Bionic fusion perspective: Audiovisual-motivated integration network for solar irradiance prediction

Bionic fusion perspective: Audiovisual-motivated integration network for solar irradiance prediction

Han Wu[1, 2,*], Xiaozhi Gao[2], Jiani Heng[3], Xiaolei Wang[4], and Xiaoshu Lü[4, 5]

[1]School of Automation, Northwestern Polytechnical University, Xi'an 710129, China

[2]School of Computing, University of Eastern Finland, Kuopio FI-70211, Finland

[3]School of Mathematics and Statistics, Beijing Technology and Business University, Beijing 100048, China

[4]School of Technology and Innovations, University of Vaasa, Vaasa FI-65200, Finland.

[5]School of Engineering, Aalto University, Espoo FI-02150, Finland.

[*]Corresponding author: wuhan@mail.nwpu.edu.cn

Keywords: Audiovisual integration cognition; multi-scale feature; residual connection; solar irradiance prediction; Transformer.

Abstract:

Accurate and reliable prediction of solar irradiance (SI) is an important requirement to develop solar energy while a challenging task due to stochastic and nonlinear data characteristics. Additionally, most deep networks show powerful prediction capabilities but lack the supports from biological science, reflecting that bionically-inspired networks in SI analysis are still not enough explored. To this end, this paper proposes an audiovisual-motivated Transformer-CNN integration network, called ATI-net, for predicting SI. The audiovisual cognition gives a superior design framework for ATI-net with signal capture, signal analysis, and prediction blocks. In the first block, through mimicking the function of

both eye and ear in external signal conversion, multi-scale features are extracted by incorporating multi-branch convolutions with varying kernels, where the Mish function addresses the problem that traditional ReLU function stops learning when the input is negative. In the second block, through mimicking the function of left and right hemispheres in neuronal signal analysis, two structures triggered by Transformers and convolutions are designed to remember temporal evolutionary rules, where residual connections are beneficial to mine deep information and avoid forgetting. In the third block, through mimicking the function of a higher brain region in generating understanding, the above information is integrated to make the SI prediction. Besides, the nonlinear dependencies and linear relationships are independently extracted and integrated into the ATI-net, which not only reduces information interference but is consistent with the "divide and conquer" idea. Experimental results show that the ATI-net outperforms 18 benchmarks, and average improvements of root mean squared error (RMSE) are 26.28% and 26.01% for two datasets, respectively. In summary, the ATI-net is one of the reliable alternatives to SI prediction.

1. Introduction

With the rapid development of the world economy and people's living standards, global energy consumption is gradually increasing by 1.6% per year anticipated by the International Energy Agency (IEA) [1], where fossil fuels are still the main source [2]. Excessive consumption of fossil fuels not only depletes energy sources but also leads to climate change problems [3]. To reduce global warming and meet global energy demand, most countries focus on the development and utilization of renewable energy sources, where solar energy is a flexible, sustainable, promising, and distributed resource [4]. The average solar irradiance (SI) received by the surface of the earth is 1367 W/m^2 about 1.8×10^8 GW fully year [5]. Besides, based on a report from REN21, total renewable energy generation in 2020

reaches 2838 GW across the globe, where solar energy reaches 760 GW, accounting for 26.77% [6]. Accurate and reliable prediction of SI plays a significant role in promoting solar energy generation and integration, supporting efficient decision-making of sectors, and minimizing operational uncertainties. However, due to the stochastic, nonlinear, and intermittent properties of SI, its high-precise modeling is still a challenging task [7].

Nowadays, there are two groups for predicting SI [8, 9]: physical and data-driven methods. Physical methods infer future SI based on numerical weather prediction (NWP) and physical rules, where detailed weather conditions, such as rainfall, relative humidity, temperature, etc., and geographic information are taken as input [10]. The above mathematical equations present clear relationships between SI and corresponding weather conditions, being suitable for long-term and large-scale SI simulation. However, they are excessively complex and require professional background knowledge, high computing costs, and advanced hardware [11]. Data-driven methods using regression techniques with NWP data or accumulated historical data, and are post-processing methods based on classical statistical and machine learning (ML) techniques.

Classical statistic techniques first look for internal trends based on autoregressive or moving strategies and then estimate the SI, being suitable for short-term prediction. As widespread applications, the autoregressive moving average (ARMA) [12], autoregressive integrated moving average (ARIMA) [13], vector autoregressive (VAR) [14], generalized autoregressive conditional heteroskedasticity (GARCH) [15] are designed to forecast SI. For example, the ARIMA-based model encapsulated the temporal dependencies based on a few historical data and then made SI prediction [13]. Generally, the above models are simple and easy-implement but have some strict constraints, such as ARMA-based models require stationary data [16]. Thus, they are able to handle linear data but fail for complex

nonlinear SI data [11].

ML techniques directly approximate complex nonlinear functions by mapping from input to output [17], resulting in it gradually penetrating into various time series analysis and achieving the most attention. Based on the layer number, ML generally contains deep and shallow ML [18]. The support vector regression (SVR), being one of the mainstream shallow ML methods, handled nonlinear dependencies and showed strong learning capabilities in SI prediction [19]. Extreme learning machine (ELM) was another famous shallow ML and integrated with teaching learning optimization for parameter selection, improving SI estimation capability [20]. Extreme gradient boosting (XGBoost) generated the SI prediction by minimizing residuals of successive iterations of multiple decision trees, being better than SVR [21]. Generally, shallow ML models are able to cope with nonlinear data but lack the introduction of deeply hidden features and temporal dependencies due to small layers and no gating mechanism. The transition from “shallow mining” to “deep mining” may be interesting.

Deep ML shows powerful feature engineering and self-learning capabilities in the face with complex and variable SI data. According to the information transfer structure, the deep ML methods are divided into three groups: recursive, parallel, and hybrid-recursive-parallel methods [22]. Recursive methods are proposed based on the gating mechanism and the sequential input, which controls the feature flow for remembering core information, such as the recurrent neural network (RNN) [23], long short-term memory network (LSTM) [24], and gated recurrent unit (GRU) [25]. For instance, through mining relationships between SI and related weather data, the LSTM-based model was established for predicting one-hour SI, showing less overfitting and better generalization capability than the backpropagation network (BPNN) [24]. Based on the parameter optimization of the grid search technique, GRU and LSTM were respectively adopted to predict SI, where GRU had a faster

convergence [25]. Moreover, for capturing bidirectional sequential dependencies, the bidirectional LSTM (BI-LSTM) [26] and bidirectional GRU (BI-GRU) [27] integrated the forward and backward layers for SI prediction. Some studies also found that LSTM and GRU had preferences, and their hybrid use was beneficial for remembering dynamics of time series, thereby designing the BI-LSTM-GRU model [28]. To mitigate gradient explosion and vanishing of recursive methods, some attention turns to the parallel and hybrid-recursive-parallel methods. On the first branch, the multilayer perceptron (MLP) [29] and convolutional neural network (CNN) [30] presented flexible and superior feature extraction capabilities, where one-dimensional convolution and two/three-dimensional convolution were suitable for handling time-related and image-related SI data, respectively. Based on the residual connection, the information in the shallow layer was transmitted to the deep layer without loss [31], which learned more hidden information and improved MLP/CNN. Moreover, the temporal convolutional network (TCN) [32] with parallel causal-dilated convolutions and Transformer [33] with parallel multi-head self-attention were used to forecast SI. On another branch, the hybrid-recursive-parallel methods, including CNN-LSTM, CNN-GRU, LSTM-CNN, GRU-CNN, etc., are proposed to fit the SI series. For example, through fusing the advantages of CNN and LSTM layers, CNN-LSTM was proposed for forecasting SI and better than LSTM, where the dependencies were mined at the feature level [34]. Besides, the LSTM-CNN first extracted time-related dependency from SI data and then captured internal relationships [4]. Generally, the deep ML models show advantages in the prediction accuracy of SI, especially for the Transformer framework, but most of them start from the engineering or mathematical tools instead of the support from biological science [35]. Bionically-inspired ideas not only improve the performance of deep networks but also increase their interpretability at the biological level.

Of late, it is prospective to design deep networks by mimicking the advanced bio-cognitive processes, especially for time series analysis. For example, based on the left-right brain interaction process, a bionic network is explored to forecast significant wave height and performs the gate, collaboration, and inhibition functions [36]. Starting from the bio-auditory cognition process, an auditory-circuit-motivated network (ACM-Net) is designed to forecast short-term electricity price and shows higher accuracy than CNN-LSTM and TCN [37]. Drawing inspiration from the bio-visual multi-scale analysis process, a bionic MSV-Net is built to forecast short-term electricity price and realizes the visual sensitivity at the time level [38]. Motivated by the process of traditional Chinese medicine pulse diagnosis, a bionic multi-feature extraction network is developed to predict electricity load, showing that different activation functions have preferences and advanced activation functions improve the model performance [39]. Generally, the above models indicate that it is effective to design deep prediction networks via mimicking bio-cognition processes, but bionically-inspired networks in SI analysis are still not enough explored.

Experienced hunters successfully capture prey through audiovisual integration recognition, where the prey trajectories are time series with stochastic and nonlinear changes. Thus, this paper explores an audiovisual-motivated Transformer-CNN integration network (ATI-net) for predicting SI. Six experiments and six discussions under two cases show the effective prediction and stability of the ATI-net. In general, the main contributions of this study are given as follows:

- The audiovisual integration cognition provides a bionically-inspired framework to design the ATI-net, which leverages powerful self-learning to present SI fluctuations.
- The nonlinear dependencies and linear relationships are independently extracted and then integrated into the ATI-net, reducing the information interference.

- Various receptive field sizes by integrating multi-branch convolutions with kernels are designed to handle multi-scale features, where the advanced Mish function addresses the problem that the traditional ReLU function stops learning when the input is negative.
- Two structures driven by Transformers and convolutions are proposed to mine change rules, and residual connections increase the network depth and fuse the information of shallow and deep layers, being useful to avoid information forgetting.

The remainder of this paper is organized as follows. In Section 2, research problems of this paper and basic principles of Transformer are introduced, respectively. In Section 3, the ATI-net is explored and implemented. In Section 4, six experiments under two cases are presented. In Section 5, six discussions of the ATI-Net are carried out. Finally, conclusions and future works are summarized in Section 6.

2. Research problems and basic principles

This section first introduces the research problems of this study and then presents the basic principles of Transformer.

2.1. Research problems

The prediction problem of SI is depicted as a multivariate time series prediction. The input contains the global horizontal index (GHI) and related weather conditions, such as wind speed, relative humidity, temperature, etc., and the output is the GHI in the next time stamp. The general definition is as follows:

$$\hat{g}_{i+n+1} = \boldsymbol{F}\left(\boldsymbol{s}_i, \boldsymbol{s}_{i+1}, ..., \boldsymbol{s}_{i+n}\right) \tag{1}$$

where $(\boldsymbol{s}_i, \boldsymbol{s}_{i+1}, ..., \boldsymbol{s}_{i+n})$ is the input multivariate series, including GHI and corresponding weather conditions; $\hat{g}_{i+n+1}$ is the predicted GHI at time $i+n+1$; $n+1$ is the input size of sliding window; $\boldsymbol{F}$ is the fitting function of SI.

This paper starts from the bionic fusion perspective to re-construct the $\boldsymbol{F}$ function and aims to make high-precise prediction. As shown in Fig. 1, there are two core reasons why it is important to combine the audiovisual integration system and the deep network for GHI prediction. 1) From Fig. 1, the prey trajectories are complex and time-related multivariate series, and the GHI prediction is depicted as a multivariate series prediction. 2) From Fig. 1, trajectories of the prey have several movement modes, such as "Movement" and "Stop". The goal GHI series also has several evolutionary rules corresponding to the "Movement" and "Stop". Thus, based on the above similarities, it is reasonable to mimic this biological audiovisual system to design the deep network for GHI prediction. The prediction problem of SI is re-constructed via mimicking the bio-audiovisual system:

$$\begin{aligned} \hat{g}_{i+n+1} &= \boldsymbol{ATI\text{-}net}\left(\boldsymbol{s}_i, \boldsymbol{s}_{i+1}, ..., \boldsymbol{s}_{i+n}\right) \\ &= \boldsymbol{f}_4\left\{\boldsymbol{f}_2\left[\boldsymbol{f}_1\left(\boldsymbol{s}_i, \boldsymbol{s}_{i+1}, ..., \boldsymbol{s}_{i+n}\right)\right], \boldsymbol{f}_3\left(g_i, g_{i+1}, ..., g_{i+n}\right)\right\} \end{aligned} \tag{2}$$

where $(\boldsymbol{s}_i, \boldsymbol{s}_{i+1}, ..., \boldsymbol{s}_{i+n})$ contains historical GHI and related weather conditions; $\boldsymbol{ATI\text{-}net}(*)$ is the designed ATI-net; $(g_i, g_{i+1}, ..., g_{i+n})$ contains the historical GHI data; $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ functions are designed based on the bio-visual system; $\boldsymbol{f}_3$ function is designed via the bio-auditory system; $\boldsymbol{f}_4$ is the integration function of bio-visual and bio-auditory systems. Compared with Eq. (1), the complex $\boldsymbol{F}$ function is transmitted into multiple simple sub-functions, which not only brings more refined designs but also reduces difficulty in function approximation.

In summary, how to explore the ATI-net via mimicking the audiovisual cognition system for GHI prediction is an open issue.

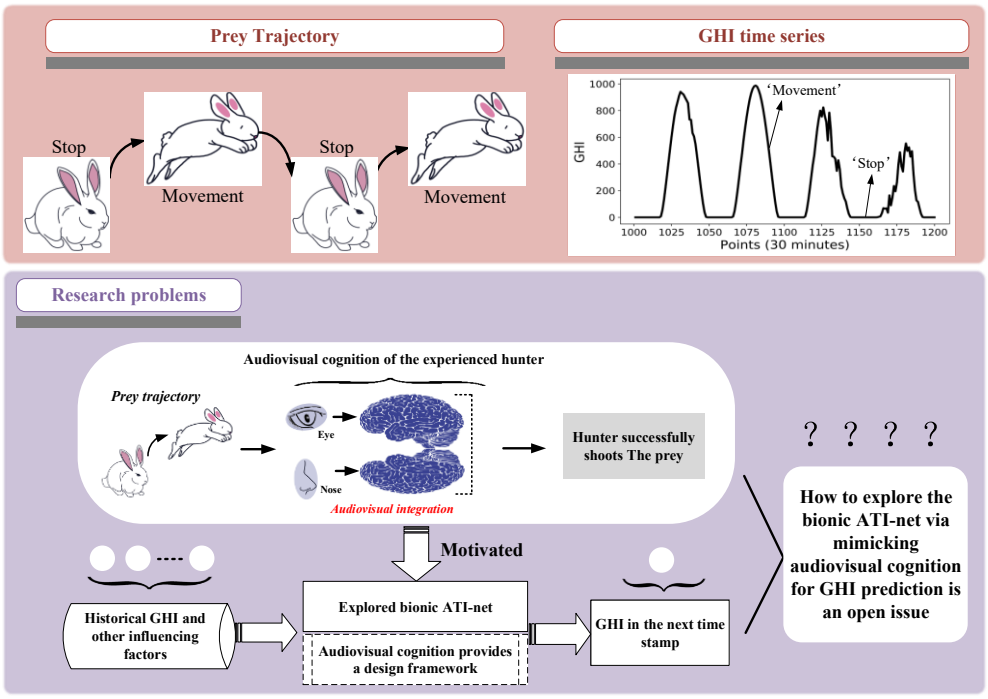


Fig. 1: Description of research problems in this paper.

2.2. Transformer

The Transformer is first proposed in [40] and contains two core sections with an encoder and decoder. The first section encodes the input series through repeated applications of dot-product attention within the input series. The second section generates the output by applying dot-product attention between the encoded input and the output as it is being formed. Since this paper mainly focuses on the short-term SI prediction, we only adopt the Transformer encoder section to reduce the model complexity as performed in [41]. Moreover, the Transformer is effective to mine temporal dependencies based on the multi-head self-attention and is injected into the ATI-net. As shown in Fig. 2, the Transformer encoder has three components, including the multi-head self-attention, add & dropout, and feed-forward layer [41].

1) Multi-head self-attention

Multi-head self-attention optimizes the synergistic relationships of multiple variables fed to the

Transformer encoder and augments the performance [42], thereby enhancing the accuracy and efficiency. The multi-head self-attention relies heavily on the scaled dot-product attention, which assigns weights for input multiple variables and further highlights their relationships. Specific formulas are given by:

$$\boldsymbol{Q} = \boldsymbol{S}\boldsymbol{w}^{Q}, \boldsymbol{K} = \boldsymbol{S}\boldsymbol{w}^{K}, \boldsymbol{V} = \boldsymbol{S}\boldsymbol{w}^{V} \tag{3}$$

$$\text{self-attention}\left(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\right)=\boldsymbol{Z}=Softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{T}}{\sqrt{d}}\right)\boldsymbol{V} \tag{4}$$

where $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ are the query, key, and value matrices, which are calculated by the linear transformation of input $\boldsymbol{S}$; $\boldsymbol{w}^{Q}$, $\boldsymbol{w}^{K}$, and $\boldsymbol{w}^{V}$ are linear transformation matrices; $\boldsymbol{Z}$ is the weight matrix based on the scaled dot-product attention; $Softmax(*)$ is the Softmax function; d is the dimension to prevent excessively large inner product values.

The multi-head self-attention first performs the linear transformation based on query, key, and value matrices, and then independently feeds transformed matrices into multiple scaled dot-product attentions. These output matrices are concatenated as the output of multi-head self-attention. The specific formula is as follows:

$$\boldsymbol{M}=\text{Concat}\left(\boldsymbol{Z}_1,\boldsymbol{Z}_2,...\boldsymbol{Z}_n\right)\boldsymbol{w}^{O} \tag{5}$$

where $\boldsymbol{M}$ is the output of multi-head self-attention; $\boldsymbol{w}^{O}$ is the transformation matrix.

2) Add & dropout

The add & dropout has two parts with “add” and “dropout”, where the “add” belongs to the residual connection, mitigates the network degradation, and makes sure the effective information flow. The “dropout” is a strategy to speed up the network convergence and avoid overfitting. The specific formula of this component is as follows:

$$\boldsymbol{A}=dropout\left(\boldsymbol{S}+\boldsymbol{M}\right) \tag{6}$$

where $+$ is the "add" operation; $dropout\left(*\right)$ is the "dropout" operation; $\boldsymbol{A}$ is the output after add & dropout component.

3) Feed-forward layer

The feed-forward layer is a two-layer fully connected network. The first network layer introduces the nonlinear features via the rectification of the linear unit (ReLU) activation function, and the second network layer only makes two linear transformations. The specific formula of this component is as follows:

$$\boldsymbol{B}=ReLU\left(0,\boldsymbol{A}\boldsymbol{w}_a+\boldsymbol{b}_a\right)\boldsymbol{w}_b+\boldsymbol{b}_b \tag{7}$$

where $\boldsymbol{w}$ and $\boldsymbol{b}$ are the trainable weights and bias; $ReLU\left(*\right)$ is the ReLU activation function; $\boldsymbol{B}$ is the output of this component.

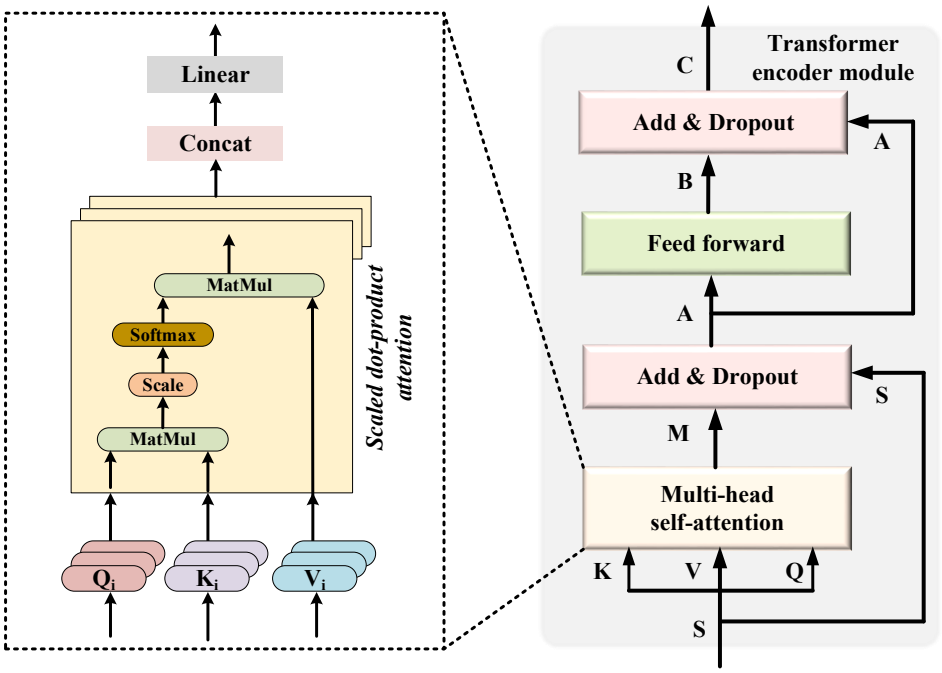


Fig. 2: Description of Transformer encoder.

3. The explored ATI-net

3.1. Bionic modeling of the ATI-net

Fig. 3 presents the specific framework of the audiovisual cognition system [43], which is divided into three blocks based on the function, including the signal capture, signal analysis, and prediction. In particular, the prey trajectory is a kind of external stimulus, where light and sound stimulus are respectively captured by the eye and ear and then are converted into multiple neuronal signals for transferring information flow and suppressing irrelevant interference. After that, these electricity signals are independently processed by right and left hemispheres in the primary cerebral cortex, where left and right hemispheres not only remember core information but also show some complementarity. Finally, the processed information is integrated into a higher cerebral cortex for shooting the prey and understanding the environment. Furthermore, our previous work [36-39] proves that it is promising to design deep prediction networks via learning from bio-cognitive processes. Thus, the audiovisual cognition system provides a bionically-inspired framework for ATI-net in predicting GHI.

As shown in Fig. 3, the ATI-net inherits the advantages of the framework and layer function involved in the bio-audiovisual cognition system. Specifically, the ATI-net has three blocks with signal capture, signal analysis, and prediction, which are related to the blocks of bio-audiovisual cognition system. The layer-by-layer processing of ATI-net mimics the framework. Moreover, each block of ATI-net has a clear layer function: 1) Through mimicking that the eye and ear respectively capture light and sound stimulus and then convert them into multiple electrical signals, the signal capture block performs multi-branch convolutions with varying kernels to extract multi-scale features and enhance the feature engineering, where Mish and Linear functions are used for nonlinear and linear features, respectively. 2) Through mimicking that the right and left hemispheres independently handle electrical signals and remember the core information, the signal analysis block designs two structures triggered by

Transformers and CNNs to capture nonlinear and linear evolutionary rules, where residual connections avoid the information loss. 3) Through mimicking that a higher cerebral cortex fuses processed information for shooting the prey and understanding the environment, the prediction block integrates the nonlinear and linear information to make the GHI prediction. Overall, the hierarchical processing and collaboration of the above blocks ensure that ATI-net captures the dynamics of the GHI series and forecasts them effectively.

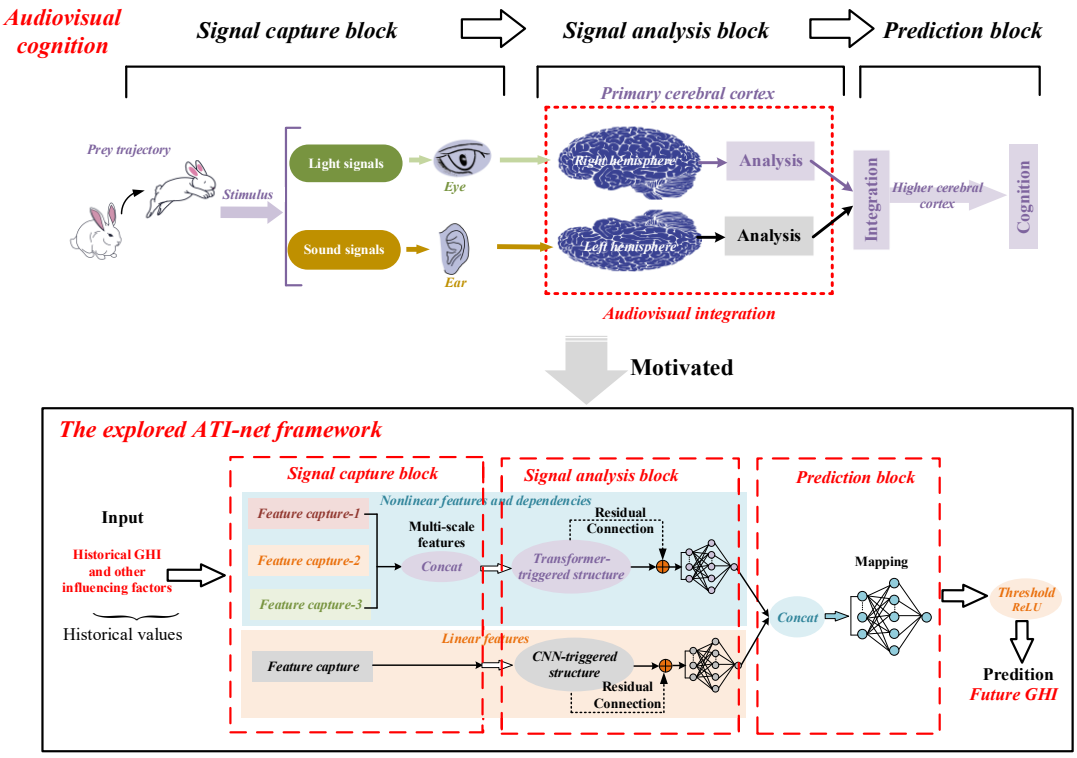


Fig. 3: The audiovisual-motivated framework for the ATI-net.

3.2. Internal construction of the ATI-net

Internal constructions of signal capture, signal analysis, and prediction blocks are presented in detail.

1) Signal capture block

As shown in Fig. 4, the light and sound stimulus are respectively captured by the eye and ear and then are converted into neuronal electrical signals for transferring information and suppressing

irrelevant interference. For example, the eye converts the light stimulus into electricity signal counting frequencies, electricity signal density frequencies, group electricity signal activity frequencies, etc., to transfer important information. Thus, the main function of this block is to convert the input multivariate series into useful coded features. One-dimensional convolution is famous for powerful parallel feature extraction and is applied in this block. From [38], the bio-visual system captures multi-scale time-related features, and multi-branch one-dimensional convolutions with various kernels are designed, where a big convolution kernel learns overall features on a big time scale, and a small convolution kernel learns specific features on a small time scale. As performed in [44], the kernel sizes are set to 1×1, 1×3, and 1×5. Besides, the traditional ReLU activation function has the problem of stopping passing information when the input is negative. The advanced Mish activation function addresses this problem via setting a small disturbance [45] and is taken to mine nonlinear features. Table 1 describes the difference between ReLU and Mish functions. The bio-auditory system does not have multi-scale features and is mimicked by the one-dimensional convolutions with fixed convolution kernel and Linear activation function for mining linear relationships. From [37], the kernel sizes are set to 1×3. Details of this block are:

$$\boldsymbol{S}_i = \left(\boldsymbol{s}_i, \boldsymbol{s}_{i+1}, ..., \boldsymbol{s}_{i+n}\right) \tag{8}$$

$$\boldsymbol{G}_i = \left(g_i, g_{i+1}, ..., g_{i+n}\right) \tag{9}$$

$$\begin{cases} \mathbf{A} = Mish\left(\boldsymbol{S}_i * \boldsymbol{w}_1 + \boldsymbol{b}_1\right) & //\text{ Kernel size: } 1\times1 \\ \mathbf{B} = Mish\left(\boldsymbol{S}_i * \boldsymbol{w}_2 + \boldsymbol{b}_2\right) & //\text{ Kernel size: } 1\times3 \\ \mathbf{C} = Mish\left(\boldsymbol{S}_i * \boldsymbol{w}_3 + \boldsymbol{b}_3\right) & //\text{ Kernel size: } 1\times5 \\ \mathbf{D} = Concact\left(\mathbf{A}, \mathbf{B}, \mathbf{C}\right) \end{cases} \tag{10}$$

$$\mathbf{E} = Linear\left(\boldsymbol{G}_i * \boldsymbol{w}_4 + \boldsymbol{b}_4\right) \quad //\text{ Kernel size: } 1\times3 \tag{11}$$

where $\boldsymbol{S}_i$ is the input multivariate series, including the GHI and related weather conditions; $\boldsymbol{G}_i$ is the input GHI series; $Mish(*)$ is the Mish function $Linear(*)$ is the Linear function; $\mathbf{D}$ is the

extracted multi-scale nonlinear features; **E** is the captured linear features; * is the convolution product.

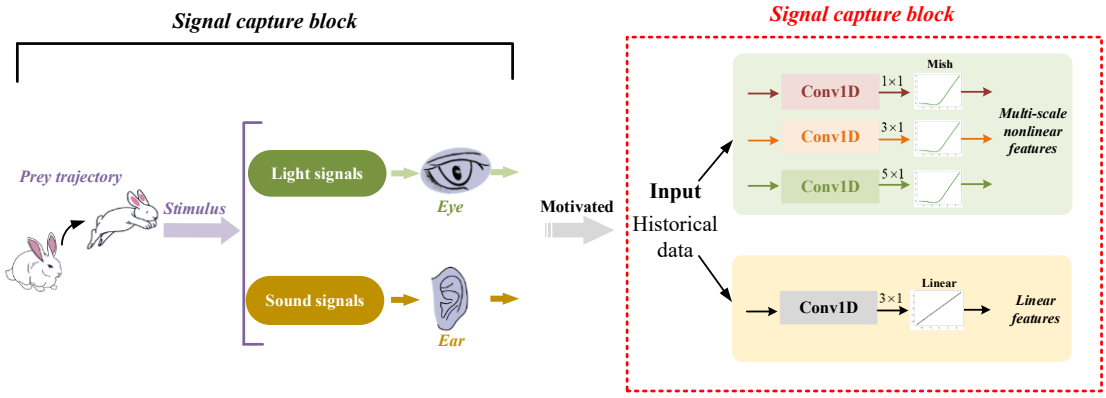
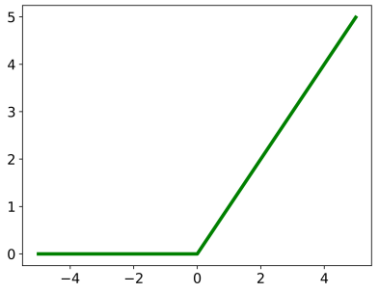
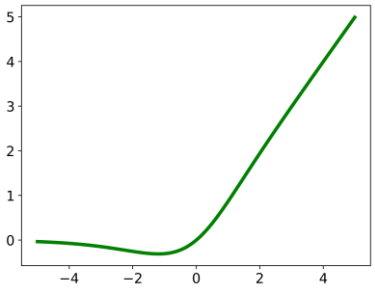


Fig. 4: Description of the signal capture block.

Table 1: Descriptions of ReLU and Mish activation functions

ReLU: $f(x)=\max(0,x)$	Mish: $f(x)=x\tanh\left(\ln\left(1+e^{x}\right)\right)$
	

2) Signal analysis block

As shown in Fig. 5, electricity signals are independently processed via right and left hemispheres in the primary cerebral cortex, where left and right hemispheres not only remember core information but also show some complementarity. Thus, the main functions of this block are the mining of time evolutionary rules, the extraction of deep-level linear relationships and the integration of them. The Transformer shows a powerful capability of extracting temporal dependencies, and CNN demonstrates the capability of layer-by-layer handling features. Since there are some similarities and complementarities, the right and left hemispheres are mimicked by Transformer encoders and CNNs,

respectively. In addition, since different regions in the biological brain have preferences, the activation functions of Transformer encoders and CNNs divide input information into two groups for respectively handling. The residual connection is adopted to fuse shallow and deep information, which allows not only the flexible application of intermediate information but also the rapid transfer of shallow information to deep information [31]. In summary, this block combines the advantages of Transformer, CNN, and residual connection. On the one hand, this block flexibly utilizes important information in time and feature dimensions. On the other hand, it transmits data from the shallow layer to deep layer for improving the backpropagation and reusing features. Moreover, the flatten layer and fully connected layer are adopted for converting the feature matrix. Details of this block are:

$$\begin{cases} \mathbf{F} = Transformer\left(\mathbf{D}\right) & \text{// Right hemisphere} \\ \mathbf{H} = Transformer\left(\mathbf{F}\right) \\ \mathbf{I} = Transformer\left(\mathbf{H}\right) \\ \mathbf{J} = \mathbf{F} + \mathbf{H} + \mathbf{I} \\ \hat{\mathbf{J}} = Flatten(\mathbf{J}) \\ \mathbf{K} = \hat{\mathbf{J}} \square \boldsymbol{w}_5 + \boldsymbol{b}_5 \end{cases} \tag{12}$$

$$\begin{cases} \mathbf{L} = Linear\left(\mathbf{E} * \boldsymbol{w}_6 + \boldsymbol{b}_6\right) & \text{// Left hemisphere} \\ \mathbf{M} = Linear\left(\mathbf{L} * \boldsymbol{w}_7 + \boldsymbol{b}_7\right) \\ \mathbf{N} = \mathbf{E} + \mathbf{L} + \mathbf{M} \\ \hat{\mathbf{N}} = Flatten(\mathbf{N}) \\ \mathbf{O} = \hat{\mathbf{N}} \square \boldsymbol{w}_8 + \boldsymbol{b}_8 \end{cases} \tag{13}$$

where $Transformer\left(*\right)$ is the Transformer encoder; $Flatten\left(*\right)$ is the flatten layer; $\square$ is the fully connected product; $\mathbf{K}$ is the output of extracted temporal dependencies; $\mathbf{O}$ is the output of capturing deep-level linear relationships.

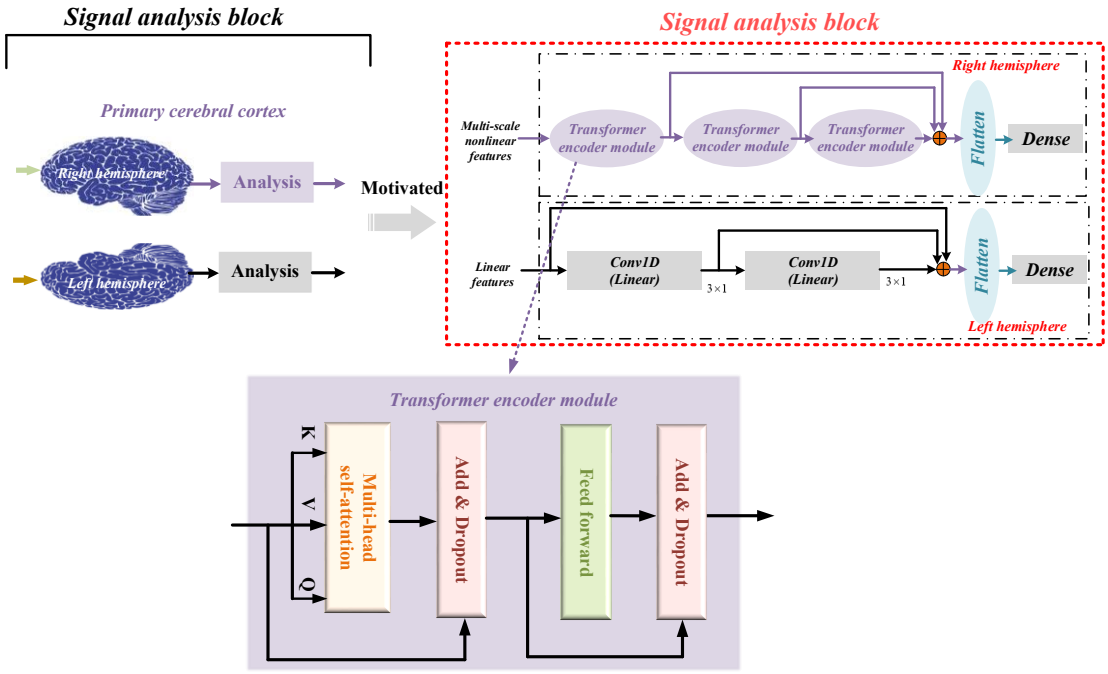


Fig. 5: Description of the signal analysis block.

3) Prediction block and summary

The processed information from right and left hemispheres is integrated in a higher cerebral cortex for shooting the prey and understanding the environment. Thus, the main function of this block is to integrate the nonlinear dependencies and linear relationships for making GHI predictions. The divide-and-conquer strategy attenuates the loss of useful linear and nonlinear information. There are four integration strategies for features, including Concat [39], Attention [46], SENet [47], and Self-attention [48] integration, as shown in Fig. 6. The Concat integration preserves features based on equal weights and produces no loss of features. Other integration strategies all assign weights to features through their preferences, thus enhancing what they consider to be core features and ignoring what they consider to be ineffective features. Based on experimental results in Subsection 5.3, the Concat integration shows the best results and is used in this block. The flatten layer and fully connected layer map the relationships between features and the target. Additionally, according to the domain knowledge of GHI, we set the ReLU function as the threshold to avoid the negative prediction. Details

of this block are:

$$\begin{cases} \mathbf{P} = Contact(\mathbf{K},\ \mathbf{O}) \\ \widehat{\mathbf{P}} = Flatten(\mathbf{P}) \\ \tilde{g}_{i+n+1} = \widehat{\mathbf{P}} \square \boldsymbol{w}_9 + \boldsymbol{b}_9 \\ \hat{g}_{i+n+1} = ReLU\left(\tilde{g}_{i+n+1}\right) \end{cases} \tag{14}$$

where $\hat{g}_{i+n+1}$ is the predicted GHI; $Contact(*)$ is the Concat layer.

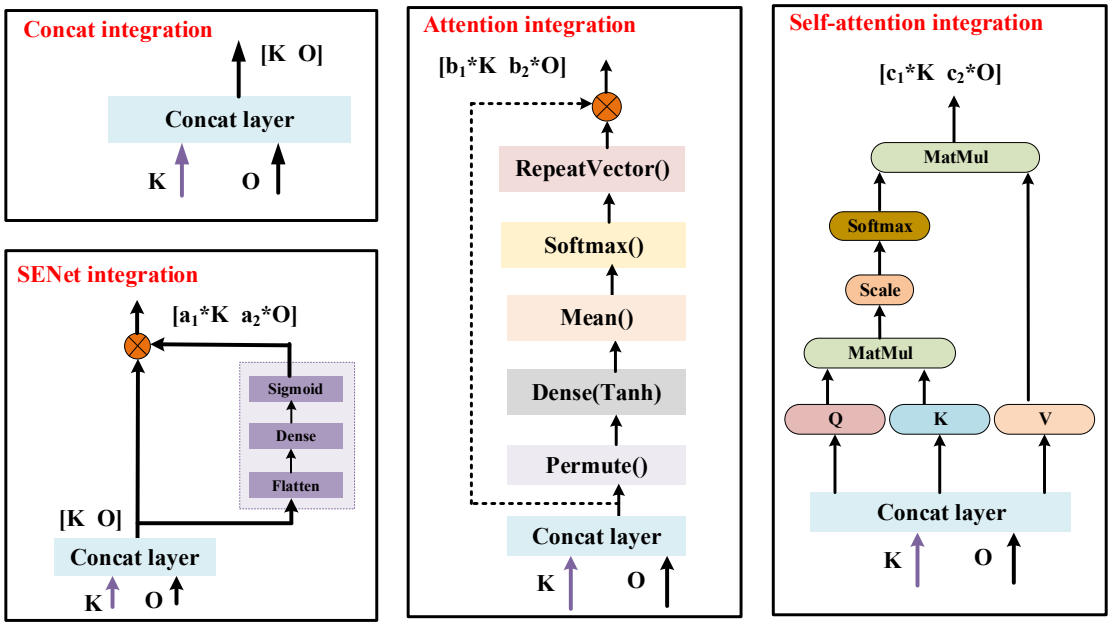


Fig. 6: Description of four integration strategies.

Fig. 7 presents the specific construction of the ATI-net for GHI prediction. Overall, the explored ATI-net solves some research gaps: 1) Compared to most uninterpretable deep networks, the design framework of the ATI-net is supported by the audiovisual integrated system, bringing interpretability at the biological level. 2) Compared to the fixed receptive field of most deep networks, the ATI-net incorporates multi-branch convolutions with varying kernels and Mish function for increasing sensitivity to time horizons. 3) Compared to the single use of deep networks, the ATI-net integrates the advantages of Transformer, CNN, and residual connection, not only flexible using important information in time and feature dimensions but also combining data of the shallow and deep layer for improving the backpropagation and reusing features. Thus, the ATI-net is one of the reliable

alternatives to GHI prediction.

Remark: When conducting time series forecasting, the LSTM and its variants, as Markov decision processes in essence, generally cannot learn the global information moment from all input time steps well [49]. The contribution of the last input time step is usually larger than the time step ahead. The proposed method directly establishes the mapping relationships between the prediction and all inputs, which effectively captures global information.

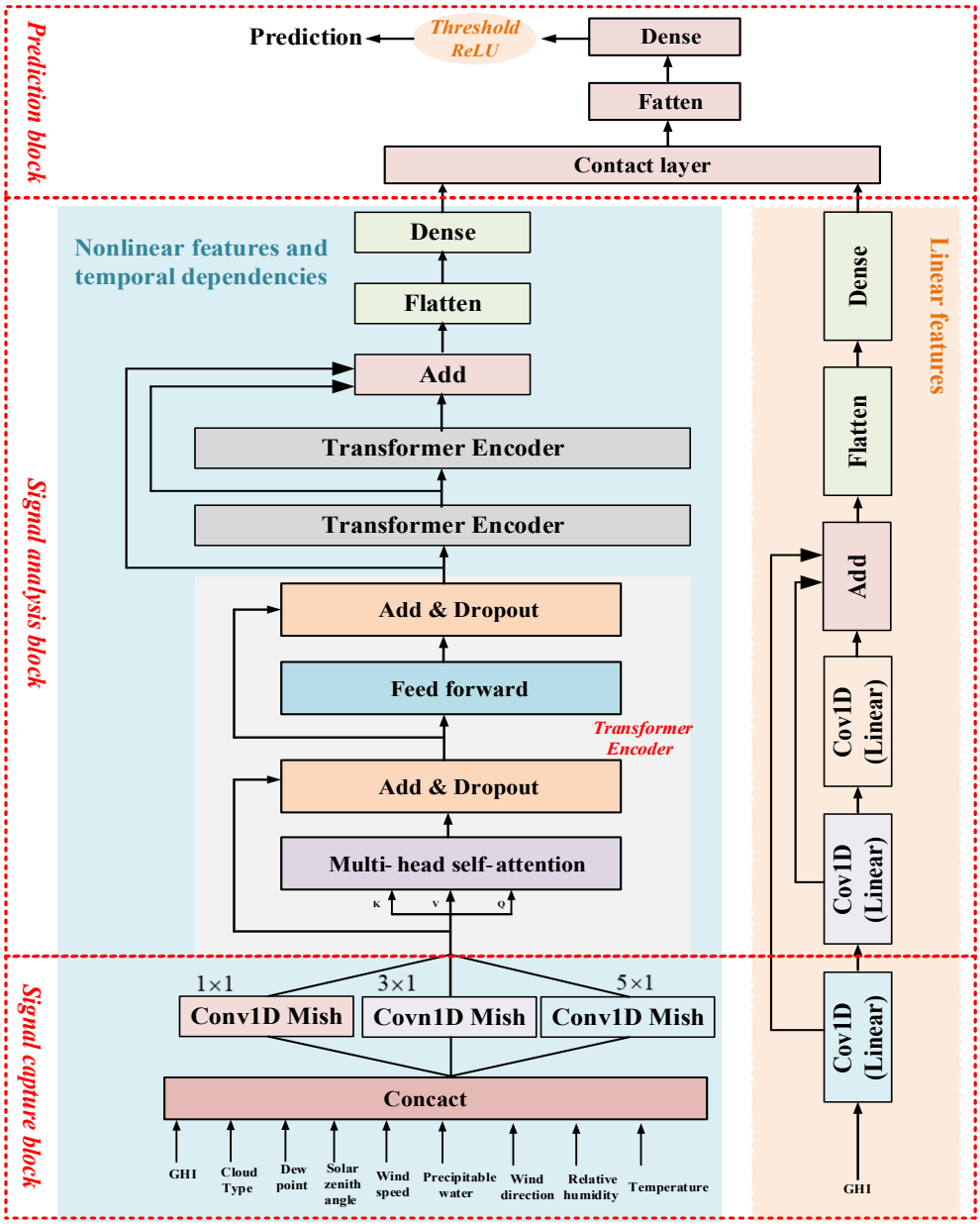


Fig. 7: Construction of the ATI-net for GHI prediction.

4. Data description and experiments

Six kinds of experiments are applied to show the predictive advantages of the ATI-net in the SI series, and the description of experimental arrangement is given in Table 2.

Table 2: Description of experimental arrangement.

Experiments	Goals
Experiment A: Comparison with recursive methods Experiment B: Comparison with parallel methods Experiment C: Comparison with hybrid methods Experiment D: Comparison with existing bionic prediction methods	Evaluate the prediction accuracy
Experiment E: Ablation studies	Evaluate the network structure
Experiment F: Multi-step-ahead prediction	Evaluate the multi-step prediction performance

4.1. Data description

From [50], the 30-minute SI prediction is one of the more interesting and challenging issues. The real SI datasets from New York and Hawaii are used to validate the proposed ATI-net and benchmarks, where the sampling time is 30 minutes. The above datasets are publicly available from Kaggle and contain 9 variables: GHI, Cloud Type, Dew Point, Solar Zenith Angle, Wind Speed, Precipitable water, Wind Direction, Relative Humidity, and Temperature. The data quality check is the primary step after collecting data. As performed in [51, 52], the data quality check is beneficial to train prediction models and includes the removal of missing data and abnormal recordings, etc. When making the prediction, the input contains all historical variables, and the output is the GHI in the next time stamp. All prediction experiments are carried out by the sliding window method. There are three common types for dividing training and test samples, including 80% : 20%, 75% : 25%, and 70% : 30%. As performed in [51], this paper selects the division type of 80% training samples and 20% test samples. The number of total samples, training-validation samples, and test samples are 3000, 2400, and 600, where the training and validation samples are 1932 and 468, respectively. As performed in [53, 54], the training

and validation samples are applied to select the hyperparameters for the proposed method. Detailed training-validation data, test data, and partial autocorrelation function (PACF) values are shown in Fig. 8. We can see that: 1) As performed in [38], since PACF values after 59 and 61 fall in the confidence interval, they are set as input window sizes for two datasets, respectively. 2) The GHI series have strong nonlinearity and randomness. Fig. 9 shows the Pearson and Spearman coefficients of different variables, where Pearson and Spearman coefficients present linear and nonlinear relationships, respectively. As performed in [46], due to the small number, all the variables are kept. Table 3 reports the statistical results of the above datasets. We can clearly see that Jarque-Bera statistics are larger than the zero value, which means that the GHI series do not belong to the normally distributed.

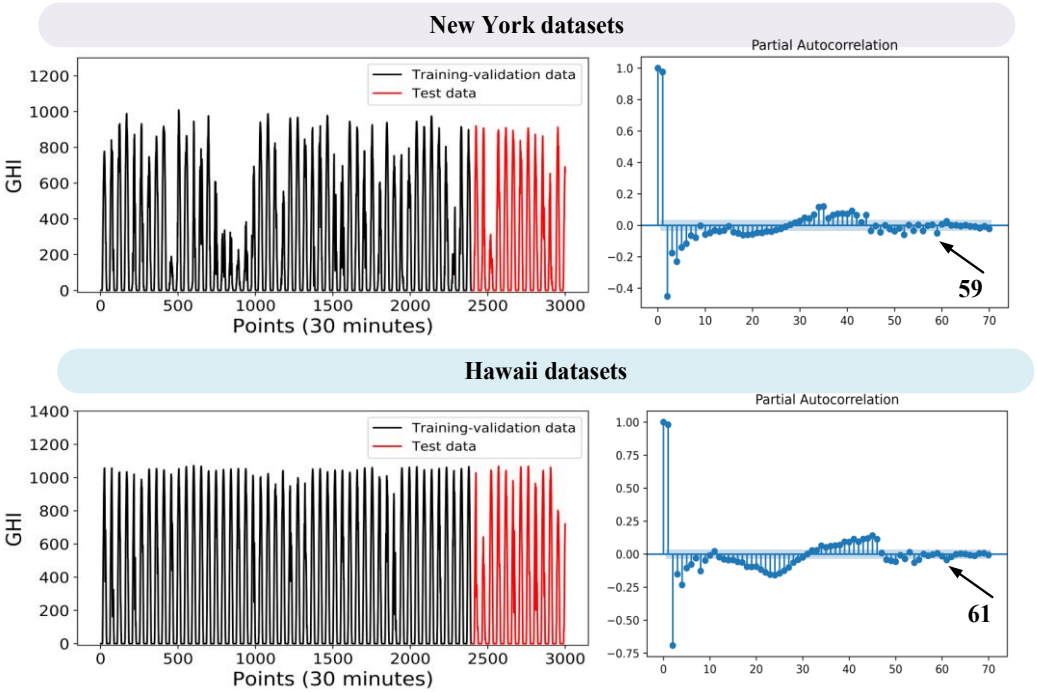


Fig. 8: Detailed descriptions of two GHI datasets.

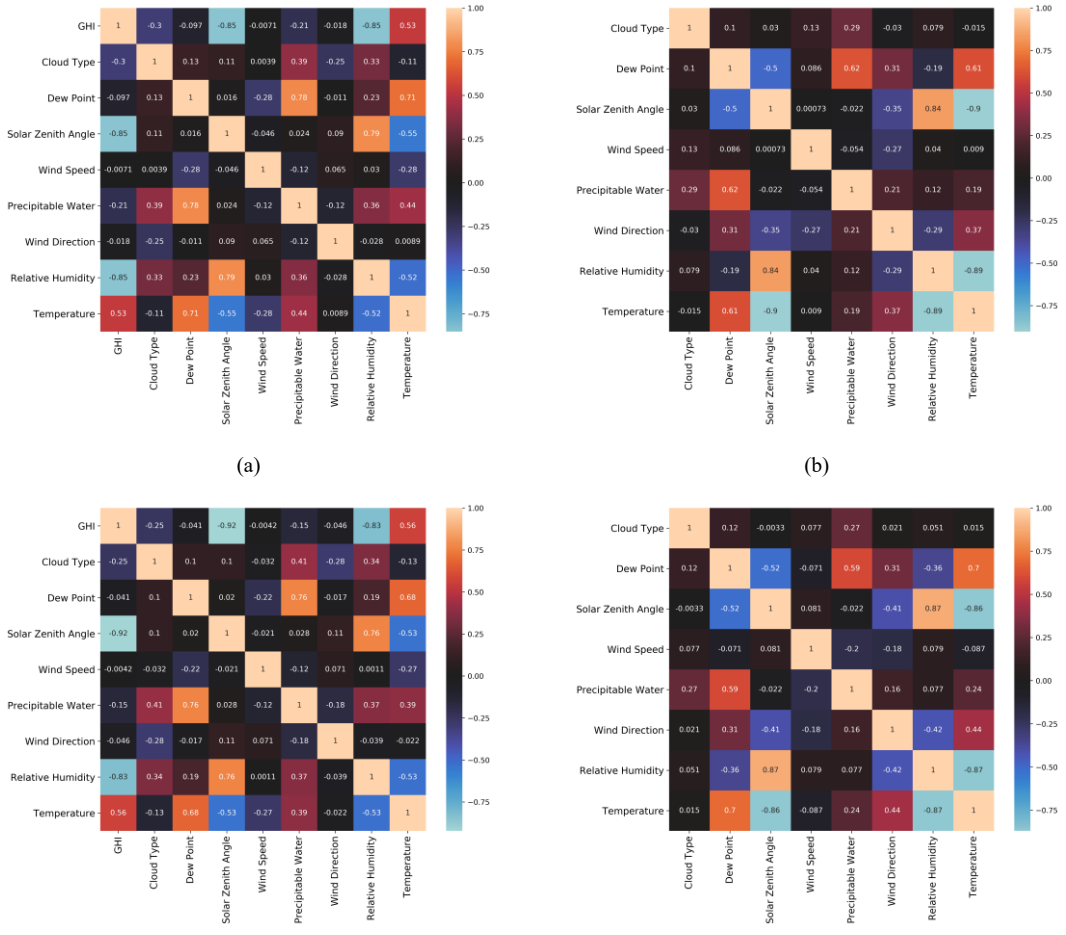


Fig. 9: (a) Pearson coefficient of New York datasets; (b) Pearson coefficient of Hawaii datasets; (c) Spearman coefficient of New York datasets; (d) Spearman coefficient of Hawaii datasets.

Table 3: Statistical characteristics of two datasets.

Datasets	Parameters	Min	Max	Mean	Median	STD	Skewness	Kurtosis	Jarque-Bera
Training-validation datasets in New York	GHI	0.00	1009.00	251.83	90.50	309.14	0.93	-0.59	1634.52
	Cloud Type	0.00	9.00	3.09	3.00	3.16	0.30	-1.59	2145.03
	Dew Point	5.70	26.10	18.95	19.50	3.77	-0.67	0.17	978.36
	Solar Zenith Angle	17.29	118.57	72.18	75.31	33.06	-0.19	-1.36	1918.56
	Wind Speed	0.10	4.90	1.82	1.80	0.79	0.50	0.27	843.81
	Precipitable water	0.90	6.20	3.56	3.50	1.35	-0.05	-1.15	1720.91
	Wind Direction	0.00	359.00	209.55	216.50	84.90	-0.45	-0.41	1239.04
	Relative Humidity	44.80	100.00	83.18	87.33	14.00	-0.58	-0.89	1647.99
	Temperature	9.50	31.90	22.20	22.40	4.15	-0.19	-0.36	1140.76
Test datasets in New York	GHI	0.00	920.00	254.71	73.00	309.44	0.85	-0.82	435.66
Training-validation datasets	GHI	0.00	1072.00	324.60	78.00	386.78	0.70	-1.13	1903.21

in Hawaii	Cloud Type	0.00	8.00	0.64	0.00	1.89	2.91	1.17	5135.38
	Dew Point	13.20	20.50	16.46	16.40	1.15	0.14	-0.23	1048.87
	Solar Zenith Angle	1.00	139.69	79.02	82.95	42.51	-0.20	-1.27	1840.17
	Wind Speed	0.20	6.90	1.62	1.50	0.93	1.53	4.62	1198.51
	Precipitable water	1.10	3.40	2.23	2.20	0.45	0.05	-0.34	1113.58
	Wind Direction	0.00	359.00	104.95	86.00	66.07	1.99	3.95	1682.49
	Relative Humidity	60.71	100.00	92.59	100.00	10.45	-1.13	-0.06	1440.35
	Temperature	13.30	23.10	17.80	17.00	2.40	0.48	-1.04	1724.83
Test datasets in Hawaii	GHI	0.00	1069.00	281.91	61.00	356.72	0.91	-0.65	416.58

4.2. Benchmarks, evaluation metrics, and network parameters

- *Benchmarks*

From [11], since the GHI series usually is recognized as the nonlinear time series, we do not select the classical statistical methods. The proposed ATI-net is compared with 18 benchmarks, which include two shallow ML methods (SVR [19] and XGBoost [21]), five recursive deep ML methods (LSTM [24], BI-LSTM [26], GRU [25], BI-GRU [27], and BI-LSTM-GRU [28]), five parallel deep ML methods (MLP [29], ResNet, CNN [30], TCN [32], and Transformer [33]), and six hybrid-recursive-parallel deep ML methods (CNN-LSTM [34], CNN-BILSTM, CNN-GRU, CNN-BIGRU, LSTM-CNN [4], and GRU-CNN [4]).

- *Evaluation metrics*

The performance of the ATI-net and benchmarks are evaluated by several key evaluation metrics, including the root mean squared error (RMSE), mean absolute error (MAE), relative root mean squared error (RRMSE), Theil inequality coefficient (TIC), and index of agreement (IA). The above-used metrics give measures of the method accuracy, precision, and consistency, allowing for the objective assessment of the prediction performance. Through considering these metrics, researchers and practitioners gain valuable insights into the quality of the forecasts and make informed decisions based on the results [55]. In general, the smaller the RMSE, MAE, RRMSE, and TIC values are, and the

higher the IA value is, the better the prediction model is. Specific formulas of the above metrics are defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(g_j - \hat{g}_j\right)^2} \tag{15}$$

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|g_j - \hat{g}_j\right| \tag{16}$$

$$RRMSE = \frac{RMSE}{\overline{g}} \times 100\% \tag{17}$$

$$TIC = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(g_j - \hat{g}_j\right)^2} \Bigg/ \left(\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(g_j\right)^2} + \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\hat{g}_j\right)^2}\right) \tag{18}$$

$$IA = 1 - \frac{\sum_{j=1}^{n}\left(g_j - \hat{g}_j\right)^2}{\sum_{j=1}^{n}\left(\left|\hat{g}_j - \overline{g}\right| + \left|g_j - \overline{g}\right|\right)^2} \tag{19}$$

where g_j is the actual GHI value; $\hat{g}_j$ is the prediction value; n is the sample number; $\overline{g}_j$ is the average value.

- *Network parameters*

Specific parameters of the ATI-net for predicting GHI are presented in Table 4. Before training all prediction models, the inputs are standardized via the following formula to accelerate the convergence:

$$\tilde{g}_i = \frac{g_i - \mu}{\sqrt{\sigma^2}} \tag{20}$$

where μ is the average value; σ is the standard deviation value; g_i is the actual value; $\tilde{g}_i$ is the value after standardization.

Moreover, numbers of training epoch and batch size are set to 200 and 50, learning rate changes from 0.3 to 0.00001 with training epochs, optimizer and loss function are set to Adam and MSE, and L2 regularization is applied to avoid overfitting. The above hyperparameters affect the prediction performance and are not changed in the training processes for the New York and Hawaii datasets to guarantee the fairness in comparison. Experimental results in Section 5 show that the above

hyperparameters are reasonable.

Table 4: Parameters of the explored ATI-net

Blocks	Layer type	Layer name	Two GHI datasets		
			Kernel	Filter	Activation
Signal capture block	Convolution	Visual System	32	1×1	Mish
	Convolution		32	3×1	Mish
	Convolution		32	5×1	Mish
	Convolution	Auditory system	32	3×1	Linear
Signal analysis block	Transformer encoder	Right hemisphere	Numbers of head and hidden size of feedforward network are set to 4 and 32		
	Transformer encoder				
	Transformer encoder				
	Flatten		/	/	/
	Dense		1	/	/
	Convolution	Left hemisphere	32	3×1	Linear
	Convolution		32	3×1	Linear
	Flatten		/	/	/
	Dense		1	/	/
Prediction block	Flatten	/	/	/	/
	Output (Dense)	/	1	/	/

4.3. Experiment A: Comparison with recursive methods

This experiment presents the prediction results of two shallow ML methods (SVR and XGBoost), five recursive methods (LSTM, BI-LSTM, GRU, BI-GRU, and BI-LSTM-GRU), and the proposed method under New York and Hawaii datasets. Tables 5 and 6 give results of five evaluation metrics, and the best results are marked in bold.

We can clearly see that: 1) The proposed ATI-net performs better than two shallow ML methods and five recursive methods in all test situations. Details are $\mathrm{RMSE}=46.8092$, $\mathrm{MAE}=25.7922$, $\mathrm{RRMSE}=18.3778$, $\mathrm{TIC}=0.0586$, and $\mathrm{IA}=0.9941$ for New York datasets, and are $\mathrm{RMSE}=59.8059$, $\mathrm{MAE}=27.0089$, $\mathrm{RRMSE}=21.2147$, $\mathrm{TIC}=0.0657$, and $\mathrm{IA}=0.9929$ for Hawaii datasets, indicating that the ATI-net effectively presents the nonlinear fluctuations of GHI. Bionically-inspired deep networks are promising to enhance GHI prediction performance. 2) XGBoost

shows better prediction than SVR and some recursive methods, implying it is wise to cope with nonlinear time series via integrating multiple decision trees. 3) Except for BI-GRU in New York datasets, bidirectional dependency extractions of BI-LSTM and BI-GRU are more useful than unidirectional dependency extractions of LSTM and GRU in predicting GHI. 4) BI-LSTM-GRU is better than BI-LSTM under two datasets, which highlights the integrated advantages of BI-LSTM and GRU. 5) GRU and BI-GRU are better than LSTM and BI-LSTM, suggesting that fewer gating mechanisms attenuate the model's forgetting of core information and thus improve predictive performance. 6) The prediction differences between the LSTM and GRU are significant. The GRU unit only has two gates, including reset and update gates, and the LSTM unit has three gates, including input, forget, and output gates. The gate mechanism mainly controls the information flow. Thus, when stacking multiple LSTM or GRU units, the prediction limitation of GRU is weaker than that of LSTM, and GRU's performance may be much better than that of LSTM.

Furthermore, to clearly demonstrate the performance, detailed prediction results of the XGBoost, BI-LSTM, BI-GRU, BI-LSTM-GRU, and the proposed method are shown in Figs. 10-11 for New York and Hawaii datasets. Sum values of prediction absolute errors are also presented in Figs. 10-11. It can be observed that: The prediction results and sum values of absolute errors for the proposed method are closer to the actual values and small values, respectively.

Remark A: Experimental results show that the proposed method significantly improves the prediction accuracy compared with 7 benchmarks, and the audiovisual cognitive process is promising to provide a design framework for the deep network in predicting GHI. Due to the bionically-inspired idea, the proposed method has relatively small time lags when predicting. In addition, the GRU or the integration of LSTM and GRU is more interesting than a single LSTM in handling the GHI series. In

summary, the study provides valuable insights for practitioners in the field of solar power generation.

Table 5: Results for GHI prediction compared with recursive models in New York

Methods	RMSE	MAE	RRMSE	TIC	IA
SVR	102.0658	67.4048	40.0721	0.1263	0.9709
XGBoost	54.0299	28.9959	21.2127	0.0676	0.9922
LSTM	103.9752	57.8142	40.8218	0.1315	0.9697
BI-LSTM	80.7345	49.0266	31.6973	0.1010	0.9823
GRU	51.4109	27.5101	20.1845	0.0635	0.9931
BI-GRU	56.6706	31.4372	22.2495	0.0699	0.9915
BI-LSTM-GRU	67.4493	39.6314	26.4813	0.0834	0.9879
The proposed method	**46.8092**	**25.7922**	**18.3778**	**0.0586**	**0.9941**

Table 6: Results for GHI prediction compared with recursive models in Hawaii

Methods	RMSE	MAE	RRMSE	TIC	IA
SVR	123.1661	70.8199	43.6901	0.1348	0.9679
XGBoost	73.7453	32.7579	26.1593	0.0813	0.9891
LSTM	88.2828	44.6846	31.3161	0.0969	0.9843
BI-LSTM	84.8221	43.0288	30.0885	0.0940	0.9852
GRU	65.6093	30.1416	23.2733	0.0719	0.9914
BI-GRU	63.7641	30.2667	22.6187	0.0706	0.9918
BI-LSTM-GRU	66.1098	32.8578	23.4508	0.0730	0.9911
The proposed method	**59.8059**	**27.0089**	**21.2147**	**0.0657**	**0.9929**

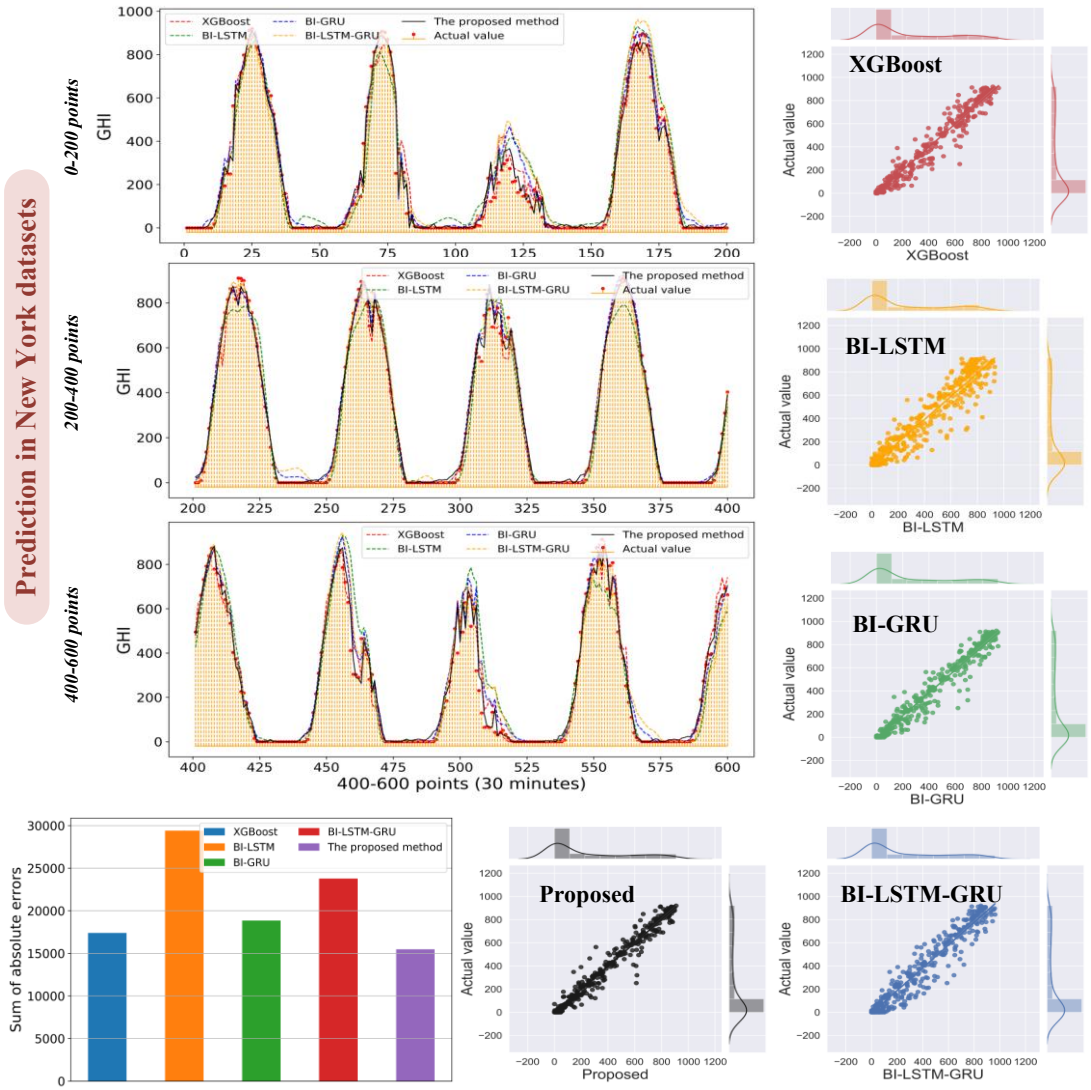


Fig. 10: Detailed results of GHI prediction compared with recursive models under New York datasets.

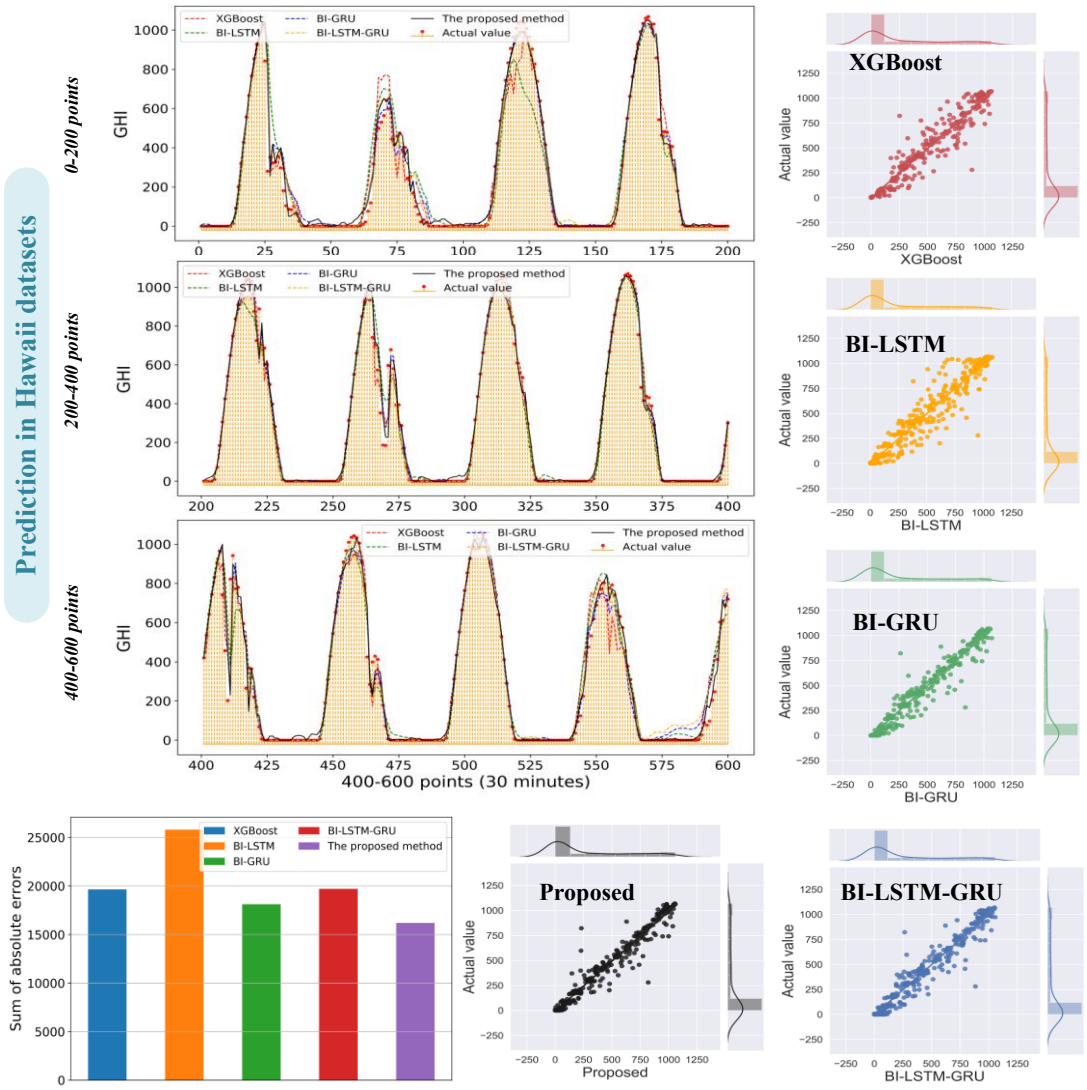


Fig. 11: Detailed results of GHI prediction compared with recursive models under Hawaii datasets.

4.4. Experiment B: Comparison with parallel methods

This experiment assesses the prediction results of two shallow ML methods (SVR and XGBoost), five parallel methods (MLP, ResNet, CNN, TCN, and Transformer), and the proposed method under New York and Hawaii datasets. Tables 7 and 8 list the corresponding results of evaluation indicators. Fig. 12 presents detailed prediction results of the ResNet, CNN, TCN, Transformer, and the proposed method.

Details are: 1) It is evident that the proposed ATI-net outperforms other parallel methods in terms of prediction and correlation with the actual data, which indicates that the ATI-net is capable to provide reliable and precise GHI predictions. The superiority of the ATI-net is further illustrated in Fig. 12, which showcases its capability to accurately predict the GHI based on New York and Hawaii datasets. 2) The ResNet with residual connections is better than MLP under two datasets, implying that residual connections are beneficial to fuse shallow and deep information and further avoid information forgetting. It is effective to combine residual connections with Transformers, enhancing the remember capability. 3) The CNN is better than MLP and ResNet, which implies that parallel convolutional features are more efficient than features extracted by fully connected layers in fitting GHI. 4) The Transformer and CNN rank second and third, being better to construct the ATI-net. 5) When conducting time series forecasting, the MLP and CNN all directly establish the mapping relationships between the prediction and all inputs, being suitable to mine the global information from all inputs. Thus, some results of MLP and CNN are better than those of LSTM.

Remark B: Experimental results show that the proposed ATI-net is better than parallel ML methods. Moreover, the Transformer and CNN are promising compared with other parallel methods and are suitable to be injected into ATI-net for improving GHI prediction. The high prediction accuracy of ResNet highlights that the residual connection is an effective strategy to avoid information forgetting.

Table 7: Results for GHI prediction compared with parallel models in New York

Methods	RMSE	MAE	RRMSE	TIC	IA
SVR	102.0658	67.4048	40.0721	0.1263	0.9709
XGBoost	54.0299	28.9959	21.2127	0.0676	0.9922
MLP	77.4159	43.6902	30.3943	0.0951	0.9841
ResNet	72.4074	41.8097	28.4279	0.0892	0.9861
CNN	53.7075	29.7626	21.0862	0.0662	0.9924
TCN	73.4525	48.9597	28.8383	0.0901	0.9856

Transformer	49.4463	27.8573	19.4132	0.0612	0.9936
The proposed method	**46.8092**	**25.7922**	**18.3778**	**0.0586**	**0.9941**

Table 8: Results for GHI prediction compared with parallel models in Hawaii

Methods	RMSE	MAE	RRMSE	TIC	IA
SVR	123.1661	70.8199	43.6901	0.1348	0.9679
XGBoost	73.7453	32.7579	26.1593	0.0813	0.9891
MLP	88.9821	40.6536	31.5642	0.0963	0.9845
ResNet	88.0544	40.9043	31.2351	0.0954	0.9847
CNN	83.7086	40.0458	29.6936	0.0910	0.9860
TCN	93.0717	50.8160	33.0149	0.1012	0.9826
Transformer	72.1568	33.9764	25.5958	0.0789	0.9896
The proposed method	**59.8059**	**27.0089**	**21.2147**	**0.0657**	**0.9929**

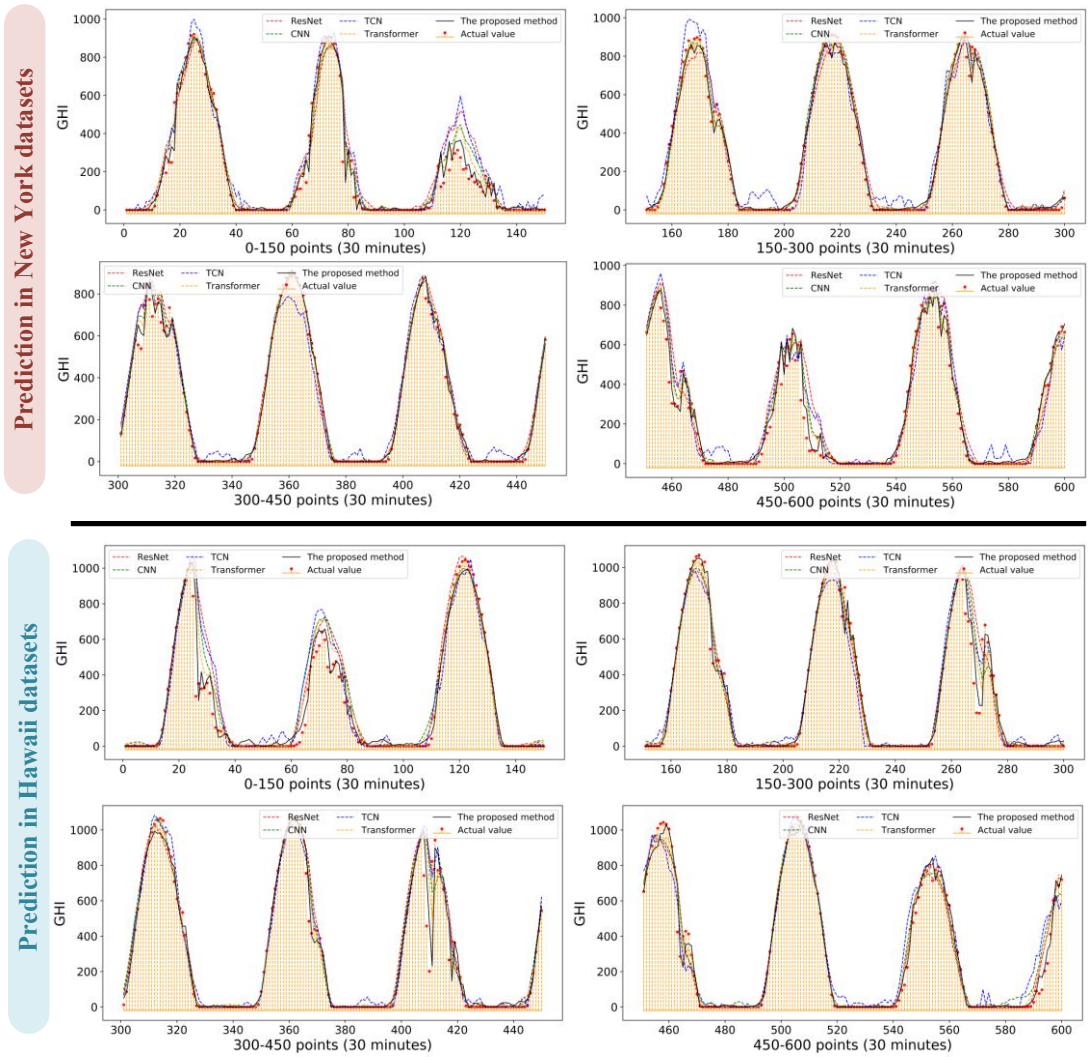


Fig. 12: Detailed results of GHI prediction compared with parallel models under New York and Hawaii datasets.

4.5. Experiment C: Comparison with hybrid methods

This experiment evaluates the performance of two shallow ML methods (SVR and XGBoost), six hybrid methods (CNN-LSTM, CNN-BILSTM, CNN-GRU, CNN-BIGRU, LSTM-CNN, and GRU-CNN), and the proposed method. Tables 9 and 10 reports results in terms of RMSE, MAE, RRMSE, TIC, and IA. The detailed prediction results of the CNN-LSTM, CNN-BIGRU, LSTM-CNN, GRU-CNN, and the proposed method are shown in Fig. 13.

Details are: 1) The proposed method outperforms six hybrid methods and integrates the advantages of CNN and Transformer. Specifically, the proposed method uses the power of multi-scale feature extraction, residual connection, and Transformer for accurate estimation. The multi-scale feature extraction enables the ATI-net to capture complex time-related information, the residual connection avoids the forgetting of important information, and Transformer helps the extraction of meaningful evolutionary rules. Based on the coordination, the proposed method effectively handles the complex GHI series and make effective predictions. 2) Hybrid methods, such as LSTM-CNN and GRU-CNN, are worse than hybrid methods, such as CNN-LSTM and CNN-GRU, which implies that temporal dependency extraction at the feature level is more favorable for GHI prediction. 3) Hybrid methods, such as CNN-LSTM and CNN-GRU, are better than hybrid methods, such as CNN-BILSTM and CNN-BIGRU, which means extraction of bidirectional dependency at the feature level does not apply to GHI prediction.

Remark C: Compared with six hybrid methods, the proposed ATI-net shows the best prediction accuracy. Through integrating the architectures of CNNs and Transformers, the ATI-net effectively handles the GHI intricacies and makes accurate predictions. Moreover, temporal dependency extraction at the feature level is better than that at the data layer, and it is also the reason that Transformers are

connected after multi-scale feature extraction in the ATI-net.

Table 9: Results for GHI prediction compared with hybrid models in New York

Methods	RMSE	MAE	RRMSE	TIC	IA
SVR	102.0658	67.4048	40.0721	0.1263	0.9709
XGBoost	54.0299	28.9959	21.2127	0.0676	0.9922
CNN-LSTM	54.6448	30.8804	21.4541	0.0673	0.9922
CNN-BILSTM	55.4497	30.8682	21.7702	0.0680	0.9920
CNN-GRU	48.2539	26.7879	18.9450	0.0598	0.9939
CNN-BIGRU	51.8246	29.6212	20.3469	0.0640	0.9929
LSTM-CNN	103.2979	64.1503	40.5559	0.1260	0.9714
GRU-CNN	62.3598	40.5393	24.4831	0.0763	0.9898
The proposed method	**46.8092**	**25.7922**	**18.3778**	**0.0586**	**0.9941**

Table 10: Results for GHI prediction compared with hybrid models in Hawaii

Methods	RMSE	MAE	RRMSE	TIC	IA
SVR	123.1661	70.8199	43.6901	0.1348	0.9679
XGBoost	73.7453	32.7579	26.1593	0.0813	0.9891
CNN-LSTM	79.8220	36.6607	28.3149	0.0868	0.9874
CNN-BILSTM	87.0945	39.7594	30.8946	0.0949	0.9848
CNN-GRU	70.2318	32.8936	24.9130	0.0767	0.9902
CNN-BIGRU	71.3419	31.5703	25.3068	0.0781	0.9899
LSTM-CNN	119.4558	65.0360	42.3740	0.1274	0.9721
GRU-CNN	82.9958	45.5930	29.4407	0.0899	0.9864
The proposed method	**59.8059**	**27.0089**	**21.2147**	**0.0657**	**0.9929**

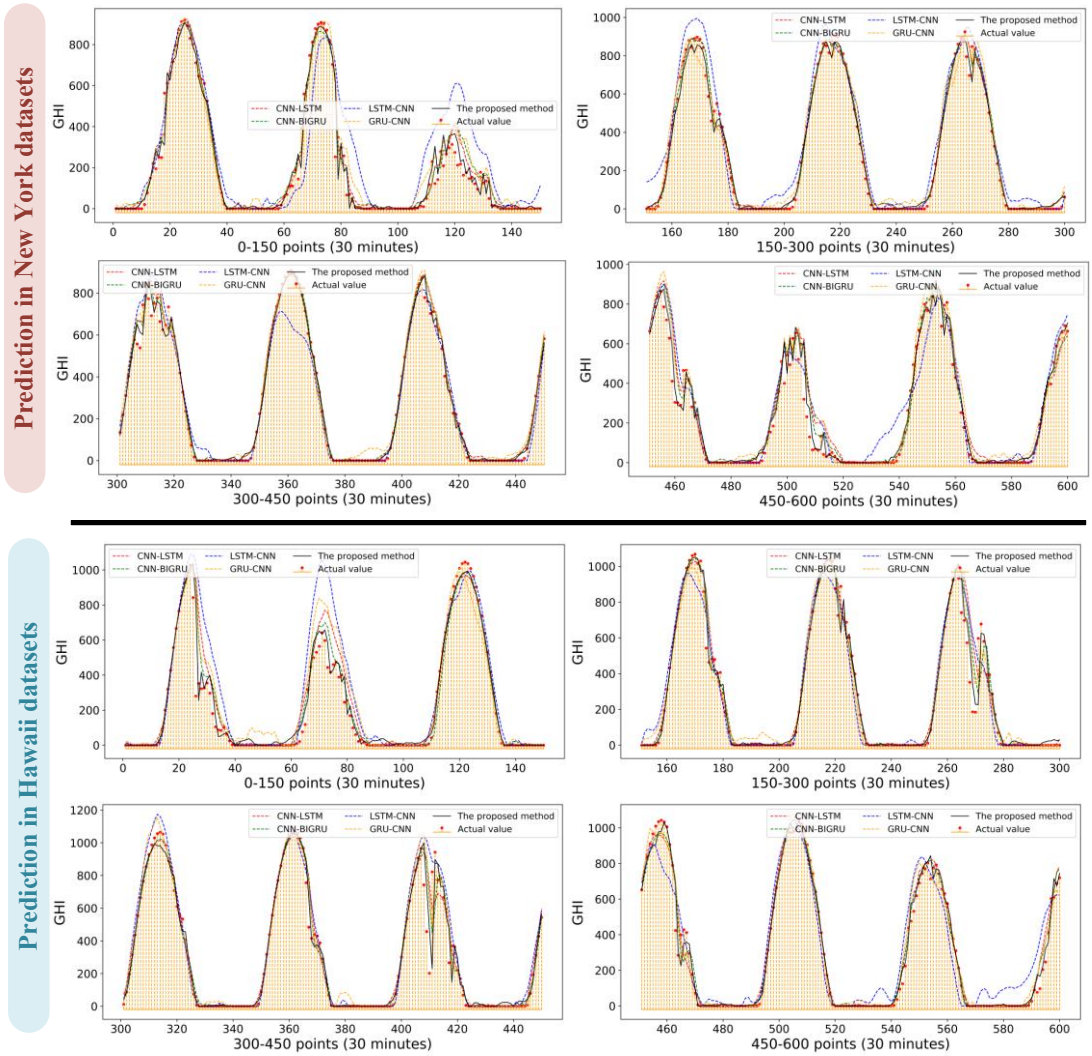


Fig. 13: Detailed results of GHI prediction compared with hybrid models under New York and Hawaii datasets.

4.6. Experiment D: Comparison with existing bionic prediction methods

The audiovisual cognition provides a superior framework to design the ATI-net for GHI prediction, and the ATI-net essentially is a bionically-inspired network. This experiment evaluates the prediction accuracy of the ATI-net and other four bionic deep prediction networks. This experiment re-sets these network parameters and re-trains these networks based on the New York and Hawaii datasets. Tables 11 and 12 give the prediction results in terms of RMSE, MAE, RRMSE, TIC, and IA.

Details are: 1) The ATI-net outperforms other four bionically-inspired deep networks, and the

AMFC-Net ranks second. The results imply that although bio-cognitive processes provide guidelines for network designs, specific applications need to corresponding cognitive processes. In other words, since the GHI series are similar to the prey trajectories, the audiovisual system is imitated to design the ATI-net. 2) The AMFC-Net and ACM-Net only imitate the bio-auditory system, and the MSV-Net only mimics the bio-visual system. Their results are worse than those of the ATI-net, implying that the audiovisual system is capable of integrating useful information.

Remark D: These methods have some predictive differences. The main reasons are that these bionic methods have different bionic imitations and motivations from specific applications. For example, the MSV-Net only imitates the bio-visual system for forecasting electricity prices, and the AMFC-Net only imitates the bio-auditory system for forecasting oil prices. The designs of the above two networks involve different data characteristics and bionic systems. In addition, the audiovisual cognition and the GHI prediction have some similarities, so this paper proposes such a bionic network. Experimental results imply that the audiovisual integrated system is better to be imitated to design networks for GHI prediction.

Table 11: Results for GHI prediction compared with existing bionic models in New York

Methods	RMSE	MAE	RRMSE	TIC	IA
ACM-Net in [37]	51.6760	26.8931	20.2886	0.0634	0.9931
MSV-Net in [38]	48.5697	27.5513	19.0690	0.0603	0.9938
PME-Net in [39]	49.0177	28.0271	19.2449	0.0607	0.9937
AMFC-Net in [47]	47.2365	26.5128	18.5456	0.0586	0.9941
The proposed method	**46.8092**	**25.7922**	**18.3778**	**0.0586**	**0.9941**

Table 12: Results for GHI prediction compared with existing bionic models in Hawaii

Methods	RMSE	MAE	RRMSE	TIC	IA
ACM-Net in [37]	77.7180	40.0400	27.5685	0.0847	0.9879
MSV-Net in [38]	81.8272	41.6752	29.0262	0.0890	0.9866
PME-Net in [39]	75.1787	34.1199	26.6678	0.0819	0.9888

AMFC-Net in [47]	69.9801	32.0878	24.8237	0.0764	0.9903
The proposed method	**59.8059**	**27.0089**	**21.2147**	**0.0657**	**0.9929**

4.7. Experiment E: Ablation studies

The experiment conducts ablation studies to show effects of different components in ATI-net for predicting GHI, and results are given in Tables 13 and 14. Details are: 1) Compared with other four methods, the ATI-net presents the highest prediction accuracy, which implies that all components perform positive effects. The ATI-net's structure is reasonable. 2) The results of Without linear module is the worst. It is better to deal with linear and nonlinear features separately. 3) The ATI-net is better than the ReLU function based method, and the reason is that the Mish function addresses the problem of the ReLU function stopping learning when the input is negative. Thus, it is reasonable to select the Mish function for building the ATI-net. A good predictive model requires a suitable activation function. 4) Multi-scale feature extraction and residual connection attenuate the information loss in the transmission process of deep network.

Remark E: The superior prediction accuracy of the ATI-net demonstrates that all components have positive effects in predicting GHI, and the corresponding network structure is reasonable. Furthermore, The activation function passes information layer by layer and deserves deep consideration in the network design.

Table 13: Ablation study results for GHI prediction in New York

Methods	RMSE	MAE	RRMSE	TIC	IA
ReLU function based method	46.9538	26.7077	18.4346	0.0588	0.9941
Without linear module	47.9163	28.0076	18.8125	0.0596	0.9939
Without multi-scale feature extraction	46.9672	26.1360	18.4398	0.0586	0.9941
Without residual connection	46.8521	25.8953	18.3947	0.0587	0.9941
The proposed method	**46.8092**	**25.7922**	**18.3778**	**0.0586**	**0.9941**

Table 14: Ablation study results for GHI prediction in Hawaii

Methods	RMSE	MAE	RRMSE	TIC	IA
ReLU function based method	61.6292	28.2121	21.8614	0.0676	0.9924
Without linear module	67.9419	31.0984	24.1007	0.0741	0.9908
Without multi-scale feature extraction	64.2370	29.3178	22.7865	0.0705	0.9918
Without residual connection	61.9361	27.7095	21.9703	0.0679	0.9924
The proposed method	**59.8059**	**27.0089**	**21.2147**	**0.0657**	**0.9929**

4.8. Experiment F: Multi-step-ahead prediction

This experiment compares the prediction accuracy between the proposed method and the other four benchmarks (GRU, CNN-LSTM, GRU-CNN, and TCN) under the three-step-ahead prediction. The three-step-ahead prediction experiment is performed based on the multiple-input multiple-output (MIMO) strategy [56], and the above prediction models are all re-trained. The formula of MIMO is:

$$\left(\hat{g}_{i+n+1}, \hat{g}_{i+n+2}, \hat{g}_{i+n+3}\right) = \boldsymbol{ATI\text{-}net}\left(\boldsymbol{s}_i, \boldsymbol{s}_{i+1}, ..., \boldsymbol{s}_{i+n}\right) \tag{21}$$

where $\hat{g}_{i+n+1}$, $\hat{g}_{i+n+2}$, and $\hat{g}_{i+n+3}$ are the 1-step, 2-step, and 3-step predictions; $\boldsymbol{ATI\text{-}net}(*)$ is the proposed ATI-net.

Tables 15 and 16 report the three-step-ahead prediction results in two datasets, and the details are: 1) In the New York datasets, the proposed method generates the best performance in 1-step, 2-step, and 3-step predictions. In the Hawaii datasets, the proposed method ranks first, second, and third in 1-step, 2-step, and 3-step predictions, respectively. 2) The GRU also obtains relatively good predictions in two datasets, which successfully mines long-term change rules. 3) As the number of prediction steps increases, the prediction accuracy of all methods decreases significantly.

Remark F: The multi-step-ahead prediction results show that the proposed method is more suitable for one-step prediction, and the main reason is that the proposed method only introduces a few bionic components in the audiovisual integrated system for the long-term memory. It is impossible that the

proposed method outperforms those benchmarks in all cases. In summary, the application of the proposed method is for one-step and single-position GHI prediction.

Table 15: Comparison with GRU, CNN-LSTM, GRU-CNN, and TCN for three-step-ahead prediction in New York

	Methods	RMSE	MAE	RRMSE	TIC	IA
1-step	GRU	66.5502	37.6152	26.3619	0.0827	0.9881
	CNN-LSTM	88.9799	53.1256	35.2468	0.1131	0.9773
	GRU-CNN	90.1821	49.6952	35.7230	0.1096	0.9790
	TCN	105.2776	68.9365	41.7026	0.1320	0.9684
	The proposed method	**47.0559**	**26.9082**	**18.6398**	**0.0590**	**0.9940**
2-step	GRU	92.6335	49.9861	36.5274	0.1145	0.9771
	CNN-LSTM	96.9353	56.0114	38.2237	0.1239	0.9728
	GRU-CNN	104.8084	59.8520	41.3282	0.1274	0.9712
	TCN	122.4090	80.3420	48.2685	0.1536	0.9564
	The proposed method	**71.3761**	**47.0422**	**28.1451**	**0.0890**	**0.9859**
3-step	GRU	111.6544	59.9459	43.8367	0.1381	0.9663
	CNN-LSTM	100.9866	59.5492	39.6484	0.1273	0.9709
	GRU-CNN	116.1358	66.1426	45.5962	0.1409	0.9643
	TCN	135.4425	89.6710	53.1762	0.1690	0.9462
	The proposed method	**89.2146**	**61.3704**	**35.0266**	**0.1117**	**0.9774**

Table 16: Comparison with GRU, CNN-LSTM, GRU-CNN, and TCN for three-step-ahead prediction in Hawaii

	Methods	RMSE	MAE	RRMSE	TIC	IA
1-step	GRU	63.3877	29.4869	22.6741	0.0700	0.9919
	CNN-LSTM	89.9442	38.3295	32.1735	0.0977	0.9841
	GRU-CNN	94.6250	49.4016	33.8478	0.1032	0.9821
	TCN	103.1349	55.3091	36.8919	0.1121	0.9787
	The proposed method	**59.7352**	**25.3077**	**21.3676**	**0.0655**	**0.9930**
2-step	GRU	**88.9107**	41.5252	**31.6737**	**0.0977**	**0.9841**
	CNN-LSTM	96.5398	42.9285	34.3915	0.1051	0.9815
	GRU-CNN	109.4755	57.2401	38.9997	0.1195	0.9757
	TCN	114.9494	61.5096	40.9497	0.1241	0.9737
	The proposed method	90.3447	**40.4698**	32.1845	0.0981	0.9839
3-step	GRU	**103.1875**	50.8779	**36.6032**	0.1130	0.9785
	CNN-LSTM	104.3009	**46.0488**	36.9981	**0.1129**	**0.9786**
	GRU-CNN	116.8125	61.0114	41.4363	0.1271	0.9724
	TCN	123.4496	66.9438	43.7907	0.1321	0.9700
	The proposed method	108.3859	50.4585	38.4472	0.1163	0.9771

5. Discussion

This section discusses the method improvement, method stability, different integration strategies, different dual-stream frameworks based on Transformer, hyperparameter selection, and prediction under more training-validation and test samples, respectively.

5.1. Method improvement discussion

This paper uses the P_{RMSE} to show the improvement percentage of the ATI-net compared to recursive, parallel, and hybrid benchmarks. The formula of P_{RMSE} is:

$$P_{RMSE} = \left| \frac{RMSE_{compared} - RMSE_{proposed}}{RMSE_{compared}} \right| \times 100\% \tag{22}$$

where $RMSE$ is the evaluation metric of RMSE.

Fig. 14 presents the improvement indicator of P_{RMSE}, and we can find that: 1) Overall average improvements of P_{RMSE} are 26.28% and 26.01% for two datasets, respectively. The prediction accuracy is significantly promoted by the ATI-net, which is one of the reliable alternatives for SI prediction. 2) Average improvement percentages of P_{RMSE} are 31.64%, 25.87%, and 20.37% for comparing with recursive, parallel, and hybrid benchmarks under New York datasets. Average improvement percentages of P_{RMSE} are 22.38%, 29.26%, and 27.55% for comparing with recursive, parallel, and hybrid benchmarks under Hawaii datasets. Compared with the above three kinds of ML, the bionic ATI-net has the superiority in prediction accuracy.

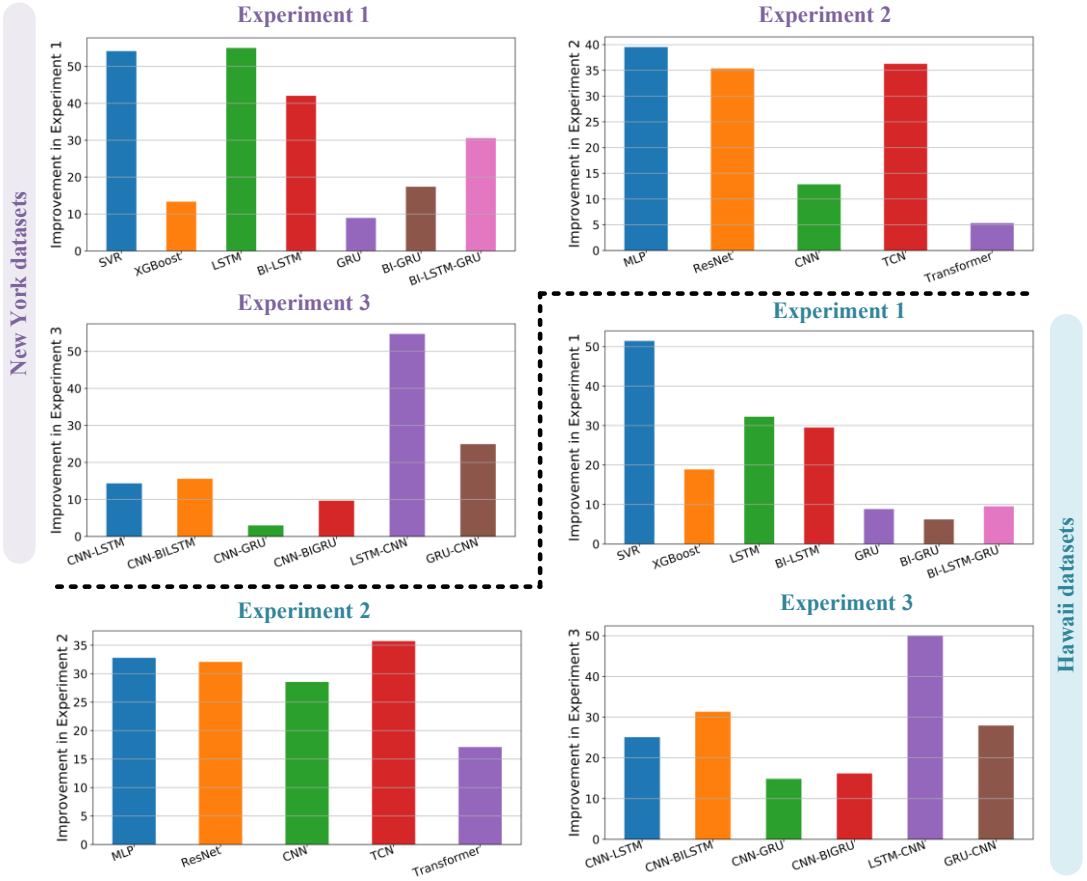


Fig. 14: Improvement percentage of P_{RMSE} (%) under two datasets.

5.2. Method stability discussion

A good prediction model requires not only high prediction accuracy but also stable prediction results. This paper uses the STD values of prediction errors to show the model prediction stability, where smaller STD values imply better model predictive stability. Table 17 lists the STD values under two datasets, and we can see that: 1) The ATI-net has the smallest STD values with 46.6174 and 59.4223 for two datasets, meaning the best model prediction stability. 2) The CNN-based hybrid methods, such as CNN-LSTM, CNN-GRU, and CNN-BILSTM, have better stability than single recursive methods, such as LSTM, GRU, and BILSTM, implying that the feature extraction by CNN promotes the predictive stability.

Table 17: Stability analysis of STD under two datasets

Methods	STD in New York datasets	STD in Hawaii datasets
SVR	98.3538	119.9735
XGBoost	53.9237	73.6800
LSTM	103.9716	87.9725
BI-LSTM	80.4924	84.7650
GRU	50.2409	65.1904
BI-GRU	54.7347	63.7633
BI-LSTM-GRU	66.0798	65.9839
MLP	74.6031	87.1410
ResNet	69.9929	86.0824
CNN	51.7249	81.6104
TCN	69.2423	90.6412
Transformer	47.8461	70.9508
CNN-LSTM	52.4935	78.0477
CNN-BILSTM	52.4475	85.4794
CNN-GRU	47.1122	69.1615
CNN-BIGRU	50.1803	70.7458
LSTM-CNN	98.3645	114.6329
GRU-CNN	57.7073	80.2921
The proposed method	**46.6174**	**59.4223**

5.3. Integration strategy discussion

This part discusses the results of four integration strategies, including Attention, SENet, Self-attention, and Concat integrations. The Concat integration preserves features based on equal weights and produces no information loss. Other integration strategies all assign weights to features through their preferences, thus enhancing what they consider to be core features and ignoring what they consider to be ineffective features. Table 18 reports the prediction results, and we can find that: 1) The Concat integration used in the ATI-net has the best results and is suitable for predicting SI series. 2) Other integration strategies have varying degrees of degradation, where some useful features are not weighted higher.

Table 18: Integration strategy analysis

Datasets	Methods	RMSE	MAE	RRMSE	TIC	IA
New York	Attention integration in [46]	47.5588	26.8722	18.6721	0.0596	0.9939
	SENet integration in [47]	50.7829	29.5962	19.9379	0.0633	0.9931
	Self-attention integration	49.4198	27.5991	19.4028	0.0613	0.9936
	The proposed method	**46.8092**	**25.7922**	**18.3778**	**0.0586**	**0.9941**
Hawaii	Attention integration in [46]	65.0241	28.7722	23.0657	0.0712	0.9916
	SENet integration in [47]	68.4020	32.3855	24.2639	0.0750	0.9906
	Self-attention integration	69.6858	31.1349	24.7193	0.0758	0.9905
	The proposed method	**59.8059**	**27.0089**	**21.2147**	**0.0657**	**0.9929**

5.4. Transformer framework discussion

Dual-stream frameworks are promising in handling nonlinear time series. This part compares the ATI-net (Transformer-CNN) and other four dual-stream methods based on Transformer, including the Transformer-LSTM, Transformer-GRU, Transformer-TCN, and Transformer-Transformer. Table 19 gives specific prediction results of the above dual-stream methods, and we can see that: 1) Except for MAE in New York datasets, the ATI-net (Transformer-CNN) has the best results, implying that CNN outperforms LSTM, GRU, TCN, and Transformer in extracting linear relationships from SI series. 2) Transformer-TCN ranks second in the New York datasets, and Transformer-Transformer ranks second in the Hawaii datasets. The above methods may be interesting to be further explored in predicting SI series.

Table 19: Transformer-based framework analysis

Datasets	Methods	RMSE	MAE	RRMSE	TIC	IA
New York	Transformer-LSTM	47.5425	27.4172	18.6657	0.0591	0.9940
	Transformer-GRU	48.0130	26.1034	18.8504	0.0599	0.9939
	Transformer-TCN	47.2086	**25.7018**	18.5346	0.0591	0.9940
	Transformer-Transformer	47.5982	26.8672	18.6876	0.0594	0.9939
	The proposed method	**46.8092**	25.7922	**18.3778**	**0.0586**	**0.9941**
Hawaii	Transformer-LSTM	69.0708	30.4519	24.5012	0.0756	0.9905
	Transformer-GRU	60.8920	29.5267	21.5999	0.0668	0.9926

	Transformer-TCN	61.2634	28.7911	21.7317	0.0671	0.9925
	Transformer-Transformer	60.0951	27.6932	21.3172	0.0657	0.9929
	The proposed method	**59.8059**	**27.0089**	**21.2147**	**0.0657**	**0.9929**

5.5. Selection of hyperparameters

Hyperparameters affect the prediction of deep ML methods. As performed in [53, 54], the grid search technique is applied to select the locally optimal hyperparameters. In this paper, two optimizers (Adagrad and Adam) and three learning rates (0.3, 0.00001, and from 0.3 to 0.00001) are compared. Table 20 lists the prediction results under validation samples, and these models are all trained based on training samples. It is found that: 1) The Adam optimizer and the learning rate changed from 0.3 to 0.00001 provides the best results, which implies that the training hyperparameters selected in this paper are reasonable. 2) The optimization of more hyperparameters may generate a better prediction, but it may significantly increase the cost.

Table 20: Hyperparameter selection based on the grid search technique

Datasets	Optimizer	Learning rate	RMSE	MAE	RRMSE	TIC	IA
New York	Adagrad	0.3	58.3757	33.8895	23.3399	0.0740	0.9907
		0.00001	58.3755	33.8900	23.3398	0.0740	0.9907
		From 0.3 to 0.00001	58.3739	33.8887	23.3392	0.0740	0.9907
	Adam	0.3	53.4642	31.9380	21.3762	0.0678	0.9922
		0.00001	51.5449	28.0873	20.6088	0.0651	0.9929
		From 0.3 to 0.00001	**51.5087**	**28.0559**	**20.5944**	**0.0650**	**0.9929**
Hawaii	Adagrad	0.3	48.1234	25.5043	14.8822	0.0481	0.9959
		0.00001	48.1238	25.5045	14.8824	0.0481	0.9959
		From 0.3 to 0.00001	48.1241	25.5039	14.8824	0.0481	0.9959
	Adam	0.3	39.3677	17.0044	12.1745	0.0393	0.9973
		0.00001	37.9009	16.9281	11.7209	0.0379	0.9975
		From 0.3 to 0.00001	**37.8699**	**16.8157**	**11.7113**	**0.0379**	**0.9975**

5.6. An extended experiment with more training-validation and test samples

Since datasets used in the above experiments are not long, this experiment assesses the prediction

performance under more training-validation and test samples. The GHI and related factors are collected from New York, where the sampling time is 12 hours. This dataset starts from 2015/1/1 at 12:00, and includes 4379 samples, where the training-validation and test samples are 3503 (80%) and 876 (20%), respectively. The period of training-validation samples is more than four years, and the period of test samples is more than one year. The input is the multivariate time series of GHI and other factors, and the input window size is set to 54 based on PACF values in Fig. 15. Table 21 shows specific predictions between the proposed method, Persistence, and the other seven deep ML methods. Moreover, Fig. 15 presents detailed predictions of Persistence, LSTM, GRU, Transformer, and the proposed method. Details are: 1) The proposed method still shows the best prediction accuracy compared to Persistence and the other seven deep ML methods. The proposed method inherits some effective analytical capabilities from the audiovisual integration system and leverages powerful self-learning to present GHI fluctuations. 2) Transformer and GRU present relatively good prediction performance, where Transformer grasps core information based on self-attentions, and GRU remembers core information via gated mechanisms. 3) The deep ML methods are better than the Persistence, implying that these deep ML methods capture useful rules from the historical GHI. In summary, the proposed method is suitable for predicting GHI.

Table 21: Prediction results under more training-validation and test samples

Methods	RMSE	MAE	RRMSE	TIC	IA
Persistence	554.8433	472.0594	235.3672	0.7071	0.0936
LSTM	178.2376	113.1698	75.6093	0.2389	0.8975
GRU	162.3091	103.2444	68.8523	0.2151	0.9178
BI-LSTM-GRU	190.9239	121.0628	80.9909	0.2454	0.8901
TCN	167.6626	105.1341	71.1233	0.2200	0.9148
Transformer	162.0370	104.2021	68.7369	0.2164	0.9158
CNN-LSTM	167.7974	104.9955	71.1805	0.2256	0.9092
CNN-GRU	166.6353	104.2209	70.6875	0.2213	0.9123

The proposed method	**159.8261**	**102.6623**	**67.7990**	**0.2133**	**0.9185**

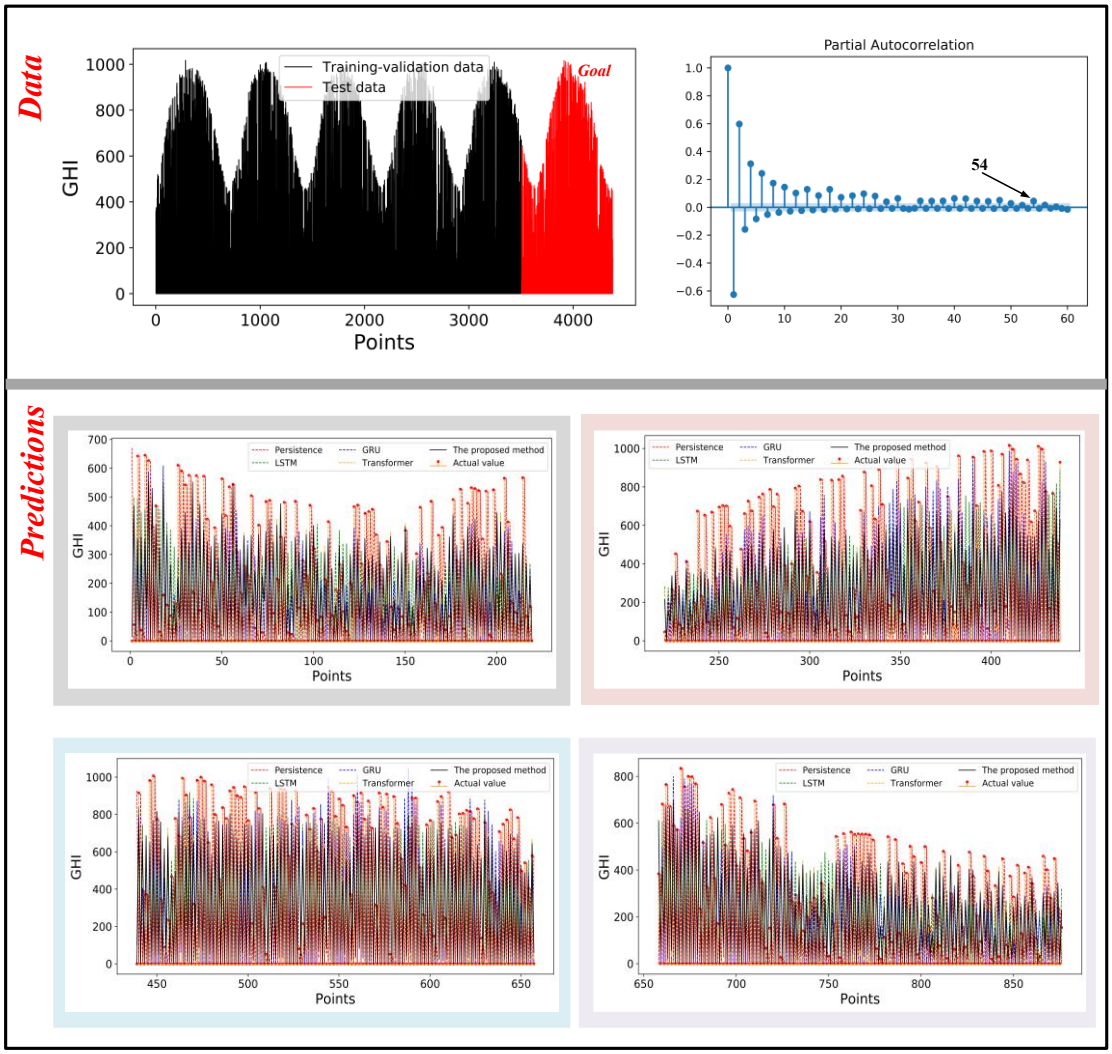


Fig. 15: Data analysis, and predictions of Persistence, LSTM, GRU, Transformer and the proposed method funder more training-validation and test samples.

6. Conclusion

Accurate prediction of SI is crucial to promote the utilization of solar energy and stabilize the operation of solar power generation. In this paper, starting from the bio-cognitive process, an audiovisual-motivated Transformer-CNN integration network, namely ATI-net, is designed for predicting SI. Specifically, the ATI-net has three blocks with signal capture, signal analysis, and prediction. The first block extracts multi-scale features via introducing several one-dimensional

convolutions with various kernels and Mish functions, improving the feature learning capability. The second block adopts two structures triggered by Transformers and one-dimensional convolutions to not only mimic right and left hemispheres but also mine sequential dependencies. The third block integrates information of nonlinear dependencies and linear relationships to make the SI prediction. Two cases from New York and Hawaii are used to illustrate that the ATI-net has better prediction accuracy than 18 benchmarks from shallow, recursive, parallel, and hybrid ML. Based on experimental analysis and discussions, the proposed method is more suitable for one-step and single-position GHI prediction. The conclusions of this study are summarized as follows:

1) Average RMSE improvements are 26.28% and 26.01% for two datasets, respectively. The explored ATI-net not only has a bionically-inspired design framework but also combines the advantages of Transformer and CNN, significantly improving the prediction accuracy. This paper suggests that prediction networks designed by bio-cognitive processes are expected to further enhance nonlinear time series analysis.

2) Compared with 18 benchmarks, the ATI-net has the best prediction stability with 46.6174 and 59.4223 STD values for two datasets. The reasons are the multi-scale feature extraction, residual connection, and independent treatment of linear and nonlinear relationships. This paper suggests that reducing the loss of useful information from feature engineering and network connection is useful for promoting the prediction accuracy and stability.

3) The proposed ATI-net has a dual-stream structure and is better than other Transformer-based dual-stream networks, such as Transformer-LSTM, Transformer-TCN, and Transformer-Transformer, which prove its superiority and effectiveness. This paper suggests that the dual-stream framework based on Transformer may be promising in handling SI series.

In the future, some key areas are worth studying: 1) More evaluation experiments under different seasons and weather conditions, such as cloudy, rainy, and sunny, may be conducted to demonstrate the generalization of the ATI-net. 2) The lightweight prediction framework may be developed, further reducing hardware requirements and improving industrial application scenarios. 3) The ATI-net may need more bionic components to learn the long-term change rules of GHI and to make multi-step-ahead predictions.

Declaration of conflicting interests

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.

Acknowledgment

The research is supported in part by Youth Program of National Natural Science Foundation of China through the grant 72103186, China Scholarship Council through the grant 202306290148, Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University through the grant CX2023063, and Outstanding Doctoral Dissertation Cultivation Fund of the School of Automation, Northwestern Polytechnical University.

References

[1] K. Veerabhadrappa, B. G. Suhas, C. K. Mangrulkar, R. S. Kumar, V. S. Mudakappanavar, Narahari, and K. N. Seetharamu, “Power generation using ocean waves: A review,” *Global Transitions Proceedings*, vol. 3, no. 2, pp. 359-370, 2022.

[2] J. Hu, Q. Tang, Z. Wu, B. Zhang, C. He, and Q. Chen, “Optimization and assessment method for total energy system retrofit in the petrochemical industry considering clean energy

substitution for fossil fuel," *Energy Conversion and Management*, vol. 284, article 116967, 2023.

[3] Y. Liang, Y. Ju, J. Qin, and W. Pedrycz, "Multi-granular linguistic distribution evidential reasoning method for renewable energy project risk assessment," *Information Fusion*, vol. 65, pp. 147-164, 2021.

[4] Z. A. Khan, T. Hussain, and S. W. Baik, "Dual stream network with attention mechanism for photovoltaic power forecasting," *Applied Energy*, vol. 338, article 120916, 2023.

[5] A. S. B. M. Shah, H. Yokoyama, and N. Kakimoto "High-precision forecasting model of solar irradiance based on grid point value data analysis for an efficient photovoltaic system," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 474-481, 2015.

[6] REN21, Renewable 2020 global status report. https://www.ren21.net/wp-content/uploads/2019/05/GSR2021_Full_Report.pdf.

[7] Y. Yu, T. Niu, J. Wang, and H. Jiang, "Intermittent solar power hybrid forecasting systems based on pattern recognition and feature extraction," *Energy Conversion and Management*, vol. 277, article 116579, 2023.

[8] L. Massidda, F. Bettio, and M. Marrocu, "Probabilistic day-ahead prediction of PV generation. A comparative analysis of forecasting methodologies and of the factors influencing accuracy," *Solar Energy*, vol. 271, article 112422, 2024.

[9] J. Liu, X. Huang, Q. Li, Z. Chen, G. Liu, and Y. Tai, "Hourly stepwise forecasting for solar irradiance using integrated hybrid models CNN-LSTM-MLP combined with error correction and VMD," *Energy Conversion and Management*, vol. 280, article 116804, 2023.

[10] A. Murata, H. Ohtake, and T. Oozeki, "Modeling of uncertainty of solar irradiance forecasts on numerical weather predictions with the estimation of multiple confidence intervals," *Renewable Energy*, vol. 117, pp. 193-201, 2018.

[11] J. Tong, L. Xie, S. Fang, W. Yang, and K. Zhang, "Hourly solar irradiance forecasting based on encoder-decoder model using series decomposition and dynamic error compensation," *Energy Conversion and Management*, vol. 270, article 116049, 2022.

[12] B. Singh and D. Pozo, "A guide to solar power forecasting using ARMA models," *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pp. 1-4, 2019.

[13] S. Atique, S. Noureen, V. Roy, V. Subburaj, S. Bayne, and J. Macfie, "Forecasting of total daily solar energy generation using ARIMA: A case study," *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0114-0119, 2019.

[14] Z. Jiang, J. Che, and L. Wang, "Ultra-short-term wind speed forecasting based on EMD-VAR model and spatial correlation," *Energy Conversion and Management*, vol. 250, article 114919, 2021.

[15] M. David, F. Ramahatana, P. J. Trombe, and P. Lauret, "Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models," *Solar Energy*, vol. 133, pp. 55-72, 2016.

[16] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65-76, 2013.

[17] J. Tong, L. Xie, S. Fang, W. Yang, and K. Zhang, “Hourly solar irradiance forecasting based on encoder-decoder model using series decomposition and dynamic error compensation,” *Energy Conversion and Management*, vol. 270, article 116049, 2022.

[18] X. Du, J. Xue, L. Yu, W. Lei, H. Ma, C. Cao, C. Shu, and Y. Li, “Coal damage and energy characteristics during shallow mining to deep mining,” *Energy*, vol. 291, article 130375, 2024.

[19] A. Zendehboudi, M. A. Baseer, and R. Saidur, “Application of support vector machine models for forecasting solar and wind energy resources: A review,” *Journal of Cleaner Production*, vol. 199, pp. 272-285, 2018.

[20] R. K. Sahu, B. Shaw, J. R. Nayak, and Shashikant, “Short/medium term solar power forecasting of Chhattisgarh state of India using modified TLBO optimized ELM,” *Engineering Science and Technology, an International Journal*, vol. 24, no. 5, pp. 1180-1200, 2021.

[21] X. Li, L. Ma, P. Chen, H. Xu, Q. Xing, J. Yan, S. Lu, H. Fan, L. Yang, and Y. Cheng, “Probabilistic solar irradiance forecasting based on XGBoost,” *Energy Reports*, vol. 8, no. 5, pp. 1087-1095, 2022.

[22] H. Wu, X. Gao, and J. Heng, “Bio-multisensory-inspired gate-attention coordination model for forecasting short-term significant wave height,” *Energy*, vol. 294, article 130887, 2024.

[23] H. Su, E. Zio, J. Zhang, M. Xu, X. Li, and Z. Zhang, “A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model,” *Energy*, vol. 178, pp. 585-597, 2019.

[24] X. Qing, and Y. Niu, “Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM,” *Energy*, vol. 148, article 117715, 2018.

[25] S. Tajjour, S. S. Chandel, M. A. Alotaibi, H. Malik, F. P. G. Marquez, and A. Afthanorhan, "Short-term solar irradiance forecasting using deep learning techniques: A comprehensive case study," *IEEE Access*, vol. 11, pp. 119851-119861, 2023.

[26] M. Ehteram, M. A. Nia, F. Panahi, and A. Farrokhi, "Read-first LSTM model: A new variant of long short-term memory neural network for predicting solar radiation data," *Energy Conversion and Management*, vol. 305, article 118267, 2024.

[27] A. Mellit, A. M. Pavan, and V. Lughi, "Deep learning neural networks for short-term photovoltaic power forecasting," *Renewable Energy*, vol. 172, pp. 276-288, 2021.

[28] M. Chhetri, S. Kumar, P. P. Roy, and B. G. Kim, "Deep BLSTM-GRU model for monthly rainfall prediction: A case study of Simtokha, Bhutan," *Remote Sensing*, vol. 12, no. 19, 2021.

[29] Y. Kashyap, A. Bansal, and A. K. Sao, "Solar radiation forecasting with multiple parameters neural networks," *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 825-835, 2015.

[30] G. Terren-Serrano, and M. Martinez-Ramon, "Deep learning for intra-hour solar forecasting with fusion of features extracted from infrared sky images," *Information Fusion*, vol. 95, pp. 42-61, 2023.

[31] C. Yildiz, H. Acikgoz, D. Korkmaz, and U. Budak, "An improved residual-based convolutional neural network for very short-term wind power forecasting," *Energy Conversion and Management*, vol. 228, article 113731, 2021.

[32] Y. Li, L. Song, S. Zhang, L. Kraus, T. Adcox, R. Willardson, A. Komandur, N. Lu, "A TCN-based hybrid forecasting framework for hours-ahead utility-scale PV forecasting," *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 4073-4085, 2023.

[33] J. Liu, H. Zang, L. Cheng, T. Ding, Z. Wei, and G. Sun, "A Transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting," *Applied Energy*, vol. 342, article 121160, 2023.

[34] S. M. J. Jalali, S. Ahmadian, A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "Automated deep CNN-LSTM architecture design for solar irradiance forecasting," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 52, no. 1, pp. 54-65, 2022.

[35] Y. Chen, Z. Wei, H. Gou, H. Liu, L. Gao, X. He, and X. Zhang, "How far is brain-inspired artificial intelligence away from brain? " *Frontiers in Neuroscience*, vol. 16, article 1096737, 2022.

[36] H. Wu, Y. Liang, and X. Gao, "Left-right brain interaction inspired bionic deep network for forecasting significant wave height," *Energy*, vol. 278, article 127995, 2023.

[37] H. Wu, Y. Liang, X. Gao, and P. Du, "Auditory-circuit-motivated deep network with application to short-term electricity price forecasting," *Energy*, vol. 288, article 129729, 2024.

[38] H. Wu, Y. Liang, J. Heng, C. Ma, and X. Gao, "MSV-net: Multi-scale visual-inspired network for short-term electricity price forecasting," *Energy*, vol. 291, article 130350, 2024.

[39] H. Wu, Y. Liang, and J. Heng, "Pulse-diagnosis-inspired multi-feature extraction deep network for short-ter, electricity load forecasting," *Applied Energy*, vol. 339, article 120995, 2023.

[40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. U. Kaiser, and I. Polosukhin, "Attention is all you need," *Advanced in Neural Information Processing Systems*, vol. 30, 2017.

[41] L. Jiang, T. Zhang, W. Lei, K. Zhuang, and Y. Li, "A new convolutional dual-channel Transformer network with time window concatenation for remaining useful life prediction of rolling bearings," *Advanced Engineering Informatics*, vol. 56, article 101966, 2023.

[42] S. Sun, Y. Liu, Q. Li, T. Wang, and F. Chu, "Short-term multi-step wind power forecasting based on spatio-temporal correlations and transformer neural networks," *Energy Conversion and Management*, vol. 283, article 116916, 2023.

[43] H. Park, and C. Kayser, "The context of experienced sensory discrepancies shapes multisensory integration and recalibration differently," *Cognition*, vol. 225, article 105092, 2022.

[44] K. Zhao, Z. Jia, F. Jia, and H. Shao, "Multi-scale integrated deep self-attention network for predicting remaining useful life of aero-engine", *Engineering Applications of Artificial Intelligence*, vol. 120, article 105860, 2023.

[45] D. Misra, "Mish: A self regularized non-monotonic activation function", *arXiv: 1908.08681v3*, 2020.

[46] H. Wu, Y. Liang, and J. Zuo, "Human-inspired spatiotemporal feature extraction and fusion network for weather forecasting," *Expert Systems with Applications*, vol. 207, article 118089, 2022.

[47] H. Wu, Y. Liang, X. Gao, and J. Heng, "Bionic-inspired oil price prediction: Auditory multi-feature collaboration network," *Expert Systems with Applications*, vol. 244, article 122971, 2024.

[48] S. Pan, B. Yang, S. Wang, Z. Guo, L. Wang, J. Liu, and S. Wu, "Oil well production prediction based on CNN-LSTM model with self-attention mechanism," *Energy*, vol. 284, article 128701, 2023.

[49] D. Xiao, C. Qin, J. Ge, P. Xia, Y. Huang, and C. Liu, "Self-attention-based adaptive remaining useful life prediction for IGBT with Monte Carlo dropout," *Knowledge-Based Systems*, vol. 239, article 107902, 2022.

[50] S. Ghimire, R. C. Deo, N. Raj, and J. Mi, "Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms," *Applied Energy*, vol. 253, article 113541, 2019.

[51] P. Singla, M. Duhan, and S. Saroha, "A dual decomposition with error correction strategy based improved hybrid deep learning model to forecast solar irradiance," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 44, no. 1, pp. 1583-1607, 2022.

[52] S. Saroha, and S. K. Aggarwal, "Wind power forecasting using wavelet transforms and neural networks with tapped delay," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 2, pp. 197-209, 2018.

[53] P. Singla, M. Duhan, and S. Saroha, "A point and interval forecasting of solar irradiance using different decomposition based hybrid models," *Earth Science Informatics*, vol. 16, no. 3, pp. 2223-2240, 2023.

[54] P. Singla, M. Duhan, and S. Saroha, "An integrated framework of robust local mean decomposition and bidirectional long short-term memory to forecast solar irradiance," *International Journal of Green Energy*, vol. 20, no. 10, pp. 1073-1085, 2023.

[55] D. Diaz-Bedoya, M. Gonzalez-Rodriguez, J. Clairand, X. Serrano-Guerrero, and G. Escriva-Escriva, "Forecasting univariate solar irradiance using machine learning models: A case study of two Andean Cities," *Energy Conversion and Management*, vol. 296, article 117618, 2023.

[56] C. Deng, Y. Huang, N. Hasan, and Y. Bao, "Multi-step-ahead stock price index forecasting using long short-term memory model with multivariate empirical mode decomposition," *Information Sciences*, vol. 607, pp. 297-321, 2022.