

A temporal fusion transformer augmented GeoAI framework for estimating hourly land surface solar irradiation

Xuan Liao ^a , Man Sing Wong ^{a,b,c,*} , Rui Zhu ^d , Zhe Wang ^e

^a Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, PR China

^b Research Institute for Sustainable Urban Development, The Hong Kong Polytechnic University, Hong Kong, PR China

^c Research Institute for Land and Space, The Hong Kong Polytechnic University, Hong Kong, PR China

^d Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Singapore 138632, Republic of Singapore

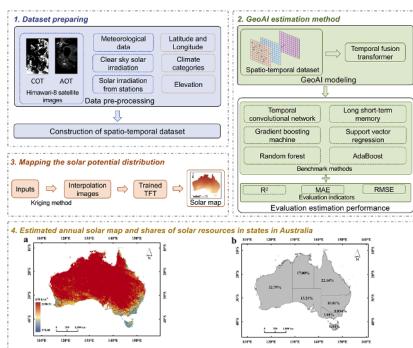
^e Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, PR China



HIGHLIGHTS

- Constructing the GeoAI dataset for spatio-temporal data.
- Proposing a temporal fusion transformer-augmented GeoAI framework.
- Offering interpretable and highly accurate estimation results.
- Using geo-visualization technology to generate the estimated solar maps.

GRAPHICAL ABSTRACT



ARTICLE INFO

Key words:

Land surface solar irradiation
Spatial-temporal data
Geospatial artificial intelligence
Temporal fusion transformer
Geographic Information Science
Remote sensing

ABSTRACT

Accurate estimation of land surface solar irradiation is critical for effective solar energy utilization and planning of solar photovoltaic planning. Although traditional machine learning methods have been demonstrated to estimate solar irradiation effectively, they face challenges in modeling over large regions, as well as lacking of ability to model spatial diversity and temporal dynamics of solar irradiation, and providing limited interpretability. To address these limitations, this study proposed a geospatial artificial intelligence framework augmented by Temporal Fusion Transformer for hourly estimation of land surface solar irradiation. As a case study in Australia, the results demonstrate superior performance with the coefficient of determination, the mean absolute error, and Root Mean Square Error as high as 0.90, 0.25(kWh/m²), and 0.63(kWh/m²), showing improvements of 21.62–66.67 %, 78.37–85.98 %, and 62.81–73.25 %, respectively, compared to the benchmarks of other methods, including Support Vector Regression, Random Forest, Gradient Boosting Machine, AdaBoost, Long Short-Term Memory, Temporal Convolutional Network, ConvLSTM, Transformer, and Graph Neural Network. Furthermore, interpretability results of the model indicate that among the temporal variables, observed solar irradiation and clear sky solar irradiation significantly contribute to the model's performance. The results show this framework enhanced accuracy and interpretability for solar irradiation estimation over large areas,

* Corresponding author.

E-mail address: Ls.charles@polyu.edu.hk (M.S. Wong).

providing valuable insights for future studies and supporting decision-making for developing the renewable energy industry.

1. Introduction

The pursuit of carbon neutrality as a global objective is aimed at mitigating greenhouse gas emissions and laying the foundation for a sustainable future in the world. In this challenging endeavor, the utilization of renewables, such as solar energy, wind power, and ocean energy, is regarded as one of the most promising and effective approaches to achieve carbon neutrality [1]. Among these sources, solar energy stands out as one of the most promising due to its unique advantages, including its clean, renewable, and ubiquitous nature [2]. A powerful method to harness solar energy is the rapidly developing photovoltaic technology [3,4]. Deploying and optimizing solar photovoltaic (PV) systems require the precise estimation of solar potential to maximize solar farming while minimizing PV areas [5]. Thus, precise solar irradiation estimation is of significance to maximize the utilization of solar energy.

The present studies used for estimating solar irradiation can be divided into three categories, including empirical, time series, and artificial intelligence algorithms [6]. Empirical models utilize mathematical functions to establish correlations between solar radiation and other relevant parameters [7], such as bright sunshine hours [8–10], cloud cover [11–13], and temperature [14–16]. Although empirical methods have been adopted to estimate solar irradiation, these approaches are limited by the locational dependence and the empirical coefficients in the estimation. This limitation arises due to the variation of regression coefficients in the models across different locations.

Moreover, plenty of studies have focused on traditional time series methods for solar irradiation estimation, such as autoregressive integrated moving average (ARIMA) [17], autoregressive moving average (ARMA) [18], an autoregressive moving average model with exogenous variables (ARMAX) [19], and autoregressive fractionally integrated moving average (ARFIMA) [20]. Although these models have generally demonstrated success in estimating solar irradiation in certain regions, they pose certain drawbacks when applied to non-stationary time series solar irradiation. Specifically, ARIMA and ARMA base estimations solely relied on historical solar irradiation data, resulting in limited transferability. This limitation arises because solar irradiation exhibits spatiotemporal variations across regions, and trends observed in one region may not be applicable to others. Moreover, ARMAX and ARFIMA share this limitation. Despite both methods requiring external variables for solar irradiation estimation, the complex parameter settings are difficult to transfer to untrained regions.

With the advancement of computer science, the use of Artificial Intelligence (AI) algorithms has become widespread. Many studies have confirmed that AI algorithms have provided reliable and highly accurate results [6]. Some traditional machine learning methods, such as Support Vector Regression (SVR), Random Forest (RF), and Gradient Boosting Machine (GBM), show experienced quality assurance in some application tasks. For example, Liao et al. [21] conducted station-by-station modeling using four methods (i.e., SVR, RF, GBM, and Multi-Layer Perceptron (MLP)) and selected the optimal model to estimate land surface solar irradiation in Australia and China. The results suggest that GBM outperformed other methods in terms of four indicators, including a coefficient of determination (R^2), normalized Root Mean Square Error (nRMSE), normalized mean bias error (nMBE), and consumption of time (t). Although Liao et al. [21] proposed a reliable method for generating land surface solar irradiation maps in Australia and China, this study has two limitations: i) station-by-station modeling used in this study increased the computational cost; ii) optimal model, trained with data from a single station, may not fully capture the complete range of solar irradiation across large area. Therefore, using this model beyond the

station's dataset could lead to the reduction in accuracy in estimating solar irradiation. Furthermore, the abovementioned methods investigated the non-linear relationship between many input parameters and solar irradiation and have not considered the time series characteristic of solar irradiation, which may lead to suboptimal performance in solar irradiation estimation. To tackle this problem, several time series deep learning methods are proposed and show good performance in extracting time series features, such as Long Short-term Memory (LSTM), Temporal Convolutional Network (TCN), and Recurrent Neural Network (RNN). Alper et al. [22] used LSTM, MLP, and adaptive neuro-fuzzy inference system with grid partition, and fuzzy c-means to estimate the one-hour-ahead solar irradiation in Tarsus, Turkey. The results illustrate that the LSTM model in 1-h-ahead solar irradiation estimating yielded the highest accuracy performance. While these methods yield highly accurate estimation results, they are limited in exploring the effects of geographical heterogeneity on solar irradiation, especially for large area estimation. Specifically, several studies have highlighted the significance of geographical factors such as latitude, solar position, and topography in influencing solar irradiation levels [23–25]. To address geospatial problems, Geospatial Artificial Intelligence (GeoAI) is proposed, which integrates spatial studies and AI, primarily focusing on incorporating spatial thinking and concept into deep learning models [26]. GeoAI models have been demonstrated to outperform traditional non-spatial machine/deep learning models in several energy-related tasks. For example, Huang et al. [27] proposed a 3D ConvLSTM-CNN network to extract the temporal and spatial features of the sky image for solar irradiation estimating. The experimental results show the proposed method has a promising performance for 5-min ahead global horizontal irradiation estimation, achieving 28.2 %, 34.8 %, 19.9 %, 42.7 %, and 68.3 % improvement on nRMSE, mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE), median absolute percentage error (MedAPE), and R^2 , suggesting taking the merits of spatial features and time-series features can improve the accuracy of solar irradiation estimation. Although the GeoAI methods could overcome the limitation of geographic heterogeneity by integrating the spatial features with the temporal features, model interpretability is still a major challenge in GeoAI research because of the black-box nature. Recent studies have focused on the interpretability of AI. For example, Gao et al. [28] used a convolutional neural network (CNN) to extract the feature maps for the hourly solar radiation estimation in Tokyo, and the extracted feature maps boosted the interpretability of convolution layers. Similarly, the Graph Neural Network [29], the attention mechanism [30], and a transformer model [31] are able to offer explanatory results for estimating solar radiation in some studies. However, these interpretable AI models have not considered the impact of geographic heterogeneity on solar radiation.

Based on the literature review, we found that utilizing traditional machine learning for station-by-station modeling not only complicates the computation but also struggles to account for the spatial diversity and temporal dynamics of solar irradiation across large area. Although some machine learning methods can conduct simplified spatio-temporal modeling of solar irradiation, these models often exhibit poor interpretability, making it challenging for individuals to comprehend the contributions of spatio-temporal variables in model development. To fill these research gaps, we propose an interpretable GeoAI framework augmented by Temporal Fusion Transformer (TFT) for estimating hourly solar irradiation in Australia. The primary contributions of this study are outlined as follows.

- (1) This study introduces a novel approach for constructing a comprehensive spatio-temporal dataset by integrating multi-

source data from multiple stations, including aerosol optical thickness (AOT) and cloud optical thickness (COT) from Himawari-8 satellite images, measured solar irradiation, meteorological data, clear sky solar irradiation (CSI), and the geographic attributes of stations such as position, elevation, and climate categories. Utilizing Geographic Information Science (GIS) technology, this dataset allows for a unified representation of diverse data types across multiple stations. This innovative GIS-based dataset construction facilitates the use of the TFT model for unified modeling, overcoming the tedious process of station-by-station modeling and enhancing the model's generalization capabilities across large regions.

- (2) This study develops a GeoAI framework that integrates both temporal and spatial data into a single dataset and employs the TFT model for training. This framework is a relatively novel approach in solar irradiation estimation research. The TFT model handles both temporal data and static geographic spatial information within the network structure effectively, thereby addressing the challenge of geographic heterogeneity. This capability enables more accurate large area solar irradiation estimation, providing significant improvements over traditional methods.
- (3) The TFT method offers interpretable results regarding the significant internal processing of the modeling. This is achieved by calculating the importance of time-varying variables and static variables in the Encoder, Decoder, and the static layer, along with attention weights assigned to various time steps. These calculated values provide a comprehensive understanding of the contributions of all variables in the TFT modeling process. Furthermore, the geo-visualization technology is performed to generate hourly land surface solar irradiation maps in Australia.

This paper is organized as follows. Section 2 presents the GeoAI framework based on TFT for estimating land surface solar irradiation. Section 3 compares the estimation performance and analyzes the results. Finally, Section 4 shows the conclusion.

2. Methodology

2.1. Overview

Fig. 1 demonstrates the mechanism of the GeoAI framework proposed in this study. The GeoAI framework begins by constructing a spatio-temporal dataset, followed by applying the TFT method to train a GeoAI model utilizing this dataset. Subsequently, leveraging this trained GeoAI model, the framework proceeds to estimate hourly land surface solar irradiation and subsequently express the estimations through geospatial visualization techniques. Specifically, we first cleaned data. Then, the spatio-temporal dataset was constructed in GIS. Next, the GeoAI estimation method was utilized to estimate hourly land surface solar irradiation. The TFT was employed for modeling the hourly land surface solar irradiation using the constructed spatio-temporal dataset. To evaluate the estimation performance of the TFT model, nine methods (i.e., RF, GBM, SVR, LSTM, AdaBoost, TCN, ConvLSTM, Transformer, and Graph Neural Network (GNN)) were selected as the benchmarking models to compare the estimation performance with the TFT model, based on three evaluation indicators. Finally, the Geo-visualization method was used to map the estimated hourly land surface solar irradiation in Australia from 8 am to 4 pm within a day.

2.2. Study area and data

Australia, with its large spatiotemporally heterogeneous solar distribution [21,32], is an ideal study area for thoroughly testing the network under diverse conditions. A dataset was generated utilizing 13 stations in Australia, encompassing the necessary hourly datasets spanning six years from 2015 to 2020. **Fig. 2** shows the distribution of the 13 meteorological stations in Australia, and **Table 1** introduces the geographical coordinates, elevation, climate category, and ranges of observed solar irradiation of these stations.

Many studies have demonstrated that meteorological data, such as temperature, wind speed, and atmosphere pressure, can greatly affect the estimation performance using machine learning methods [34–38]. Additionally, solar radiation experiences attenuation after passing through clouds and aerosols before reaching the Earth's surface. Therefore, this study employs COT, AOT, CSI, meteorological

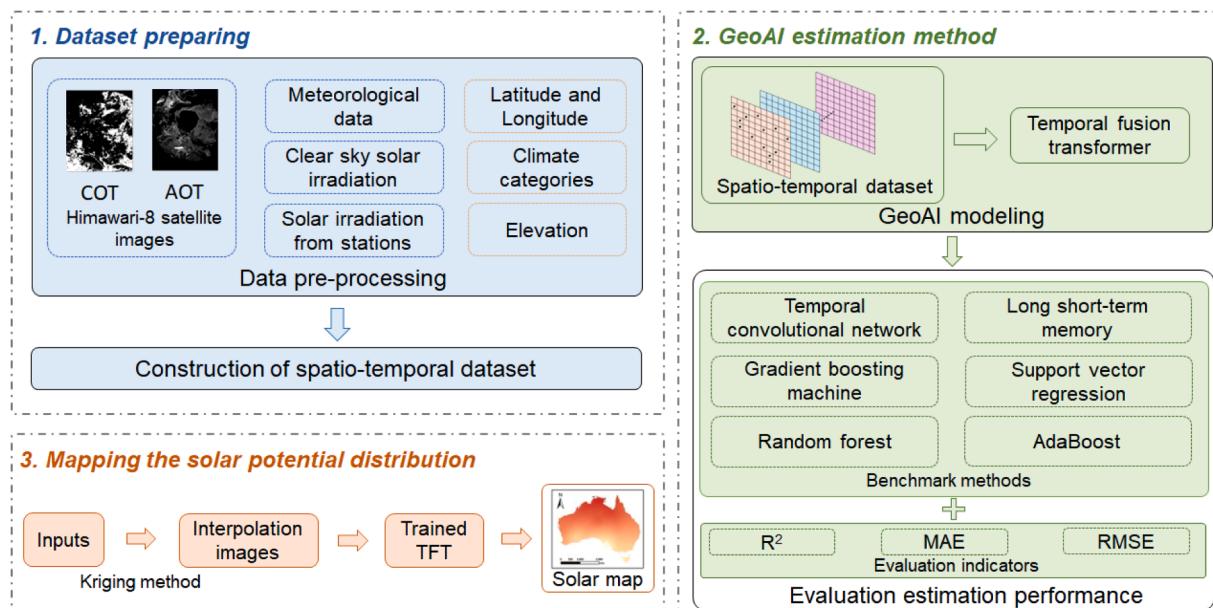


Fig. 1. The mechanism of GeoAI framework.

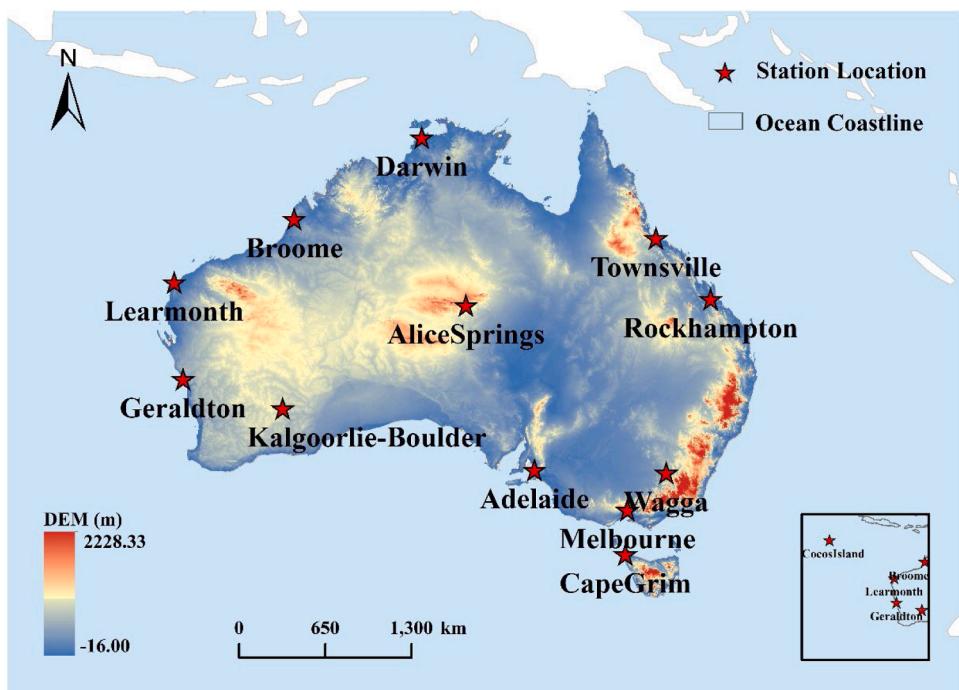


Fig. 2. The distribution of the 13 meteorological stations in Australia. The background is Digital Elevation Model (DEM) with a 250 m resolution [33].

Table 1

Geographical coordinates, elevation, climates, and ranges of observed solar irradiation of the 13 meteorological stations in Australia. (M: Mediterranean, SHD: Subtropical hot desert, HSA: Hot semi-arid, TO: Temperate oceanic, TR: Tropical rainforest, TS: Tropical savanna, SA: Semi-arid, HS: Humid subtropical).

Station no.	Station name	Latitude ($^{\circ}$ S)	Longitude ($^{\circ}$ E)	Elevation (m)	Climate	Range of observed solar irradiation (kWh/m^2)
1	Adelaide	34.95	138.51	2.00	M	0–1.38
2	Alice Springs	23.79	133.88	546.00	SHD	0–1.48
3	Broome	17.94	122.23	7.42	HSA	0–1.44
4	Cape Grim	40.68	144.68	95.00	TO	0–1.31
5	Cocos Island	12.18	96.83	3.00	TR	0–1.37
6	Darwin	12.42	130.89	30.40	TS	0–1.45
7	Geraldton	28.80	114.69	29.70	M	0–1.44
8	Kalgoorlie-Boulder	30.78	121.45	365.30	SA	0–1.39
9	Learmonth	22.24	114.09	5.00	HAS	0–1.36
10	Melbourne	37.66	144.83	113.40	TO	0–1.41
11	Rockhampton	23.37	150.47	10.40	HS	0–1.51
12	Townsville	19.24	146.76	4.34	TS	0–1.57
13	Wagga	35.15	147.45	212.00	HS	0–1.43

parameters (MIs) (i.e., air temperature, humidity, wind speed, atmospheric pressure, dew point, minimum temperature, maximum temperature, and apparent temperature) as the inputs, and the observed land surface solar irradiation as the ground truth. Zang et al. [23] utilized spatial features extracted from stations and temporal features from historical solar irradiation time-series data to estimate short-term global horizontal irradiation, emphasizing the enhanced estimating performance achieved through the integration of spatial and temporal features. In this study, we used the geographical latitude and longitude, the elevation of the stations, climate category, and meteorological parameters, as the inputs in the model building.

Since the Himawari-8 [39] products are freely available with high spatio-temporal resolutions, we extracted AOT data and COT data from the Himawari-8 satellite images with an hourly updated temporal resolution and the 5 km spatial resolution. Additionally, hourly meteorological parameters are obtained from the Openweather website [40]. Hourly CSI in 13 stations is calculated using a Python online library called Pysolar [41]. The hourly solar irradiation data was measured by 13 meteorological stations in Australia [40]. Table 2 lists the specific category, source, and spatial resolution of all data.

Table 2

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

Data name	Data category	Data source	Resolution
AOT	Temporal data	Himawari-8 satellite images	5 km
COT	Temporal data	Himawari-8 satellite images	5 km
CSI	Temporal data	Calculation values by Pysolar	Discrete-point data, no spatial resolution
MIs	Temporal data	Openweather website	Discrete-point data, no spatial resolution
Observed hourly solar irradiation	Target data	Meteorological stations in Australia	Discrete-point data, no spatial resolution
Elevation	Spatial data	Meteorological stations in Australia	Discrete-point data, no spatial resolution
Latitude	Spatial data	Meteorological stations in Australia	Discrete-point data, no spatial resolution
Longitude	Spatial data	Meteorological stations in Australia	Discrete-point data, no spatial resolution
Climate category	Spatial data	Meteorological stations in Australia	Discrete-point data, no spatial resolution

2.3. Construction of spatio-temporal dataset

2.3.1. Spatial data and temporal data

The GeoAI model in this study was trained using a dataset generated from data corresponding to 13 stations. Since this study employed multi-source data, including satellite images and text data, these raw data differ in format, representation, and resolution. To organize the multi-source data into a training dataset, we first extract pixel values corresponding to the geographic coordinates of stations from the remote sensing images and remove parts of the text data irrelevant to model training, and then reconstruct and standardize the data. GIS technology was employed to reconstruct these data, transforming them into spatial data and temporal data, as illustrated in Fig. 3. Specifically, temporal data refers to information that varies over time at specific geographic locations, integrating both spatial and temporal elements to capture how certain attributes at a site evolve across different time periods. The structure of temporal data is organized around the geographic coordinates of each station, which serve as the spatiotemporal correlation. This structure typically includes the variable name, the station's geographic coordinates, and the corresponding time series of observed values. As illustrated in Table 2, temporal data encompass MIs, observed solar irradiation from stations, CSI, COT, and AOT. In contrast, spatial static data pertains to information that remains constant over time at a given location. This type of data consists of the geographic position (such as latitude and longitude) and its associated attributes, like elevation and climate category. The organization of spatial static data also relies on the geographic coordinates of each station as the spatial correlation, and includes station name, latitude and longitude, elevation, and climate category. As shown in Table 2, spatial data include geographic coordinates, elevation, station name, and climate category.

2.3.2. GeoAI dataset

We adopted the spatio-temporal dataset construction framework proposed by Liao et al. [42], as depicted in Fig. 3. The process began

with the integration of reconstructed spatial and temporal data, following these key steps: i) the spatial data and temporal data were merged based on their spatio-temporal correlation, which is the geographic coordinates of the data, ii) data imputation was performed on the merged dataset. Specifically, the MissForest method [43], a machine learning-based technique for simulating missing data, was utilized to fill in the gaps in the dataset. The missing values comprised only 0.02 % of the entire dataset, and iii) Since the TFT model distinguishes between static and dynamic input variables via separate network layers, we explicitly labeled the spatial (static) and temporal (dynamic) variables in the dataset. Consequently, the resulting spatio-temporal dataset encompasses geographic coordinates, time stamps, temporal variable types (AOT, COT, CSI, and MIs), their corresponding values, climate category, elevation, and station names. Furthermore, hourly observed land surface solar irradiation measured at each station was designated as the target variable for model training within this dataset. After that, the whole dataset is divided into samples which consist of 45 continuous time steps (where each time step represents an hour). Within each sample, the previous 36 time steps are used for training, and the final nine time steps are used for testing. 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 data, respectively.

2.4. GeoAI method

The introduction of the GeoAI framework aims to address the challenge faced by traditional machine learning models in estimating solar irradiation due to their limited ability to handle geographical heterogeneity. This geographical heterogeneity arises from the lack of consideration for static geographical spatial relationships, such as geographic location, climate category, and elevation, and their impact on solar irradiation distribution. To tackle this geographical heterogeneity, the AI methods employed within the GeoAI framework should be

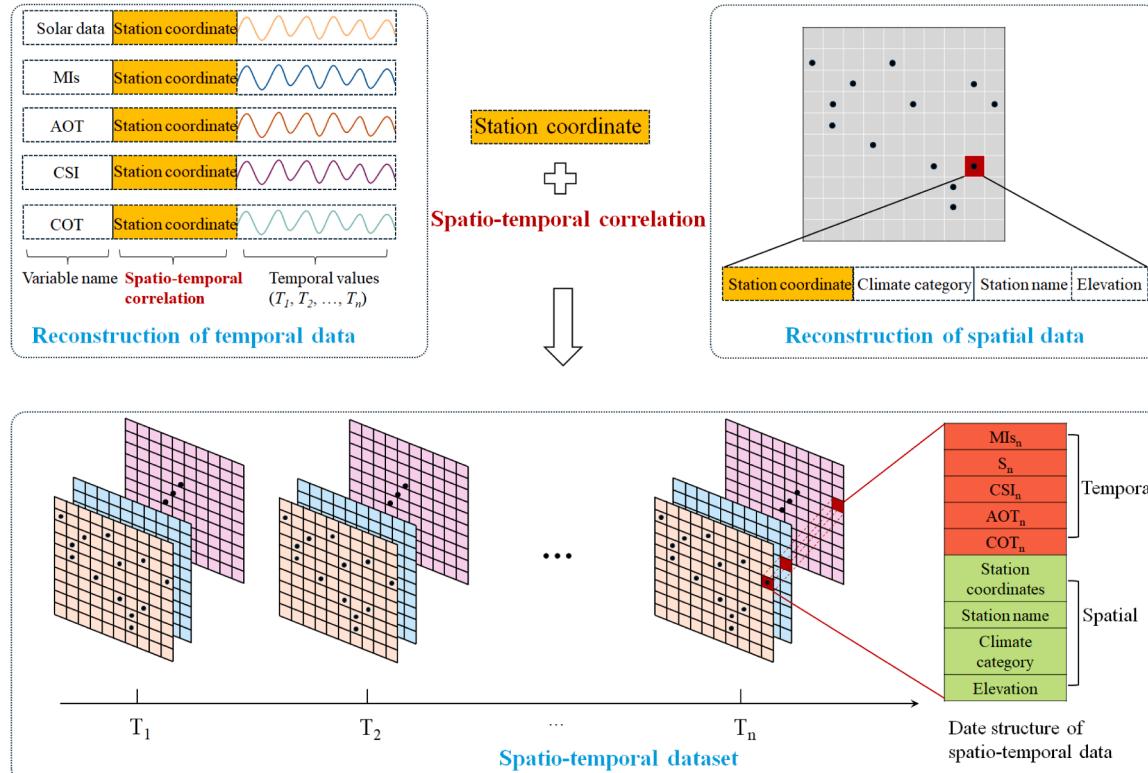


Fig. 3. The process of the GIS representation for constructing the spatio-temporal dataset [42].

capable of modeling the nonlinear temporal relationships of solar irradiation and the static geographical spatial relationships. The Temporal Fusion Transformer (TFT) model [44], which leverages attention-based deep learning architecture, precisely captures the temporal dependencies of time-series variables and the influences of static variables within the network. Hence, in this study, we employed the TFT model as the GeoAI method to estimate hourly land surface solar irradiation.

2.4.1. Temporal fusion transformer modeling

TFT model mainly consists of Gated Residual Network (GRN), Variable Selection Network (VSN), Static Covariate Encoders (SCE), and Temporal Fusion Decoder (TFD). The GRN is used as a building block of TFT to filter out unnecessary components of the architecture and can be flexibly applied to non-linear processing only where needed. The SCE is used for processing spatial data from GeoAI dataset and feeding the spatial characteristics into LSTM encoder-decoder layer for integrating the temporal data, the TFD for capturing the spatiotemporal characteristics from GeoAI dataset, which helps TFT model to estimate the distribution of hourly solar resources in different geographical locations accurately. Additionally, VSN is used for selecting relevant inputs to preserve the most critical features as well as to reduce computational costs and improve overfitting and multicollinearity problems. The importance of each input variable can be calculated through VSN, which offers a feature-based interpretation of this model. Also, the attention mechanism used in TFD allocates weights to each time step of the model, reflecting the importance of features at those respective time steps in the model's estimations. In other words, the attention mechanism quantifies the importance of input variables at different time steps for the model's performance, thereby aiding in a better understanding of the model. The mechanism of TFT model is detailed in [Appendix A](#).

The implementation of the TFT model involves the use of Python 3.8 along with TensorFlow 2.12.0, PyTorch-forecasting 0.10.3, and PyTorch-lightning 1.8.6. We employed the Python library “Time-SeriesDataset” to split the data. Early stopping was utilized to prevent overfitting. The computations were executed on a high-performance computer equipped with an intel (R) Core (TM) i7-6800 K CPU, operating at 3.40 GHz, 6.0 TB RAM, and running on the Ubuntu 16.04 LTS system.

2.4.2. Evaluation method

To evaluate the estimation performance of the TFT model, the most frequently used evaluation indicators, i.e., the coefficient of the determination (R^2), the mean absolute error (MAE), and Root Mean Square Error (RMSE) were adopted, given as:

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

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (2a)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3a)$$

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

To evaluate the estimation performance of TFT model, we employed the commonly used machine learning methods (i.e., RF, SVR, GBM, and AdaBoost) and deep learning methods (i.e., LSTM, TCN, convLSTM, Transformer, and GNN) to compare the estimation results of the TFT model based on the R^2 , MAE, and RMSE. To evaluate the correlation between observed surface solar irradiation and the variables discussed in [Section 2](#) (including MIs, AOT, COT, and CSI), we employed the Pearson correlation coefficient (PCC). This statistical measure quantifies the degree and direction of linear dependence between two variables

[45]. Mathematically, PCC is calculated as the covariance of variables X and Y, normalized by the product of their standard deviations. This normalization ensures the coefficient ranges between -1 and 1, indicating the strength and polarity of their linear association. This coefficient can also be equivalently defined as follows:

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \sum (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \quad (4a)$$

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. Mapping the solar potential distribution

To better illustrate the diurnal variation of solar irradiation, we generated the hourly land surface solar irradiation map from 8 am to 4 pm on July 31st 2020. The MIs, COT, AOT, and CSI images are well prepared and combined as the multi-attribution image. MIs data from each station was interpolated into grids with 5 km by 5 km spatial resolution using the Kriging interpolation method to derive a set of MIs images. The climate category and the pixel ID are the attribution values assigning for each corresponding pixel in this input image. The attributions in this input image were classified into static features or time-varying features. The labeled input image was put into the trained TFT model to generate the solar maps.

3. Results and discussion

3.1. The estimation performance of TFT

3.1.1. The results of 13 individual datasets

Liao et al. [21] constructed four individual machine learning models for each station using RF, SVR, GBM, and AdaBoost. It is found that the four models have similar estimation performance. Notably, using the GBM model, all stations in Australia achieved R^2 values of ≥ 0.7 , with approximately 38 % of the stations in Australia reaching R^2 values of ≥ 0.8 . The results suggest that the estimation models are reliable with high estimation accuracy, which indicates that the proposed method can effectively estimate land surface solar irradiation over large regions.

To evaluate the estimation performance of the TFT model, we also used the same datasets to estimate the land surface solar irradiation. The geo-visual comparison results of R^2 , RMSE, and MAE are presented in [Fig. 4](#), [Fig. 5](#), and [Fig. 6](#), respectively. It is found that 82 % of stations achieved R^2 values of ≥ 0.8 using TFT model, while only one station showed a relatively low accuracy result, with $R^2=0.64$. It is noted that approximately 62 % of the stations reached R^2 values of ≥ 0.9 . Especially, three stations present ultra-high accuracy, achieving $R^2=0.97$. Likewise, the RMSE and MAE of the TFT model in all stations are less than those of the GBM model. Compared to the results of the machine learning methods used in Liao et al. [21], the TFT model outperforms these machine learning methods. This suggests that the TFT model is highly effective in estimating time-series land surface solar irradiation with superior accuracy.

3.1.2. Overview of the constructed dataset

The remarkable innovation of the TFT method lies in its ability to integrate static spatial data and temporal data, allowing for the extraction and utilization of a more comprehensive dataset. In order to assess the TFT method's capability to process spatio-temporal data, we aggregated the datasets from the 13 stations into a single entire dataset to train the TFT model. Additionally, we applied four machine learning methods and five deep learning methods to estimate land surface solar irradiation using this combined dataset. The grid search method [46] was used to optimize the hyperparameters, and the optimization of

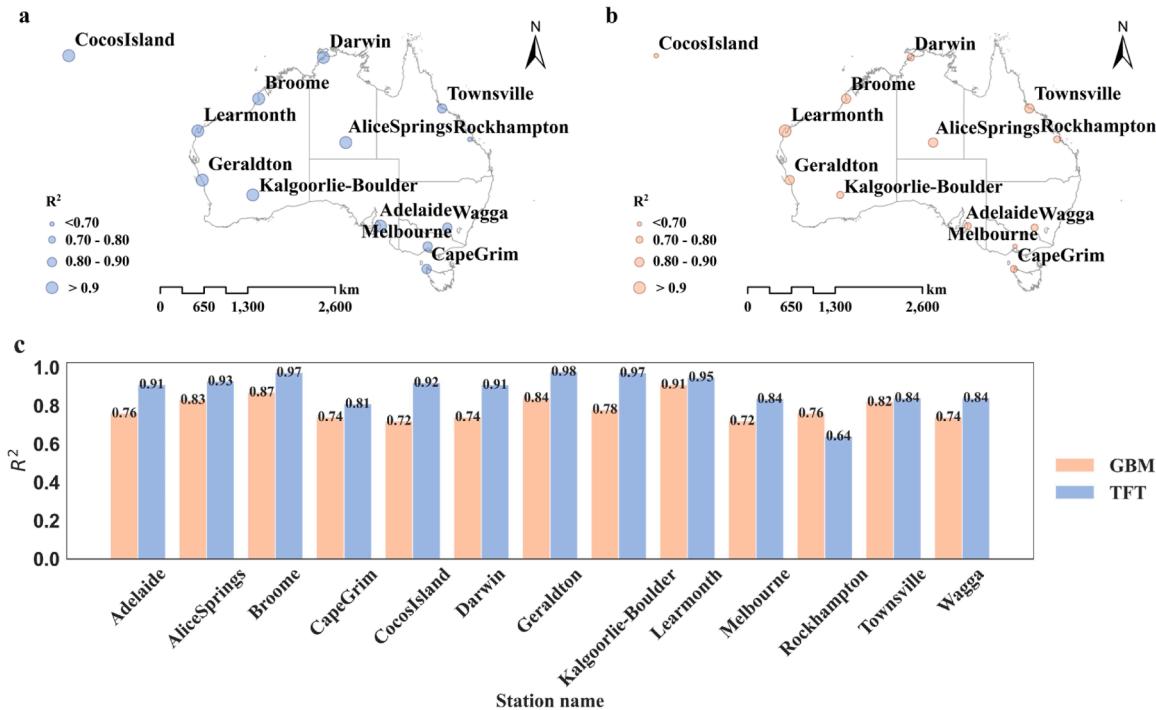


Fig. 4. The geo-visual comparison results of R^2 using GBM model and TFT model. (a) the visual result of the TFT model, (b) the visual result of the GBM model, (c) the detailed comparison bar results.

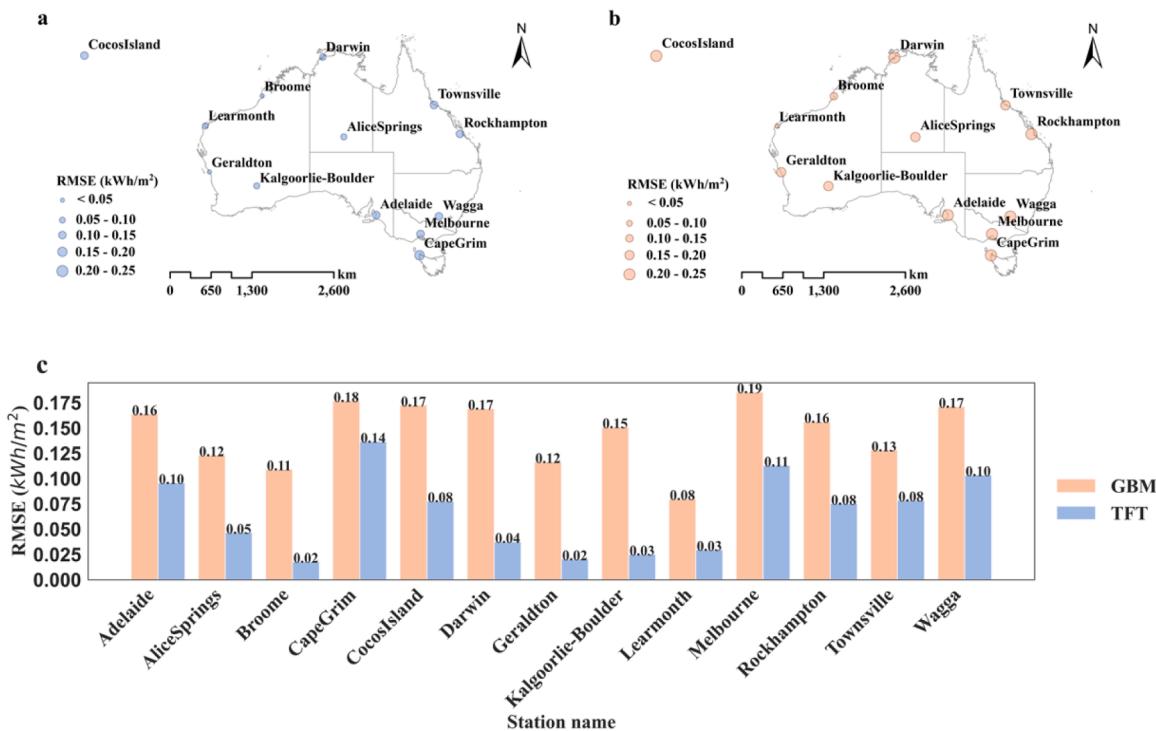


Fig. 5. the geo-visual comparison results of RMSE using GBM model and TFT model. (a) the visual result of the TFT model, (b) the visual result of the GBM model, (c) the detailed comparison bar results.

hyperparameters of RF, SVR, GBM, AdaBoost, LSTM, TCN, Transformer, convLSTM, GNN, and TFT are listed in Table 3. The results from all methods are depicted in Fig. 7. Overall, the TFT method outperforms other methods, achieving the highest estimation accuracy with $R^2=0.90$, $\text{MAE}=0.025(\text{kWh/m}^2)$, and $\text{RMSE}=0.063(\text{kWh/m}^2)$. Significant improvements can be observed when comparing the performance of the

TFT method with methods based on LSTM, TCN, RF, GBM, AdaBoost, Transformer, convLSTM, GNN, and SVR, with improvements between 21.62–66.67 %, 78.37–85.98 %, and 62.81–73.25 % observed for R^2 , MAE , and RMSE , respectively. It is noticed that only the network of the TFT model can learn spatial features and temporal features among these mentioned methods. Therefore, the highest accurate results from the

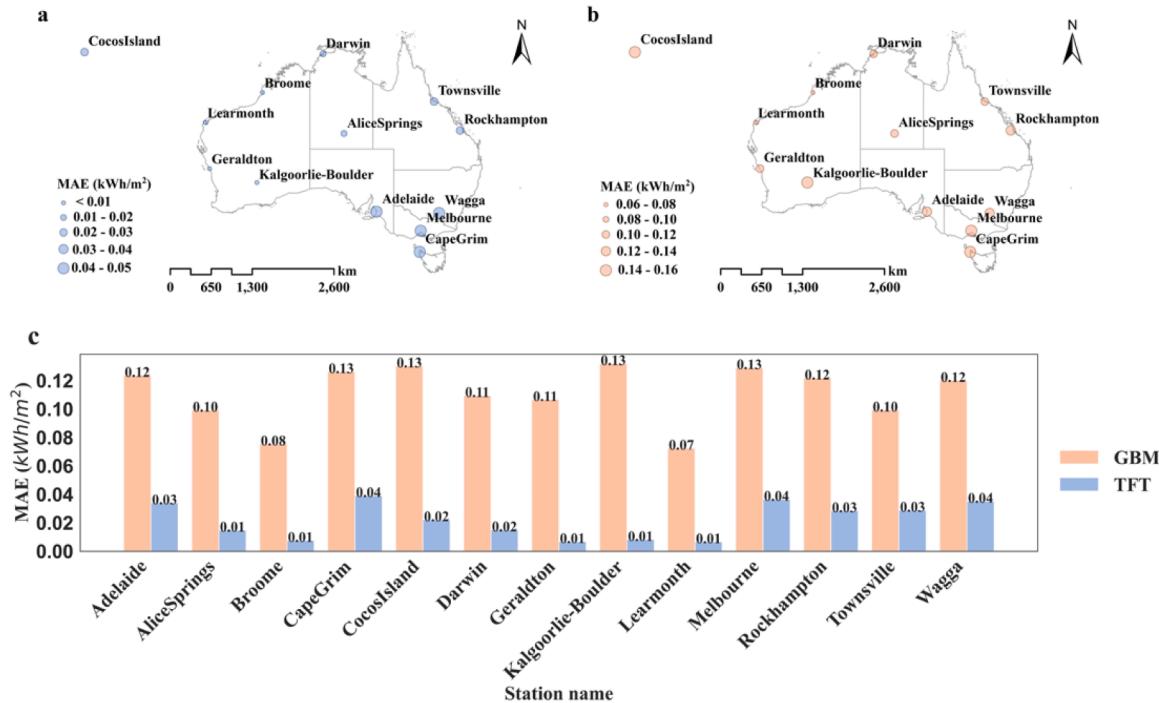


Fig. 6. The geo-visual comparison results of MAE using GBM model and TFT model. (a) the visual result of the TFT model, (b) the visual result of the GBM model, (c) the detailed comparison bar results.

Table 3

The hyperparameters of RF, SVR, GBM, AdaBoost, LSTM, TCN, Transformer, convLSTM, GNN, and TFT.

RF	n estimators: 100, min samples leaf:1, min samples split:2
GBM	n estimators: 200, learning rate: 0.1, max depth:4, subsample:0.9
AdaBoost	n estimators: 100, learning rate: 1, base estimator: deprecated
SVR	C:100, epsilon: 0.1, kernel: poly,
LSTM	num layers:4, units:50, dropout:0.2
TCN	Kernel size:5, num layers:3, dropout:0.3
Transformer	Learning rate=0.01, attention head size=4, hidden size=128, dropout:0.3
convLSTM	Learning rate=0.01, num layers:3, kernel size:5, dropout:0.3
GNN	Learning rate=0.01, num layers:3, hidden dim: 64, dropout:0.3
TFT	Learning rate=0.03, hidden size=16, attention head size=1, dropout=0.1, hidden continuous size=8

TFT model illustrate the superiority of combining spatial and temporal correlations using the TFT model.

Furthermore, we compared the accuracy of the TFT model and the GBM model using both the entire dataset and 13 individual datasets. As depicted in Figs. 4 through 7, the TFT model exhibits superior performance on the entire dataset compared to the average performance on individual datasets. Specifically, there are notable improvements of 16.28 %, 2.27 %, and 4.79 % in terms of R^2 , MAE, and RMSE, respectively. Conversely, the GBM model shows a degradation in performance when applied to the entire dataset, with increases in MAE and RMSE by 25.65 % and 26.70 %, respectively, and a decrease in R^2 by 11.54 %. The GBM model trained on a dataset composed of data from a single station demonstrates high precision but exhibits decreased accuracy when applied to datasets comprising multiple stations. In contrast, the TFT model performs exceptionally well on complex spatio-temporal datasets encompassing multiple stations. This indicates that the TFT model learns the spatial diversity among different stations within multi-station spatio-temporal datasets effectively, thereby showing the efficacy of the proposed GeoAI framework in addressing the geographic heterogeneity issue in spatio-temporal data estimation.

The training times of different models were calculated, as presented in Table 4. Overall, the Adaboost model exhibits the shortest training time, indicating its swift computational speed when handling extensive datasets. However, the SVR model requires the longest training time, suggesting a greater time investment when processing large datasets. Despite the prolonged training duration of the TFT model, its notably high estimation accuracy, as depicted in Fig. 7, positions this training time within an acceptable range. Consequently, selecting the TFT model as the estimation model is a rational decision, as its protracted training time is offset by its superior accuracy, particularly in addressing the demands of large datasets.

3.1.3. Analysis of seasonal variations

Following the training of the TFT model, we utilized the trained model to estimate land surface solar irradiation in all 13 stations within a 18-hour period during daytime in Spring (October 1st and October 2nd), Summer (January 1st and January 2nd), Autumn (April 1st and April 2nd), and Winter (July 1st and July 2nd) of 2019. The selected timeframe spans from 8 am to 4 pm. The estimated values and corresponding measured values are compared in Fig. 8, Fig. 9, Fig. 10, Fig. 11, respectively. Overall, the errors between estimated values and corresponding measured values are relatively small in four seasons, suggesting the estimation accuracy of TFT model is high. Additionally, the estimated land surface solar irradiation is the highest during Summer in all sites, while it is the lowest during Winter. Furthermore, the trend of land surface solar irradiation within a day across these sites is similar, with solar irradiation peaking around noon in all locations. These findings underscore Australia's abundant solar potential, and the consistent trend of solar irradiation throughout the day provides valuable insights for effectively distributing solar PV systems.

We have created seasonal solar irradiation maps for Australia (Fig. 13). The data show seasonal patterns, with the highest irradiation in summer and the lowest in winter. However, the seasonal differences are relatively small, confirming Australia's excellent solar energy potential. In terms of spatial distribution, there are some regional variations. The southwestern coastal areas receive slightly less solar irradiation than other regions, but overall, the differences across the

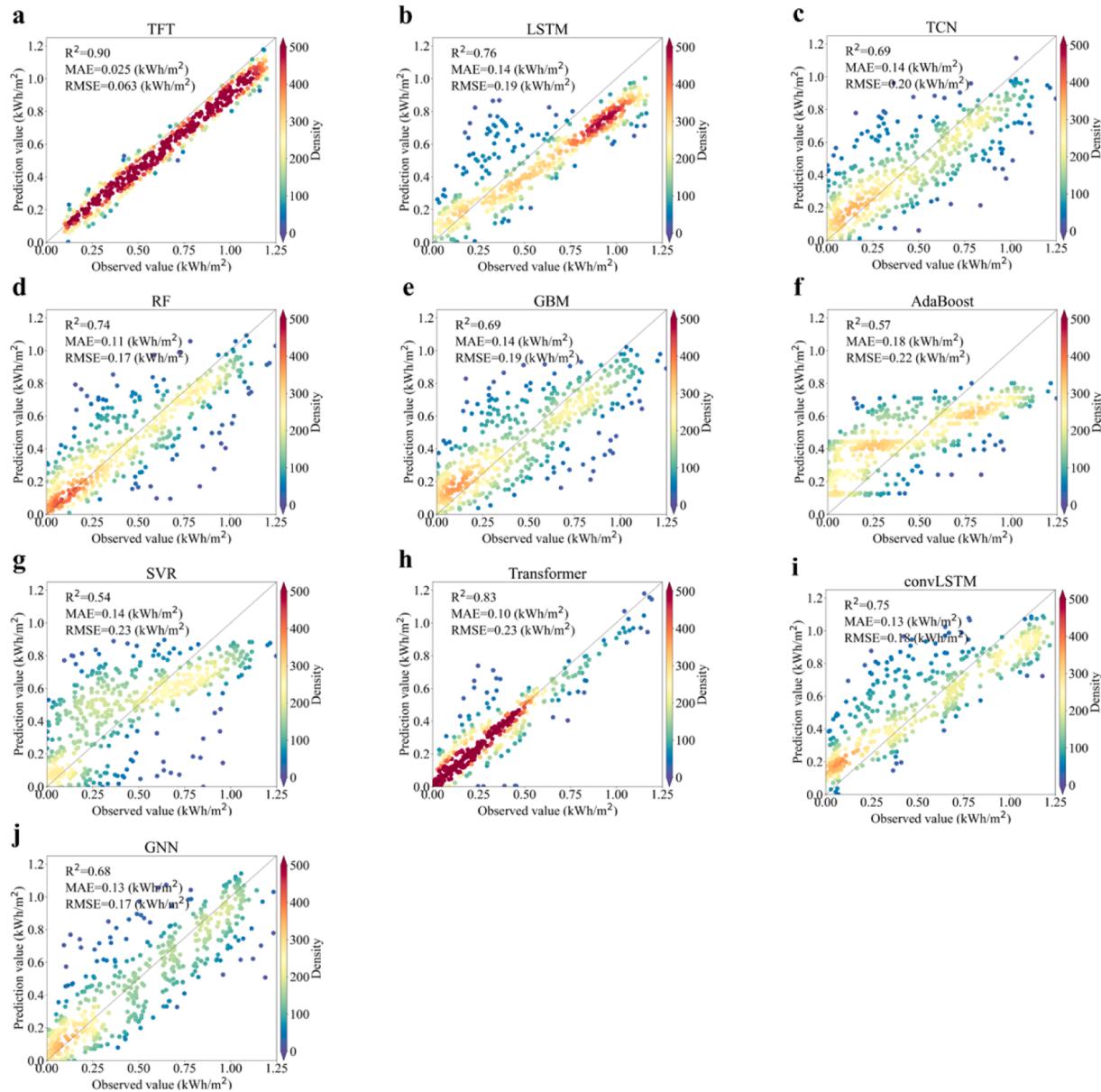


Fig. 7. The estimation results of TFT method, LSTM method, TCN method, RF method, GBM method, Adaboost method, SVR method, Transformer method, convLSTM method, and GNN method using the entire dataset.

Table 4

The comparison of training time of TFT method, LSTM method, TCN method, RF method, GBM method, Adaboost method, and SVR method using the entire dataset.

Model	Time (s)	Model	Time (s)
TFT	2123.58	AdaBoost	7.73
LSTM	1328.03	SVR	2322.59
TCN	797.09	Transformer	2063.23
RF	170.97	convLSTM	1967.36
GBM	40.27	GNN	2107.73

country are minor.

To evaluate the accuracy of the TFT model across the four seasons, we calculated the mean hourly absolute errors between the estimated and measured values for each season. Fig. 12 displays the results. The errors for all stations are less than 0.07 kWh/m², indicating that the TFT model performs with high accuracy. This consistently low error across different seasons suggests that the TFT model could be deemed as robust

and reliable for various seasonal conditions, providing accurate estimations year-round.

3.2. TFT interpretability

3.2.1. Correlation analysis

Fig. 14 presents the PCC values among these variables using a heatmap. CSI exhibits the strongest positive correlation with observed solar irradiation, followed by variables related to temperature, while humidity displays the strongest negative correlation. This outcome suggests that CSI, humidity, and variables associated with temperature exert significant influence on land surface solar irradiation. These variables emerge as strong potential factors for estimating solar potential.

3.2.2. Importance of input features

The TFT model displays a strong interpretable capability by the calculation of the variable importance scores for all input variables and the representation of the attention weight patterns. Fig. 15 illustrates the

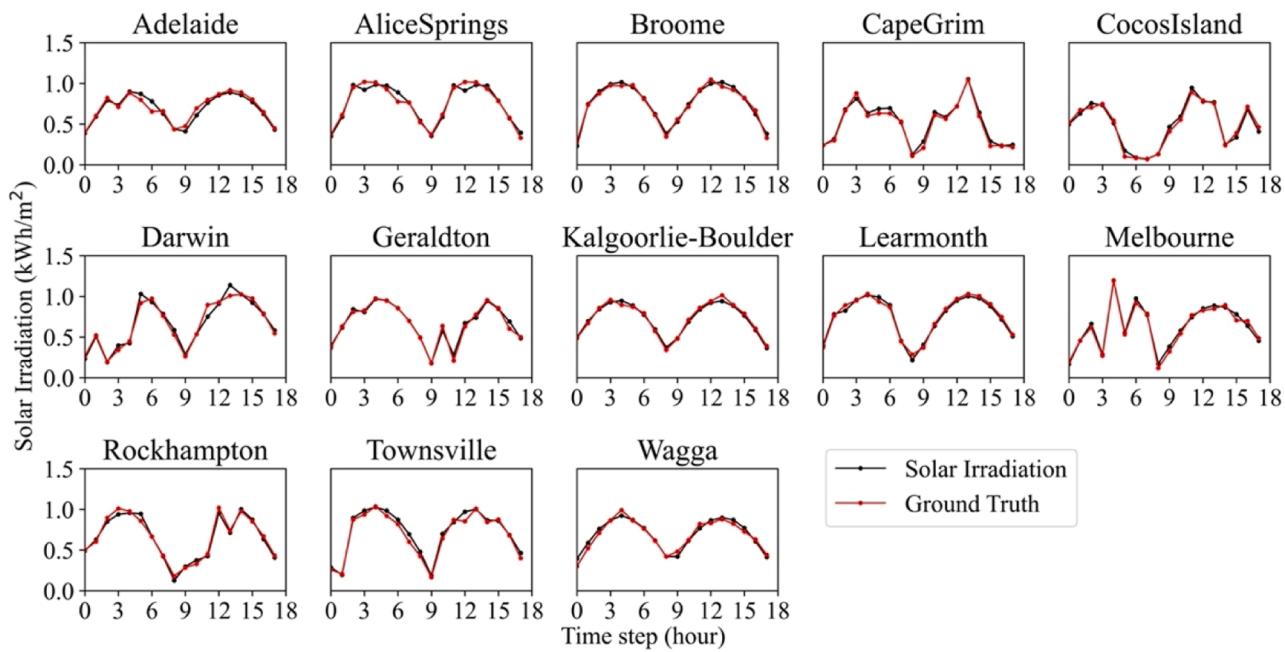


Fig. 8. The estimation results of 13 sites within 18 h of daytime in spring.

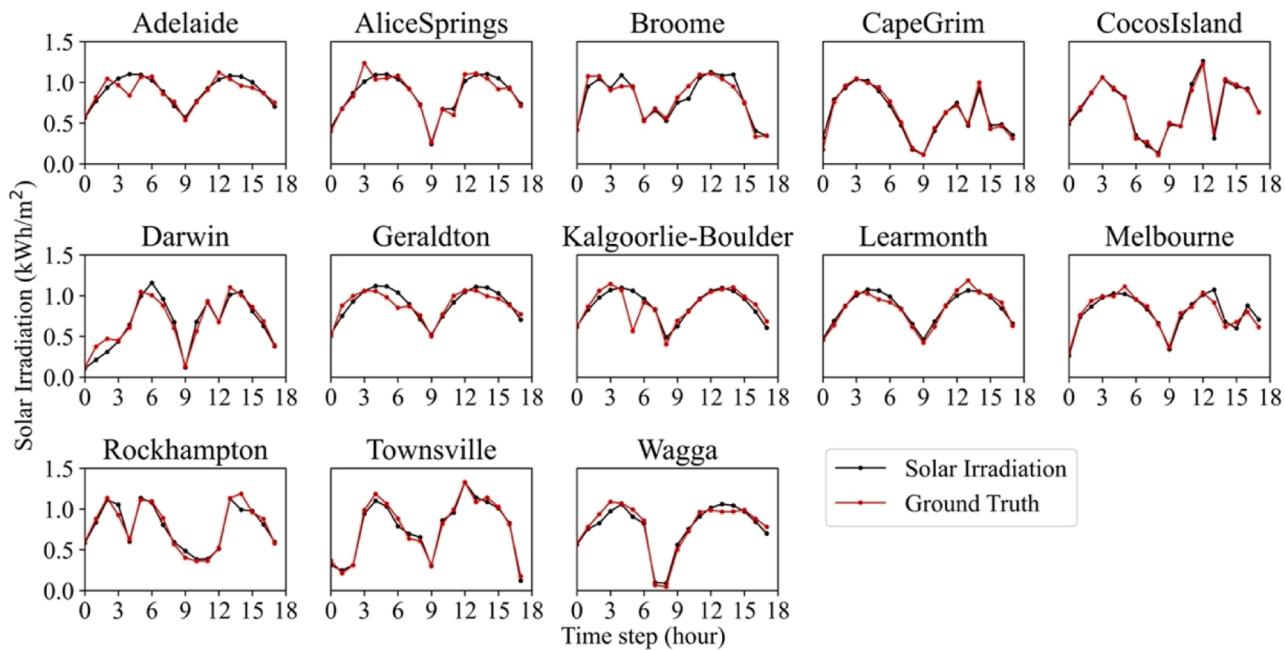


Fig. 9. The estimation results of 13 sites within 18 h of daytime in summer.

importance of the time varying variables in the Encoder and the Decoder, and the importance of the static variables, respectively.

The encoder variables represent the inputs for which past values are known at estimation time, including MIS variables, the observed land surface solar irradiation, and relative time index which represents the relative time sequence of the sample in the dataset. The importance of each encoder variable can be seen in Fig. 15(a). Notably, solar irradiation shows the highest importance in the Encoder layer, with over 40 % importance, following the CSI. The remaining variables just display relatively small importance. The findings from Fig. 15(a) suggest that CSI and observed solar irradiation play significant roles in training the model.

Decoder variables represent inputs for which future values are

known at estimation time, including MIS variables and relative time index. The importance of each decoder variable for the estimation model can be observed in Fig. 15(c). It is noticed that the relative time index and maximum temperature account for higher importance in the Decoder layer, both exceeding 13 %. The variables with importance ranging from 7 % to 11 % include apparent temperature, dew point, humidity, and mean temperature, and the remaining variables show similar importance, at about 6 %. These results suggest that the temporal features of the time series play an important role in the estimation of the model, and variables related to the temperature significantly contribute to the Decoder layer, greatly aiding the TFT model in estimating solar irradiation.

The results from both the encoder and decoder weighted importance

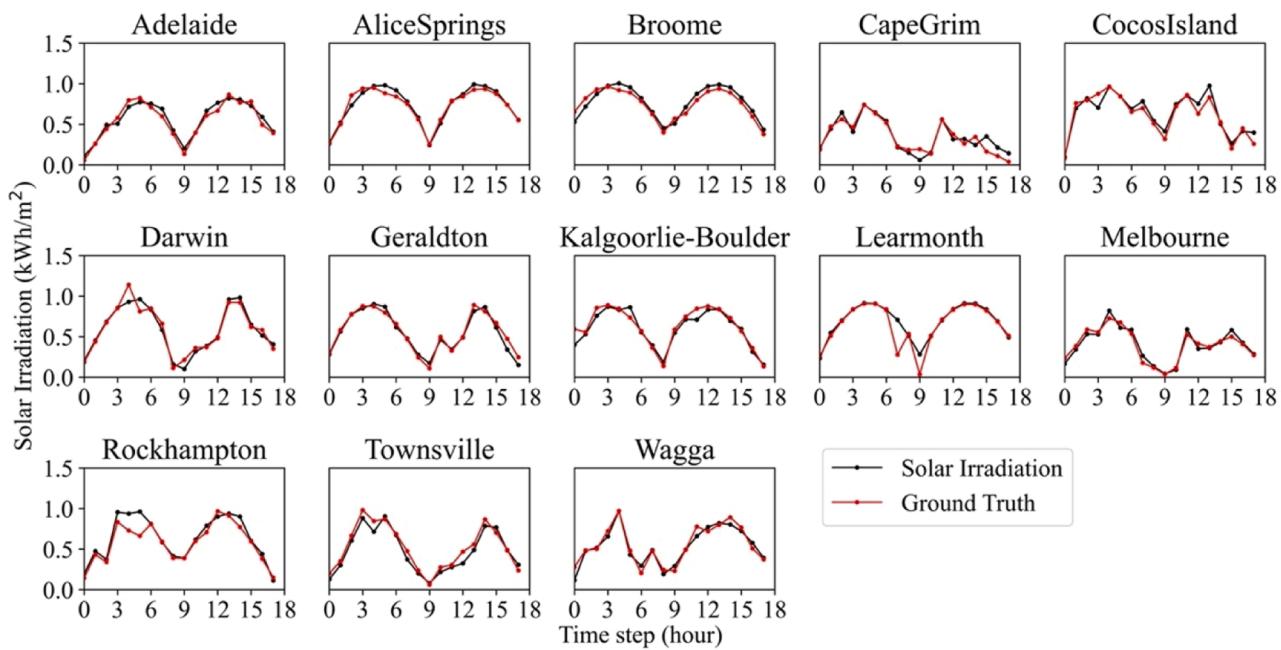


Fig. 10. The estimation results of 13 sites within 18 h of daytime in fall.

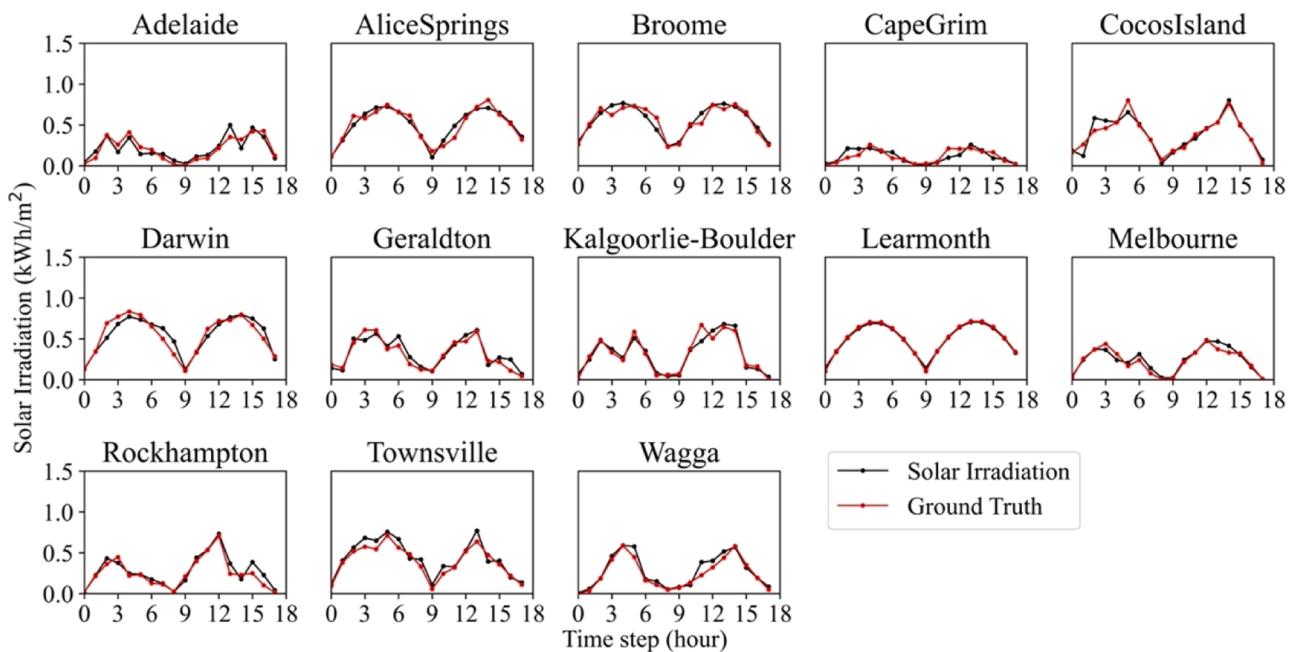


Fig. 11. The estimation results of 13 sites within 18 h of daytime in winter.

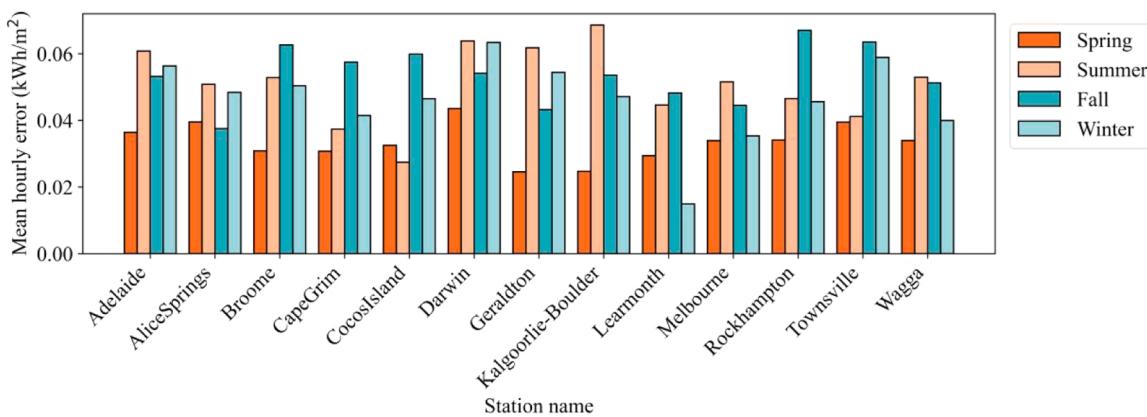


Fig. 12. The mean hourly absolute error of 13 stations in four seasons.

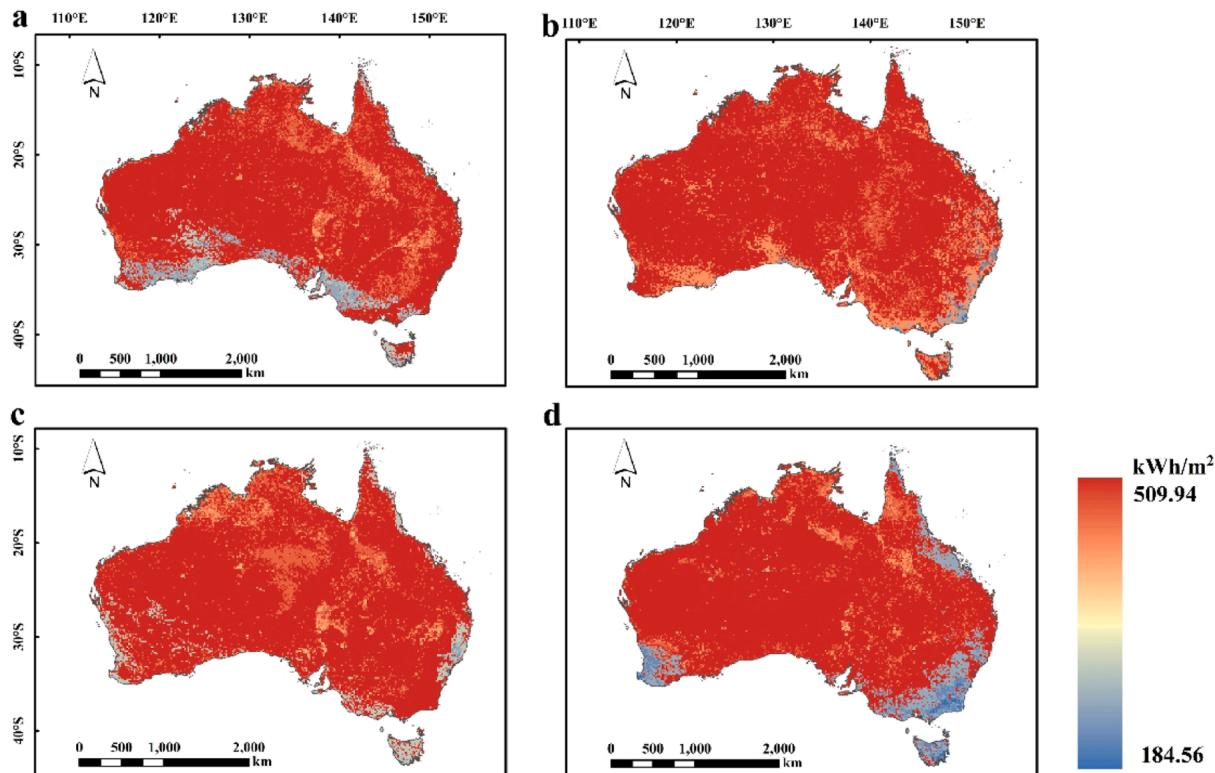


Fig. 13. Seasonal solar irradiation maps of four seasons. (a) map in Spring; (b) map in Summer;(c) map in Fall;(d) map in Winter.

highlight the necessity of having a good representation of factors related to solar irradiation and factors related to temperature to achieve high estimation performance. Combining PCC results with the calculated importance of all variables using the TFT model, we found that both results are highly consistent, which suggests that the interpretability of the TFT by calculation of the variable importance scores for all input variables is reliable.

The TFT dataset incorporates three types of static inputs: target center, target scale, and the identification of the facility, providing additional information and context to the model. In this case, the first two, $\text{solar}_{\text{center}}$, and $\text{solar}_{\text{scale}}$ are related to the standardization of the land surface solar irradiation, both serving as static variables in the model. $\text{solar}_{\text{center}}$ has a value of 0, and $\text{solar}_{\text{scale}}$ is the median of the time series. Furthermore, Station ID is used to identify each station in the

dataset. And $\text{encoder}_{\text{length}}$ is the default static input of TFT. In Fig. 15(b), we found that the $\text{solar}_{\text{center}}$ received the majority of the importance, accounting for almost 23 %, following by $\text{solar}_{\text{scale}}$, $\text{encoder}_{\text{length}}$, and Longitude of the station.

3.2.3. Hidden feature analysis

The TFT model not only shows good interpretability by providing the importance of variables in the different layers but also reflects the interpretability through attention values. Attention values reveal the model's dynamic focus on input variables at different time steps.

To greatly explain the calculated attention values, three typical attention value trend and values of their corresponding temporal variables are presented in Fig. 16. The training period of each sample is 36 h which consists of four-day data. Overall, the trend of the attention values

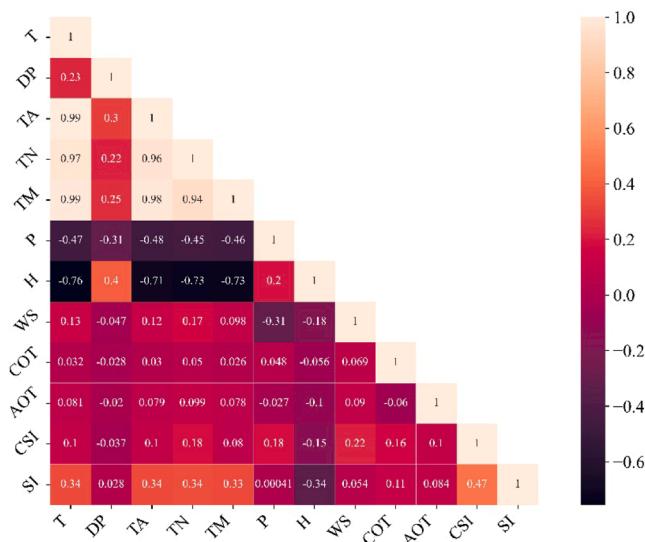


Fig. 14. Heatmap of PCC between target value (observed land surface solar irradiation) and the meteorological variables, AOT, COT, and CSI. (SI: solar irradiation, T: air temperature, H: humidity, WS: wind speed, P: atmospheric pressure, DP:dew point, TN: minimum temperature, TM: maximum temperature, and TA:apparent temperature).

is similar, which is peaking at about 4 pm (18th time step) on the second day and is bottom at about 4 pm (27th time step) on the third day. This indicates that the attention mechanism in the model exhibits a high level of concern for the first 18 time steps in each sample, shows low sensitivity from around the 18th time step to approximately the 27th time step, and then begins to display a higher degree of focus again in the final 9 time steps.

From this attention trend, we can observe that the model's attention mechanism effectively captures the diurnal periodicity of solar radiation changes. Specifically, the daily surface solar radiation exhibits similar variations. After learning consecutive two days of data, the model has preliminarily grasped the characteristics of daily solar radiation changes, leading to reduced attention to the data on the third day. Therefore, when the model learns the data for the fourth day, the existing parameters cannot accurately simulate this variability. Consequently, the model needs to re-adjust its weights to adapt to the data on the fourth day.

3.3. Generation of hourly land surface solar maps

We utilized the trained model to generate hourly land surface solar irradiation maps at a 5-km spatial resolution for Australia on October 1st, 2020, spanning from 8 am to 4 pm. The inputs for the trained model included COT, AOT, CSI, meteorological images, and static geographic locations, along with climate categories attribute information associated with these images. Specifically, the COT and AOT images were obtained from the Himawari-8 satellite, CSI images were calculated using Pysolar, and a set of meteorological images were acquired through the Kriging interpolation method. The results are presented in Fig. 17. In general, solar irradiation progressively increases from 8 am to 12 pm, reaching its peak at noon. Subsequently, it gradually decreases, and by 4 pm, solar irradiation is at its lowest. Furthermore, solar irradiation gradually decreases from northwest areas to southeast in each hourly map. These hourly solar irradiation maps provide precise solar power generation forecast data. These forecast data can be utilized for real-time grid operation adjustments to ensure supply-demand balance.

Additionally, to evaluate the estimation accuracy of the generated maps, we calculated the absolute error between estimated values and measured values in 13 stations. The result shows in Fig. 18. It is noticed

that most of errors are less than 0.08 kWh/m^2 , suggesting the proposed method has the good estimation accuracy for mapping the hourly solar irradiation in the large-scale region. Moreover, distinct error distributions were observed across different stations, likely due to variations in data quality. Some stations exhibited high levels of missing ground truth data, leading to the use of temporal mean imputation during data pre-processing, which might have introduced errors. For example, higher error distribution appears in Cocos Island station. The station with the highest missing rate in observed solar irradiation data (complete absence from September to December 2019) was processed using temporal mean imputation, where missing values were replaced by the average of corresponding periods in other years. This interpolation approach likely introduced bias, contributing to the significantly higher prediction errors observed at this station compared to others.

Moreover, The TFT model was employed to calculate the annual total land surface solar irradiation map, as shown in Fig. 19(a). Specifically, the total land surface solar irradiation for the 15th day of each month in 2020 was first calculated. This value was then used as the monthly average to estimate the total solar irradiation for each month. By summing the values for all twelve months, a rough estimate of the annual total solar irradiation was obtained, as shown in Figure 19(a). The distribution indicates that Australia's overall solar irradiation is high, with the northern regions receiving slightly higher levels than the southern regions. To evaluate the accuracy of the generated map with the published map by Solargis [47], Fig. 20(a) shows the spatial distribution of the annual land surface solar irradiation map from Solargis. A comparison of the quantitative ranges and distribution patterns shows that our results are consistent with the published map. Additionally, we calculated the absolute error map between the generated map using our method and the map from Solargis. As shown in Fig. 20(b), the majority of error values fall within the $0\text{--}40 \text{ kWh/m}^2$ range, with only a few values in the $160\text{--}200 \text{ kWh/m}^2$ range. Overall, the error values are relatively small, further demonstrating the high accuracy of our estimated method.

To evaluate the generalization capability of our model, we also conducted a leave-one-station-out validation, where involved excluding a station during the training phase and subsequently using it for testing. This validation process was carried out for all stations, and the results are detailed in Table 5. The results demonstrate a high level of estimation accuracy based on R^2 , MAE, and RMSE, indicating that the approach employed in this study effectively addresses prediction challenges related to spatio-temporal data. Furthermore, we applied our model to estimate annual solar irradiation in Japan, and the results are shown in Fig. 21. We also compared the generated map to the published map from Solargis. From Fig. 21(b), it is found that the absolute errors are small, indicating the map generating by our model has high accuracy.

Moreover, the analysis of solar resource distribution across different states was conducted based on their respective geographical areas, as shown in Figure 18(b). The Northern Territory, situated close to the equator, exhibited the highest average annual solar irradiation of 2012.34 kWh/m^2 , representing 17 % of the total national solar irradiation resources. In contrast, Tasmania recorded the lowest mean annual solar irradiation levels at 1238.84 kWh/m^2 , contributing only 0.94 % to the overall national solar irradiation resources. These findings underscore the significant solar energy potential in the northern regions of Australia, suggesting targeted solar energy development initiatives could be beneficial. This data can inform policymakers in designing and implementing focused solar energy projects in these promising northern areas.

4. Discussion and conclusion

This study proposes the temporal fusion transformer-augmented GeoAI framework to provide a highly accurate and interpretable estimation of hourly land surface solar irradiation. This framework employs

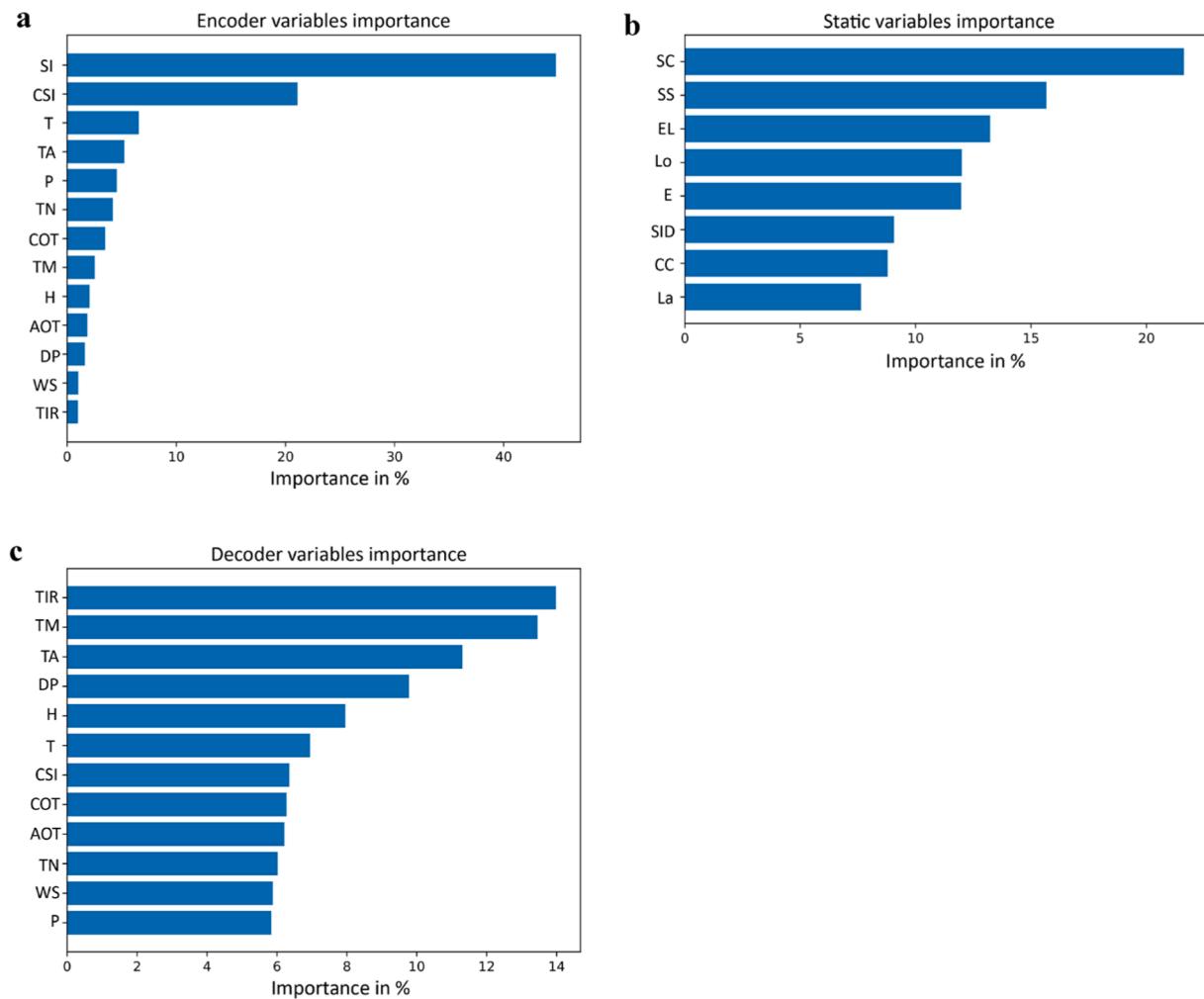


Fig. 15. The importance of variables in the different layers. (a) Encoder variables importance; (b) static variables importance; (c) Decoder variables importance. (SI: solar irradiation, T: air temperature, H: humidity, WS: wind speed, P: atmospheric pressure, DP: dew point, TN: minimum temperature, TM: maximum temperature, TA: apparent temperature, TIR: relative time index, SC: solar_{center} (normalized variable), SS: solar_{scale} (normalized variable), EL: encoder_{length}, La: latitude, Lo: longitude, SID: station ID, CC: climate categories, E: elevation).

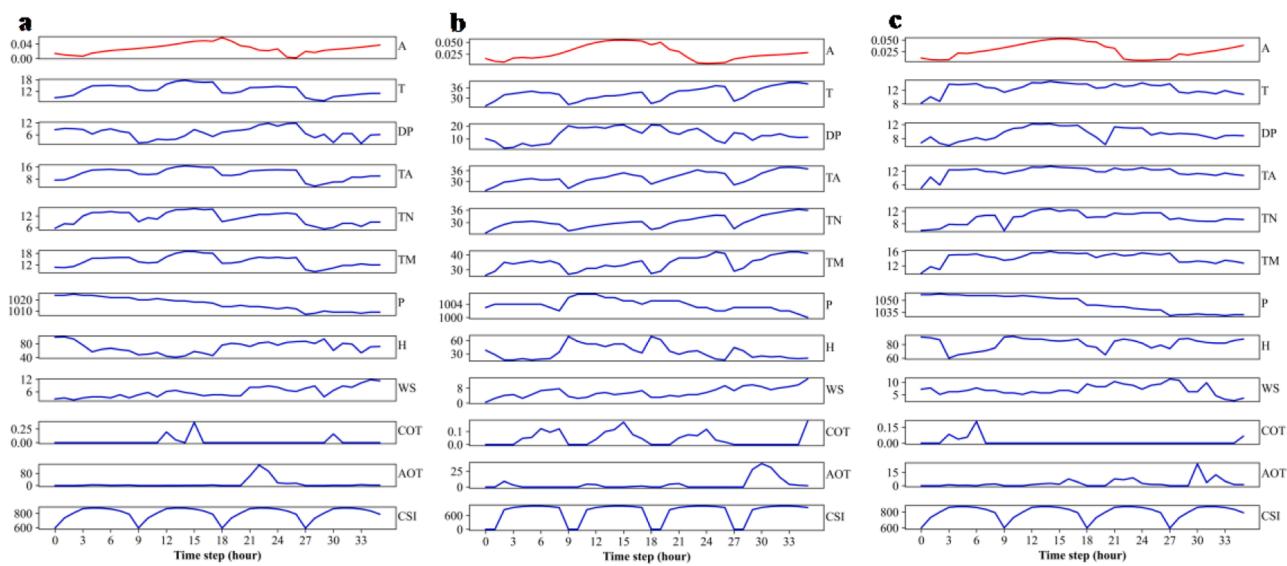


Fig. 16. Three selected typical attention values of TFT and values of all temporal input variables from three periods (A: attention weight). (a) four days from January 1st, 2019 to January 4th, 2019; (b) four days from May 1st, 2019 to May 4th, 2019; (c) four days from September 1st, 2019 to September 4th, 2019.

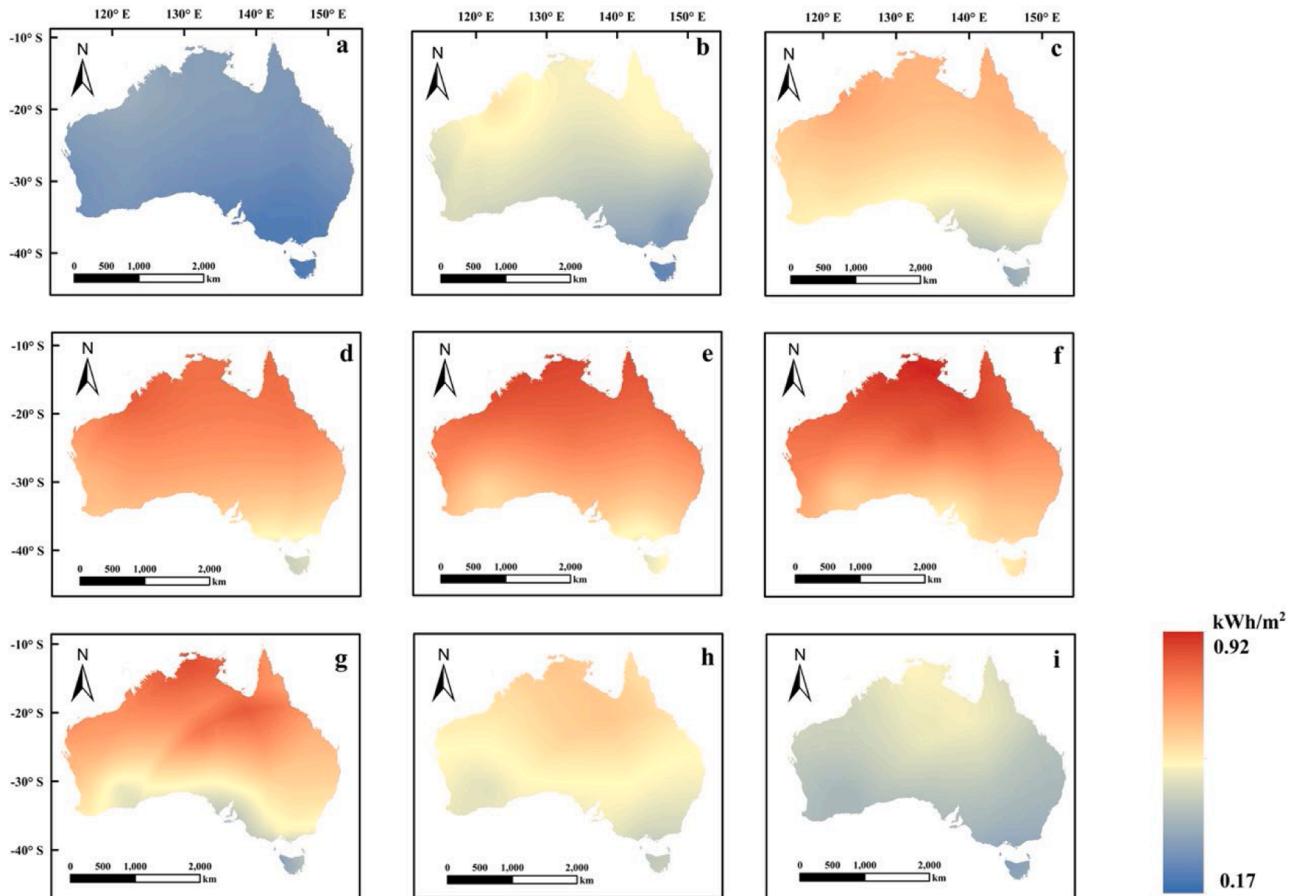


Fig. 17. The estimated hourly land surface solar map from 8 am to 4 pm in Australia. (a) hourly solar map on 8 am; (b) hourly solar map on 9 am; (c) hourly solar map on 10 am; (d) hourly solar map on 11 am; (e) hourly solar map on 12 am; (f) hourly solar map on 1 pm; (g) hourly solar map on 2 pm; (h) hourly solar map on 3 pm; (i) hourly solar map on 4 pm.

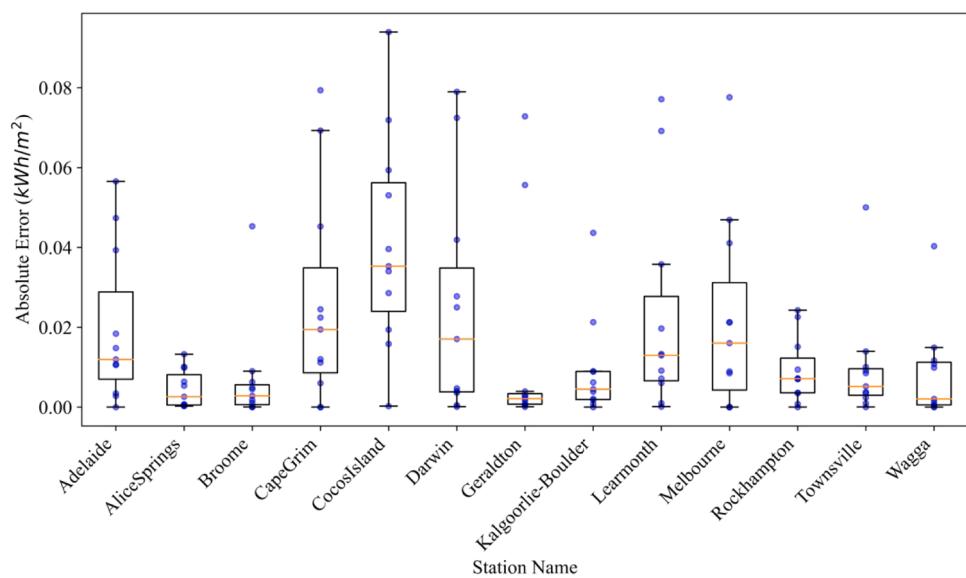


Fig. 18. Boxplot of the absolute error between estimated values and measured values in 13 stations from 8 am to 4 pm.

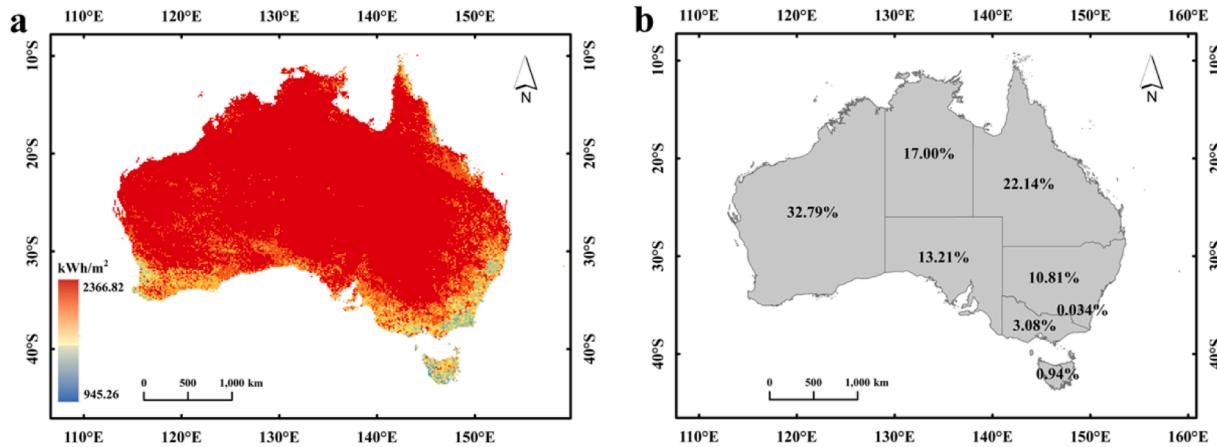


Fig. 19. Annual solar radiation resources in 2020 in Australia. (a) Spatial distribution of annual land surface solar irradiation across Australia. (b) Share of solar irradiation resources in states.

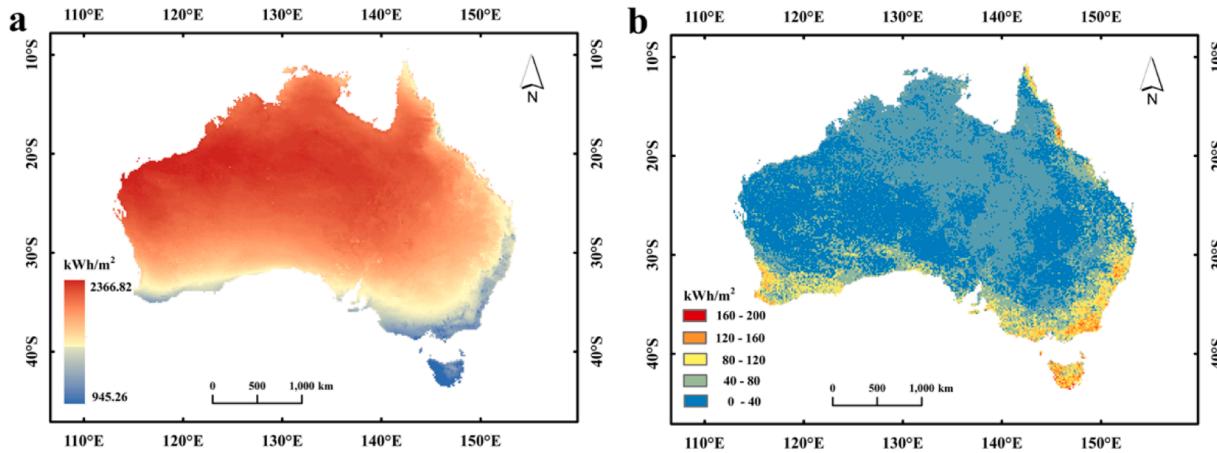


Fig. 20. Accuracy Evaluation between the generated annual land surface solar irradiation map using our method and the annual map from Solargis in Australia. (a) annual land surface solar irradiation map from Solargis. (b) absolute error map between the generated map using our method and the map from Solargis.

Table 5

The results of leave-one-station-out validations for 13 stations.

Leave out station name	R ²	MAE (kWh/m ²)	RMSE (kWh/m ²)
Adelaide	0.89	0.076	0.10
Alice Springs	0.83	0.089	0.16
Broome	0.86	0.074	0.15
Cape Grim	0.90	0.025	0.063
Cocos Island	0.91	0.025	0.056
Darwin	0.90	0.036	0.076
Geraldton	0.86	0.096	0.17
Kalgoorlie-Boulder	0.82	0.11	0.16
Learmonth	0.87	0.099	0.12
Melbourne	0.84	0.11	0.15
Rockhampton	0.89	0.086	0.096
Townsville	0.86	0.10	0.11
Wagga	0.87	0.093	0.096

the state-of-the-art TFT deep learning model using spatio-temporal dataset constructed by GIS technology to explore the non-linear relationship between multi-source variables (i.e., COT, AOT, CSI, MIs, geographical coordinates, elevation, station name, and climate category) and land surface solar irradiation. As a case study in Australia, compared to benchmark models, the TFT model is the optimal method for estimating time series hourly land surface solar irradiation, which has super high accuracy with R², MAE, RMSE equaling 0.90, 0.25(kWh/m²), and 0.63(kWh/m²), respectively. Furthermore, this study provides

a method for generating continuous solar potential maps with high accuracy over large areas for the analysis of solar source distribution.

One of the advantages of using the TFT model to estimate solar irradiation is its ability to provide highly accurate estimation results. Specifically, looking at the results from Fig. 4 to Fig. 7, the estimation accuracy of the TFT model is higher than that of other benchmark models, whether applied to an entire dataset or 13 individual datasets. Examining the absolute estimation errors for the four seasons from Fig. 8 to Fig. 11, it can be seen that the estimation errors for each hour of each season are relatively small. Moreover, to validate the accuracy of using the TFT model to generate solar radiation maps, we calculated the estimation errors for each station, and it can be seen from Fig. 18 that the absolute errors are also relatively small.

The framework of GeoAI proposed in this study effectively addresses the issue of geographical heterogeneity that arises when estimating spatio-temporal data. Geographical heterogeneity refers to the problem of only considering the impact of temporal features of the data while neglecting the influence brought about by geographical spatial differences. In this framework, the TFT model we utilize can overcome this problem, as it can simultaneously learn both the temporal features of the dataset and the spatial geographic attributes. To validate the effectiveness of the TFT model in addressing geographical heterogeneity, we conducted a comparative experiment. We trained the TFT model using both the entire dataset, which includes geographical spatial information, and 13 individual datasets without such information. Results from

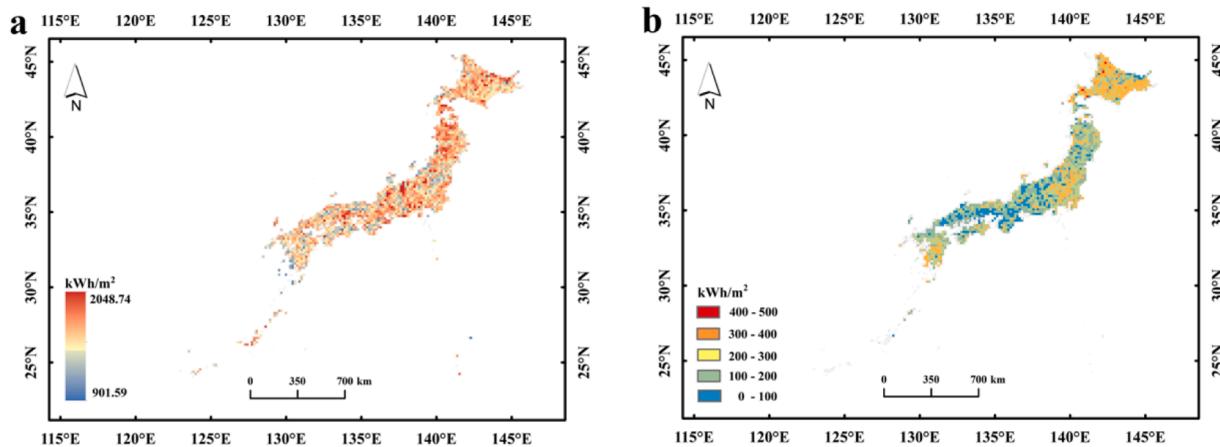


Fig. 21. The model transfer results in Japan. (a) Spatial distribution of annual land surface solar irradiation across Japan; (b) absolute error map between the generated map using our method and the map from Solargis.

Fig. 4 to **Fig. 7** indicate that although the estimation result of the TFT method using the entire dataset is slightly lower than the results obtained using the individual station datasets (approximately 61 % of stations), the values of R^2 , MAE, and RMSE strongly suggest that the TFT method can effectively learn geographic spatial features from the constructed multi-station spatio-temporal dataset and it is suitable for the large area estimation.

Furthermore, the TFT method successfully addresses the black-box limitation commonly associated with traditional machine learning methods. It achieves interpretability by calculating the importance scores of all variables and the attention weights for each time step. In our study, we observed that solar irradiation and CSI play significant roles in the Encoder, while the relative time index and variables related to temperature contribute more to the Decoder. The quantified importance scores of these variables enhance the transparency and interpretability of the TFT model. Additionally, the model calculates attention values for each step in past time, allowing for an understanding of the most crucial time steps that the TFT model focused on.

The one limitation of our study is the low resolution of the Himawari-8 satellite. In this study, the spatial resolution of the COT and AOT images is 5 km, indicating that each pixel represents a geographical area of 5 km × 5 km. However, although the COT and AOT values within a pixel unit are assumed to be identical in the images, the actual situation is that there is variability in the true COT and AOT within the geographical extent of a pixel. The values of COT and AOT are extracted from the corresponding pixels in the satellite images based on the geographical locations of the study sites. However, this extraction method may lead to discrepancies with actual site data, resulting in errors between the true and extracted values of COT and AOT. Apart from COT and AOT, the meteorological data, CSI data, and the observed land surface solar irradiation data are obtained from on-site measurements in the study, thereby avoiding such errors. These data discrepancies may introduce biases in the estimation of solar irradiation in the study.

Due to the high cost associated with acquiring high spatio-temporal resolution reanalysis data images, we utilized the low-cost Kriging interpolation method to produce spatially continuous meteorological images. Although the rough interpolation result appears to slightly reduce the accuracy of the solar map estimation, the reliability of our method is supported by the absolute error analysis presented in **Fig. 18** and the comparison with a published map. These findings suggest that the impact of the rough interpolation result on the overall estimation performance is minimal, thereby affirming the credibility of our generated solar map.

Furthermore, the transferring capability of this model is represented in the accessibility of data. In our study, we found that meteorological factors are important input parameters for estimating solar irradiation. The reliance on non-free hourly meteorological data could increase the challenge of applying the TFT model in other regions, where obtaining such data might pose difficulties. Additionally, although the coverage area of the Himawari-8 satellite images is limited, the other meteorological satellite images including cloud and aerosol data in other regions where Himawari-8 cannot cover can be used to replace the Himawari-8 satellite images.

This study provides a simple but high-accuracy model to estimate hourly land surface solar irradiation over large areas at low cost. The merits of low cost and high accuracy of this method enable it competitive. The high-accuracy, interpretable, and reliable resulting estimations hold significant value for policymakers, energy planners, and investors, contributing to the sustainable development of renewable energy and the realization of carbon neutrality goals.

CRediT authorship contribution statement

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

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.

Acknowledgements

This project is substantially funded by the General Research Fund (Grant No. 15603923 and 15603920), and the Collaborative Research Fund (Grant No. C5062–21GF) and Young Collaborate Research Fund (Grant No. C6003–22Y) from the Research Grants Council, Hong Kong, China. The authors acknowledge the funding support (Grant No. BBG2 and CD81) from the Research Institute for Sustainable Urban Development, Research Institute of Land and Space, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China.

Appendix A. Temporal fusion transformer algorithm

The Temporal Fusion Transformer (TFT) model [13] is a newly attention-based deep learning architecture specifically designed for handling multi-dimensional time series data, as shown in Fig. A.1. The inputs of model consist of past dynamic inputs, static inputs, and future dynamic inputs. The past dynamic inputs and future dynamic inputs were fed into temporal processing layers, including variable selection, LSTM layer, and gate layer, for learning the useful temporal features. The static inputs were fed into static layers to learn the significant static features to generate the static vectors, and these static vectors were integrated with the temporal features. After integrating temporal features with static features, Encoder output was concatenated with decoder output to be fed into the multi-head attention. After that, a series of layers were used to enhance the model performance.

Let I represents unique entities in land surface solar irradiation. Each entity i consists of static metadata s_i , time series inputs $X_{i,t}$, and solar targets $y_{i,t}$ at time step t , $t \in [0, T_i]$. Time series inputs $X_{i,t}$ can be classified into two categories, $X_{i,t} = [z_{i,t}^T, x_{i,t}^T]^T$. Past inputs $z_{i,t}$ denote that these variables can only be measured at each step and are unknown beforehand, and known future inputs $x_{i,t}$ represent they can be pre-determined and the value of these variables are known before time step t . The estimation function is defined as follows:

$$\hat{y}_i(t, \tau) = f(\tau, y_{i,t-k:t}, z_{i,t-k:t}, x_{i,t-k:t+\tau}, s_i) \quad (1)$$

Where $y_{i,t-k:t} = \{y_{i,t-k}, \dots, y_{i,t}\}$ denotes targets until the time t , $z_{i,t-k:t} = \{z_{i,t-k}, \dots, z_{i,t}\}$ denotes past inputs, $x_{i,t-k:t+\tau} = \{x_{i,t-k}, \dots, x_{i,t+\tau}\}$ denotes known future inputs across the full range, and τ represents the estimation time point.

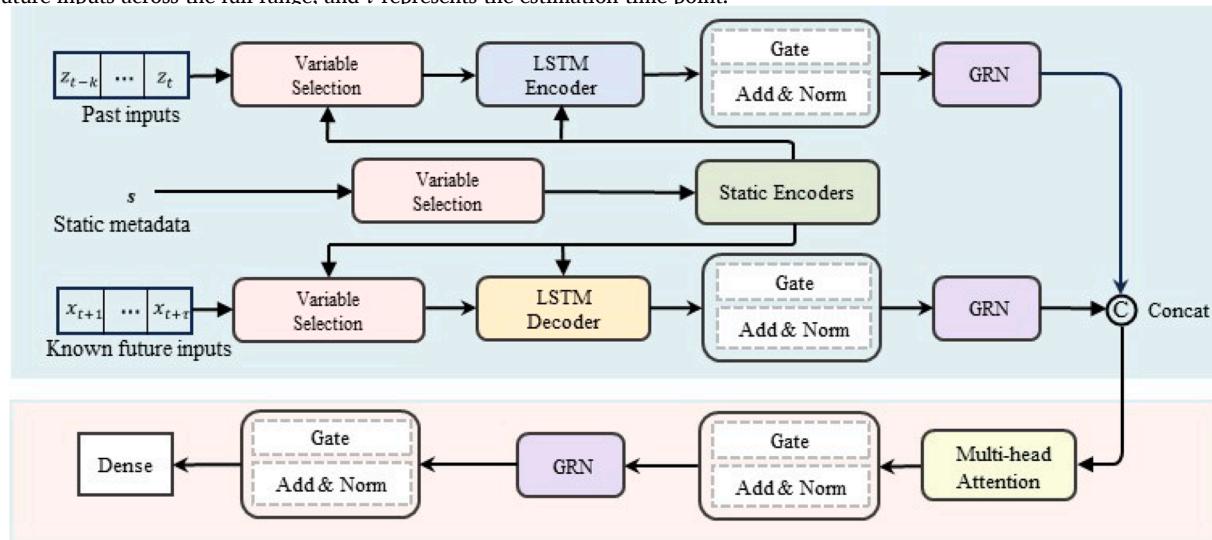


Fig. A.1. Temporal fusion transformer architecture.

A.1. Gating mechanisms

Previous studies [10,48] suggested that the estimation of solar irradiation could be affected by several factors, such as a series of meteorological factors, aerosol, and cloud coverage. The mechanism of machine learning methods is to construct the non-linear relationship between multi-source variables and the target. However, a complex and redundant variable dataset would make the model more complex, which leads to an increase in the computation time. Therefore, the selection of reasonable variables is essential for building an effective model to estimate land surface solar irradiation.

Gating mechanisms can filter out unnecessary components of the architecture and can be flexibly applied to non-linear processing only where needed. To achieve this, Gated Residual Network (GRN) is used as a building block of TFT. The architecture of the GRN is demonstrated in Fig. A.2. The GRN consists of two inputs, namely, a primary a and an optional context vector c . The GRN is described as follows:

$$GRN_\omega(a, c) = LayerNorm(a + GLU_\omega(\eta_1)) \quad (2)$$

$$\eta_1 = W_{1,\omega}\eta_2 + b_{1,\omega} \quad (3)$$

$$\eta_2 = ELU(W_{2,\omega}a + W_{3,\omega}c + b_{2,\omega}) \quad (4)$$

$$GLU_\omega(\gamma) = \sigma(W_{4,\omega}\gamma + b_{4,\omega}) \odot (W_{5,\omega}\gamma + b_{5,\omega}) \quad (5)$$

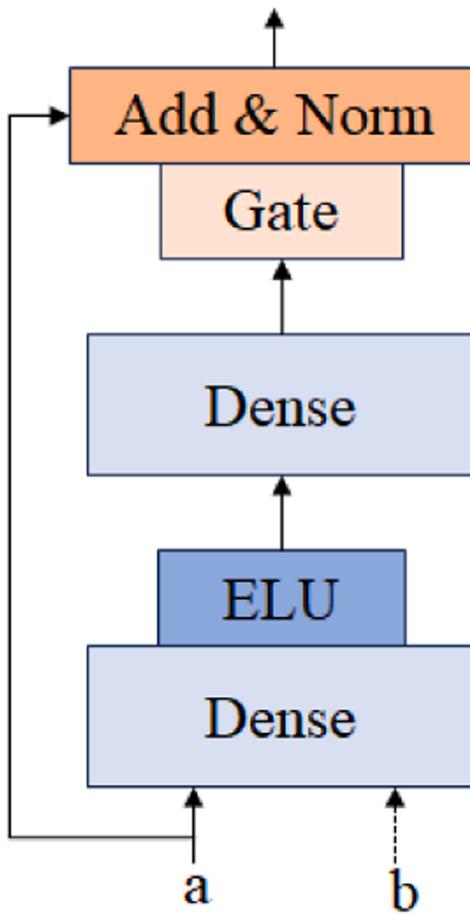


Fig. A.2. GRN architecture.

where LayerNorm is standard layer normalization η_1 and η_2 are intermediate layers, ϖ is an index to represent weight sharing, $W_{(\cdot)}$ and $b_{(\cdot)}$ denote the weights and biases, ELU is the exponential linear unit activation function, \odot is the element-wise Hadamard product, and $\sigma_{(\cdot)}$ is the sigmoid activation function. GLU enables TFT to regulate the degree to which the GRN contributes to the original input, potentially by passing the entire layer if required. This can occur when the GLU outputs are predominantly close to 0, effectively suppressing the nonlinear contribution.

A.2. Variable selection network

The variable selection network not only assesses the importance of each variable for selecting relevant input variables but also enables TFT to reduce variables that demonstrate a detrimental effect on the estimation performance.

The static and time-series continuous variables are transformed into feature representations and dimensional vectors, respectively. Let $\xi_t^{(j)}$ represent the transformed j th variable at time t . At each time step, each $\xi_t^{(j)}$ goes through its own GRN, as shown in Eq. (6). $\tilde{\xi}_t^{(j)}$ is the corresponding processed feature vector for the j th variable. All past inputs are transformed into flattened vectors $[I]_t = \left[\xi_t^{(1)^T}, \dots, \xi_t^{(m_\chi)^T} \right]^T$. The variable selection weights v_{χ_t} are calculated using Eq. (7), where c_s denote an external context vector. Finally, the processed features are weighted by their variable selection weights and combined as shown in Eq. (8).

$$\tilde{\xi}_t^{(j)} = GRN_{\tilde{\xi}_{(j)}}(\xi_t^{(j)}) \quad (6)$$

$$v_{\chi_t} = \text{Softmax}(GRN_{v_\chi}([I]_t, c_s)) \quad (7)$$

$$\tilde{\xi}_t = \sum_{j=1}^{m_\chi} v_{\chi_t}^{(j)} \tilde{\xi}_t^{(j)} \quad (8)$$

A.3. Static covariate encoder

In contrast to other time-series deep learning methods, the static covariate encoder is designed to integrate static features extracted from static metadata into TFT network. Individual GRN encoders are utilized to generate four different context vectors, namely, c_s , c_e , c_c , and c_h . c_s is the context vector, which is from static covariate encoder, (c_c, c_h) is locally process temporal features for to initialize the cell state and hidden state respectively for the first LSTM in the layer, and c_e is the Enrichment of temporal features. These four context vectors are integrated with temporal features in the TFT network.

A.4. Interpretable multi-head attention module

TFT employs a self-attention mechanism that modify from the multi-head attention mechanism in the standard transformer architectures proposed by Vaswani et al. [49]. This modification enables TFT to capture long-term relationships across various time steps and enhances explainability. The Attention mechanism is described as follows:

$$\text{Attention}(Q, K, V) = A(Q, K)V \quad (9)$$

$$A(Q, K) = \text{Softmax}\left(QK^T / \sqrt{d_{\text{attn}}}\right) \quad (10)$$

where Q is the “query”, K is the “key”, V is the “value”, and $A(\cdot)$ denotes a normalization function. In general, the multi-head attention mechanism is defined in Eq. (11), (12).

$$\text{MultiHead}(Q, K, V) = [H_1, \dots, H_{mH}]W_H \quad (11)$$

$$H_h = \text{Attention}\left(QW_Q^{(h)}, KW_K^{(h)}, V W_V^{(h)}\right) \quad (12)$$

where $W_Q^{(h)}$, $W_K^{(h)}$, $W_V^{(h)}$ denote head-specific weights for queries, keys, and values, respectively. Additionally, W_H is the combination of outputs concatenated from all heads H_h .

The values learned in each head using multi-head attention are different, so attention weights would not represent the importance of specific variables which enables the model to decrease explainability. To enable the attention weights as the unique value for each variable, the TFT model modifies multi-head attention to share values learned from each head and uses additive aggregation of all heads. The Interpretable multi-head attention is defined in Eq. (13), (14).

$$\text{InterpretableMultiHead}(Q, K, V) = \tilde{H}W_H \quad (13)$$

$$\tilde{H} = \tilde{A}(Q, K)VW_V = 1/H \sum_{h=1}^{mH} \text{Attention}\left(QW_Q^{(h)}, KW_K^{(h)}, V W_V^{(h)}\right) \quad (14)$$

A.5. Loss functions

TFT is trained through joint minimization of quantile loss [50]. The outputs of all quantiles are then added, with the formula given as follows:

$$L(\Omega, W) = \sum_{y_t \in \Omega} \sum_{q \in \varrho} \sum_{T=1}^{T_{\max}} \frac{QT(y_t, \hat{y}(q, t-T), q)}{MT_{\max}} \quad (15)$$

$$QL(y, \hat{y}, q) = q(y - \hat{y})_+ + (1-q)(\hat{y} - y)_+ \quad (16)$$

Where Ω is the domain of training data containing samples, ϱ is the set of output quantiles, W is the weights of TFT, and $(.)_+$ means $\max(0, .)$.

Data availability

Data will be made available on request.

References

- [1] Wang F, Harindintwali JD, Yuan Z, Wang M, Wang F, Li S, Yin Z, Huang L, Fu Y, Li L. Technologies and perspectives for achieving carbon neutrality. *Innovation* 2021;2:4.
- [2] Dutta A, Farooq S, Karimi IA, Khan SA. Assessing the potential of CO₂ utilization with an integrated framework for producing power and chemicals. *J CO₂ Util* 2017;19:49–57.
- [3] Cardoso A, Jurado-Rodríguez D, López A, Ramos MI, Jurado JM. Automated detection and tracking of photovoltaic modules from 3D remote sensing data. *Appl Energy* 2024;367:123242.
- [4] Sahu BK. A study on global solar PV energy developments and policies with special focus on the top ten solar PV power producing countries. *Renew Sustain Energy Rev* 2015;43:621–34.
- [5] Zhu R, Kwan MP, Perera A, Fan H, Yang B, Chen B, Chen M, Qian Z, Zhang H, Zhang X. GIScience can facilitate the development of solar cities for energy transition. *Adv Appl Energy* 2023;100129.
- [6] Gürel AE, Bulut U, Ag, Bakır H, Ergun A, Yıldız G. A state of art review on estimation of solar radiation with various models. *Heliyon* 2023;9:2.
- [7] Besharat F, Dehghan AA, Faghih AR. Empirical models for estimating global solar radiation: a review and case study. *Renewable Sustain Energy Rev* 2013;21:798–821.
- [8] Hassan GE, Youssef ME, Mohamed ZE, Ali MA, Hanafy AA. New temperature-based models for predicting global solar radiation. *Appl Energy* 2016;179:437–50.
- [9] Mohammadi K, Shamshirband S, Anisi MH, Alam KA, Petković D. Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Conversion Manag* 2015;91:433–41.
- [10] Manzano A, Martín M, Valero F, Armenta C. A single method to estimate the daily global solar radiation from monthly data. *Atmos Res* 2015;166:70–82.
- [11] Ahmed MS, Guo H, Tanino K. Cloud cover-based models for estimation of global solar radiation: a review and case study. *Int J Green Energy* 2022;19:175–89.
- [12] Bayrakçı HC, Demircan C, Keçebaş A. The development of empirical models for estimating global solar radiation on horizontal surface: a case study. *Renew Sustain Energy Rev* 2018;81:2771–82.

- [13] Sarkar MNI. Estimation of solar radiation from cloud cover data of Bangladesh. *Wind Water Sol energy* 2016;3:1–15.
- [14] Fan J, Wu L, Zhang F, Cai H, Ma X, Bai H. Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China. *Renew Sustain Energy Rev* 2019; 105:168–86.
- [15] Feng Y, Gong D, Zhang Q, Jiang S, Zhao L, Cui N. Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. *Energy Convers Manag* 2019;198:111780.
- [16] Chen JL, He L, Yang H, Ma M, Chen Q, Wu SJ, Zi Xiao. Empirical models for estimating monthly global solar radiation: a most comprehensive review and comparative case study in China. *Renew Sustain Energy Rev* 2019;108:91–111.
- [17] Shadab A, Ahmad S, Said S. Spatial forecasting of solar radiation using ARIMA model. *Remote Sens Appl: Soc Environ* 2020;20:100427.
- [18] Belmahi B, Louzazni M, El Bouardi A. One month-ahead forecasting of mean daily global solar radiation using time series models. *Optik* 2020;219:165207.
- [19] 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.
- [20] Ismail MT, Karim SA A. Time series models of high frequency solar radiation data. *Pract Ex Energy Optim Models* 2020;1:79–89.
- [21] 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.
- [22] Yildirim A, Bilgili M, Ozbek A. One-hour-ahead solar radiation forecasting by MLP, LSTM, and ANFIS approaches. *Meteorol Atmos Phys* 2023;135:10.
- [23] Zang H, Cheng L, Ding T, Cheung KW, Wang M, Wei Z, Sun G. Application of functional deep belief network for estimating daily global solar radiation: a case study in China. *Energy* 2020;191:116502.
- [24] Ismail MS, Moghavvemi M, Mahlia T. Analysis and evaluation of various aspects of solar radiation in the Palestinian territories. *Energy Convers Manag* 2013;73: 57–68.
- [25] Park JK, Das A, Park JH. A new approach to estimate the spatial distribution of solar radiation using topographic factor and sunshine duration in South Korea. *Energy Convers Manag* 2015;101:30–9.
- [26] Liu P, Biljecki F. A review of spatially-explicit GeoAI applications in Urban Geography. *Int J Appl Earth Obs Geoinf* 2022;112:102936.
- [27] Huang X, Liu J, Xu S, Li C, Li Q, Tai Y. A 3D ConvLSTM-CNN network based on multi-channel color extraction for ultra-short-term solar irradiance forecasting. *Energy* 2023;272:127140.
- [28] Gao Y, Miyata S, Akashi Y. Interpretable deep learning for hourly solar radiation prediction: a real measured data case study in Tokyo. *J Build Eng* 2023;79: 107814.
- [29] Gao Y, Miyata S, Akashi Y. Interpretable deep learning models for hourly solar radiation prediction based on graph neural network and attention. *Appl Energy* 2022;321:119288.
- [30] Xiao Z, Gao B, Huang X, Chen Z, Li C, Tai Y. An interpretable horizontal federated deep learning approach to improve short-term solar irradiance forecasting. *J Clean Prod* 2024;436:140585.
- [31] Gao Y, Miyata S, Matsunami Y, Akashi Y. Spatio-temporalinterpretableneuralnetwork for solar irradiation prediction using transformer. *Energy Build* 2023;297:113461.
- [32] Deo RC, Şahin M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renew Sustain Energy Rev* 2017;72:828–48.
- [33] Digital elevation model in Australia. <https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search#/metadata/66006>; 2024 [Accessed 25 March 2024].
- [34] Dahmani K, Notton G, Voyant C, Dizene R, Nivet ML, Paoli C, Tamas W. Multi-layer perceptron approach for estimating 5-min and hourly horizontal global irradiation from exogenous meteorological data in locations without solar measurements. *Renew Energy* 2016;90:267–82.
- [35] Biazar SM, Rahmani V, Isazadeh M, Kisi O, Dinipashoh Y. New input selection procedure for machine learning methods in estimating daily global solar radiation. *Arab J Geosci* 2020;13:1–17.
- [36] Deo RC, Wen X, Qi F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl Energy* 2016;168:568–93.
- [37] Rabehi A, Guermoui M, Lalmi D. Hybrid models for global solar radiation prediction: a case study. *Int J Ambient Energy* 2020;41:31–40.
- [38] 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.
- [39] JAXA Himawari monitor P-tree system. <https://www.eorc.jaxa.jp/ptree/>; 2021 [Accessed 15 December 2023].
- [40] Open weather. url, <http://reg.bom.gov.au/index.php/>; 2023 [Accessed 15 December 2023].
- [41] Pysolar. url, <https://pysolar.readthedocs.io/en/latest/>; 2021 [Accessed 15 December 2023].
- [42] Liao X, Wong MS, Zhu R. Dual-gate temporal fusion transformer for estimating large-scale land surface solar irradiation. *Renew Sustain Energy Rev* 2025;214: 115510.
- [43] 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.
- [44] Lim B, Arik SO, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int J Forecast* 2021;37:1748–64.
- [45] Cohen I, Huang Y., Chen J., Benesty J., Benesty J., Chen J., Huang Y., Cohen I. Pearson correlation coefficient. Noise reduction in speech processing 2009; 1–4.
- [46] Hutter F, Kotthoff L, Vanschoren J. *Automated Machine Learning: Methods, Systems, Challenges*. Springer Nature; 2019.
- [47] Solar maps of Solargis in Australia; 2024. <https://solargis.com/maps-and-gis-data/download/australia> [Accessed 25 May 2024].
- [48] Huang Z, Zhu R, You L. Unravelling the effects of dynamic urban thermal environment on utility-scale floating photovoltaic electricity generation. *Sustain Cities Society* 2023;99:104964.
- [49] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst* 2017;30:1.
- [50] Wen R., Torkkola K., Narayanaswamy B., Madeka D. A multi-horizon quantile recurrent forecaster. *arXiv preprint arXiv* 2017; 11053.