Energy 275 (2023) 127430

Contents lists available at ScienceDirect

Energy

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

A novel approach based on integration of convolutional neural networks and echo state network for daily electricity demand prediction

Sujan Ghimire [a], Thong Nguyen-Huy [b], Mohanad S. AL-Musaylh [c], Ravinesh C. Deo [a], David Casillas-Pérez [d,*], Sancho Salcedo-Sanz [a,e]

[a] School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD, 4300, Australia
[b] Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, 4350, QLD, Australia
[c] Department of Information Technologies, Management Technical College, Southern Technical University, Basrah 61001, Iraq
[d] Department of Signal Processing and Communications, Universidad Rey Juan Carlos, Fuenlabrada, 28942, Madrid, Spain
[e] Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, 28805, Madrid, Spain

ARTICLE INFO

Dataset link: https://www.energex.com.au

Keywords:
Electricity demand forecasting
Sustainable energy
Artificial intelligence
Deep learning
Echo state networks
Convolutional neural networks
Hybrid algorithms

ABSTRACT

Predicting electricity demand data is considered an essential task in decisions taking, and establishing new infrastructure in the power generation network. To deliver a high-quality electricity demand prediction, this paper proposes a hybrid combination technique, based on a deep learning model of Convolutional Neural Networks and Echo State Networks, named as CESN. Daily electricity demand data from four sites (Roderick, Rocklea, Hemmant and Carpendale), located in Southeast Queensland, Australia, have been used to develop the proposed hybrid prediction model. The study also analyzes five other machine learning-based models (support vector regression, multilayer perceptron, extreme gradient boosting, deep neural network, and Light Gradient Boosting) to compare and evaluate the outcomes of the proposed deep learning approach. The results obtained in the experimental study showed that the proposed hybrid deep learning model is able to obtain the highest performance compared to other existing models developed for daily electricity demand data forecasting. Based on the statistical approaches utilized in this study, the proposed hybrid approach presents the highest prediction accuracy among the compared models. The obtained results showed that the proposed hybrid deep learning algorithm is an excellent and accurate electricity demand forecasting method, which outperformed the state of the art algorithms that are currently used in this problem.

1. Introduction

Supporting the growth of energy security policies and plans is considered an essential task that can be achieved by relatively reducing the predicted errors of the electricity demand (G) data [1]. In fact, errors in the estimation of G data may lead to affect energy policies and systems [2,3]. For example, a report published in 2017 by Finkel [4] identified erroneous demand forecasts as one of the primary causes of several Australian electrical grid issues.

Electricity demand forecasting errors may have different consequences depending on the prediction time-horizon where the problem is defined: In long-term scales, demand overestimation may result in wasteful and ineffective over-investment in assets, raising costs that are then passed on to customers, raising this way the cost of energy. On the other hand, underestimating the electricity demand can result in insufficient investment, which might lead to problems with energy security and reliability. In short-term scales, prediction errors are associated with issues in the electricity supply and transportation, as well as punctual shortages due to underestimation of the demand. Note that the use of renewable energy has made this problem even harder, due to the intrinsic variability of renewable sources [5,6].

To address this imbalance in power supply and demand under such conditions, it is vital to properly predict G. To this end, a particularly accurate predictive model is required. In recent years, several prediction models have been utilized to address this complex forecasting problem of electricity demand time series in different countries such as China [7], Brazil [8], Iran [9] or countries of the West Africa [10], among others. The most commonly used approaches may be further classified into statistical models and Artificial Intelligence/Machine Learning (AI/ML) based models. For example, a variety of statistical models, such as regression with seasonality, Exponential Smoothing,

* Corresponding author.
E-mail addresses: sujan.ghimire@usq.edu.au (S. Ghimire), thong.nguyen-huy@usq.edu.au (T. Nguyen-Huy), mohanad.al-musaylh@stu.edu.iq (M.S. AL-Musaylh), ravinesh.deo@usq.edu.au (R.C. Deo), david.casillas@urjc.es (D. Casillas-Pérez), sancho.salcedo@uah.es (S. Salcedo-Sanz).

https://doi.org/10.1016/j.energy.2023.127430
Received 8 October 2022; Received in revised form 31 March 2023; Accepted 2 April 2023
Available online 8 April 2023

Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA), and Nonlinear AutoRegressive Moving Average model with exogenous inputs (NARMAX) [11–14], have been previously used to predict *G*. These statistical methods have demonstrated good empirical performances. The key benefit of statistical models over ML-based models is that these statistical models did not require intensive computing as required by ML-based techniques. However, the non-linear and non-normal distribution of *G* time series makes it challenging for statistical models to produce accurate predictions [15]. Therefore, in order to increase prediction accuracy, researchers have used a number of ML-based models that are very excellent at capturing the non-linear characteristics of *G* time series, despite the fact that they are computationally costly. In previous works, different ML models have been used to predict *G* time series. For example, multivariate adaptive regression spline (MARS), classical support vector regression (SVR), hybrid particle swarm optimized (PSO) merged with SVR, M5 trees model, artificial neural network (ANN), online sequential extreme learning machine (OS-ELM) or multiple linear regression [16]. A recent study has applied ML approaches to predict *G* time series in three-time steps of half-hourly, hourly and daily, showing that MARS is an excellent approach in half-hourly and hourly prediction time steps whereas for daily, the SVR model exceeded the comparison models [2]. Subsequently, *G* prediction accuracy was improved by integrating the improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) method, used to decompose the *G* time series before developing the prediction model. In this case, the PSO procedure was utilized to determine the parameters of the SVR model creating hybrid ICEEMDAN-PSO-SVR model [17]. In [18] a recent approach is proposed which also involves the decomposition of *G* times series into different sub-components, and the prediction of future *G* values using an Extreme Learning Machine. In addition, note that climate variables have shown to have a significant effect on predicting *G* data. Thus, a wide range of climate and Reanalysis datasets [19] have been employed within ML methods to find the relationships between predictive and *G* data. Finally, in [20] a new methodology for *G* prediction was introduced, by using maximal overlap discrete wavelet transform as a significant technique used before the OS-ELM model to decompose the inputs *G* data into a number of filters for more accurate *G* prediction.

Despite the benefits that ML models have shown in different prediction problems, including electricity demand as shown above, deep learning (DL) strategies are rapidly developed and have demonstrated a higher prediction performance compared to ML models in many cases [21]. The first and most important benefit of DL is their ability to create new features to address complex prediction tasks without human involvement in the training phase, unlike the ML methods, which require human participation [22]. Second, DL approaches consider multiple layers that make the model more effective in the training part and work better with unstructured datasets [23–25]. Finally, DL algorithms have a significant ability to deal with enormous datasets as they can execute a large number of computations in valuable time and cost [26,27]. In terms of *G* prediction, Deep Belief Networks (DBN) comprised of layers of Boltzmann machines, have been previously applied, and the results obtained showed that the DBN outperformed the traditional ANN model [28]. Long Short-term Memory Network (LSTM) [29] was also proposed for short-term electricity demand prediction and Deep neural networks [30] was used for demand management and load prediction, resulting in superior performance compared to ANNs. Furthermore, in [31] a Deep Recurrent Neural Network (DRNN) based on pooling was proposed to predict a short-term electricity demand problem. In this case, traditional RNNs were outperformed by the DRNN model. Additionally, numerous studies have demonstrated that hybrid models, which are created by integrating a few different prediction models and data processing methods, may effectively increase the predictive accuracy of ML and DL methods. For instance, [32] proposed an ensemble model for predicting electrical load based on enhanced kernel-ELM, extended Kalman filter, and empirical mode decomposition (EMD). A Bayesian neural network and wavelet transform ensemble electrical load prediction model was developed in [33], and empirical results support the model's strong predicting performance. In order to estimate electricity load, [34] introduced an SVR model integrated with EMD and autoregressive methods. That study demonstrated that the provided model can concurrently perform prediction with excellent accuracy. In order to predict *G*, [35] developed an ensemble technique based on EMD and DBN. The simulation results showed that the proposed model outperforms traditional ML models. Furthermore, in order to accurately estimate household energy consumption, [36] proposed a CNN-LSTM neural network that can extract spatial and temporal data. Compared to other traditional DL and ML models, the proposed CNN-LSTM method yielded the lowest root mean square error.

This study presents a hybrid model which integrates a convolutional neural network (CNN) and Echo State Network (ESN), for the accurate prediction of *G* time series, which we called CESN. This specific combination of techniques is used to create a novel CESN hybrid model, an advanced tool that shows excellent results in the prediction of electricity demand time series. The ESN which forms part of the proposed CESN approach is a type of Recurrent Neural Network (RNN), related to the idea of reservoir computing, where the traditional hidden layer is replaced by a dynamical reservoir [37]. Unlike traditional Neural Networks, the ESN is computationally efficient, since there is no back-propagation phase on the reservoir and it does not have the vanishing-gradient problem [38]. Furthermore, ESN can handle chaotic time series like *G* [39] and are not affected by bifurcations [40,41]. The CNN has also been a good approach for predicting chaotic and real-world time series [42], so its hybridization with the ESN tries to further exploit these properties of both networks. In this study, we have also employed an improved optimization approach based on the Bayesian optimization technique for optimizing the CNN hyperparameters as well as ESN reservoir parameters, to obtain high accuracy and better prediction capabilities of the final hybrid approach.

The rest of the paper has been structured in the following way: the next section is dedicated to the method's theoretical synopsis. Section 3 demonstrates the method, datasets, model development and model evaluation used in this paper. Section 4 presents the results obtained and a discussion with comparative analysis against other ML algorithms. Finally, Section 5 illustrates the paper's conclusion and possible future research directions. Appendix A shows a table of acronyms to ease the reading of the paper, Appendix B provides more implementation details about the CESN architecture and Appendix C shows further experimentation carried out to show the performance the proposed model.

2. Review of ML and DL techniques theoretical framework

2.1. Fundamental concepts of CNNs

CNNs are efficient tools for time series prediction. In CNN, convolution layers are the primary distinction between CNN and Neural Network [43]. These convolution layers have the ability to automatically extract characteristics from input data, collecting those that are crucial for mapping the connection between input and output variables. The CNN model can be run with multi-layer architectures, involving the pooling layer, convolutional layer, flattening layer, fully connected or dense layer and output layer. In particular, the convolution layer in the CNN structure learns the input data features through convolution operations, the pooling layer compiles the features obtained by the convolution operations and implements dimensionality reduction and secondary feature extraction operations, the flattening layer is used to transform the multi-dimensional feature maps received from the previous pooling layer into a one-dimension array to meet the data processing requirements of the next fully connected layer, and the

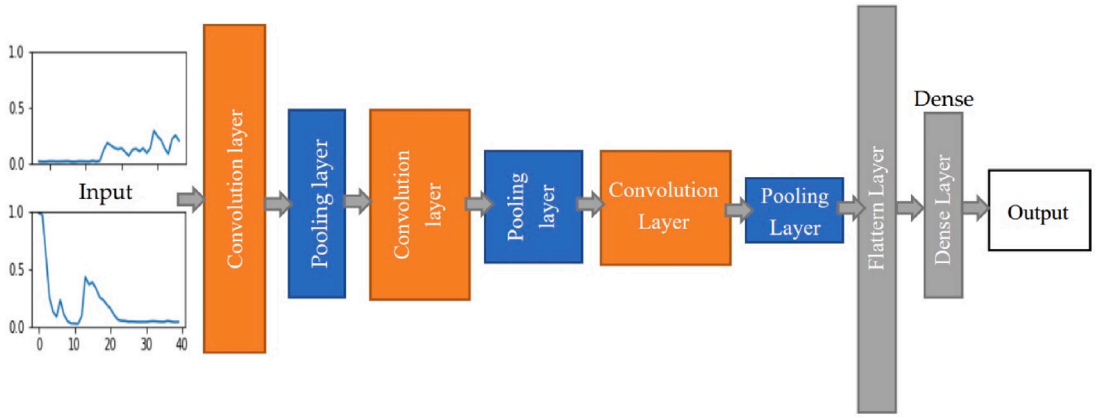

Fig. 1. A typical 1D CNN architecture.

Table 1
Descriptive statistics of daily electricity demand G at four substations of Southeast Queensland where the proposed deep learning hybrid CESN model is implemented.

Statistical parameters	Roderick	Rocklea	Hemmant	Carpendale
Median (MW)	531.83	380.08	273.74	29.98
Mean (MW)	534.84	380.13	278.27	30.28
Standard deviation (MW)	98.74	97.64	80.13	5.44
Variance	9748.67	9533.75	6420.99	29.58
Maximum (MW)	1003.88	869.09	703.87	60.75
Minimum (MW)	65.23	38.04	40.11	5.74
Range	938.65	831.05	663.76	55.01
Interquartile range	118.05	164.44	131.34	6.89
Skewness	0.30	0.12	0.06	0.56
Kurtosis	3.50	2.73	2.30	4.14

regression prediction is ultimately accomplished by the fully connected or dense layer, which collects the input data characteristics gained through the pooling and flattening operation [44]. CNN is widely used in image recognition in which 2D convolutional filters are used. However in time series prediction tasks CNN with 1D convolutional filters are used [45], a CNN model with three 1D convolutional layers and one fully connected layer is presented in Fig. 1.

This study has employed the activation function as "ReLU" to improve the computing efficiency of nonlinear data ([46,47]). The below equation describes the calculation process of the convolution layer.

$$f = \varphi\left(\omega^n \oplus x + b^n\right) \tag{1}$$

where ω, n, x and b, represent the weighting factors in the kernels, the number of kernels, the vector of the input series and the vector of the bias, respectively. The symbol $\oplus$ denotes the convolution operation and φ represents the rectified linear unit (ReLU) activation function, which with default values, returns the following element-wise;

$$\sigma(y) = \max(0, y) \tag{2}$$

where σ is a function of y over the set of zero and the input y. This function is linear for values greater than zero, however, it is a nonlinear function as negative values are always output as zero. The retrieved feature may have a sizeable number of dimensions after the convolution procedure, therefore to efficiently minimize the number of feature dimensions, a pooling layer is added after the convolution layer. As the performance of the maximum pooling operation in time series prediction tasks is typically superior to that of the average pooling operation, it is used in this paper instead of the average pooling layer. Eq. (3) also describes how to calculate the maximum polling operation.

$$Y_P = Pool_{\max}\left(Y_c\right) \tag{3}$$

where Y_P represents the output of the pooling layer, and $Pool_{\max}$ represents the maximum pooling function.

2.2. Fundamental concepts of ESN

ESN is a specific type of RNN that replaces RNN hidden layer with a reservoir. The ESN model, in contrast to RNN models, uses a fixed recurrent neural network as a reservoir and simply modifies the output weight matrix using a linear learning method, allowing for increment in computational efficiency [37]. Fig. 2 depicts the basic ESN structure, which comprises of a reservoir, an input layer, and an output layer. The numbers of input neurons, reservoir neurons, and output neurons are denoted by K, N, and L, respectively. Let $u = u(n)$ denote the external input, $x = x(n)$ the reservoir state, and $y = y(n)$ the output vector. W_{in}, W, W_{back}, and W_{out} denote the weight of the matrices of the input, reservoir, feedback, and output, respectively, and the size of the matrices is successively given by $N \times K$, $N \times N$, $N \times L$, and $L \times (K+N)$. The ESN model is trained by a supervised learning process, which involves two main steps, the reservoir states wast first updated, and then the reservoir-to-output layer's weight matrix W_{out} was learned. The dynamics of the ESN system can be defined as follows [48]:

$$x(t+1) = f\left(W_{in}u(t+1) + Wx(t) + W_{back}y(t)\right) \tag{4}$$

$$\mathbf{y}(t+1) = f^{out}\left(W_{out}x(t+1)\right) \tag{5}$$

where f denotes a sigmoid function inside the reservoir state and f^{out} denotes the output activation function.

During the ESN model training phase, the matrices W_{in} and W are initialized randomly. The inputs are projected into the high-dimensional state spaces in the reservoir, and the matrix can be learned by linear regression [49,50]. Studies have concluded that the reservoir performs the same role as the kernel in kernel-based learning algorithm [51,52] and is able to take into account the temporal information available in the inputs [53]. Thus, training an ESN is both simple and fast, and the ESN can avoid the vanishing gradient problem associated with NN and RNN, and reduce the computational complexity for modeling time series data. Additionally, [54] have identified the number of neurons in the reservoir (N), the connectivity rate ($\alpha \approx [0.01 - 0.05]$) and the spectral radius ($\rho \approx [0.1 - 0.99]$) are crucial parameters of ESN. The spectral radius α of an ESN is the maximum of all eigenvalues of the reservoir weights and the connectivity rate α represents the connection situation among the neurons in the reservoir. These hyperparameters have a significant impact on how well ESN performs. In spite of certain suggestions made by earlier research [55,56], it might be challenging to choose the right parameter values for a given application. In this study, we use the Bayesian approach [57] to find parameter values appropriate for the particular application in order to achieve good prediction performance.

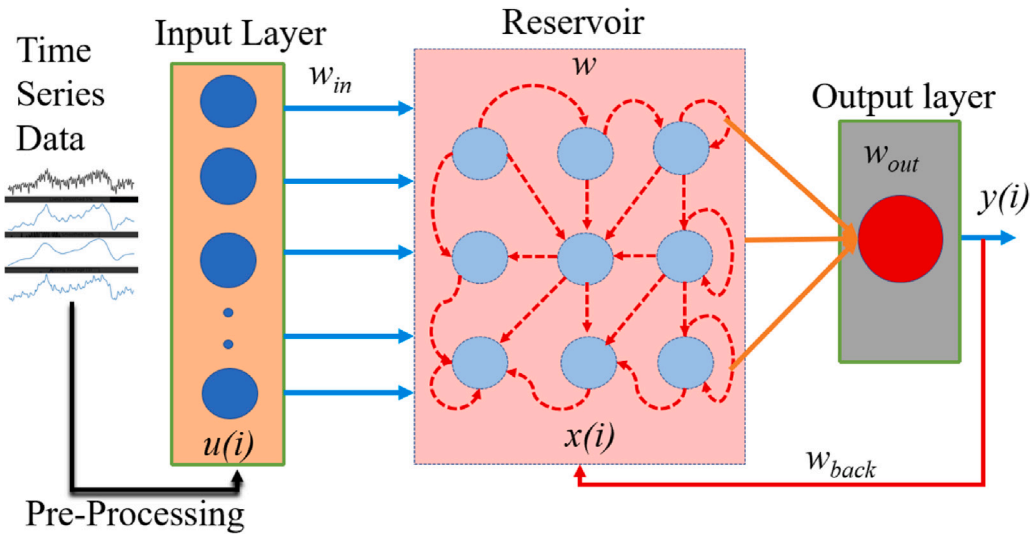

Fig. 2. Schematic view of Echo State Network (ESN).

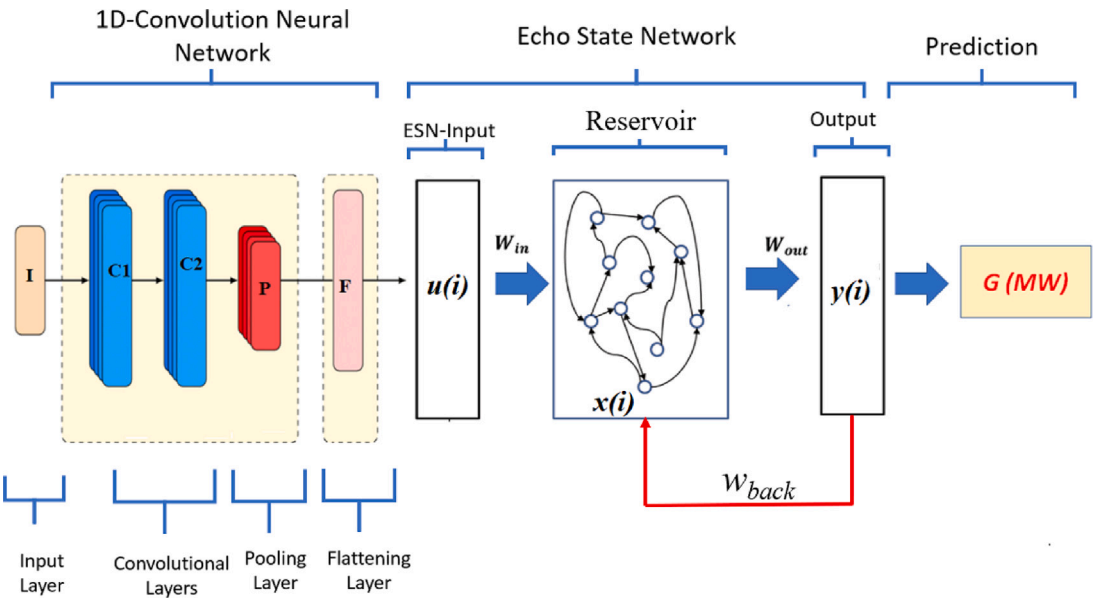

Fig. 3. Schematic view of 1D-Convolution Neural Network (CNN) integrated with Echo State Network (ESN) to develop the hybrid Deep Learning model CESN.

2.3. Proposed CNN-echo state network (CESN)

The architecture of the proposed deep hybrid model is shown in Fig. 3. The ESN model was integrated with CNN to form a hybrid model called CESN for the prediction of daily electricity demand G. In our proposed G prediction framework, the CNN model is used for extracting data features from the lagged series of the G. The CNN portion of the model extracts both features of the common and local trend that need various periods of time from the G time series. The extracted data features from the CNN layers are flattened out before passing it through the ESN model, which worked as the predictive regression operator for G data. One of the advantages of this deep hybrid CESN model is its capacity to extract complex spatial and temporal information from a lagged G time series and retain these elements for prediction. As aforementioned, the ReLU was used as an activation function for CNN, Adam was used as the back propagation algorithm, and the FCL was replaced by the ESN model in this proposed CESN model. A detailed summary of the CESN architecture, including all its parameters is shown in Appendix B.

2.4. Benchmark models

Several ML benchmark models have been considered for comparison with the proposed CESN approach. We describe here the most important characteristics of these ML approaches.

2.4.1. Deep Neural Network (DNN)

The Deep Neural Network (DNN), also known as the multi-layer feed-forward ANN, is an advanced form of the ANN. DNN model comprises of the input layer, hidden layers and the output layer [58]. The input layer is used to receive signals from peripherals or systems, and the output layer makes decisions about inputs. Hidden layers exist between the input and output layers, which are the actual computational engines of the DNN model [59]. DNN techniques are well suited for big data analysis and have useful applications in time series forecasting, pattern and speech recognition, computer vision, and natural language processing. Furthermore, DNN methods are mostly used in the literature for nonlinear prediction tasks [60–63].

Fig. 4. Map of Southeast Queensland showing the exploded section of the study sites where daily electricity Demand (G) prediction is done.

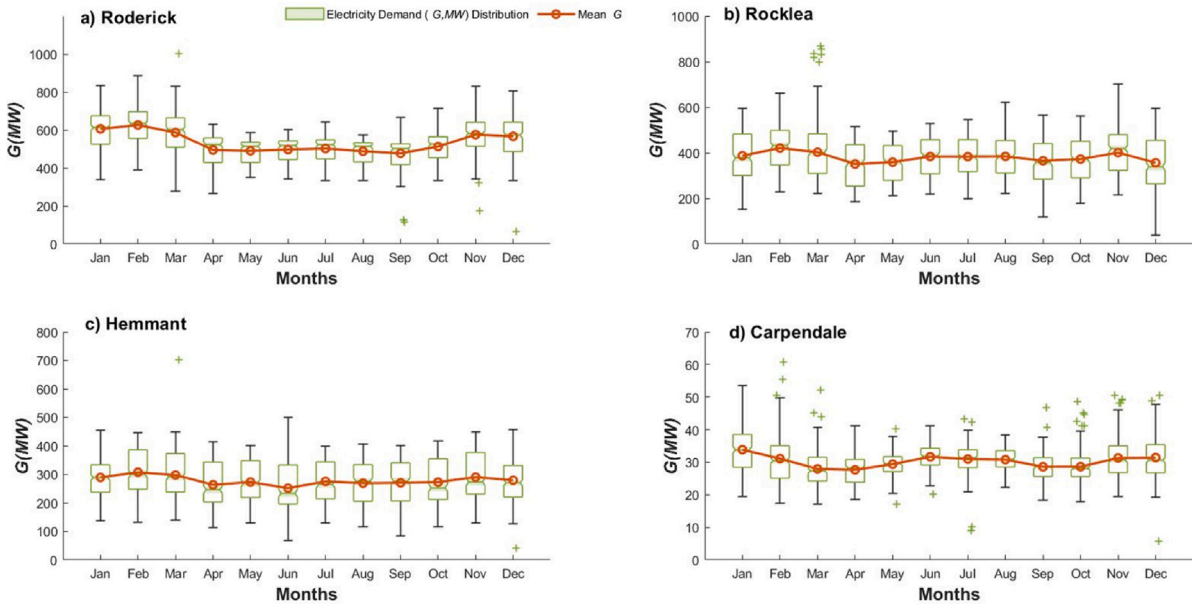

Fig. 5(a). Monthly distribution of G in the Study substations based on months of years.

2.4.2. Support Vector Regression (SVR)

SVR, which can be applied to both classification and regression tasks, is one of the widely statistical machine learning algorithms and used the structural risk minimization (SRM) principle [64,65]. SVR employs the kernel function to map lower-dimensional data into higher-dimensional data, this kernel function can be of a different form (Gaussian, Polynomial, and linear). The SVR model in this study utilizes the Gaussian kernel function, also known as Radial Basis Function (RBF) [66]. RBF is a popular choice for regression tasks, primarily due to its capability of handling intricate error boundaries and its simple tuning process.

2.4.3. Xtreme Gradient Boosting regressor (XGB)

Xtreme Gradient Boosting (XGB) belongs to the family of decision tree-based boosting algorithms [67]. The XGB model's less computational cost and precision have prompted us to compare its performance to the CESN for G prediction. The trees or weak learners are added in sequential order, and only the residuals are fed to the next weaker

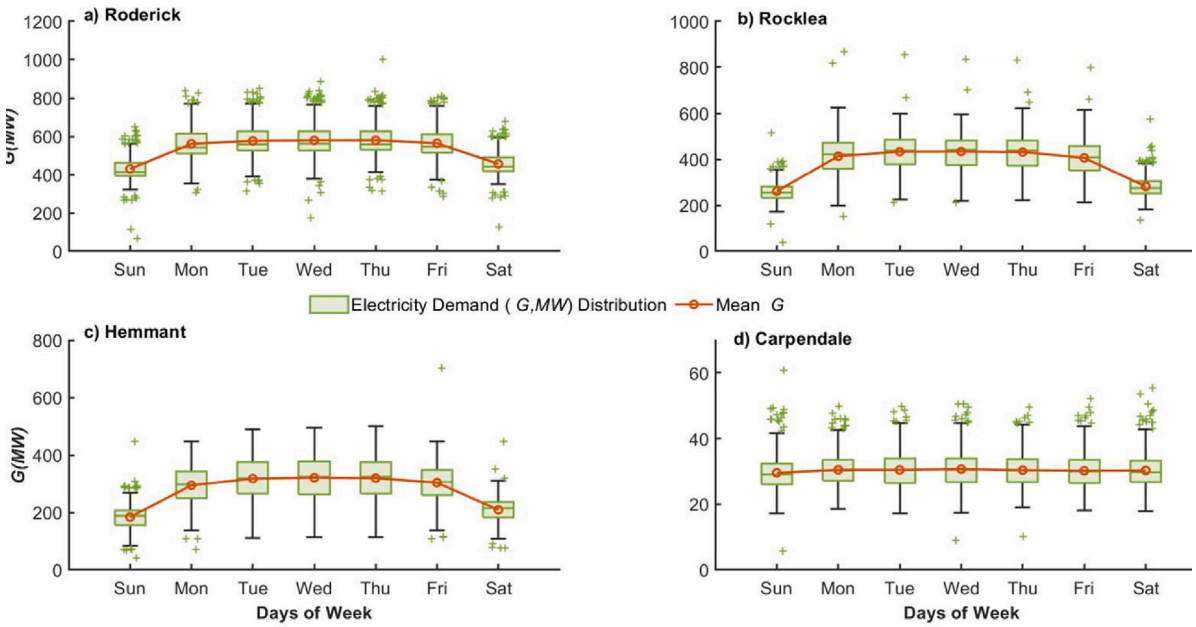

Fig. 5(b). Box plot showing the distribution of G in the study substations based on days of the week.

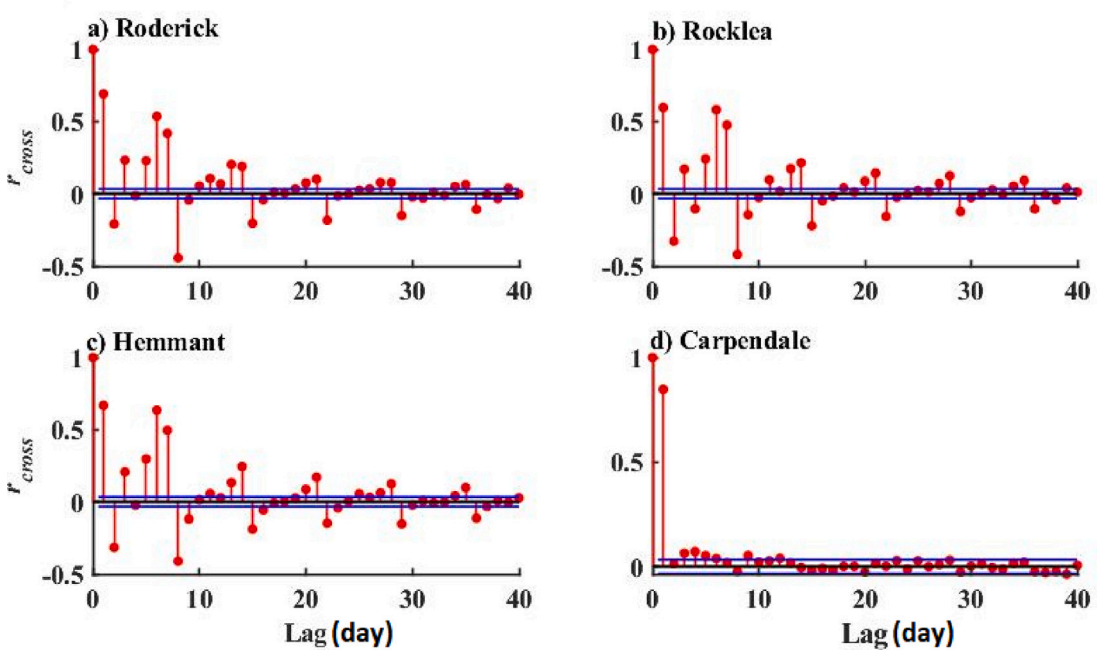

Fig. 6(a). Partial autocorrelation plot of daily (G) time-series data for four substations. Red bars that extend across the horizontal blue lines are statistically significant. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

learner thereby reducing errors [68]. Contrary to RNN and NN with back propagation, the XGB model is not prone to the vanishing gradient problem as it employs the Newton boosting based on the Newton–Raphson technique which accelerates the approach to global minima thereby eliminating the vanishing gradient problem.

2.4.4. Light Gradient Boosting decision tree model (LGBM)

Microsoft released the Light Gradient Boosting decision tree (LGBM) model as an open source model in 2017 [69]. LGBM is a boosting ensemble model that transforms coupled weak learners into a potential model [43]. LGBM improves upon Gradient Boosted Decision Trees (GBDT) models in terms of faster execution time and reduced memory consumption while maintaining high accuracy [70]. As the amount of data grows, the accuracy of traditional GBDT-based models decreases and the prediction speed drops significantly [71]. LGBM models use histogram-based algorithms to reduce the impact of high-dimensional data, speed up computation time, and prevent overfitting of predictive systems [72]. Recent studies have used the LGBM as a prediction model for wind speed [73–75] prediction, solar radiation prediction [76] and also in electricity load prediction problems [77].

2.4.5. Multi-Layer Perceptron model (MLP)

Artificial Neural Network architectures encompass many kinds of algorithms with exceptional designs to fits well-described applications [47]. The distinction among those algorithms is typically associated with the information processing strategies adopted. The MLP

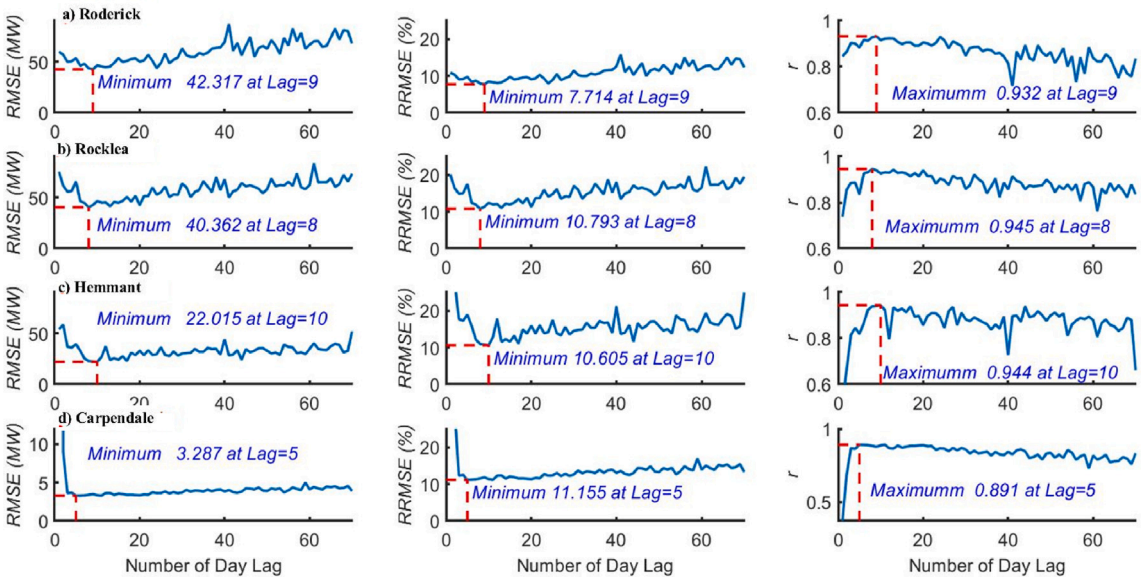

Fig. 6(b). Lag period optimization for daily electricity prediction based on *RMSE* (MW), *RRMSE* (%) and correlation coefficient r for all substations.

model is mostly used in prediction tasks [78]. This data-driven method analyzes the embedded feature patterns to produce a meaningful description of the next time, steps [79,80]. The MLP structure includes 3 layers of perception, namely, the input layer, the hidden layer, and the output layer. Each layer has several processing units called neurons, and each neuron is fully interconnected with weighted connections to the unit in the next layer [81]. The learning process comes from the error of back-propagation using the gradient descent research approach. The benefits of the MLP model in prediction tasks are: quicker implementation, fast training, robustness to outliers and/or missing data; moreover, when an MLP is designed with an appropriate number of neurons, it is a universal approximator [82]. The general expression of the network can be given as follows [83]:

$$y = f_2\left(\sum_{j=1}^{N}\left(w_j f_1\left(\sum_{i=1}^{n} h_{ij} + b_j\right) + b_0\right)\right) \tag{6}$$

where y is the predicted output, h_{ij}, b_j, and f_1 are respectively, the weight matrix, the bias vector, and the hidden layer activation function, and w_j, b_o, and f_2 are the weight vector, the bias scalar, and the output layer activation function respectively.

3. Material and methods

In this section, the electricity demand (G) dataset and the key processes used in the development of the deep hybrid CESN model are briefly introduced.

3.1. Electricity demand data

Electricity demand data (G) for this study was requested from Energex, the federal government-owned electricity distribution company based in South East Queensland (SEQ). The Energex network extends across an area of over 25,000 square kilometers. More than 53,000 km of overhead and subterranean power lines, 288 substations, and around 50,000 distribution transformers all contribute to the delivery of electricity. The company's system operation records half-hourly G in Megawatts (MW), and therefore this research has utilized the data in MW to create predictive models for electricity demand. Note that for the interpretation of the model performance in terms of the electricity demand that is normally measured in MWh, the respective timescale over which results are presented should be applied (with Energy (MWh) = Power (MW) × Time (hours)). This will help normalize the different experiments where temporal scales are different. This database provides the company with enough information about actual demand every half-hour, considered during maintenance planning. Table 1 shows the statistical characteristics of the substations G dataset. Besides this, Fig. 7 in Section 4 shows examples of electricity demand profile G generated by the proposed CESN model.

The training procedure of the proposed CESN model uses data from four different substations (Roderick, Rocklea, Hemmant and Carpendale) between 01/07/2011 and 30/06/2021 at 30-minute, three hours, and daily prediction levels. Fig. 4 shows the location of these stations. It is noteworthy that half-hour Energex data were converted to 3-hourly by calculating the sum of every 6 values and similarly for the daily interval comprised of 48 values for the day.

Fig. 5(a)(a–d) shows a boxplot with the distribution of G at four substations over 1/07/2011 to 30/06/2021, including the median, first and third quartile, minimum and maximum value of electricity demand (MW). The electricity demand during the summer months (Dec, Jan and Feb) and winter (Jun, Jul and Aug) were relatively higher compared with spring months (Sept, Oct and Nov) and Autumn (Mar, Apr and May). Furthermore, weekend G consumption was lower than that of weekdays: the mean G on weekdays was 560, 420, 360, and 28 MW compared to 420, 260, 180, and 27 MW for Roderick, Rocklea, Hemmant and Carpendale, respectively (see Fig. 5(b)).

3.2. Predictive model development

3.2.1. Selection and initialization of G time-series data

In this study, the antecedent lagged inputs based on the G data are used as predictor variables to train the proposed deep hybrid CESN, as well as the benchmark models. This was done by first determining the salient lagged inputs based on G time-series using partial autocorrelation function (PACF), and then further validating the lagged inputs by using Extreme Learning Machine model. The PACF shows the correlation, r_{corr}, between the G time series at L_t for up to a specific number of lags. Fig. 6(a) shows the statistically significant lagged data of each substation, the X-axis is the time of the lags, and the blue line represents the confidence interval. Any PACF outside the confidence level represents a strong partial autocorrelation for the series at certain lags. As per Fig. 6(a), it can be seen that the lags at 1, 2, 3, 5, 6, 7, 8, 11, 13, 14 and 15 (for Roderick), 1, 2, 3, 5, 6, 7, 8, 9, 13, 14 and 15 (for Rocklea), 1, 2, 3, 5, 6, 7, 8, 9, 13, 14 and 15 (for Hemmant), and

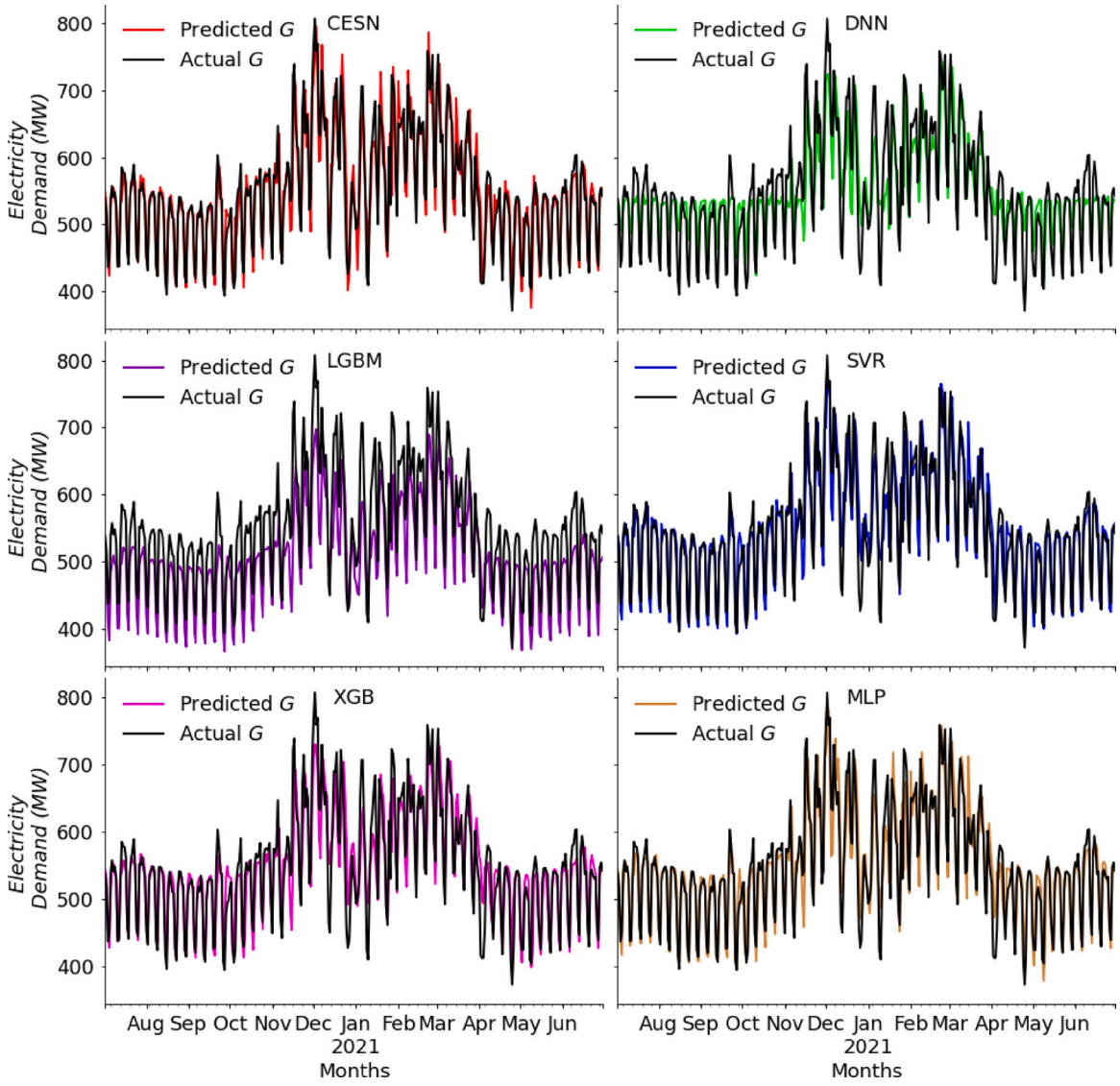

Fig. 7. Performance of our proposed CESN model for daily electricity demand prediction compared with six benchmark models SVR, MLP, XGB, DNN, and LGBM in Roderick substation.

1, 2, 3, 4 and 5 (for Carpendale) have a high correlation with the G at time t. Furthermore, to validate the *PACF* criteria for selecting the order of lag variable, the Extreme Learning Machine (ELM) model was utilized. Fig. 6(b) shows the ELM model (Transfer function= *Sigmoid* and number of hidden neuron= *50*) performance in terms of root mean square error (*RMSE*, MW), relative root-mean square error (*RRMSE*, %) and correlation coefficient (r) with varying lag (*1–40*) for the 4 substations at SEQ.

It is important to note that from, Fig. 6(b), the optimal lagged input combination was 9, 8, 10 and 5 for daily G data for Roderick, Rocklea, Hemmant and Carpendale, respectively, with the higher magnitude of r and lower magnitude of *RMSE* and *RRMSE*. Therefore, the input and output variables of the deep hybrid CESN and benchmark models can be formulated as follows (example for 10 lags):

$$input = \left\{ \begin{array}{c} x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10} \\ x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11} \\ \vdots \\ x_{t-1}, x_{t-2}, \ldots, x_{t-9}, x_{t-10} \end{array} \right\}, \quad t \geq 10 \tag{7}$$

$$output = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_t \end{pmatrix}^T, \quad t \geq 10 \tag{8}$$

3.2.2. Preprocessing and data segregation

In this study, there is no missing G data between the period 01/07/2011 to 30/06/2021, therefore no imputation mechanism is required. The normalization of the lagged G data is done by $Z - score$ normalization. A z-score or standard score is obtained when the population mean is subtracted from an individual raw score. The difference is then divided by the population standard deviation to obtain a dimensionless quantity, which is the standard score. The ability to calculate prediction intervals is one of $z - score$ normalization's strong points. Normalized attributes are less impacted by outliers since the $z - score$ normalization is based on the standard deviation rather than the range. The Z-score can be defined as:

$$G' = \frac{G - \mu}{\sigma} \tag{9}$$

where G' stands for the normalized G, and μ and σ are mean and standard deviation of G respectively. Once the Lagged data is normalized, they are further segregated into training, validation and testing. The training data is used to train CESN as well as benchmark models. The testing data have no effect on training and provide an independent measure of network performance during and after training. The validation data are used for measuring the generalization capability of the network and stop training when generalization comes to an end. In this

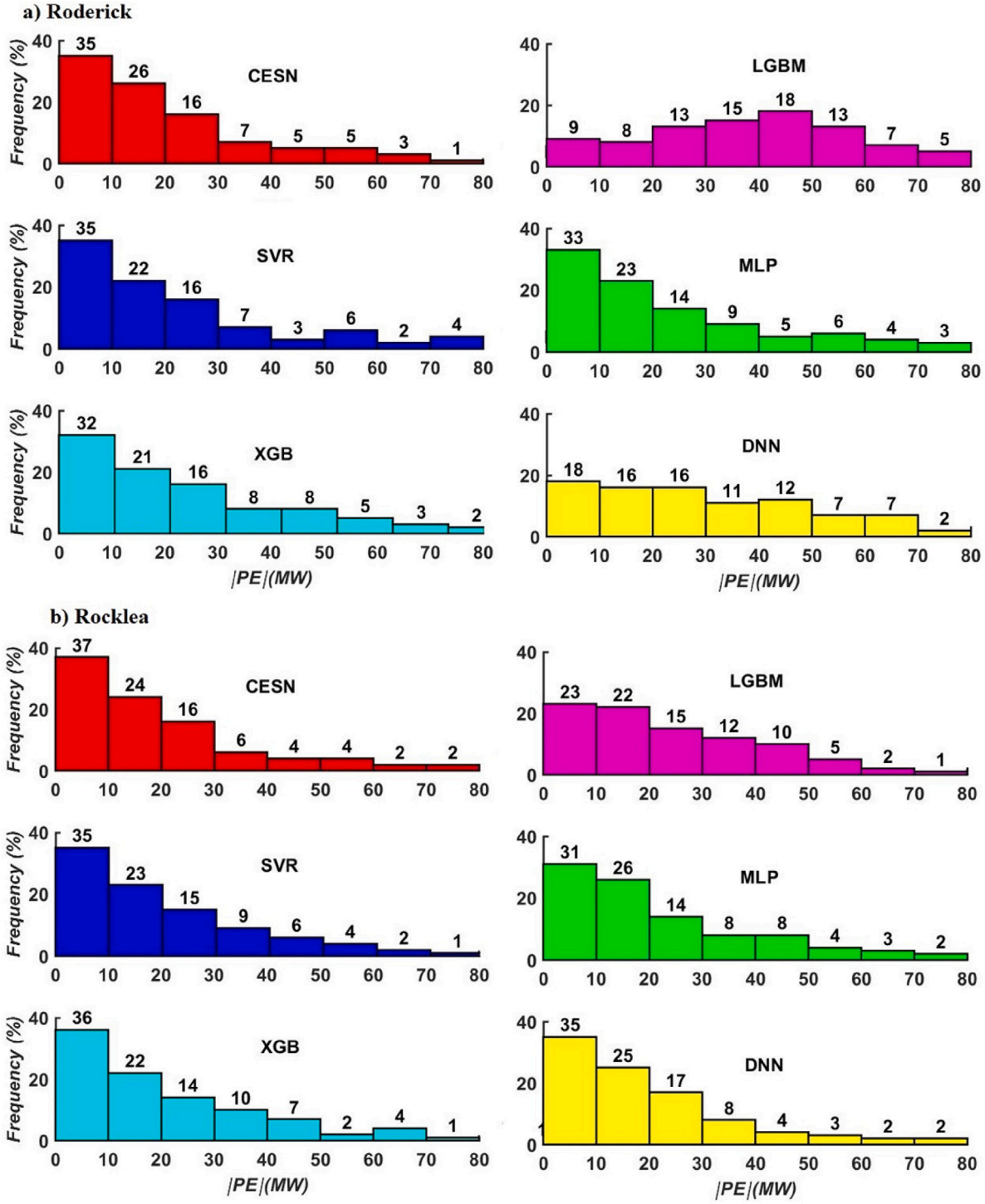

Fig. 8. Histogram illustrating the frequency (in percentages) of absolute Prediction errors ($|PE|, MW$) of the best performing CESN model during test period compared with benchmarked models for the prediction of daily electricity demand (G, MW) at (a) Roderick, (b) Rocklea, (c) Hemmant, and (d) Carpendale.

study, the data from 01/07/2011 to 30/06/2020 (3652 data points) are used for training and validation whereas the data from 01/07/2020 to 30/06/2021 (365 data points) are used for testing. During model training, the validation split was set as 0.2 *i.e.* 20%, which means the 20% of the training data are used for validation purposes.

3.2.3. CESN model development

As previously stated, this article combines CNN and ESN to propose a deep hybrid CESN prediction model. CNN is used to perform feature extraction on the preprocessed data, and the extracted feature vectors are used as inputs into the ESN for model prediction. The model combines CNN's ability to extract data features, improve the prediction accuracy of the model, and uses ESN to predict the model, which improves the efficiency of model training. A summarized CESN algorithm (Fig. 3) involves the following eight steps:

- Convert normalized G dataset to signal vector as input to the CNN model.

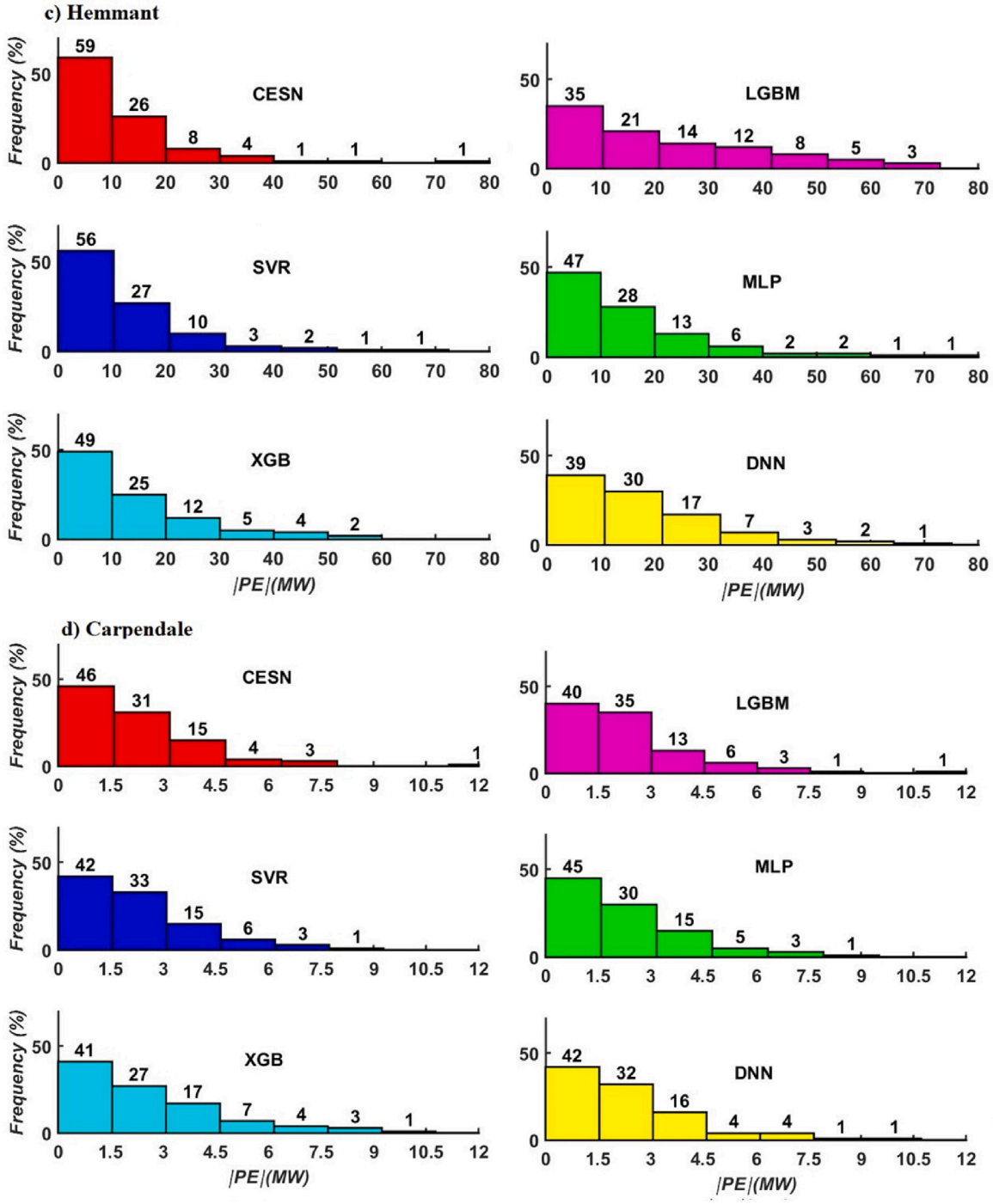

Fig. 8. (*continued*).

- Initialize the CNN parameters, such as the maximum epoch of iteration, the learning rate, the optimizer and the number of filters.
- Implement convolution and pooling.
- Fine tune the stacked model using the Bayesian optimization algorithm.
- Finish the optimization and acquire the features of the data set
- Use the extracted features for regression by ESN.
- Initialize the ESN parameters, such as the total number of neurons in the reservoir, Spectral radius, and the Connectivity rate.
- Get the final optimized CESN model and predict on test data-set.

As aforementioned, this study has utilized one of the popular ML hyperparameter optimization (HPO) [57] algorithms known as Bayesian optimization. Bayesian optimization is a sequential design strategy and can minimize the objective function efficiently. The Hyperopt python library was employed in this study. Hyperopt offers the sequential model-based optimization (SMBO) method that uses the Tree of Parzen estimators (TPE) algorithm to optimize the hyperparameters. Table 2 displays the hyperparameters search space for the CESN and benchmark models'. Before selecting the ideal model hyper-parameter, HyperOpt takes into account all possible combinations within this search space. Apart from HPO, when developing the deep hybrid CESN model for

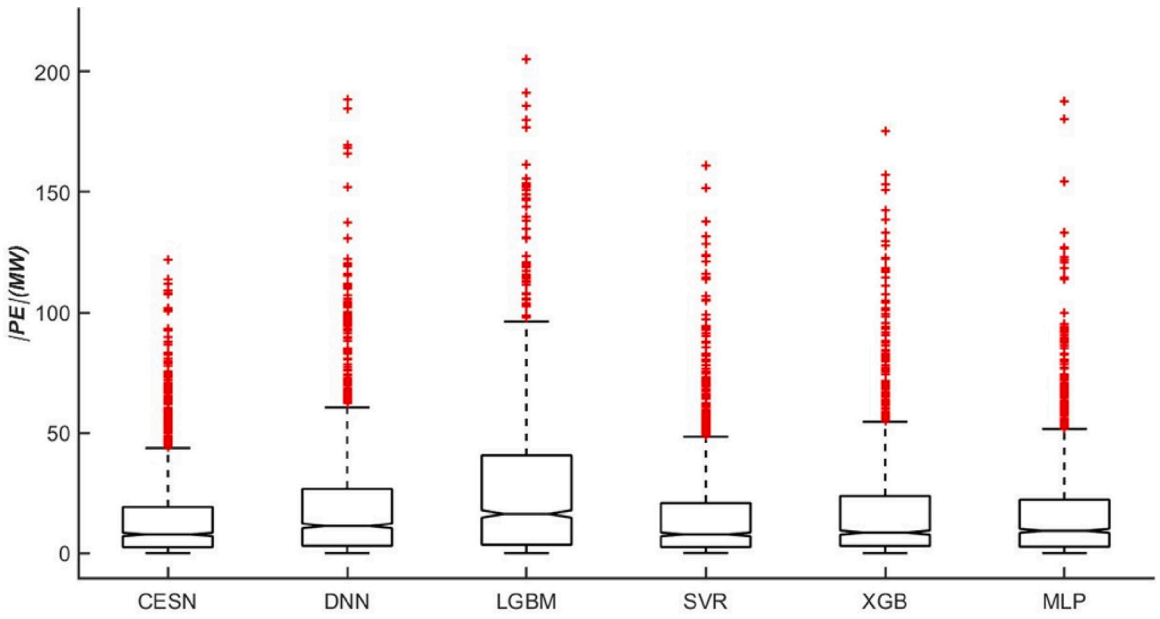

Fig. 9. Box plots of the prediction error (PE) generated by CESN vs other benchmarked models in the testing phase. For notations and model names, please refer to Table 3.

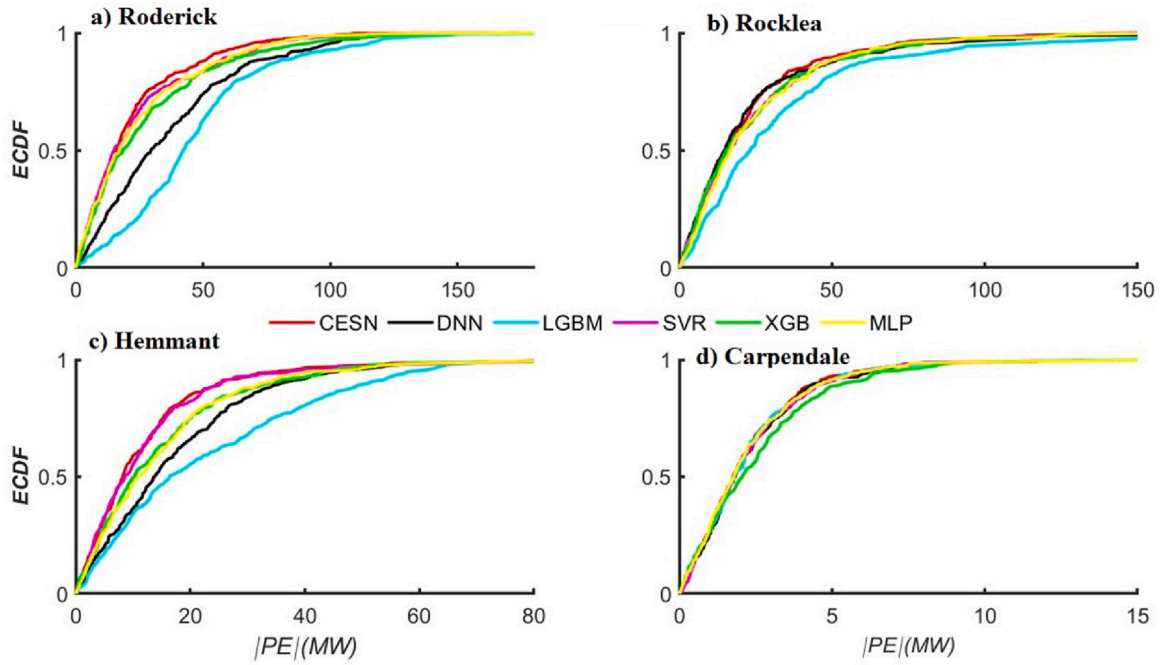

Fig. 10. Empirical cumulative distribution function (ECDF) of the prediction error generated by the CESN model vs. DNN, LGBM, SVR, XGB and MLP models. (a) Roderick, (b) Rocklea, (c) Hemmant, and (d) Carpendale.

daily G prediction, this study also used the following regularization strategy to minimize the overfitting.

- The parameter of dropout is set to 0.1, which makes 10% of random neuron nodes in each hidden layer to be discarded. This dropout regularization is used to weaken the strong dependence of some nodes and distribute the back-propagation correction value to each parameter in a balanced manner.
- The learning rate is reduced by using the callback function called $\varepsilon ReduceLROnPlateau\varepsilon$ to further improve the performance of the neural network when the evaluation indicators stop changing in the network. The specific operation of the callback method is that if the model performance in the validation set stops improving, the learning rate is reduced at a rate of 0.7 times. The lower limit of the learning rate is set to 0.00001.
- Further, in order to reduce model overfitting when training goes too long, we used a technique called "early stopping (es)". This es method tracks the model performance against validation data and halts the training after 10 consecutive epochs without additional improvement.

3.2.4. Benchmark model development

Regarding alternative ML for comparison, five other well-known prediction models based on the Deep Neural Network (DNN), Light Gradient Boosting Machine (LGBM), Xtreme Gradient Boosting(XGB), Support Vector Regression (SVR), and Multilayer Perceptron (MLP)

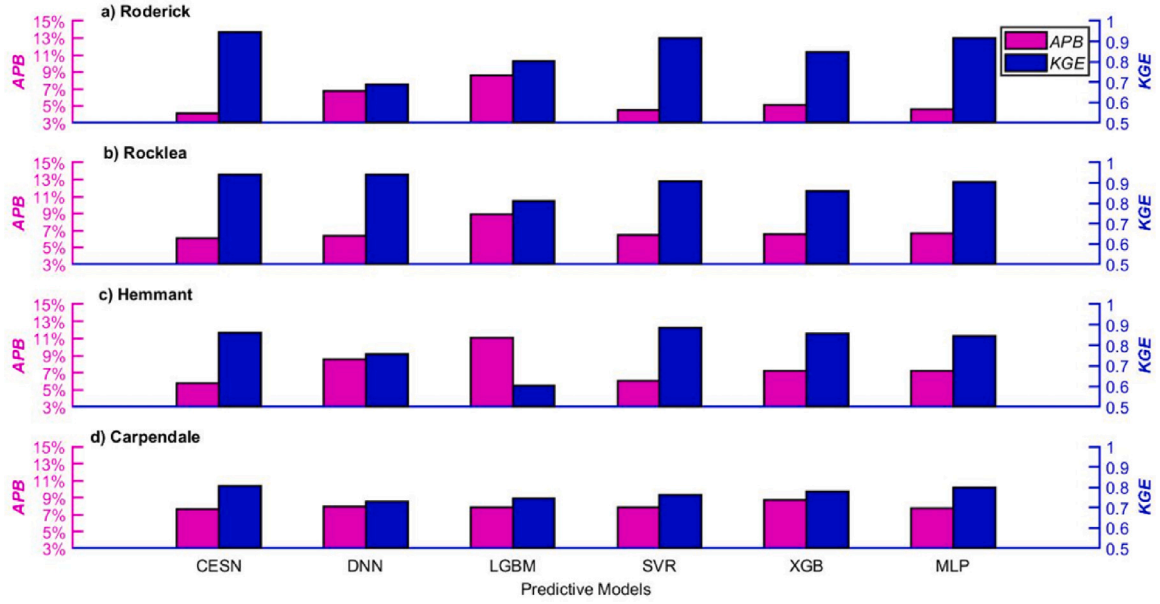

Fig. 11. Bar chart showing a comparison of the optimal models in terms of their absolute percentage bias ($APB, \%$) and the Kling–Gupta efficiency (KGE) in the testing phase. For notations and model names, please refer to Table 3.

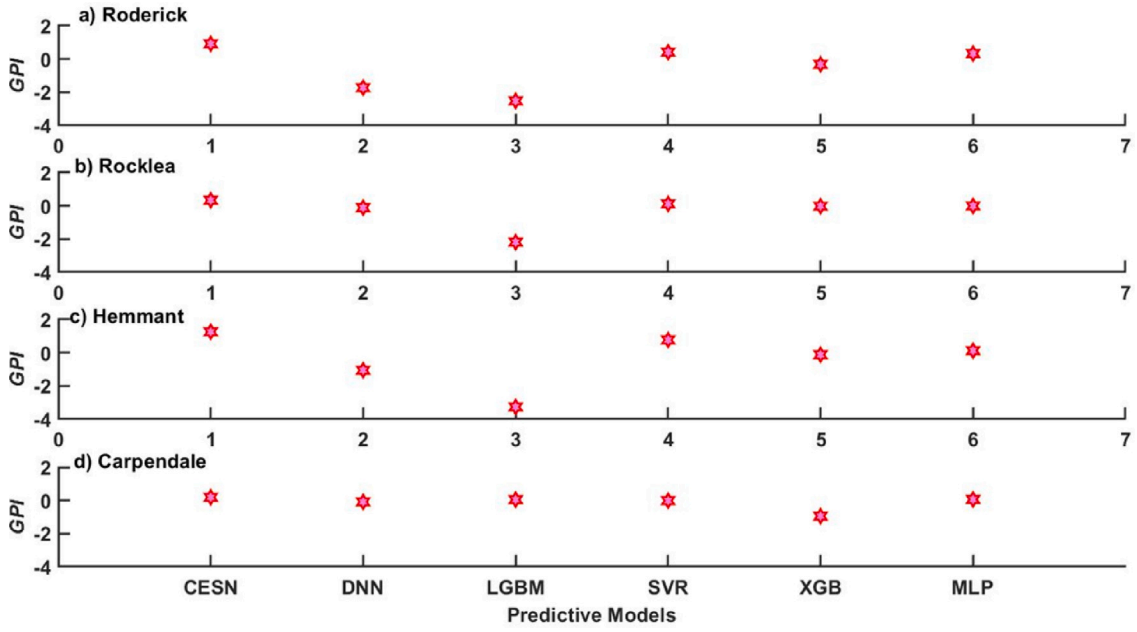

Fig. 12. The evaluation of the proposed CESN model against the five other benchmarked models based on Global performance indicator (GPI).

were developed. All of them were built using Keras 2.2.4 [84] on TensorFlow 1.13.1 [85] backend in Python 3.6. They were trained on a system with 32 GB RAM and an Intel Core i7 processor.

3.3. Performance evaluation metrics considered

Several metrics have been used to evaluate model efficiency. Each metric informs about different quantitative behavior and has its own strengths and weaknesses. This work uses the following set of common statistical metrics [47,65,86–89]: Correlation (r) (10), Root Mean Square Error (RMSE) (11), Mean Absolute Error (MAE) (12), Root Mean Square Percentage Error (RMSPE) (13), Mean Absolute Percentage Error (MAPE)(14), Willmot's Index (WI) (15), Nash–Sutcliffe Equation (NS) (16), Legates and McCabe's (LM) (17), Explained Variance Score (E_{var}) (18), Uncertainty at 95% ($U_9 5$) (19) and t estimator(19).

$$r = \frac{\sum_{i=1}^{n}(G^m - \langle G^m \rangle)(G^p - \langle G^p \rangle)}{\sqrt{\sum_{i=1}^{n}(G^m - \langle G^m \rangle)^2}\sqrt{\sum_{i=1}^{n}(G^p - \langle G^p \rangle)^2}} \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(G^p - G^m)^2} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|G^p - G^m| \tag{12}$$

$$RMSPE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(G^p - G^m)^2}}{\langle G^m \rangle} \tag{13}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|G^p - G^m|}{G^p} \tag{14}$$

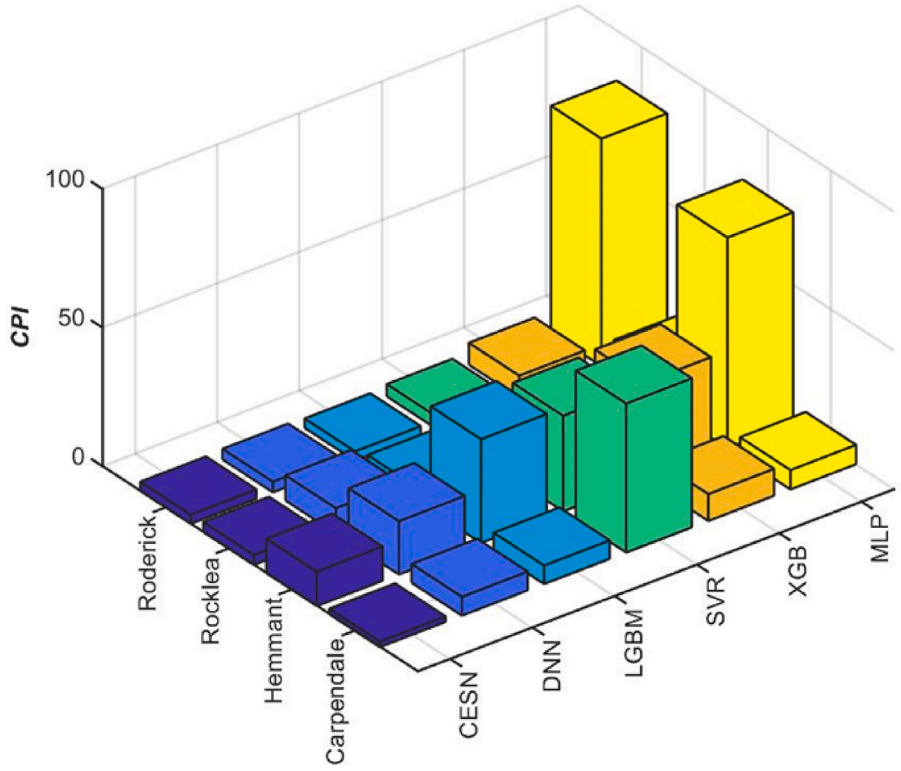

Fig. 13. 3D-bar plot of the Combined Performance Index (CPI) of CESN vs. the comparative DNN, LGBM, SVR, XGB and MLP models used in prediction of daily electricity demand (G). Names for each model are provided in Table 3.

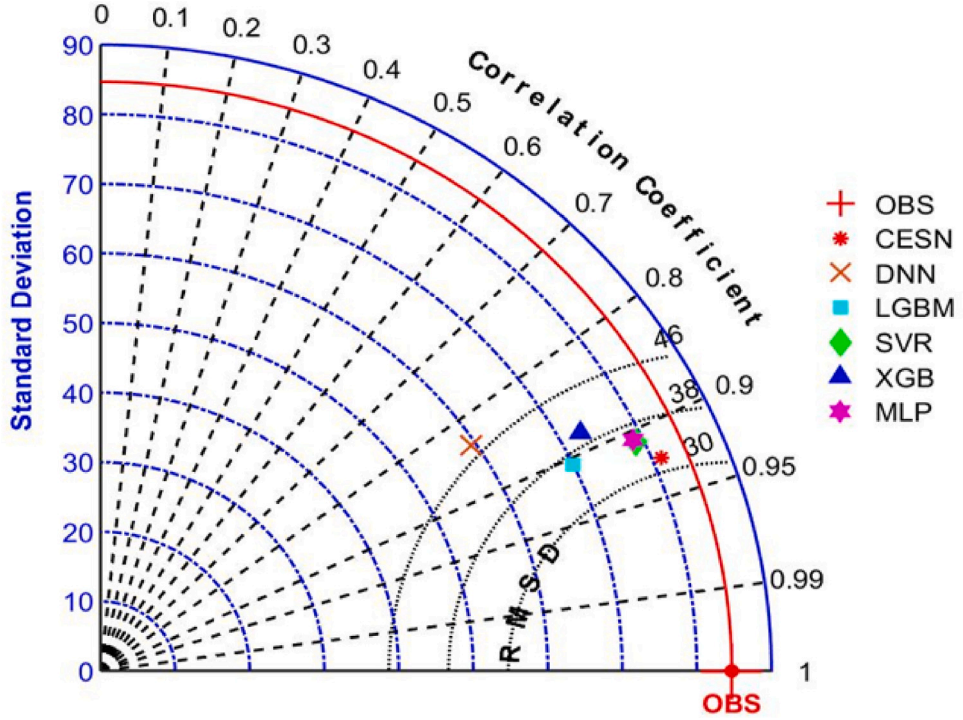

Fig. 14. Taylor diagram depicting the correlation coefficient, standard deviation and root mean square deviation ($RMSD$) of CESN vs. DNN, LGBM, SVR, XGB and MLP models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$WI = 1 - \frac{\sum_{i=n}^{n}(G^m - G^p)^2}{\sum_{i=n}^{n}(|G^p - \langle G^m \rangle| + |G^m - \langle G^m \rangle|)^2} \quad (15)$$

$$NS = 1 - \frac{\sum_{i=1}^{n}(G^m - G^p)^2}{\sum_{i=1}^{n}(G^m - \langle G^m \rangle)^2} \quad (16)$$

$$LM = 1 - \frac{\sum_{i=1}^{n}|G^m - G^p|}{\sum_{i=1}^{n}|G^m - \langle G^m \rangle|} \quad (17)$$

$$E_{var} = 1 - \frac{\mathrm{Var}(G^m - G^p)}{\mathrm{Var}(G^m)} \quad (18)$$

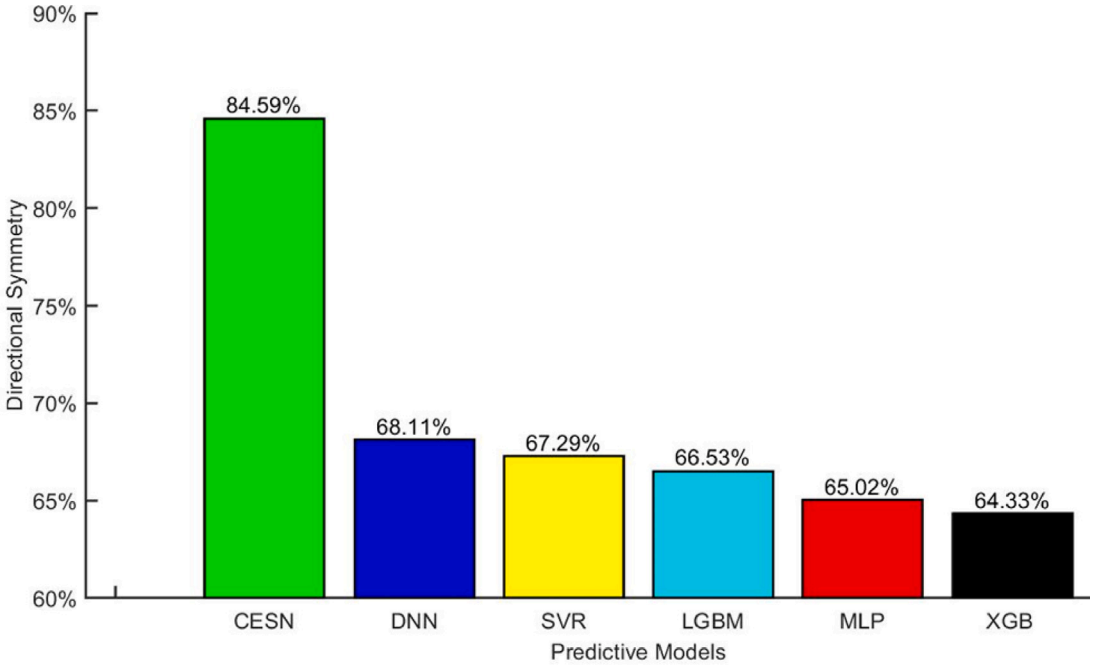

Fig. 15. The performance of the proposed CESN model in comparison to other models using directional symmetry (DS) criteria.

Table 2
Architecture of the CNN integrated with ESN (CESN) model vs. SVR, MLP, XGB, DNN, and LGBM models developed for daily electricity demand prediction at four sites of Southeast Queensland. Note: - ReLU = Rectified Linear Units; Adam = Adaptive Moment Estimation, gbdt = traditional Gradient Boosting Decision Tree, rbf = Radial Basis Function, logistic = Logistic Sigmoid Function, tanh = Hyperbolic Tangent Activation Function, loguniform = In the log-uniform distribution, points are sampled uniformly between $log(a)$ and $log(b)$, where log is most frequently the logarithm with base 10.

Predictive models	Model hyperparameters	Hyperparameter selection	Roderick	Rocklea	Hemmant	Carpendale
Convolution Neural Network Integrated with Echo State Network (CESN)	Filter1	[50, 80, 100, 200]	**100**	**50**	**100**	**50**
	Filter 2	[40, 50, 60, 70, 80]	**60**	**50**	**40**	**40**
	Filter 3	[20, 10, 30, 5]	**5**	**30**	**10**	**5**
	Reservoir neurons	[50, 80, 100, 200]	**100**	**80**	**100**	**50**
	Spectral radius	[0.5, 0.55, 0.6, 0.7, 0.75, 0.8, 0.9]	**0.6**	**0.55**	**0.7**	**0.5**
	Connectivity rate	[0.01, 0.02, 0.03, 0.04, 0.05]	**0.01**	**0.03**	**0.01**	**0.01**
	Epochs	[1000]				
	Solver	['adam']				
	Batch size	[5, 10, 15, 20, 25, 30]	**5**	**10**	**5**	**5**
Deep Neural Network (DNN)	Hiddenneuron 1	[100, 200, 300, 400, 50]	**300**	**200**	**100**	**100**
	Hiddenneuron 2	[20, 30, 40, 50, 60, 70]	**50**	**40**	**20**	**20**
	Hiddenneuron 3	[10, 20, 30, 40, 50]	**50**	**10**	**10**	**20**
	Hiddenneuron 4	[5, 6, 7, 8, 12, 15, 18]	**18**	**5**	**5**	**5**
	Batch size	[5, 10, 15, 20, 25, 30]	**10**	**5**	**5**	**5**
	Solver	['adam']				
	Epochs	[1000]				
Light Gradient Boosting Machine (LGBM)	Boosting Type	'gbdt'				
	Maximum tree leaves	[40, 50, 60, 70]	**50**	**40**	**50**	**40**
	Boosting learning rate	[0.01, 0.1, 0.001, 0.005]	**0.001**	**0.001**	**0.005**	**0.001**
	Maximum tree depth	[5, 8, 10, 20, 25]	**20**	**10**	**20**	**10**
	Boosted trees to fit.	[20, 50, 70, 80, 100]	**70**	**50**	**70**	**50**
Extreme Gradient Boosting (XGB)	Boosting learning rate	[0.01, 0.1, 0.001, 0.005]	**0.001**	**0.01**	**0.005**	**0.001**
	Maximum tree depth	[5, 8, 10, 20, 25, 30, 40]	**40**	**25**	**40**	**10**
	Boosted trees to fit.	[30, 50, 70, 80, 90, 100]	**30**	**70**	**90**	**30**
Support Vector Regression (SVR)	Kernel	'rbf'				
	Epsilon	0.0001				
	Cost Function (C)	loguniform (−40, 10)	**161.516**	**21510.1**	**521.72**	**20646.7**
	Penalty function (Gamma)	loguniform (−40, 10)	**0.03304**	**0.0688**	**0.118**	**0.0683**
Multilayer Perceptron (MLP)	Hidden neuron	[50, 60, 70, 80, 90, 100]	**60**	**80**	**70**	**80**
	Activation function	['relu', 'logistic', 'tanh']	**logistic**	**relu**	**relu**	**tanh**
	Learning rate	[0.001, 0.002, 0.005, 0.006]	**0.001**	**0.005**	**0.002**	**0.001**
	Solver	['adam']				

$$U_{95} = 1.96(SD^2 - RMSE^2)^{0.5} \tag{19}$$

$$t = \sqrt{\frac{(n-1)\cdot MBE^2}{RMSE^2 - MBE^2}} \tag{20}$$

We also use the Theil's inequality coefficient (TIC) (21), Standard Deviation of Relative Error (STDRE) (22), the skill score (SS) (23), the relative RMSE ($RMSE_r$) (24), Kolmogorov–Smirnov Test Integral (KSI) (25), the Critical Limit Over-estimation (OVER) (26) and the Combined Performance Index (CPI) (27).

$$TIC = \frac{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(G^p - G^m)^2}}{\left(\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(G^m)^2} + \sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(G^p)^2}\right)} \tag{21}$$

$$STDRE = \left(\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{G^p - G^m}{G^m}\right)^2\right)^{1/2} \tag{22}$$

$$SS = 1 - \frac{RMSE(p,x)}{RMSE(pr,x)} \tag{23}$$

$$RMSE_r = \frac{RMSE(p,x)}{RMSE(r,x)} \tag{24}$$

$$KSI = \frac{100}{A_c}\int_{X_{min}}^{X_{max}} D_n dx \tag{25}$$

$$OVER = \frac{100}{A_c}\int_{X_0}^{X_1} \max(D_n - D_c, 0) dx \tag{26}$$

$$CPI = \frac{KSI + OVER + 2RMSE}{4} \tag{27}$$

where G^m and G^p are the observed and predicted value of G, $\langle G^m\rangle$ and $\langle G^p\rangle$ are the observed and predicted mean of G, p stands for the model prediction, x for the observation, pr for perfect prediction (persistence), and r for the reference prediction.

The above metrics have the following skills:

- r are in the interval $[-1, 1]$. MAE and $RMSE$ are 0 in perfect predictions and ∞ in a worst fit.
- Standard deviation of relative error ($STDRE$) calculates the data dispersion. A smaller value of $STDRE$ indicates that the data is less scattered [90].
- Models with lower U_{95} and t-statistics value close to 0 perform better predictions.
- The value of Theil's inequality coefficient (TIC) is in the interval $[0, 1]$ where TIC close to 0 indicates good predictions [91].
- RMSPE and MAPE have ranged from 0% to 100%. Model performance is excellent when $RRMSE$ is lower than 10%, good if $RRMSE$ is between 10% and 20%, fair if $RRMSE$ is between 20% and 30%, and poor if $RRMSE$ is greater than 30% [92].
- WI is a metric that overcomes the insensitivity issues of RMSE and MAE. WI values close to 0 show inaccurate models and those close to 1 [93].
- NS evaluates both the observed and predicted variance G and it has ranges from $-\infty$ (the worst case) to 1 (perfect fit) [94].
- LM is a more robust metric developed to address the limitations of both WI and NS [95] and has value ranges between 0 and 1 (the best case).
- E_{var} computes the biased variance to explain the fraction of variance. It has a range from 0 to 1.
- $RMSE$, $OVER$ and KSI are combined into a single metric named CPI which takes into account dispersion, absolute bias (through RMSE) and likeness of distributions (through KSI and OVER). KSI gives information about the similarity between the distributions of the measured and modeled diffuse fractions and discriminates well between different models. OVER describes the relative frequency of exceedance situations, when the normalized distribution of modeled data points in specific bins exceeds the critical limit that would make it statistically undistinguishable from the reference distribution. A CPI value close to 0 shows an excellent performance [96].

Furthermore, the overall performance of the model was evaluated by the Global Performance Indicator (GPI) that can be calculated by the six metrics as follows [97].

$$\mathrm{GPI}_i = \sum_{j=1}^{6}\alpha_j(g_j - y_{ij}) \tag{28}$$

where α_j is the median of scaled values of statistical indicator, $j = 1$ for RMSE, MAE, MAPE, RMSPE and MBE ($j = 1, 2, 3, 4, 5$), -1 for r; g_j is a scaled value of the statistical indicator j for model i with larger GPI indicating a better performance.

We evaluated the model performance with Kling–Gupta Efficiency (KGE) [98] and Absolute Percentage Bias (APB; %) [99]. Mathematically, these metrics are stated as follows:

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\langle G^p\rangle}{\langle G^m\rangle} - 1\right)^2 + \left(\frac{CV_p}{CV_m}\right)^2} \tag{29}$$

$$APB = \frac{\sum_{i=1}^{n}((G^m - G^p) * 100)}{\sum_{i=1}^{n} G^m}, \tag{30}$$

where r is the correlation coefficient, CV is the coefficient of variation.

Additionally, the performance to predict the direction of movement was measured by a Directional Symmetry (DS) as follows:

$$DS = \frac{1}{n}\sum_{t=2}^{n} d_t \times 100\% \tag{31}$$

where:

$$d_t = \begin{cases} 1 & \text{if } (G_t^m - G_{t-1}^m)(G_t^p - G_{t-1}^m) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

We use the Diebold–Mariano (DM) and Harvey, Leybourne, and Newbold (HLN) test to evaluate the statistical significance of the models. When DM and HLN test have values greater or equal than 0, then the alternative models overcome the comparative models [100–102].

4. Results and discussion

The performance of the proposed CESN model for daily electricity demand prediction has been tested and verified through a comparative analysis against six benchmark models including SVR, MLP, XGB, XGB, DNN, and LGBM, see Fig. 7 for the Roderick substation (results on the other three substations: Rockle, Hemmant, and Carpendale are placed in Appendix C). In general, the proposed CESN model consistently produced the best performance based on the different evaluation metrics considered, detailed in Section 3.3, over the four study sites. SVR and DNN were the second-best models depending on performance indicators and study sites. The LGBM and XGB models generated the lowest predictive accuracy. We detail these results in this section.

Table 3 shows the correlation coefficient values for all considered models in daily electricity demand prediction.

The results obtained clearly show that the proposed CESN outperforms the competitive models over four study sites. For instance, CESN produced the highest correlation coefficient values $r \sim 0.926$, 0.925, 0.932 and 0.818 for each substation studied, considerably better than the obtained values of next best model SVR with $r \sim 0.909$, 0.920, 0.918 and 0.810 and far from the worst case, the LGBM with $r \sim 0.905$, 0.849, 0.866 and 0.815. The average error values from Table 3 also indicate a better predictive performance of CESN relative to the other models. Specifically, CESN predicted daily electricity demand at four study sites with $RMSE \sim 32.126$, 34.195, 17.688, and 3.050 MW and $MAE \sim 22.527$, 22.755, 12.002, and 2.259 MW, which are between

Table 3
Descriptive statistics of daily electricity demand G at four substations of Southeast Queensland where the proposed deep learning hybrid CESN model is implemented.

Study site	Predictive model	Model performance metrics		
		r	*RMSE (MW)*	*MAE (MW)*
Roderick	**CESN**	**0.926**	**32.126**	**22.527**
	SVR	0.909	35.293	24.642
	MLP	0.907	35.742	25.291
	XGB	0.883	39.835	28.097
	DNN	0.838	47.655	37.08
	LGB	0.905	56.679	47.286
Rocklea	**CESN**	**0.925**	**34.195**	**22.755**
	SVR	0.920	35.138	24.338
	MLP	0.917	35.748	24.883
	XGB	0.917	35.979	24.459
	DNN	0.912	36.892	23.704
	LGB	0.849	47.413	33.103
Hemmant	**CESN**	**0.932**	**17.688**	**12.002**
	SVR	0.918	19.244	12.571
	MLP	0.913	20.972	15.06
	XGB	0.902	21.648	14.995
	DNN	0.887	23.764	17.683
	LGB	0.866	30.222	22.962
Carpendale	**CESN**	**0.818**	**3.0503**	**2.2598**
	SVR	0.810	3.1108	2.3256
	MLP	0.812	3.0993	2.2899
	XGB	0.812	3.4756	2.5816
	DNN	0.812	3.1175	2.3355
	LGB	0.815	3.0865	2.3039

Table 4
The testing performance of the CNN integrated with ESN (CESN) model vs. SVR, MLP, XGB, DNN, and LGBM models as measured by skill score (SS), standard deviation of relative error ($STDRE$) and explained variance (E_{var}).

Study site	Predictive model	Model performance metrics		
		Skill Score (SS)	STDRE	E_{var}
Roderick	**CESN**	**75.44%**	**4.07%**	**0.856**
	SVR	70.36%	4.68%	0.826
	MLP	69.60%	4.52%	0.822
	XGB	62.24%	5.14%	0.779
	DNN	23.56%	5.90%	0.683
	LGB	45.96%	4.67%	0.814
Rocklea	**CESN**	**85.64%**	**8.53%**	**0.854**
	SVR	84.84%	8.64%	0.847
	MLP	84.31%	8.77%	0.841
	XGB	84.11%	9.21%	0.838
	DNN	72.40%	9.58%	0.830
	LGB	83.29%	12.26%	0.720
Hemmant	**CESN**	**86.85%**	**7.62%**	**0.864**
	SVR	84.43%	9.38%	0.842
	MLP	81.51%	7.95%	0.831
	XGB	80.30%	10.04%	0.813
	DNN	61.60%	11.44%	0.771
	LGB	76.26%	14.98%	0.685
Carpendale	**CESN**	**16.80%**	**7.93%**	**0.669**
	SVR	13.46%	9.13%	0.655
	MLP	14.10%	8.44%	0.658
	XGB	8.02%	8.00%	0.659
	DNN	14.81%	9.82%	0.655
	LGB	13.10%	9.46%	0.661

3 to 43% of RMSE and between 4 to 52% lower than the previous comparative models, the SVR and LGB respectively.

The frequency histograms illustrated in Fig. 8 offer the possibility of a better understanding of prediction error distributions for all models. In general, a right-skewed pattern with more absolute prediction errors ($|PE|$) close to zero indicates a better predictive capacity. The visual comparisons clearly show that, across the four study sites, the proposed CESN model yields the best performance, followed by the SVR. The MLP and XGB models were in the moderate group while LGBM and DNN produced the worst performance. These analyses were also confirmed by the notched box plots and empirical cumulative distribution function (ECDF) of $|PE|$ in Fig. 9 and Fig. 10, respectively. These figures clearly show that the boxes of CESN and SVR represent lower medians, interquartile ranges, and outliers meaning the distributions of $|PE|$ values were narrower, with the majority of error values closer to zero. By contrast, the ECDF of the LGBM is shifted right the furthest towards higher values, indicating that it provides the most prediction errors.

The prediction capacity for daily electricity demand of the proposed CESN model and comparative candidates was further measured by evaluating its skill score (SS), standard deviation of relative error ($STDRE$) and explained variance (E_{var}). The higher SS and E_{var} values or lower STDRE values demonstrate a better performance. From Table 4, it is evident that the CESN yielded the best performance for all sites with $SS \sim 86.85\%$ (Hemmant substation), $STDRE \sim 4.07\%$ (Roderick substation), and $E_{var} \sim 0.864$ (Hemmant substation). The SVR model generally generated the second-best predictive performance, except for the Carpendale substation. The worst performances are DNN (Roderick and Carpendale substation) and LGBM (Rocklea and Hemmant substation).

Similarly, for the statistical performance criteria including $RMSPE$, $MAPE$, WI, NS and LM, the advantages of the proposed CESN model over other benchmark models can also be seen in Tables 5 and 6. The CESN model generally produces the lowest values of RMSPE and MAPE, and the highest values of $WI, NS\ and\ LM$, indicating its best performance for daily electricity demand prediction. Consistent results were also found for the second-best and worst performance, i.e., SVR and LGBM models, respectively, over the Roderick, Rocklea and Hemmant sites. However, the findings are slightly different at the

Table 5
The geographic comparison of the accuracy of the CNN integrated with ESN (CESN) model vs. other comparative models in terms of the relative errors ($RMSPE, \%$) and ($MAPE, \%$) computed within the test sites. Note that the best model is boldfaced.

Study site	Predictive model	Model performance metrics	
		RMSPE	MAPE
Roderick	**CESN**	**5.86%**	**4.04%**
	SVR	6.43%	4.47%
	MLP	6.52%	4.55%
	XGB	7.26%	5.07%
	DNN	8.69%	6.95%
	LGB	10.33%	8.37%
Rocklea	**CESN**	**9.15%**	**6.64%**
	SVR	9.40%	7.02%
	MLP	9.57%	7.28%
	XGB	9.63%	7.26%
	DNN	9.87%	6.95%
	LGB	12.69%	10.04%
Hemmant	**CESN**	**8.52%**	**6.06%**
	SVR	9.27%	6.70%
	MLP	10.11%	7.46%
	XGB	10.43%	8.12%
	DNN	11.45%	10.15%
	LGB	14.56%	13.72%
Carpendale	**CESN**	**10.33%**	**8.02%**
	SVR	10.54%	8.15%
	MLP	10.50%	8.07%
	XGB	11.77%	8.61%
	DNN	10.56%	8.39%
	LGB	10.45%	8.20%

Carpendale site, varying depending on the criteria considered. The XGB model yields the highest values of $WI \approx 0.807$ compared to $WI \approx 0.798$ from the CESN model as the second-best performance. By contrast, based on the RMSPE, MAPE, NS and LM criteria, the LGBM model produced the second-best performance while XGB was the worst model.

Table 7 provides the geographic comparison of the accuracy of all considered models, using the uncertainty at 95% ($U95$), t-statistics (t) and Thiel's inequality coefficient (TIC). The $U95$ is an efficient quantitative criterion for selecting the optimal predictive model among

Table 6
The performance of the CNN integrated with ESN (CESN) model vs. SVR, MLP, XGB, DNN, and LGBM models using the Willmott's Index (WI), Nash–Sutcliffe Coefficient (NS) and the Legates & McCabe's (LM) Index of Agreement. Note that the best model is boldfaced.

Study site	Predictive model	Model performance metrics		
		WI	*NS*	*LM*
Roderick	**CESN**	**0.916**	**0.856**	**0.651**
	SVR	0.898	0.826	0.618
	MLP	0.894	0.822	0.608
	XGB	0.863	0.779	0.565
	DNN	0.735	0.683	0.426
	LGB	0.826	0.645	0.268
Rocklea	**CESN**	**0.937**	**0.854**	**0.712**
	SVR	0.932	0.846	0.692
	MLP	0.927	0.841	0.685
	XGB	0.923	0.838	0.690
	DNN	0.926	0.830	0.700
	LGB	0.858	0.720	0.581
Hemmant	**CESN**	**0.935**	**0.862**	**0.700**
	SVR	0.919	0.837	0.685
	MLP	0.916	0.810	0.623
	XGB	0.890	0.797	0.625
	DNN	0.835	0.755	0.557
	LGB	0.637	0.629	0.425
Carpendale	**CESN**	0.798	**0.669**	**0.455**
	SVR	0.789	0.655	0.439
	MLP	0.791	0.658	0.448
	XGB	**0.807**	0.605	0.378
	DNN	0.765	0.654	0.437
	LGB	0.781	0.661	0.445

Table 7
The comparison of the prediction accuracy of the CNN integrated with ESN (CESN) model vs. other comparative models in terms of the uncertainty at 95% (U95), t-statistics (t) and Theil's inequality coefficient (TIC) during testing. Note that the best model is boldfaced.

Study site	Predictive model	Model performance metrics		
		U95	*t*	*TIC*
Roderick	**CESN**	**89.104**	**−0.292**	**0.029**
	SVR	97.875	−0.523	0.032
	MLP	99.130	−0.371	0.032
	XGB	110.460	−0.648	0.036
	DNN	132.180	0.325	0.043
	LGB	132.240	−22.585	0.053
Rocklea	**CESN**	**94.828**	**0.185**	**0.045**
	SVR	97.331	−1.416	0.046
	MLP	99.138	0.538	0.047
	XGB	99.795	−0.571	0.047
	DNN	102.300	0.609	0.048
	LGB	131.500	−0.403	0.062
Hemmant	**CESN**	**48.882**	**−2.334**	**0.042**
	SVR	52.923	3.567	0.045
	MLP	56.264	−7.327	0.050
	XGB	58.586	6.207	0.050
	DNN	64.533	5.727	0.055
	LGB	79.051	10.154	0.069
Carpendale	**CESN**	**8.457**	**0.014**	**0.051**
	SVR	8.629	0.833	0.052
	MLP	8.592	0.902	0.052
	XGB	**9.128**	−9.728	0.060
	DNN	8.636	1.355	0.052
	LGB	8.557	0.885	**0.051**

different models by offering useful information on a particular model's deviance. The statistical t test indicates the desired model while the TIC is useful for evaluating the generalization ability of the model. Values of $U95, t$ *and* TIC approaching zero signify a high prediction accuracy. The results from Table 7 also indicated the best performance of CESN against the compared models.

Table 8
Evaluation of the CNN integrated with ESN (CESN) model against comparison models in terms of the Diebold–Mariano (DM) test statistic. The column of the table is compared with the rows, and if the result is positive, the model in the rows outperforms the one in the column; on the contrary, if it is negative, then the one in the column is superior.

	CESN	DNN	LGB	SVR	XGB	MLP
CESN		**0.678**	**1.469**	**1.048**	**3.669**	**2.072**
DNN			0.786	0.584	3.253	0.241
LGB				−0.126	2.717	−0.373
SVR					3.321	−0.241
XGB						3.311

Table 9
Evaluation of the CNN integrated with ESN (CESN) model against comparison models in terms of the Harvey–Leybourne–Newbold (HLN) test statistic. The column of the table is compared with the rows, and if the result is positive, the model in the rows outperforms the one in the column; on the contrary, if it is negative, then the one in the column is superior.

	CESN	DNN	LGB	SVR	XGB	MLP
CESN		**0.712**	**1.542**	**1.100**	**3.852**	**2.175**
DNN			0.825	0.614	3.415	0.253
LGB				−0.132	2.852	−0.392
SVR					3.487	−0.253
XGB						−3.476

The proposed CESN and the relative five benchmarked models' performances were also evaluated using the absolute percentage bias (APB), Kling–Gupta efficiency (KGE), Global performance indicator (GPI), and combined performance index (CPI). The results represented in Fig. 11 clearly show that for every study site, the CESN model generated lower APB and higher KGE values compared to other models. The GPI and CPI values illustrated in Fig. 12 and Fig. 13 also supported that CESN was able to produce daily electricity demand prediction better than SVR, MLP, XGB, XGB, DNN, and LGBM models. Furthermore, according to the CPI (Fig. 13), DNN was the second-best model for predicting daily electricity demand over Hemmant and Carpendale sites while MLP produced the worst performance over Roderick and Hemmant sites. Fig. 14 plotted the Taylor diagram for better understanding of how well distributions of predicted and observed data match each other in terms of standard deviation, correlation, and root-mean-square difference ($RMSD$). Over the four study sites, compared to other models, CESN has a higher correlation meaning better agreement between simulated and observed data. Furthermore, the simulated data from the CESN model were also closer to the standard deviation (red line) of the observations and $RMSD$ line showing the higher quality of the simulation process. This statistical summary revealed that the performance of DNN was far away from other models.

Finally, the performance was evaluated for each pair of models using the Diebold–Mariano (DM), Harvey–Leybourne–Newbold (HLN) test statistics (Tables 8 and 9) and the ratio of root mean square error ($RMSE_r$) (Table 10). The interpretation of these performance indicators is based on a comparison between columns and rows. When the result is positive (Tables 8 and 9) or greater than 1 (Table 10), the model placed in the rows exceeded the model in the columns, and conversely. The findings consistently indicated that the proposed CESN model is able to generate the daily electricity demand prediction with the highest level of accuracy. Similarly, Fig. 15 shows that deep hybrid CESN model DS (i.e., directional prediction accuracy) is more accurate than others, with an average of 84.59%. Congruently with earlier findings, DM, HLN, and DS test provide consistent results, which indicate that deep hybrid CESN predicts daily electricity demand G more accurately than other models.

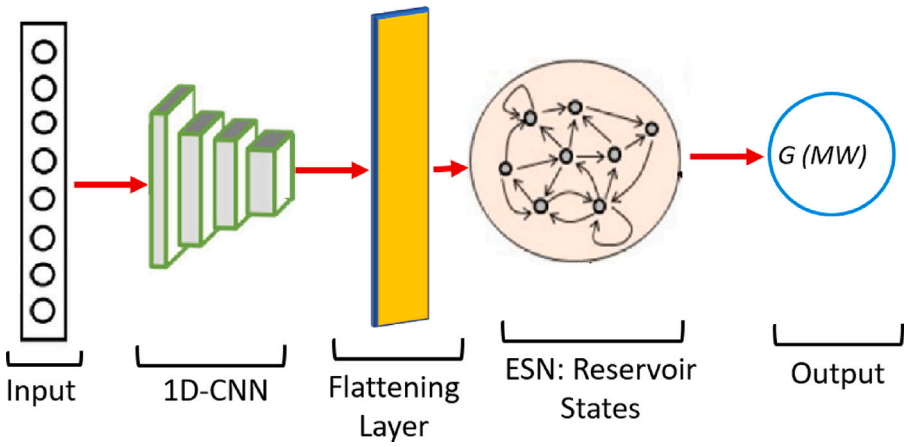

Fig. B.16. CESN architecture.

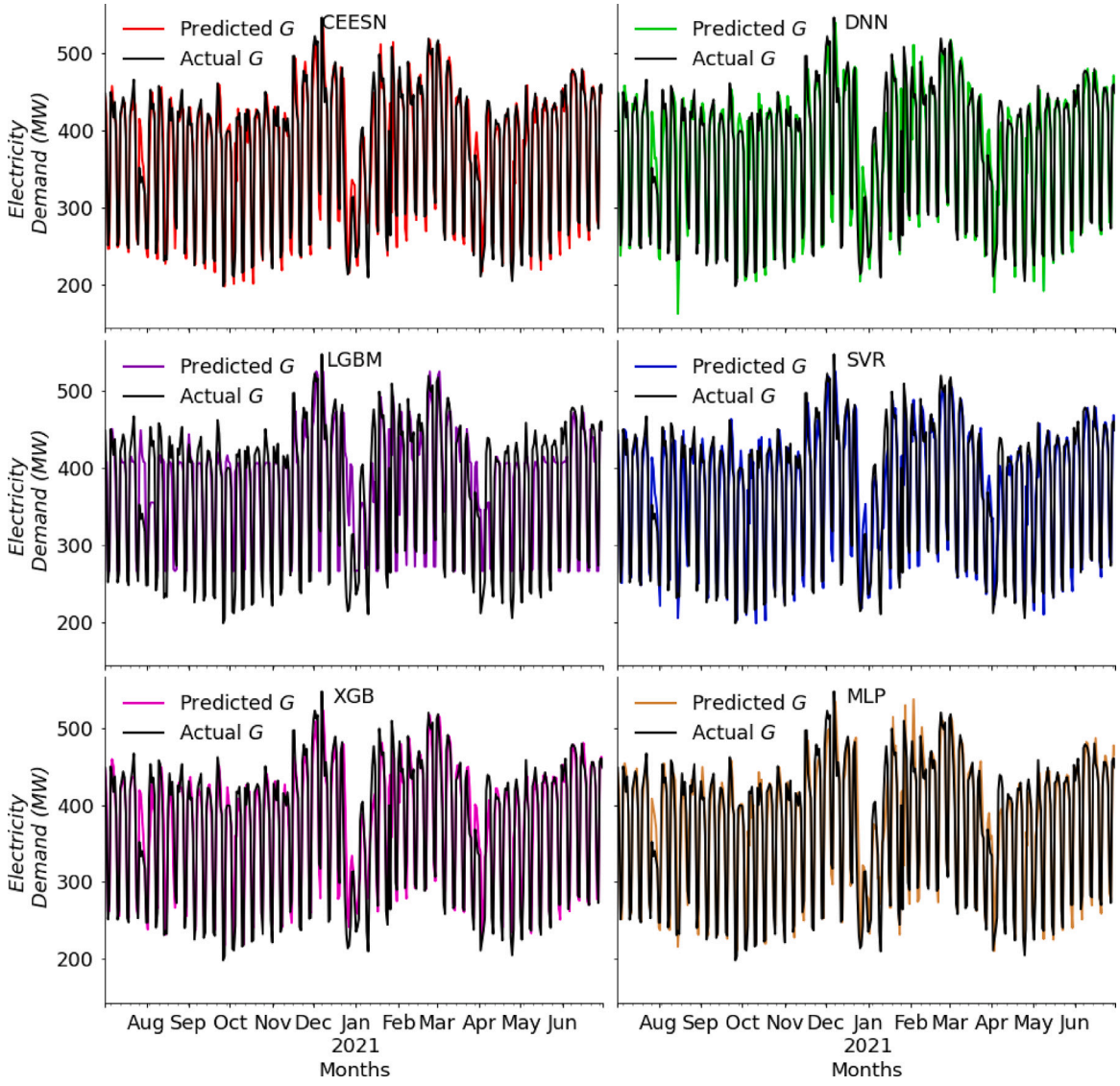

Fig. C.17. Comparison of the proposed CESN model for daily electricity demand prediction analysis with six benchmark models including SVR, MLP, XGB, DNN, and LGBM in Rocklea substation.

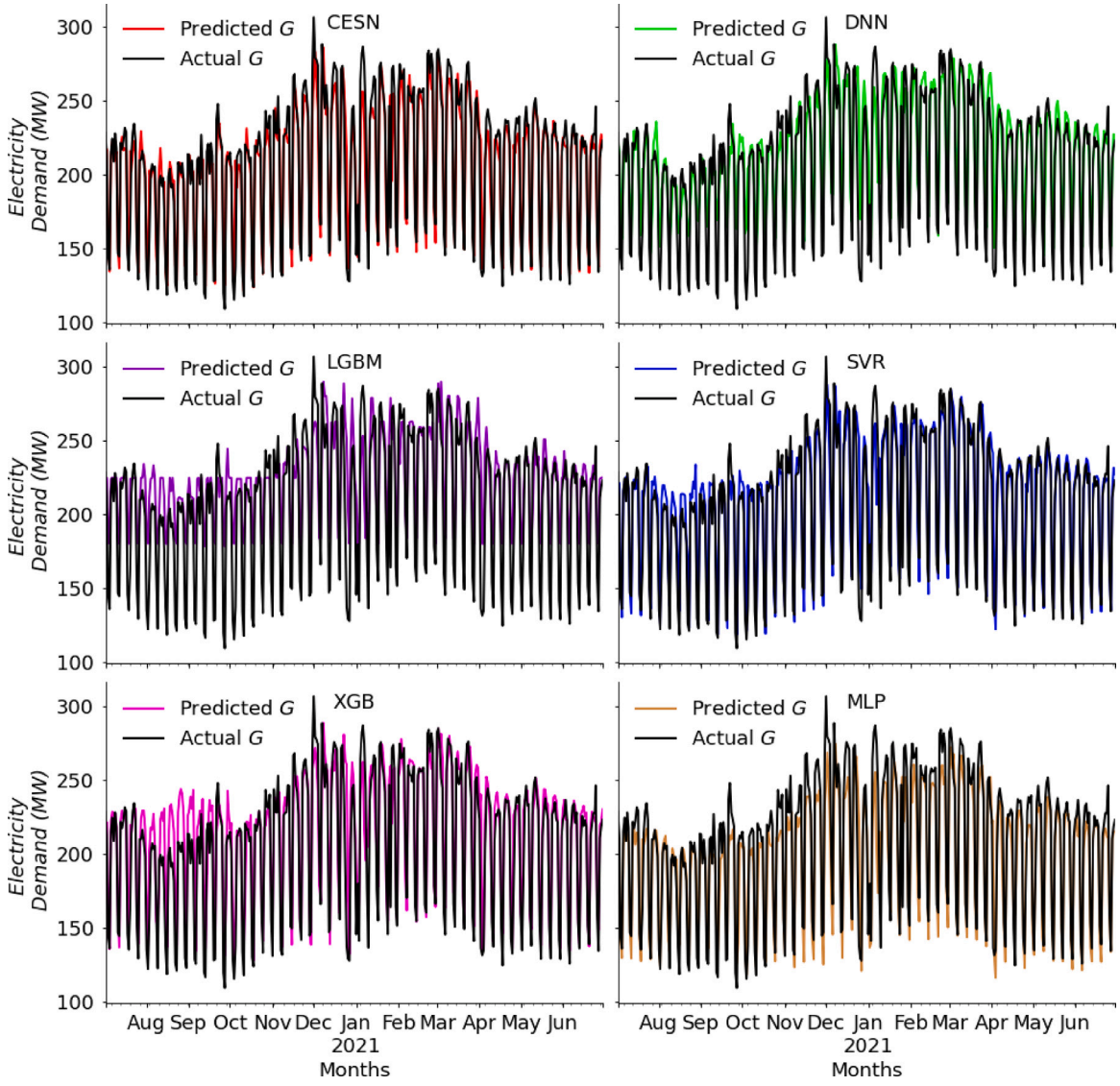

Fig. C.18. Comparison of the proposed CESN model for daily electricity demand prediction analysis with six benchmark models including SVR, MLP, XGB, DNN, and LGBM in Hemmant substation.

Table 10
The performance of the CNN integrated with ESN (CESN) model with comparative benchmark models in the test period measured by the ratio of root mean square error ($RMSE_r$). The column of the table is compared with the rows, and if the result is > 1, the model in the row outperforms the one in the column; on the contrary, if it is < 1, then the one in the column is superior.

	CESN	DNN	LGB	SVR	XGB	MLP
CESN		**1.024**	**1.045**	**1.040**	**1.298**	**1.032**
DNN			1.020	1.016	1.268	1.008
LGB				0.996	1.243	0.988
SVR					1.248	0.993
XGB						0.795

5. Conclusions, limitations, and future research directions

An artificial intelligence technique based on the deep hybrid CESN model, which integrates a convolutional neural network (CNN) model with the Echo State Network (ESN) process, has been introduced in this study for the prediction of daily electricity demand data (G). Five benchmark models (support vector regression (SVR), multi layer perceptron (MLP), Xtreme gradient boosting (XGB), deep neural network (DNN), and Light Gradient Boosting decision tree model (LGBM)) were also developed to evaluate the prediction accuracy of the CESN model, based on data collected for four stations in southeast Queensland, Australia. Based on the results obtained, it has been shown that the proposed CESN model was able to obtain better G prediction performance not only than the benchmark models developed in this research study, but also when it is compared with previous works [2,20,103] on G prediction. The main reason behind the better performance of the proposed CESN model is that, apart from learning the load trend characterization, CESN integrates the advantage of the contextual features yielded from the timestamp and the temporal information in the historical electricity demand data to achieve better predictive accuracy.

Despite the performance of the proposed CESN model has been consistently demonstrated using different statistical tests and criteria over the four study studies, we have also shown that it is different from the other comparative models based on ML. The predictive capacity of SVR, MLP, XGB, XGB, DNN and LGBM models varied depending on the performance indicators and study sites. While the inconsistent results emphasize the importance of applying multiple performance

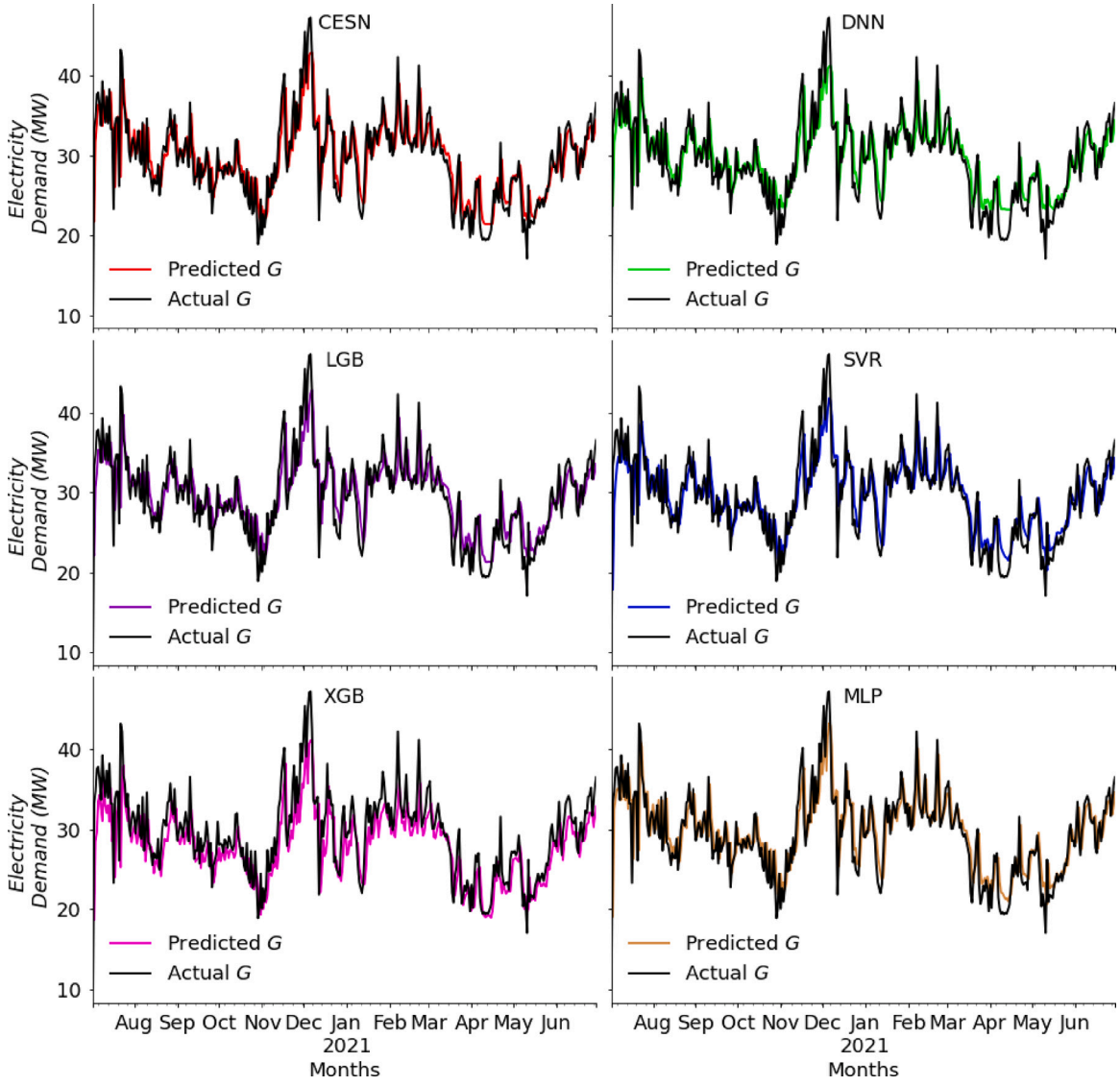

Fig. C.19. Comparison of the proposed CESN model for daily electricity demand prediction analysis with six benchmark models including SVR, MLP, XGB, DNN, and LGBM in Carpendale substation.

evaluation matrices, further investigation of the pre-processing data, patterns of data at study sites, uncertainty prediction analysis, and model configurations of all considered models are required.

In general, we have shown that the proposed hybrid CESN model can deliver better performance in electricity demand prediction than classical ML methods. The proposed CESN model can also support electricity markets and provide a precise decision support tool valuable to improve policies applied and decision making processes.

Regarding limitations and future lines of research, note that noise information can significantly affect the performance of predictive models. Time-series denoising techniques, for example, multivariate empirical mode decomposition (MEMD) [104] or variational mode decomposition (VMD) [105], can be applied to enhance prediction accuracy. The original electricity demand time series can be decomposed completely into several components, where the noise term can be removed. Another factor that can also affect the model performance is the seasonal pattern, which is a common component in the electricity demand time series. Fast Fourier Transform (FFT) [106] might be an effective method to address the seasonal problem. The FFT technique measures the seasonal length and then computes the seasonal indexes by the seasonal adjustment method to complete the seasonal factor eliminating process.

Uncertainty analysis of short-term electricity demand prediction is also an essential step because of the intermittent, unpredictability and variations of the distributed generations and loads in microgrid systems. The interval prediction method, such as least square support vector regression (LSSVR) [107], can be applied to provide information related to prediction uncertainty. This uncertainty analysis will provide the confidence levels of the proposed predictive models in capturing both the future tendency of the electricity demand and the best coverage probability and prediction interval width.

Finally, incorporating causality in electricity demand prediction is also suggested for future research to improve and maintain prediction at a high level of prediction accuracy in different time periods. Additional information may include, but is not limited to, economic factors, and weather condition forecasting. However, these additional inputs might have different effects on short-term and long-term electricity prediction. In addition, when incorporating various input variables, a feature selection process can be performed to select the most influential predictors and eventually form hybrid models.

Table A.11
List of Acronyms.

Term	Acronyms
Artificial Intelligence	AI
Artificial Neural Network	ANN
Autoregressive Integrated Moving Average	ARIMA
Auto-Regressive Moving Average	ARMA
Convolutional Neural Network	CNN
Decision Trees	DT
Deep Belief Networks	DBN
Deep Learning	DL
Deep Neural Network	DNN
Deep Recurrent Neural Network	DRNN
Empirical Mode Decomposition	EMD
Empirical Risk Minimization	ERM
Extreme Learning Machine	ELM
Gated Recurrent Unit	GRU
Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise	ICEEMDAN
K-Nearest Neighbors	KNN
Light Gradient Boosting decision tree	LGBM
Long Short-Term Memory	LSTM
Machine Learning	ML
Multi-Layer Perceptron model	MLP
Partial Autocorrelation Function	PACF
Prediction Interval	PI
Particle Swarm Optimized	PSO
Radial Basis Function	RBF
Recurrent Neural Networks	RNN
Rectified Linear Unit	ReLU
SHapley Additive exPlanations	SHAP
Structural Risk Minimization	SRM
Support Vector Regression	SVR
South-East Queensland	SEQ
Xtreme Gradient Boosting	XGB

Table B.12
Parameters of the CESN architecture.

	Roderick	Rocklea	Hemmant	Carpendale
1D CNN Block (3 CNN Layers)				
Train size *(Seq. length, T. steps, Feats.)*	(2922, 1, 9)	(2922, 1, 8)	(2922, 1, 10)	(2922, 1, 5)
Test size *(Seq. length, T. steps, Feats.)*	(365, 1, 9)	(365, 1, 8)	(365, 1, 10)	(365, 1, 5)
Validation size *(Seq. length, T. steps, Feats.)*	(730, 1, 8)	(730, 1, 10)	(730, 1, 5)	
CNN input shape *(T. steps, Feats)*	(1, 9)	(1, 8)	(1, 10)	(1, 5)
CNN Layer1: Filter	100	50	100	50
Kernel Size	10	10	10	10
kernel_initializer	glorot_unif.	glorot_unif.	glorot_unif.	glorot_unif.
Padding	Same	Same	Same	Same
Pooling Layer	MaxPool 1	MaxPool 1	MaxPool 1	MaxPool 1
Activation	Relu	Relu	Relu	Relu
CNN Layer2	60	50	40	40
Kernel Size	1	1	1	1
kernel_initializer	glorot_unif.	glorot_unif.	glorot_unif.	glorot_unif.
Padding	same	same	same	same
Pooling Layer	MaxPool 1	MaxPool 1	MaxPool 1	MaxPool 1
Activation	Relu	Relu	Relu	Relu
CNN Layer3	5	30	10	5
Kernel Size	1	1	1	1
kernel_initializer	glorot_unif.	glorot_unif.	glorot_unif.	glorot_unif.
Padding	same	same	same	same
Pooling Layer	MaxPool 1	MaxPool 1	MaxPool 1	MaxPool 1
Activation	Relu	Relu	Relu	Relu
Final Layer	Flattening l.	Flattening l.	Flattening l.	Flattening l.
Final CNN Layer Output	(2922, 5)	(2922, 30)	(2922, 10)	(2922, 10)
Solver	Adam	Adam	Adam	Adam
Epochs	1000	1000	1000	1000
Batch size	5	10	5	5
ESN Block				
Input Layer	(2922, 5)	(2922, 30)	(2922, 10)	(2922, 10)
Spectral Radius	0, 6	0, 55	0, 7	0, 5
Connectivity rate	0, 01	0, 03	0, 01	0, 01
Output Layer	(365, 1)	(365, 1)	(365, 1)	(365, 1)

CRediT authorship contribution statement

Sujan Ghimire: Data curation, Methodology, Software, Validation, Visualization, Writing – reviewing, Conceptualization, Investigation. **Thong Nguyen-Huy:** Writing – review & editing, Investigation, Methodology, Reviewing. **Mohanad S. AL-Musaylh:** Writing – review & editing, Investigation, Methodology, Reviewing. **Ravinesh C. Deo:** Methodology, Visualization, Writing – reviewing, Investigation, Conceptualization, Supervision, Reviewing. **David Casillas-Pérez:** Writing, Visualization, Editing, Methodology, Conceptualization, Investigation, Reviewing. **Sancho Salcedo-Sanz:** Writing – editing, Visualization, Conceptualization, Methodology, Investigation, Reviewing, Supervision.

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 were acquired from ENERGEX (https://www.energex.com.au).

Acknowledgments

The authors thank the data providers, all the reviewers and the Editor for their thoughtful comments, suggestions and the review process. Partial support of this study is through the project PID2020-115454GB-C21 of the Spanish Ministry of Science and Innovation (MICINN).

Appendix A. Acronyms

Table A.11 provides the acronyms used in this paper:

Appendix B. CESN architecture

Fig. B.16 shows the architecture of our proposal CESN divided into blocks.

Table B.12 details the parameters of each block in order to be implemented.

Appendix C. Further results

Figs. C.17, C.19 and C.18 show the prediction results of our proposed CELM method together with the other six methods of the comparison, see Section 4.

References

[1] Hu Y, Li J, Hong M, Ren J, Man Y. Industrial artificial intelligence based energy management system: Integrated framework for electricity load forecasting and fault prediction. Energy 2022;244:123195.
[2] Al-Musaylh MS, Deo RC, Li Y, Adamowski JF. Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting. Appl Energy 2018;217:422–39.
[3] Castillo VZ, De Boer H-S, Muñoz RM, Gernaat DE, Benders R, van Vuuren D. Future global electricity demand load curves. Energy 2022;258:124741.
[4] Finkel A. Independent review into the future security of the national electricity market: preliminary report. Australia: Department of the Environment and Energy; 2016.
[5] Yang J, Yang Z, Duan Y. Optimal capacity and operation strategy of a solar-wind hybrid renewable energy system. Energy Convers Manage 2021;244:114519.
[6] Lipu MH, Miah MS, Ansari S, Hannan M, Hasan K, Sarker MR, et al. Data-driven hybrid approaches for renewable power prediction toward grid decarbonization: Applications, issues and suggestions. J Clean Prod 2021;328:129476.
[7] Chen Y, Zhang L, Xu P, Di Gangi A. Electricity demand response schemes in China: Pilot study and future outlook. Energy 2021;224:120042.
[8] Velasquez CE, Zocatelli M, Estanislau FB, Castro VF. Analysis of time series models for Brazilian electricity demand forecasting. Energy 2022;247:123483.
[9] Zamanipour B, Ghadaksaz H, Keppo I, Saboohi Y. Electricity supply and demand dynamics in Iran considering climate change-induced stresses. Energy 2023;263:126118.
[10] Adeoye O, Spataru C. Modelling and forecasting hourly electricity demand in west African countries. Appl Energy 2019;242:311–33.
[11] Clements AE, Hurn A, Li Z. Forecasting day-ahead electricity load using a multiple equation time series approach. European J Oper Res 2016;251(2):522–30.
[12] McLoughlin F, Duffy A, Conlon M. Evaluation of time series techniques to characterise domestic electricity demand. Energy 2013;50:120–30.
[13] Hussain A, Rahman M, Memon JA. Forecasting electricity consumption in Pakistan: The way forward. Energy Policy 2016;90:73–80.
[14] Sarkodie SA. Estimating Ghana's electricity consumption by 2030: An ARIMA forecast. Energy Sources Part B: Econ Plan Policy 2017;12(10):936–44.
[15] Lee Y-S, Tong L-I. Forecasting nonlinear time series of energy consumption using a hybrid dynamic model. Appl Energy 2012;94:251–6.
[16] Al-Musaylh MS, Deo RC, Adamowski JF, Li Y. Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF reanalysis atmospheric predictors in southeast Queensland, Australia. Renew Sustain Energy Rev 2019;113:109293.
[17] Mohanad SA-M, Ravinesh CD, Yan L. Particle swarm optimized–support vector regression hybrid model for daily horizon electricity demand forecasting using climate dataset. In: E3S web of conferences, Vol. 64. EDP Sciences; 2018, p. 08001.
[18] Gao T, Niu D, Ji Z, Sun L. Mid-term electricity demand forecasting using improved variational mode decomposition and extreme learning machine optimized by sparrow search algorithm. Energy 2022;261:125328.
[19] Salcedo-Sanz S, Ghamisi P, Piles M, Werner M, Cuadra L, Moreno-Martínez A, et al. Machine learning information fusion in earth observation: A comprehensive review of methods, applications and data sources. Inf Fusion 2020;63:256–72.
[20] Al-Musaylh MS, Deo RC, Li Y. Electrical energy demand forecasting model development and evaluation with maximum overlap discrete wavelet transform-online sequential extreme learning machines algorithms. Energies 2020;13(9):2307.
[21] Ghimire S, Deo RC, Wang H, Al-Musaylh MS, Casillas-Pérez D, Salcedo-Sanz S. Stacked LSTM sequence-to-sequence autoencoder with feature selection for daily solar radiation prediction: A review and new modeling results. Energies 2022;15(3):1061.
[22] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. Improved complete ensemble empirical mode decomposition with adaptive noise deep residual model for short-term multi-step solar radiation prediction. Renew Energy 2022;190:408–24.
[23] Kumari P, Toshniwal D. Deep learning models for solar irradiance forecasting: A comprehensive review. J Clean Prod 2021;318:128566.
[24] Ghimire S, Nguyen-Huy T, Prasad R, Deo RC, Casillas-Pérez D, Salcedo-Sanz S, et al. Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction. Cogn Comput 2022;1–27.
[25] Jayasinghe WLP, Deo RC, Ghahramani A, Ghimire S, Raj N. Deep multi-stage reference evapotranspiration forecasting model: Multivariate empirical mode decomposition integrated with the Boruta-random forest algorithm. IEEE Access 2021;9:166695–708.
[26] Alkhayat G, Mehmood R. A review and taxonomy of wind and solar energy forecasting methods based on deep learning. Energy AI 2021;4:100060.
[27] Jayasinghe WLP, Deo RC, Ghahramani A, Ghimire S, Raj N. Development and evaluation of hybrid deep learning long short-term memory network model for pan evaporation estimation trained with satellite and ground-based data. J Hydrol 2022;607:127534.
[28] Dedinec A, Filiposka S, Dedinec A, Kocarev L. Deep belief network based electricity load forecasting: An analysis of Macedonian case. Energy 2016;115:1688–700.
[29] Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans Smart Grid 2017;10(1):841–51.
[30] Ryu S, Noh J, Kim H. Deep neural network based demand side short term load forecasting. Energies 2016;10(1):3.
[31] Shi H, Xu M, Li R. Deep learning for household load forecasting—A novel pooling deep RNN. IEEE Trans Smart Grid 2017;9(5):5271–80.
[32] Liu N, Tang Q, Zhang J, Fan W, Liu J. A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. Appl Energy 2014;129:336–45.
[33] Ghofrani M, Ghayekhloo M, Arabali A, Ghayekhloo A. A hybrid short-term load forecasting with a new input selection framework. Energy 2015;81:777–86.
[34] Fan G-F, Peng L-L, Hong W-C, Sun F. Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. Neurocomputing 2016;173:958–70.

[35] Qiu X, Ren Y, Suganthan PN, Amaratunga GA. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. Appl Soft Comput 2017;54:246–55.

[36] Kim T-Y, Cho S-B. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019;182:72–81.

[37] Wang L, Su Z, Qiao J, Deng F. A pseudo-inverse decomposition-based self-organizing modular echo state network for time series prediction. Appl Soft Comput 2022;116:108317.

[38] Roberts C, Lara JD, Henriquez-Auba R, Bossart M, Anantharaman R, Rackauckas C, et al. Continuous-time echo state networks for predicting power system dynamics. Electr Power Syst Res 2022;212:108562.

[39] Li Y, Li Y. Predicting chaotic time series and replicating chaotic attractors based on two novel echo state network models. Neurocomputing 2022;491:321–32.

[40] Lukoševičius M. A practical guide to applying echo state networks. In: Neural networks: tricks of the trade. Springer; 2012, p. 659–86.

[41] Ceni A, Ashwin P, Livi L. Interpreting recurrent neural networks behaviour via excitable network attractors. Cogn Comput 2020;12(2):330–56.

[42] Han Z, Zhao J, Leung H, Ma KF, Wang W. A review of deep learning models for time series prediction. IEEE Sens J 2019;21(6):7833–48.

[43] Ju Y, Sun G, Chen Q, Zhang M, Zhu H, Rehman MU. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. IEEE Access 2019;7:28309–18.

[44] Ghimire S, Yaseen ZM, Farooque AA, Deo RC, Zhang J, Tao X. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. Sci Rep 2021;11(1):1–26.

[45] Ghimire S, Nguyen-Huy T, Deo RC, Casillas-Perez D, Salcedo-Sanz S. Efficient daily solar radiation prediction with deep learning 4-phase convolutional neural network, dual stage stacked regression and support vector machine CNN-REGST hybrid model. Sustain Mater Technol 2022;32:e00429.

[46] Hoseinzade E, Haratizadeh S, Khoeini A. U-cnnpred: A universal CNN-based predictor for stock markets. 2019, arXiv preprint arXiv:1911.12540.

[47] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S, Sharma E, Ali M. Deep learning CNN-LSTM-MLP hybrid fusion model for feature optimizations and daily solar radiation prediction. Measurement 2022;111759.

[48] Ma Q, Shen L, Chen W, Wang J, Wei J, Yu Z. Functional echo state network for time series classification. Inform Sci 2016;373:1–20.

[49] Wang H, Liu Y, Wang D, Luo Y, Tong C, Lv Z. Discriminative and regularized echo state network for time series classification. Pattern Recognit 2022;108811.

[50] Wang H, Liu Y, Lu P, Luo Y, Wang D, Xu X. Echo state network with logistic mapping and bias dropout for time series prediction. Neurocomputing 2022;489:196–210.

[51] Büsing L, Schrauwen B, Legenstein R. Connectivity, dynamics, and memory in reservoir computing with binary and analog neurons. Neural Comput 2010;22(5):1272–311.

[52] Yang X, Liu Q, Liu X, Xue J. An improved deep echo state network inspired by tissue-like P system forecasting for non-stationary time series. J Membr Comput 2022;1–10.

[53] Verstraeten D, Schrauwen B, d'Haene M, Stroobandt D. An experimental unification of reservoir computing methods. Neural Netw 2007;20(3):391–403.

[54] Wang L, Hu H, Ai X-Y, Liu H. Effective electricity energy consumption forecasting using echo state network improved by differential evolution algorithm. Energy 2018;153:801–15.

[55] Xu X, Niu D, Fu M, Xia H, Wu H. A multi time scale wind power forecasting model of a chaotic echo state network based on a hybrid algorithm of particle swarm optimization and Tabu search. Energies 2015;8(11):12388–408.

[56] Zhong S, Xie X, Lin L, Wang F. Genetic algorithm optimized double-reservoir echo state network for multi-regime time series prediction. Neurocomputing 2017;238:191–204.

[57] Bergstra J, Yamins D, Cox DD, et al. Hyperopt: A Python library for optimizing the hyperparameters of machine learning algorithms. In: Proceedings of the 12th python in science conference, Vol. 13. Citeseer; 2013, p. 20.

[58] Armeniakos G, Zervakis G, Soudris D, Henkel J. Hardware approximate techniques for deep neural network accelerators: A survey. ACM Comput Surv 2022.

[59] Kabir HD, Abdar M, Khosravi A, Jalali SMJ, Atiya AF, Nahavandi S, et al. Spinalnet: Deep neural network with gradual input. IEEE Trans Artif Intell 2022.

[60] Jiang W. Internet traffic prediction with deep neural networks. Internet Technol Lett 2022;5(2):e314.

[61] Bas E, Egrioglu E, Kolemen E. Training simple recurrent deep artificial neural network for forecasting using particle swarm optimization. Granul Comput 2022;7(2):411–20.

[62] Gasparin A, Lukovic S, Alippi C. Deep learning for time series forecasting: The electric load case. CAAI Trans Intell Technol 2022;7(1):1–25.

[63] Wu R, Kim T, Tian DJ, Bianchi A, Xu D. {*DnD*}: A {*Cross − Architecture*} deep neural network decompiler. In: 31st USENIX security symposium. 2022, p. 2135–52.

[64] Castillo-Botón C, Casillas-Pérez D, Casanova-Mateo C, Ghimire S, Cerro-Prada E, Gutierrez P, et al. Machine learning regression and classification methods for fog events prediction. Atmos Res 2022;272:106157.

[65] Ghimire S, Bhandari B, Casillas-Pérez D, Deo RC, Salcedo-Sanz S. Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in queensland, Australia. Eng Appl Artif Intell 2022;112:104860.

[66] Salcedo-Sanz S, Rojo-Álvarez JL, Martínez-Ramón M, Camps-Valls G. Support vector machines in engineering: An overview. Wiley Interdisc Rev: Data Min Knowl Discov 2014;4(3):234–67.

[67] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM-KDD international conference on knowledge discovery and data mining. 2016, p. 785–94.

[68] Joshi A, Kumar V, Warrior H. Modeling the sea-surface pCO2 of the central bay of Bengal region using machine learning algorithms. Ocean Model 2022;102094.

[69] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. Lightgbm: A highly efficient gradient boosting decision tree. Adv Neural Inform Process Syst (NIPS) 2017;30.

[70] Massaoudi M, Refaat SS, Chihi I, Trabelsi M, Oueslati FS, Abu-Rub H. A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for short-term load forecasting. Energy 2021;214:118874.

[71] Jiajun H, Chuanjin Y, Yongle L, Huoyue X. Ultra-short term wind prediction with wavelet transform, deep belief network and ensemble learning. Energy Convers Manage 2020;205:112418.

[72] Wang C, Yu K, Qu F, Bu J, Han S, Zhang K. Spaceborne GNSS-R wind speed retrieval using machine learning methods. Remote Sens 2022;14(14):3507.

[73] Wang Y, Zou R, Liu F, Zhang L, Liu Q. A review of wind speed and wind power forecasting with deep neural networks. Appl Energy 2021;304:117766.

[74] Khochare J, Rathod J, Joshi C, Laveti RN. A short-term wind forecasting framework using ensemble learning for indian weather stations. In: 2020 IEEE international conference for innovation in technology. IEEE; 2020, p. 1–7.

[75] Maliyeckel MB, Sai BC, Naveen J. A comparative study of LGBM-SVR hybrid machine learning model for rainfall prediction. In: 2021 12th international conference on computing communication and networking technologies. IEEE; 2021, p. 1–7.

[76] Nwokolo SC, Obiwulu AU, Ogbulezie JC, Amadi SO. Hybridization of statistical machine learning and numerical models for improving beam, diffuse and global solar radiation prediction. Clean Eng Technol 2022;9:100529.

[77] Nziyumva E, Hu R, Hsu C-Y, Niyogisubizo J. Electrical load forecasting using hybrid of extreme gradient boosting and light gradient boosting machine. In: The international conference on image, vision and intelligent systems. Springer; 2022, p. 1083–93.

[78] Almutairi K, Almutairi M, Harb K, Marey O. Optimal sizing grid-connected hybrid PV/Generator/Battery systems following the prediction of CO2 emission and electricity consumption by machine learning methods (MLP and SVR): Aseer, Tabuk, and eastern region, Saudi Arabia. Front. Front Energy Res 2022;10:879373.

[79] Carneiro TC, Rocha PA, Carvalho PC, Fernández-Ramírez LM. Ridge regression ensemble of machine learning models applied to solar and wind forecasting in Brazil and Spain. Appl Energy 2022;314:118936.

[80] Vikas T, Malemnganbi R, Shimray BA. MLP-BP based optimal ranking of solar power plant site. Energy Exergy Sustain Clean Environ 2022;1:33–42.

[81] Melin P, Monica JC, Sanchez D, Castillo O. Multiple ensemble neural network models with fuzzy response aggregation for predicting COVID-19 time series: The case of Mexico. In: Healthcare, Vol. 8. MDPI; 2020, p. 181.

[82] Haykin S. Neural networks and learning machines, 3/E. Pearson Education India; 2009.

[83] He T, Dong Z, Meng K, Wang H, Oh Y. Accelerating multi-layer perceptron based short term demand forecasting using graphics processing units. In: 2009 transmission & distribution conference & exposition: asia and pacific. IEEE; 2009, p. 1–4.

[84] Chollet F, et al. Keras (2015). 2017.

[85] Goldsborough P. A tour of tensorflow. 2016, arXiv preprint arXiv:1610.01178.

[86] Deo RC, Ghimire S, Downs NJ, Raj N. Optimization of windspeed prediction using an artificial neural network compared with a genetic programming model. In: Research anthology on multi-industry uses of genetic programming and algorithms. IGI Global; 2021, p. 116–47.

[87] Ahmed AAM, Deo RC, Ghimire S, Downs NJ, Devi A, Barua PD, et al. Introductory engineering mathematics students' weighted score predictions utilising a novel multivariate adaptive regression spline model. Sustainability 2022;14(17):11070.

[88] Ghimire S, Deo RC, Downs NJ, Raj N. Global solar radiation prediction by ANN integrated with European centre for medium range weather forecast fields in solar rich cities of Queensland Australia. J Clean Prod 2019;216:288–310.

[89] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. Boosting solar radiation predictions with global climate models, observational predictors and hybrid deep-machine learning algorithms. Appl Energy 2022;316:119063.

[90] Abdi J, Hadipoor M, Hadavimoghaddam F, Hemmati-Sarapardeh A. Estimation of tetracycline antibiotic photodegradation from wastewater by heterogeneous metal-organic frameworks photocatalysts. Chemosphere 2022;287:132135.

[91] Ehsan BMA, Begum F, Ilham SJ, Khan RS. Advanced wind speed prediction using convective weather variables through machine learning application. Appl Comput Geosci 2019;1:100002.

[92] Pan T, Wu S, Dai E, Liu Y. Estimating the daily global solar radiation spatial distribution from diurnal temperature ranges over the Tibetan Plateau in China. Appl Energy 2013;107:384–93.
[93] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30(1):79–82.
[94] Mandeville A, O'connell P, Sutcliffe J, Nash J. River flow forecasting through conceptual models part III-the ray catchment at grendon underwood. J Hydrol 1970;11(2):109–28.
[95] Legates DR, McCabe Jr GJ. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res 1999;35(1):233–41.
[96] Espinar B, Ramírez L, Drews A, Beyer HG, Zarzalejo LF, Polo J, et al. Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. Sol Energy 2009;83(1):118–25.
[97] Despotovic M, Nedic V, Despotovic D, Cvetanovic S. Review and statistical analysis of different global solar radiation sunshine models. Renew Sustain Energy Rev 2015;52:1869–80.
[98] Gupta HV, Kling H, Yilmaz KK, Martinez GF. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J Hydrol 2009;377(1–2):80–91.
[99] McKenzie J. Mean absolute percentage error and bias in economic forecasting. Econom Lett 2011;113(3):259–62.
[100] Sun S, Qiao H, Wei Y, Wang S. A new dynamic integrated approach for wind speed forecasting. Appl Energy 2017;197:151–62.
[101] Diebold FX, Mariano RS. Comparing predictive accuracy. J Bus Econom Statist 2002;20(1):134–44.
[102] Costantini M, Pappalardo C. Combination of forecast methods using encompassing tests: An algorithm-based procedure. Tech. rep., Reihe Ökonomie/Economics Series; 2008.
[103] Al-Musaylh MS, Deo RC, Adamowski JF, Li Y. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. Adv Eng Inform 2018;35:1–16.
[104] Murawwat S, Asif HM, Ijaz S, Malik MI, Raahemifar K. Denoising and classification of arrhythmia using MEMD and ANN. Alex Eng J 2022;61(4):2807–23.
[105] Li Y, Tang B, Yi Y. A novel complexity-based mode feature representation for feature extraction of ship-radiated noise using VMD and slope entropy. Appl Acoust 2022;196:108899.
[106] Zhao Z, Wang H, Liu C, Xu X, Sun L, Wang J, et al. An FFT-based method for uncertainty quantification of Nomex honeycomb's in-plane elastic properties. Compos Struct 2022;116217.
[107] Jiang P, Li R, Liu N, Gao Y. A novel composite electricity demand forecasting framework by data processing and optimized support vector machine. Appl Energy 2020;260:114243.