

An attention fused sequence -to-sequence convolutional neural network for accurate solar irradiance forecasting and prediction using sky images

Chiagoziem C. Ukwuoma ^{a,b,c}, Dongsheng Cai ^{a,b,*}, Olusola Bamisile ^{a,b}, Hongbo Yin ^{a,b}, Grace Ugochi Nneji ^c, Happy N. Monday ^c, Ariyo Oluwasanmi ^d, Qi Huang ^{a,b}

^a College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Sichuan P.R, 610059, China

^b Sichuan Engineering Technology Research Center for Industrial Internet Intelligent Monitoring and Application, Chengdu University of Technology, Sichuan P.R, 610059, China

^c OBU, Sino-British Collaborative Education, Chengdu University of Technology, Sichuan P.R, 610059, China

^d School of Information and Software Engineering, University of Electronic Science and Technology of China, Sichuan P.R, 610059, China

ARTICLE INFO

Keywords:

Solar energy
Photovoltaic
Forecasting
Attention mechanism
Sky images
Hybrid model

ABSTRACT

The challenge of accurately forecasting ultra-short-term solar irradiance for photovoltaic systems is complicated by rapidly changing weather, and while ground-based sky images offer potential improvements, effectively extracting spatiotemporal data from these images remains a significant hurdle for current computer vision models. A new hybrid model, "An Attention Fused Sequence-to-Sequence Convolutional Neural Network," is being developed to address this challenge. The model predicts intra-hour GHI, DNI, and DHI with a 10-min lead time by combining a Convolutional Neural Network (for spatial feature extraction from sky images), an attention mechanism (to focus on relevant regions), and a sequence-to-sequence model (for temporal feature extraction from time-series data). The proposed Model is trained using the NREL Solar Radiation Research Laboratory Dataset while evaluating the model with the Mean Bias Error, Mean Absolute Error, Root Mean Squared Error, R Squared, and forecasting skill score. The 10-min and sequence length 2 interval is considered to be the best performing across most of the evaluation metrics with an MBE value is 2.321 W/m², MAE value of 39.490 W/m², RMSE value of 62.086 W/m², R² value is 0.909 W/m², FSS value of 23.589 W/m² and an MBE of 4.876 W/m², MAE of 56.887 W/m², RMSE of 85.346 W/m², R² of 0.834 and FSS of 24.881 W/m² respectively. Furthermore, the sensitivity analysis reveals that the proposed model's performance is influenced by both the sequence length and the lead time. The proposed framework outperforms other techniques for ultra-short-term PV generation forecasting, demonstrating its potential for practical deployment in PV systems to improve grid reliability and energy management.

Nomenclature

(continued)

Abbreviations	Symbols
PV	Photovoltaic
NWP	Numerical Weather Prediction
GHI	Global Horizontal Irradiance
DNI	Direct Normal Irradiance
DHI	Diffuse Horizontal Irradiance
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
CNN	Convolutional Neural Network
CDF	Cloud Distribution Features
ReLU	Rectified Linear Unit
MBE	Mean Bias Error
%	Percentage
i_t	Input Gate
f_t	Forget Gate
o_t	Output Gate
c_t	Cell State
g_t	Cell State Update
h_t	Hidden State
H	Forecast Horizon
R	Forecast Resolution
L	Forecast Lead Time
U	Forecast Update Rate

(continued on next column)

MAE	Mean Absolute Error	W/m ²	Watts Per Square Meter
RMSE	Root Mean Squared Error	R ²	Coefficient of Determination
ELM	Extreme Learning Machine		
RNN	Recurrent Neural Networks		
DNN	Deep Neural Networks		
AM	Attention Mechanism		
AE	Auto-Encoder		
ViT	Vision Transformer		
NREL	National Renewable Energy Laboratory's		

* Corresponding author. College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Sichuan P.R, 610059, China.
E-mail address: caidongsheng@cdut.edu.cn (D. Cai).

1. Introduction

1.1. Background

The fast development of solar energy, wind power, and other clean energy sources has grown into a crucial tactic in addressing the changing conditions as the world's energy and climate concerns grow. Solar energy is perhaps the most viable renewable resource among them because of its availability, cost, mobility, cleanliness, and positive effects on the natural world [1,2]. However, the main issues with integrating solar electricity are that it is sporadic and unpredictable [3], and production changes seriously impede power networks' ability to operate steadily and economically [4]. Incorporating solar photovoltaic (PV) plants into power networks has necessitated reliable PV generation forecasts over a short time horizon for network planning and staffing purposes. Furthermore, acquiring global horizontal irradiance (GHI) with excellent spatial and time precision can greatly improve the reliability of PV power forecasts [5,6]. Therefore, the modelling of PV electricity is essential for forecasting short-term PV variations. PV generation is heavily reliant on the sun's rays, and while it can be somewhat accurate on bright days, it may not attain optimum accuracy on gloomy, wet, or cloudy days due to variations and inconsistencies.

Over the last two generations, numerous PV generation predictions have been developed [7,8], but predicting remains difficult due to variables such as sunlight, wind, and cloud covering [9,10]. PV generation forecasting methods are broadly classified into two types: utilising past PV output data and image-based analysis and forecasting. Numerical Weather Prediction (NWP) can help predict PV generation across intermediate and long horizons [11], but it does not work for ultra-short-term forecasting due to local atmospheric variables at the PV facility [12,13]. The PV generation time series displays seasonal and cyclical oscillations [14]. PV generation estimates have been constructed using a variety of statistical and machine learning approaches, including outlier identification [15], Extreme Learning Machine (ELM) [16], and day-ahead solar irradiance forecasting [17]. Nevertheless, these solutions need high-precision real-time meteorological data and computing complexity. Deep learning (DL) approaches [18,19], including multi-layer feed-forward artificial neural networks, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Convolutional Neural Network (CNN)-LSTM, have been used to forecast PV generation. Nonetheless, these approaches may fail to correctly capture variations that are unpredictable in the ultra-short term. Numeric data-driven approaches necessitate more data and computationally complex processes. Sky-imagers, which employ ground-based sensors to gather regional sky images and mirror actual time atmospheric conditions above the PV plant, can enhance forecast accuracy on an ultra-short time scale. Cloud images have been successfully used to improve PV generation forecast accuracy. Satellite and sky images are employed in image-based analysis and forecasting because they both contain local weather data as well as underlying spatiotemporal correlations that may be used to predict future local environmental conditions. As a result, a powerful deep learning model is required to grasp and expose the information encoded in sky images, as well as to combine these details with historical PV generation data to provide reliable ultra-short-term PV generation predictions [20].

1.2. Related works

Empirical approaches and DL approaches are the two main categories into which image-derived solar irradiance forecasts may be broadly divided [21]. Key informative carriers, such as cloud cover indices (CCI), cloud cluster matching (CCM), and cloud movement vectors (CMV), are extracted using empirical approaches from ground-based sky images [22]. A 20-min forecasting algorithm for

clear-sky score prediction was developed by Victor et al. [23]. It performs similarly to conventional models and has a forecast skill of 5–20 % for the majority of sky situations, but it has trouble with variations in irradiance and cloudy days. To lessen hardware functional management, Guillermo and Manel [24] devised a data processing technique for removing cloud dynamic elements from sky images and observations of solar irradiance. They then integrated this technique into solar forecasting methods. Zhang et al. [25] precisely measure and monitor variations in solar irradiance using photographs of the sky taken from the ground. It creates a superior joint point-interval real-time estimating model that yields prediction intervals and predictable forecasting outcomes. Regardless of the weather, the model generates tight intervals while retaining excellent quality and durability. However, because of their poor versatility, these approaches are not sturdy enough to withstand abrupt changes in the weather [26]. On the other hand, competitive outcomes can be obtained with DL techniques, which do not require intricate pre-processing, due to their robust feature extraction and quadratic extrapolation ability [27]. This adaptability presents several chances to improve predicting accuracy. Federated learning, for example, allows the grouping of a global model without exchanging raw data, preserving forecasting accuracy while safeguarding data privacy [28]. Significantly enhancing model efficiency can be achieved by transfer learning, which is the transfer of conceptual information from a particular field to a different one, especially when training data is scarce [29]. Models can integrate new data while remembering what they have already learnt thanks to incremental learning [30]. Big data learning also makes use of huge datasets for training purposes to improve the generalisation of DL models [31]. Even with these developments, there is still work to be done in the field of efficiently combining multi-source data (such as sky photos and historical time series).

CNNs are a popular architecture for feature extraction from image information and are the foundation of image-based solar forecasting because of their common weights and local perceptive mechanisms [32]. Cong et al. [33,34] developed a SolarNet deep CNN model to forecast GHI from 10 to 60 min in advance without using numerical data or feature engineering. Quentin et al. analyzed four popular DL architectures for solar irradiance prediction based on a set of spherical sky images and external factors [35]. Hao et al. [36] identified a 3D-CNN that can extract characteristics from sky images using successful training procedures. Fei et al. [37] put forward a real-time approach for estimating solar irradiance up to the minute by extracting RGB values and pixel positions from sky images. Zhao et al. [38] developed a hybrid mapping technique based on deep learning for predicting ultra-short-term solar PV output. In his methods, the k-means clustering analysis using feature extraction from the convolutional autoencoder was carried out first followed by capturing the connections between the surface irradiance and the sky image by the hybrid model. Zhao Zhen [39] proposes a novel convolutional auto-encoder for extracting cloud distribution features (CDFs). In this work, a feature fusion by time step LSTM-FUSION irradiance prediction model is developed, accounting for the one-to-one relationship between CDFs and GHI. An approach for minutely irradiance prediction utilising multivariate extraction of features of all-sky images is proposed by Wu et al. [40]. The optical flow method is used to generate the cloud displacement vector after pre-processing the initial images into four weather kinds. For irradiance projection models, convolutional neural networks extract multivariate data and blend them with historical irradiance and weather conditions. A technique for estimating sun irradiance using multivariate extracted features from all-sky images is presented by Wu et al. [41]. Utilising an enhanced clustering-boundary rectification technique, the technique distinguishes between cloud and sky pixels, divides all-sky images into four categories based on cloud-sky types, and uses an image RGB grid and convolutional neural network to extract global as well as local characteristics. It then builds irradiance projection models for four cloud-sky types by combining these features with historical irradiance and meteorological variables. In comparison to benchmarks, the

strategy reduces mean absolute percentage error by 0.21 %, 20.21 %, 2.53 %, and 5.30 %.

Yet, the lack of spatial or temporal characteristic extraction limits the approaches' capacity to attain a high degree of accuracy. Two CNN-based frameworks were built for processing sky imagery [34], with the first model performing 2D convolution on layered sky images and the 3D CNN doing direct 3D convolution on the sky image sequence. The results indicated that convolutional operations failed to perform well across channels, and the 2D convolutional design outperformed the 3D design. In Ref. [42], both spatial and temporal characteristics of sky images were evaluated, with PhyDNet and ConvLSTM running in parallel. However, the image-generating approach involves over-detailed features that are concealed in the inner layer, which may have an impact on the accuracy of the model. This study's work is more in line with the model proposed by Jonathan et al. [43]. They proposed an Attention-embedded Convolutional Neural Network to anticipate GHI, DNI, and DHI for intra-hour solar forecasting using sequences of sky images. Six lead times and four sequence lengths were employed, and the optimal combination was found to be four sequences and 20 min. With low RMSE (62.75 W/m^2), low MBE (2.71 W/m^2), and a high FSS 26 (38.81), this combination yields an appropriate balance and optimal result that demonstrates good accuracy, little bias, and a high skill score. A hybrid multipurpose ensemble learning model is proposed by Wang et al. [44] to anticipate solar irradiance in the short-term using sky images and past data. The model extracts temporal features and provides cloud cover statistics by using XGBoost to gather logical connections between input features and subsequent observations. A novel technique for stochastic end-to-end forecasting with an ImageNet pre-trained deep neural network is presented by Chaaraoui et al. [45]. The two-phase method entails modifying and updating an elementary model for forecasting after it has been trained to determine irradiance from all-sky imager images. When used for predicting, the model accounts for diffuse horizontal, global horizontal, and direct normal irradiance and gets a general positive skill score of 18.6 % in contrast to an intelligent consistency forecast. To achieve real-time precision in solar energy forecasting, Nijhum et al. [46] describe a unique technique for estimating solar irradiance using infrared sky pictures and a CNN-regression model. Improved solar irradiance estimation is achieved by employing a lightweight, pre-trained MobileNetV2 model, which reduces the RMSE to 16.18 W/m^2 . Sebena et al. [47] investigate if sky images and meteorological information may be used to forecast how much electricity solar panels would produce. To reduce unpredictability and maximise control, it makes use of the SkyCam dataset and suggests three scenarios that combine block-based and sequential processing of sky photos and meteorological data using CNN and LSTM networks.

A transformer-based multimodal-learning system was created by Lie et al. [48] to predict global horizontal irradiance in the ultra-short period. The approach uses a generative decoder for multi-step prediction, incorporates historical and experimentally calculated GHI, and converts ground-based sky images into optical circulation maps. According to the empirical findings, the proposed strategy accomplishes 10-min-ahead forecasting with a normalised root mean square error (NRMSE) of 4.28 %. By employing a transformer-based structure, Zhang et al. [32] enhanced the accuracy of ramp occurrences, surpassing the gate architecture-based model and the conventional CNN model by 9.43 % and 3.91 % on 2-min and 6-min scales, accordingly. According to Thomas et al. [49], irradiance can be precisely estimated using sky images and a machine learning model based on vision transformers without the need for further data. The model incorporates sky images from a conventional optical and fish-eye camera, as well as 17 years of GHI, DNI, and DHI from a high-accuracy pyranometer and sun-tracked device. For both DHI and GHI ($\text{RMSE} = 31 \text{ W/m}^2$), the proposed model produced extremely accurate results. A unique cloud image-based ultra-short-term prediction system is presented by Xu et al. [50]. Multi-layer perception is used for one-step PV generation forecasting, and the Vision Transformer model and Gated Recurrent Unit encoder are

integrated for highly dimensional hidden analysis of features.

1.3. Knowledge gaps and scientific contributions

Even with the advancements in current research, several approaches still need to be looked at more. Recurrent neural networks (RNNs) are commonly used in conventional techniques for representing historical time-series data. RNNs are susceptible to feature deterioration via forward propagation, though, because of their sequential design [51]. Furthermore, not every time step in the input data sequence is significant when modelling lengthy sequences, which might cause uncertainty and inefficiencies if the entire time series is employed [52]. Therefore, it is imperative to apply techniques that can pinpoint and highlight the critical time stages that most accurately depict the sequential properties. CNN-based techniques for extracting spatiotemporal information from sky images have been the subject of more and more recent research. CNNs a type of deep neural network with convolutional layers process local data quickly and effectively via sliding windows, offering robust visual capabilities while preserving a low level of model structure. Nevertheless, CNNs are less successful in gathering global information [53], which makes it difficult to identify broad trends in sky images, especially when there is a lot of cloud motion. Consequently, a viable substitute is the creation of attention-based learning techniques for improving global feature acquisition. Furthermore, there is a dearth of studies on the cooperative learning of feature relationships over many methods. Differences in solar irradiance are tightly connected with changing factors (cloud cover, solar status, and cloud opacity) that are extracted from sky photographs. Still, not much research has been done on how historical data and sky image sequences interact. An important difficulty is that most contemporary image- and numerical-based forecasting models are responsive instead of proactive, as noted by Paletta et al. [35]. In particular, based only on sky images, these models frequently fall short of predicting when solar ramp events will occur. Therefore, to improve forecasting accuracy, more study is required to fully examine the connection between various feature modalities.

This study tackles these difficulties in a unified framework by investigating and constructing a temporal-specialized deep learning mechanism and an exceptionally well spatial feature extraction technique. The proposed model integrates the powerful properties of a CNN, a Sequence-to-Sequence model, and an attention mechanism. The CNN is for spatial feature extraction, an attention mechanism to focus on the relevant regions, and a sequence-to-sequence model to capture temporal dependencies for accurate solar irradiance forecasting and prediction using sky images. The proposed method is designed to predict the future values of global horizontal irradiance (GHI), direct normal irradiance (DNI), and diffuse horizontal irradiance (DHI) for utility operations. To accurately define the temporal structure of the predictions, forecast horizon (H), forecast resolution (R), forecast lead time (L), and forecast update rate (U) features were analyzed. The NREL Solar Radiation Research Laboratory Dataset including Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), and forecasting skill score (FSS) evaluation metrics were used in this paper for analysis. To confirm the validity of the proposed model, the findings obtained are further compared with SOTA research. The following are this manuscript's originality and main contribution.

- ❖ A novel Attention Fused Sequence-to-Sequence Convolutional Neural Network which aims at enhancing the precision of solar irradiance forecasts through the analysis of sky images is proposed.
- ❖ A spatiotemporal analysis forecasting framework is developed for capturing the dynamic changes over time in sequential sky images, enabling a more nuanced understanding of temporal patterns.
- ❖ A fusion of CNN and Attention mechanisms is designed to process spatial features effectively enabling the model to comprehensively

interpret the overarching patterns within the static spatial attributes of sky imagery.
 ♦ A hierarchical three-tier model structure is established to process both sky images and additional exogenous data in a cohesive end-to-end manner enhancing the stability and accuracy of clear-sky irradiance predictions, minimizing the risk of anomalies and bolstering the dependability of the forecasted clear-sky irradiance values.

The subsequent sections of this paper are structured in the following manner. Section 2 discusses in detail the materials and methodologies employed in this manuscript including the proposed model, dataset, and data preprocessing and evaluation metrics. Section 3, which is the results and analysis, the implementation hyperparameters, proposed model forecast results, components analysis of the proposed model and result comparison with the state-of-the-art models on the same dataset, result discussion and limitations, and future works were all discussed. The conclusion is done in section 4.

2. Materials and Methodology

This section talks about the materials and the proposed method used in this manuscript including the dataset, evaluation metrics, and implementation setup. The primary objective of this study is to accurately forecast the GHI, DNI and DHI within an intra-hour timescale using a hybrid approach that combines image data from sky cameras and numerical weather data. The forecasting model leverages sequence images of the sky to capture cloud dynamics and uses numerical weather data as an additional input to improve the accuracy of solar irradiance predictions. Fig. 1 illustrates the overall workflow in this study. The dataset used consists of image data and numerical data. The image data are sky images captured at consistent intervals (e.g., every minute), used

to extract cloud dynamics and patterns whereas the numerical data incorporates features like temperature, humidity, wind speed, and solar angles, essential for accurate irradiance predictions. These data types are synchronized to ensure that for every image, corresponding numerical features are available. Data processing techniques such as resizing (all images are resized to a uniform dimension), normalization (pixel values are scaled between 0 and 1) and augmentation (techniques such as random rotations, flips, and zooms are applied to improve model robustness) were deployed for the image dataset while for the numerical data pre-processing, numerical features are standardized, ensuring all inputs have a similar scale and contribute equally during training. Furthermore, Time-based feature engineering is applied to extract temporal information like time of day and day of year. Image and numerical data are synchronized by timestamp and labelled with corresponding GHI, DNI, and DHI values for the proposed model training. The proposed model integrates convolutional layers, attention mechanisms, and an LSTM to capture spatiotemporal dependencies in the data.

2.1. Proposed model

The proposed Attention-Fused Sequence-to-Sequence Convolutional Neural Network is a hybrid model that processes image data through a series of convolutional layers, applies attention mechanisms, and integrates a sequence-to-sequence network for temporal pattern extraction. The proposed model then fuses the image-based features with numerical data for final predictions as seen in Fig. 2. The proposed model comprises several major blocks including the input layer, CNN block, attention block, sequence-to-sequence block, numerical data integration and output layer. The first block which is the input layer accepts an image input tensor of shape (H, W, C). Several convolutional layers with ReLU activation are seen in the CNN block to capture spatial

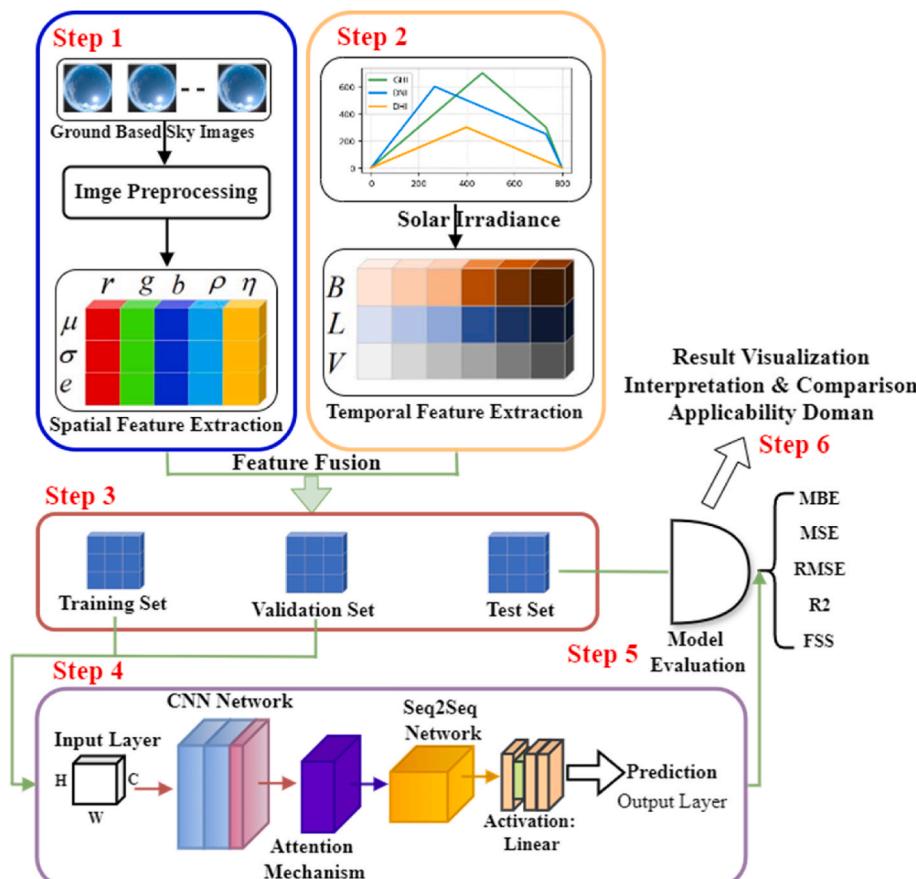


Fig. 1. Technical Diagram of the Work done in this study.

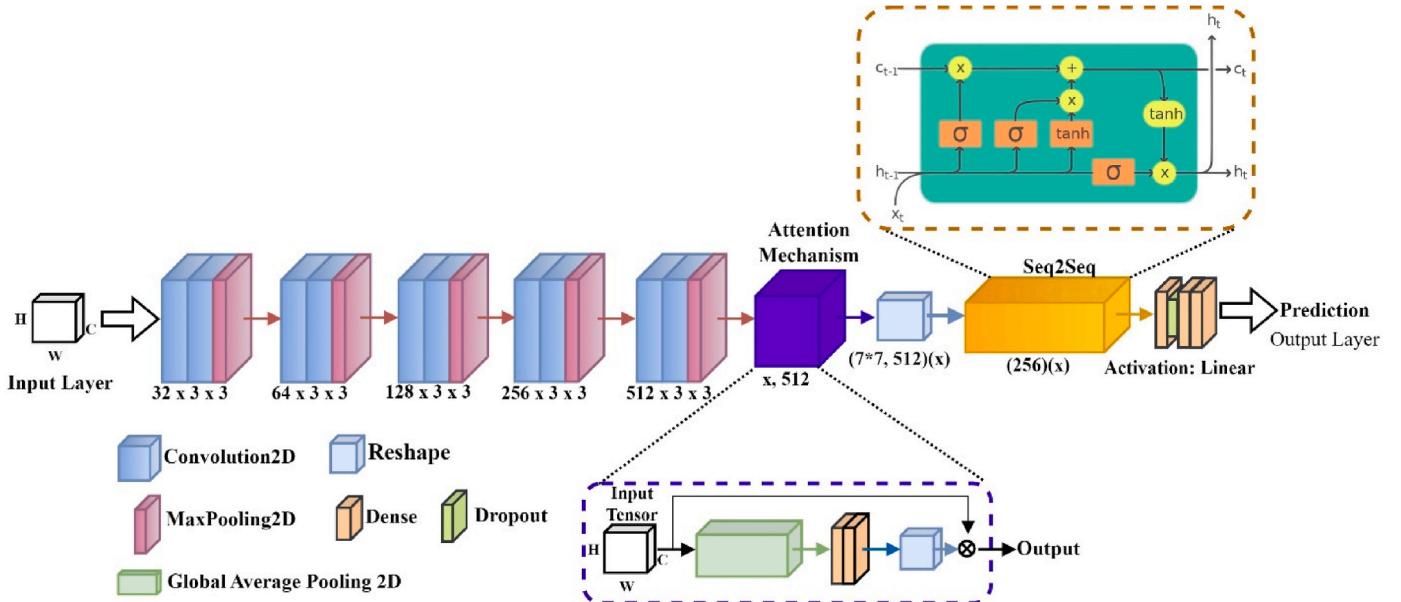


Fig. 2. Proposed Model architecture.

features from the input block. Feature extraction starts with lower-level patterns (e.g., edges) and progresses to more complex representations (e.g., cloud formations). Layers are followed by MaxPooling2D operations to reduce dimensionality while preserving relevant features. Specifically, the CNN block comprises convolutional layers with filters ranging from 32, 64, 128, 256 and 512, kernel size of 3, and ReLU activation. Max pooling follows each convolutional block with the Batch normalization layer at the 3rd and 5th convolution while a drop-out layer is added at the 4th and after the dense and Sequence to sequence layer. Max-pooling helps in increasing the model's receptive field and computational efficiency by reducing the number of parameters. The CNN blocks are responsible for extracting hierarchical features from the input sky images specifically spatial patterns and structures such as cloud formations, sky colour, and other relevant features. These learned features help the model understand the complex relationships between sky conditions and solar irradiance levels.

Mathematically, let's denote the output of the convolutional layer at each stage as x_i where i denotes the stage number. This study denotes the convolution operation $\text{Conv2D}(X_i)$ with a kernel k_i can be represented as:

$$x_i = \text{ReLU}(x_i * k_i + b_i) \quad (1)$$

where $*$ is the convolution operation and b_i is the bias term. Each convolutional layer is followed by a Maxpooling operation which reduces the spatial dimensions of the feature maps which is denoted as x'_i .

$$x'_i = \text{MaxPool}(x_i) \quad (2)$$

The third block is the attention block which is incorporated to weigh important regions of the feature maps generated by the CNN blocks. The attention mechanism uses global average pooling and a dense layer to create a feature vector that reweights the original convolutional output, enhancing the model's focus on critical regions. By assigning higher weights to these regions, the attention mechanism enables the model to focus on the most informative parts of the sky images, improving the accuracy of irradiance predictions. Let x_j be the output of the convolutional layer just before the attention block. The attention block is represented as:

$$A = \text{GlobalAvgPool}(x_j) \quad (3)$$

$$W_1 = \text{ReLU}(W_1 \bullet A + b_1) \quad (4)$$

$$W_2 = \text{ReLU}(W_2 \bullet W_1 + b_2) \quad (5)$$

$$M = \text{Reshape}(W_2) \quad (6)$$

$$x'_i = \text{Multiply}(x_j, M) \quad (7)$$

The output from the attention block is reshaped into a sequence and passed to a sequence-to-sequence model block. The sequence-to-sequence model captures temporal dependencies in cloud movement across image sequences. By leveraging the sequential nature of sky image data, the sequence-to-sequence layer enhances the model's ability to make accurate predictions by considering how irradiance levels evolve. Assuming the output of the attention map is denoted as x_{att} , the sequence-to-sequence layer is calculated thus;

$$x_{\text{seq2seq}} = \text{Seq2Seq}(x_{att}) \quad (8)$$

The extracted image features are combined with processed numerical data (e.g., temperature, and wind speed) in the fully connected layers. Finally, the output of the proposed model is the prediction layer which is denoted as;

$$y = \text{Linear}(x'_{\text{Seq2Seq}}) \quad (9)$$

The final output layer consists of three neurons with a linear activation function, predicting GHI, DNI, and DHI. In general, the proposed model is computed as

$$y = \text{OutputLayer}(\text{Seq2Seq}(\text{Att}_{\text{Block}}(\text{Conv}_{\text{Layers}}(x)))) \quad (10)$$

2.1.1. Spatial features extraction block (Block Two)

Assuming $x^{(0)}$ represents the input tensor, $x^{(i)}$ is the output tensor after the i - th layer, the proposed feature extraction block (Fig. 3) accepts input features as thus;

$$x^{(0)} = \text{input}(\text{shape} = \text{input}_{\text{shape}}) \quad (11)$$

where the input shape is denoted $224 \times 224 \times 3 * \text{nos}$ of image. The layers are represented as $x^{(1,2,3,\dots,n)}$ where n stands for the number of layers. Layer $x^{(1,2,4,5,8,\text{and } 10)}$ share the same configurations including ReLU, Con2D, Input features, biases, and padding as seen in Equation (12).

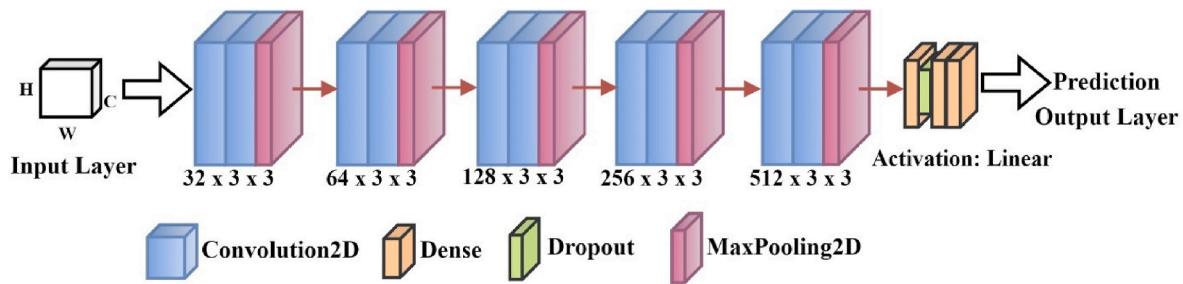


Fig. 3. Proposed spatial feature extraction Model (block two).

$$x^{(n)} = \text{ReLU} \left(\text{Conv2D} \left(x^{(0)}, W_{\text{conv}}^{(i)}, b_{\text{conv}}^{(i)}, \text{padding}=\text{'same'} \right) \right) \quad (12)$$

Where $x^{(n)}$ is the number of convolutions, $W_{\text{conv}}^{(i)}$ are the weights and $b_{\text{conv}}^{(i)}$ is the biases of the i -th convolutional layer, $\text{ReLU}(\bullet)$ signifies the Rectified Linear Unit activation function. Layer $x^{(7 \text{ and } 11)}$ share the same configurations with other convolutional layers with the inclusion of kernel regularizer = L2(0.05). Layer $x^{(3,6,9 \text{ and } 12)}$ Share the same configurations including MaxPooling2D, input features, and pool size as seen in Equation (13).

$$x^{(n)} = \text{MaxPooling2D}(x^{(0)}, \text{pool_size} = (2, 2)) \quad (13)$$

where MaxPooling2D(\cdot) represents the max-pooling operation with a pool size of (2, 2).

2.1.2. Relevant regions focus block

Fig. 4 shows the proposed attention mechanism. The first operation is applying a global average pooling 2D to the input tensor as seen in Equation (14).

$$x_{\text{avg}} = \text{GlobalAveragePooling2D}(\text{input_tensor}) \quad (14)$$

This operation calculates the average value along each channel of the input tensor, resulting in a global representation of the input features. x_{avg} is passed into the next layer as seen below.

$$x_{\text{dense}_1} = \text{ReLU}(\text{Dense}(x_{\text{avg}}, W_{\text{dense}_1}, b_{\text{dense}_1})) \quad (15)$$

$$x_{\text{dense}_2} = \text{ReLU}(\text{Dense}(x_{\text{dense}_1}, W_{\text{dense}_2}, b_{\text{dense}_2})) \quad (16)$$

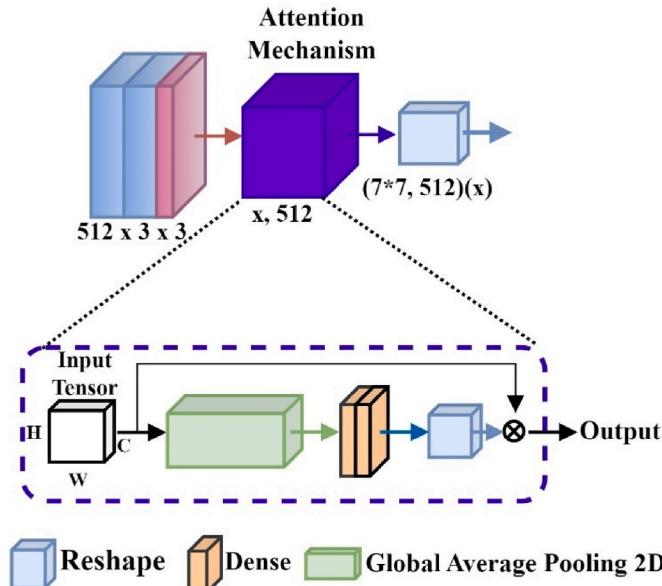


Fig. 4. Proposed attention mechanism.

Where W_{dense_n} and b_{dense_n} represents the weights and biases and ReLU activation functions are applied after each dense layer. The reshaping operation is carried out on the x_{dense_2} thus:

$$x_{\text{reshape}} = \text{Reshape}(x_{\text{dense}_2}) \quad (17)$$

The reshape operation reshapes the output of the second dense layer into a $1 \times 1 \times \text{in_channels}$ tensor for an element-wise multiplication with the original input tensor.

$$x_{\text{attention}} = \text{Multiply}(\text{input_tensor}, x_{\text{reshape}}) \quad (18)$$

This operation performs element-wise multiplication between the original input tensor and the reshaped tensor from the dense layers. Each element of the input tensor is multiplied by the corresponding element of the reshaped tensor, applying attention weights obtained from the dense layers. The output of $x_{\text{attention}}$ represents the input tensor modulated by attention weights. This mechanism allows the model to focus on important regions of the input features while suppressing less relevant areas.

2.1.3. Sequence to sequence model

The Sequence-to-sequence model (Fig. 5) contains several gates to control the flow of information including the input gate (i_t), forget gate (f_t), output gate (o_t) and cell state (c_t). The input gate (i_t), forget gate (f_t), cell state update (g_t), new cell state (c_t), output gate o_t and hidden state (h_t) which are represented mathematically as:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (19)$$

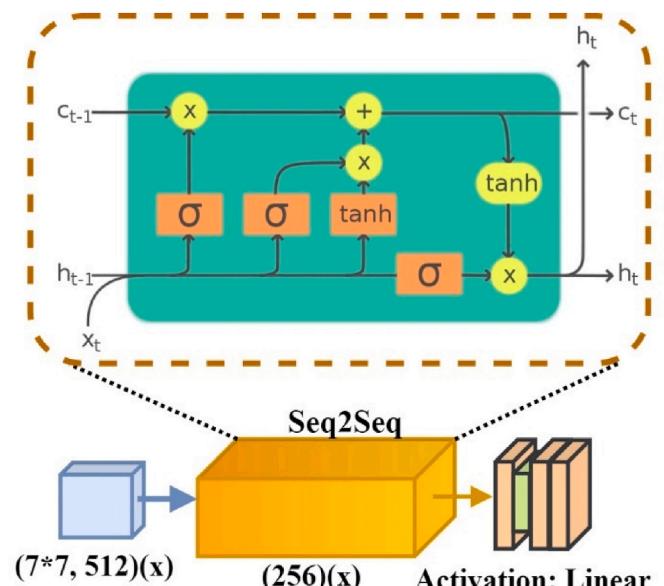


Fig. 5. Proposed sequence-to-sequence Model.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (20)$$

$$g_t = \tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \quad (21)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (22)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (23)$$

$$h_t = o_t \odot \tanh(c_t) \quad (24)$$

where x is the input at time step t , h_{t-1} is the previous time step hidden state, c_{t-1} is the previous time step cell state, $((W_{xi}, W_{xf}, W_{xg}), (W_{hi}, W_{hf}, W_{hg})$ and (W_{ci}, W_{cf}, W_{cg}) represents the input, hidden state, and cell state weight matrices respectively. (b_i, b_f, b_o) represent bias and the different activation functions include σ (sigmoid) and \tanh (hyperbolic tangent). \odot signifies elementwise multiplication. In Equation (22), f_t regulates the amount of information to retain from the previous state whereas the i_t focuses on the amount of new information to select. In Equation (23), o_t calculates what percentage of c_t to output after it must have undergone \tanh function. The step-by-step operation of the proposed sequence-to-sequence model is depicted from Equation (19)–(24). Through these steps, the seq2seq carefully selects or disregards information over long sequences thereby making it suitable for accurate solar irradiance forecasting and prediction using sky images.

2.2. Dataset and data preprocessing

The NREL Solar Radiation Research Laboratory Dataset was used in this paper for analysis [54,55]. NREL stands for the National Renewable Energy Laboratory. The dataset which incorporates both ASI-16 and numerical measurement is said to be the largest dataset in this domain that is made publicly available for research around the globe. The image data is of 1536 x 1536 pixels per image. The dataset covers the period from October 1, 2017, to December 31, 2022, which is five years and three months. The training set, validation set, and testing set are the three subsets into which it is divided. Data from the first three years (October 1, 2017, to December 31, 2020) are included in the training set. The data from January 1, 2021, to December 31, 2021, is then set aside for validation, and the testing set is the data from January 1, 2022,

to December 31, 2022, the remaining year. The numerical data consists of atmospheric pressure, DNI, DHI, GHI, relative humidity, temperature, and wind speed. The dataset contains information and a basic Python code for preprocessing thus, this paper made use of the Python code to process the data for uniformity. For the numerical features, the DNI, DHI, and GHI were used in the proposed model as labels, while others were used in benchmarks. Owing to the proposed model's ability to train in an end-to-end manner, the original image pixels of 1536 x 1536 were resized to 224 x 224 pixels. Furthermore, six various lead durations ranging from 10 min ahead to 60 min ahead were tested, along with four different image sequence lengths: $2^0, 2^1, 2^2$, and 2^3 images. Fig. 6 shows how the dataset is measured and formed from the satellites while the sample of the original sky image and its corresponding processed image is depicted in Fig. 7.

2.3. Evaluation metrics

This paper made use of 5 evaluation metrics namely the Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), and forecasting skill score (FSS). The average of the variations between each data point's true and forecasted values is used to compute MBE which is mathematically calculated as;

$$MBE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i) \quad (25)$$

where n = number of samples, O_i and P_i are the observed values and predicted values of the i th item respectively. Unlike the MBE, by dividing the total number of observations by the sum of all errors, the MAE determines the exact difference between the actual and anticipated values mathematically represented as;

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - O_i| \quad (26)$$

where $|\bullet|$ = absolute value. The RMSE corresponds to the square root of the average squared error, and its measurement unit aligns with that of the dependent variable as represented in Equation (27).

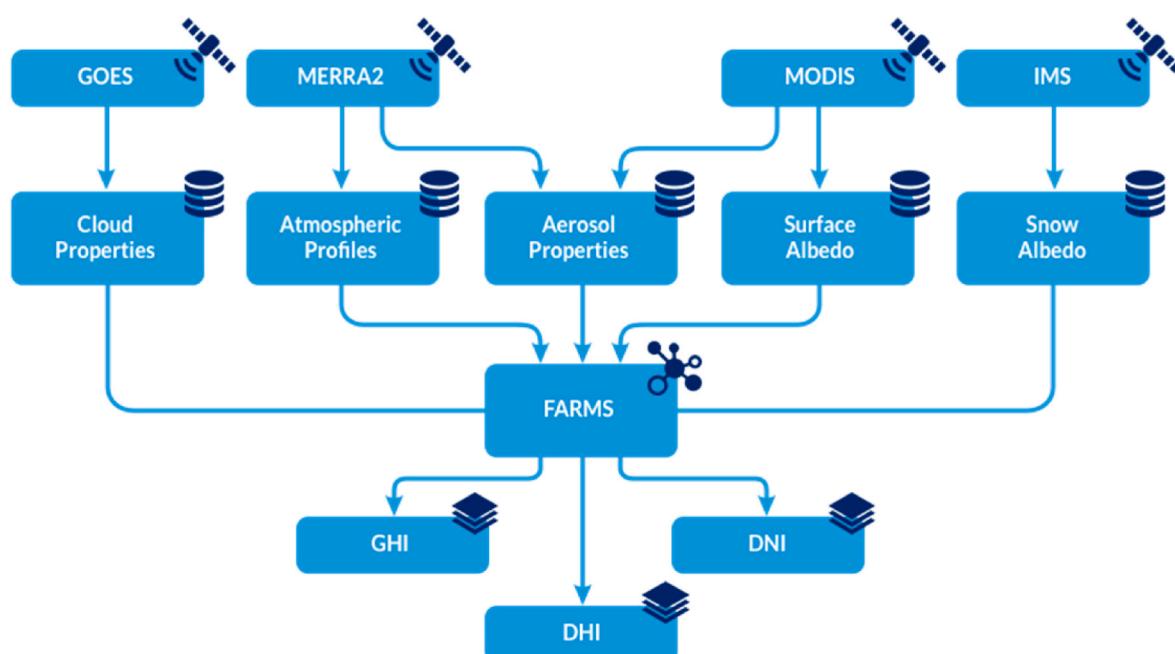
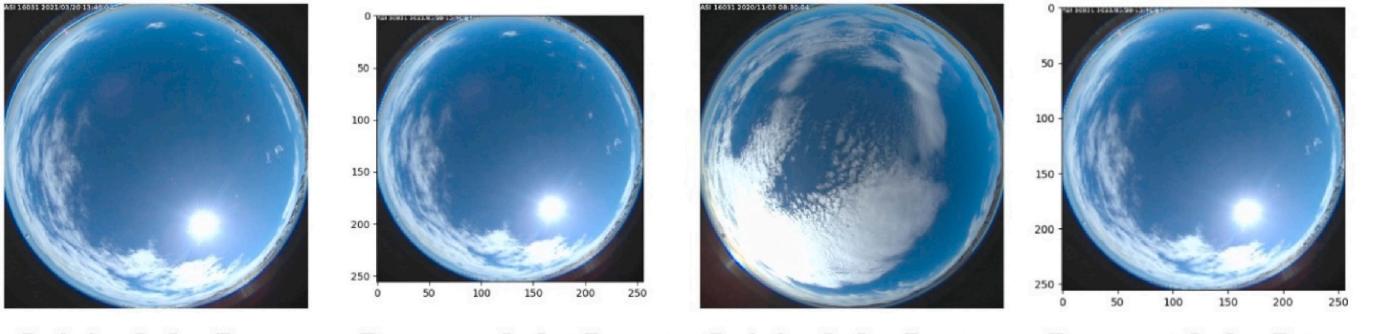


Fig. 6. Data flow from Satellite to solar radiation measurement.



Original sky Image Processed sky Image Original sky Image Processed sky Image

Fig. 7. Sample of the sky Image (unprocessed vs. Pre-processed).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (27)$$

R^2 measures the percentage of variability in the dependent variable that can be explained by the independent variables in the model mathematically represented as;

$$R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{the total sum of squares (SST)}} = 1 - \frac{\sum (P_i - O_i)^2}{\sum (P_i - \bar{O})^2} \quad (28)$$

Note: $R^2 = 0$ means that the model does not explain any of the variability in the dependent variable whereas an $R^2 = 1$ indicates that the model perfectly explains all of the variability in the dependent variable.

$$\text{FSS} = \left(1 - \frac{\text{RMSE}_{\text{model}}}{\text{RMSE}_{\text{reference}}} \right) \quad (29)$$

Meteorology and atmospheric science employ the FSS as a criterion to evaluate forecasting model performance. The spatial consistency of predicted and observed patterns is measured. Elevated FSS values signify precise forecasts, whilst reduced values imply inadequate alignment. FSS is useful for evaluating the spatial accuracy of weather, climate, and environmental forecasts and may be adjusted to various situations. It offers information about how well a model can represent spatial patterns, which is essential for producing precise forecasts.

2.4. Implementation hyperparameters

The proposed model was trained based on the specific hyperparameters in Table 1. MAE is used as the loss function since it is robust to outliers and provides a more interpretable error metric in the context of irradiance forecasting. The Adam optimizer is employed due to its adaptive learning rate capabilities, leading to faster convergence. A validation set

Table 1
Proposed model training hyperparameters.

Training Parameters		Attention Mechanism Parameters	
Input Image Size	224, 224, 3 * no of images	Activation	ReLU
No of Image	1, 2, 4, 6, 8	Kernel	L2 =
		regularizer	0.05
Train Batch Size	16	Activation	Linear
Validation Batch Size	16	Kernel	L2 =
		regularizer	0.05
Test Batch Size	1	Optimizer	Adam
Data Generator Batch Size	2	Learning Rate	1e-4
Number of minutes	10, 20, 30, 40, 50, 60	Beta 1, Beta 2	0.7, 0.9
Early Stopping	Monitor = val_loss, patience = 10	decay	0.001
Epoch	185, 200	Loss, Metrics	MAE

is used to monitor the model's performance during training and prevent overfitting. Early stopping is applied based on validation loss.

3. Result and analysis

The section talks about the experimental results and analysis carried out in this manuscript including the results and analysis, statistical and sensitivity analysis, proposed model component analysis, applicability domain using the Williams plot, comparison with state-of-the-art models, and finally concludes with the result discussion, identified limitations, and future works.

3.1. Proposed model forecast result

Table 2 and Fig. 8 show the performance of the proposed model for solar irradiance forecasting and prediction using sky images. Lower values of MBE, MAE, and RMSE indicate better accuracy, while higher R^2 values mean a better fit of the model to the data. The results in Table 2 show that as the sequence length increases, the performance of the proposed model tends to degrade slightly in terms of MAE, RMSE, and R^2 suggesting that longer sequences might introduce more

Table 2
Forecast results.

Analysis	MBE ↓	MAE ↓	RMSE ↓	R^2 ↑
10 Minutes Sequence 1	5.18142	35.108	63.127	0.943
10 Minutes Sequence 2	1.36329	26.979	49.375	0.962
10 Minutes Sequence 4	1.99512	46.206	67.044	0.875
10 Minutes Sequence 8	0.74282	49.667	68.797	0.854
20 Minutes Sequence 1	7.74272	55.411	87.457	0.835
20 Minutes Sequence 2	8.85133	51.828	80.334	0.864
20 Minutes Sequence 4	7.52518	52.464	74.707	0.877
20 Minutes Sequence 8	2.07154	56.069	76.725	0.818
30 Minutes Sequence 1	3.23372	61.721	93.460	0.821
30-min Sequence 2	1.83530	59.773	86.409	0.816
30 Minutes Sequence 4	7.12370	78.274	105.742	0.673
30 Minutes Sequence 8	2.13401	60.190	81.435	0.765
40 Minutes Sequence 1	9.95460	84.804	121.292	0.683
40 Minutes Sequence 2	2.81683	61.711	91.032	0.825
40 Minutes Sequence 4	0.92924	65.568	90.477	0.749
40 Minutes Sequence 8	5.72550	69.912	93.417	0.654
50 Minutes Sequence 1	3.59415	69.812	103.650	0.785
50 Minutes Sequence 2	4.85653	67.545	99.326	0.777
50 Minutes Sequence 4	3.78572	78.665	106.640	0.667
50 Minutes Sequence 8	1.91867	58.725	87.432	0.812
60 Minutes Sequence 1	5.43587	75.217	109.321	0.745
60 Minutes Sequence 2	9.53313	73.485	105.598	0.760
60 Minutes Sequence 4	1.98154	80.758	107.921	0.616
60 Minutes Sequence 8	8.76475	80.372	105.891	0.653
50 Minutes Sequence 8 was trained using 350 epochs without early stopping				

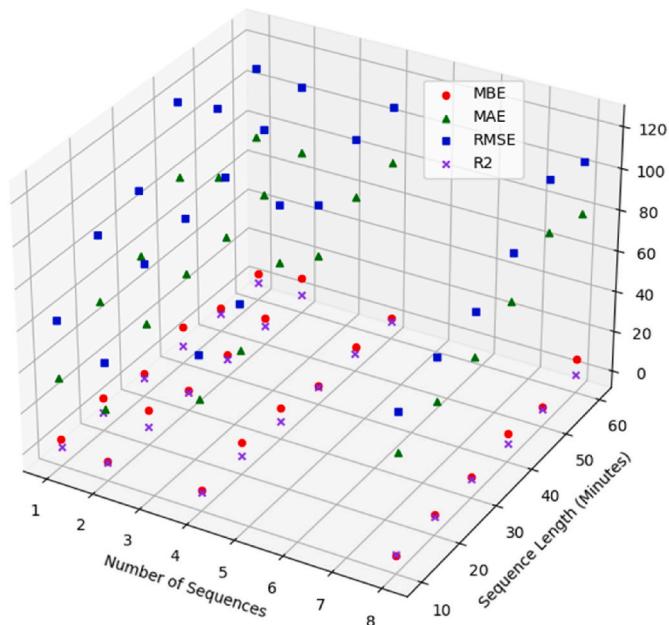


Fig. 8. 3D Forecast Result visualization.

complexity or noise, making predictions slightly less accurate. For instance, comparing Sequence 1 with Sequence 8, there is often a noticeable increase in MAE, RMSE, and a decrease in R^2 when the number of sequences increases indicating that while additional sequences might provide more information for the model, it could also introduce more variability or noise, leading to slightly less accurate predictions. Among the configurations tested, certain setups consistently perform better than others. For example, Timeseries 10 Minutes Sequence 2 and Timeseries 20 Minutes Sequence 4 tend to have lower MAE, RMSE, and higher R^2 compared to other configurations within their respective sequence lengths. These configurations are considered the most optimal setups for the proposed model in terms of accuracy and predictive power.

On the other hand, some configurations exhibit higher errors and lower R^2 values, indicating areas where the proposed model might struggle to accurately predict solar irradiance. Understanding the factors contributing to these discrepancies projects the future work of this study via model improvement, such as refining the architecture, fine-tuning hyperparameters, or incorporating additional features. Notwithstanding, the recorded results demonstrate the effectiveness of the proposed model in solar irradiance forecasting using sky images. To represent the forecast results both from the image and numerical feature, the dataset was reprocessed by combining the length and time of the sequence into rainy and sunny days. The times were adjusted as seen in Fig. 9. Following related works approaches, the GHI were selected and the graph shows that the proposed model is capable of correctly predicting the ground truth images.

3.2. Component analysis of the proposed model

The proposed model component analysis is shown in Table 3. In each minute, the best configuration performance is used for this analysis. For example, Timeseries 10 Minutes Sequence 2, Timeseries 20 Minutes Sequence 4, Timeseries 30 Minutes Sequence 2, Timeseries 40 Minutes Sequence 2, Timeseries 50 Minutes Sequence 2, and Timeseries 60 Minutes Sequence 2 have lower MAE, RMSE, and higher R^2 compared to other configurations within their respective sequence lengths. These configurations are considered the most optimal setups for the proposed model in terms of accuracy and predictive power.

Table 3, Figs. 10 and 11 explain the impact of each component of the

proposed model (baseline, attention mechanism, sequence to sequence model) on the performance metrics (MAE, RMSE, R^2) individually and in combination within the proposed model. Across all time intervals, the baseline model demonstrates a moderate level of performance, serving as a reference point for evaluating the effectiveness of enhancements. It provides a baseline level of accuracy in solar irradiance forecasting, but its performance can be further improved with additional features. When adding the attention mechanism to the baseline model, we observe improvements in MAE, RMSE, and R^2 , indicating that attention helps the model to focus on relevant features in the input data, enhancing prediction accuracy. The improvements are notable across various time intervals, demonstrating the effectiveness of attention in capturing important spatial information from sky images.

Incorporating sequence information into the baseline model leads to further improvements in performance metrics, particularly in MAE and RMSE suggesting that accounting for temporal dependencies in the data enhances the model's ability to capture patterns and make accurate forecasts over time. The proposed model, which integrates attention mechanisms and a sequence-to-sequence model demonstrates the highest level of performance across all metrics and time intervals. Combining attention and sequence information synergistically improves the model's predictive capabilities, resulting in the most accurate forecasts of solar irradiance. The proposed model achieves the lowest MAE and RMSE values and the highest R^2 values, indicating superior accuracy and model fit compared to individual components and the baseline model. The component analysis highlights the complementary nature of attention mechanisms and the sequence-to-sequence model in enhancing the baseline model's performance for solar irradiance forecasting. While each component contributes to improvements in prediction accuracy independently, their integration within the proposed model results in the most significant enhancements, underscoring the importance of considering both spatial and temporal aspects of the data for accurate forecasting.

3.3. Statistical analysis of the proposed model

The statistical analysis of the proposed model based on time (10, 20, 30, 40, 50 and 60 min respectively) is shown in Table 4. Examining the MBE values, we can observe that the model's bias fluctuates across the time intervals, with the lowest bias occurring at the 10-min interval and the highest at the 20-min interval. This suggests that the model's ability to accurately predict the target variable may be influenced by the time frame considered. Similarly, the MAE and RMSE values generally increase as the time interval grows longer, indicating that the model's overall error and deviation from the observed values become more pronounced with longer forecasting horizons. The R^2 values, which represent the model's coefficient of determination, showcase a declining trend as the time interval increases. This suggests that the model's ability to explain the variance in the target variable diminishes over longer periods, potentially due to the increased complexity and uncertainty inherent in longer-term forecasts. Lastly, the FSS values, which measure the model's spatial forecasting ability, exhibit relatively stable performance across the different time intervals, with only minor fluctuations observed. This indicates that the model maintains a consistent level of skill in capturing the spatial patterns and distributions of the target variable, regardless of the time frame considered. The 10-min interval is considered to be the best performing across most of the evaluation metrics with an MBE value is 2.321 W/m², MAE value of 39.490 W/m², RMSE value of 62.086 W/m², R^2 value is 0.909 W/m² and FSS value of 23.589 W/m².

The statistical analysis across different sequence lengths (1, 2, 4, and 8) for solar irradiance forecasting and prediction is presented in Table 5. Looking at the MBE results, the model shows a decreasing bias as the sequence length increases. The MBE values range from 5.857 W/m² for sequence length 1–3.543 W/m² for sequence length 8, indicating that the model's tendency to over- or under-estimate the solar irradiance

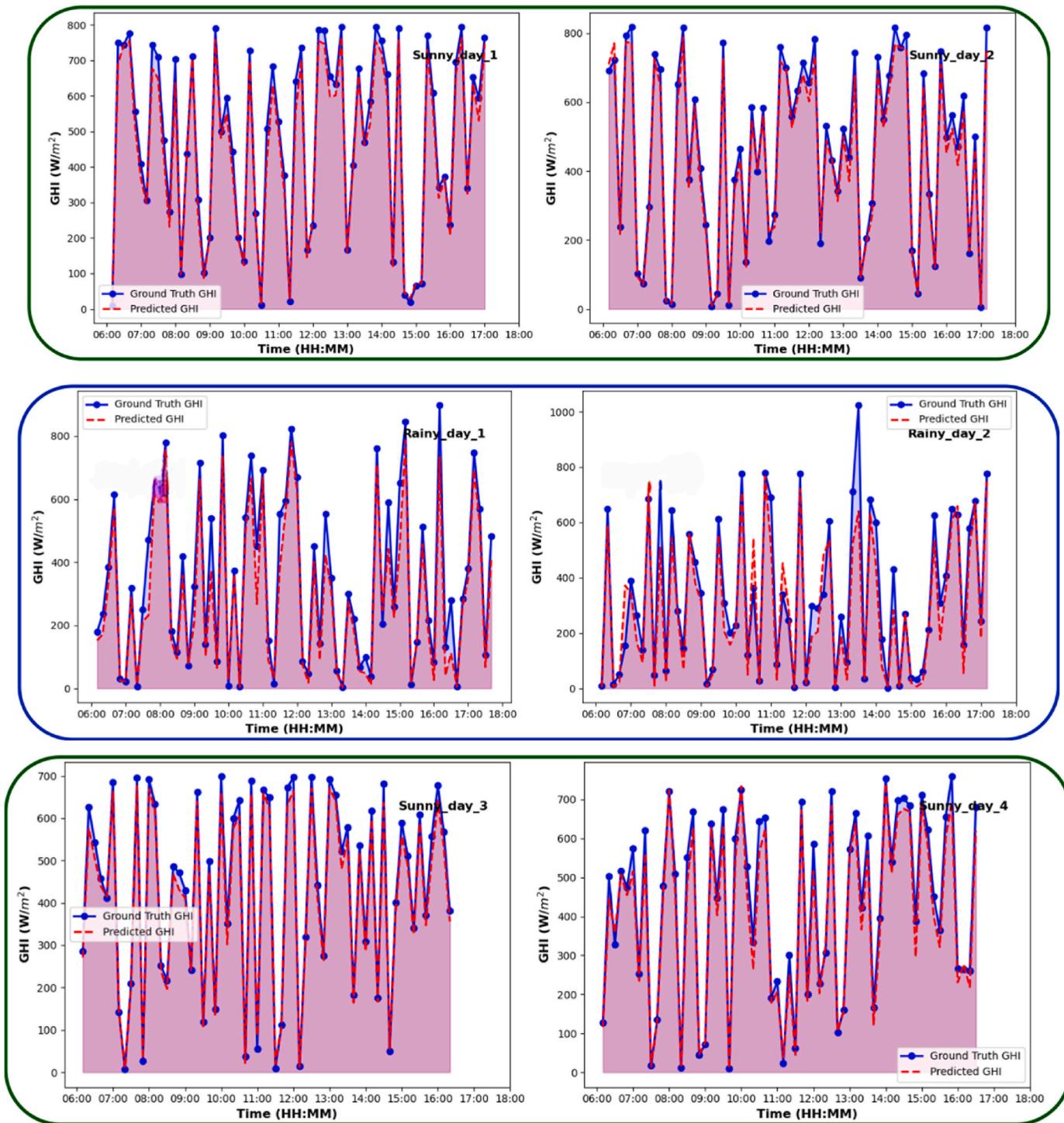


Fig. 9. Forecasting Results Rainy Days vs Sunny Days.

decreases with longer input sequences. The 95 % confidence intervals also become narrower, suggesting that the model's bias becomes more consistent as the sequence length increases. The MAE values range from 63.679 W/m² for sequence length 1–67.194 W/m² for sequence length 8, indicating that the model maintains a relatively consistent level of absolute error across different sequence lengths. The RMSE values follow a similar trend, with the lowest value of 85.346 W/m² for sequence length 2 and the highest value of 96.385 W/m² for sequence lengths 1 and 4. The 95 % confidence intervals for both MAE and RMSE suggest that the model's performance is statistically significant. The R² values range from 0.802 for sequence length 1 to 0.653 for sequence

length 8, indicating that the model explains a larger proportion of the variance in the solar irradiance data for shorter sequence lengths. The FSS values range from 16.348 for sequence length 1 to 27.0 for sequence length 8, suggesting that the model's spatial forecasting skill improves as the sequence length increases. These results demonstrate that the proposed model is capable of accurately forecasting and predicting solar irradiance, with improved performance in terms of bias, absolute error, and spatial forecasting skill as the input sequence length is increased. The trade-off is a slight decrease in the model's coefficient of determination, indicating that longer sequences may capture more complex relationships in the data, but the model's overall predictive power

Table 3
Proposed model component analysis.

	Analysis	MBE ↓	MAE ↓	RMSE ↓	R ² ↑
Baseline	10 Minutes	1.9515	86.758	107.921	0.616
Baseline + Attention	Sequence 2	1.7587	61.952	87.086	0.873
Baseline + Sequence		1.7353	59.773	86.409	0.816
Proposed (baseline + Attention + Sequence)		1.3633	26.979	49.375	0.962
Baseline	20 Minutes	8.7648	80.372	105.891	0.453
Baseline + Attention	Sequence 4	7.7924	57.622	88.371	0.869
Baseline + Sequence		7.7427	55.411	87.457	0.835
Proposed (baseline + Attention + Sequence)		7.5252	52.464	74.707	0.877
Baseline	30-min	2.0715	76.069	107.921	0.518
Baseline + Attention	Sequence 2	1.9815	65.758	97.725	0.616
Baseline + Sequence		1.9921	66.206	96.044	0.775
Proposed (baseline + Attention + Sequence)		1.8353	59.773	86.409	0.816
Baseline	40 Minutes	1.9951	96.206	99.044	0.675
Baseline + Attention	Sequence 4	1.8412	71.123	92.409	0.723
Baseline + Sequence		1.8353	70.773	93.112	0.716
Proposed (baseline + Attention + Sequence)		0.9292	65.568	90.477	0.749
Baseline	50 Minutes	2.1340	97.190	131.435	0.765
Baseline + Attention	Sequence 8	1.8552	91.773	118.870	0.816
Baseline + Sequence		1.8454	90.773	119.409	0.816
Proposed (baseline + Attention + Sequence)		1.91867	58.725	87.432	0.812
Baseline	60 Minutes	2.8168	101.711	131.032	0.325
Baseline + Attention	Sequence 4	2.3924	97.622	120.371	0.469
Baseline + Sequence		2.3427	95.411	121.457	0.435
Proposed (baseline + Attention + Sequence)		1.9815	80.758	107.921	0.616

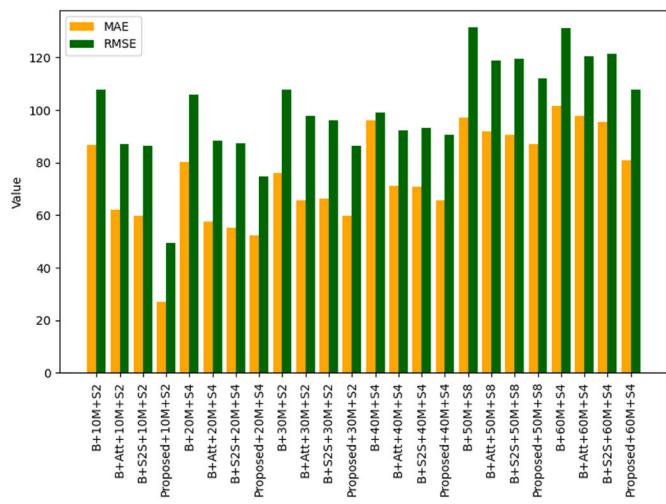


Fig. 10. Component Analysis MAE vs RMSE.

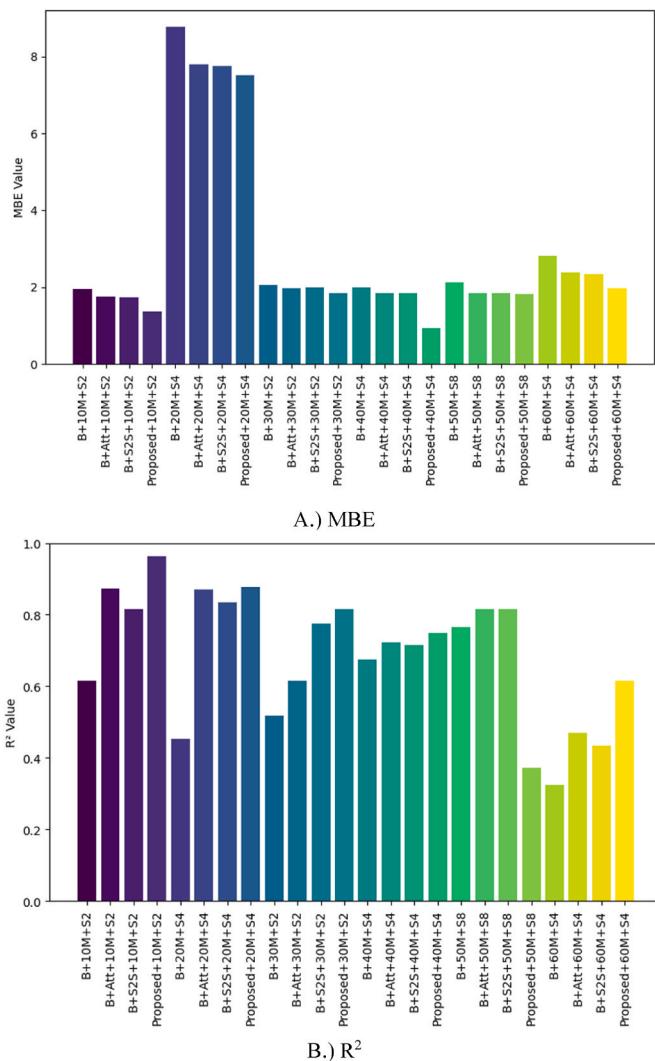


Fig. 11. Component analysis.

remains strong. Overall, the sequence length 2 appears to be the best performing overall, with an MBE of 4.876 W/m², MAE of 56.887 W/m², RMSE of 85.346 W/m², R² of 0.834 and FSS of 24.881 W/m². The model's spatial forecasting ability is also strong at this sequence length.

3.4. Sensitivity analysis of the proposed model

Fig. 11 shows the sensitivity analysis of the proposed model to understand the proposed model behaviour in short and long-term forecasting. Fig. 12A shows the relationship between the MBE and the R² for the different sequence lengths and lead times. As the sequence length increases, the MBE and R² values generally move towards higher MBE and lower R², indicating a tradeoff between bias and predictive performance. The shorter lead times (10–30 min) tend to exhibit lower MBE and higher R², suggesting better model performance for shorter-term forecasts. The longer lead times (40–60 min) show a wider spread in MBE and R² values, indicating more variability in model performance for longer-term forecasts.

Fig. 12B. Focuses on the relationship between the MAE and the lead time for the different sequence lengths. As the lead time increases, the MAE generally increases for all sequence lengths, indicating that the model's accuracy degrades over longer forecast horizons. Shorter sequence lengths (Seq = 1, 2) tend to have lower MAE values compared to longer sequence lengths (Seq = 4, 8), especially for the shorter lead times (10–30 min). The spread in MAE values increases with longer lead

Table 4

Statistical analysis based on time in minutes.

Metric	Minutes	Mean	Standard Deviation	Min	Max	95 % Confidence Interval
MBE	10	2.321	1.975	0.742	5.181	[-0.821, 5.463] W/m ²
	20	6.548	3.040	2.072	8.851	[1.710, 11.385] W/m ²
	30	3.582	2.437	1.835	7.1237	[-0.296, 7.459] W/m ²
	40	4.857	3.930	0.929	9.955	[-1.396, 11.110] W/m ²
	50	3.514	1.259	1.819	4.857	[1.510, 5.517] W/m ²
	60	6.428	3.457	1.981	9.533	[0.928, 11.930] W/m ²
MAE	10	39.490	10.399	26.979	49.667	[22.943, 56.037] W/m ²
	20	53.943	2.108	51.828	56.069	[50.5881, 57.298] W/m ²
	30	64.989	8.896	59.773	78.274	[50.834, 79.144] W/m ²
	40	70.499	10.108	61.711	84.804	[54.415, 86.582] W/m ²
	50	75.743	8.880	67.545	86.952	[61.613, 89.875] W/m ²
	60	77.458	3.660	73.485	80.758	[71.634, 83.282] W/m ²
RMSE	10	62.086	8.799	49.375	68.797	[48.084, 76.087] W/m ²
	20	79.806	5.607	74.707	87.457	[70.884, 88.727] W/m ²
	30	91.762	10.546	81.435	105.742	[74.981, 108.542] W/m ²
	40	99.055	14.880	90.477	121.292	[75.377, 122.731] W/m ²
	50	105.426	5.360	99.326	112.086	[96.896, 113.955] W/m ²
	60	107.183	1.760	105.598	109.321	[17.195, 28.485] W/m ²
R^2	10	0.909	0.052	0.854	0.962	[0.826, 0.991] W/m ²
	20	0.848	0.027	0.818	0.877	[0.805, 0.891] W/m ²
	30	0.769	0.069	0.673	0.821	[0.659, 0.8780] W/m ²
	40	0.728	0.076	0.654	0.825	[0.607, 0.848] W/m ²
	50	0.650	0.193	0.373	0.785	[0.644, 0.957] W/m ²
	60	0.6435	0.143	0.453	0.760	[0.417, 0.870] W/m ²
FSS	10	23.589	7.824	16.790	34.850	[11.141, 36.038] W/m ²
	20	23.423	7.292	15.250	31.510	[11.820, 35.025] W/m ²
	30	20.855	9.752	9.170	32.580	[5.338, 36.372] W/m ²
	40	21.005	13.075	1.480	29.190	[0.200, 41.809] W/m ²
	50	21.590	1.099	20.630	22.700	[19.840, 23.340] W/m ²
	60	22.840	3.547	18.600	27.270	[17.195, 28.485] W/m ²

Table 5

Statistical analysis based on sequence length.

Metric	Sequence	Mean	Standard Deviation	Min	Max	95 % Confidence Interval
MBE	1	5.857	2.568	3.233	9.955	[3.162–8.552] W/m ²
	2	4.876	3.556	1.363	9.533	[1.142, 8.611] W/m ²
	4	3.890	2.817	0.929	7.525	[0.933, 6.846] W/m ²
	8	3.543	3.066	0.743	8.765	[0.326, 6.760] W/m ²
MAE	1	63.679	17.352	35.108	84.804	[45.469, 81.889] W/m ²
	2	56.887	16.375	26.979	73.485	[39.702, 74.072] W/m ²
	4	66.989	14.820	46.206	80.758	[51.436, 82.542] W/m ²
	8	67.194	14.501	49.667	86.952	[51.976, 82.412] W/m ²
RMSE	1	96.385	20.173	63.127	121.292	[75.214, 117.555] W/m ²
	2	85.346	19.789	49.375	105.598	[64.578, 106.113] W/m ²
	4	96.385	20.173	63.127	121.292	[75.214, 117.555] W/m ²
	8	89.726	17.033	68.797	112.086	[71.850, 107.601] W/m ²
R^2	1	0.802	0.088	0.683	0.943	[0.709, 0.895] W/m ²
	2	0.834	0.073	0.760	0.962	[0.758, 0.910] W/m ²
	4	0.743	0.112	0.616	0.877	[0.626, 0.860] W/m ²
	8	0.653	0.199	0.373	0.854	[0.444, 0.862] W/m ²
FSS	1	16.348	7.736	1.480	22.37	[8.230, 24.467] W/m ²
	2	24.881	5.305	19.770	34.85	[19.314, 30.449] W/m ²
	4	20.638	6.869	9.170	27.16	[13.429, 27.847] W/m ²
	8	27.0	5.198	20.66	32.58	[21.545, 32.455] W/m ²

times, suggesting that the model's performance becomes more variable for longer-term forecasts. Overall, the sensitivity analysis reveals that the proposed model's performance is influenced by both the sequence length and the lead time. Shorter sequence lengths and shorter lead times generally result in better bias, predictive accuracy, and consistency in the model's performance. However, longer sequence lengths and lead times introduce more variability and tradeoffs between different performance metrics, which is important to consider when selecting the appropriate model configuration for a given application or forecasting scenario.

3.5. Applicability domain of the proposed model

The underlying data characteristics of the deployed dataset are visualised using the Williams plot as shown in Fig. 13. This is to identify potential sources of bias or instability in the model and make informed decisions on data pre-processing and feature engineering to optimize the performance and reliability of our proposed model for solar irradiance forecasting. The provided Williams plots offer a comprehensive visualization and analysis of the relationship between leverage and standardized residuals for different solar irradiance components: DNI, DHI,

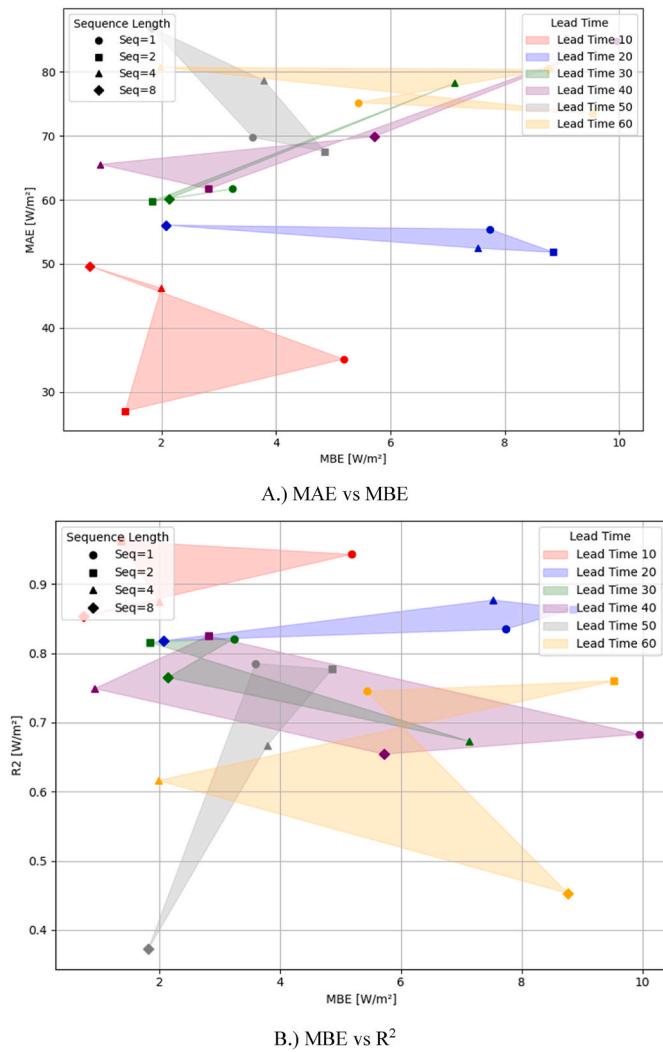


Fig. 12. Sensitivity analysis.

and GHI.

Examining the Williams plot for DNI, we can observe a distinct separation between the valid data points and the suspected data points. The valid data points are clustered within a well-defined region, while the suspected data points exhibit a wider spread and are predominantly located outside the upper leverage threshold. This pattern suggests that the suspected data points may be subject to higher leverage and potentially greater influence on the model's performance, warranting further investigation or potential exclusion. The Williams plot for DHI presents a similar scenario, with the valid data points occupying a more confined space within the leverage thresholds, while the suspected data points exhibit a broader distribution and a significant number falling outside the upper leverage limit. This indicates that the suspected data points for DHI may also have a higher degree of leverage and influence on the model's behaviour. The Williams plot for GHI showcases a more diverse distribution of data points, with both valid and suspected data points spanning a wider range of leverage values. While the valid data points are predominantly concentrated within the leverage thresholds, a considerable number of suspected data points are found outside the upper leverage limit. This suggests that the GHI data may have a higher degree of variability and potential outliers, which could impact the model's performance and reliability.

3.6. Comparison with state-of-the-art models on the same dataset

This study compared the performance of the proposed model using different ML models including two ANNs with varied back-propagation (BP), three gradient boosting machines (GBMs) with different loss functions, and a random forest (RF) model. This paper only recorded the best result from the several evaluated ANNs, GBMs, and RF models. The same input features were used for this comparison including DNI, DHI, GHI, clear sky DNI, clear sky DHI, clear sky GHI, infrared radiation, dry bulb temperature, wind chill temperature, relative humidity, wind speed, peak wind speed, and pressure from the preceding 4 h. For the comparisons, three evaluation metrics were used including the MBE, RMSE, and Skill Scores as highlighted in the subsequent tables.

The MBE is a significant metric for assessing the performance of solar irradiance forecasting models as seen in Table 6 and Fig. 14. Comparing the MBE results of the proposed model with state-of-the-art models reveals noteworthy insights. Across various sequence lengths and time intervals, the proposed model consistently outperforms most of the compared models in terms of MBE. For instance, at 10 min sequence length 1, the proposed model achieves an MBE of 5.18, which is notably lower than most state-of-the-art models such as ANN1 (23.64), ANN2 (-5.39), GBM1 (-0.38), GBM2 (-0.65), GBM3 (-2.02), RF (-1.32), and SCNN (19.91). This trend persists across different sequence lengths and time intervals, indicating the superiority of the proposed model in accurately forecasting solar irradiance compared to existing models. Furthermore, the proposed model consistently demonstrates competitive performance across various time intervals, showcasing its robustness and effectiveness in handling different forecasting horizons. Overall, these results highlight the effectiveness and superiority of the proposed model in solar irradiance forecasting tasks compared to state-of-the-art models.

Table 7 and Fig. 15 present the RMSE result of the proposed models against the state-of-the-art models. Across the different sequence lengths and models, there are noticeable variations in performance. For instance, in the first row (L = 10), the "Proposed" algorithm consistently performs lower compared to other algorithms. Conversely, in some cases, like at L = 60, it outperforms certain other algorithms suggesting that the optimal hyperparameter setting of the proposed model was not met. Additionally, it is noticed that as the sequence length increases, the performance of some algorithms either remains stable or slightly deteriorates, while others exhibit more significant fluctuations. This could indicate varying degrees of robustness of the algorithms to longer sequences.

Table 8 and Fig. 16 provide the forecast skill scores comparison between the state-of-the-art models and the proposed model. The forecast skill score indicates the performance of each algorithm in making predictions. As the sequence length increases, there are fluctuations in the performance of different algorithms. For instance, for ANN1, ANN2, GBM1, and GBM2, there is a slight improvement in forecast skill scores as the sequence length increases, especially up to a certain point. Notably, there are instances where the forecast skill scores decrease unpredictably with increasing sequence length, indicating that longer sequences might not necessarily lead to better predictions for certain algorithms. The RF, SCNN, and 3DCNN models achieve relatively high forecast skill scores compared to other algorithms in several scenarios while GM2 and GM3, exhibit lower forecast skill scores across multiple sequence lengths. The interaction between sequence length and algorithm performance is complex and depends on the specific algorithm. Some algorithms may be more suitable for handling longer sequences, while others may perform better with shorter sequences. The "Proposed" model forecast skill scores were relatively higher compared to the proposed models. The analysis of the forecast skill scores across different sequence lengths provides valuable insights into the performance of various algorithms.

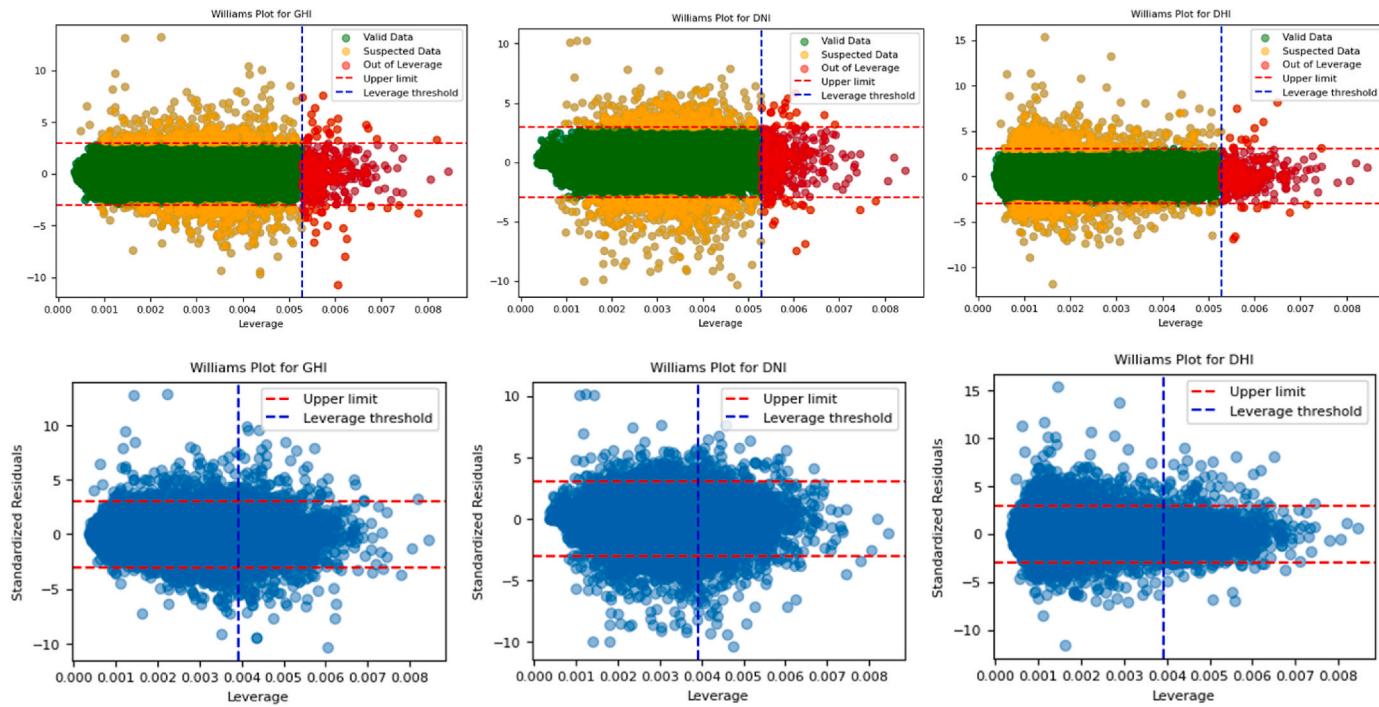


Fig. 13. Applicability domain of the proposed Model based on william plot.

Table 6
Forecast MBEs [proposed model vs. State-of-the-art models].

L[min]	Sequence Length	ANN1	ANN2	GBM1	GBM2	GBM3	RF	SCNN	3DCNN	Proposed
10	1	23.64	-5.39	-0.38	-0.65	-2.02	-1.32	-1.37	19.91	5.18
	2	6.74	-6.24	-0.79	-1.13	-2.48	-0.03	1.29	6.08	1.36
	4	1.13	-12.04	-1.55	-1.91	-3.39	0.07	10.35	2.31	1.99
	8	13.12	56.54	-3.17	-3.49	-4.69	-2.02	9.15	1.94	0.74
20	1	-7.41	9.86	-0.49	-0.41	0.15	-1.10	-13.18	10.01	7.74
	2	40.12	-28.42	-1.43	-1.13	-0.48	-0.44	9.22	-1.32	8.85
	4	5.00	-19.33	-3.03	-2.56	-2.30	-0.28	-5.86	1.09	7.53
	8	-6.63	-20.42	-6.22	-5.79	-4.85	-3.17	10.35	-4.90	2.08
30	1	-2.89	-77.21	-0.96	0.42	2.79	-2.24	1.14	-19.49	3.23
	2	-6.21	-38.40	-2.13	-0.48	2.13	-0.76	4.26	-2.66	1.83
	4	21.58	-10.58	-4.48	-2.62	-0.76	-1.59	-49.17	0.29	7.12
	8	-11.88	-23.78	-9.34	-7.26	-5.54	-5.78	6.66	6.30	2.13
40	1	-17.61	28.96	-1.23	1.08	4.79	-0.17	3.63	-1.61	9.95
	2	-29.19	-36.05	-2.70	-0.23	3.82	-0.76	-2.66	10.60	2.82
	4	-36.37	-30.31	-5.68	-3.11	0.62	-2.80	-2.92	11.70	0.93
	8	2.59	-56.42	-12.29	-9.06	-5.91	-8.21	-4.29	8.85	5.73
50	1	16.25	-28.99	-1.70	0.84	5.67	-1.02	-11.75	-14.07	3.59
	2	-8.79	-44.02	-3.49	-0.66	4.56	-2.05	8.30	2.79	4.86
	4	-1.76	-29.78	-7.05	-3.85	0.93	-6.01	-6.45	-1.08	3.79
	8	-20.52	-40.17	-15.35	-11.51	-6.71	-12.26	19.58	7.41	1.82
60	1	36.68	-15.45	-2.30	0.65	6.01	-1.41	-6.18	-6.86	5.44
	2	-1.62	-66.35	-4.27	-0.85	4.80	-2.72	-8.14	4.38	9.53
	4	80.22	-89.03	-8.65	-4.64	1.50	-6.99	2.84	4.44	1.98
	8	-29.40	-27.53	-18.69	-13.94	-7.10	-15.20	-1.23	0.30	8.76

3.7. Limitations and future works

Despite the encouraging results yielded by the proposed model, the result analysis reveals areas where the model could need further exploration which include.

- ❖ Robustness and generalizability are impacted by changes in input data, such as differences in weather conditions or image quality.
- ❖ The model relies solely on sky image sequences for prediction, without incorporating any other measurements or data sources.

❖ The increased complexity of the model leads to higher computing costs and longer training times, which may be a tradeoff for improved feature capture and prediction accuracy.

To address the proposed model's limitations, future works will enhance robustness by using data augmentation and diverse datasets. Incorporate additional data sources for improved accuracy, optimize complexity through model simplification, and establish continuous evaluation to refine performance over time.

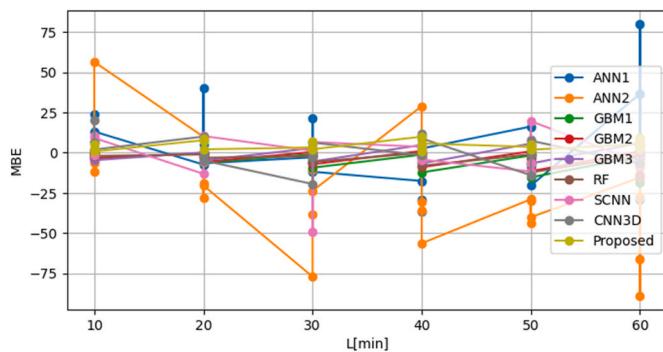


Fig. 14. Mbe result comparison with state-of-the-art models.

4. Conclusions

This study presents an Attention Sequence -To-Sequence Convolutional Neural Network to predict the future values of GHI, DNI, and DHI for utility operations using sky images. A convolutional neural network is for spatial feature extraction, an attention mechanism to focus on the relevant regions, and a sequence-to-sequence model to capture temporal dependencies to accurately define the temporal structure of the predictions. The model's accuracy and predictive power were optimized by considering both spatial (through attention mechanisms) and temporal (through sequence information) aspects of the data. The main findings of this study are summarized thus.

- ❖ Based on an extensive evaluation using five years of publicly available data, the proposed model, which integrates a Convolutional Neural Network (CNN), attention mechanism, and sequence-to-sequence architecture, outperformed the benchmark methods across various evaluation metrics.
- ❖ Furthermore, the study findings indicate that the proposed model exhibits a high degree of robustness to factors such as the length of the input image sequence and the forecast horizon. In contrast, the benchmark models exhibited much greater instability, with a significant decrease in forecasting accuracy as the lead time increased.

❖ The integration of the attention mechanism and the sequence-to-sequence model into the CNN framework has significantly enhanced the model's capability to identify and assimilate the relevant features from the input sky images, leading to improved accuracy in solar irradiance forecasting.

❖ The sequence length 2 appears to be the best performing overall, with an MBE of 4.876 W/m^2 , MAE of 56.887 W/m^2 , RMSE of 85.346 W/m^2 , R^2 of 0.834 and FSS of 24.881 W/m^2 . The model's spatial forecasting ability is also strong at this sequence length.

❖ The 10-min interval is considered to be the best performing across most of the evaluation metrics with an MBE value is 2.321 W/m^2 , MAE value of 39.490 W/m^2 , RMSE value of 62.086 W/m^2 , R^2 value is 0.909 W/m^2 and FSS value of 23.589 W/m^2 .

❖ As the sequence length increases, the MBE and R^2 values generally move towards higher MBE and lower R^2 , indicating a tradeoff between bias and predictive performance. The shorter lead times (10–30 min) tend to exhibit lower MBE and higher R^2 , suggesting better model performance for shorter-term forecasts. The longer lead times (40–60 min) show a wider spread in MBE and R^2 values, indicating more variability in model performance for longer-term forecasts.

❖ The sensitivity analysis reveals that the proposed model's performance is influenced by both the sequence length and the lead time.

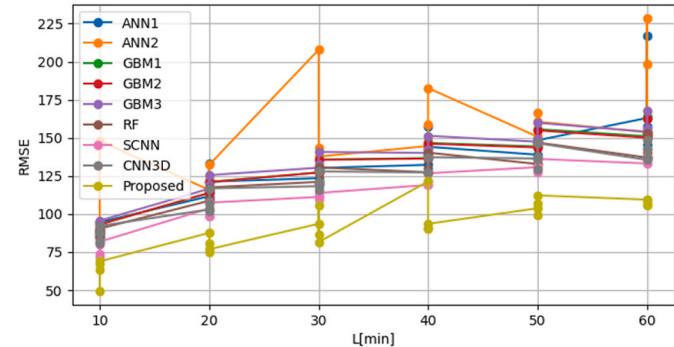


Fig. 15. Rmse result comparison with state-of-the-art models.

Table 7
Forecast RMSE [proposed model vs. State-of-the-art models].

L[min]	Sequence Length	ANN1	ANN2	GBM1	GBM2	GBM3	RF	SCNN	3DCNN	Proposed
10	1	94.52	88.92	87.23	87.16	88.89	85.36	79.90	81.82	63.13
	2	85.28	88.71	87.80	87.67	89.53	83.66	71.30	80.28	49.38
	4	84.51	90.97	88.91	88.79	91.04	85.39	73.76	87.38	67.04
	8	94.49	147.89	93.07	92.98	95.53	90.26	81.67	92.04	68.80
20	1	111.53	115.71	113.70	113.77	116.50	108.58	103.53	102.87	87.46
	2	132.79	122.81	114.09	114.05	117.17	108.98	98.53	101.73	80.33
	4	110.47	121.09	115.28	115.13	118.99	109.60	98.27	106.86	74.71
	8	121.12	132.17	120.96	120.84	125.45	117.33	107.35	116.69	76.73
30	1	123.42	207.87	127.22	127.16	130.30	121.01	111.07	118.16	93.46
	2	119.41	143.39	127.52	127.35	130.74	119.59	109.33	115.73	86.41
	4	130.69	126.96	128.90	128.60	132.64	121.63	111.72	121.11	105.74
	8	130.12	137.62	135.70	135.43	140.66	130.52	113.69	127.89	81.44
40	1	132.24	144.59	136.64	136.36	140.00	127.34	118.98	127.24	121.29
	2	142.73	148.00	136.92	136.56	140.33	127.54	119.35	127.58	91.03
	4	157.26	158.46	138.54	138.09	142.12	129.70	121.48	126.67	90.48
	8	144.06	182.62	146.56	146.07	151.32	140.21	126.66	137.21	93.42
50	1	138.85	150.71	144.22	143.58	147.36	132.75	130.67	136.38	103.65
	2	148.82	166.43	144.63	144.01	147.77	133.32	127.49	129.49	99.33
	4	136.49	148.22	146.32	145.65	149.41	135.62	131.29	137.43	106.64
	8	148.34	160.51	155.61	154.82	159.82	146.86	136.10	146.43	112.09
60	1	162.98	153.64	150.92	150.01	153.91	136.96	133.12	135.49	109.32
	2	145.11	198.57	151.37	150.52	154.45	137.75	135.43	137.05	105.60
	4	216.87	228.69	153.17	151.99	156.00	140.29	137.55	142.77	107.92
	8	157.38	162.25	164.01	162.69	167.57	152.97	142.95	148.23	105.89

Table 8

Forecast skill scores [%] [proposed model vs. State-of-the-art models].

L[min]	Sequence Length	ANN1	ANN2	GBM1	GM2	GM3	RF	SCNN	3DCNN	Proposed
10	1	-4.20	1.97	3.84	3.91	2.01	5.90	11.92	9.80	21.93
	2	6.05	2.27	3.27	3.42	1.37	7.83	21.45	11.56	34.85
	4	7.82	0.77	3.02	3.15	0.70	6.86	19.54	4.69	16.79
	8	-0.40	-57.15	1.11	-1.51	1.20	4.09	13.21	2.20	20.79
20	1	9.21	5.81	7.45	7.39	5.17	11.62	15.73	16.26	15.25
	2	-7.41	0.66	7.72	7.75	5.23	11.85	20.30	17.71	19.77
	4	10.34	1.72	6.44	6.56	3.43	11.05	20.24	13.27	27.16
	8	3.22	-5.61	3.35	-0.24	3.44	6.25	14.22	6.76	31.51
30	1	14.68	-43.70	12.06	12.10	9.93	16.35	23.22	18.32	18.46
	2	18.03	1.57	12.46	12.58	10.25	17.90	24.94	20.55	23.21
	4	11.43	13.95	12.64	12.84	10.11	17.57	24.28	17.92	9.17
	8	7.50	2.17	3.53	3.73	0.01	7.22	19.18	9.09	32.58
40	1	21.01	13.64	18.38	18.55	16.38	23.94	28.93	24.00	1.48
	2	15.29	12.17	18.74	18.96	16.72	24.31	29.17	24.29	26.27
	4	7.94	7.24	18.90	19.17	16.81	24.08	28.89	25.85	27.08
	8	4.97	-20.47	3.32	3.64	0.18	7.51	16.44	9.49	29.19
50	1	28.03	21.89	25.25	25.58	23.62	31.20	32.28	29.32	22.37
	2	23.36	14.29	25.52	25.84	23.90	31.34	34.35	33.31	22.70
	4	30.66	24.70	25.67	26.01	24.10	31.10	33.30	30.18	20.63
	8	8.21	0.68	3.71	4.20	1.11	9.13	15.79	9.39	20.66
60	1	26.53	30.74	31.96	32.37	30.61	38.26	39.99	38.92	18.60
	2	34.99	11.03	32.18	32.56	30.80	38.28	39.32	38.60	22.49
	4	4.14	-1.08	32.30	32.82	31.05	37.99	39.20	36.89	23.00
	8	14.40	11.75	10.79	11.51	8.86	16.80	22.24	19.37	27.27

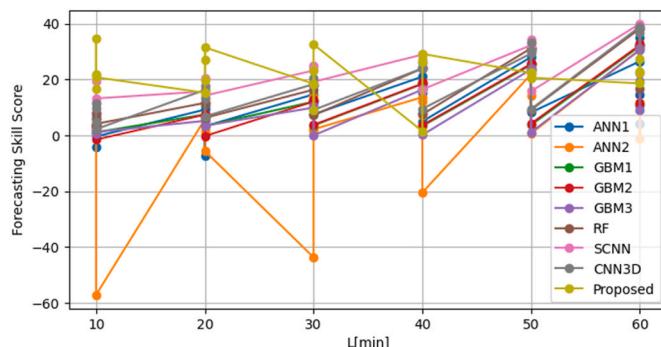


Fig. 16. FSS result comparison with state-of-the-art models.

Shorter sequence lengths and shorter lead times generally result in better bias, predictive accuracy, and consistency in the model's performance. However, longer sequence lengths and lead times introduce more variability and tradeoffs between different performance metrics, which is important to consider when selecting the appropriate model configuration for a given application or forecasting scenario.

- ❖ Using the Williams plot of DHI, DNI, and GHI, the application domain analysis revealed that a credible model is indicated by the majority of data points having leverage values below 0.3 and falling within ± 3 standardised residuals. To guarantee the correctness and resilience of the model, additional research is necessary as a result of the identification of potentially significant outliers.
- ❖ The results explain that longer sequences may introduce complexity or noise, leading to slightly less accurate predictions. Certain setups, such as Timeseries 10 Minutes Sequence 2 and Timeseries 20 Minutes Sequence 4, demonstrated superior performance within their respective sequence lengths. The proposed model's component analysis highlighted the complementary nature of attention mechanisms and sequence to the to-sequence model in enhancing the baseline model's performance.

CRediT authorship contribution statement

Chiagoziem C. Ukwuoma: Writing – original draft, Methodology, Data curation, Conceptualization. **Dongsheng Cai:** Supervision, Funding acquisition, Conceptualization. **Olusola Bamisile:** Validation, Resources. **Hongbo Yin:** Validation, Formal analysis. **Grace Ugochi Nneji:** Validation, Software, Data curation. **Happy N. Monday:** Visualization, Validation, Software, Data curation. **Ariyo Oluwasanmi:** Writing – review & editing. **Qi Huang:** Supervision, Funding acquisition, Conceptualization.

Informed consent

All participants included in the study provided informed consent.

Ethical approval

None of the authors conducted studies involving human participants or animals in this article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors gratefully acknowledge the support of the National Natural Science Foundation of China (Grant No. 52007025), Natural Science Foundation of Sichuan, China (2024NSFSC0119) and the Fundamental Research Funds for the Central Universities, China (Grant No. ZYGX2019J034).

Data availability

The TensorFlow/Keras code employed in our experiment is presently unavailable to the wider audience. However, it will be made accessible

upon publication of the study on our GitHub page.

References

- [1] E. Kabir, P. Kumar, S. Kumar, A.A. Adelodun, K.-H. Kim, Solar energy: potential and future prospects, *Renew. Sustain. Energy Rev.* 82 (Feb. 2018) 894–900, <https://doi.org/10.1016/j.rser.2017.09.094>.
- [2] V. Hosseinkhani, M. Sarvi, A new five-level inverter with reduced leakage current for photovoltaic system applications, *Prot. Control Mod. Power Syst.* 7 (1) (May 2022), <https://doi.org/10.1186/s41601-022-00240-3>.
- [3] C. Hu, Z. Cai, Y. Zhang, R. Yan, Y. Cai, B. Cen, A soft actor-critic deep reinforcement learning method for multi-timescale coordinated operation of microgrids, *Prot. Control Mod. Power Syst.* 7 (1) (Aug. 2022), <https://doi.org/10.1186/s41601-022-00252-z>.
- [4] C. Zhang, L. Hua, C. Ji, M. Shahzad Nazir, T. Peng, An evolutionary robust solar radiation prediction model based on WT-CEEMDAN and IASO-optimized outlier robust extreme learning machine, *Appl. Energy* 322 (Sep. 2022) 119518, <https://doi.org/10.1016/j.apenergy.2022.119518>.
- [5] S. Zhang, J.J.Q. Yu, Bayesian deep learning for dynamic power system state prediction considering renewable energy uncertainty, *J. Mod. Power Syst. Clean Energy* 10 (4) (2022) 913–922, <https://doi.org/10.35833/mpce.2020.000939>.
- [6] L. Cheng, H. Zang, Z. Wei, G. Sun, Secure multi-party household load scheduling framework for real-time demand-side management, *IEEE Trans. Sustain. Energy* 14 (1) (Jan. 2023) 602–612, <https://doi.org/10.1109/tste.2022.3221081>.
- [7] V. Le Guen, N. Thome, A deep physical model for solar irradiance forecasting with fisheye images, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 630–631.
- [8] J. Khan, M.H. Arsalan, Solar power technologies for sustainable electricity generation—A review, *Renew. Sustain. Energy Rev.* 55 (2016) 414–425.
- [9] R. Ahmed, V. Sreram, Y. Mishra, M.D. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization, *Renew. Sustain. Energy Rev.* 124 (2020) 109792.
- [10] R. Zhang, H. Ma, T.K. Saha, Photovoltaic nowcasting with bi-level spatio-temporal analysis incorporating sky images, *IEEE Trans. Sustain. Energy* 12 (3) (2021) 1766–1776.
- [11] D.P. Larson, L. Nonnenmacher, C.F.M. Coimbra, Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest, *Renew. Energy* 91 (2016) 11–20.
- [12] A. Mellit, A. Massi Pavan, E. Ogliari, S. Leva, V. Lugh, Advanced methods for photovoltaic output power forecasting: a review, *Appl. Sci.* 10 (2) (2020) 487.
- [13] M. Diagne, M. David, P. Lauret, J. Boland, N. Schmutz, Review of solar irradiance forecasting methods and a proposition for small-scale insular grids, *Renew. Sustain. Energy Rev.* 27 (2013) 65–76.
- [14] Y. Dodge, *The Concise Encyclopedia of Statistics*, Springer Science & Business Media, 2008.
- [15] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, S. Wang, Short-term solar power forecasting based on weighted Gaussian process regression, *IEEE Trans. Ind. Electron.* 65 (1) (2017) 300–308.
- [16] Y. Ma, Q. Lv, R. Zhang, Y. Zhang, H. Zhu, W. Yin, Short-term photovoltaic power forecasting method based on irradiance correction and error forecasting, *Energy Rep.* 7 (2021) 5495–5509.
- [17] F.P.M. Kreuwel, W. Knap, M. Schmeits, J.V.-G. de Arellano, C.C. van Heerwaarden, Forecasting day-ahead 1-minute irradiance variability from numerical weather predictions, *Sol. Energy* 258 (2023) 57–71.
- [18] A. Mellit, A.M. Pavan, V. Lugh, Deep learning neural networks for short-term photovoltaic power forecasting, *Renew. Energy* 172 (2021) 276–288.
- [19] X. Meng, F. Gao, T. Xu, K. Zhou, W. Li, Q. Wu, Inverter-data-driven second-level power forecasting for photovoltaic power plant, *IEEE Trans. Ind. Electron.* 68 (8) (2020) 7034–7044.
- [20] Y. Chu, M. Li, C.F.M. Coimbra, D. Feng, H. Wang, Intra-hour irradiance forecasting techniques for solar power integration: a review, *iScience* 24 (10) (2021).
- [21] F. Lin, Y. Zhang, J. Wang, Recent advances in intra-hour solar forecasting: a review of ground-based sky image methods, *Int. J. Forecast.* 39 (1) (2023) 244–265.
- [22] S. Park, Y. Kim, N.J. Ferrier, S.M. Collis, R. Sankaran, P.H. Beckman, Prediction of solar irradiance and photovoltaic solar energy product based on cloud coverage estimation using machine learning methods, *Atmosphere* 12 (3) (2021) 395.
- [23] V.A.M. Lopez, G. van Urk, P.J.F. Doodkorte, M. Zeman, O. Isabella, H. Ziar, Using sky-classification to improve the short-term prediction of irradiance with sky images and convolutional neural networks, *Sol. Energy* 269 (2024) 112320.
- [24] G. Terrén-Serrano, M. Martínez-Ramón, Processing of global solar irradiance and ground-based infrared sky images for solar nowcasting and intra-hour forecasting applications, *Sol. Energy* 264 (2023) 111968.
- [25] X. Zhang, et al., Solar irradiance prediction interval estimation and deterministic forecasting model using ground-based sky image, in: *2022 IEEE/IAS 58th Industrial and Commercial Power Systems Technical Conference (I&CPS)*, 2022, pp. 1–8.
- [26] L. Cheng, H. Zang, Z. Wei, T. Ding, R. Xu, G. Sun, Short-term solar power prediction learning directly from satellite images with regions of interest, *IEEE Trans. Sustain. Energy* 13 (1) (2021) 629–639.
- [27] W. Kong, Y. Jia, Z.Y. Dong, K. Meng, S. Chai, Hybrid approaches based on deep whole-sky-image learning to photovoltaic generation forecasting, *Appl. Energy* 280 (2020) 115875.
- [28] X. Zhang, F. Fang, J. Wang, Probabilistic solar irradiation forecasting based on variational Bayesian inference with secure federated learning, *IEEE Trans. Ind. Informatics* 17 (11) (2020) 7849–7859.
- [29] T. Niu, J. Li, W. Wei, H. Yue, A hybrid deep learning framework integrating feature selection and transfer learning for multi-step global horizontal irradiation forecasting, *Appl. Energy* 326 (2022) 119964.
- [30] B.K. Puah, et al., A regression unsupervised incremental learning algorithm for solar irradiance prediction, *Renew. Energy* 164 (2021) 908–925.
- [31] H. Liu, Z. Duan, C. Chen, Wind speed big data forecasting using time-variant multi-resolution ensemble model with clustering auto-encoder, *Appl. Energy* 280 (2020) 115975.
- [32] L. Zhang, R. Wilson, M. Sumner, Y. Wu, Advanced multimodal fusion method for very short-term solar irradiance forecasting using sky images and meteorological data: a gate and transformer mechanism approach, *Renew. Energy* 216 (2023) 118952.
- [33] C. Feng, J. Zhang, SolarNet: a sky image-based deep convolutional neural network for intra-hour solar forecasting, *Sol. Energy* 204 (2020) 71–78.
- [34] C. Feng, J. Zhang, W. Zhang, B.-M. Hodge, Convolutional neural networks for intra-hour solar forecasting based on sky image sequences, *Appl. Energy* 310 (2022) 118438.
- [35] Q. Paletta, G. Arbod, J. Lasenby, Benchmarking of deep learning irradiance forecasting models from sky images—An in-depth analysis, *Sol. Energy* 224 (2021) 855–867.
- [36] H. Yang, L. Wang, C. Huang, X. Luo, 3D-CNN-Based sky image feature extraction for short-term global horizontal irradiance forecasting, *Water* 13 (13) (2021) 1773.
- [37] F. Wang, et al., A minutely solar irradiance forecasting method based on real-time sky image-irradiance mapping model, *Energy Convers. Manag.* 220 (2020) 113075.
- [38] Z. Zhen, et al., Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image, *IEEE Trans. Ind. Appl.* 56 (4) (2020) 3385–3396.
- [39] Z. Zhen, et al., Ultra-short-term irradiance forecasting model based on ground-based cloud image and deep learning algorithm, *IET Renew. Power Gener.* 16 (12) (2022) 2604–2616.
- [40] X. Wu, Y. Su, Y. Sun, H. Yang, Z. Zhen, F. Wang, A minutely solar irradiance forecasting method based on multidimensional feature extraction using all-sky images, in: *2023 IEEE/IAS 59th Industrial and Commercial Power Systems Technical Conference (I&CPS)*, 2023, pp. 1–7.
- [41] X. Wu, et al., Multidimensional feature extraction based minutely solar irradiance forecasting method on using all-sky images, *IEEE Trans. Ind. Appl.* 60 (3) (2024) 4494–4504.
- [42] H. Gao, M. Liu, Short-term solar irradiance prediction from sky images with a clear sky model, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2475–2483.
- [43] A.L. Jonathan, D. Cai, C.C. Ukwuoma, N.J.J. Nkou, Q. Huang, O. Bamisile, A radiant shift: attention-embedded CNNs for accurate solar irradiance forecasting and prediction from sky images, *Renew. Energy* (2024) 121133.
- [44] Z. Wang, L. Wang, C. Huang, X. Luo, A hybrid ensemble learning model for short-term solar irradiance forecasting using historical observations and sky images, *IEEE Trans. Ind. Appl.* 59 (2) (2022) 2041–2049.
- [45] S. Chaaraoui, S. Houben, S. Meilinger, Probabilistic end-to-end irradiance forecasting through pre-trained deep learning models using all-sky-images, *Adv. Sci. Res.* 20 (2024) 129–158.
- [46] I.R. Nijhum, P. Kenny, S. Dev, Intra-hour solar irradiance estimation using infrared sky images and MobileNetV2-based CNN regression, in: *2023 IEEE 7th Conference on Energy Internet and Energy System Integration (EI2)*, 2023, pp. 3534–3539.
- [47] M. Sebena, J. Kacur, Prediction of solar energy from sky images and meteorological data using neural networks, in: *2023 International Symposium ELMAR*, 2023, pp. 189–192.
- [48] J. Liu, H. Zang, L. Cheng, T. Ding, Z. Wei, G. Sun, A Transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting, *Appl. Energy* 342 (2023) 121160.
- [49] T.M. Mercier, A. Sabet, T. Rahman, Vision transformer models to measure solar irradiance using sky images in temperate climates, *Appl. Energy* 362 (2024) 122967.
- [50] S. Xu, R. Zhang, H. Ma, C. Ekanayake, Y. Cui, On vision transformer for ultra-short-term forecasting of photovoltaic generation using sky images, *Sol. Energy* 267 (2024) 112203.
- [51] A. Vaswani, et al., Attention is all you need.(nips), *arXiv Prepr. arXiv1706.03762* 10 (2017) S0140525X16001837, 2017.
- [52] H. Zhou, et al., Informer: beyond efficient transformer for long sequence time-series forecasting, *Proc. AAAI Conf. Artif. Intell.* 35 (12) (2021) 11106–11115.
- [53] A. Dosovitskiy, et al., An image is worth 16x16 words: transformers for image recognition at scale, *arXiv Prepr. arXiv2010.11929*, 2020, pp. 1–22.
- [54] T. Stoffel, A. Andreas, Nrel Solar Radiation Research Laboratory (Srrl): Baseline Measurement System (Bms); Golden, Colorado (Data), 1981.
- [55] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, J. Shelby, The national solar radiation data base (NSRDB), *Renew. Sustain. Energy Rev.* 89 (Jun. 2018) 51–60, <https://doi.org/10.1016/j.rser.2018.03.003>.