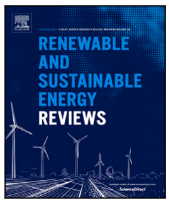
Renewable and Sustainable Energy Reviews 214 (2025) 115510

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser



Original research article

Dual-gate Temporal Fusion Transformer for estimating large-scale land surface solar irradiation

Xuan Liao [a], Man Sing Wong [a,b,*], Rui Zhu [c]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China
[b] Research Institute for Sustainable Urban Development, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China
[c] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Singapore 138632, Republic of Singapore

ARTICLE INFO

Keywords:
Dual-gate Temporal Fusion Transformer
Hourly land surface solar irradiation estimation
Interpretable deep learning network
GeoAI
Geographical heterogeneity

ABSTRACT

An accurate estimation of land surface solar irradiation (LSSI) is crucial to address the solar intermittency for optimizing solar photovoltaic (PV) installation and mitigrating PV curtailment. This involves enhancing solar photovoltaic (PV) system efficiency by optimizing layout and maximizing solar energy capture and conversion. While deep learning methods have significantly improved the rapid and accurate estimation of solar irradiation, they face challenges in handling geographical heterogeneity and providing interpretable results. To address these challenges, this study proposes the Dual-gate Temporal Fusion Transformer (DGTFT), a novel interpretable deep learning network, to improve LSSI estimation. By integrating the Temporal Fusion Transformer with the Dual-gate Gated Residual Network and Dual-gate Multi-head Cross Attention, the optimal network achieved R^2=0.93, MAE=0.022 (kWh/m^2), RMSE=0.038 (kWh/m^2), rRMSE=0.13, and nRMSE=0.048 through ablation experiments. When applied to datasets observed from Australia, China, and Japan, DGTFT outperformed traditional machine learning methods with a minimum R^2 increase of 23.88%, MAE decrease of 43.18%, RMSE decrease of 9.09%, rRMSE decrease of 32.25%, and nRMSE decrease of 62.79%. Furthermore, the interpretability results of the DGTFT model indicate that clear-sky solar irradiation significantly contributed to the model's performance from Australia and Japan; and the maximum temperature and humidity were the largest importance variables in the Chinese dataset. Accurately estimating LSSI, providing interpretable results, and generating continuous solar irradiation maps for large-scale areas, this study aids in quantifying solar potential and offers scientific guidance for the PV industry's development.

1. Introduction

Achieving carbon neutrality goals and advancing the development of renewable energy sources are critical imperatives in contemporary times [1]. Solar energy, as a clean and renewable energy source, offers notable advantages in this regard. Accurately estimating solar irradiation facilitates a comprehensive understanding of its distribution patterns, thereby providing essential insights for optimizing the deployment of photovoltaic (PV) systems. Through the optimization of PV system layouts, maximal utilization of solar resources can be achieved, leading to enhanced system efficiency and capacity. This not only reduces reliance on traditional energy sources but also mitigates carbon emissions, fostering environmental preservation and sustainable development. Hence, accurate estimation of solar irradiation plays a pivotal role in driving the development of renewable energy sources and the realization of carbon neutrality objectives.

Recently, methods for solar irradiation estimation have been developed using empirical methods [2–4], time series statistical methods [5–7], and artificial intelligence (AI) methods [8,9]. They suggested that AI methods can achieve high accuracy and fast computation in solar radiation estimation, and time series AI methods, such as Long Short Term Memory (LSTM), are capable of estimating solar radiation via effectively capturing the time-dependence relationship of solar data.

As deep learning networks have shown competitiveness in constructing dynamic non-linear relationships between multi-factors and the target, deep learning algorithms represent a promising approach to modeling and estimating solar radiation more accurately and comprehensively [10], such as Recurrent Neural Network (RNN), LSTM, and Temporal Convolutional Network (TCN). However, general deep learning methods still encounter difficulties in spatio-temporal series solar radiation estimation. Previous studies indicate that (i) in solar

* Corresponding author at: Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China.
E-mail address: Ls.charles@polyu.edu.hk (M.S. Wong).

https://doi.org/10.1016/j.rser.2025.115510
Received 1 July 2024; Received in revised form 12 February 2025; Accepted 14 February 2025
Available online 26 February 2025

List of Abbreviations

R^2	Coefficient of determination
AdaBoost	Adaptive Boosting
AI	Artificial intelligence
AOT	Aerosol optical thickness
ARIMA	Auto-Regressive Integrated Moving Average
COT	Cloud optical thickness
CSI	Clear-sky solar irradiation
DGMCA	Dual-gate Multi-head Cross Attention
DGRN	Dual-gate Gated Residual Network
DGTFT	Dual-gate Temporal Fusion Transformer
DP	Dew point
GBM	Gradient Boosting Machine
GeoAI	Geospatial Artificial Intelligence
H	Humidity
LSSI	Land surface solar irradiation
LSTM	Long Short Term Memory
MAE	Mean absolute error
MASE	Mean absolute scaled error
MIs	Meteorological indice
MLP	Multi-Layer Perceptron
nRMSE	Normalized Root Mean Square Error
P	Atmospheric pressure
PV	Photovoltaic
RF	Random Forest
RMSE	Mean square error
RNN	Recurrent Neural Network
rRMSE	Relative Root Mean Square Error
SI	Solar irradiation
T	Air temperature
TA	Apparent temperature
TCN	Temporal Convolutional Network
TFT	Temporal Fusion Transformer
TN	Minimum temperature
WS	Wind speed
XGBoost	Extreme Gradient Boosting

radiation estimation using machine learning, feature selection is typically employed to reduce network complexity and redundancy and it is generally conducted before modeling, but this may not ensure optimal parameter selection, potentially overlooking critical features or interactions relevant to solar radiation [11]; (ii) they are also limited in investigating the impact of geographic variability on solar irradiation [12]; and (iii) it is challenging for end-users to understand and trust machine learning methods because of their black-box nature [13].

Furthermore, sophisticated Transformer architectures were used to estimate solar irradiation. One study developed a vision transformer-based machine learning model to measure solar irradiation, which produces highly accurate estimates for both global horizontal irradiance (RMSE = 52 W/m^2) and diffuse irradiance (RMSE = 31 W/m^2) [14]. Most recently, Temporal Fusion Transformer (TFT), which is a type of Transformer employing an LSTM structure and utilizing Multi-head Attention, has been used to predict time-series data, such as solar PV power [15,16], traffic speed [17], building energy consumption [18], and wind speed [19]. These studies indicated that TFT was competitive compared to models such as RNN, LSTM, and TCN. For example, a further study [15] examined the hourly day-ahead solar PV power estimation performance of several models using data from six different facilities located in Germany and Australia, including the Auto-Regressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), Multi-Layer Perceptron (MLP), Extreme Gradient Boosting (XGBoost), and TFT. It found that TFT outperformed the other four models in terms of root mean square error (RMSE), mean absolute error (MAE), mean absolute scaled error (MASE), coefficient of determination (R^2), and quantile loss. It was also noticed that the TFT algorithm can learn long-term and short-term temporal relationships, respectively, and also helps to efficiently build feature representations of static variables, observed and known time-varying inputs, whereas the other four models only learn the temporal features from the dataset and face challenges in geographic heterogeneity. Therefore, TFT is preferable to ARIMA, LSTM, MLP, and XGBoost for estimating spatio-temporal solar data.

Current research indicates that the variability of solar radiation is influenced by various factors, including meteorological time series variables and static spatial variables such as climate categories and geographical locations [12]. One of the mainstream approaches to address this geographic variability is through the utilization of GeoAI, which applies artificial intelligence techniques to geospatial data like solar irradiation. However, the current application of the TFT method for solar radiation estimation exhibits certain limitations. Existing studies employing the TFT algorithm lack sufficient geographic spatial attributes in their static variables, thus hindering the network's capacity to fully capture the geographic spatial characteristics inherent in solar radiation variations. Consequently, if TFT models are to be employed for large-scale continuous solar irradiation estimation, it becomes imperative to bolster their capability to learn static geographic spatial attributes. As TFT has demonstrated the best performance, dependability, and interpretability among the studied deep learning methods, this study will propose a novel deep learning network based on TFT to address the above limitations. Specifically, this study will preserve the overall framework of the TFT to ensure that the optimized network retains interpretability and feature selection capabilities, enhance the GRN network layer of the original model to improve the estimation accuracy, and optimize the attention layer to strengthen the network's ability to learn the spatiotemporal features from the solar dataset, thereby effectively addressing geographical heterogeneity issues. This study aims to introduce an interpretable deep learning network designed to improve land surface solar irradiation (LSSI) estimation using spatio-temporal data. The contributions of this work are threefold. Firstly, a novel spatio-temporal deep learning network is developed, termed DGTFT, for accurate LSSI estimation. DGTFT demonstrates competitiveness with basic TFT methods and other state-of-the-art networks in terms of both estimation accuracy and model interpretability. Secondly, a GeoAI framework is proposed to effectively address geographical heterogeneity challenges. Thirdly, the well-trained network enables transfer learning for solar irradiation estimation across different datasets, facilitating the generation of large-scale continuous solar potential maps.

2. Methodology

2.1. Research framework

Fig. 1 describes the research framework of this study. Firstly, we cleaned the collected multi-source data. Then, the geographical spatio-temporal dataset was constructed in GIS. Next, novel interpretable deep learning networks with improved structures were proposed to improve the estimation capability of spatio-temporal land surface solar irradiation, and the optimal network was determined based on a series of ablation experiments. After that, to evaluate the capabilities of transfer learning and the effectiveness of the proposed networks, the optimal network was trained using the hourly dataset in Australia, and the well-trained network was applied to the hourly dataset in Japan and the daily dataset in China. Additionally, the interpretability of the models applied in three countries was offered. Finally, the annual continuous LSSI in three countries was generated using the proposed network.

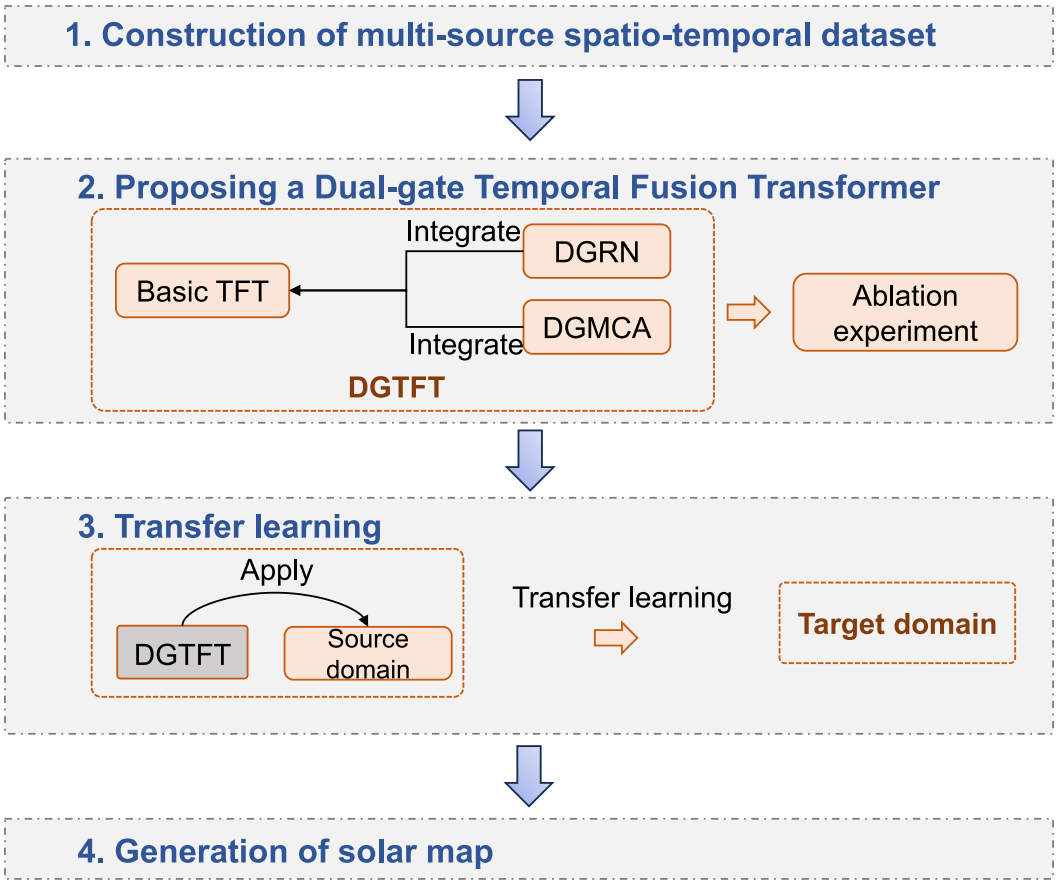


Fig. 1. The framework of this study.

2.2. Study area and data

2.2.1. Data introduction and data preprocessing

Building upon the work of Liao et al. [8], who investigated the influence of various variables (including air temperatures, humidity, wind, atmospheric pressure, aerosol optical thickness (AOT), cloud optical thickness (COT), and clear-sky solar irradiation (CSI)) on LSSI in Australia and China, this study extends the analysis to include five additional stations in Japan. However, it is noteworthy that the data source for Japan differs from that of the other regions. Specifically, the observed hourly LSSI data in Japan is sourced from the Japan Meteorological Agency, which provides direct and diffuse solar irradiation data separately, without global solar irradiation data. Consequently, the sum of direct and diffuse solar irradiation is utilized as a proxy for global solar irradiation in the Japanese dataset. Table 1 lists the specific category, source, and spatial resolution of all data.

The highest temporal resolution that can be freely obtained from Australia, China, and Japan are 10-minute, daily, and hourly, but MIs from these countries are hourly updated. Therefore, to obtain the same resolution for building the deep learning models, all data in each country are aggregated to the same temporal resolution, with the lowest resolution serving as the benchmark, i.e., daily in China and hourly in Australia and Japan. The original AOT and COT data have a temporal resolution of 10 min, and these data and MIs were accumulated daily in China and hourly in Australia and Japan. The original data of observed solar irradiation from Australia was updated every minute, and this study rescaled the temporal resolution to hourly-based updates for the constancy of the dataset in Japan. Furthermore, data imputation was performed on the merged dataset. Specifically, the MissForest method [20], a machine learning-based technique for simulating missing data, was utilized to fill in the gaps in the dataset. The proportion of missing values in the datasets from Australia, China, and Japan was found to be 0.02%, 0.001%, and 0.01%, respectively.

2.2.2. Research area

To assess the transfer learning capability of the proposed DGTFT model, we conduct experiments on three distinct datasets: the hourly dataset from Australia, the daily dataset from China, and the hourly dataset from Japan. The selection of Australia, China, and Japan as the study regions is estimated on the incorporation of Himawari-8 satellite images within our dataset, given its coverage of these three nations. Furthermore, the substantial geographical disparities among these countries serve to enhance the validation of the generalizability of the proposed model. Given that the temporal resolution of LSSI data varies across the three countries, each dataset exhibits different temporal resolutions. These datasets consist of observations from 28 publicly available meteorological stations spanning six consecutive years from 2015 to 2020. Specifically, there are 13 stations in Australia (Fig. 2(d)), 10 stations in China (Fig. 2(b)), and five stations in Japan (Fig. 2(c)). Table 2 provides detailed information on the climates and observed solar irradiation ranges at the 28 meteorological stations.

2.3. Construction of spatio-temporal datasets

2.3.1. Spatial data and temporal data

The data can be categorized into two main types: spatial data and temporal data. As illustrated in Fig. 3, temporal data includes meteorological indices (MIs), solar irradiation measurements from stations, clear-sky solar irradiation (CSI), cloud optical thickness (COT), and aerosol optical thickness (AOT). On the other hand, spatial data comprises geographical coordinates, station names, and climate categories. In this study, each meteorological station is treated as a discrete

Table 1
The category, source, and spatial resolution of all data used in this study.

Data name	Data category	Data source	Resolution
AOT	Temporal	Himawari-8 satellite images [21]	5 km
COT	Temporal	Himawari-8 satellite images	5 km
CSI	Temporal	Calculation values by Pysolar [22]	Discrete-point data, no spatial resolution
MIs	Temporal	Openweather website [23]	Discrete-point data, no spatial resolution
Observed solar irradiation	Target	Meteorological stations [24–26]	Discrete-point data, no spatial resolution
Elevation	Spatial	Meteorological stations	Discrete-point data, no spatial resolution
Latitude	Spatial	Meteorological stations	Discrete-point data, no spatial resolution
Longitude	Spatial	Meteorological stations	Discrete-point data, no spatial resolution
Climate category	Spatial	Meteorological stations	Discrete-point data, no spatial resolution

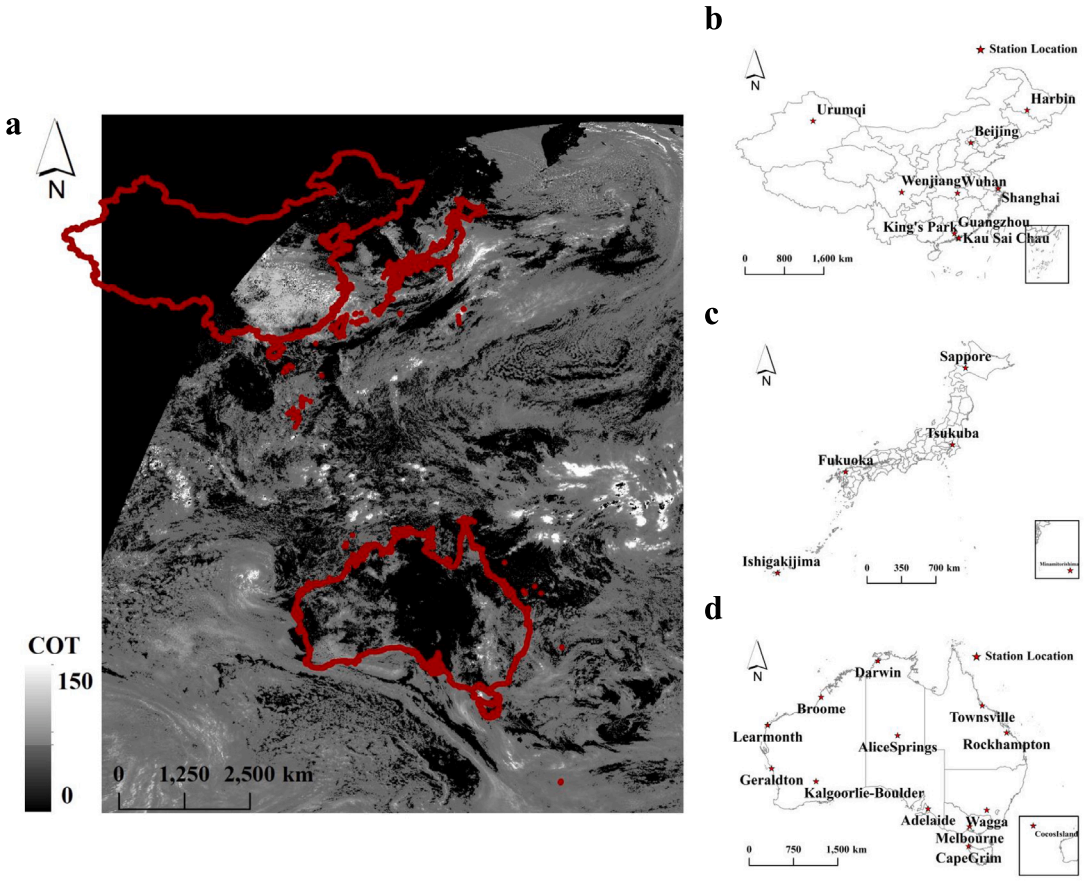


Fig. 2. The distribution of 28 stations in three countries. (a) the geographical positions of Australia, China, and Japan in the Himawari-8 satellite COT image; (b) 10 stations in China; (c) five stations in Japan; (d) 13 stations in Australia.

geographic point. Spatial attributions were then assigned to these geographic points, as depicted in Fig. 3. Specifically, temporal data refers to data that change over time at specific geographic coordinates, encompassing both geographic and temporal information to describe the changes in attributes of a location over different time points. The structure of temporal data is based on the geographic coordinates of the station as the spatio-temporal correlation, including the data variables, i.e., category name, the geographic coordinates, and a time-series of data. As shown in Table 1, temporal data include MIs, observed solar irradiation from stations, CSI, COT, and AOT. Spatial static data refers to time-invariant data related to specific geographic coordinates or space, containing geographic coordinates (e.g., latitude and longitude) and associated attribute information (e.g., elevation and climate category). The structure of such data also uses the geographic coordinates of the station as the spatio-temporal correlation, including the station's geographic coordinates, station name, elevation, and climate category. As shown in Table 2, spatial data include geographic coordinates, elevation, station name, and climate category.

Table 2
Climates and ranges of observed solar irradiation of the 28 meteorological stations.

Country	Station name	Climate	Range of observed solar irradiation (kWh/m^2)
Australia	Adelaide	Mediterranean	0–1.38
	Alice Springs	Subtropical hot desert	0–1.48
	Broome	Hot semi-arid	0–1.44
	Cape Grim	Temperate oceanic	0–1.31
	Cocos Island	Tropical rainforest	0–1.37
	Darwin	Tropical savanna	0–1.45
	Geraldton	Mediterranean	0–1.44
	Kalgoorlie-Boulder	Semi-arid	0–1.39
	Learmonth	Hot semi-arid	0–1.36
	Melbourne	Temperate oceanic	0–1.41
	Rockhampton	Humid subtropical	0–1.51
	Townsville	Tropical savanna	0–1.57
	Wagga	Humid subtropical	0–1.43
China	Beijing	Humid continental	0–9.66
	Guangzhou	Humid subtropical	0.24–7.81
	Harbin	Humid continental	0.13–12.13
	Kau Sai Chau	Humid subtropical	0–1.09
	King's Park	Humid subtropical	0–1.08
	Shanghai	Humid subtropical	0.16–8.65
	Urumqi	Continental cold semi-arid	0–11.75
	Wenjiang	Humid subtropical	0.21–8.39
	Wuhan	Humid subtropical	0.14–8.40
Japan	Fukuoka	Humid subtropical	0.00–1.09
	Ishigakijima	Humid subtropical	0.00–1.14
	Minamitorishima	Tropical savanna	0.00–1.10
	Sapporo	Humid continental	0.00–1.14
	Tsukuba	Humid continental	0.00–1.12

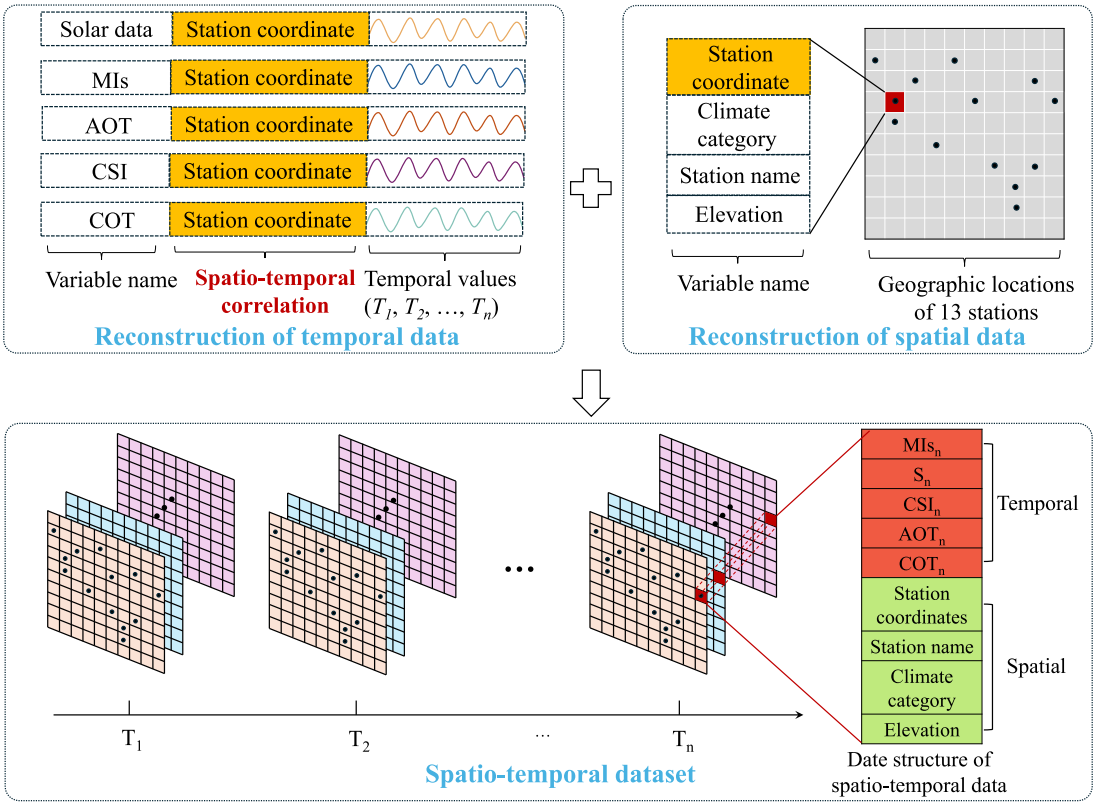


Fig. 3. The process of the GIS representation for constructing the spatio-temporal dataset.

2.3.2. GeoAI dataset

Fig. 3 illustrates the processing for constructing the spatio-temporal dataset. After the reconstruction of spatial and temporal data, the following method was employed for constructing the spatio-temporal dataset: (i) the spatial data and temporal data were merged based on their spatio-temporal correlation, which is the geographic coordinates of the data, (ii) Given that the DGTFT model offers distinct network layers for processing static and time-varying inputs, spatial and temporal input variables were labeled accordingly in the dataset. Therefore, the constructed spatio-temporal dataset comprises geographical coordinates, time information, types of temporal data variables (AOT, COT, CSI, and MIs), corresponding data values, climate category, elevation,

and their respective station names. Additionally, the hourly/daily observed LSSI from the stations was assigned as the training target in this dataset.

This study created three datasets for Australia, China, and Japan, with 1825 samples, 261 samples, and 1825 samples. Each sample from Australia and Japan consists of a nine-hour timestep, while that from China consists of a seven-day timestep. The structure of each hour's data includes geographical coordinates, time information, types of temporal data variables (AOT, COT, CSI, and MIs), corresponding data values, climate category, elevation, and their respective station names. To facilitate model training and evaluation, the entire dataset was divided into three subsets: a training dataset, a validation dataset, and a test dataset, constituting 80%, 10%, and 10% of the samples, respectively.

2.4. Dual-gate temporal fusion transformer

In this study, a novel framework for estimating LSSI is proposed named Dual-gate Temporal Fusion Transformer (DGTFT), which advances the backbone using the TFT [13] module. To greatly forecast time-series solar data, we propose: (i) a novel Dual-gate Gated Residual Network (DGRN) that modifies from the GRN of the original TFT for more accurate estimation performance. (ii) a novel Dual-gate Multi-head Cross Attention (DGMCA) that integrates the interpretable Multi-head Attention that inherits the TFT with Cross Attention [20] for effectively learning the spatio-temporal features from the dataset and greatly integration the static spatial features with the temporal features.

2.4.1. Model overview

As shown in Fig. 4, the proposed DGTFT is composed of a multi-data encoder and a temporal fusion decoder. There are three modules in the multi-data encoder, namely, a static encoder, a past-observed encoder, and a future-known decoder. The input data is classified into three categories (i.e., static metadata, past inputs, and known-future inputs) for feeding into the corresponding layers, and this aims to greatly distinct and extract useful static and temporal features. In the static encoder, the static metadata is first embedded and fed into the variable selection, and then the output is transformed into four static context vectors for integrating with time-varying features. In the past-observed encoder and future-known decoder, the data processing is the same. Specifically, the inputs are also embedded and fed into the variable selection, and then the LSTM module is employed for learning temporal features. The variable selection module and LSTM model inherit from the TFT [19]. After the multi-data encoder, the outputs are fed into the temporal fusion decoder. The temporal fusion decoder is composed of a DGRN, a DGMCA, and position-wise feed-forward layer. The static context vectors are integrated with the outputs of the past-observed encoder and future-known decoder using the DGRN for the static enrichment, respectively, and then the two outputs of the DGRN are concatenated to be fed into the DGMCA for picking up long-range dependencies. Finally, non-linear processing in position-wise feed-forward layer is applied to the outputs of the DGMCA.

2.4.2. Dual-gate gated residual network

GRN plays a crucial role in TFT to flexibly provide non-linear processing, which is applied in data encoding, variable selection, and enhancing the temporal features with static data. Although the simple design of GRN aims to enable the model flexible to give precise insights into the non-linear relationship between inputs and targets, the excessively simple structure of this design may not accurately describe the non-linear relationship. Therefore, we propose a novel Dual-gate Gate Residual Network (DGRN) to improve the non-linear processing ability of GRN. In this section, we detailed the proposed DGRN, which is composed of two branches of non-linear processing. Since the DGRN is applied in different modules for processing the single input X and dual-branch inputs (i.e., X and static context c_s), the DGRN contains two modules to greatly process the inputs, namely, a single-input module and a dual-input module. In the single-input module, to greatly construct the non-linear relationship, X is fed into two branches in parallel and each branch contains one Linear layer and a Tanh activation function. Then, the outputs of the two branches are concatenated to be fed into one Linear layer and the Tanh activation function. To avoid the degradation of the model, the residual connection is conducted, and the output is fed into the gate layer. In the dual-input module, the inputs contain X and static context c_s, X is also fed into two branches for processing one Linear layer and a Tanh activation function. c_s is also fed into two branches for integrating with the features of X. After that, the outputs of both branches are fed into the layers that are the same as those in the single-input module.

2.4.3. Dual-gate multi-head cross attention

In this section, we detail the proposed DGMCA, which is composed of a self attention and a cross attention. The output of a past-observed encoder and the output of a future-known decoder are fed into the DGMCA to learn long-term temporal dependency. To greatly learn the information of past time and the estimation information, we design a dual-gate structure using a self attention and a cross attention. Since a self attention module and a cross attention module are in parallel, the output of a past-observed encoder and the output of a future-known decoder are fed into two modules. Specifically, in the self attention module, the output of a past-observed encoder is first concatenated with the output of a future-known decoder, and then the concatenated output C_{ts} are transformed into the query, the key, and the value for performing the self attention. In the cross attention module, only the output of a future-known decoder serves as the query, and the output of a past-observed encoder are transformed as the key and the value. After this dual-gate attention structure, the output of the self attention module is concatenated with the output of the cross attention module. We detail a self attention and a cross attention next.

The self attention is performed using the concatenated output C_{ts} of a past-observed encoder and a future-known decoder. To enhance the forecasting performance, the query is transformed from intercepted C_{ts} related to the known-future time-series data, and the key and value are transformed from C_{ts}.

Cross attention (CA) is performed between the output of a past-observed encoder E_p and the output of a future-known decoder D_f. Mathematically, CA can be expressed as

$$q = D_f W_q, \quad k = E_p W_k, \quad v = E_p W_v \tag{1}$$

$$A = softmax(qk^T / \sqrt{C/h}) \tag{2}$$

$$CA = Av \tag{3}$$

where W_q, W_k, W_v are learnable parameters, C and h are the embedding dimension and number of heads, and A denotes the attention map. It is noticed that the computation and memory complexity of generating A in cross attention are linear rather than quadratic as in all-attention because we only employ D_f in the query, and it leads to enhanced efficiency of the entire process [27]. Furthermore, as in self attention, a multi-head mechanism is also used in CA.

2.4.4. Implementation details

The TFT model was implemented using Python 3.8 in conjunction with TensorFlow 2.12.0, PyTorch-forecasting 0.10.3, and PyTorch-lightning 1.8.6. Data splitting was conducted using the Python library "TimeSeriesDataset". To prevent overfitting, early stopping techniques were employed. The computations were performed on a high-performance computer featuring an Intel (R) Core (TM) i7-6800K CPU, operating at 3.40 GHz, with 6.0 TB of RAM, and running on the Ubuntu 16.04 LTS system.

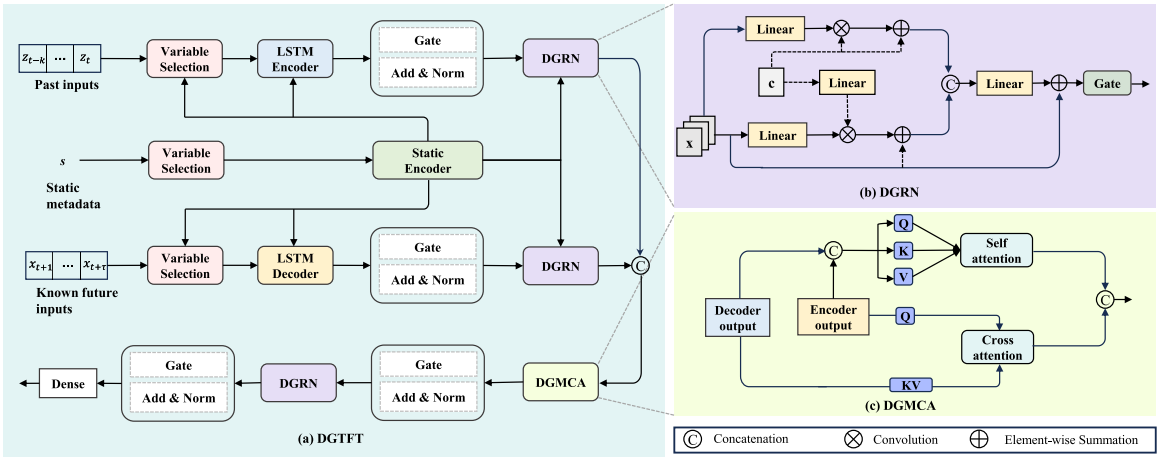


Fig. 4. The proposed network architecture. (a)DGTFT architecture; (b) DGRN architecture; (c) DGMCA architecture. The proposed network employed TFT architecture as the backbone, mainly including Gated Residual Network, Variable Selection Network, Static Covariate Encoders, and Temporal Fusion Decoder. The specific process is shown in (a). The network is advanced by two improved architectures, including DGRN and DGMCA. The aim of dual-gate design is to greatly process two types of features (temporal features and static features) and to increase the accuracy of estimation.

2.4.5. Evaluation metrics

In assessing the proposed network's estimation performance, widely used evaluation metrics were employed, including the coefficient of the determination (R^2), the mean absolute error (MAE), Root Mean Square Error (RMSE), Relative Root Mean Square Error (rRMSE), and the normalized Root Mean Square Error (nRMSE) were adopted, given as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{5}$$

$$nRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{\frac{1}{n}\sum_{i=1}^{n} y_i} \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{7}$$

$$rRMSE = \frac{\sqrt{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{\sum_{i=1}^{n} y_i^2} \tag{8}$$

where $\hat{y}_i$ and y_i are estimated and observed LSSI values, respectively. $\bar{y}_i$ is the average value of observed land surface solar irradiation.

To assess the correlation between observed land surface solar irradiation, MIs, AOT, COT, and CSI, we calculated Pearson correlation coefficient (PCC) between the target variable (LSSI) and variables mentioned in Section 2. PCC is a statistical metric that measures the strength and direction of a linear relationship between two random variables [28]. The Pearson correlation coefficient, which measures the linear relationship between two variables, x and y, is formally defined as the covariance of these two variables divided by the product of their standard deviations (serving as a normalization factor). This coefficient can also be equivalently defined as follows:

$$r_{xy} = \frac{\sum(x_i - \bar{x})\sum(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \tag{9}$$

where $\bar{x}$ denotes the mean of x. $\bar{y}$ denotes the mean of y. The coefficient r_{xy} ranges from −1 to 1 and it is invariant to linear transformations of either variable.

2.5. Generation annual land surface solar irradiation maps

After training the models using datasets in Australia, China, and Japan, these well-trained models were employed to generate annual land surface solar maps at a 5-km spatial resolution in three countries in 2020. The meteorological, COT, AOT, and CSI images are well-prepared and used as the input parameters of the trained model. In addition, a set of meteorological images are obtained by using the Kriging interpolation method.

3. Results and discussion

3.1. Ablation study

To verify the effectiveness of each component in the proposed DGTFT, we conduct ablation studies on the dataset in Australia. We use the TFT as the backbone, and we substitute a novel DGRN and DGMCA for the original GRN and interpretable multi-head attention, respectively. The components being evaluated contain DGRN and DGMCA. We also further conduct ablation studies on the Linear layers in DGRN with different activation functions (i.e., null, Tanh, Sigmoid, and Softmax) to explore the optimal combination of the Linear layer and the activation function. Five indicators are employed to evaluate the performance of different combinations, namely, R^2, MAE, RMSE, rRMSE, and nRMSE. The results are shown in Table 3.

Overall, the combination of "Baseline+DGRN+ DGMCA" shows the best prediction performance based on five indicators, with $R^2 = 0.9260$, MAE = 0.02198 (kWh/m^2), RMSE = 0.03823 (kWh/m^2), rRMSE = 0.1338, and nRMSE = 0.04845, followed by the "Baseline+DGMCA". Although the nRMSE value of this combination is slightly higher than that of the "Baseline+DGMCA", it outperforms other combinations based on the values of R^2 and MAE. Therefore, "Baseline+DGRN+ DGMCA" shows the best performance for predicting the LSSI based on the comprehensive evaluation of these indicators. Furthermore, it outperforms the benchmark "Baseline" by 2%, 13%, 7%, 9%, and 7% for R^2, MAE, RMSE, rRMSE, and nRMSE. These results suggest that the "Baseline+DGRN+ DGMCA" effectively improves the estimation capability for land surface solar irradiation.

3.1.1. Effect of DGMCA

Compared to the benchmark, the "Baseline+DGMCA" increases by 2% for R^2 and decreases by 12%, 3%, 6%, and 8% for MAE, RMSE, rRMSE, and nRMSE. This suggests that the designed DGMCA module is able to learn better long-term temporal dependance and spatial features than the original TFT model.

Table 3
The performance of different components of our model on the test dataset of the dataset in Australia.

Architecture	R^2	MAE (kWh/m^2)	nRMSE	RMSE (kWh/m^2)	rRMSE
Baseline	0.9091	0.02535	0.05253	0.04089	0.1475
Baseline+DGRN	0.9150	0.02411	0.05054	0.04067	0.1409
Baseline+DGMCA	0.9257	0.02226	**0.04828**	0.03971	0.1391
Baseline+DGRN +DGMCA	**0.9260**	**0.02198**	0.04845	**0.03823**	**0.1338**
Baseline+DGRN + DGMCA+Tanh	0.9186	0.02318	0.05053	0.04098	0.1396
Baseline+DGRN + DGMCA+sigmoid	0.9166	0.02270	0.05073	0.04103	0.1388
Baseline+DGRN + DGMCA+softmax	0.9197	0.02326	0.05015	0.03980	0.1409

3.1.2. Effect of DGRN

Compared to the benchmark, the "Baseline+DGRN" increases by 1% for R^2 and decreases by 5%, 3%, 6%, and 4% for MAE RMSE, rRMSE, and nRMSE, which indicates that the proposed DGRN module improve the prediction capability. Furthermore, we give the insights into the effect of the combination of DGRN and DGMCA. The result of the "Baseline+DGRN+ DGMCA" is superior in R^2 and MAE than the "Baseline", "Baseline+DGRN", and "Baseline+DGMCA". Additionally, we also investigate the impact of the commonly used activation functions on the prediction performance, including Tanh, Sigmoid, and Softmax. From the results, although the performance of these three combinations is better than that of the benchmark, their performance is worse than that of the "Baseline+DGRN+ DGMCA". And this suggests that these activation functions are not suitable for adding the DGRN module.

3.2. Evaluation of the performance of DGTFT

3.2.1. The performance of transfer learning

To evaluate the capability of transfer learning of the proposed DGTFT, we employ three datasets in Australia, China, and Japan to calculate the estimation accuracy based on R^2, MAE, RMSE, nRMSE, rRMSE, and execution time. The results are shown in Tables 4, 5, and 6. To greatly illustrate the estimation results, the scatter plots are provided in Figs. 5, 6, and 7. Overall, the performance of the proposed DGTFT is superior to other traditional machine learning methods(i.e., Adaptive Boosting (Adaboost), Gradient Boosting Machine (GBM), Multi-Layer Perceptron (MLP), and Random Forest (RF)) and time series deep learning methods (LSTM, Transformer), which suggests that the DGTFT can provide a high accurate and reliable prediction performance and has the excellent capability of transfer learning. The capability of integrating static spatial data with temporal data of the DGTFT may lead to highly accurate estimation performance. Tradition machine learning methods make it difficult to use static information to enhance the model learning ability. Distinct from the methods [8] that train the individual model for each station, we just train the one model for each dataset. Therefore, we can notice that machine learning methods are limited in processing spatio-temporal data, while the DGTFT shows a good capability to investigate this nonlinear relationship integrated static spatial data with temporal data. Additionally, the estimation performance of time-series deep learning methods outperformed traditional machine learning methods in the datasets of these three countries, but they were surpassed by the DGTFT model. These findings indicate the DGTFT model exhibits advantages in estimating spatio-temporal data. Furthermore, we found that our method and Transformer have longer execution time, while GBM requires the shortest amount of time. Despite the prolonged execution time of our model, its notably high estimation accuracy positions this time within an acceptable range.

Additionally, it is noticed that there are multiple data points with very close predicted values but differing observed values in Figs. 5(a), 6(e), and 7(a). This phenomenon may arise from the poor generalization capability of some machine learning models and overfitting due to constraints imposed by the training data. Specifically, from the results of Figs. 5 and 7, we found that AdaBoost performed the worst, suggesting its struggle to accurately capture complex relationships among spatio-temporal data features in multi-station long time series data, leading to limited generalization in prediction. As for the result of Fig. 6(e), it is highly likely that the poorer performance is due to the smaller sample size of the dataset from China compared to the other two countries.

Table 4
The estimation performance of the dataset from Australia using the DGTFT.

Model	R^2	MAE (kWh/m^2)	RMSE (kWh/m^2)	nRMSE	rRMSE	Execution time (s)
AdaBoost	0.57	0.18	0.22	0.14	0.50	220.83
RF	0.74	0.12	0.17	0.11	0.39	171.93
GBM	0.69	0.14	0.19	0.13	0.43	**46.98**
MLP	0.68	0.14	0.19	0.17	0.44	187.64
LSTM	0.76	0.11	0.17	0.12	0.37	1346.12
Transformer	0.86	0.083	0.13	0.09	0.25	2344.81
Our method	**0.93**	**0.022**	**0.038**	**0.048**	**0.13**	2123.58

Table 5
The estimation performance of the dataset from China using the DGTFT.

Model	R^2	MAE (kWh/m^2)	RMSE (kWh/m^2)	nRMSE	rRMSE	Execution time (s)
AdaBoost	0.35	3.98	5.66	0.18	0.38	22.39
RF	0.46	3.76	4.45	0.19	0.43	21.62
GBM	0.64	2.32	3.88	0.13	0.48	**4.54**
MLP	0.19	4.75	5.02	0.33	0.79	23.15
LSTM	0.68	2.42	3.01	0.14	0.43	320.24
Transformer	0.75	**1.04**	2.85	0.12	0.33	530.45
Our method	**0.88**	1.72	**2.08**	**0.09**	**0.21**	518.68

Furthermore, among the three datasets, the estimation results of the dataset in Australia are better than those of the other two datasets, which indicates that the DGTFT model is slightly more adaptable to the Australian dataset than the other two datasets. it is noticed that the estimation results using our model are far better than other traditional machine learning methods using the dataset in China. The smaller dataset size in China compared to the other datasets may contribute to this discrepancy. Given that the dataset in China has a daily temporal resolution while the others are hourly, it is evident that the dataset size in China is relatively smaller when considering the consistent study time. These findings suggest that traditional machine learning methods may struggle with smaller datasets, whereas DGTFT exhibits robustness regardless of dataset size. Particularly, when faced with a limited number of training samples, DGTFT's estimation accuracy far surpasses that of other machine learning models, underscoring the robustness of the proposed network architecture to smaller sample sizes.

3.2.2. Generation of annual land surface solar irradiation maps

Figs. 8(a), 9(a), and 10(a) describe the distribution of annual LSSI in Australia, China, and Japan, respectively. Overall, the land surface solar irradiation levels across Australia predominantly reside within

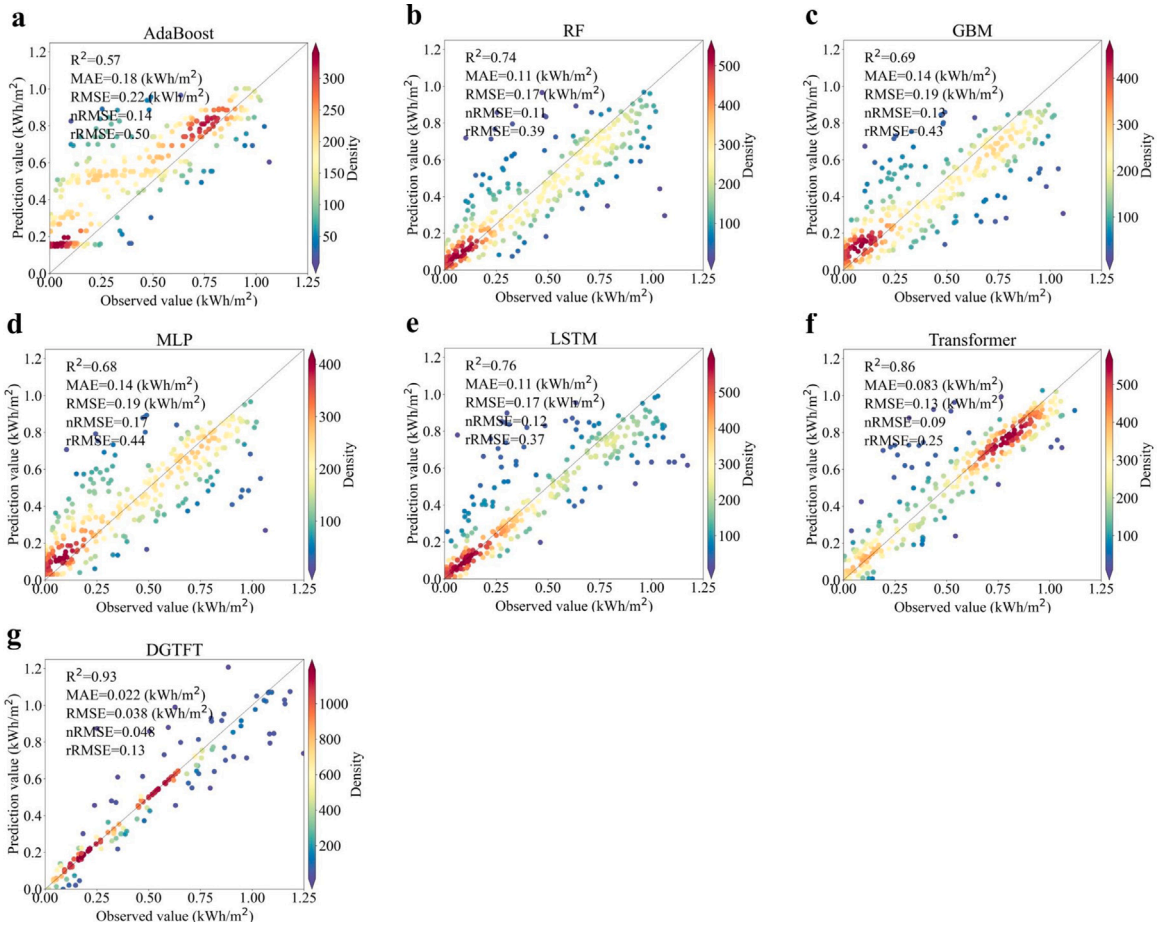


Fig. 5. Distribution of annual land surface solar irradiation in Australia.

Table 6
The estimation performance of the dataset from Japan using the DGTFT.

Model	R^2	MAE (kWh/m^2)	RMSE (kWh/m^2)	nRMSE	rRMSE	Execution time (s)
AdaBoost	0.54	0.17	0.28	0.21	0.58	224.74
RF	0.66	0.15	0.17	0.18	0.49	113.60
GBM	0.64	0.15	0.17	0.19	0.51	**67.64**
MLP	0.63	0.16	0.18	0.19	0.52	124.37
LSTM	0.74	0.11	0.17	0.12	0.38	1419.83
Transformer	0.81	0.11	0.15	0.10	0.31	2156.32
Our method	**0.83**	**0.077**	**0.12**	**0.087**	**0.21**	2315.33

higher ranges, contrasting with Japan where they mostly fall within lower ranges, with China positioned intermediate to the two. This underscores Australia's abundant solar energy resources. Specifically, in Australia, land surface solar irradiation levels are generally high, except for a small portion near the southern coastal areas where values are relatively lower. Across China, land surface solar irradiation diminishes gradually from the northwest to the southeast. Conversely, in Japan, solar irradiation levels predominantly register within lower ranges, with sporadic higher values scattered across its southern and central regions. Furthermore, based on the respective areas of each region, the share of solar resources for each region was analyzed, as depicted in Figs. 8(b), 9(b), and 10(b). Notably, in Australia, the area located near the equator, had the highest mean annual solar irradiation of 2023.94 kWh/m^2, representing 16.46% of the nation's solar irradiation resources. In contrast, Tasmania had the lowest mean annual solar irradiation resources, with 1250.90 kWh/m^2, representing 0.95% of the national solar irradiation resources. In China, Xinjiang, Inner Mongolia, and Tibet provinces possess the highest solar energy resources, accounting for 16.27%, 14.58%, and 11.89% of the national total, respectively. Across the prefectures of Japan, there is minimal variation in the proportion of solar energy contribution. These results indicate that the northern regions of Australia and the northwest regions of China possess a high potential for solar energy development. Policymakers can utilize this information to design targeted solar energy harvesting development.

To evaluate the estimation accuracy of the generated maps, we calculated the annual cumulative absolute errors between estimated values and measured values in 28 stations in three countries. Fig. 11 shows the results. Overall, the annual cumulative absolute error values across these 28 stations are relatively small, with 92.86% of stations exhibiting annual cumulative error values below 400 (kWh/m^2). Specifically, the annual cumulative error values at Australian sites are slightly lower compared to those in China and Japan. These findings suggest the high precision of our trained model in generating large-scale solar irradiation maps, thus affirming the strong generalization capability and broad applicability of the proposed neural network model. Furthermore, we compared our generated maps with the published maps by Solargis [29]. Solargis develops proprietary algorithms and provides high-quality solar data and energy evaluation software based on satellite image processing and atmospheric and meteorological models. The model of Solargis [30] has been validated for 189 sites worldwide, and the results show that the Standard deviation of Global Horizontal Irradiation is ±3.0%, suggesting its high accuracy. The products of Solargis have been applied in the fields of site selection, energy yield simulation, optimization of power plant design, evaluation of power plant performance, forcasting, and ground data verification [31]. Therefore, the map from Solargis is the reliable benchmark. Figs. 12(a),

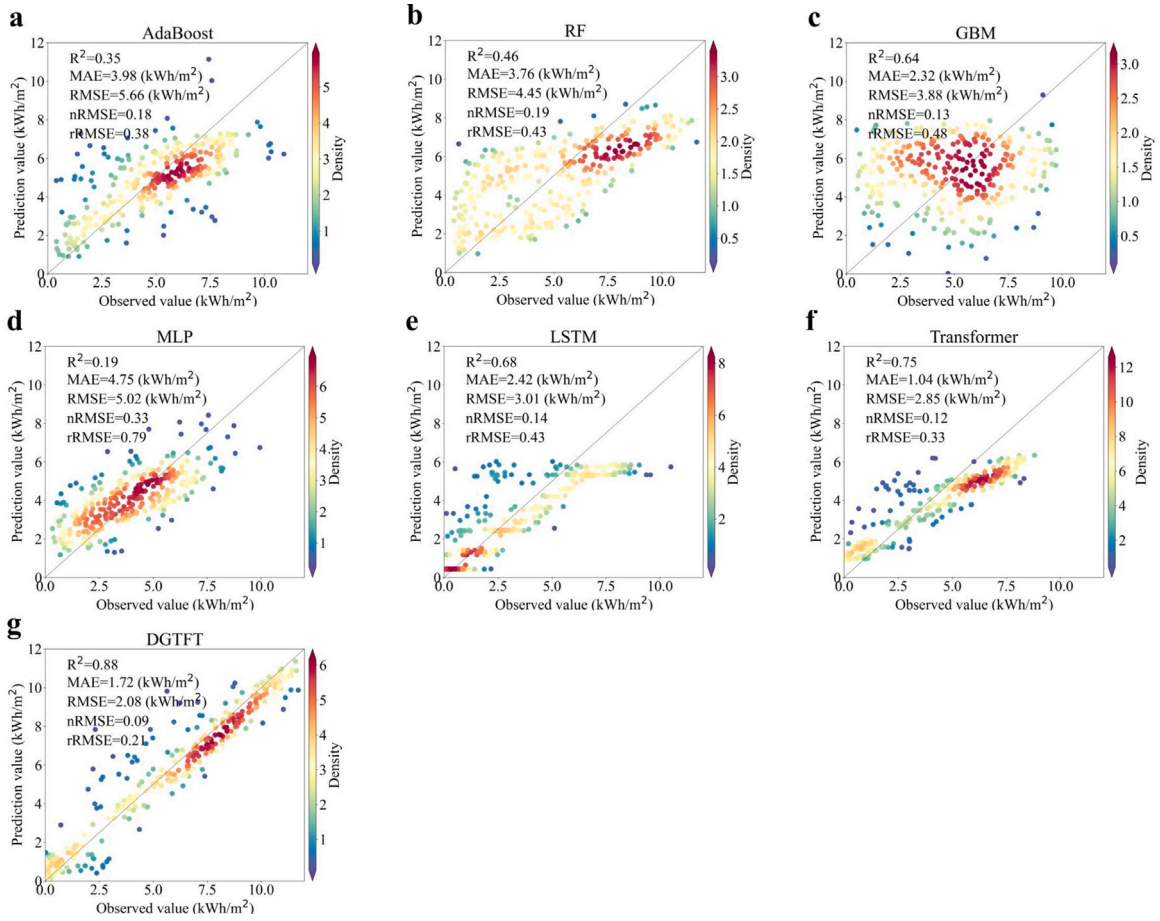


Fig. 6. Distribution of annual land surface solar irradiation in China.

13(a), and 14(a) show the spatial distribution of the annual land surface solar irradiation map from Solargis in Australia, China, and Japan, respectively. A comparison of the quantitative ranges and distribution patterns shows that our results of three countries are consistent with the published maps. Additionally, we calculated the absolute error maps between the generated maps using our method and the maps from Solargis in Australia, China, and Japan, respectively. As shown in Figs. 12(b), 13(b), and 14(b), the majority of error values in Australia and Japan fall within the 0–100 range, and those in China are mostly under 300 (kWh/m^2). Overall, the error values are relatively small, further demonstrating the high accuracy of our estimated method.

3.3. Interpretability of DGTFT

3.3.1. Importance of input variables

The DGTFT enables its network structure interpretive by quantifying the importance of variables in the different layers, including past-observed encoder, future-known decoder, and static encoder. Figs. 15, 16, and 17 show the importance of variables in the past-observed encoder, future-known decoder, and static encoder of the models trained by datasets in Australia, China, and Japan. In the Decoder network layer, for models trained on the Australian and Japanese datasets, CSI emerged as the variable contributing the most to network training, with importance indices of approximately 40% and 50%, respectively. Conversely, for the model trained on the Chinese dataset, the variables of highest importance were the maximum temperature and humidity, with importance indices exceeding 20%. In the Encoder network layer, solar irradiation emerged as the variable contributing the most to network training for models trained on the Australian and Japanese datasets, with importance indices of approximately 85% and 40%, respectively. However, for the model trained on the Chinese dataset, CSI was the most important variable, with an importance index of approximately 24%. In the Static network layer, it incorporates four types of static inputs: target center, target scale, the identification of the facility providing additional information and context to the model, and spatial variables mentioned in Section 2. In this case, the first two, $solar_{center}$, and $solar_{scale}$ are related to the standardization of the land surface solar irradiation, both serving as static variables in the model. $solar_{center}$ has a value of 0, and $solar_{scale}$ is the median of the time series. For the model trained on the Australian dataset, the most important variable was $solar_{scale}$, with an importance index of approximately 33%; for the model trained on the Chinese dataset, the most important variable was $Station_{ID}$, with an importance index of approximately 86%; and for the model trained on the Japanese dataset, the most important variables were Longitude and $solar_{center}$, with importance indices of approximately 23%. These findings elucidate the varying contributions of different variables across different network layers during model training, thereby enhancing the interpretability of deep learning networks.

We also conducted the correlation analysis among the time-varying variables using datasets in Australia, China, and Japan. Fig. 18 describes correlation heatmaps. Overall, CSI has the strongest positive correlation with the observed solar irradiation in all datasets, followed by the variables related to the temperature. Furthermore, the humidity shows the strongest negative correlation with the observed solar irradiation in datasets in Australia and Japan, whereas the air pressure shows the strongest negative correlation with the observed solar irradiation in the dataset in China. Combining these results with the importance results in Figs. 15, 16, and 17, it is noted that both results are consistent, which suggests that the interpretability of the

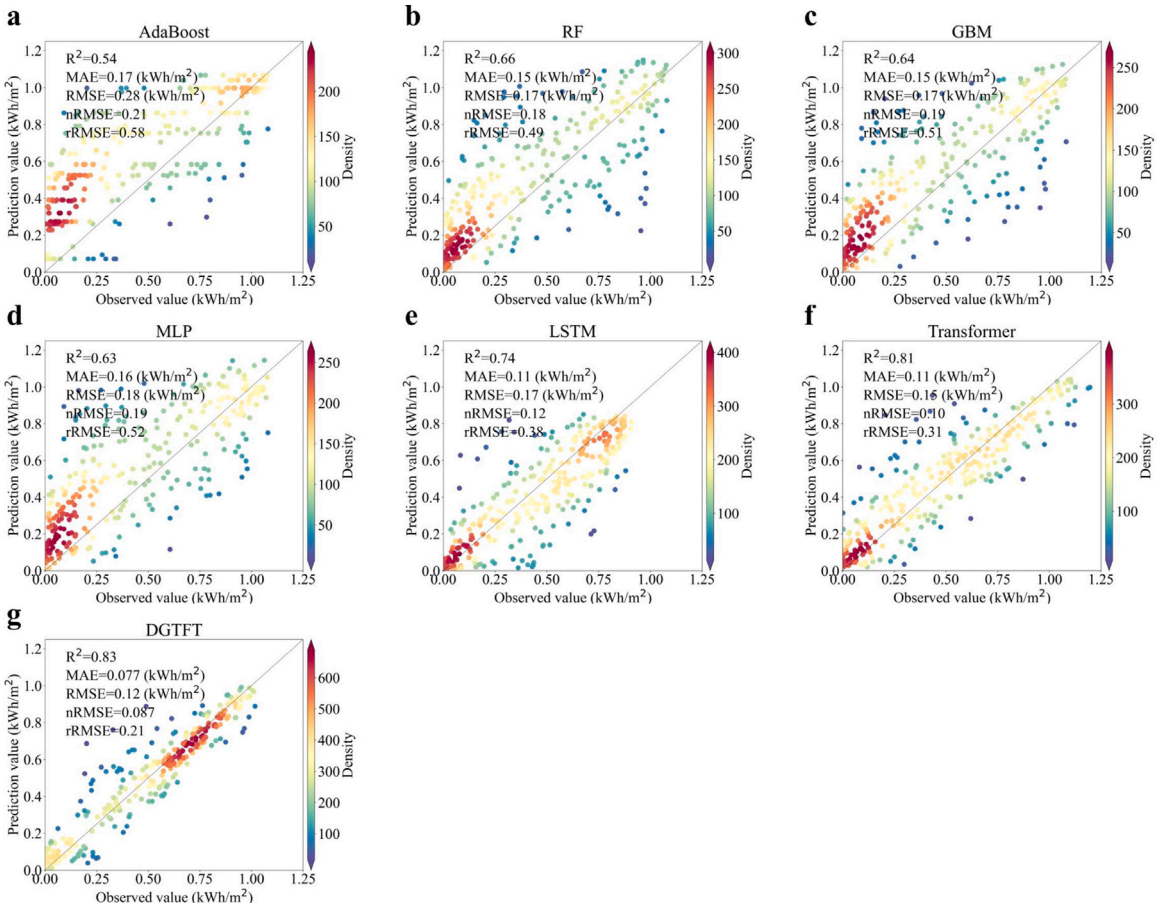


Fig. 7. Distribution of annual land surface solar irradiation in Japan.

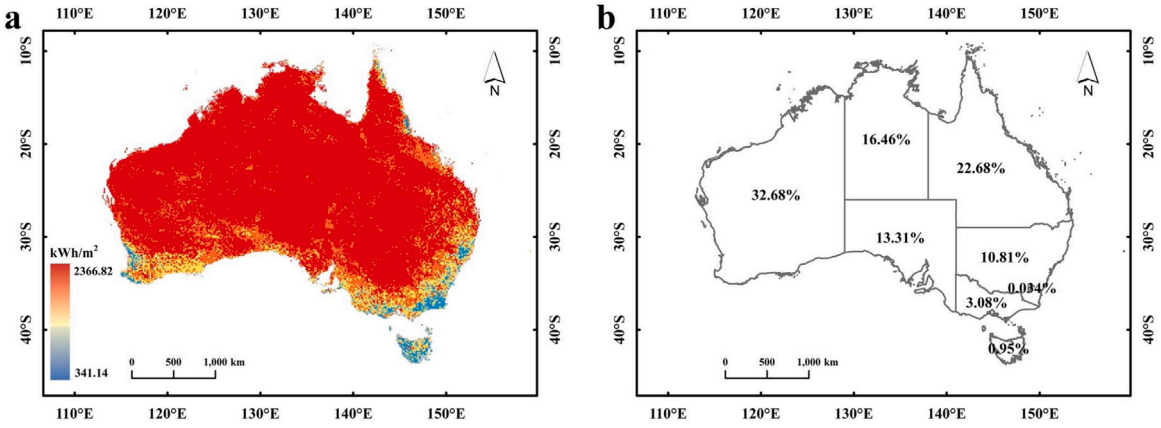


Fig. 8. Annual solar radiation resources in 2020 in Australia. (a) Spatial distribution of annual total land surface solar irradiation across Australia. (b) Share of solar irradiation resources in states.

DGTFT by providing the variable's importance in different layers is reliable.

3.3.2. *Hidden feature analysis*

The DGTFT model not only shows good interpretability by providing the importance of variables in the different layers but also reflects the interpretability through attention values. It is possible to notice what DGTFT has learned by showing the attention values for each time step.

Three typical attention value trend and their corresponding samples of the dataset in each country are presented in Figs. 19, 20, and 21. The observation period of each sample for datasets in Australia and Japan is 36 h(four days × nine hours), which consists of four-day data, and that in China is 28 days, which consists of four-week data(four weeks × seven days). Overall, the trend of Attention weight of the three samples in each country is similar. In Australia, the attention weight gradually increases over time, reaching its peak at around the 18th hour, which is the end of the second day. Subsequently, it declines and reaches its lowest point at the 25th hour before slowly rising again. In China, the attention weight exhibits cyclical variations, with approximately a two-day cycle. In Japan, the attention weight reaches its maximum at the beginning and then immediately drops to a minimum close to zero, maintaining this low level until about the 20th hour. Afterward, slight fluctuations appear, but it remains at the lowest level.

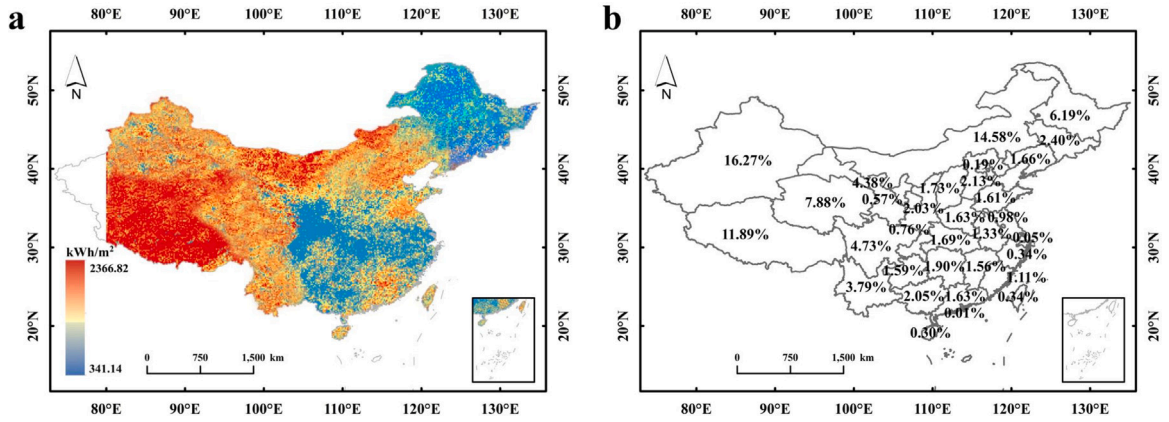


Fig. 9. Annual solar radiation resources in 2020 in China. (a) Spatial distribution of annual total land surface solar irradiation across China.(b) Share of solar irradiation resources in provinces.

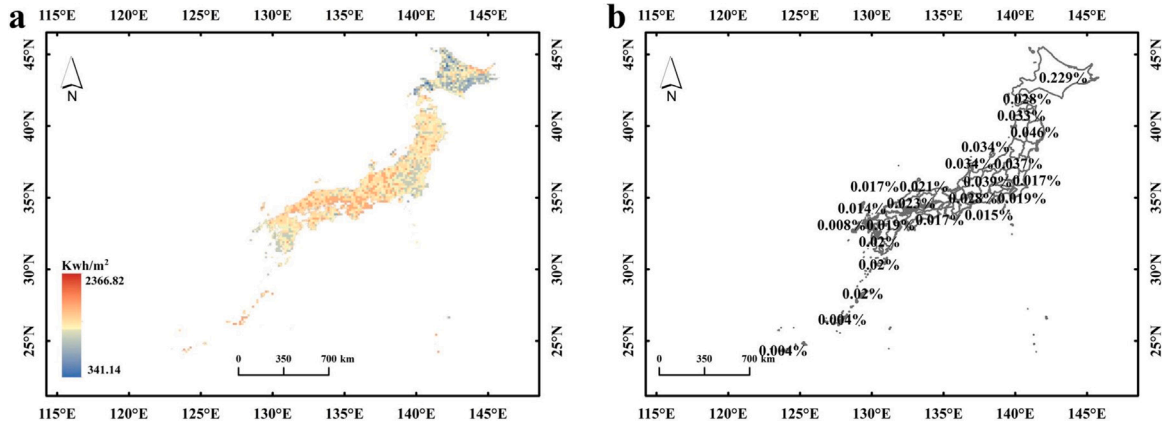


Fig. 10. Annual solar radiation resources in 2020 in Japan. (a) Spatial distribution of annual total land surface solar irradiation across Japan. (b) Share of solar irradiation resources in prefectures.

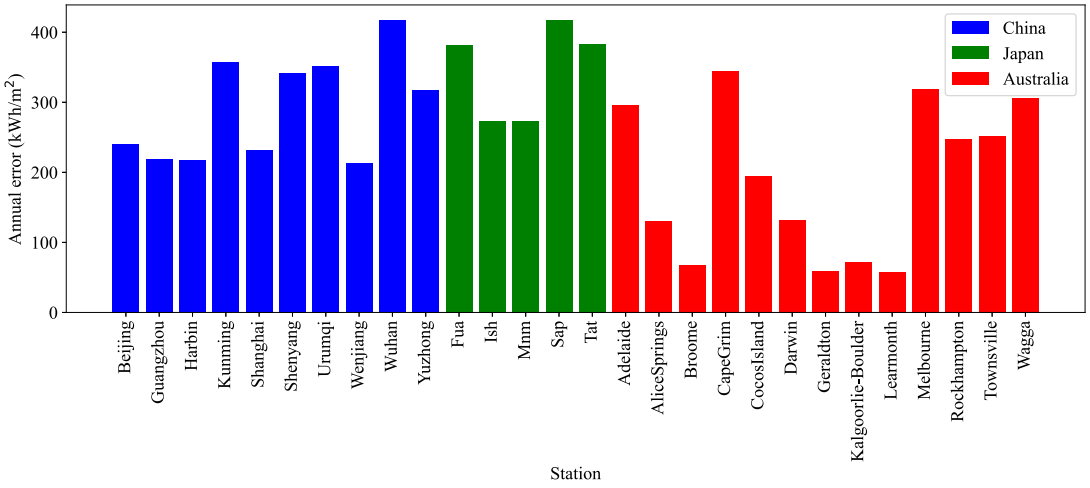


Fig. 11. Annual absolute errors between estimated values and measured values of 28 stations in Australia, China, and Japan.

4. Disscusion and conclusion

This study proposes the state-of-the-art deep learning model DGTFT to provide the interpretive and high-accuracy method for estimating LSSI. With a series of well-designed ablation experiments, the optimal network is obtained that exceeds TFT in terms of R^2, MAE, RMSE, rRMSE, and nRMSE by 2%, 13%, 7%, 9%, and 7%, respectively. The developed network remains competitive compared to the traditional machine learning models(i.e., Adaboost, GBM, MLP, and RF) and time series deep learning methods (LSTM and Transformer) using datasets in Australia, China, and Japan. These improvements mainly come from the capability of the proposed network for extracting and

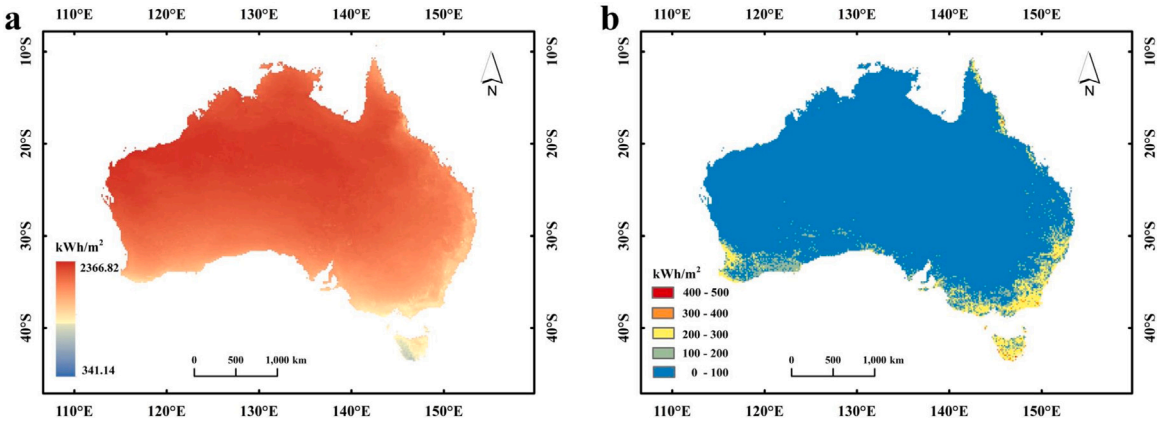


Fig. 12. Accuracy evaluation between the generated annual land surface solar irradiation map using our method and the annual map from Solargis in Australia. (a) annual land surface solar irradiation map from Solargis. (b) absolute error map between the generated map using our method and the map from Solargis.

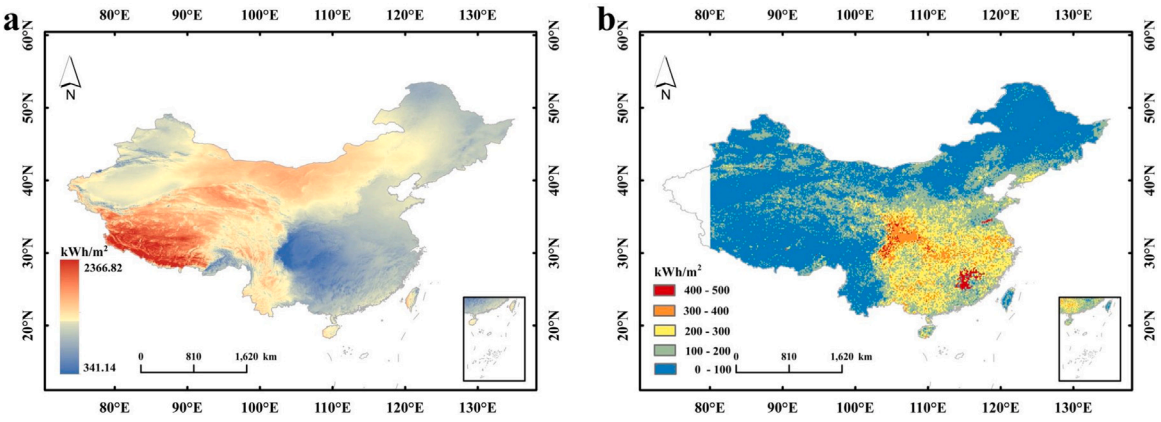


Fig. 13. Accuracy evaluation between the generated annual land surface solar irradiation map using our method and the annual map from Solargis in China. (a) annual land surface solar irradiation map from Solargis. (b) absolute error map between the generated map using our method and the map from Solargis.

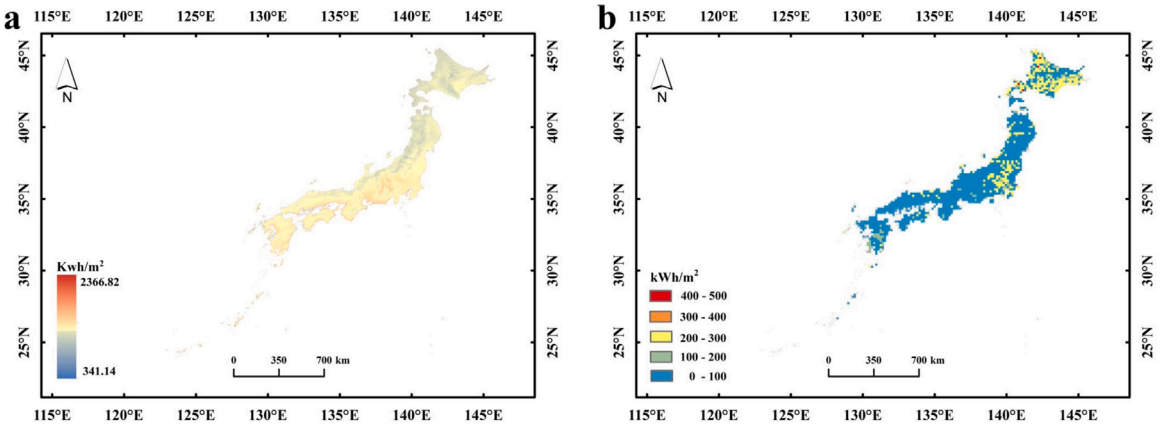


Fig. 14. Accuracy evaluation between the generated annual land surface solar irradiation map using our method and the annual map from Solargis in Japan. (a) annual land surface solar irradiation map from Solargis. (b) absolute error map between the generated map using our method and the map from Solargis.

learning spatio-temporal features from geo-datasets, which illustrates a significant contribution to accurately estimating LSSI at large-scale regions.

Four important findings are revealed from this study. First, the improved network based on the Dual-gate mechanism can effectively increase the estimation accuracy of land surface solar irradiation. Specifically, the estimation performances were all improved using "baseline+DGRN", "baseline+DGMCA", and "baseline+ DGRN + DGMCA", respectively. DGRN network employed the Dual-gate mechanism to deepen the data processing in the network, and DGMAC integrated the cross Attention with self Attention to increase the channels for extracting and learning the spatial features and temporal features from the geo-dataset. Second, the network demonstrates strong transferability. In this study, datasets from three different countries were utilized to evaluate the performance of the proposed network in

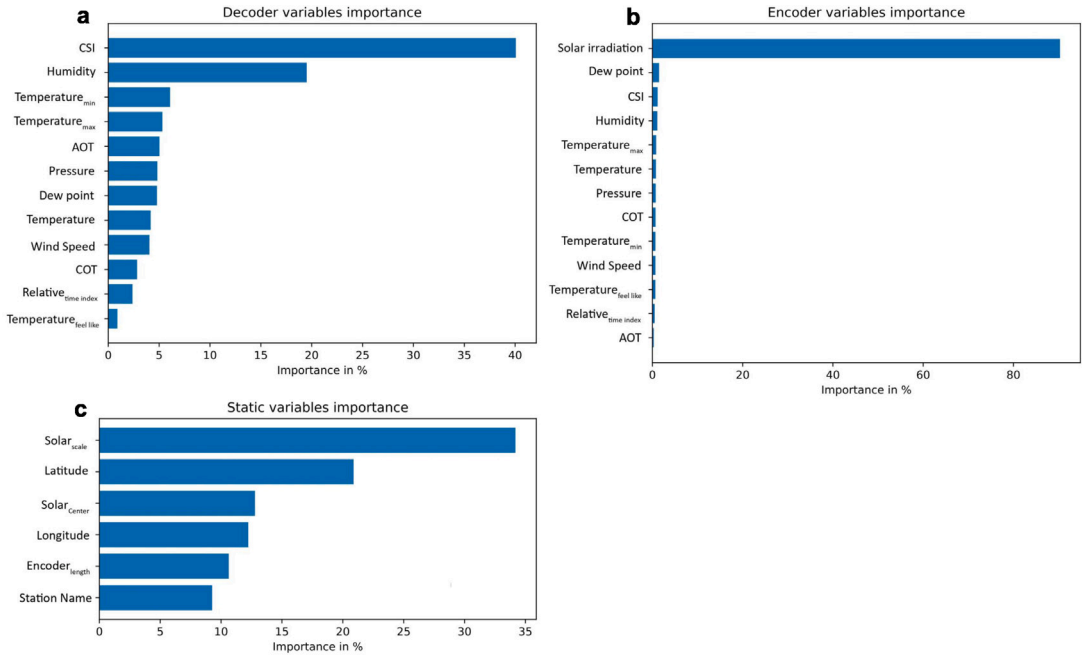


Fig. 15. The importance of variables in the past-observed encoder, future-known decoder, and static encoder in Australia.

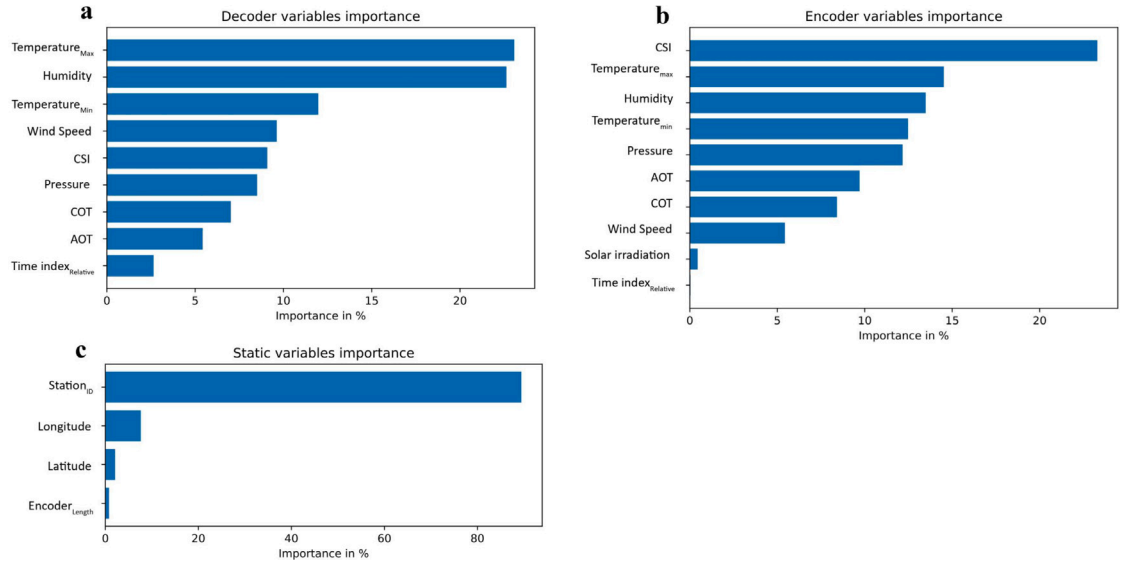


Fig. 16. The importance of variables in the past-observed encoder, future-known decoder, and static encoder in China.

estimation. Observed LSSI in these datasets was sourced from local meteorological stations in each respective country, with varying temporal resolutions. Despite these disparities among the datasets, the proposed network exhibited remarkably high estimation accuracy across all three datasets, significantly surpassing other machine learning models. Third, based on the importance values of various variables computed by the network, it is evident that CSI and variables related to temperature play crucial roles in estimating solar irradiation. This provides scientific guidance for studying factors influencing solar irradiation. Finally, the generated maps using our method have been compared with Solargis maps and ground station measurements, as shown in Figs. 11, 12, 13, and 14. The results indicate that our method exhibits high accuracy and reliability. In terms of estimated accuracy, the solar irradiation distribution in three countries is fairly consistent with the Solargis maps, although China shows slightly higher errors compared to Australia and Japan, possibly due to the fewer training samples available in China. Despite Solargis' capability to provide high-accuracy solar irradiation distribution maps, our proposed method holds a competitive edge. The merits of our method lie in its high data accessibility and strong transferability, enabling the cost-effective acquisition of high-temporal-resolution solar irradiation maps.

This study is significant in three aspects. First, this study is innovative in accurately estimating different temporal resolution LSSI across different countries using the proposed network. The significant improvement in R^2, nRMSE, RMSE, rRMSE, and MAE in three datasets suggests the new network has the strong capability of transfer learning. Second, this study is vital for providing a reliable and effective GeoAI framework to handle the issue of geographical heterogeneity. Specifically, by constructing spatio-temporal datasets to annotate static geographic spatial attributes and temporal variables and employing a dual-gate mechanism in the novel network to enhance learning of both static and temporal features, the well-trained model can effectively

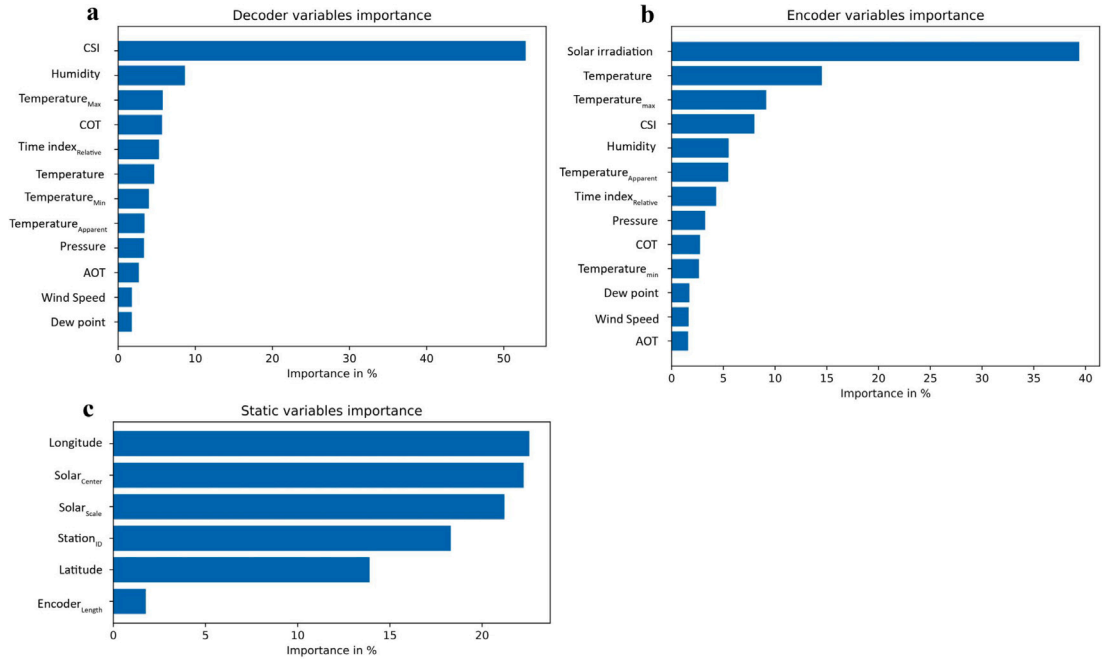


Fig. 17. The importance of variables in the past-observed encoder, future-known decoder, and static encoder in Japan.

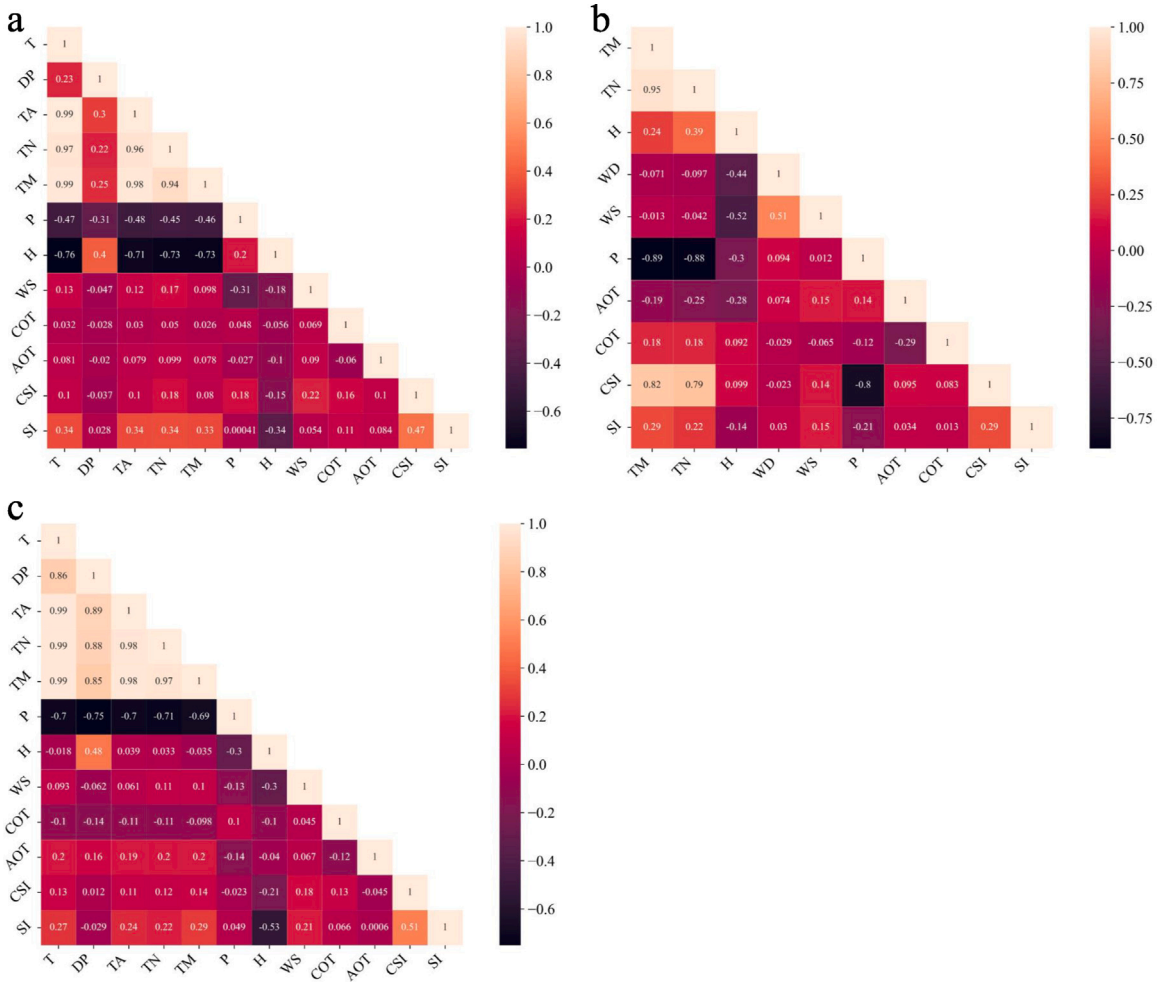


Fig. 18. The correlation heatmap among used variables in three countries. (a)the dataset in Australia; (b) the dataset in China; (c) the dataset in Japan.

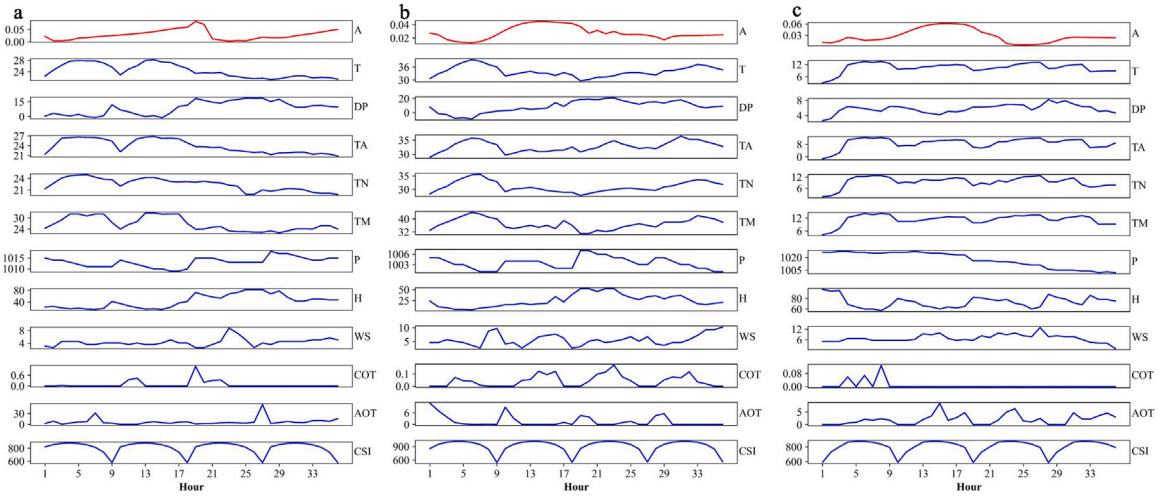


Fig. 19. Three selected typical attention values of DGTFT using the dataset in Australia.(A: attention weight).

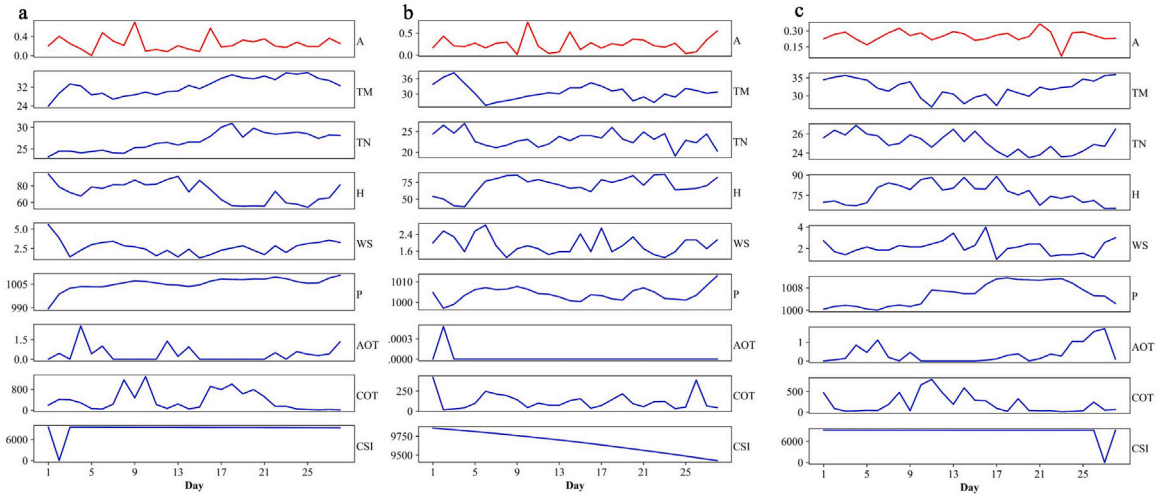


Fig. 20. Three selected typical attention values of DGTFT using the dataset in China.(A: attention weight).

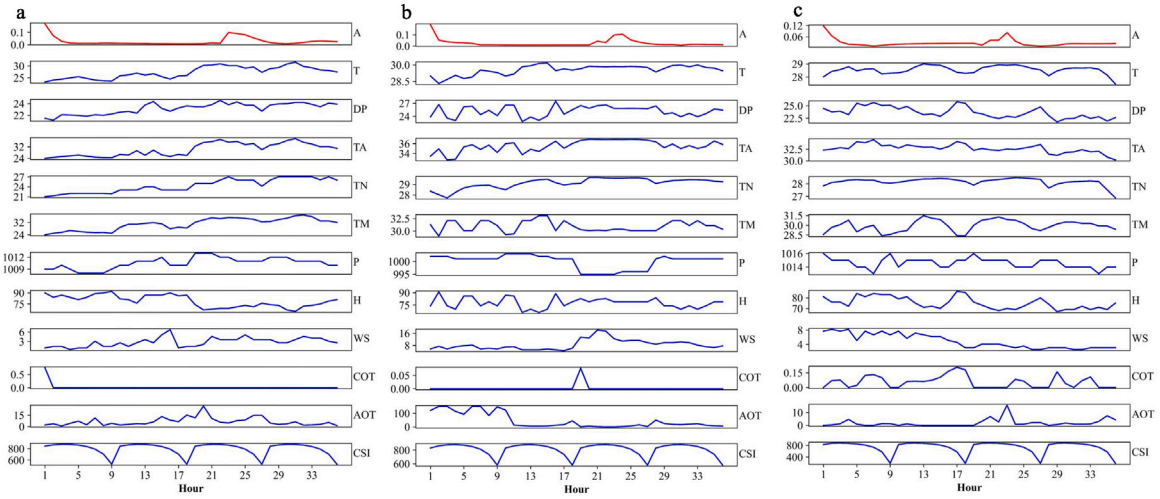


Fig. 21. Three selected typical attention values of DGTFT using the dataset in Japan.(A: attention weight).

capture geographic spatial relationships and temporal dependencies within the dataset, thus addressing geographic heterogeneity issues. Third, this study holds significant implications for the interpretability of deep learning. The proposed network not only computes the important contributions of various variables to different layers of the network but also calculates the attention of the network at different time steps. This aids users in gaining a concrete understanding of the model's decision-making process, thereby enhancing the model's credibility.

Although our proposed DGTFT model performs well in estimating spatio-temporal solar radiation data, there is still room for improvement in the algorithm. This algorithm requires the construction of a GeoAI dataset from multiple sources, including remote sensing imagery, station data, and geographical information data. Additionally, it necessitates the annotation of whether variables in the dataset are temporal or static, a process that is rather complex and increases the difficulty of algorithm generalization. In future algorithm enhancements, the input structure will be optimized to accommodate multiple data formats.

In conclusion, the DGTFT model proposed in this study can provide high-accuracy, interpretive, and reliable estimation results for land surface solar irradiation, which can provide a reliable reference for the design of solar power generation systems. The new network contributes to studies related to solar potential estimation, which is also deliverable for other spatio-temporal data estimation.

CRediT authorship contribution statement

Xuan Liao: Conceptualization, Methodology, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing. **Man Sing Wong:** Funding acquisition, Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Rui Zhu:** Methodology, Visualization, Writing – original draft, Writing – review & editing.

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.

Acknowledgments

This work is supported by the funding support from the General Research Fund (Grant No. 15602619 and 15603920), and the Collaborative Research Fund (Grant No. C5062-21GF) and the Young Colloboartive Research Fund (C6003-22Y) from the Research Grants Council, Hong Kong, China; and the funding support from the Research Institute for Sustainable Urban Development, The Hong Kong Polytechnic University, Hong Kong, China (Grant No. 1-CD81). All authors acknowledge the advices provided by Dr Wang Zhe from Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.

Data availability

Data will be made available on request.

References

[1] Wang F, Harindintwali JD, Yuan Z, Wang M, Wang F, Li S, , Chen JM. Technologies and perspectives for achieving carbon neutrality. Innov 2021;2:4.
[2] Hargreaves GH, Samani ZA. Estimating potential evapotranspiration. J Irrig Drain Div 1982;108:225–30.
[3] Allen RG. Self-calibrating method for estimating solar radiation from air temperature. J Hydrol Eng 1997;2:56–67.
[4] Bristow KL, Campbell GS. On the relationship between incoming solar radiation and daily maximum and minimum temperature. Agric For Meteorol 1984;31:159–66.
[5] Shadab A, Ahmad S, Said SE. Spatial forecasting of solar radiation using ARIMA model. Remote Sens Appl: Soc 2020;20:100427.
[6] Ji W, Chee KC. Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. Sol Energy 2011;85:808–17.
[7] Silva VLGd, Oliveira Filho D, Carlo JC, Vaz PN. An approach to solar radiation prediction using ARX and ARMAX models. Front Energy Res 2022;10:822555.
[8] Liao X, Zhu R, Wong MS. Simplified estimation modeling of land surface solar irradiation: A comparative study in Australia and China. Sustain Energy Technol Assess 2022;52:102323.
[9] Gürel AE, Ağbulut Ü, Bakır H, Ergün A, Yıldız G. A state of art review on estimation of solar radiation with various models. Heliyon 2023;9:2.
[10] Ajith M, Martínez-Ramón MJR. Deep learning algorithms for very short term solar irradiance forecasting: A survey. Renew Sustain Energy Rev 2023;182:113362.
[11] Liao X, Zhu R, Wong MS, Heo J, Chan P, Kwok CYT. Fast and accurate estimation of solar irradiation on building rooftops in Hong Kong: A machine learning-based parameterization approach. Renew Energy 2023;216:119034.
[12] Zang H, Liu L, Sun L, Cheng L, Wei Z, Sun G. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. Renew Energy 2020;160:26–41.
[13] Lim B, Arık SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. Int J Forecast 2021;37:1748–64.
[14] Mercier TM, Sabet A, Rahman T. Vision transformer models to measure solar irradiance using sky images in temperate climates. Appl Energy 2024;362:122967.
[15] López Santos M, García-Santiago X, Echevarría Camarero F, Blázquez Gil G, Carrasco Ortega P. Application of temporal fusion transformer for day-ahead PV power forecasting. Energies 2022;15:5232.
[16] Mazen FMA, Shaker Y, Abul Seoud RA. Forecasting of solar power using GRU–temporal fusion transformer model and DILATE loss function. Energies 2023;16:8105.
[17] Zhang H, Zou Y, Yang X, Yang H. A temporal fusion transformer for short-term freeway traffic speed multistep prediction. Neurocomputing 2022;500:329–40.
[18] Zheng P, Zhou H, Liu J, Nakanishi Y. Interpretable building energy consumption forecasting using spectral clustering algorithm and temporal fusion transformers architecture. Appl Energy 2023;349:121607.
[19] Wu B, Wang L, Zeng YR. Interpretable wind speed prediction with multivariate time series and temporal fusion transformers. Energy 2022;252:123990.
[20] Arriagada P, Karelovic B, Link O. Automatic gap-filling of daily streamflow time series in data-scarce regions using a machine learning algorithm. J Hydrol 2021;598:126454.
[21] JAXA Himawari monitor P-tree system. 2024, https://www.eorc.jaxa.jp/ptree/ [Accessed 14 November 2024].
[22] Pysolar. 2021, https://pysolar.readthedocs.io/en/latest/ [Accessed 14 November 2024].
[23] OpenWeahter. 2024, https://openweathermap.org/ [Accessed 14 November 2024].
[24] Australian government bureau of meteorology. 2024, http://reg.bom.gov.au/index.php/ [Accessed 14 November 2024].
[25] China national meteorological information center. 2024, http://data.cma.cn/ [Accessed 14 November 2024].
[26] Japan meteorological agency. 2024, https://www.jma.go.jp/jma/indexe.html [Accessed 14 November 2024].
[27] Chen CFR, Fan Q, Panda R. Crossvit: Cross-attention multi-scale vision transformer for image classification. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021, p. 357–66.
[28] Rodgers JL. Thirteen ways to look at the correlation coefficient data. Amer Statist 1988;41:59.
[29] Solargis map. 2024, https://solargis.com/maps-and-gis-data [Accessed 14 November 2024].
[30] SolarGIS solution. 2024, https://solargis.com/solutions [Accessed 14 November 2024].
[31] Perez R, Cebecauer T, Suri M. Solar energy forecasting and resource assessment. Academic Press; 2013.