meanwhile the MSE loss can help the model converge because the gradient of MSE loss decreases as the loss decreases. Sometimes the GHI sequence changes dramatically in the short-term due to extreme weather conditions, which results in the outliers.

The huber loss is the combination of the Mean Average Error (MAE) loss and the MSE loss. It not only has the advantage of the MAE loss which is not sensitive to outliers, but also it has the advantage of the MSE loss whose gradient decreases as it decreases to avoid the divergence caused by a large gradient. Assuming that y is the real next hour GHI value and the $f(x^m)$ is the prediction of the next hour GHI value, then the huber loss is given by the following equation:

$$L_{huber}(y, f(x^m)) = \begin{cases} \frac{1}{2}(y - f(x^m))^2 & \text{for } |y\text{-}f(x^m)| \le \delta \\ \delta\,|y\text{-}f(x^m)| - \frac{1}{2}\delta^2 & \text{otherwise.} \end{cases} \tag{16}$$

where δ is a hyper-parameter. When the Huber loss is between [0-δ, 0+δ], it is equivalent to MSE loss, and when it is at [-∞, δ] and [δ, +∞], it is equivalent to rRMSE loss.

The proposed FADF consists of the main GRU model for hourly GHI forecasting and the Feature Attention Sub-network for feature weighting at each time step. The weighted input derived from the Feature Attention Sub-network directly influences the forecasting performance of the main GRU model and the prediction error loss guides the training of the Feature Attention Sub-network. Due to the mutual influence of these two networks, the optimization is rather difficult. Therefore, a training strategy of FADF is proposed and described in Algorithm 1.

Algorithm 1 Training strategy of the FADF

Input: Training Dataset (X^m,Y), Randomly Initialized Feature Attention Sub-network G_{FASN}, Randomly Initialized Decoder G_{MLP}, Randomly Initialized Main GRU model G_{GRU}

Output: Trained Feature Attention Sub-network G''_{FASN}, Trained Main GRU model G''_{GRU}

1. The Feature Attention Sub-network serves as the encoder and a MLP serves as the decoder. Pre-train the Feature Attention Sub-network with the Auto-encoder framework using the MSE loss, i.e., min $\|X^m - G_{MLP}(G_{FASN}(X^m))\|^2$, and get the pre-trained Feature Attention Sub-network G'_{FASN}.

2. Fix the pre-trained Feature Attention Sub-network G'_{FASN}. Train the main GRU model G_{GRU} according to the huber loss, i.e., $L_{huber}(Y,G_{GRU}(G'_{FASN}(X^m)))$, and get the pre-trained main GRU model G'_{GRU}.

3. Jointly fine-tune the pre-trained Feature Attention Sub-network G'_{FASN} (obtained from Step 1) and the main GRU model G'_{GRU} (obtained from Step 2) according to the huber loss, i.e., $L_{huber}(Y,G'_{GRU}(G'_{FASN}(X^m)))$, and get the trained Feature Attention Sub-network G''_{FASN} and Main GRU model G''_{GRU}.

2.3. Time Complexity Analysis

The time complexity of the proposed hybrid method in the testing phase consists of two parts, namely the time complexity of assigning a GHI sequence of the testing sample to the corresponding cluster by DTC and, the time complexity of the next hour GHI forecasting with the testing sample using the FADF of the corresponding cluster. Note that the GRU model is the basic feature learning component of the DTC and FADF, thus the time complexity of the GRU model without the output layer is given first:

$$O(T(d_1(3d_x+3d_1)+\sum_{i=1}^{L-1}d_{i+1}(3d_i+3d_{i+1})) \tag{17}$$

where T, L, d_x, and d_i denote the length of the input sequence, the total number of the hidden layers, the feature dimension of the input layer, and the dimension of the i^{th} hidden layer, respectively. Note that the time complexity of the element-wise operation is ignored here to simplify the expression.

For the DTC, a GHI sequence is served as the input and the output of the final hidden layer is served as the feature vector. The distances between the feature vector and each cluster center vector are calculated to assign the corresponding cluster. Therefore, the time complexity of the DTC is given as follows:

$$O(T(d_1(3+3d_1)+\sum_{i=1}^{L^{DTC}-1}d_{i+1}(3d_i+3d_{i+1}))+O(Kd_L^{DTC}) \tag{18}$$

where K, L^{DTC}, and d_L^{DTC} denote the number of cluster centers, the number of the hidden layers of the DTC model, and the dimension of the final hidden layer of the DTC model.

Two GRU models are utilized for FADF. They are the Feature Attention Sub-network (FASN) and the main GRU (MGRU). The FASN takes the input sequences to produce the corresponding feature vectors to calculate the feature weight of each time step. The MGRU takes the weighted input to forecast the next-hour GHI. Therefore, the time complexity of the FADF is calculated with the following equation:

$$\begin{aligned}&O(T(d_1(3d_x+3d_1)+\sum_{i=1}^{L^{FASN}-1}d_{i+1}(3d_i+3d_{i+1}))+\\&O(T(d_1(3d_x+3d_1)+\sum_{i=1}^{L^{MGRU}-1}d_{i+1}(3d_i+3d_{i+1}))+\\&O(T(2d_xd_h+d_L^{FASN}))+O(d_L^{MGRU})\end{aligned} \tag{19}$$

where LFASN, LMGRU, d_L^{FASN}, d_L^{MGRU}, and d_h denote the number of hidden layer of the FASN, the number of hidden layer of the MGRU, the dimension of the final hidden layer of the FASN, the dimension of the final hidden layer of the MGRU, and the dimension of the hidden layer in the attention module, respectively.

3. Simulations and Results

3.1. Simulation Setup

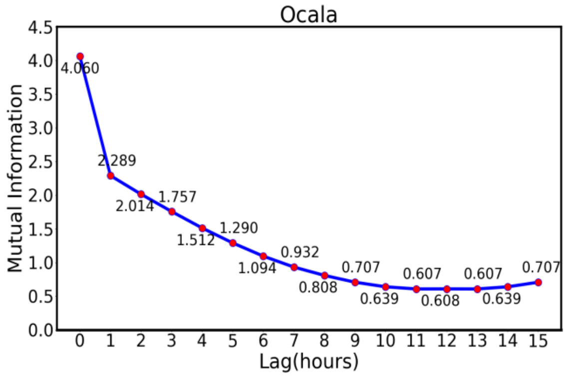
1) Data

This study employs two 12-year (from 2005 to 2017) hourly datasets collected from the National Solar Radiation Database (Sengupta et al., 2018) to train and test the 1-hour ahead GHI forecasting models. One dataset is based on Itupiranga (latitude = 5.15° S, longitude = 49.34° W), Brazil and the other one is based on Ocala (latitude = 29.17° N, longitude = 82.14° W), Marion, Florida, United States.

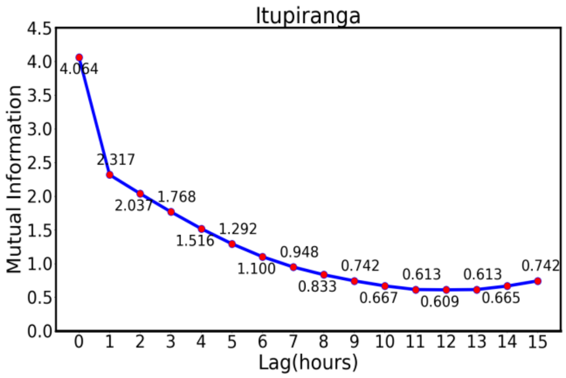
In this work, all the hourly data from 2005 to 2014 were used as the training set, the data from 2015 to 2016 were utilized as the validation set for determining the hyper-parameters (e.g., the number of clusters), and data from 2017 were served as the testing set.

2) Implementation Details

The features from the two information sources were utilized to establish the data-driven models for hourly GHI forecasting. They are: (i) the GHI features: Historical GHI, clear-sky GHI, clear-sky index and solar zenith angle; (ii) the meteorological features: Temperature, relative humidity, wind speed, wind direction, and pressure.



(a)



(b)

Fig. 2. Mutual information of various time-lags in a) Ocala and b) Itupiranga.

Moreover, the length (i.e., the window size of sliding window) of lagged data plays a crucial role in determining the optimal structure of data-driven models for hourly GHI forecasting. Obviously, if the window size is too small, then the historical information may be not rich enough for a model to forecast the next hour GHI correctly; if the window size is too large, then too much redundant information will be fed into the data-driven model to cause the model over-fitting the training data. In this work, the mutual information was used to calculate both the linear and nonlinear cross-correlation of the GHI time series with itself at different time steps while the first minimum criterion usually considered in evaluating in mutual information (Ghimire et al., 2019) was adopted to determine the optimal window size. The main idea of the first minimum criterion is that two samples can be considered statistically independent if they are delayed by a number of samples, equal to the time needed for the mutual information to reach the first minimum. Therefore, as shown in Fig. 2, the sliding window size for the Ocala dataset and the Itupiranga dataset were set as 11 hours and 12 hours, respectively.

To validate the effectiveness of our proposed method, several recent works were chosen to compare with our proposed method. For simplicity, our proposed method is denoted as HYBRID in the following section. The brief introduction of comparison methods is given as follows:

(i) Smart Persistence (Yang et al., 2019): The smart persistence is the persistence model which is often referred to as the baseline in previous studies. Its main idea is to predict the next hour GHI by assuming that the next hour clear-sky index is the same as the current hour clear-sky index, as shown below:

$$\hat{y}(t+1) = k_{cs}(t) I_{cs}(t+1) \tag{20}$$

where $k_{cs}(t)$, $I_{cs}(t+1)$ and $\hat{y}(t+1)$ represent the clear-sky index at time t, the clear-sky GHI at time $t+1$, and the predicted GHI at time $t+1$ respectively.

(ii) TB_K-means+MLP (Azimi et al., 2016): A new clustering method TB_K-means is proposed to partition the GHI time series data into k clusters and each cluster has its corresponding GHI predictor which is the Multiple Layer Perceptron (MLP).

(iii) LSTM (Qing et al., 2018): The LSTM with meteorological features as input is utilized to achieve the day ahead GHI forecasting.

(iv) C_LSTM (Ghimire et al., 2019): The C_LSTM exploits a CNN to extract local temporal features and then a LSTM takes these local temporal features as input to forecast the GHI.

(v) ResInceptionGRUAttn (Yan et al., 2020): The ResInceptionGRUAttn uses two CNN-based Inception structures (Inception_ResNet and InceptionV3) to achieve feature extraction and then a two-layer GRU with attention mechanism is utilized to make predictions.

(vi) XGBDNN (Kumari et al., 2021): Two advance base models (extreme gradient boosting and deep neural network) are utilized for solar forecasting. Multiple extreme gradient boosting are used to build an XGB forest and the ridge regression is utilized to integrate the XGB forest and the DNN to avoid the over-fitting problem.

In order to make a fair comparison, the features fed to all data-driven models (except the Smart Persistence) are the same.

3）Evaluation Criteria

Several commonly used evaluation metrics are employed to validate the forecasting performances of GHI prediction models. They are the Root Mean Square Error (RMSE), Relative Root Mean Square Error (rRMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), Maximum Error ($Error_{max}$), Minimum Error ($Error_{min}$), and Forecast Skill (FS) as shown in (21), (22), (23), (24), (25), (26), and (27) respectively.

$$\text{RMSE}(\text{W/m}^2)=\sqrt{\frac{1}{U}\sum_{t=1}^{U}(f(t)-y(t))^2} \tag{21}$$

$$\text{rRMSE}(\%)=\frac{\sqrt{\frac{1}{U}\sum_{i=1}^{U}\left(f(t)-y(t)\right)^2}}{\frac{1}{U}\sum_{i=1}^{U}y(t)}\times 100 \tag{22}$$

$$\text{MAE}(\text{W/m}^2)=\frac{1}{U}\sum_{t=1}^{U}\left|f(t)-y(t)\right| \tag{23}$$

$$\text{R}^2(\%)=\left(1-\frac{\sum_{t=1}^{U}(f(t)\text{-}y(t))^2}{\sum_{t=1}^{U}(y(t)\text{-}\frac{1}{U}\sum_{t=1}^{U}y(t))^2}\right)\times 100 \tag{24}$$

$$\text{Error}_{\max}(\text{W/m}^2)=\max\left|f(t)-y(t)\right| \tag{25}$$

$$\text{Error}_{\min}(\text{W/m}^2)=\min\left|f(t)-y(t)\right| \tag{26}$$

$$\text{FS}(\%)=\left(1\text{-}\frac{\text{rRMSE}_{\text{proposed}}}{\text{rRMSE}_{\text{baseline}}}\right)\times 100 \tag{27}$$

where $f(t)$, $y(t)$, and U represent the predicted GHI of t^{th} testing sample, the real GHI of t^{th} testing sample, and the total number of testing samples, respectively. The $\text{rRMSE}_{\text{proposed}}$ is the rRMSE of the model under evaluation and $\text{rRMSE}_{\text{baseline}}$ is the rRMSE of the baseline i.e., the Smart Persistence method.

3.2. Results and Analysis

1) Ablation Study

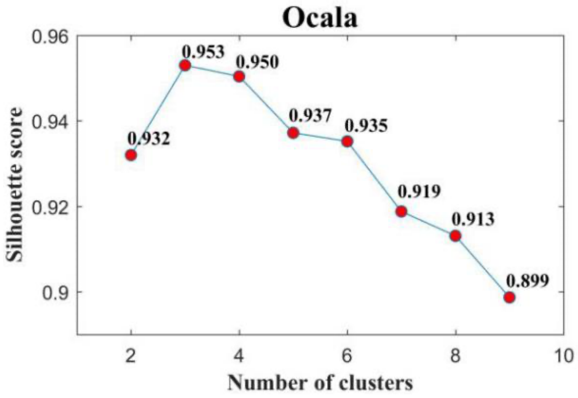
The proposed hourly GHI forecasting method HYRBID mainly consists of two components. They are the DTC and the FADF. To validate the effectiveness of each component, an ablation study was conducted. As shown in Table 1, the performance of FADF is better than the GRU, which lacks the feature attention mechanism compared with the FADF on two testing sets under the RMSE and rRMSE metrics (the best performance is marked in bold black). It shows the validity of the feature weighting idea and the successful design of the feature attention sub-network. The feature attention sub-network can dynamically assign a higher weight to the more important feature for hourly GHI forecasting at each time step, thus FADF achieves higher forecasting accuracies compared with the single GRU. Furthermore, after combining the FADF with the DTC (thus resulting in the HYBRID), the performance is further boosted (i.e., lower RMSE and lower rRMSE). Training a single FADF model on the data with several GHI time series patterns may force the model to learn the common and universal features for next hour GHI forecasting and thus the model fails to achieve the performance as good as the HYBRID. It indicates that grouping the data with similar GHI time series patterns into the same clusters using the deep learning technique and forecasting the next hour GHI by the corresponding experts (i.e., the FADF of the corresponding cluster) are worthy doing and beneficial for improving the GHI forecasting performance.

Table 1. The daytime one-hour ahead GHI forecasting performance of the proposed method

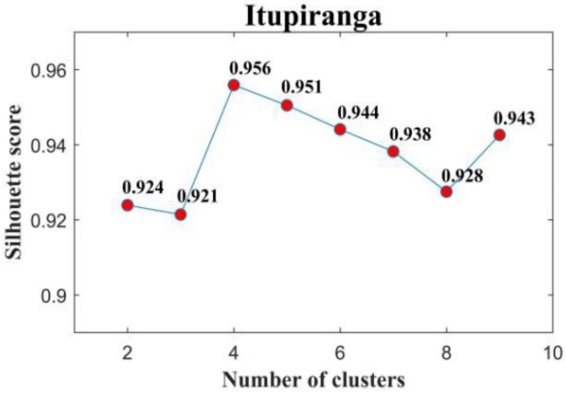
Locations	Models	Performance Metrics	
		RMSE(W/m^2)	rRMSE (%)
Ocala	GRU	117.35	27.29
	FADF	115.50	26.93
	HYBRID	**112.60**	**26.18**
Itupiranga	GRU	120.28	26.47
	FADF	119.41	26.28
	HYBRID	**117.71**	**25.91**

2) Evaluation of DTC

The proposed DTC can be categorized as the prototype-based clustering algorithm. The prototype-based clustering assumes that the clustering architecture can be characterized by a set of prototypes. This section further validates the superiority of the DTC compared with other prototype-based clustering methods. They are the K-means++ clustering algorithm, the Fuzzy C-Means (FCM) clustering algorithm, and the Gaussian Mixture Model (GMM) clustering algorithm. The single FADF trained on the whole training data is treated as the benchmark to the hybrid methods. These hybrid methods include the FADF combined with the K-means++ clustering algorithm (FADF+K-means++), the FADF combined with the (FCM clustering algorithm (FADF+FCM), the FADF combined with the GMM clustering algorithm (FADF+GMM), and the proposed HYBRID (the FADF combined with the DTC). In this work, the optimal number of clusters of each clustering algorithm is determined by the silhouette score which is one of the measurements for evaluating the performance of the clustering methods. For the adopted DTC method, the number of clusters for Ocala and Itupiranga are 3 and 4 respectively, as shown in Fig. 3.



(a)



(b)

Fig. 3. Silhouette score of DTC in different number of clusters for a) Ocala and b) Itupiranga.

As seen from Table 2, all hybrid methods achieve lower RMSE and rRMSE scores than the benchmark on both testing datasets (the best performance is marked in bold black). Overall, the proposed HYBRID achieves the best performance among all the hybrid methods and the benchmark. The advantage of the DTC is that it maps the original GHI time series into the Euclidean feature space through the deep learning and measures the distance between samples in the feature space, while the distance measurements between samples adopted by the traditional prototype-based clustering methods may easily ignore the characteristics of time series data.

Table 2. The daytime one-hour ahead GHI forecasting performance of the FADF combined with clustering methods

Locations	Models	Performance Metrics	
		RMSE (W/m^2)	rRMSE (%)
Ocala	FADF	115.50	26.93
	FADF+K-means++	114.15	26.54
	FADF+FCM	114.87	26.71
	FADF+GMM	113.56	26.41
	HYBRID	**112.60**	**26.18**
Itupiranga	FADF	119.41	26.28
	FADF+K-means++	118.04	25.98
	FADF+FCM	118.15	26.00
	FADF+GMM	117.92	25.95
	HYBRID	**117.71**	**25.91**

3) Computational Cost

Table 3. The computational cost of the proposed method

	Training (second/sample)	Testing (second/sample)
Ocala	0.66	1.77E-4
Itupiranga	0.59	2.42E-4

Table 3 shows the training and testing computational cost of the proposed method on the Ocala and Itupiranga datasets. The experiment was conducted with RTX 2080 Ti graphics processing unit. It is worth mentioning that the computational cost of the proposed method mainly comes from the DTC (while the DTC needs additional training round to determine the optimal clustering number) and the FADF. Both training of the DTC and FADF set the training maximum epochs as 1000 while the early-stop technique was also used in the experiment and the patience of the early-stop technique was set as 15. Table 3 shows that although the training of the proposed method takes time in both Ocala and Itupiranga datasets, the testing time of the proposed method is short enough which means the proposed method can be used in real-time. Note that the testing time of the proposed method on the Ocala dataset is shorter than that on the Itupiranga dataset mainly because the optimal clustering number for Ocala is 3 while Itupiranga is 4 and the forecasting process is conducted sequentially in terms of clustering id in this work.

4) Performance Analysis

Tables 4 and 5 show the 1-hour ahead daytime GHI forecasting results of different models in different locations with different performance metrics; the best performance is marked in bold black and the second-best performance is marked in bold blue. The values in brackets are the variances of results of the three runs, indicating the uncertainty of the model for the prediction results. Note that all models are implemented with the same dataset to provide a fair comparison, and the adjustments of hyper-parameters of comparative methods are referred to the corresponding papers (Azimi et al., 2016; Qing et al., 2018; Ghimire et al., 2019; Yan et al., 2020; Kumari et al., 2021).

In terms of the RMSE, R^2, and FS metrics, the XGBDNN achieves the best performance in both Ocala and Itupiranga dataset, while the proposed HYBRID achieves the second-best. Note that the XGBDNN is an ensemble model integrating multiple extreme gradient boosting trees and one deep neural network through the ridge regression. The proposed HYBRID is only a combination of the clustering method DTC and the forecasting model FADF. To further improve the forecasting performance, the ensemble of multiple FADF models in each cluster is ensured as a potential research direction inspired by the idea of the XGBDNN. In addition, the proposed HYBRID yields lower average MAEs than the XGBDNN in both Ocala and Itupiranga. It achieves the lowest average MAE (71.31 W/m^2) in the Itupiranga dataset. This is mainly because the huber loss is utilized to train the FADF of each clusters so that the trained FADF is not that sensitive to outliers.

The Smart Persistence model yields the worst performance in all kinds of metrics except for $Error_{min}$. It yields the zero $Error_{min}$ which is smaller than those from all other neural network-based models. None of the neural network-based methods can achieve the zero $ERROR_{min}$ because they can only approximate the target as closed as possible. TB_K-means+MLP achieves a better performance than the Smart Persistence model, benefiting from the clustering technique to de-trend the GHI time series into several clusters and the MLP for GHI forecasting being developed for each cluster. However, the cluster selection strategy and the limited forecasting ability of such shallow model (i.e., MLP) may affect the forecasting accuracy. The deep learning based models LSTM, C_LSTM, and ResInceptionGRUAttn show a comparable performance with TB_K-means+MLP.

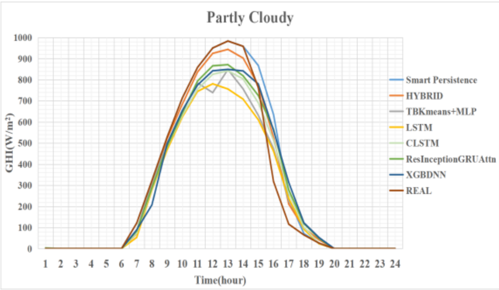
Table 4. One-hour ahead daytime GHI forecasting performance comparisons of different models in Ocala

Models	Performance Metrics						
	RMSE (W/m^2)	rRMSE (%)	MAE (W/m^2)	R^2 (%)	$ERROR_{max}$ (W/m^2)	$ERROR_{min}$ (W/m^2)	FS(%)
Smart Persistence	125.93 (±0.00)	29.97 (±0.00)	**63.84 (±0.00)**	79.73 (±0.00)	725.00 (±0.00)	**0.00 (±0.00)**	0.00 (±0.00)
TB_K-means+MLP	114.00 (±1.19)	26.51 (±0.28)	75.76 (±1.65)	83.40 (±0.36)	619.54 (±17.70)	0.01 (±0.01)	11.54 (±0.94)
LSTM	113.85 (±1.71)	26.48 (±0.40)	79.08 (±1.49)	83.43 (±0.47)	610.05 (±5.54)	0.02 (±0.02)	11.65 (±1.32)
C_LSTM	113.46 (±1.15)	26.38 (±0.27)	77.47 (±3.56)	83.53 (±0.35)	**602.89 (±35.50)**	0.02 (±0.01)	11.96 (±0.88)
ResInception GRUAttn	112.70 (±0.82)	26.21 (±0.19)	72.07 (±0.32)	83.73 (±0.23)	653.65 (±35.50)	**4.00E-3 (±4.00E-3)**	12.54 (±0.64)
XGBDNN	**111.97 (±0.78)**	**26.04 (±0.18)**	72.28 (±0.34)	**83.93 (±0.25)**	628.68 (±6.16)	7.33E-3 (±7.51E-3)	**13.12 (±0.60)**
HYBRID	**112.60 (±0.57)**	**26.18 (±0.13)**	**65.86 (±0.35)**	**83.80 (±0.17)**	**607.71 (±12.45)**	0.04 (±0.06)	**12.65 (±0.44)**

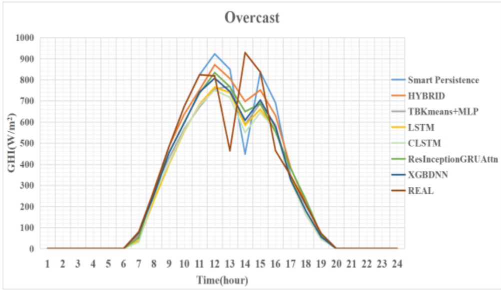
Table 5. One-hour ahead daytime GHI forecasting performance comparisons of different models in Itupiranga

Models	Performance Metrics						
	RMSE (W/m^2)	rRMSE (%)	MAE (W/m^2)	R^2 (%)	$ERROR_{max}$ (W/m^2)	$ERROR_{min}$ (W/m^2)	FS(%)
Smart Persistence	133.55 (±0.00)	29.39 (±0.00)	**71.37 (±0.00)**	78.18 (±0.00)	840.00 (±0.00)	**0.00 (±0.00)**	0.00 (±0.00)
TB_K-means+MLP	119.30 (±0.73)	26.25 (±0.16)	79.42 (±1.28)	82.60 (±0.20)	**673.76 (±33.71)**	0.02 (±0.02)	10.64 (±0.54)
LSTM	118.47 (±0.22)	26.07 (±0.05)	79.83 (±1.43)	82.77 (±0.05)	**680.22 (±9.32)**	0.01 (±0.01)	11.22 (±0.12)
C_LSTM	118.62 (±1.05)	26.11 (±0.23)	80.11 (±4.11)	82.77 (±0.32)	692.79 (±18.22)	**4.33E-3 (±1.15E-3)**	11.17 (±0.78)
ResInception GRUAttn	117.81 (±0.37)	25.93 (±0.81)	77.19 (±2.24)	83.00 (±0.10)	719.17 (±34.20)	0.02 (±0.02)	11.78 (±0.27)
XGBDNN	**117.14 (±0.562)**	**25.78 (±0.12)**	77.44 (±0.84)	**83.20 (±0.17)**	708.33 (±14.14)	0.03 (±0.02)	**12.27 (±0.42)**
HYBRID	**117.71 (±0.47)**	**25.91 (±0.10)**	**71.31 (±0.41)**	**83.03 (±0.12)**	703.38 (±7.33)	4.33E-3 (±6.65E-3)	**11.88 (±0.31)**

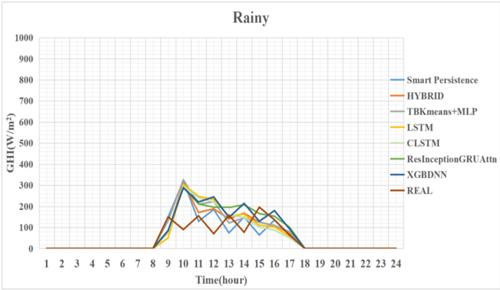
Fig. 4 shows the real GHI series (brown lines) and predicted GHI series from different models under different weather conditions, namely, partly cloudy, overcast, and rainy in Itupiranga. The subfigures demonstrate the effectiveness of the proposed method to forecast the GHI with different characteristics and variations. The HYBRID can forecast the real GHI with less error under different weather conditions and yields the smallest MAE values with 46.75, 81.88, and 71.06 W/m^2 for the partly cloudy, overcast, and rainy days, respectively. Therefore, the proposed HYBRID is more robust to different weather conditions than other compared methods.



(a) Partly cloudy (30th June 2017)



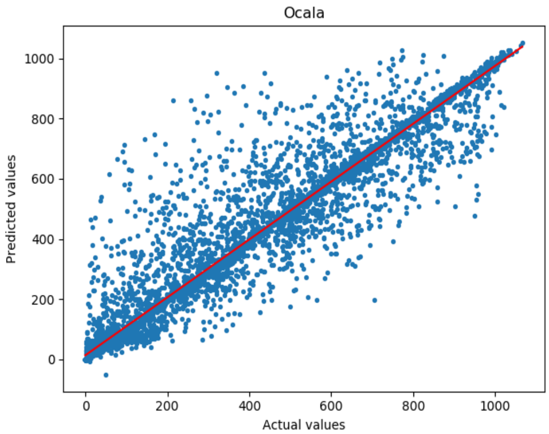
(b) Overcast (29th July 2017)



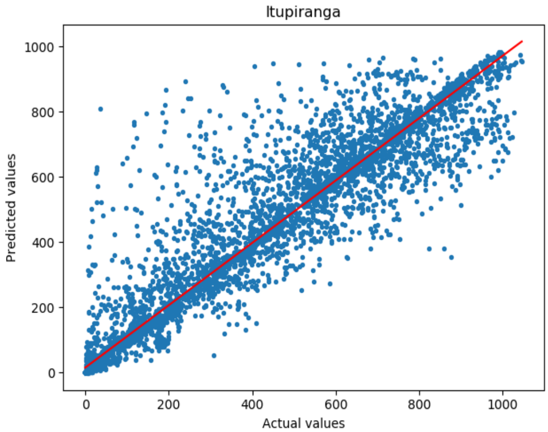
(c) Rainy (28th December 2017)

Fig. 4. 1-hour ahead GHI forecasting results of different models for the partly cloudy, overcast, and rainy days in Ocala.

Fig. 5 shows the correlation between real values and HYBRID forecasted values of GHI during 2017. The red lines in the plots indicate real values and blue dots indicate forecasted values by the HYBRID. The high density of the blue dots around the red lines shows the small forecasting errors of the HYBRID. However, there are still some big forecasting errors existing mainly caused by the wrong cluster assignment of the DTC, which needs further research to improve the prediction accuracy of the clustering method such that the final forecasting error can be reduced.



(a)



(b)

Fig. 5. Scatter plots of the actual and predicted GHI for the proposed model for the whole year of 2017 in a) Ocala and b) Itupiranga.

The performance of the proposed HYBRID method depends on the clustering method and the 1-hour ahead GHI forecasting performance of the FADF. However, the imbalanced weather type issue (Lai et al., 2019) where training samples for common weather events may influence the clustering performance of the DTC. Furthermore, the neural network-based FADF is sensitive to the perturbation of the training data, and thus it may jeopardize the robustness of the FADF (Yeung et al., 2007). The above-mentioned problems will be addressed in future works.

4. Conclusions

In this paper, we propose a deep learning based hybrid method for 1-hour ahead GHI forecasting. The proposed method adopts the Deep Time-series Clustering (DTC) to group the GHI time series data into multiple clusters to better identify its irregular patterns and thus to provide a better clustering performance. Then, the Feature Attention Deep Forecasting (FADF) model which is capable of dynamically weighting the input features and using the weighted features to forecast the next hour GHI is utilized for each clustered 1-hour ahead forecasting sub-task. Simulation results on the National Solar Radiation Database show that the proposed method achieves the smallest solar forecasting error compared to the smart persistence and other recently published methods.

Acknowledgements

This work is supported by National Natural Science Foundation of China under Grants 61876066 and 61572201,Guangzhou Science and Technology Plan Project 201804010245, Department of Finance and Education of Guangdong Province 2016 [202]: Key Discipline Construction Program, China; the Education Department of Guangdong Province: New and Integrated Energy System Theory and Technology Research Group [Project Number 2016KCXTD022]; Brunel University London BRIEF Funding, UK.

References

Azimi R., *et al*. (2016). A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting. *Energy Conversion and Management*, 118, 331-344.

Boukelia T., *et al*. (2017). Potential assessment of a parabolic trough solar thermal power plant considering hourly analysis: ANN-based approach. *Renewable Energy*, 105, 324-333.

Chung J., *et al.* (2014). 'Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv Preprint ArXiv:1412.3555*.

David M., *et al*. (2016). Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Solar Energy*, 133, 55-72.

Fermín R., *et al*. (2018). Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renewable Energy*, 126, 855-864.

Feng C., *et al*. (2018). Unsupervised clustering-based short-term solar forecasting.

IEEE Transactions on Sustainable Energy, 10, 2174-2185.

Faizollahzadeh A. S., *et al.* (2018). Computational intelligence approach for modeling hydrogen production: A review. *Engineering Applications of Computational Fluid Mechanics*, 12, 438-458.

Fotovatikhah F., *et al.* (2018). Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work. *Engineering Applications of Computational Fluid Mechanics*, 12, 411-437.

Fu L., *et al.* (2019). A regional photovoltaic output prediction method based on hierarchical clustering and the mRMR criterion. *Energies*, 12(20), 3817.

Feng C. and Zhang J. (2020). SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. *Solar Energy*, 204, 71-78.

Ghimire S., *et al*. (2019). Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy*, 253, 113541.

Guo X., *et al*. (2020). Adaptive self-paced deep clustering with data augmentation. *IEEE Transactions on Knowledge and Data Engineering*, 32, 1680-1693.

Huang C., *et al*. (2018). Data-driven short-term solar irradiance forecasting based on information of neighboring sites. *IEEE Transactions on Industrial Electronics*, 66, 9918-9927.

Jeon B.-K. and Kim E.-J. (2020). Next-day prediction of hourly solar irradiance using local weather forecasts and LSTM trained with non-local data. *Energies*, 13, 5258.

Kumari P., *et al* (2021). Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. *Journal of Cleaner Production*, 279, 123285.

Li S., *et al*. (2017). Typical solar radiation year construction using k-means clustering and discrete-time Markov chain. *Applied Energy*, 205, 720-731.

Lai C. S., *et al*. (2017a). A comprehensive review on large-scale photovoltaic system with applications of electrical energy storage. *Renewable and Sustainable Energy Reviews*, 78, 439-451.

Lai, C. S., *et al.* (2017b). Daily clearness index profiles cluster analysis for photovoltaic system. *IEEE Transactions on Industrial Informatics*, 13, 2322-2332.

Lai C. S., *et al*. (2019). A robust correlation analysis framework for imbalanced and dichotomous data with uncertainty. *Information Sciences*, 470, 58-77.

Ma J., *et al.* (2017). Novel field-support vector regression-based soft sensor for accurate estimation of solar irradiance. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 64, 3183-3191.

Mosavi A., *et al.* (2019). State of the art of machine learning models in energy systems, a systematic review. *Energies*, 12, 1301.

Niu, D., *et al*. (2020). Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: A case study. *Applied Soft Computing*, 93, 106389.

Pan K., *et al.* (2020). Photovoltaic output power estimation and baseline prediction approach for a residential distribution network with behind-the-meter systems. *Forecasting*, 2, 470–487.

Qing X., and Niu Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*, 148, 461-468.

Qadir, Z., *et al.* (2021). Predicting the energy output of hybrid PV – wind renewable energy system using feature selection technique for smart grids. *Energy Reports,* https://doi.org/10.1016/j.egyr.2021.01.018

Sun H., *et al*. (2016). Assessing the potential of random forest method for estimating solar radiation using air pollution index. *Energy Conversion and Management*, 119, 121-129.

Shakya A. *et al.* (2017). Solar irradiance forecasting in remote microgrids using Markov switching model. *IEEE Transactions on Sustainable Energy*, 8, 895-905.

Sun Y., and Sz. (2018). Solar PV output prediction from video streams using convolutional neural networks. *Energy & Environmental Science*, 11, 1811-1818.

Salcedo-Sanz S., *et al.* (2018a). An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the sunshine state of Australia. *Applied Energy*, 209, 79-94.

Salcedo-Sanz S., *et al.* (2018b). Feature selection in machine learning prediction systems for renewable energy applications. *Renewable and Sustainable Energy Reviews,* 90, 728-741.

Sun S., *et al.* (2018). A decomposition-clustering-ensemble learning approach for solar radiation forecasting. *Solar Energy*, 163, 189-199.

Song S., *et al.* (2018). Spatio-temporal attention-based LSTM networks for 3d action recognition and detection. *IEEE Transactions on Image Processing*, 27, 3459-3471.

Sengupta M., *et al*. (2018). The national solar radiation database (NSRDB). *Renewable and Sustainable Energy Reviews*, 89, 51-60.

Theocharides S., *et al*. (2020). Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing. *Applied Energy*, 268, 115023.

Wang Y. F., *et al*. (2018). Resilience-constrained hourly unit commitment in electricity grids. *IEEE Transactions on Power Systems*, 33, 5604-5614.

Wang Y. F., *et al.* (2019). Impact of cascading and common cause outages on resilience-constrained economic operation of power systems in extreme conditions. *IEEE Transactions on Smart Grid*, 11, 590-601.

Wu X., *et al.* (2020). Optimal kernel ELM and variational mode decomposition for probabilistic PV power prediction. *Energies*, 13(14), 3592.

Xie J., *et al.* (2016). Unsupervised deep embedding for clustering analysis. *Proceedings of the In International conference on machine learning*, 478-487.

Yeung D. S., *et al*. (2007). Localized generalization error model and its application to architecture selection for radial basis function neural network. *IEEE Transactions on Neural Networks*, 18, 1294-1305.

Yang D. (2019). A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES). *Journal of Renewable and Sustainable Energy*, 11, 022701.

Yan K., *et al* (2020). Short-term solar irradiance forecasting based on a hybrid deep learning methodology. *Information*, 11(1), 32.

Zhang X., *et al*. (2018). A solar time based analog ensemble method for regional solar power forecasting. *IEEE Transactions on Sustainable Energy*, 10, 268-279.

Zhen Z., *et al*. (2020). Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image. *IEEE Transactions on Industry Applications*, 56, 3385-3369.