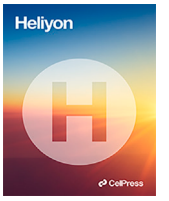
Heliyon 9 (2023) e18097

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

On predicting annual output energy of 4-terminal perovskite/silicon tandem PV cells for building integrated photovoltaic application using machine learning

Dong C. Nguyen [a,b,*], Yasuaki Ishikawa [a,*]

[a] College of Science and Engineering, Aoyama Gakuin University, Sagamihara, Kanagawa 252-5258, Japan
[b] Institute of Materials Science, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Cau Giay, Hanoi 100000, Viet Nam

ARTICLE INFO

Keywords:
Building integrated photovoltaic
Tandem PV cell
Atlas simulation
Machine learning
Photovoltaic
Surrogate algorithm

ABSTRACT

Building integrated photovoltaic (BIPV), based on tandem PV cells, is considered a new alternative for combining solar energy with buildings. Accurately predicting the BIPV-harvested annual output energy ($E_{out,annual}$) is crucial for evaluating the BIPV performance. Machine learning (ML) is a potential candidate for solving such a problem without the time-consuming process of experimental investigations. This contribution proposes an artificial neural network (ANN) to predict the $E_{out,annual}$ of 4-terminal perovskite/silicon (psk/Si) PV cells under realistic environmental conditions. The input variables of the proposed model consist of the input solar irradiance (P_{in}), incident light's angle (A_{in}), the PV module's temperature (T_{mod}), the psk absorber's thickness (Th_{psk}), and the psk absorber's bandgap (B_{psk}). The input data were received from the simulated results. This work also evaluates the degree of importance of each input variable and optimizes the architecture of the ANN using the surrogate algorithm before predictions. The optimized ANN-3 (three hidden layers) model shows superior performance indicators, including a mean squared error of MSE = 0.02283, correlation coefficient R = 0.99999, and Willmott's index of agreement I_w = 0.99999. Consequently, the predicted highest $E_{out,annual}$ at B_{psk} of 1.71 eV is 297.73, 115.01, 193.98, and 97.6 kWh/m^2 for the rooftop, east, south, and west facades, respectively.

1. Introduction

A building-integrated photovoltaic (BIPV) system has been designed based on the basic requirements of construction works to replace conventional construction products, such as rooftops, walls, windows, facades, and so on [1]. The BIPV systems could generate electricity for buildings while protecting them from adverse climatic conditions [2,3]. In addition, BIPV is considered a promising option that contributes to meeting the carbon-neutral energy system. Although BIPV has appeared since the early 1990s [4], its growth rate is relatively slow. By 2030, it is estimated that the BIPV system could achieve an approximate efficiency level of 22% [5]. So far, two PV cell categories in the market are composed of crystalline silicon (c-Si) wafer-based technology and thin-film technology (amorphous silicon (a-Si), chalcogen, and organic PV cells). The primary technology for BIPV is based on silicon wafers due to technological advancements and low cost of production [6]. The use of perovskite/silicon (psk/Si) tandem PV cells, which

* Corresponding authors at: College of Science and Engineering, Aoyama Gakuin University, Sagamihara, Kanagawa 252-5258, Japan.
E-mail addresses: dongnc@ee.aoyama.ac.jp (D.C. Nguyen), yishikawa@ee.aoyama.ac.jp (Y. Ishikawa).

https://doi.org/10.1016/j.heliyon.2023.e18097
Received 31 March 2023; Received in revised form 4 July 2023; Accepted 6 July 2023

Available online 13 July 2023

Nomenclature

Lists of Abbreviations

2T	2-terminal
4T	4-terminal
ANN	Artificial neural network
ANN-1	One-hidden-layer artificial neural network
ANN-2	Two-hidden-layer artificial neural network
ANN-3	Three-hidden-layer artificial neural network
ANN-4	Four-hidden-layer artificial neural network
ANN-5	Five-hidden-layer artificial neural network
BIPV	Building-integrated photovoltaics
c-Si	Crystalline silicon
DM	Data mining
FL	Fuzzy logic
GPR	Gaussian process regression
MAE	Mean absolute error
ML	Machine learning
MSE	Mean squared error
RF	Random forest
RMSE	Root mean squared error
SHJ	Silicon heterojunction
SI	Supplementary information
STC	Standard test conditions
SVR	Support vector regression

List of Symbols

η	Power conversion efficiency
A_{in}	Incident light's angle
B_{psk}	psk absorber's bandgap
$E_{in,annual}$	Annual input energy
$E_{out,annual}$	Annual output energy
I_w	Willmott's index of agreement
P_{in}	Input irradiance
$P_{out,bot}$	Bottom-cell output power density
$P_{out,tandem}$	Tandem-cell output power density
$P_{out,top}$	Top-cell output power density
T_{mod}	PV module's temperature
Th_{psk}	psk absorber's thickness
b	Bias
R	Correlation coefficient
w	Weight

have a theoretical efficiency potential of up to 44%, is considered a viable choice for the BIPV system [7]. So far, record efficiency is 30.1% for a 4-terminal (4T) psk/Si tandem PV cell [8] and 33.7% for a 2-terminal (2T) psk/Si tandem PV cell [9]. Although the 4T tandem PV cell has a record efficiency not so high as the 2T tandem PV cell, the advantages of the 4T tandem PV cell compared to the 2T tandem PV cell are easily processable and have no current matching constraint. Hence, the 2T and 4T tandem PV cells are emerging as good substitutes for the traditional c-Si PV cell in the context of the BIPV system.

Some recent works presented numerical methods for estimating the annual output energy ($E_{out,annual}$) of the 2T psk/Si tandem PV cell for the BIPV under natural operating conditions composed of input solar irradiance (P_{in}), incident light's angle (A_{in}), PV module's temperature (T_{mod}). The first method used one-dimensional optical simulation (GenPro4 program) combined with electrical calculations based on the one-diode model [10]. The second method employed two-dimensional optical and electrical simulations (Atlas framework) [11]. Because of the incredibly long simulation duration (several years) and a lack of complex refractive index (n,k) data for a continuous psk absorber's bandgap range, the $E_{out,annual}$ estimation for the continuous psk absorber's bandgap range is impossible. So, both these methods were conducted for a sole psk absorber's bandgap. The third method utilized the combination of Atlas simulation and an artificial neural network (ANN) model [12], which could overcome the shortcoming of both methods mentioned above. However, no previous investigation has presented estimating the annual output energy ($E_{out,annual}$) of the 4T psk/Si tandem PV cells for the BIPV under realistic environmental conditions. It is, therefore, crucial to precisely simulate and estimate the $E_{out,annual}$ of the 4T tandem PV cells to improve their efficiency and assess the tandem-based BIPV performance. Furthermore, the estimation is also necessary to compare the efficiency of psk/Si tandem PV cells with the different numbers of terminals.

This work presents a $E_{out,annual}$ predictive method of the 4T psk/Si tandem PV cell using an ANN based on the atlas simulation results. Different machine learning (ML) techniques can provide predictive solutions with superior accuracy for various applications in renewable energy fields, including wind energy, power generation from energy, power demand, etc. In numerous investigations, ML models have been performed extensively for PV power generation predictions based on solar irradiance. The applied ML models are very diverse and include fuzzy logic (FL) models [13–15], artificial neural network (ANN) models [15–17], support vector regression (SVR) models [18–20], Gaussian process regression (GPR) models [21,22], random forest (RF) models [23–27]. According to the survey of the available literature, the authors propose that ANNs are simple but provide superior performance in terms of the estimation error for predicting power generation from solar energy. Therefore, the authors deploy the ANN for estimating the generation of the annual output energy ($E_{out,annual}$) of 4T tandem PV cells for the BIPV under realistic environmental conditions. Determining the optimal architecture of an ANN relies heavily on determining the number of hidden layers and their corresponding neurons. This step is crucial as it directly impacts the network's performance. Nevertheless, arriving at the optimal architecture remains a complex and challenging problem. In previous studies, researchers have employed different methods to determine the topologies of ANNs, including approaches that rely solely on the number of input and output neurons [28,29], trial and error [30,31], and the rule of thumb [32,33]. However, these methods were solely conducted for one or two hidden layers. In this work, the authors utilize the surrogate algorithm for optimizing the ANN architecture with various hidden layers up to five. Fig. 1 depicts this work's general framework for predicting the tandem-based BIPV $E_{out,annual}$ under realistic environmental conditions.

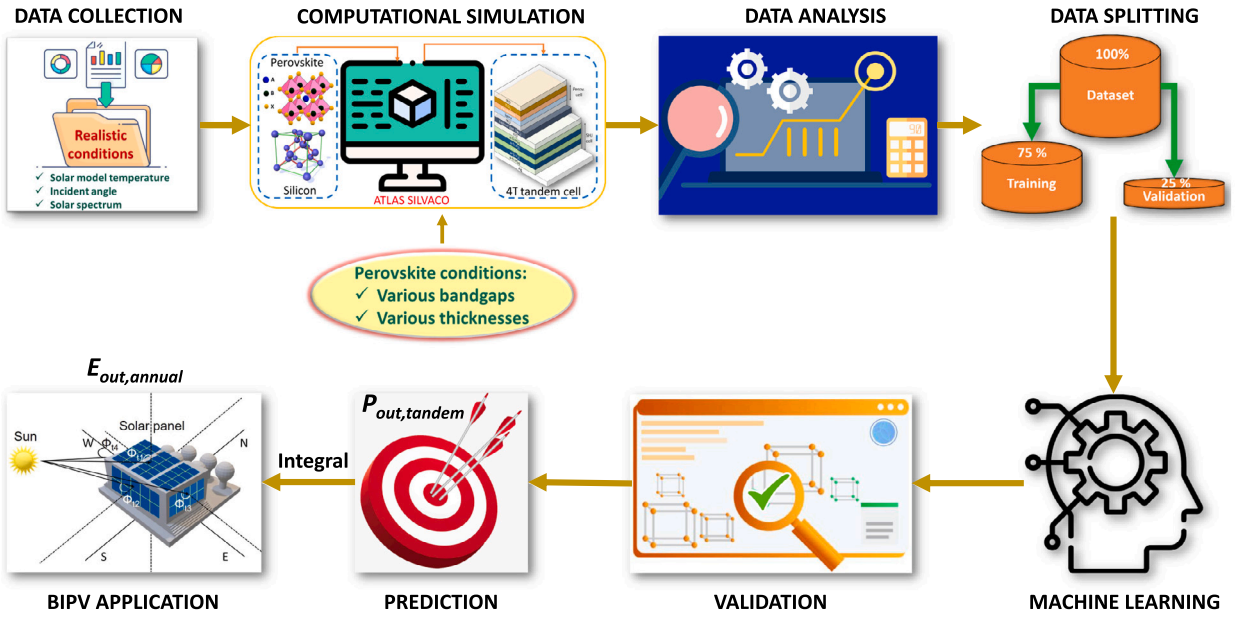


Fig. 1. The general framework of the intelligent system for predicting the tandem $E_{out,annual}$ for the BIPV.

2. Methodology

2.1. 4T psk/Si tandem PV cell architecture and its simulation method

The physical architecture of the 4T psk/Si tandem PV cell includes the top (psk PV cell) and bottom (silicon heterojunction (SHJ) PV cell) sub-cells separated by an optical gap, as demonstrated in Fig. 2. The psk sub-cell includes a thin film's lithium fluoride (LiF) as an anti-reflective coating, a thin film's indium tin oxide (ITO) as a front contact, a thin film's titanium oxide (TiO_2) as an electron transport layer, a psk absorber, a thin film's spiro-OMeTAD as a hole transport layer, and an ITO rear contact layer. The SHJ sub-cell comprises a front contact's ITO thin film, n^+ doped and intrinsic (i) hydrogenated a-Si layers, the c-Si wafer, i, and p^+ doped a-Si layers and a rear contact's silver thin film. The optical gap between the psk and SHJ sub-cells is a silicon nitride (SiN_x) thin film of 16 μm (instead of air gap), which separates the two sub-cells electrically and manages near-infrared (NIR) light absorption between the two sub-cells [34]. The simulation results presented the outstanding advantage of the SiN_x optical gap in enhancing the tandem PV cell performance compared to the air gap, as shown in Fig. S.2 (SI). In this study, the authors assumed that all 4T tandem PV cell interfaces were optically flat and without any surface imperfections.

The (n,k) data used to simulate the 4T tandem PV cells were obtained from previously published literature sources: LiF [35], ITO [36], TiO_2 [36], the psk absorber with various bandgaps (1.55, 1.62, 1.67, 1.70, and 1.73 eV) [37,38], Spiro-OMeTAD [36], p^+-a-Si [39], a-Si(i) [39], c-Si(n) [40], n^+-a-Si [39]. The SiN_x thin film's the (n,k) data (refractive index 2.71 at 600 nm) were estimated by an ellipsometric system for the film deposited by catalytic chemical vapor deposition and depicted in Fig. S.1 of the supplementary information (SI). The input parameters of the psk sub-cell layers [41,42], the SHJ cell's layers [43,44] in the 4T tandem PV cell are listed in Table S.1 (SI). Also, we took into consideration the characteristics of defect states at both the TiO_2/psk interface, as described in previous works [45–47], and the a-Si/c-Si interface, as reported in a study by Lu et al. [43] during our simulations (Table S.2 of SI). The other material parameters were set as default values in the Atlas module. Besides, the simulations utilized a fixed surface recombination velocity ($\upsilon_{surf,n} = \upsilon_{surf,p} = 10$ cm/s) for the a-Si layers. Series resistance (R_s) of the top and bottom sub-cells referenced from recent work [48] were set as 1.1 and 3.8 Ω. The Atlas simulations were performed on a Dell Precision 5820 Tower Workstation running Red Hat Linux version 7.9. The workstation was equipped with a 3.7 GHz Intel(R) Xeon(R) W-2135 processor and 64 GB DDR4 RAM, providing the computational resources necessary for the simulations. Various Atlas physical models for this work's numerical simulations could be referred to in the literature [11]. All simulated results examine the influence of parameters P_{in}, A_{in}, T_{mod}, the psk absorber's thickness (Th_{psk}), and the psk absorber's bandgap (B_{psk}) on the power conversion efficiency (η) of the 4T psk/SHJ tandem PV cells, as presented in Figure S.3 and 4 (SI).

2.2. Data collection and pre-processing for building an ANN

Selecting dependent or independent variables is crucial in building and optimizing an ANN and should be conducted thoroughly. In this work, the authors characterized the influence of the output psk ($P_{out,top}$) and SHJ ($P_{out,bot}$) power densities on the structural parameters composed of Th_{psk} and B_{psk}, and environmental conditions comprised P_{in}, A_{in}, and T_{mod}. In the recent study on the $E_{out,annual}$ prediction of the 2T psk/Si tandem PV cell for the BIPV, they used two input predictive variables of visible-spectral (P_{vis})

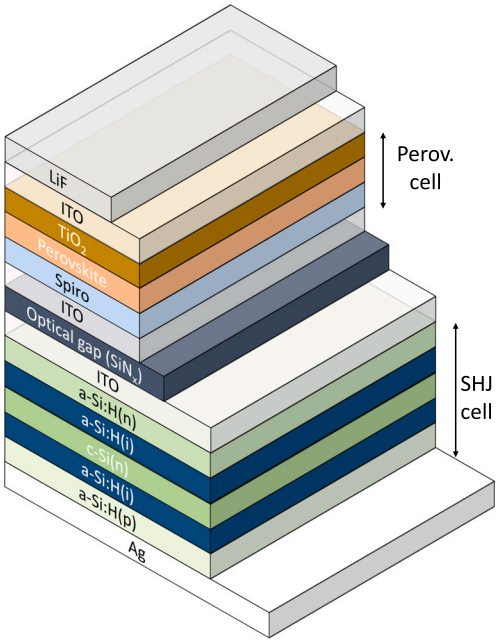


Fig. 2. Architecture of the 4T psk/Si tandem PV cell.

and near-infrared-spectral (P_{nir}) solar irradiances as instead of using one P_{in} variable in the proposed ANN because they proved that the P_{out} value of the 2T tandem PV cell depends on both irradiance and shape of the solar spectrum [12,49]. In this work, the spk and SHJ sub-cells of the 4T tandem PV cell are optically and electrically independent, so the behavior of the output power (P_{out}) of the 4T tandem PV cell is not influenced by changes in the shape of the solar spectrum. The mean squared error (MSE) indicator does not exhibit a significant difference between the two ANNs when considering the effect of solar spectral shapes versus not considering it. This observation is demonstrated in Figure S.5 of the supplementary information (SI). Note that the MSE indicator of an ANN, as presented in Sub-section 2.3.3, is employed to assess the predictive performance of an ANN. Consequently, the P_{in} variable is sufficient for the ANN in predicting the 4T tandem $E_{out,annual}$.

All input environmental variables for building (training and validation) the ANN and predicting the $P_{out,top}$ and $P_{out,bot}$ values were measured, collected, and estimated in Gifu prefecture, located at a latitude of 35.42°N and a longitude of 136.76°E, Japan, from 6 to 18 o'clock daily in 2015. Note that the 2015 data of real operating conditions in Gifu, Japan, is just the representative to make a database for the ANN. The data of any other year could also be employed to replace the input predictive variables of this work without affecting the ANN performance. The training and validation data set for the ANN consists of data specifically from August 2015 for the sole direction (rooftop) of buildings, as shown in Fig. 3 (a-c). Meanwhile, the data set for predicting the $P_{out,top}$ and $P_{out,bot}$ values consists of the data for the whole year of 2015 in four building orientations of the rooftop, east, south, and west [10]. The P_{in} variable values for the ANN are the experimentally measured data by a project under Japan's New Energy and Industrial Technology Development Organization (NEDO), as depicted in Fig. 3 (d). The website pvlighthouse.com was utilized to estimate the values of the variable A_{in}, as illustrated in Fig. 3 (e-h) (referred to [11] (SI)). The T_{mod} variable values were calculated by an estimation method for the silicon-wafer-based PV cell [50], as shown in Fig. 3 (i-l) (referred to [11] (SI)). The B_{psk} variable values comprise 1.56, 1.62, 1.67, 1.70, and 1.73 eV. In contrast, the Th_{psk} variable values are systematically investigated across the 400 to 700 nm range, with a step size of 20 nm applied for each B_{psk} point. The simulated $P_{out,top}$, $P_{out,bot}$ values are the target responses, which are utilized to compare the predicted $P_{out,top}$ and $P_{out,bot}$ values. The data set used to construct the ANN consisted of 32,240 data points, randomly split into two portions: 75% for training and 25% for validation.

2.3. ML-based prediction method

2.3.1. ANN architecture

An ANN includes three primary input, hidden, and output layers with many artificial neurons (nodes) in each layer, as shown in Fig. 4. In this contribution, the input layer has five neurons as predictive variables composed of P_{in}, A_{in}, T_{mod}, Th_{psk}, and B_{psk}, which are provided for the ANN to learn and derive the final results. In an ANN model, the hidden layer positioned between the input and output layers performs computations and identifies hidden features and patterns. It is possible to have more than one hidden layer in the ANN to enhance its ability to capture complex relationships. Optimizing the number of hidden layers and neurons is critical in achieving an optimal and accurate architecture for the ANN model. The output layer in the ANN model presents the final result derived from all the performed computations. This work's output layer comprises two neurons of $P_{out,top}$ and $P_{out,bot}$ responses. An

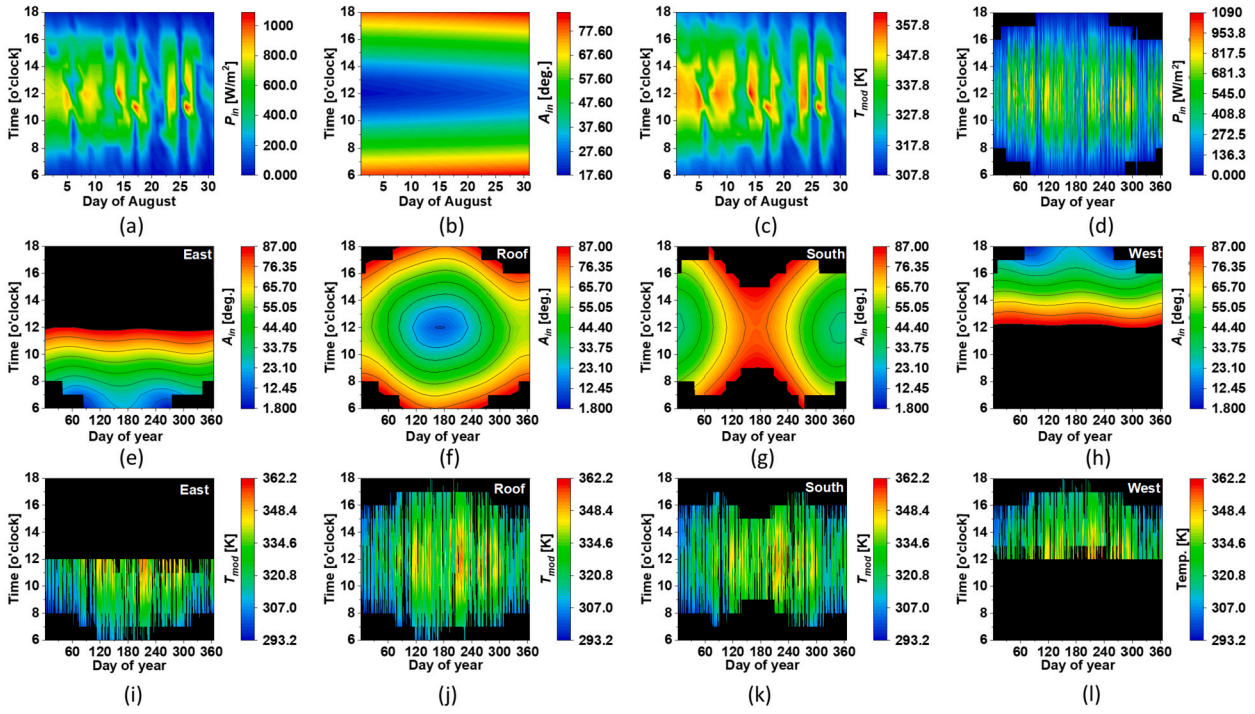


Fig. 3. Input environmental variables for building an ANN composed of P_{in}, A_{in}, and T_{mod} (a-c), input environmental variables for predicting the $P_{out,top}$ and $P_{out,bot}$ values based on the trained ANN for four building orientations of the rooftop, east, south, and west in 2015 (d-l). Note that the P_{in} value is simultaneously the same for four building orientations.

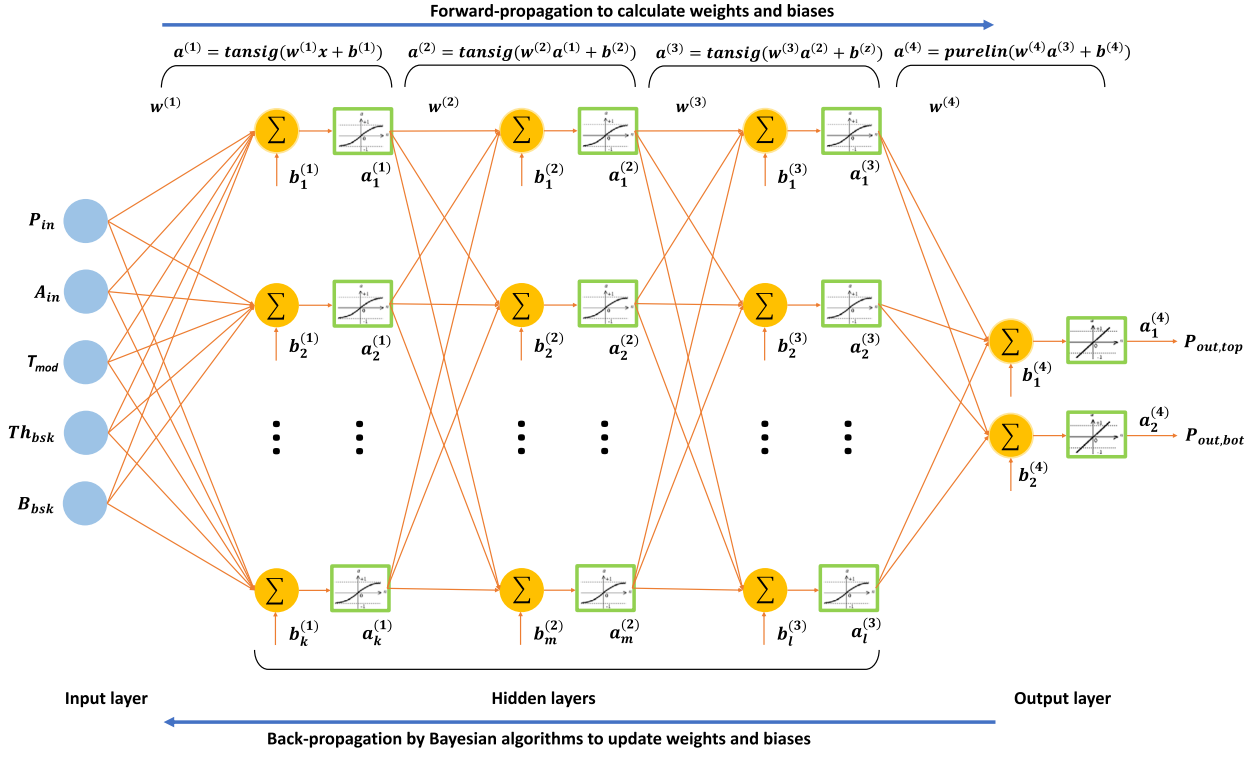


Fig. 4. The ANN composed of one input layer with five neurons (P_{in}, A_{in}, T_{mod}, Th_{bsk}, and B_{bsk}), three hidden layers with the numbers of corresponding neurons (k, m, and l), one output layer with two neurons ($P_{out,top}$ and $P_{out,bot}$).

ANN performs two propagation processes composed of forward-propagation [51–53] to compute weights (w) and biases (b) of the ANN and back-propagation [54] to diminish the error between the target and prediction values by adapting the w and b values.

The forward-propagation process conducts four steps as follows:

- Step 1: Scale all input and output data to a normalized range −1 to 1 via Eq. (1):

$$y = \frac{(y_{max} - y_{min})(x - x_{min})}{x_{max} - x_{min}} + y_{min} \tag{1}$$

  where x is the actual values of the predictive variables whose x_{min} and x_{max} are the minimum and maximal values. Furthermore, $y_{min} = -1$, $y_{max} = 1$, and the value of x is normalized and denoted as y.
- Step 2: The weight (w) and bias (b) values are randomly initialized. A hyperbolic tangent sigmoid transfer function is applied to compute the values within the hidden layers, which can be implemented using a MATLAB function presented by Eq. (2):

$$Tansig(x) = \frac{2}{1 + exp(-2 \times x) - 1} \tag{2}$$

- Step 3: Compute the normalized output values using a purelin function (a linear transfer function in MATLAB), denoted by $purelin(x) = x$.
- Step 4: Re-scale the normalized output values to the predicted actual responses using the mapminmax function (a MATLAB function), which processes matrices by normalizing the minimum and maximum values of each row to $[y_{min}, y_{max}]$.

The back-propagation process employs the Bayesian algorithm to minimize squared errors and adjust the weights (w) accordingly. This algorithm aims to determine the optimal combination of weights that results in a well-generalized ANN [55]. The Bayesian algorithm can help an ANN obtain superior performance and address the over-fitting issue compared to the Levenberg-Marquardt algorithm. The ANN with the Bayesian algorithm introduces weights into the objective function $F(w)$, as denoted by Eq. (3):

$$F(w) = \alpha \sum w^2 + \beta \sum e^2 \tag{3}$$

where w and e are weights and errors of the ANN, respectively. Herein, α and β are objective parameters. The $\sum e^2$ value is minimized by the Levenberg-Marquardt algorithm [12,54], then computing $F(w)$. According to the Bayesian rule, the objective function parameters α and β are optimized by Eq. (4) [56]:

$$\alpha^* = \frac{\gamma}{2\sum (w^*)^2} \qquad \beta^* = \frac{N - \gamma}{2\sum (e^*)^2} \qquad \gamma = N - \alpha^* trace^{-1}(H(w^*)) \tag{4}$$

where α^*, β^*, and e^* are the optimal α, β, and e values at the minimum point w^*, respectively. In this context, the symbol γ represents the number of effective ANN parameters utilized to reduce the error function. The work by Kumar et al. [57] discussed the concept of effective parameters and their role in quantifying the complexity or expressive capacity of the ANN. In order to determine the α^*, β^*, and γ values, it is necessary to calculate the Hessian matrix $H(w^*)$ of the objective function $F(w^*)$ at the w^* point using the Gauss-Newton approximation [58]. Fig. 5 (a) illustrates the flowchart of the Bayesian back-propagation algorithm. For this study, we utilized the back-propagation Bayesian algorithm, which was implemented using the neural fitting toolbox of MATLAB R2022a software. The training, validation, and deployment of the Artificial Neural Network (ANN) for prediction purposes were conducted on a desktop computer with Windows 11 Pro OS (Microsoft), equipped with a 2.9 GHz Intel Core i7-10700 processor and 16 GB RAM.

2.3.2. Optimization of the number of hidden layers and the number of neurons in each layer

Designing an optimal ANN architecture involves a crucial step of determining the number of hidden layers and their corresponding neurons because it dramatically impacts the performance of the ANN [59–61]. However, determining the optimal ANN architecture is still a challenging and complex issue that needs to be solved [33,62]. Suppose the determined ANN architecture does not match the needs, which results in under-fitting or over-fitting of the ANN, leading to the reduction of ANN performance. Some methods have been performed to determine the ANN architecture based on the rule of thumb [32,33], input and output attributes [28,29], trial and error [31], and K-means clustering and principal component analysis [63].

So, this work presents the optimization of the number of hidden layers and their neurons to improve the performance of the ANN and prevent its over-fitting and under-fitting problems. Because the number of hidden layers and the number of neurons in each layer are integer values, an algorithm solving an optimization problem related to integer variables must be applied. Therefore, this work optimizes the number of hidden layers and the number of neurons in each layer using the surrogate algorithm [64], available in the MATLAB Global Optimization Toolbox. The surrogate algorithm aims to find the best possible outcome of an objective function with minimal evaluations by balancing exploration and speed during optimization. The surrogate optimization algorithm performs two following processes:

- Process 1: Construct the surrogate algorithm composed of three steps:
  - Step 1: Choose random initial points within the given boundaries.
  - Step 2: Assess the objective function at these specified points.

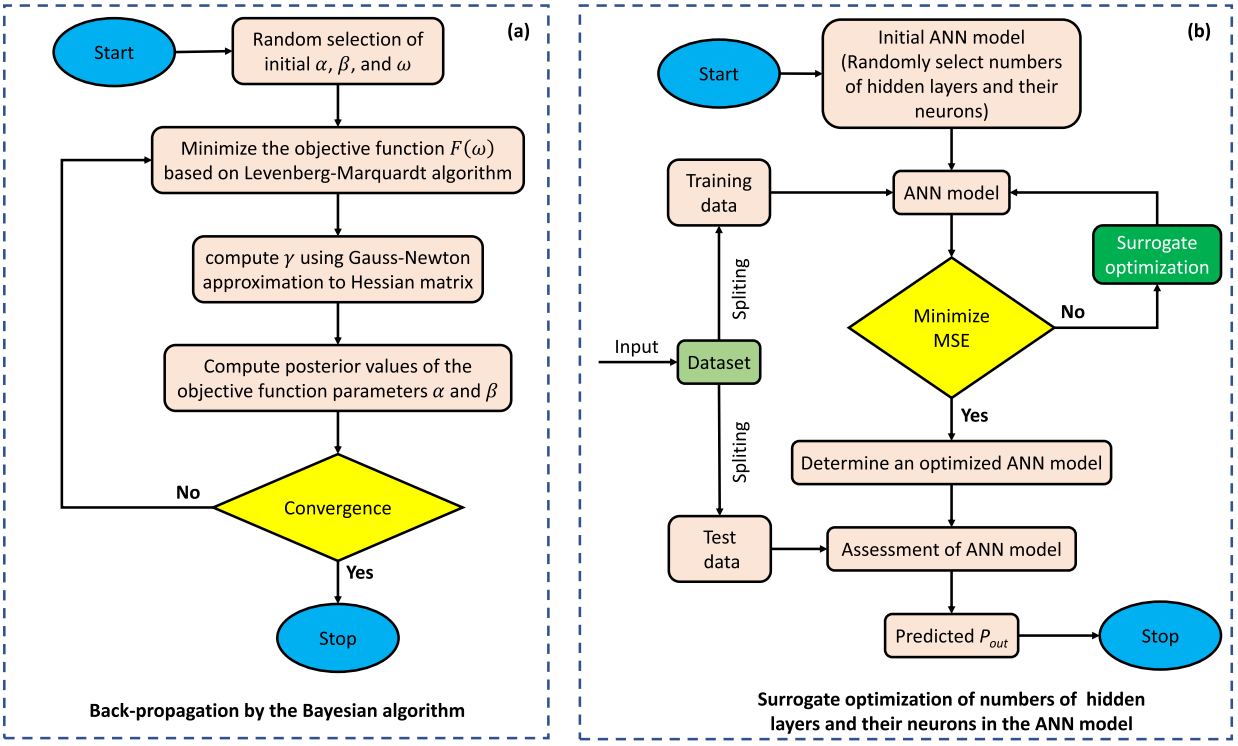


Fig. 5. The Bayesian back-propagation algorithm (a) and the flowchart of the surrogate algorithm deployed in an ANN to optimize the number of hidden layers and the number of neurons in each layer (b).

  - Step 3: Create a substitute for the objective function using a radial basis function to interpolate through the given points.
- Process 2: Look for a minimal objective value composed of five steps:
  - Step 1: Collect a few thousand random points within the boundaries.
  - Step 2: Assess a merit function using the surrogate value obtained from the previously assessed points, then consider exploring other points within the distances where the objective function has been assessed.
  - Step 3: Identify the optimal point based on its merit function measurement.
  - Step 4: Assess the objective function at the optimal point.
  - Step 5: Modify the surrogate and utilize the updated value for all future searches.

The surrogate algorithm could solve an integer-variable optimization problem in terms of the below Eq. (5):

$$\min_{x} f(x) \; such \; that \begin{cases} lower \; bounds \leq x \leq upper \; bounds \\ x = \{x_1, x_2, x_3, ..., x_i\} \;,\; x_i \in integer \\ i : number \; of \; variables \;,\; i = 1, 2, 3... \end{cases} \tag{5}$$

where $f(x)$ and x_i are objective function and integer variables, respectively. All variables have to be an integer within finite bounds. In this contribution, the performance of the ANN is determined by the mean squared error (MSE) value denoted by $f(x)$, where x_i represents the number of neurons in the ANN's i^{th} hidden layer. This helps to evaluate the ANN performance accurately. The number of hidden layers is examined for optimization from 1 to 5. The number of corresponding hidden layer neurons is varied from 1 to 30. Thus, this work optimizes the number of hidden layers and the number of neurons in each layer to obtain the minimum value of the validation-stage MSE indicator. Each hidden layer has a minimum and maximum number of neurons known as the lower and upper bounds, respectively. The flowchart of the surrogate optimization algorithm is depicted in Fig. 5 (b).

2.3.3. Assessment of ANN performance

Five employed statistical indicators are of MSE, root mean squared error (RMSE), mean absolute error (MAE), correlation coefficient (R), and Willmott's index of agreement (I_w) [65] to assess the ANN performance. These indicators were represented by the following Eqs. (6)–(10).

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(P_{i,out(Atlas)} - P_{i,out(ANN)} \right)^2 \tag{6}$$

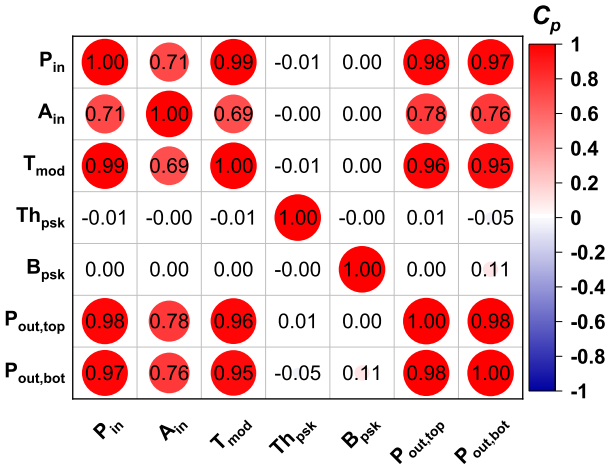


Fig. 6. Correlation matrix indicating the relationship between the input variables used for prediction and the target responses.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_{i,out(Atlas)} - P_{i,out(ANN)}\right)^2} \tag{7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|P_{i,out(Atlas)} - P_{i,out(ANN)}\right| \tag{8}$$

$$R = \frac{\sum_{i=1}^{n}\left(P_{i,out(ANN)} - \overline{P}_{out(ANN)}\right)\times\left(P_{i,out(Atlas)} - \overline{P}_{out(Atlas)}\right)}{\sqrt{\sum_{i=1}^{n}\left(P_{i,out(ANN)} - \overline{P}_{out(ANN)}\right)^2 \times \sum_{i=1}^{n}\left(P_{out(Atlas)} - \overline{P}_{out(Atlas)}\right)^2}} \tag{9}$$

$$I_w = 1 - \frac{\sum_{i=1}^{n}\left(P_{i,out(Atlas)} - P_{i,out(ANN)}\right)^2}{\sum_{i=1}^{n}\left(\left|P_{i,out(Atlas)} - \overline{P}_{out(Atlas)}\right| + \left|P_{i,out(ANN)} - \overline{P}_{out(Atlas)}\right|\right)^2} \tag{10}$$

where n is the number of data points. $P_{i,out(Atlas)}$ and $P_{i,out(ANN)}$ are the i^{th} simulated and predicted output power densities, respectively. Furthermore, $\overline{P}_{out(Atlas)}$ and $\overline{P}_{out(ANN)}$ are the mean values of simulated and predicted output power densities, respectively.

3. Results and discussions

3.1. Analysis of the importance degree of input variables to an ANN

In this work, the authors examined and evaluated the impact degree of each input variable comprised P_{in}, A_{in}, T_{mod}, Th_{psk}, and B_{psk} to the predictive performance of the objective responses $P_{out,top}$ and $P_{out,bot}$ of the 4T tandem PV cells. A correlation matrix between the predictive input variables and the objective responses reveals their linear correlation, as shown in Fig. 6. The Pearson correlation (C_p) index is crucial to determine the connection between the input variables used for prediction and the objective responses. The C_p indexes of 0.98 and 0.97 (P_{in} vs. $P_{out,top}$ and P_{in} vs. $P_{out,bot}$), 0.78 and 0.76 (A_{in} vs. $P_{out,top}$ and A_{in} vs. $P_{out,bot}$), 0.96 and 0.95 (T_{mod} vs. $P_{out,top}$ and T_{mod} versus $P_{out,bot}$) demonstrate the highly linear correlation between the input environmental variables (P_{in}, A_{in}, and T_{mod}) and the responses. So, these variables are considered crucial variables for the predictive ANN. Meanwhile, the input structural variables (Th_{psk} and B_{psk}) have a nonlinear correlation with the objective responses because of their C_p indexes around zero. However, further evaluation is required to determine the nonlinear correlation between the input predictive variables and the objective responses, as stated in [66].

The authors compared and assessed the metrics of each ANN with a sole input variable. For these evaluations, five uncomplicated artificial neural networks were utilized. Each network consisted of one single-neuron input layer, one 14-neuron hidden layer, and one two-neuron output layer. Table 1 listed the metrics of each ANN with a sole input variable. An ANN performs better if it has lower indicators of MSE, RMSE, and MAE and higher indicators of R and I_w. According to Table 1, the importance degree of the variables for the predictive performance in descending order is P_{in}, T_{mod}, A_{in}, B_{psk}, and Th_{psk}. Regarding the R indicator, the environmental variables have the most impact predicting the $P_{out,top}$ and $P_{out,bot}$ of the 4T tandem PV cells. In contrast, the other structural variables have an insignificant effect. However, in terms of the I_w indicator, it is clear that the structural variables also play a not small role in building a good-performance ANN. Note that the I_w indicator can serve as a standardized measure to identify differences in the model prediction error. This indicator can detect additive and proportional variations in the target and simulated means and variances [65]. The value of I_w can range from 0 to 1. The indicator of 1 indicates a perfect match, while 0 indicates no agreement.

Table 1
Assessment of impact degree of each input variable for the predictive performance.

Metrics	Input variables				
	Sole P_{in}	Sole A_{in}	Sole T_{mod}	Sole Th_{psk}	Sole B_{psk}
MSE	56.81	607.64	85.89	1931.25	1951.73
RMSE	7.54	24.65	9.27	43.95	44.18
MAE	5.39	16.57	6.81	35.12	35.10
R	0.98	0.83	0.97	0.03	0.06
I_w	0.70	0.51	0.61	0.22	0.23

Table 2
The optimal numbers of hidden layers and their neurons based on the given boundary conditions.

ANN models	Number of hidden layers	ith hidden layer	Minimal neurons	Maximal neurons	Optimal number of neurons	Optimal validation-stage indicators		
						MSE	RMSE	MAE
ANN-1	1	1^{st}	1	30	14	0.43872	0.66236	0.44492
ANN-2	2	1^{st}	1	30	18	0.10446	0.32321	0.19883
		2^{nd}	1	30	16			
ANN-3	3	1^{st}	1	30	19	0.02283	0.15111	0.10243
		2^{nd}	1	30	19			
		3^{rd}	1	30	16			
ANN-4	4	1^{st}	1	30	19	0.05024	0.22414	0.11949
		2^{nd}	1	30	20			
		3^{rd}	1	30	19			
		4^{th}	1	30	4			
ANN-5	5	1^{st}	1	30	10	0.04131	0.20325	0.10338
		2^{nd}	1	30	20			
		3^{rd}	1	30	18			
		4^{th}	1	30	13			
		5^{th}	1	30	11			

3.2. The optimal ANN and its performance assessment

This work examined five ANNs with various hidden layers, including the 1-hidden-layer (ANN-1), 2-hidden-layer ANN (ANN-2), 3-hidden-layer (ANN-3), 4-hidden-layer (ANN-4), and 5-hidden-layer (ANN-5) ANNs. The boundary conditions of the variables of the ANN's hidden layers and the number of neurons in each hidden layer for optimizing the ANN architecture are presented in Table 2. The optimal neuron number in hidden layers are (14), (18 and 16), (19, 19, and 16), (19, 20, 19, and 4), and (10, 20, 18, 13, and 11) for the ANN-1, ANN-2, ANN-3, ANN-4, and ANN-5 models, respectively. Fig. 7 (a) depicts the progress in optimizing the validation-stage indicators (MSE, RMSE, and MAE). When the minimal validation-stage MSE values were found, the optimization progress stopped after the objective evaluation number of 30 for ANN-1, 200 for ANN-2 to ANN-4, and 250 for ANN-5. Also, Fig. 7 (b and c) shows that an ANN with a more significant number of hidden layers needs a longer time for each evaluation. The cumulative time for the objective evaluation process is 1, 9, 23, 24, and 36 h for ANN-1 to ANN-5, respectively. As shown in Fig. 7 (d), the optimal validation-stage MSE values are 0.43872, 0.10446, 0.02283, 0.05024, and 0.04131, respectively. Thus, the results reveal that the ANN-3 model with the lowest optimal validation-stage MSE values has the best predictive performance.

Table 3 lists five statistic indicators composed of MSE, RMSE, MAE, R, and I_w to quantify the validation of the ANNs and confirm an ANN with the best performance. According to the comparison of the results, the two models of ANN-3 and ANN-5 have the lowest error indicators of MSE, RMSE, and MAE and the highest indicators of R and I_w. These two models have the same indicators of R and I_w. The ANN-5 model has training-stage error indicators better than the ANN-3 model. Still, the ANN-3 model has validation-stage error indicators better than the ANN-5 model. Because the validation-stage indicators take precedence in evaluating the ANN performance, the ANN-3 model is considered the most superior model for predicting the $P_{out,top}$ and $P_{out,bot}$ values of the 4T tandem PV cells.

Fig. 8 illustrates the regression plots of five predictive ANNs for the training (a-e) and validation (f-j) stages, which graphically and quantitatively depict the consistency of the simulated (target) and predicted P_{out} (including top and bottom) values. According to the results, the ANN-3 model's predicted P_{out} values closely resemble the target P_{out} values compared to other ANNs' predictions. Moreover, the authors comprehensively performed the deviation analysis to assess the ANN accuracy better, as depicted in Fig. 9 (a-f). Herein, the relevant distribution of the deviation values was denoted by $D_i = P_{i,out(ANN)} - P_{i,out(Atlas)}$. According to Fig. 9, the deviation values generated by the ANN-3 model are much more compressed in both the training (a, c, and e) and validation (b, d, and f) stages than in the other models.

Another recent study presented that an ANN (1.25994 in MSE, 1.12247 in RMSE, 0.42133 in MAE, and 0.99979 in R) could predict the annual output energy ($E_{out,annual}$) of the 2T psk/Si tandem PV cells for the BIPV with the tiny relative error values (0.10

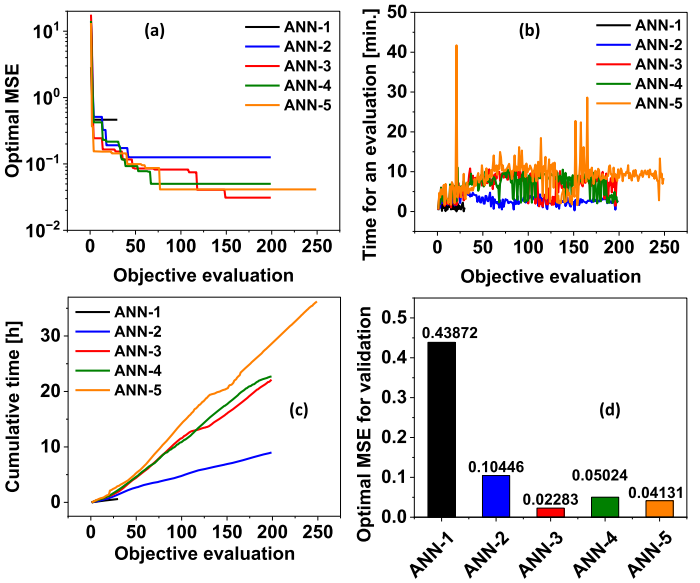


Fig. 7. Progress in optimizing the MSE values using the surrogate algorithm (a), time for each objective evaluation and cumulative time for all objective evaluations (b-c), comparison of optimal MSE values for three ANNs (d).

Table 3
Metrics in assessing the performance of five ANNs.

Performance metrics	Training					Validation				
	ANN-1 model	ANN-2 model	ANN-3 model	ANN-4 model	ANN-5 model	ANN-1 model	ANN-2 model	ANN-3 model	ANN-4 model	ANN-5 model
MSE	3.03500	0.20409	0.03775	0.06942	0.02508	0.43872	0.10446	0.02283	0.05024	0.04131
RMSE	1.74212	0.45176	0.19429	0.26348	0.15836	0.66236	0.32321	0.15110	0.22414	0.20325
MAE	0.47038	0.19885	0.10078	0.11802	0.09623	0.44492	0.19883	0.10038	0.11949	0.10243
R	0.99911	0.99995	0.99999	0.99996	0.99999	0.99994	0.99998	0.99999	0.99998	0.99999
I_w	0.99864	0.99991	0.99998	0.99997	0.99999	0.99981	0.99995	0.99999	0.99998	0.99999

for the rooftop, 0.16 for the south, 1.05 for the east, and 0.90% for the west) between the simulated and predicted $E_{out,annual}$ values for the 2T tandem PV cells at the 1.67-eV B_{psk} [12]. Compared to such an ANN, the ANN-3 model of this work has a superior performance (MSE of 0.02283, RMSE of 0.15110, MAE of 0.103383, and R of 0.99999). Consequently, the ANN-3 models have been proposed as the optimal model for predicting this work's $P_{out,top}$ and $P_{out,bot}$ values of the 4T tandem PV cells.

It is worth knowing that the ANN was built based on a database of input data (P_{in}, A_{in}, T_{mod}, Th_{psk}, and B_{psk}) and output data ($P_{out,top}$ and $P_{out,bot.}$), which does not depend on the database of any year and climatic zone. In this work, the authors just utilized the database of natural operating conditions in a specific climatic zone (Gifu of Japan) for a particular month (August 2015) to build the ANN. Such data represents the database for building the ANN. So, employing the realistic condition data of any year and climatic zone to build the ANN does not change its predictive performance. The advantage of the ANN is to predict the $E_{out,annual}$ values in any year and any climatic zone without performing any other simulations, as long as we have the input database of the desired year and climatic zone. In other words, the ANN can predict the $E_{out,annual}$ values of 4T tandem PV cells worldwide from the corresponding input database.

3.3. Tandem annual output energy predicted by the trained ANN

This study predicted the $P_{out,top}$ and $P_{out,bot}$ values using the optimal ANN-3 for the 4T tandem architecture. The predictions were made for psk materials with various bandgaps in a 1.66–1.75 eV range with a step size of 0.01 eV and different thicknesses in a 400–900 nm range with a step size of 10 nm. The tandem P_{out} value of the 4T tandem PV cell is calculated by Eq. (11):

$$P_{out,tandem} = P_{out,top} + P_{out,bot} \tag{11}$$

Then, the annual input ($E_{in,annual}$) and output ($E_{out,annual}$) energies, and η_{annual} of the 4T tandem PV cells are integrated as denoted by Eqs. (12) – (14):

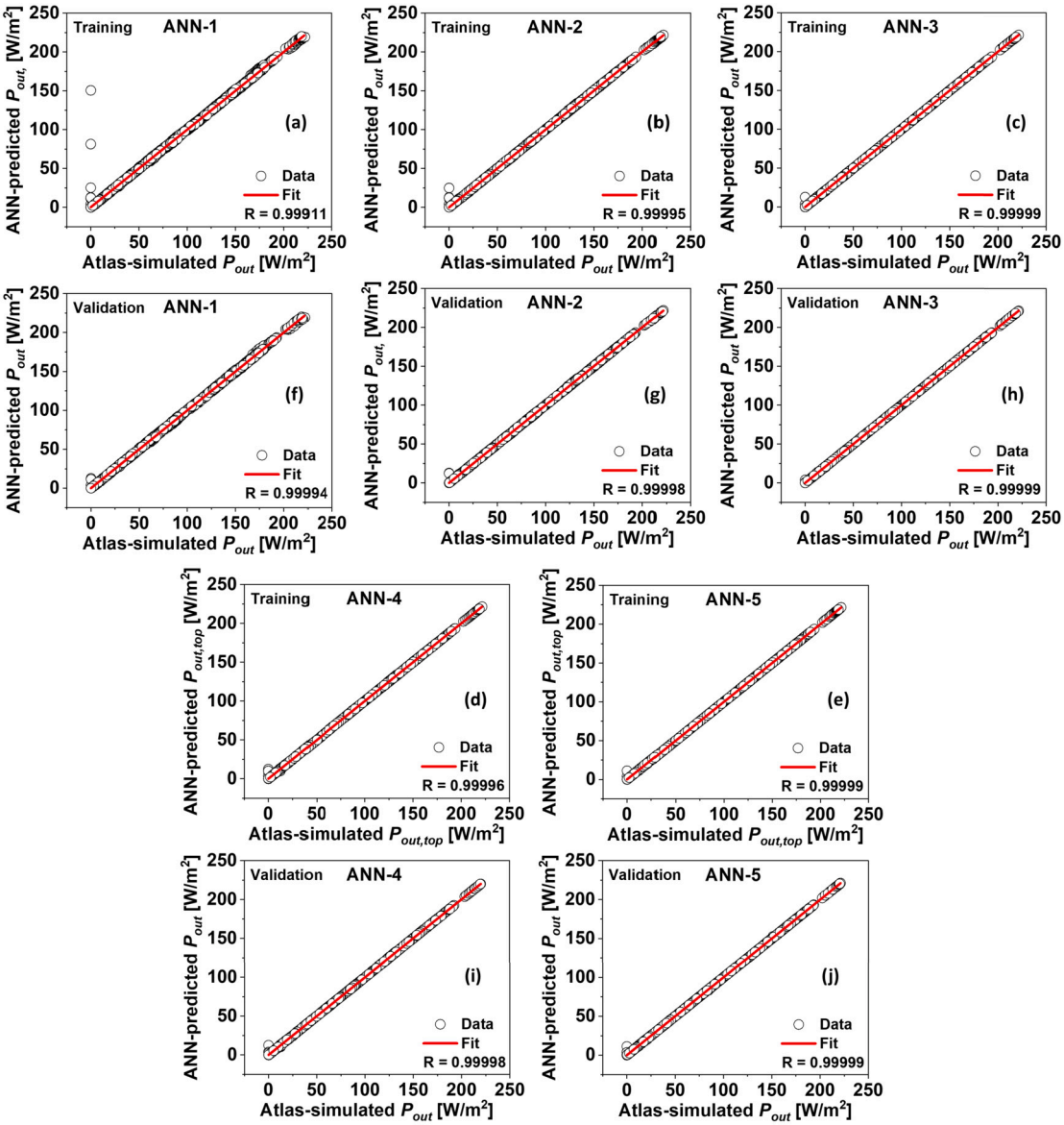


Fig. 8. Regression plots of five predictive ANNs for training (a-e) and validation stages (f-j).

$$E_{in,annual} = \int_{1}^{365}\int_{6}^{18} P_{in}(x,y)dxdy \tag{12}$$

$$E_{out,annual} = \int_{1}^{365}\int_{6}^{18} P_{out,tandem}(x,y)dxdy \tag{13}$$

$$\eta_{annual} = \frac{\eta_{out,annual}}{\eta_{in,annual}} \times 100\% \tag{14}$$

where x and y are time points of a day ($6 \le x \le 18$) and the days of a year ($1 \le y \le 365$), respectively. At a specific time on a day of the year, the values of $P_{in}(x,y)$ and $P_{out,tandem}(x,y)$ represent the P_{in} and $P_{out,tandem}$ respectively. Furthermore, η_{annual} is the annual output efficiency harvested by the 4T tandem PV cell.

Fig. 10 (a-d) illustrates the contour plots and their color mapping of the predicted $E_{out,annual}$ values of the 4T tandem PV cells for various Th_{psk} and B_{psk} variables in four different directions of buildings. We observed that the 4T tandem PV cells work well in the area at a B_{psk} range from 1.70 to 1.73 eV with various Th_{psk} values from 500 to 750 nm for both the east and west orientations, from 500 to 900 nm for the south facade, and 850-900 nm for the rooftop. The comparison of the maximal $E_{out,annual}$ and η_{annual} values

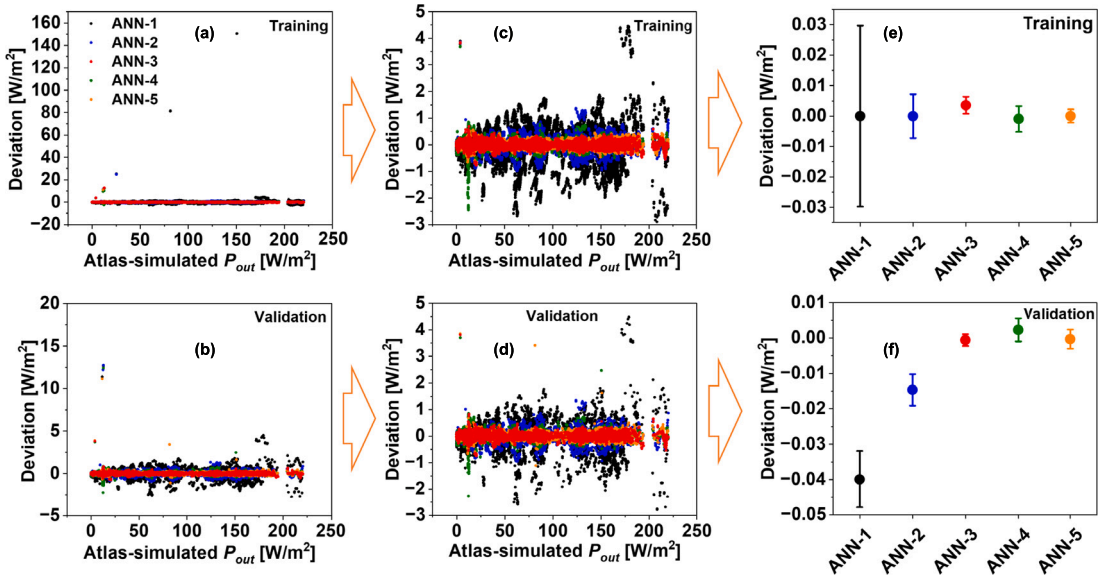


Fig. 9. The deviation values between the target and predicted P_{out} values for five predictive ANNs in the training (upper images) and validation (lower images) stages. The deviation values are in the form of data point plots for the full deviation range (a and b), for the zoom-in deviation in a range from -5 to 5 (c and d), and in the form of box-chart plots (e and f).

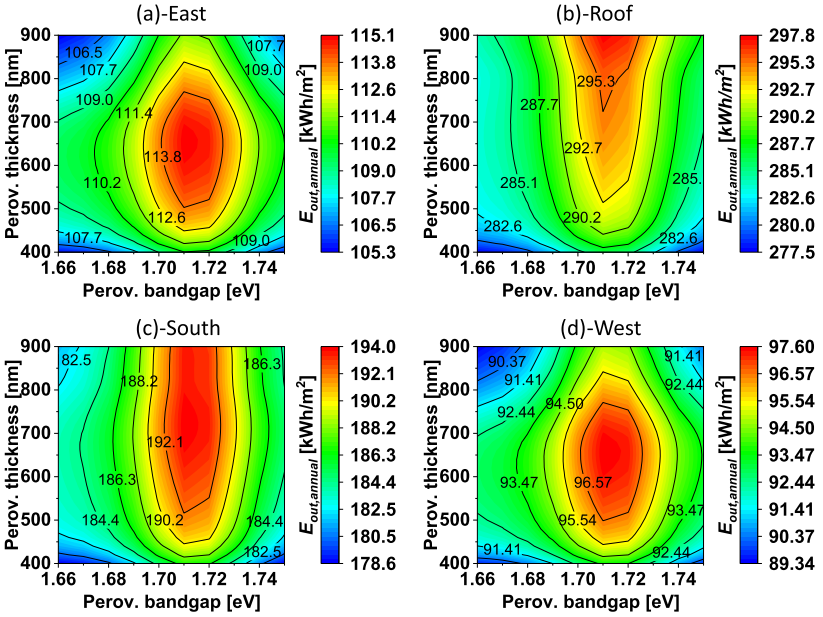


Fig. 10. Contour plots and their color mapping of the predicted $E_{out,annual}$ values of the 4T tandem PV cells for various Th_{psk} and B_{psk} variables in four different directions of buildings (a-d).

generated by the 4T tandem PV cells with the various Th_{psk} and B_{psk} values in the standard test condition (STC) and the actual operating conditions for four different directions are depicted in Table 4. Based on the available B_{psk} of 1.56, 1.62, 1.67, 1.70, and 1.73 eV in this work, the 4T tandem PV cell achieved the best-simulated power conversion efficiency (η) of 31.32% at a B_{psk} of 1.70 eV and Th_{psk} of 780 nm under the STC.

The results of η_{annual} and $E_{annual,out}$ under real operating conditions show that the optimal B_{psk} value is 1.71 eV with the optimal Th_{psk} values of 660 nm for the east and west orientations, 720 nm for the south, and 900 nm for the roof, respectively. Hence, the 4T tandem PV cells obtain the maximal $E_{out,annual}$ of 297.73, 193.98, 115.01, and 97.6 kWh/m^2 and the maximal η_{annual} of 23.90, 15.57, 9.23, and 7.84% for four building orientations of the rooftop, south, east, and west, respectively. When used for BIPV, the actual operating performance of the 4T tandem PV cell is significantly lower compared to its performance under STC because the realistic tandem performance is considerably adversely influenced by environmental factors P_{in}, A_{in}, and T_{mod}. According to Table 4

Table 4
Comparison of the maximal $E_{out,annual}$ and η_{annual} values harvested by the 2T [12] and 4T tandem PV cells with the various Th_{psk} and B_{psk} values in the standard condition and the real operating conditions for four different directions.

Tandem	Conditions	Direction	Optimal B_{psk} (eV)	Optimal Th_{psk} (nm)	Maximal η_{tandem} (%)	Maximal $\eta_{out,annual}$ (%)	Maximal $E_{out,annual}$ (kWh/m^2)
2T [12]	STC		1.73	1090	31.41		
	Realistic	East	1.72	680		8.44	105.07
		Roof	1.73	700		22.68	282.54
		South	1.72	680		14.03	174.71
		West	1.72	680		7.29	90.79
4T (this work)	STC		1.70	780	31.32		
	Realistic	East	1.71	660		9.23	115.01
		Roof	1.71	900		23.90	297.73
		South	1.71	720		15.57	193.98
		West	1.71	660		7.84	97.60

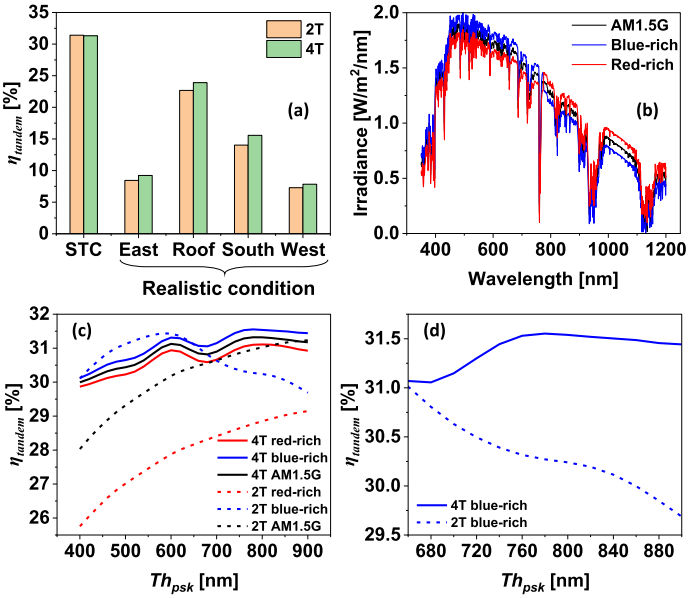


Fig. 11. Comparison of η_{tandem} values between the 2T and 4T tandem PV cells for the BIPV by real operating conditions (a), different solar spectral shapes including red-rich, AM1.5G, and blue-rich (b), comparison of η_{tandem} values between 2T (B_{psk} of 1.73 eV) and 4T (B_{psk} of 1.70 eV) tandem PV cells versus Th_{psk} under conditions composed of T_{mod} = 300 K, A_{in} = 0°, solar spectrum with P_{in} = 1000 W/m^2 and different shapes (c), zoom-in η_{tandem} values of 2T and 4T tandem PV cells versus Th_{psk} in a range from 660 to 900 nm under the blue-rich solar spectrum (d).

and Fig. 11 (a), although the maximal η_{tandem} of the 2T tandem PV cell is higher than that of the 4T tandem PV cell under the STC, the 4T tandem PV cell has the higher maximal $E_{out,annual}$ and η_{annual} values than the 2T tandem PV cell under the realistic condition. This result is attributed to the effect of solar spectral shapes on the η_{tandem} values. Fig. 11 (c) indicates comparison of η_{tandem} values between 2T (B_{psk} of 1.73 eV) and 4T (B_{psk} of 1.70 eV) tandem PV cells versus Th_{psk} under conditions composed of T_{mod} = 300 K, A_{in} = 0°, solar spectrum with P_{in} = 1000 W/m^2 and different shapes in AM1.5G spectra, red-rich spectra, and blue-rich spectra (Fig. 11 (b)). The identification of different solar spectral shapes was mentioned in the literature [12,49]. According to Fig. 11 (c), the $\eta_{tandem-4T}$ is lightly enhanced by blue-rich spectra but reduced by red-rich spectra compared to under AM1.5G. However, the solar spectral shape does not change its $\eta_{tandem-4T}$ curve shape. Blue and red light is known to be absorbed by the tandem PV cell's psk and SHJ cells, respectively. The psk sub-cell contributes significantly more to the $\eta_{tandem-4T}$ than the SHJ cell, as shown in Figure S.3 (SI). The blue-rich spectra explain the enhancement of the $\eta_{tandem-4T}$ value, while its reduction is attributed to the red-rich spectra. The shape of the $\eta_{tandem-4T}$ curve remains unchanged as the 4T tandem PV cell remains unaffected by the current density matching condition between the psk and SHJ cells. For the 2T tandem PV cell, the solar spectral shape changes the $\eta_{tandem-2T}$ behavior [12,49]. Fig. 11 (d) indicates that the $\eta_{tandem-4T}$ in a Th_{psk} range from 660 to 900 nm is significantly higher than the $\eta_{tandem-2T}$ under the blue-rich spectra. The blue-rich spectrum is the primary solar spectrum in Gifu, Japan [12]. In real operating conditions, the η_{tandem} of the 4T tandem PV cell exceeds that of the 2T tandem PV cell by a significant margin. The authors suggested that the 4T tandem PV cell work more effectively than the 2T tandem PV cell in the BIPV.

4. Conclusion and limitation

The authors examined an ANN in predicting the $E_{out,annual}$ values of the 2T psk/Si tandem PV cell for the BIPV. The proposed ANN in this work employed the input environmental (P_{in}, A_{in}, and T_{mod}) and structural (Th_{psk} and B_{psk}) variables. This work evaluated the impact degree of each input variable on the ANN performance in predicting the $P_{out,top}$ and $P_{out,bot}$ values of the 4T tandem PV cell. The results revealed that all input variables mentioned above are crucial in building a good-performance ANN. Thus, the environmental variables are the most critical factors in that contribution. Based on the optimization method using the surrogate algorithm, the ANN-3 model attained the best predictive performance (MSE of 0.02283, RMSE of 0.15110, MAE of 0.10338, R of 0.99999, and I_w of 0.99999). The authors, therefore, determined that the ANN-3 model with the architecture including one input layer (5 neurons), three hidden layers (19, 19, and 16 neurons for the 1^{st}, 2^{nd}, and 3^{rd}, respectively), and one output layer (2 neurons) are the optimal model used for the predictions in this work. The prediction result using the ANN-3 model presented that the 4T tandem PV cell can obtain the best performance within the B_{psk} range from 1.70 to 1.73 eV with various Th_{psk} ranges for different directions of buildings under natural operating conditions. Meanwhile, the optimal B_{psk} values are proposed as 1.71 eV, where the highest $E_{out,annual}$ values obtained are 297.73, 115.01, 193.98, and 97.60 kWh/m^2. The corresponding highest η_{annual} are 23.90, 9.23, 15.57, and 7.84% for four building orientations of the rooftop, east, south, and west, respectively. Furthermore, comparing the maximal $E_{out,annual}$ and η_{annual} values between the 2T and 4T psk/Si tandem PV cells revealed that the 4T tandem PV cell is better than the 2T tandem PV cell in the BIPV.

This work solely aims at the theoretical prediction by combining the simulated and predicted results. The experiment is conducted to verify the agreement between simulation and experimental results. It may be published in another work shortly. In this project, the task of developing an experimental procedure to evaluate the accuracy of the ANN still needs to be addressed. The authors have been looking for practical solutions to overcome those challenges. In parallel, the prediction of the 3T psk/Si tandem PV cell is currently underway to comprehensively evaluate the $E_{out,annual}$ and η_{annual} of the tandem PV cells with different terminals, then select the best tandem configuration for the BIPV. Furthermore, predicting the $E_{out,annual}$ and η_{annual} values of 2T, 3T, and 4T psk/Si tandem PV cells taking into account the degradation rates of psk and silicon PV cells has been considered to perform and publish somewhere.

CRediT authorship contribution statement

Dong C. Nguyen: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Yasuki Ishikawa: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

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.

Data availability

Data will be made available on request.

Acknowledgements

This work is supported by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e18097.

References

[1] N. Martín-Chivelet, K. Kapsis, H.R. Wilson, V. Delisle, R. Yang, L. Olivieri, J. Polo, J. Eisenlohr, B. Roy, L. Maturi, G. Otnes, M. Dallapiccola, W.M.U. Wijeratne, Building-integrated photovoltaic (bipv) products and systems: a review of energy-related behavior, Energy Build. 262 (2022) 111998, https://doi.org/10.1016/j.enbuild.2022.111998.

[2] A. Karthick, K. Kalidasa Murugavel, A. Ghosh, K. Sudhakar, P. Ramanan, Investigation of a binary eutectic mixture of phase change material for building integrated photovoltaic (BIPV) system, Sol. Energy Mater. Sol. Cells 207 (2020) 110360, https://doi.org/10.1016/j.solmat.2019.110360.

[3] A. Roy, A. Ghosh, S. Bhandari, S. Sundaram, T.K. Mallick, Perovskite solar cells for BIPV application: a review, Buildings 10 (2020) 129, https://doi.org/10.3390/buildings10070129.

[4] J. Benemann, O. Chehab, E. Schaar-Gabriel, Building-integrated pv modules, Sol. Energy Mater. Sol. Cells 67 (2001) 345–354, https://doi.org/10.1016/S0927-0248(00)00302-0.

[5] P. Defaix, W. Van Sark, E. Worrell, E. de Visser, Technical potential for photovoltaics on buildings in the EU-27, Sol. Energy 86 (2012) 2644–2653, https://doi.org/10.1016/j.solener.2012.06.007.

[6] A. Ghosh, S. Sundaram, T.K. Mallick, Colour properties and glazing factors evaluation of multicrystalline based semi-transparent photovoltaic-vacuum glazing for BIPV application, Renew. Energy 131 (2019) 730–736, https://doi.org/10.1016/j.renene.2018.07.088.

[7] A. Richter, M. Hermle, S.W. Glunz, Reassessment of the limiting efficiency for crystalline silicon solar cells, IEEE J. Photovolt. 3 (2013) 1184–1191, https://doi.org/10.1109/JPHOTOV.2013.2270351.
[8] Y. Zhao, M. Najafi, V. Zardetto, D. Zhang, Four terminal perovskite-silicon PV tandem devices hit 30% efficiency, https://www.solliance.eu/2022/four-terminal-perovskite-silicon-pv-tandem-devices-hit-30-efficiency/, 2022.
[9] KAUST, KAUST claims 33.7% efficiency for perovskite/silicon tandem solar cell, https://www.pv-magazine.com/2023/05/30/kaust-claims-33-7-efficiency-for-perovskite-silicon-tandem-solar-cell/, 2023.
[10] D.C. Nguyen, F. Murata, K. Sato, M. Hamada, Y. Ishikawa, Evaluation of annual performance for building-integrated photovoltaics based on 2-terminal perovskite/silicon tandem cells under realistic conditions, Energy Sci. Eng. 10 (2022) 1373–1383, https://doi.org/10.1002/ese3.1105.
[11] D.C. Nguyen, K. Sato, M. Hamada, F. Murata, Y. Ishikawa, Annual output energy harvested by building-integrated photovoltaics based on the optimized structure of 2-terminal perovskite/silicon tandem cells under realistic conditions, Sol. Energy 241 (2022) 452–459, https://doi.org/10.1016/j.solener.2022.06.018.
[12] D.C. Nguyen, Y. Ishikawa, Artificial neural network for predicting annual output energy of building-integrated photovoltaics based on the 2-terminal perovskite/silicon tandem cells under realistic conditions, Energy Rep. 8 (2022) 10819–10832, https://doi.org/10.1016/j.egyr.2022.08.233.
[13] D. Patel, S. Patel, P. Patel, M. Shah, Solar radiation and solar energy estimation using ann and fuzzy logic concept: a comprehensive and systematic study, Environ. Sci. Pollut. Res. 29 (2022) 32428–32442, https://doi.org/10.1007/s11356-022-19185-z.
[14] M. Sridharan, Short review on various applications of fuzzy logic-based expert systems in the field of solar energy, Int. J. Ambient Energy 43 (2022) 5112–5128, https://doi.org/10.1080/01430750.2021.1927839.
[15] H. Zhang, J. Shi, C. Zhang, A hybrid ensembled double-input-fuzzy-modules based precise prediction of pv power generation, Energy Rep. 8 (2022) 1610–1621, https://doi.org/10.1016/j.egyr.2022.02.298.
[16] P.M. Kumar, R. Saravanakumar, A. Karthick, V. Mohanavel, Artificial neural network-based output power prediction of grid-connected semitransparent photovoltaic system, Environ. Sci. Pollut. Res. 29 (2022) 10173–10182, https://doi.org/10.1007/s11356-021-16398-6.
[17] S.N. Mughal, Y.R. Sood, R.K. Jarial, Design and optimization of photovoltaic system with a week ahead power forecast using autoregressive artificial neural networks, Mater. Today Proc. 52 (2022) 834–841, https://doi.org/10.1016/j.matpr.2021.10.223.
[18] H. Hafdaoui, E.A.K. Boudjelthia, A. Chahtou, S. Bouchakour, N. Belhaouas, Analyzing the performance of photovoltaic systems using support vector machine classifier, Sustai. Energy Grids Netw. 29 (2022) 100592, https://doi.org/10.1016/j.segan.2021.100592.
[19] P. Lauret, C. Voyant, T. Soubdhan, M. David, P. Poggi, A benchmarking of machine learning techniques for solar radiation forecasting in an insular context, Sol. Energy 112 (2015) 446–457, https://doi.org/10.1016/j.solener.2014.12.014.
[20] M. Bouzerdoum, A. Mellit, A.M. Pavan, A hybrid model (sarima-svm) for short-term power forecasting of a small-scale grid-connected photovoltaic plant, Sol. Energy 98 (2013) 226–235, https://doi.org/10.1016/j.solener.2013.10.002.
[21] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, S. Wang, Short-term solar power forecasting based on weighted Gaussian process regression, IEEE Trans. Ind. Electron. 65 (2018) 300–308, https://doi.org/10.1109/TIE.2017.2714127.
[22] B. Zazoum, Solar photovoltaic power prediction using different machine learning methods, Energy Rep. 8 (2022) 19–25, https://doi.org/10.1016/j.egyr.2021.11.183.
[23] D. Niu, K. Wang, L. Sun, J. Wu, X. Xu, Short-term photovoltaic power generation forecasting based on random forest feature selection and ceemd: a case study, Appl. Soft Comput. J. 93 (2020) 106389, https://doi.org/10.1016/j.asoc.2020.106389.
[24] Y. Dai, Y. Wang, X. Yang, Q. Zhou, Lowess smoothing and random forest based gru model: a short-term photovoltaic power generation forecasting method, SSRN Electron. J. 256 (2021) 124661, https://doi.org/10.2139/ssrn.3966209.
[25] M. Zolfaghari, M.R. Golabi, Modeling and predicting the electricity production in hydropower using conjunction of wavelet transform, long short-term memory and random forest models, Renew. Energy 170 (2021) 1367–1381, https://doi.org/10.1016/J.RENENE.2021.02.017.
[26] D. Assouline, N. Mohajeri, J.L. Scartezzini, Large-scale rooftop solar photovoltaic technical potential estimation using random forests, Appl. Energy 217 (2018) 189–211, https://doi.org/10.1016/j.apenergy.2018.02.118.
[27] M. Zamo, O. Mestre, P. Arbogast, O. Pannekoucke, A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: deterministic forecast of hourly production, Sol. Energy 105 (2014) 792–803, https://doi.org/10.1016/j.solener.2013.12.006.
[28] M.A. Sartori, P.J. Antsaklis, A simple method to derive bounds on the size and to train multilayer neural networks, IEEE Trans. Neural Netw. 2 (1991) 467–471, https://doi.org/10.1109/72.88168.
[29] S. Tamura, M. Tateishi, Capabilities of a four-layered feedforward neural network: four layers versus three, IEEE Trans. Neural Netw. 8 (1997) 251–255, https://doi.org/10.1109/72.557662.
[30] M. Madhiarasan, Accurate prediction of different forecast horizons wind speed using a recursive radial basis function neural network, Prot. Control Mod. Power Syst. 5 (2020) 1–9, https://doi.org/10.1186/s41601-020-00166-8.
[31] M. Madhiarasan, S.N. Deepa, Comparative analysis on hidden neurons estimation in multi layer perceptron neural networks for wind speed forecasting, Artif. Intell. Rev. 48 (2017) 449–471, https://doi.org/10.1007/s10462-016-9506-6.
[32] A. Carballal, F. Cedron, I. Santos, A. Santos, J. Romero, Minimal neural network topology optimization for aesthetic classification, Neural Comput. Appl. 33 (2021) 107–119, https://doi.org/10.1007/s00521-020-05550-x.
[33] M.A. Rahman, R. chandren Muniyandi, D. Albashish, M.M. Rahman, O.L. Usman, Artificial neural network with Taguchi method for robust classification model to improve classification accuracy of breast cancer, PeerJ Comput. Sci. 7 (2021) 2–27, https://doi.org/10.7717/PEERJ-CS.344.
[34] D.C. Nguyen, Y. Ishikawa, Improvement in 4-terminal perovskite/silicon heterojunction tandem solar cells' performance with an index matching layer of silicon nitride, in: IEEE 48th Photovoltaic Specialists Conference (PVSC), 2021, pp. 0696–0700.
[35] H.H. Li, Refractive index of alkali halides and its wavelength and temperature derivatives, J. Phys. Chem. Ref. Data 5 (1976) 329–528, https://doi.org/10.1063/1.555536.
[36] E. Raoult, R. Bodeux, S. Jutteau, S. Rives, A. Yaiche, D. Coutancier, J. Rousset, S. Collin, Optical characterizations and modelling of semitransparent perovskite solar cells for tandem applications, in: 36th European Photovoltaic Solar Energy Conference and Exhibition, 2019, pp. 757–763.
[37] L.J. Phillips, A.M. Rashed, R.E. Treharne, J. Kay, P. Yates, I.Z. Mitrovic, A. Weerakkody, S. Hall, K. Durose, Dispersion relation data for methylammonium lead triiodide perovskite deposited on a (100) silicon wafer using a two-step vapour-phase reaction process, Data Brief 5 (2015) 926–928, https://doi.org/10.1016/j.dib.2015.10.026.
[38] J. Werner, G. Nogay, F. Sahli, T.C.J. Yang, M. Bräuninger, G. Christmann, A. Walter, B.A. Kamino, P. Fiala, P. Löper, S. Nicolay, Q. Jeangros, B. Niesen, C. Ballif, Complex refractive indices of cesium-formamidinium-based mixed-halide perovskites with optical band gaps from 1.5 to 1.8 eV, ACS Energy Lett. 3 (2018) 742–747, https://doi.org/10.1021/acsenergylett.8b00089.
[39] Z.C. Holman, A. Descoeudres, L. Barraud, F.Z. Fernandez, J.P. Seif, S. De Wolf, C. Ballif, Current losses at the front of silicon heterojunction solar cells, IEEE J. Photovolt. 2 (2012) 7–15, https://doi.org/10.1109/JPHOTOV.2011.2174967.
[40] M.A. Green, Self-consistent optical parameters of intrinsic silicon at 300 K including temperature coefficients, Sol. Energy Mater. Sol. Cells 92 (2008) 1305–1310, https://doi.org/10.1016/j.solmat.2008.06.009.
[41] T. Minemoto, M. Murata, Theoretical analysis on effect of band offsets in perovskite solar cells, Sol. Energy Mater. Sol. Cells 133 (2015) 8–14, https://doi.org/10.1016/j.solmat.2014.10.036.
[42] A. Nakanishi, Y. Takiguchi, S. Miyajima, Device simulation of $CH_3NH_3PbI_3$ perovskite/heterojunction crystalline silicon monolithic tandem solar cells using an n-type a-Si:H/p-type μc-$Si_{1-x}O_x$:H tunnel junction, Phys. Status Solidi A, Appl. Mater. Sci. 213 (2016) 1–6, https://doi.org/10.1002/pssa.201532946.

[43] M. Lu, U. Das, S. Bowden, S. Hegedus, R. Birkmire, Optimization of interdigitated back contact silicon heterojunction solar cells: tailoring hetero-interface band structures while maintaining surface passivation, Prog. Photovolt., Res. Appl. 19 (2011) 326–338, https://doi.org/10.1002/pip.1032.
[44] A. Froitzheim, K. Brendel, L. Elstner, W. Fuhs, K. Kliefoth, M. Schmidt, Interface recombination in heterojunctions of amorphous and crystalline silicon, J. Non-Cryst. Solids 299–302 (2002) 663–667, https://doi.org/10.1016/S0022-3093(01)01029-8.
[45] F. Zhang, W. Ma, H. Guo, Y. Zhao, X. Shan, K. Jin, H. Tian, Q. Zhao, D. Yu, X. Lu, G. Lu, S. Meng, Interfacial oxygen vacancies as a potential cause of hysteresis in perovskite solar cells, Chem. Mater. 28 (2016) 802–812, https://doi.org/10.1021/acs.chemmater.5b04019.
[46] F. Wang, S. Bai, W. Tress, A. Hagfeldt, F. Gao, Defects engineering for high-performance perovskite solar cells, npj Flex. Electron. 2 (2018) 1–14, https://doi.org/10.1038/s41528-018-0035-z.
[47] F. Azri, A. Meftah, N. Sengouga, A. Meftah, Electron and hole transport layers optimization by numerical simulation of a perovskite solar cell, Sol. Energy 181 (2019) 372–378, https://doi.org/10.1016/j.solener.2019.02.017.
[48] A. Al-Ashouri, E. Köhnen, B. Li, A. Magomedov, H. Hempel, P. Caprioglio, J.A. Márquez, A.B.M. Vilches, E. Kasparavicius, J.A. Smith, N. Phung, D. Menzel, M. Grischek, L. Kegelmann, D. Skroblin, C. Gollwitzer, T. Malinauskas, M. Jošt, G. Matič, B. Rech, R. Schlatmann, M. Topič, L. Korte, A. Abate, B. Stannowski, D. Neher, M. Stolterfoht, T. Unold, V. Getautis, S. Albrecht, Monolithic perovskite/silicon tandem solar cell with >29% efficiency by enhanced hole extraction, Science 370 (2020) 1300–1309, https://doi.org/10.1126/science.abd4016.
[49] D.C. Nguyen, Y. Ishikawa, Spectral shape changes the optimal perovskite thickness of the 2-terminal perovskite/silicon tandem solar cell, in: 2022 IEEE 48th Photovoltaic Specialists Conference (PVSC), 2022, pp. 0044–0046.
[50] J. Oh, G. TamizhMani, E. Palomino, Temperatures of building applied photovoltaic (BAPV) modules: air gap effects, in: Reliability of Photovoltaic Cells, Modules, Components, and Systems III, SPIE, 2010, pp. 33–43.
[51] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Syst. 2 (1989) 303–314, https://doi.org/10.1007/bf02551274.
[52] K.I. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural Netw. 2 (1989) 183–192, https://doi.org/10.1016/0893-6080(89)90003-8.
[53] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Netw. 2 (1989) 359–366, https://doi.org/10.1016/0893-6080(89)90020-8.
[54] W. Paul, Beyond Regression: New Tools for Prediction and Analysis in the Behavioural Sciences, Harvard University, Cambridge, MA, 1975.
[55] D.J.C. MacKay, Bayesian interpolation, Neural Comput. 4 (1992) 415–447, https://doi.org/10.1162/neco.1992.4.3.415.
[56] Z. Yue, Z. Songzheng, L. Tianshi, Bayesian regularization bp neural network model for predicting oil-gas drilling cost, in: BMEI 2011 - Proceedings 2011 International Conference on Business Management and Electronic Information, 2011, pp. 483–487.
[57] P. Kumar, S.N. Merchant, U.B. Desai, Improving performance in pulse radar detection using Bayesian regularization for neural network training, Digit. Signal Process. 14 (2004) 438–448, https://doi.org/10.1016/j.dsp.2004.06.002.
[58] F.D. Foresee, M.T. Hagan, Gauss-Newton approximation to Bayesian learning, in: IEEE International Conference on Neural Networks - Conference Proceedings, 1997, pp. 1930–1935.
[59] M.A. Nielsen, Neural Networks and Deep Learning, vol. 25, Determination Press, San Francisco, CA, USA, 2015.
[60] A. Koutsoukas, K.J. Monaghan, X. Li, J. Huan, Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data, J. Cheminform. 9 (2017) 1–13, https://doi.org/10.1186/s13321-017-0226-y.
[61] T. Nitta, Resolution of singularities introduced by hierarchical structure in deep neural networks, IEEE Trans. Neural Netw. Learn. Syst. 28 (2017) 2282–2293, https://doi.org/10.1109/TNNLS.2016.2580741.
[62] G. Naitzat, A. Zhitnikov, L.H. Lim, Topology of deep neural networks, J. Mach. Learn. Res. 21 (2020) 1–40.
[63] M.I.C. Rachmatullah, J. Santoso, K. Surendro, Determining the number of hidden layer and hidden neuron of neural network for wind speed prediction, PeerJ Comput. Sci. 7 (2021) 1–19, https://doi.org/10.7717/PEERJ-CS.724.
[64] N.V. Queipo, R.T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, P.K. Tucker, Surrogate-Based Analysis and Optimization, Progress in Aerospace Sciences, vol. 41, 2005, pp. 1–28.
[65] C.J. Willmott, Some comments on the evaluation of model performance, Bull. Am. Meteorol. Soc. 63 (1982) 1309–1313, https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2.
[66] P. Burman, A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods, Biometrika 76 (1989) 503–514, https://doi.org/10.1093/biomet/76.3.503.