

Efficient daily electricity demand prediction with hybrid deep-learning multi-algorithm approach

Sujan Ghimire^a, Ravinesh C. Deo^{a,*}, David Casillas-Pérez^b, Sancho Salcedo-Sanz^{a,c}

^a School of Mathematics, Physics, and Computing, University of Southern Queensland, Springfield, QLD, 4300, Australia

^b Department of Signal Processing and Communications, Universidad Rey Juan Carlos, Fuenlabrada, 28942, Madrid, Spain

^c 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 prediction
Sustainable energy
Artificial intelligence
Deep learning
Encode-decoder architectures
Kernel density estimation

ABSTRACT

Predicting electricity demand (G) is crucial for electricity grid operation and management. In order to make reliable predictions, model inputs must be analyzed for predictive features before they can be incorporated into a forecast model. In this study, a hybrid multi-algorithm framework is developed by incorporating Artificial Neural Networks (ANN), Encoder-Decoder Based Long Short-Term Memory (EDLSTM) and Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICMD). Following the partitioning of data, the G time-series are decomposed into multiple time-series using the ICEEMDAN algorithm, with partial autocorrelation applied to training sets to determine lagged features. We combine lagged inputs into a predictive framework where G components with the highest frequency are predicted with an ANN model, while remaining components are predicted with an EDLSTM model. To generate the results, all IMF components' predictions are merged using ICMD-ANN-EDLSTM hybrid models. A comparison is made between this objective model and standalone models (ANN, RFR, LSTM), hybrid models (CLSTM), and three decomposition-based hybrid models. Based on the results, the Relative Mean Absolute Error at Duffield Road substation was $\approx 2.82\%$, $\approx 4.15\%$, $\approx 3.17\%$, $\approx 6.41\%$, $\approx 6.60\%$, $\approx 6.49\%$, and $\approx 6.602\%$, compared to ICMD-RFR-LSTM, ICMD-RFR-CLSTM, LSTM, CLSTM, RFR, and ANN. According to statistical score metrics, the hybrid ICMD-ANN-EDLSTM model performed better than other benchmark models. Further, the results show that the hybrid ICMD-ANN-EDLSTM model can not only detect seasonality in G data, but also predict and analyze electricity market demand to add valuable insight to market analysis.

1. Introduction

As part of the power system's operation and control, electricity demand (G , MWh) prediction is an essential method of predicting fluctuations in electricity demand. Smart grids are capable of dispatching power more intelligently by predicting G accurately, although inaccurate estimation of G remains a major cause of power grid failures [1]. The seventh goal of the United Nations Sustainable Development Goals (SDG7) aims to empower the provision of affordable and reliable energy for all people and therefore calls for ensuring that energy generation and distribution are efficient in production and distribution [2]. Consequently, it is useful to study more efficient methods for G prediction, which is still a complex problem. Under-estimating G can reduce the stability of the entire electric power grid and prevent normal utilisation requirements from being met whereas overestimating G can lead to energy waste and additional operating costs [3]. Moreover, there have been studies showing that a 1% increase in electricity load forecast

error can result in millions of dollars lost [4]. In the past few decades, accurate modelling and prediction of G , which are crucial for power system management and electricity capacity scheduling, has attracted extensive attention.

Electricity demand prediction is classified into three groups based on prediction periods: short-term, mid-term, and long-term. Short-term prediction lasts from minutes to a week, whereas mid-term prediction lasts from a week to a year. Long-term prediction can have a period of more than a year. All three types of prediction are necessary instruments for smart grid control. Therefore, there have been numerous efforts by magnanimous researchers to develop precise models to predict G over last few decades. The G prediction model can be divided into four groups (a) physical model, (b) statistical model, (c) Artificial Intelligence (AI) based model and (d) hybrid models. Combining physical attributes and historical G data yields the physical

* Corresponding author.

E-mail addresses: sujan.ghimire@usq.edu.au (S. Ghimire), 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.enconman.2023.117707>

Received 6 April 2023; Received in revised form 21 August 2023; Accepted 24 September 2023

Available online 5 October 2023

0196-8904/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

model for G prediction. The majority of studies in this area show that future G data may be anticipated by studying the relationship between the original G time-series sequences and the physical information embedded within these datasets. The Unit Consumption, Load Density, and Elastic Coefficient [5] approaches are some of the physical models in literature. These physical methods, however, require a huge amount of data and are best suited for long-term G prediction. Statistical model requires historical data as their foundation, and produce short-term projections by modifying the parameters after investigating the relationship between historical and present data [6]. Although these statistical techniques benefit from their inherent improved interpretability, their predicting performance is unreliable due to their linear mapping feature [7]. The most common statistical model for G prediction are the Auto-Regressive Moving Average (ARMA) [8], Auto-Regressive Integrated Moving Average (ARIMA) [9], Generalised Autoregressive Conditional Heteroskedasticity (GARCH), Exponential Smoothing (ES), Grey Forecasting Model (GFM), Seasonal Exponential Smoothing (SES), Gray Linear Regression (GLR), Threshold Autoregressive Conditional Heteroskedasticity (TARCH), Fuzzy Logic (FL), and Kalman Filter (KF) [10]. In [8], G was predicted using ARMA with empirical findings demonstrating the effectiveness of their proposed model. In [11], a Multiplicative Seasonal ARIMA (MSARIMA) method for monthly peak G prediction for India, and empirical results show that MSARIMA outperforms other benchmark models. In [12,13], Fractional Grey Prediction model and Kalman Filter to predict G with empirical results showing that the proposed model outperformed the other comparable models.

The aforementioned statistical models are beneficial for linear sequences analysis but they could be ineffective for nonlinear model datasets. Furthermore, when used for time-series prediction, these statistical models make assumptions about stationarity. Moreover, because of their strengths in addressing problems with nonlinear and instable relationships, nonstationary time-series, and complex calculations, AI based Machine Learning (ML) and Deep Learning (DL) algorithms have become popular in predicting G [14,15]. The Artificial Neural Network (ANN) [16], Extreme Learning Machine (ELM), Random Forest Regression (RFR), Long-Short Term Memory Network (LSTM) [17], Convolutional Neural Network (CNN) [18], Gated Recurrent Unit (GRU), Kernel Ridge Regression (KRR), Deep Residual Network (ResNet), Deep Belief Network (DBN), Generalised Regression Neural Network (GRNN) [19], Back Propagation (BP) Neural Network (BPNN) [20], Recurrent Neural Network (RNN) and Support Vector Regression (SVR) are some of the most commonly used AI-based models for G prediction. The experimental results revealed that these models outperformed widely utilised statistical models for nonlinear dynamics of time series [21]. In [22], deep sequence-to-sequence Bidirectional LSTM was proposed and benchmarked with ANN, SVR and LSTM, the proposed model demonstrates the significant improve in prediction performance. Further, [23,24] compared ANN with a classic ARIMA model for G prediction to show the ANN outperformed an ARIMA model in terms of predictive power. In [25,26], it has been shown that the Deep Neural Network model (RNN, LSTM, and CNN) were found to outperform standard time-series models. There are other recently proposed relevant approaches including deep learning techniques such as [27], in which a deep learning algorithm based on an LSTM is proposed to improve electricity demand prediction including the COVID 19 lockdown. Similarly the SVR model was used by [28,29] resulting in significantly accurate predictions. However, any single AI approach alone cannot meet all of the prediction needs; consequently, one may need to combine several methodologies to thoroughly evaluate the data and increase prediction accuracy [30]. Thus, it is currently a common practice to build prediction model that use multiple AI techniques, reflecting their unique characteristics and future utility in electricity demand applications.

Based on literature, this study therefore aims to integrate improved complete ensemble empirical model decomposition, a data decomposition method used previously [31] with two AI-based predictive

methods including an encoder–decoder long short-term memory network [32] and an artificial neural network framework [33]. The integration of these methods is also based on the premise that a single AI model can have its own limitations and may not be applied to all kinds of G datasets or temporal forecasting scenarios. As a result, many scholars have proposed the concept of a hybrid model to address the existing issues. In [34], a new deep learning hybrid model combining Deep Neural Network and Historical Data Augmentation (DNN-HDA). This model has excellent generalisation and prediction accuracy compared to standalone models. Similarly, the Holt–Winters (HW) approach and an ELM model was employed in [35] to generate an HW-ELM hybrid model which considered a nonlinear hybrid model using linear forecast outputs and nonlinear residuals as the model inputs. In general, the results suggested that the HW-ELM model had a low prediction error. Other ELM-based hybrid models have been recently proposed, reporting important results, such as ANFIS-ELM [36]. Furthermore, in order to develop an effective operation control approach, [37] has suggested an Attention-based load prediction Recurrent Neural Network (Att-RNN) structure, which outperformed single RNN benchmark model in terms of precision and interpretability. Transformer's self-attention mechanism has been also explored in [38] for electricity demand. Additionally, in [39], a CNN and an LSTM model was integrated to shows increased prediction accuracy by 25%, and recently, in [40], the same hybrid method has demonstrated to have good performance for peak electricity demand prediction.

Several researchers have integrated single artificial intelligence and machine learning models with data processing methods to improve the predictive accuracy of their models [31,41,42]. These studies showed that a data decomposition-based hybrid model can break down any complex time-series such as G , into numerous modalities that represent unsteadiness in electricity use to improve the predictive accuracy and capture useful pattern required to model the fluctuations in G [43]. In [43], authors specifically proposed empirical mode decomposition (EMD) combined with an LSTM model for short-term load predictions with their results showing that the performance of the hybrid model was superior than a standalone model. In [44], the authors have built a Complementary Ensemble Empirical Mode Decomposition (CEEMD), a Multi-Objective Grey Wolf Optimiser (MOGWO), and a DBN with their proposed model showing increased predictive performance compared with standalone models. The study of [45] has developed a hybrid model comprised of Variational Mode Decomposition (VMD), a Binary Encoding Genetic Algorithm (BEGA), and an LSTM model with their results demonstrating superior performance of the hybrid model compared to the conventional benchmark models. In [46], authors developed a hybrid self-adaptive Particle Swarm Optimisation-Genetic Algorithm-Radial Basis Function model for annual electricity demand prediction, and in [47], a novel machine learning approach was developed to estimate electricity demand taking empirical evidence for the case of Thailand.

Although hybrid artificial intelligence frameworks for analysing various kinds of datasets with a feature decomposition-based approach are currently available, these methods still fail to consider unique characteristics of the G time-series, which are particularly driven by the low-and high-frequency fluctuations caused by dynamism in consumer electricity markets. To address this industry problem, the contributions of this research are as follows. (i) To fill in the important knowledge gaps in electricity demand predictions that takes into account low-and high-frequency patterns in G and further integrating novel AI methods to predict the inherently diverse G features. (ii) To improve the accuracy of G predictions by developing a hybrid framework with an ANN and an Encoder Decoder LSTM model, denoted as EDLSTM based on the Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN or ICMD) method. (iii) To comprehensively evaluate the proposed hybrid ICMD-ANN-EDLSTM model to ascertain its capability to decompose the G data into low- and high-frequency pattern in such a way that the historical patterns can

be considered for more accurate modelling of daily electricity demand than currently available models. Moreover, the proposed model is tailored to detect the anomalous features prior to forecasting the future electricity demand efficiently.

The novelty of this study is to consider the trends, chaotic behaviour and instabilities found in G data that cause inaccuracies in forecast models and therefore, to adopt multiple predictive methods to emulate these distinct components as follows.

- The original electricity demand time-series are decomposed into Intrinsic Mode Functions (IMFs) using the ICMD method.
- The Partial Autocorrelation Function, aiming to investigate the relationships between antecedent electricity demand, is employed to deduce the significant lagged data series from decomposed IMF to create an input matrix for the proposed hybrid ICMD-ANN-EDLSTM model.
- An ANN is used to predict the highest frequency components whereas the remaining IMFs are predicted by a new EDLSTM model.
- The Bayesian Optimisation (BO) method is employed to obtain optimal hyperparameters of the proposed ICMD-ANN-EDLSTM (and the comparative counterpart) models.
- The prediction of each IMF components are aggregated to generate the final prediction results and the ICMD-ANN-EDLSTM model is benchmarked with decomposition-based hybrid and standalone models.

To the best of the authors' knowledge, a careful integration of the proposed hybrid ICMD-ANN-EDLSTM model based on ICMD method in predicting daily G has not been yet explored. As the proposed ICMD-ANN-EDLSTM model integrates components from both the ICMD method and the ANN-EDLSTM model, this allows it to better capture the complex dynamics of daily G data. By combining these approaches, the proposed model is able to make more accurate predictions of the daily G than either of the standalone models.

2. Theoretical framework of deep learning models

The proposed hybrid ICMD-ANN-EDLSTM model was trained to predict daily electricity demand by combining a well-known frequency decomposition technique based on the Improved Complementary Ensemble Empirical Mode Decomposition with Adaptive Noise (ICMD), and a hybrid Deep Learning method (ANN-EDLSTM). To have a better understanding of the proposed methods, we firstly present a brief theoretical background of the models and, after that, the details of the proposed hybrid models are discussed.

2.1. Improved Complementary Ensemble Empirical Mode Decomposition with Adaptive Noise

The time series prediction decomposition model has been enhanced from the basic Empirical Mode Decomposition model (EMD) to Ensemble EMD (EEMD), Complementary EEMD (CEEMD), Complementary EEMD with Adaptive Noise (CEEMDAN). The EMD, proposed by Huang et al. [48], is a method for adaptive time-frequency analysis that is vulnerable to the mode mixing problem. EEMD [49] addresses this issue, although it is computationally inefficient, and residual noise is always present in the reconstructed signal. Further, CEEMD [50] was developed to address these constraints. Both EEMD and CEEMD have a proclivity to generate erroneous components [51]. CEEMDAN resolves such issues; nonetheless, CEEMDAN still has some issues [52]. As a result, in [53], the Improved Complementary Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN or ICMD) method was proposed. In ICMD, complex time-series signals are decomposed into finite Intrinsic Mode Functions (IMFs) and residuals. Each IMF holds the signal's local characteristics and can help minimise noise. Firstly, the local mean is utilised instead of modal estimate. Secondly, rather of employing white noise directly, the signal's local mean value is used to extract the k -order modes [42]. Further details on the ICEEMDAN implementation can be found in [53].

2.2. Artificial neural network

Artificial neural networks, designed to recreate basic biological neural systems, particularly the human brain, consist of a massive number of neurons stacked in different levels, usually three levels: an input layer, an output layer, and one or more hidden layers [54]. The input layer, which contains one neuron for each of the problem's input parameters. The second layer is the hidden layer, which is used for parameter interactions. The output layer, is used to generate expected values. Weights determined iteratively in a training stage define the link between neurons. The ANN architecture considered in this work is shown in Fig. 1(a). In the training phase, the training algorithm is used to determine weights that minimise some overall error measure, such as the Sum of Squared Errors (SSE) or root Mean Squared Errors (RMSE). As a result, ANN training is an unconstrained nonlinear minimisation problem. The training algorithm plays a vital role in the minimising the loss. In this study, an specific backpropagation (BP) algorithm, the Adam optimiser, is used. The Adam optimizer is an adaptive learning strategy which learns each weight in the neural network. It employs estimations of both the first and second moments of gradient and analyses individual learning rates for distinct parameters [55]. In turn the ReLU activation function is used in this work, since ReLU can handle large layers and deal with the vanishing gradient problem [56].

2.3. Random Forest model

The Random Forest (RFR) model is one of the most often used models in decision tree learning for regression and classification, proposed by Breiman [57]. It is extremely efficient and many time exceeds other regression models in terms of regression accuracy. The RFR model creates a large number of decorrelated decision trees during the training phase. The model's output is then calculated by averaging the output values of all individual trees [58]. In the regression problem, the RFR model did well at simulating highly non-linear relationships between a set of inputs and an output. Mathematically, the RFR model is a development of classification and regression trees (CART), which has the benefit over other techniques for dealing with non-linearities, such as NN, of always being free of over-fitting [59]. The RFR architecture is shown in Fig. 1(b). To produce the training random set and the associated decision tree, the bootstrap resampling approach is utilised.

The applications of the RFR model have demonstrated its capability in a variety of fields including hydrological, renewable and electricity demand prediction domains [33,55,58,60–68]. During the model development, three parameters n_{trees} , $mtry$, and $nodesize$ affect the model performance, and must be determined by users. The $nodesize$ option specifies the minimum number of samples in each node. A smaller $nodesize$ suggests that more trees must be produced, resulting in increase in computation time. $nodesize$ is usually set to 5 for the regression work [69].

2.4. Long Term Short Term Memory Network

The Long Term Short Term Memory Network (LSTM) model [70] is a special type of Recurrent Neural Network (RNN) with memory capabilities, that is often used to process and forecast critical events with fairly long time intervals in time-series. It has cell and hidden states that can account for the long-short term memory effects [71]. Furthermore, the LSTM model defines three gate mechanisms on the RNN structure to address the issue of not retaining information history for a long time. These gates are the input gate, the forget gate, and the output gate. LSTM outperforms other RNNs, notably Gated Recurrent Unit (GRU), by integrating a gating mechanism to govern the flow of information and the updating of states and cells [72,73]. Moreover, because of its gate structure, an LSTM layer can learn long-term dependencies and can be used to predict time series successfully.

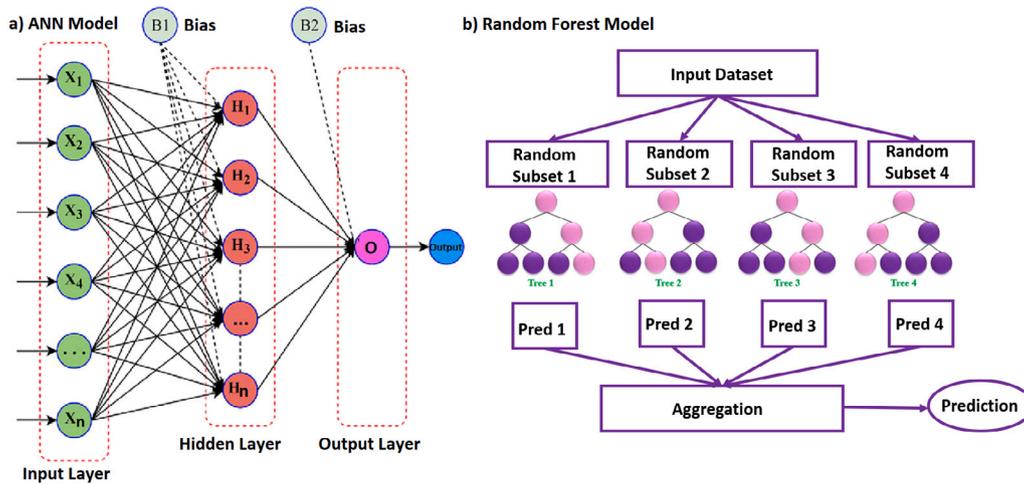


Fig. 1. Structure of Artificial Neural Network (ANN) and Random Forest (RFR) for a regression problem.

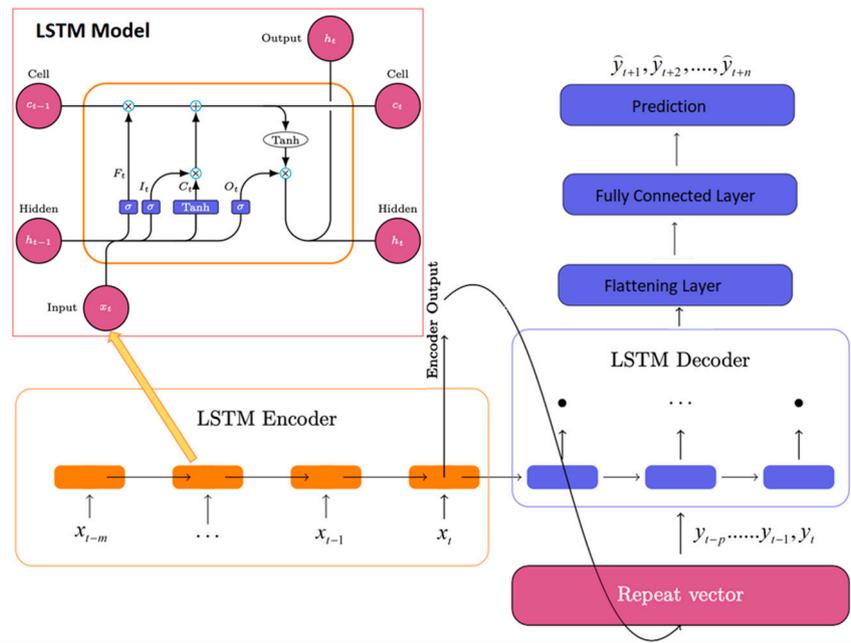


Fig. 2. Architecture of the Encoder Decoder based Long Term Short Term Memory Network (EDLSTM), The “Repeat Vector” layer connects the Encoder and Decoder by repeatedly repeating the internal representation of the input sequence, once for each time step in the output sequence.

2.5. Auto encoder–decoder architecture with LSTM

The LSTM approach, in conjunction with an encoder–decoder (EDLSTM) architecture, was first proposed in [74] for the task of natural language processing (NLP). This architecture, in conjunction with LSTM model (EDLSTM) is commonly used now to predict time-series and address sequence-to-sequence problems, in various fields like renewable, electricity demand, [32,75–80]. This model’s architecture is made up of two LSTM networks, one in each of the encoder and decoder sections (Fig. 2). The encoder section reads the sequence’s input information to extract the timing characteristics of historical data and encodes the fixed-length vector; the decoder section decodes the vector and outputs the predicted sequence. In EDLSTM, the encoder layer, unlike the fundamental LSTM structure, only outputs the hidden state from the last cell. Then, in the decoder layer, output from encoder is copied as input for each LSTM cell. It stores data gathered from the input sequence at each time step. Therefore, it may be more beneficial to employ EDLSTM to improve long-term dependencies for longer time

step prediction than the normal LSTM. An illustration of the EDLSTM network is shown in Fig. 2.

2.6. Convolution Neural Network

Convolution Neural Network (CNN) [81] is another type of Deep Neural Network. CNN is an excellent technology for automatically extracting features that has achieved remarkable success in the field of image vision. Meanwhile, they show great promise in dealing with time-series, such as automatic speech recognition, wind speed predictions and solar radiation [33,55,68,82,83]. To extract feature information, CNN models can include a convolution layer (CL) as a core layer, pooling layer (PL), fully connected (FC) layer and the output (OP) layer. The CL and PL, which are directly inspired by the basic notions of simple cells and complex cells in visual neuroscience, play a significant part in a CNN model. Multiple convolution filters are used in each CL to extract distinct characteristics. The units in a CL are connected to the previous layer’s local patches by a set of weights, and the outputs of this local weighted sum are then processed via a non-linearity function

Table 1
Descriptive statistics of daily electricity demand G (MWh) at four substations of South-east Queensland.

Statistical parameters	Duffield Road sub-station	Kedron sub-station	Kirra sub-station	Molendinar sub-station
Median (MWh)	247.62	283.87	208.69	446.83
Mean (MWh)	248.35	290.18	210.34	445.20
Standard deviation (MWh)	32.96	46.23	40.88	65.78
Variance	1086.05	2137.42	1671.06	4327.35
Maximum (MWh)	528.01	498.69	429.54	755.26
Minimum (MWh)	34.58	7.83	-0.24	174.35
Range	493.43	490.86	429.78	580.91
Interquartile range	40.97	56.35	32.01	80.85
Skewness	0.42	0.62	-2.67	0.41
Kurtosis	5.53	4.84	15.97	3.69

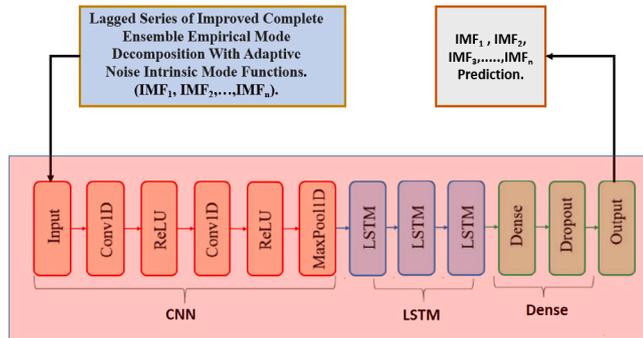


Fig. 3. Schematic of Convolution Neural Network integrated with LSTM model.

known as the activation function (AF). Furthermore, the PL is primarily utilised to reduce spatial dimension, which can significantly reduce the amount of calculation required for network training. Convolution, non-linearity, and pooling are stacked in two or three levels, followed by more CL. Finally, the features make their way through the FC layers to the OP layer.

2.7. Deep learning hybrid model: Convolution Neural Network integrated with Long Term Short Term Memory Network

CLSTM refers to a model that combines CNN and LSTM architecture. The CNN model operates on the same concept as discussed in the preceding section. The LSTM model, on the other hand, receives inputs from preceding dense layers, which receive inputs from CNN pooling layers. Fig. 3 shows the architecture of the proposed CLSTM model.

2.8. Decomposition based deep learning hybrid models

To predict electricity demand (G , MWh), this study presents a deep learning hybrid technique of ANN and EDLSTM based on ICMD. The original G data is translated into different IMF components utilising ICMD in the first stage of our proposed method. The component with the greatest frequency is then predicted using ANN, while the remaining components are predicted using EDLSTM. The prediction findings of all IMF components are merged in the final stage to generate the final prediction results. This Hybrid model is termed as ICMD-ANN-EDLSTM. This Objective model (ICMD-ANN-EDLSTM) was compared with standalone models (ANN, RFR and LSTM), deep learning hybrid model (CLSTM) and the following three decomposition based deep learning hybrid models.

Model 1: ICMD-ANN-CLSTM: Highly fluctuating IMFs are predicted using ANN and remaining IMFs by CLSTM.

Model 2: ICMD-RFR-CLSTM: Highly fluctuating IMFs are predicted using RFR and remaining IMFs by CLSTM.

Model 3: ICMD-RFR-LSTM: Highly fluctuating IMFs are predicted using RFR and remaining IMFs by LSTM.

3. Material and methods

3.1. Study region and electricity demand dataset

The G data utilised in this study is from ENERGEX (<https://www.energex.com.au>). The G data sets were selected for four Sub-stations ((a) Duffield Road, (b) Kedron, (c) Kirra, and (d) Molendinar) at South-east Queensland, Australia. Table 1 displays some descriptive information for the selected substations' daily G . The dataset contains 280,560 measurements for 30-min G from 01/07/2011 to 30/06/2021 (120 months or 3654 days). The dataset has been downsampled from 30-min interval to daily interval using Eq. (1), where G_D is a function that takes a collection of electricity demand data as input and down samples to a given period at a rate of n (i.e., $n = 48$ for daily transformation of 30-min data). It should be also noted that there are no missing data in the extracted G time-series data for four Sub-stations.

$$G_{D_i} = \sum_{j=i*n}^{(i*n)+n} G_{D_j} \quad (1)$$

3.2. Development of the ICMD-ANN-EDLSTM model

In this study, we develop a deep learning hybrid model for predicting electricity demand (G , MWh) that combines Artificial Neural Network (ANN), Encoder and Decoder based Long Short Term Memory (EDLSTM), and Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICMD). The development of the proposed ICMD-ANN-EDLSTM model involved the following steps:

Step 1: The extracted data is divided into training and testing set, 50% of data for training (i.e. 1827 data point) and remaining 50% for testing. To prevent the existence of future data in the input G time-series and the subsequent introduction of biases in the predictions, all data were partitioned independently into training and testing sets. Furthermore, the 50/50 split was done in order to get the equal number of IMFs for training and testing period.

Step 2: The Training data as well as testing data is divided into multiple Intrinsic Mode Function (IMF) components using ICMD. Each IMF component created by ICMD has unique properties and is sorted from highest to lowest frequency. The first component (IMF1) has the highest frequency. The remaining frequency IMFs reflect the periodic patterns or seasonality of the data. The final IMF component, commonly known as the residual, has the lowest frequency component. It also indicates the overall trend of the data. To generate the IMF the predefined ICMD parameters are used, the number of realisations of the Gaussian Noise or ensemble number $EN = 1000$, the added Gaussian noise was set to 0.2 times the standard deviation of the G time-series, and maximum shifting iteration (NS) was set to 5000. For Duffield Road, Kedron and Molendinar sub-stations, the ICMD algorithm decomposed the G data into eight IMF sub-series (IMF1, IMF2, ..., IMF8) and a residual component (RES) whereas for Kirra sub-station using the ICMD

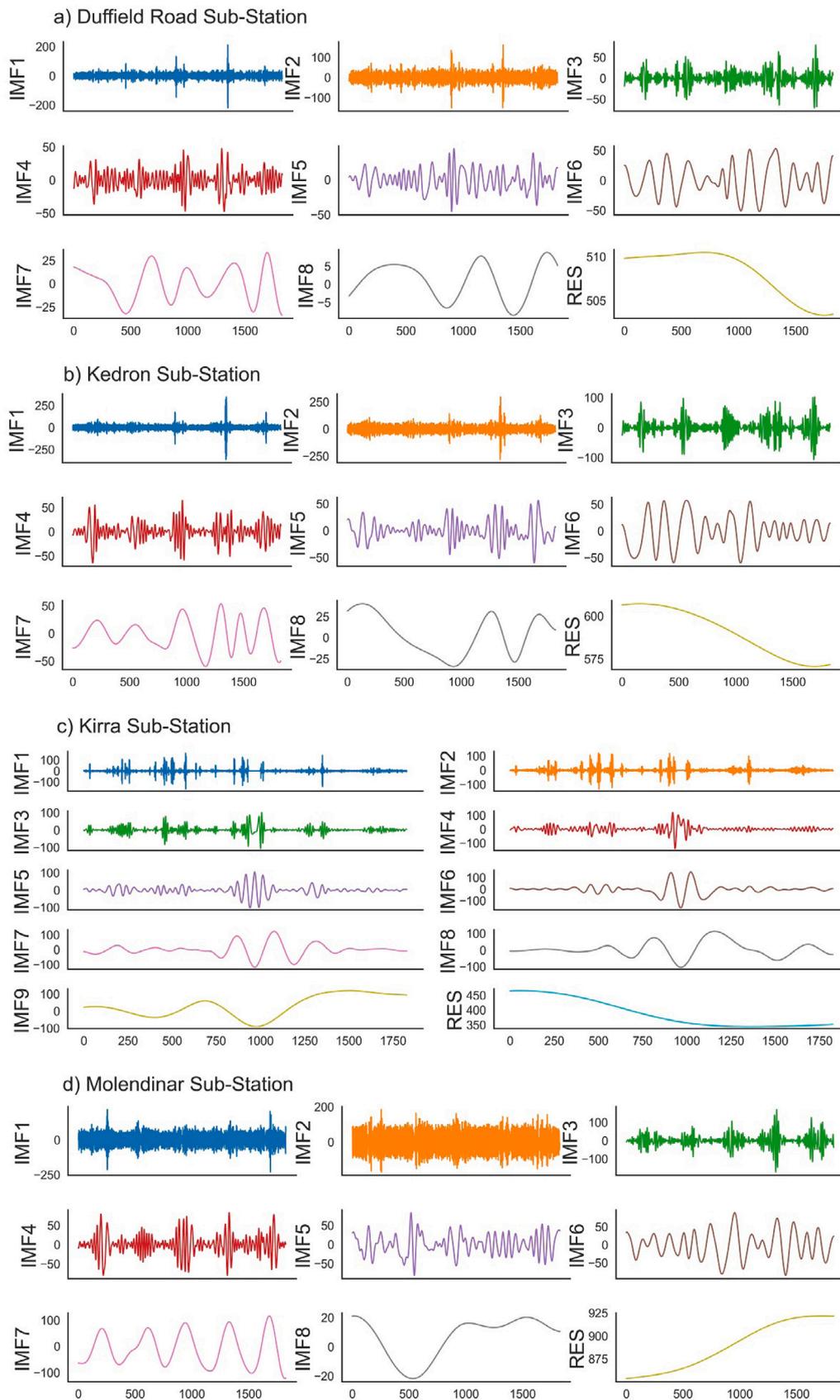


Fig. 4. Intrinsic mode functions (IMFs) and a residual component of electricity demand time-series using ICMD for (a) Duffield Road, (b) Kedron, (c) Kirra, and (d) Molendinar sub-stations during the training period.

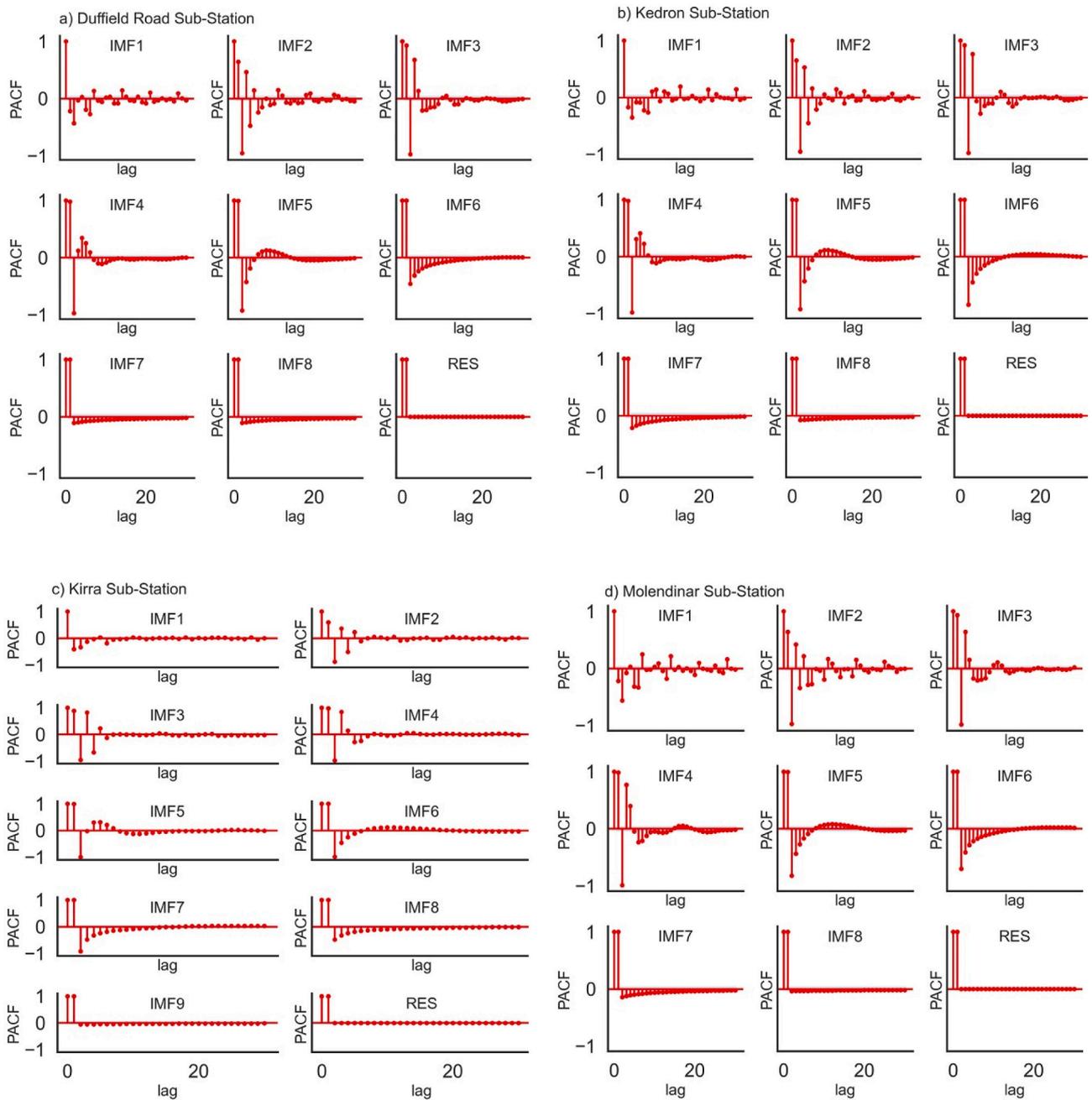


Fig. 5. Statistically significant PACF of IMFs and residual used for developing the ICMD-ANN-EDLSTM model at (a) Duffield Road, (b) Kedron, (c) Kirra, and (d) Molendinar sub-stations.

algorithm the G series was decomposed into nine IMF sub-series and a residual component. Fig. 4 depicts the ICEEMDAN decomposed G data series in IMFs with residual during training period for all four sub-stations.

Step 3: Following that, in order to create the input matrix for the deep learning hybrid model statistically significant lag time-series of respective IMFs and residual component were determined using the Partial Autocorrelation Function (PACF), as shown in Fig. 5. Table 2 shows the PACF generated significant input lags for the deep learning hybrid models.

Step 4: Further, the significant lags of IMFs are normalised using the z-score method. This normalisation is done to speed-up the training process and increase the prediction precision of model [84]. The z-score, which is derived using the arithmetic

mean and standard deviation of the given data, is the most often used score normalisation technique in Deep Learning. The normalisation process is done using Eq. (2)

$$IMF'_k = \frac{IMF_k - \mu}{\sigma} \quad (2)$$

where μ is the arithmetic mean, σ is the standard deviation of the given data, IMF'_k is normalised IMF and IMF_k is original IMF.

Step 5: The next stage in model development is data prediction, the respective IMFs, and the residual component were predicted one at a time. In this step, two prediction models are employed to predict IMF components based on their properties. As can be seen in the Fig. 5, the IMF1, IMF2, IMF3, and IMF4 series are the most complex and highly volatile. The ANN prediction model was chosen because of its ability to predict complex

Table 2
Input variables for objective(ICMD-ANN-EDLSTM) as well as benchmark models based on PACF in daily electricity demand prediction.

Significant input lag numbers at respective sub-stations				
Intrinsic mode functions	Duffield Road sub-station	Kedron sub-station	Kirra sub-station	Molendinar sub-station
IMF1	1–12	1–9,15	1–4,7	1–4,6–7,8,12,14,15
IMF2	1–12	1–9	1–6	1–8,11,12,14
IMF3	1–12	1–9	1–6	1–9
IMF4	1–12	1–10	1–7	1–5,7–9
IMF5	1–13	1–10	1–3,5,6,7	1–9
IMF6	1–12	1–10	1–6	1–10
IMF7	1–8	1–8	1–8	1–9
IMF8	1–7	1–8	1–9	1–2
IMF9			1–2	
RES	1–2	1–2	1–2	1–2

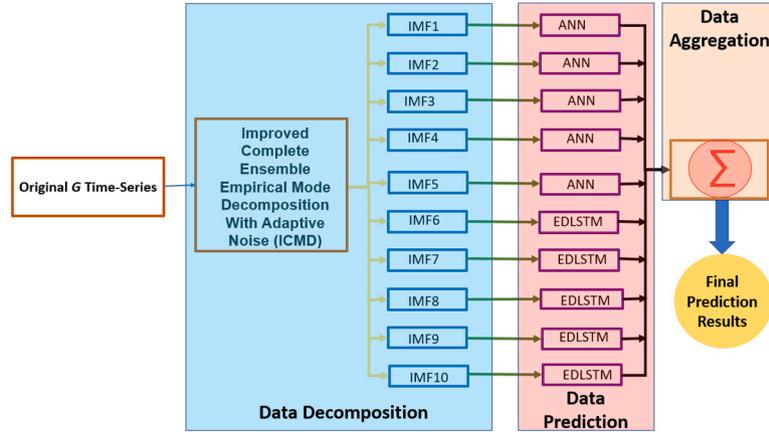


Fig. 6. Flowchart of the proposed deep learning hybrid ICMD-ANN-EDLSTM model.

time series [85,86]. The remaining IMFs (IMF5, IMF6, IMF7, IMF8, IMF9) which is periodic component of original IMF, for which the EDLSTM model is selected because this EDLSTM model has demonstrated success in predicting periodic patterns in time series data [32,87–89]. Additionally, for the residual component which indicates the long-term trend of data again the EDLSTM Model is used.

Step 6: In the last stage, the prediction outputs of all components are aggregated using summation to produce final prediction results. Fig. 6 depicts the proposed method's block diagram.

3.2.1. Benchmark model development

The deep learning hybrid ICMD-ANN-EDLSTM performance on testing set is compared against the standalone models (ANN, RFR and LSTM), deep learning hybrid model (CLSTM) and the decomposition based deep learning hybrid models (ICMD-ANN-CLSTM, ICMD-RFR-CLSTM and ICMD-RFR-LSTM). The simulation and coding for the objective model as well as benchmark models are done by the Python Jupyter Notebook [90] with the Intel i5-7200U CPU of 2.50 GHz. This software contains many libraries such as NumPy, TensorFlow, Matplotlib, Keras, etcetera [91–93].

3.2.2. Model hyperparameters tuning with Bayesian optimisation

A deep learning model performs best on a particular issue when its hyperparameters are optimised. Using various optimisation techniques, an ideal combination of hyperparameters can be obtained. Manual or automated procedures (such as grid search, random search, and so on) have typically been employed for hyperparameter optimisation. The accuracy of grid search reduces as the number of parameters increases. Manual procedure is prone to human errors and necessitates specialist knowledge [94]. Furthermore, random search is based on random distribution functions, therefore, it is common for it to miss the ideal

hyperparameters in the search. In this study, the objective models as well as benchmark models has been optimised using the Bayesian Optimisation (BO). Bayesian optimisation successfully solves the classical machine intelligence problem in sequential decision theory [95]. In order to arrive at the optimal solution in the shortest amount of time, BO determines the next evaluation position based on the information provided by the unknown objective function (f). Literature has extensively investigated BO's effectiveness over other optimisation methods [96]. As presented in Algorithm 1, BO algorithm is an iterative process. Nine hyperparameters of the proposed model (ICMD-ANN-EDLSTM)

Algorithm 1 Bayesian optimisation

- 1: **for** $t = 1, 2, \dots$ **do**
- 2: Find x_t by optimising the acquisition function u over function f :
 $x_t = \arg \max_x u(x|D_{1:t-1})$
- 3: Sample the objective function:
 $y_t = f(x_t)$
- 4: Augment the data $D_{1:t} = \{D_{1:t-1}, (x_t, y_t)\}$ and update the posterior of function f
- 5: **end for**

were optimised using BO. Table 3 displays the parameter range and the optimal values acquired after the optimisation procedure. Additionally the RFR, LSTM and CLSTM models hyperparameters are also derived from BO method, the optimal value of the hyperparameters are depicted in Table 4. Furthermore, during model training the epochs was set to 1000 but we have added the Keras checkpoint for stopping the training process if the performance starts degrading. These checkpoints are Earliestopping (es) and ReduceLRonPlateau, es is a technique that allows the users to provide an arbitrary large number of training epochs and then stop training when the model performance on the validation dataset stops improving i.e. terminate training before a model begins to overfit. Similarly, ReduceLRonPlateau is a scheduling approach that reduces the learning rate (lr) when the validation loss stops improving

Table 3

Range of hyperparameters of the proposed model (ICMD-ANN-EDLSTM) and the optimal values obtained from Bayesian optimisation algorithm. Note: ReLU = Rectified Linear Units and Adam = Adaptive Moment Estimation Algorithm.

Hyperparameters	Range of values	Duffield Road sub-station	Kedron sub-station	Kirra sub-station	Molendinar sub-station	
		Optimal value				
ANN	Hidden neuron	[50,60,70,80,90,100]	80	60	80	80
	Learning rate	[0.001,0.002,0.005,0.006,0.008]	0.005	0.001	0.005	0.008
	Epochs	1000				
	Activation function	ReLU				
	Solver	[Adam]				
EDLSTM	Number of units in Encoder Cell 1	[80,90,100,200]	80	100	200	100
	Number of units in Encoder Cell 2	[30,40,50,90,100]	50	90	50	30
	Number of units in Encoder Cell 3	[10,20,30,40,50]	40	40	20	50
	Number of units in Decoder Cell 1	[80,90,100,200]	100	90	100	80
	Number of units in Decoder Cell 2	[30,40,50,90,100]	30	100	50	40
	Number of units in Decoder Cell 3	[10,20,30,40,50]	30	10	20	40
	Batch size	[5,10,15,20,25,30]	5	5	5	5
	Epochs	1000				
	Learning rate	[0.001]				
	Optimisation algorithm	[Adam]				
	Activation function of encoder	[ReLU]				
Activation function of decoder	[ReLU]					

Table 4

Range of hyperparameters of the other benchmark models and the optimal values obtained from Bayesian optimisation algorithm. Note: ReLU = Rectified Linear Units and Adam = Adaptive Moment Estimation Algorithm.

Hyperparameters	Range of values	Duffield Road sub-station	Kedron sub-station	Kirra sub-station	Molendinar sub-station	
		Optimal value				
RFR	The maximum depth of the tree.	[5,8,10,20,25]	20	25	25	10
	The number of trees in the forest.	[50,100,150,200]	150	50	100	200
	Minimum number of samples to split an internal node	[2,4,6,8,10]	4	2	4	6
	The number of features to consider when looking for the best split.	['auto', 'sqrt', 'log2']	auto	auto	auto	auto
LSTM	LSTM layer	[1,2,3,4]	2	3	3	3
	LSTM cell (units)	[10–200]	180,60	90,50,40	100,70,20	110,50,40
	Batch size	[5,10,15,20,25,30]	5	5	5	5
	Epochs	1000				
	Learning rate	[0.001]				
	Optimisation algorithm	[Adam]				
CLSTM	Activation function	[ReLU]				
	CNN layer	[1,2,3,4]	2	3	2	2
	CNN filter	[10–150]	80,50	120,70,10	90,40	90,60
	LSTM layer	1	1	1	1	1
	LSTM cell (units)	[10–200]	60	40	80	50
	Batch size	[5,10,15,20,25,30]	10	5	10	10
	Epochs	1000				
	Learning rate	[0.001]				
Optimisation algorithm	[Adam]					
Activation function	[ReLU]					

for a longer period of time than the patience number permits. Thus, the lr is kept constant as long as it improves the validation loss, but lr is reduced when the results become stagnant. In this study, the patience for es was set to 20 and for ReduceLROnPlateau the factor was set as 0.25 ($lr_{new} = lr * 0.25$) and Patience = 10.

3.2.3. Model performance evaluation criteria

Since there are not a single statistical metric available that is completely conclusive, a comprehensive and resilient model evaluation necessitates both objective and subjective evaluations. Therefore, in this study, the performance of the objective model (ICMD-ANN-EDLSTM) and all of the other benchmark predictive models were evaluated using deterministic error metrics such as Coefficient of Determination (R^2), Mean Absolute Error (MAE (MWh)), Root Mean Square Error ($RMSE$ (MWh)), Relative Root Mean Square Error ($RRMSE(\%)$), Relative Mean Absolute Error ($RMAE(\%)$), Willmott's Index (E_{WI}), Nash–Sutcliffe Index (E_{NS}) and Legates and McCabe's Index (E_{LM}). The shortcoming of these metrics is that they quantify model assessment in a few numbers. Thus, subjective model performance assessments using various diagnostic plots such as box plots, forecasting error histograms, and Taylor plot are also performed to have a better understanding. The

mathematical expression of objective metrics are as follows [97,98]:

$$R^2 = \left(\frac{\sum_{i=1}^n (G^a - \langle G^a \rangle)(G^p - \langle G^p \rangle)}{\sqrt{\sum_{i=1}^n (G^a - \langle G^a \rangle)^2} \sqrt{\sum_{i=1}^n (G^p - \langle G^p \rangle)^2}} \right)^2 \quad (3)$$

$$RMSE \text{ (MWh)} = \sqrt{\frac{1}{N} \sum_{i=1}^N (G^p - G^a)^2} \quad (4)$$

$$MAE \text{ (MWh)} = \frac{1}{N} \sum_{i=1}^N |G^p - G^a| \quad (5)$$

$$RRMSE(\%) = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (G^p - G^a)^2}}{\langle G^a \rangle} \cdot 100 \quad (6)$$

$$RMAE(\%) = \frac{1}{N} \sum_{i=1}^N \frac{|G^a - G^p|}{G^p} \cdot 100 \quad (7)$$

$$E_{WI} = 1 - \frac{\sum_{i=1}^N (G^a - G^p)^2}{\sum_{i=1}^N (|G^p - \langle G^a \rangle| + |G^a - \langle G^p \rangle|)^2} \quad (8)$$

$$E_{LM} = 1 - \frac{\sum_{i=1}^N |G^p - G^a|}{\sum_{i=1}^N |G^a - \langle G^a \rangle|} \quad (9)$$

$$E_{NS} = 1 - \frac{\sum_{i=1}^N (G^a - G^p)^2}{\sum_{i=1}^N (G^a - \langle G^a \rangle)^2} \quad (10)$$

where G^a and G^p are the actual and the predicted G , $\langle G^a \rangle$ and $\langle G^p \rangle$ are the actual and predicted mean values of G and N is the number of tested (future) data points. For the best model performance, the range and the rationale of the metrics, in terms of their sophistication, are as follows:

- **First Order Metrics:** R^2 is bounded by $[-1, 1]$ whereas MAE and $RMSE$ (in absolute units of G (MWh)) range from 0 for a perfectly fitted to ∞ for a inferior model. R^2 determines how well the modelled data is fit to observed or actual data whereas the $RMSE$ and MAE are measure of predictive power in absolute terms.
- **Second Order Metrics:** Relative $RMSE$ and MAE are in the range of $[0, 100]\%$. If these relative metrics are less than 10% then the predictive model is excellent, and if the value is between 10% to 20% the model is good [99].
- **Normalised Metrics:**
 - The E_{WI} , bounded by $[0, 1]$, is an improvement of $RMSE$ and MAE , E_{WI} can detect additive and proportional differences in the observed and simulated means and variances [100].
 - The E_{NS} , bounded by $[-\infty, 1]$, determines the relative magnitude of the residual variance compared to the measured data variance, $-\infty$ for the worst fit to 1 for a perfectly fitted model [101].
 - The E_{LM} , bounded by $[0, 1]$, When compared to E_{NS} and E_{WI} , E_{LM} is a more robust metric that was developed to solve the shortcomings of both [102].

Furthermore the performance of the model was assessed via the Kling–Gupta Efficiency (KGE) [103] and the Absolute Percentage Bias (APB ; %) [104]. KGE is improved version of E_{NS} , which facilitates the analysis of the relative importance of its different components (correlation, bias and variability).

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\langle G^p \rangle}{\langle G^a \rangle} - 1\right)^2 + \left(\frac{CV_p}{CV_a}\right)^2} \quad (11)$$

$$APB(\%) = \left| \frac{\sum_{i=1}^n (G^a - G^p)}{\sum_{i=1}^n G^a} \right| \cdot 100 \quad (12)$$

where r was the correlation coefficient and CV was the coefficient of variation.

Additionally, the Promoting Percentage of Absolute Percentage Bias λ_{APB} , Kling–Gupta Efficiency λ_{KGE} , and Root Mean Square Error λ_{RMSE} are adopted to compare the performance of different models.

$$\lambda_{APB} = \left| \frac{(APB_1 - APB_2)}{APB_1} \right| \quad (13)$$

$$\lambda_{KGE} = \left| \frac{(KGE_1 - KGE_2)}{KGE_1} \right| \quad (14)$$

$$\lambda_{RMSE} = \left| \frac{(RMSE_1 - RMSE_2)}{RMSE_1} \right| \quad (15)$$

where APB_1 , $RMSE_1$ and KGE_1 are objective model performance metrics, and APB_2 , $RMSE_2$ and KGE_2 are benchmark model performance.

This study has also adopted the Global Performance Indicator (GPI) [105] to integrate six performance metrics into 1. The GPI is defined as below

$$GPI = \sum_{j=1}^n \alpha_j (y_j - y_{mj}) \quad (16)$$

where j is the performance indicator, whose number varies from 1 to n ; y_j is the normalised value of the performance indicator j ; and y_{mj} is the median of the normalised values of the performance indicator j . The GPI value is the distance between the normalised value of a model's performance indicator and the median of the same indicator's normalised value. The higher the deviation of a model's performance indicator value from the median of all models, the more accurate the model is when compared to the others. In this study, α_j value is -1 for R^2 and equalled 1 for the other statistical indicators.

The proposed ICMD-ANN-EDLSTM model has been evaluated from a statistical perspective based on the Diebold–Mariano (DM) test and Harvey, Leybourne and Newbold (HLN) test, both of which are excellent for evaluating predictive model statistical significance. DM and HLN statistics test assumes similar accuracy between two forecasting models. The main process of the DM and HLN test is found in [106].

4. Results

The performance of the proposed deep learning hybrid ANN-EDLSTM model based on ICEEMDAN (ICMD) for daily electricity demand (G , MWh) prediction was compared with standalone models (ANN, RFR and LSTM), deep learning hybrid model (CLSTM) and the decomposition based deep learning hybrid models (ICMD-ANN-CLSTM, ICMD-RFR-CLSTM, and ICMD-RFR-LSTM). The proposed deep learning hybrid ICMD-ANN-EDLSTM model: ICMD-ANN-EDLSTM is evaluated using a variety of model evaluation metrics. Table 5 tabulates the $RMSE$ and MAE metrics of ICMD-ANN-EDLSTM and other benchmark models. The proposed deep learning hybrid model (ICMD-ANN-EDLSTM) revealed the best performance by displaying the lowest error values of $RMSE$ (MWh) ≈ 9.39 , ≈ 9.73 , ≈ 3.51 and ≈ 17.04 for Duffield Road, Kedron, Kirra, and Molendinar sub-station, respectively during testing phase. Similarly, the deep learning hybrid (ICMD-ANN-EDLSTM) model also outperform the benchmark model in terms of MAE for all four substations. Additionally, when the MAE of the proposed model is compared with other decomposition based models (ICMD-ANN-CLSTM, ICMD-RFR-CLSTM, and ICMD-RFR-LSTM), the MAE was reduced by 37%, 34% and 50% for Duffield Road sub-station, in case of deep learning hybrid model (CLSTM) the MAE was reduced by 127%, similar trend can be seen for standalone models (ANN, RFR and LSTM) with more than 100% reduction in MAE by the proposed model. Therefore, based on $RMSE$ and MAE metrics, it was clear that the proposed decomposition based deep learning hybrid model ANN-EDLSTM are superiors compared to other comparing models (ICMD-ANN-CLSTM, ICMD-RFR-CLSTM, ICMD-RFR-LSTM, and CLSTM) and their standalone (ANN, RFR, and LSTM) counterparts.

The scatter plots, as illustrated in Fig. 7, were created to provide a more comprehensive understanding of the model prediction performance. The scatter plots also show the coefficient of determination R^2 , which can be used to further evaluate the developed models. The statistical correlation for each model was examined using a scatter plot of the actual electricity demand G_a versus predicted electricity demand G_p . During the testing phases, the proposed ICMD-ANN-EDLSTM model produced the best regression results ($R^2 \approx 0.960$), followed by the ICMD-ANN-CLSTM, ICMD-RFR-LSTM, ICMD-RFR-CLSTM, LSTM, CLSTM, RFR and ANN models, with values of $R^2 \approx 0.951$, ≈ 0.949 , ≈ 0.798 , ≈ 0.760 , ≈ 0.759 , ≈ 0.757 , and ≈ 0.727 , respectively for Duffield Road sub-station. Similarly, the proposed deep learning hybrid model has shown best regression result for the other three sub-stations. Also, in all four sub-stations the decomposition based deep learning hybrid benchmark model shows better performance than the deep learning hybrid and standalone models. Notably, when compared to the other benchmark models, the proposed deep learning hybrid (ICMD-ANN-EDLSTM) model produced the best results by exhibiting the closest line of fit (1:1, red regression line in Fig. 7). In comparison to ICMD-ANN-EDLSTM, the ANN and RFR produced inferior results due to excessively distributed scatter points around the regression

Table 5

The testing performance of the proposed Deep Hybrid ICMD-ANN-EDLSTM model vs. benchmark models as measured by Root Mean Square Error ($RMSE$, MWh) and Mean Absolute Error (MAE , MWh).

Sub-stations	Predictive model	Model performance metrics	
		$RMSE$ (MWh)	MAE (MWh)
Duffield Road	ICMD-ANN-EDLSTM	9.3993	6.7758
	ICMD-ANN-CLSTM	13.09	9.771
	ICMD-RFR-CLSTM	13.494	10.156
	ICMD-RFR-LSTM	10.3	7.6376
	CLSTM	20.453	15.433
	LSTM	20.458	15.526
	ANN	20.382	15.599
	RFR	20.729	15.615
Kedron	ICMD-ANN-EDLSTM	9.738	6.6901
	ICMD-ANN-CLSTM	10.799	7.7203
	ICMD-RFR-CLSTM	21.955	15.541
	ICMD-RFR-LSTM	11.047	7.7997
	CLSTM	23.99	17.816
	LSTM	23.976	17.546
	ANN	25.563	18.237
	RFR	24.106	17.605
Kirra	ICMD-ANN-EDLSTM	3.5145	2.3248
	ICMD-ANN-CLSTM	16.739	13.877
	ICMD-RFR-CLSTM	4.1384	2.6524
	ICMD-RFR-LSTM	4.1349	2.9442
	CLSTM	14.043	10.401
	LSTM	11.299	8.4958
	ANN	9.9397	7.2456
	RFR	9.5123	6.8965
Molendinar	ICMD-ANN-EDLSTM	17.042	11.265
	ICMD-ANN-CLSTM	30.564	22.956
	ICMD-RFR-CLSTM	27.668	19.288
	ICMD-RFR-LSTM	18.636	12.452
	CLSTM	31.461	21.651
	LSTM	31.126	20.839
	ANN	33.939	23.498
	RFR	32.184	21.809

line. Furthermore, throughout testing, the RFR model produced the worst goodness-of-fit values for all four sub-stations. The decomposition based deep learning hybrid models (ICMD-ANN-CLSTM, ICMD-RFR-LSTM, and ICMD-RFR-CLSTM), deep learning hybrid model (CLSTM) and the standalone (LSTM) model are built on a DL algorithm that processes data via several layers, allowing these models to successfully learn the properties of the G time-series and remember long-term dependencies. ANN and RFR, on the other hand, are all shallow learning networks that do not typically use deep structures like DL. As a result, the performance of decomposition based deep learning hybrid models using DL outperformed these shallow learning network (ANN and RFR).

In order to illustrate more intuitively the prediction performance of the decomposition based deep learning hybrid model, the normalised and relative metrics (E_{WI} , E_{NS} , E_{LM} , $RRMSE$, and $RMAE$) are utilised. Based on performance results listed in Tables 6 and 7, With higher magnitude of E_{WI} , E_{NS} , and E_{LM} and lower magnitude of $RRMSE$, and $RMAE$ the proposed decomposition based deep learning hybrid model outperform the standalone (LSTM, ANN and RFR) and deep learning hybrid model (CLSTM) for all four sub-stations. At Kirra sub-station for proposed decomposition based deep learning hybrid model yielded $E_{WI} \approx 0.990$, $E_{NS} \approx 0.983$ and $E_{LM} \approx 0.889$ compared to $E_{WI} \approx 0.918$, $E_{NS} \approx 0.868$ and $E_{LM} \approx 0.655$ for ANN model. Similarly, for relative errors, the proposed model yielded $RRMSE \leq 4\%$ and $RMAE \leq 3\%$ for all four sub-stations. The best $RRMSE$ ($\approx 1.66\%$) and $RMAE$ ($\approx 1.07\%$) were found for Kirra sub-station followed by Kedron, Molendinar and Duffield Road sub-stations. At Kirra sub-station, decomposition based deep learning hybrid benchmark models (ICMD-RFR-CLSTM and ICMD-RFR-LSTM) showed excellent performance with $RRMSE \approx 1.96\%$. This indicates that the predictability of the G prediction model can be significantly improved by using ICMD to decompose original G time-series. The ICMD algorithm is capable of reducing the nonstationary

Table 6

The performance of the Deep Hybrid ICMD-ANN-EDLSTM model vs. benchmark models using the Willmott's Index (E_{WI}), Nash-Sutcliffe Coefficient (E_{NS}), and the Legates & McCabe's (E_{LM}) Index of Agreement. Note that the best model is boldfaced (blue).

Sub-stations	Predictive model	Model performance metrics		
		E_{WI}	E_{NS}	E_{LM}
Duffield Road	ICMD-ANN-EDLSTM	0.957	0.930	0.751
	ICMD-ANN-CLSTM	0.912	0.863	0.642
	ICMD-RFR-CLSTM	0.907	0.855	0.627
	ICMD-RFR-LSTM	0.945	0.915	0.720
	CLSTM	0.739	0.667	0.434
	LSTM	0.714	0.669	0.430
	ANN	0.755	0.669	0.428
	RFR	0.734	0.658	0.427
Kedron	ICMD-ANN-EDLSTM	0.970	0.960	0.823
	ICMD-ANN-CLSTM	0.962	0.951	0.795
	ICMD-RFR-CLSTM	0.848	0.799	0.588
	ICMD-RFR-LSTM	0.961	0.949	0.793
	CLSTM	0.787	0.760	0.527
	LSTM	0.820	0.760	0.535
	ANN	0.800	0.731	0.516
	RFR	0.817	0.757	0.533
Kirra	ICMD-ANN-EDLSTM	0.990	0.983	0.889
	ICMD-ANN-CLSTM	0.731	0.626	0.340
	ICMD-RFR-CLSTM	0.986	0.977	0.874
	ICMD-RFR-LSTM	0.986	0.977	0.860
	CLSTM	0.749	0.736	0.505
	LSTM	0.878	0.829	0.596
	ANN	0.918	0.868	0.655
	RFR	0.924	0.879	0.672
Molendinar	ICMD-ANN-EDLSTM	0.962	0.927	0.772
	ICMD-ANN-CLSTM	0.852	0.766	0.535
	ICMD-RFR-CLSTM	0.892	0.808	0.609
	ICMD-RFR-LSTM	0.951	0.913	0.748
	CLSTM	0.851	0.752	0.561
	LSTM	0.862	0.757	0.578
	ANN	0.840	0.712	0.524
	RFR	0.845	0.740	0.558

pattern that was present in the original G time-series. The prediction accuracy can be enhanced by decomposing nonstationary original G time-series into many reasonably stationary IMF components. Furthermore, the proposed model (ICMD-ANN-EDLSTM) outperforms the other three decomposition based deep learning hybrid models (ICMD-ANN-CLSTM, ICMD-RFR-CLSTM, and ICMD-RFR-LSTM), which use identical prediction methods to predict all components. This demonstrates the efficacy of combining various prediction approaches in predicting the IMF components generated by ICMD. Thus, by taking into account the various characteristics of each IMF component and employing various methodologies to predict each IMF component, the corresponding prediction performance can be further enhanced.

In terms of the Absolute Percentage Bias ($APB\%$) error calculated during the testing phase, Table 8 demonstrates that the proposed decomposition based deep learning hybrid model produces a lower percentage value of APB when compared to the other benchmark models, and this is less than 3% for all four sub-stations. For instance, the lowest value of $APB \approx 1.10\%$ is produced by the ICMD-ANN-EDLSTM model, while the highest value of $APB \approx 6.57\%$ is produced by the ICMD-ANN-CLSTM model for the Kirra sub-station. Accordingly, all study sites appear to achieve a Kling-Gupta Efficiency APB that is closer to unity compared to all the other benchmark models. In all four sub-stations the proposed decomposition based deep learning hybrid model (ICMD-ANN-EDLSTM) yielded $KGE \geq 0.96$. Notably, these values are higher than that of other deep learning hybrid and standalone benchmark models (see Table 8). Furthermore, the efficacy of the ICMD-ANN-EDLSTM model was assessed utilising the promoting percentage (λ) of $RMSE$, APB , and KGE . The assessment of λ_{RMSE} , λ_{APB} , and λ_{KGE} provides a significant improvement of the respective parameters compared with the benchmark models, as tabulated in Table 9. For instance,

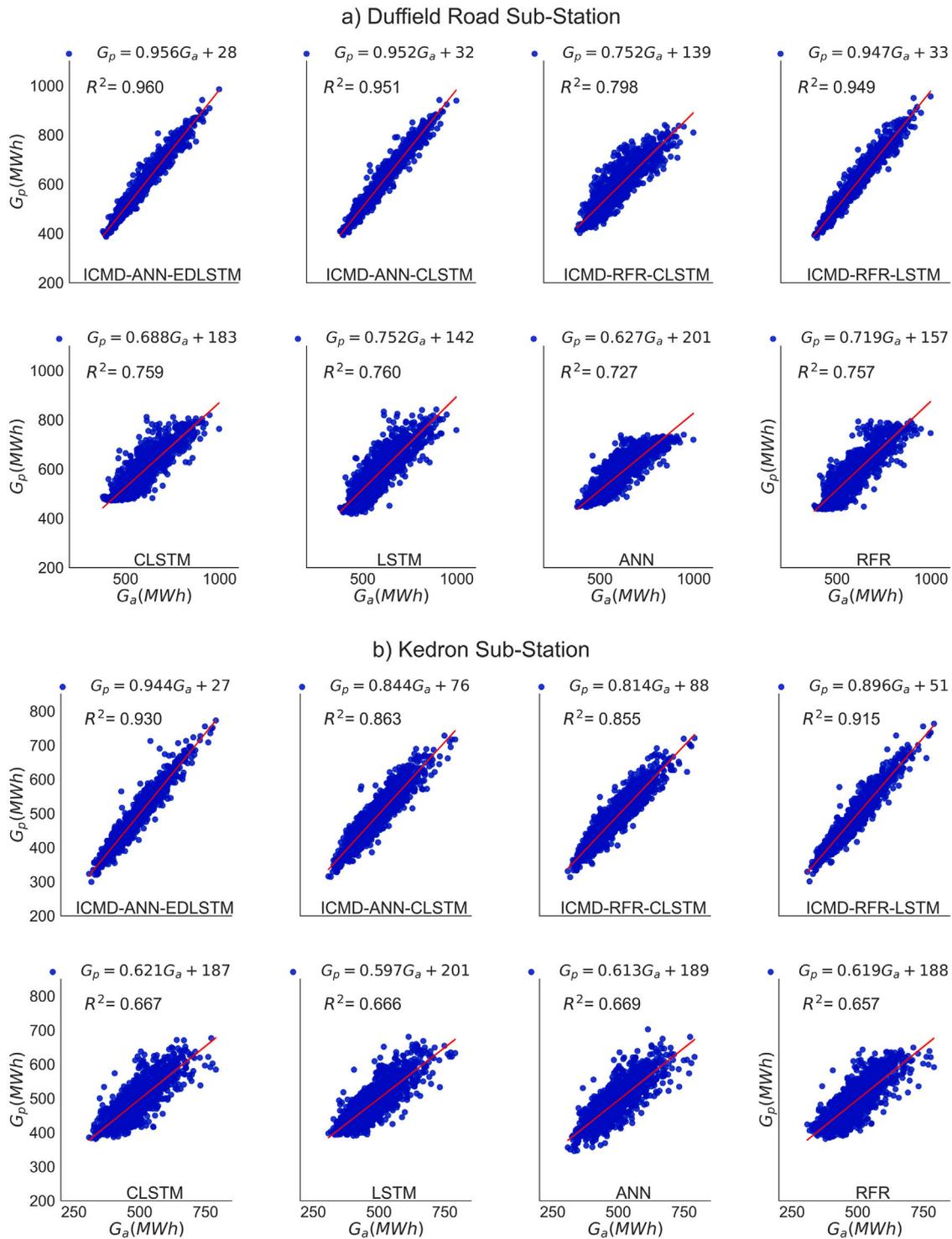


Fig. 7. Scatter plots of the actual vs. the predicted G values in the testing phase. the regression line and the R^2 values with regression equation are shown.

the RMSE of proposed model (ICMD-ANN-EDLSTM) was improved by $\approx 39\%$, $\approx 43\%$, $\approx 9\%$, $\approx 117\%$, $\approx 117\%$, $\approx 116\%$, and $\approx 120\%$ when compared with ICMD-ANN-CLSTM, ICMD-RFR-CLSTM, ICMD-RFR-LSTM, CLSTM, LSTM, ANN, and RFR respectively for Duffield Road sub-station. Surprisingly, $APB(\%)$ and $KGE(\%)$ showed a similar performance ranging from is 12 to 130% and 4 to 25%, respectively for Duffield Road sub-station. Therefore, the APB, KGE , and λ results of proposed decomposition based deep learning hybrid (ICMD-ANN-EDLSTM) model is consistent with the previous results (Tables 5, 6,

7 and Fig. 7) and substantiate its applicability for daily electricity demand prediction.

Fig. 8 shows the comparison of models absolute Prediction Error ($|PE|$) utilising box plots in the testing phase to further ascertain the enhanced results of the suggested decomposition based deep learning hybrid model. Outliers of the extreme $|PE|$ of the testing data are represented by the '+' symbol. First quartile (25th percentile) and third quartile (75th percentile) were indicated by the lower and upper lines of the boxplot, while the median (50th percentile) was indicated by the central line. The smallest and largest non-outliers were defined by

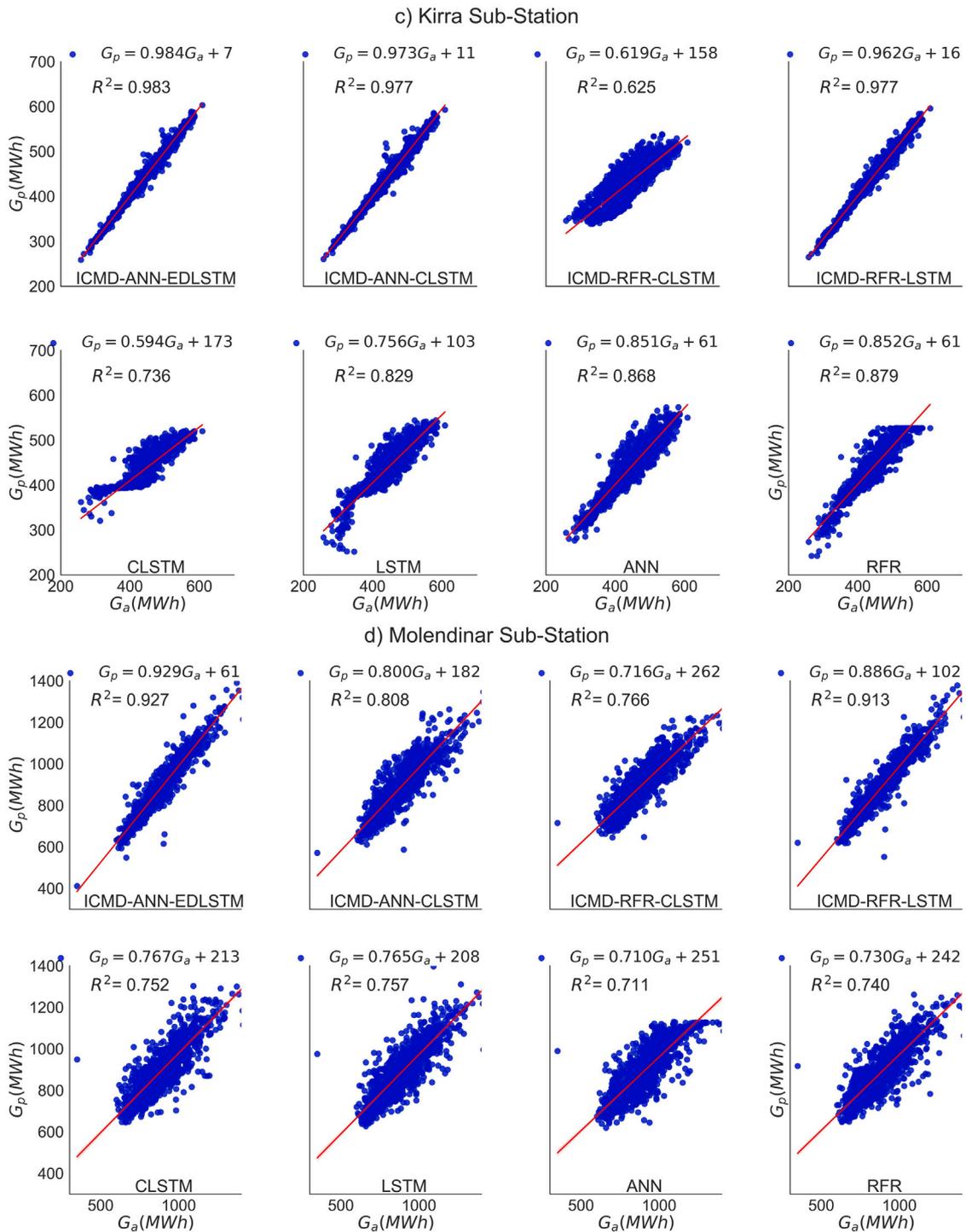


Fig. 7. (continued).

two horizontal lines, respectively, between the first and third quartiles. In accordance with the previous results (Tables 5, 6, 7, 8, 9 and Fig. 8), the smaller distribution of $|PE|$ with a smaller quartile range and therefore had better accuracy. Evidently, the proposed ICMD-ANN-EDLSTM yielded overall better performance for daily G prediction for the four sub-stations, followed by the other decomposition based deep learning hybrid models (ICMD-ANN-CLSTM, ICMD-RFR-CLSTM, and ICMD-RFR-LSTM).

To gain a more in-depth knowledge of the proposed decomposition based deep learning hybrid model, Fig. 9 illustrates a thorough interpretation by plotting the GPI values. It is revealed from Fig. 9 that the decomposition based deep learning hybrid models have comparatively

higher GPI value compared to deep learning hybrid and standalone models. In only one sub-station (Kirra) the GPI value for the ICMD-RFR-CLSTM ($GPI \approx -2.876$) model have lower value than that of the CLSTM (≈ -1.564), LSTM (≈ -0.62), ANN ($\approx -0.07\%$) and RFR (≈ -0.08) models. This lower value in GPI for ICMD-RFR-CLSTM was because of the high $RMSE$ (≈ 16.739 MWh) value of this model compared to other models.

In terms of GPI the proposed model:ICMD-ANN-EDLSTM has out-perform all other benchmark models. For instance, at Kedron substation the $GPI \approx 2.33, \approx 2.12, \approx 2.09, \approx 0.211, \approx -0.211, \approx -0.227, \approx -0.245,$ and ≈ -0.412 for ICMD-ANN-EDLSTM, ICMD-ANN-CLSTM, ICMD-RFR-LSTM, ICMD-RFR-CLSTM, LSTM, RFR, CLSTM, and ANN, respectively.

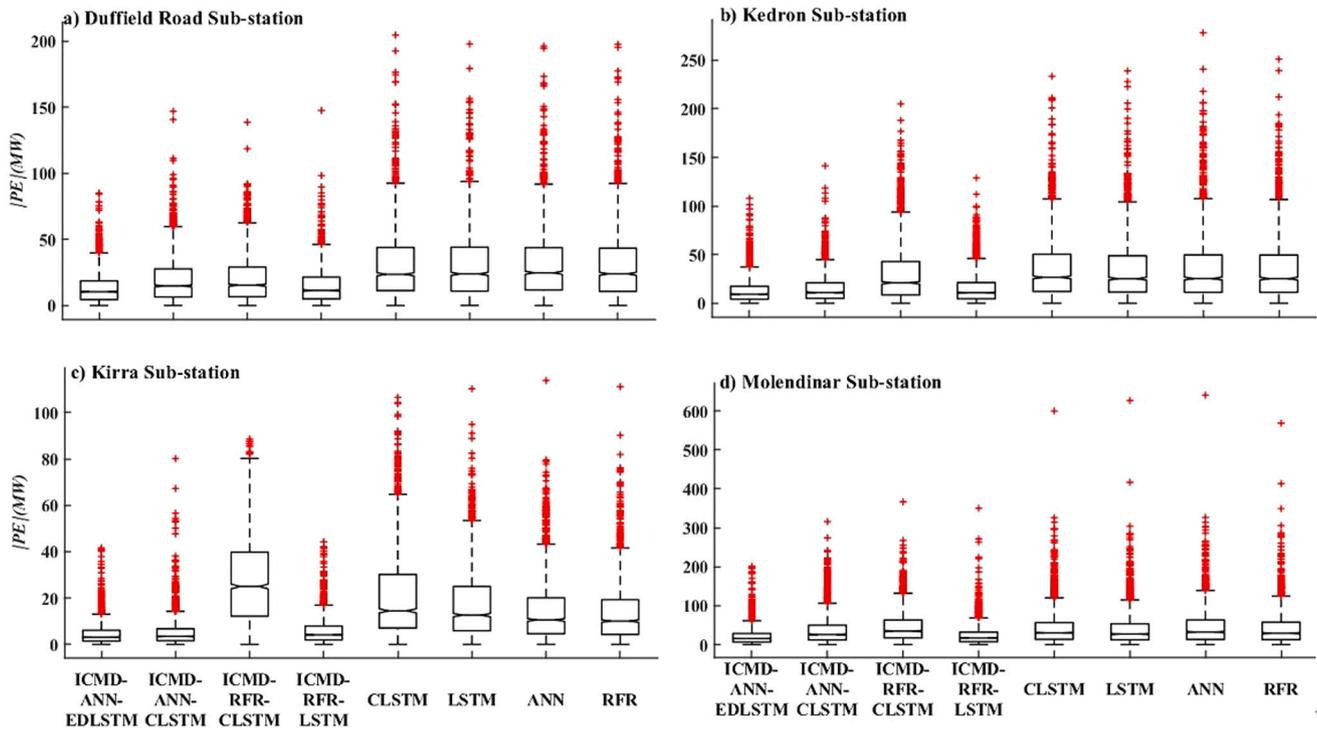


Fig. 8. Models performance comparison using box plots of $|PE|$ during the training period.

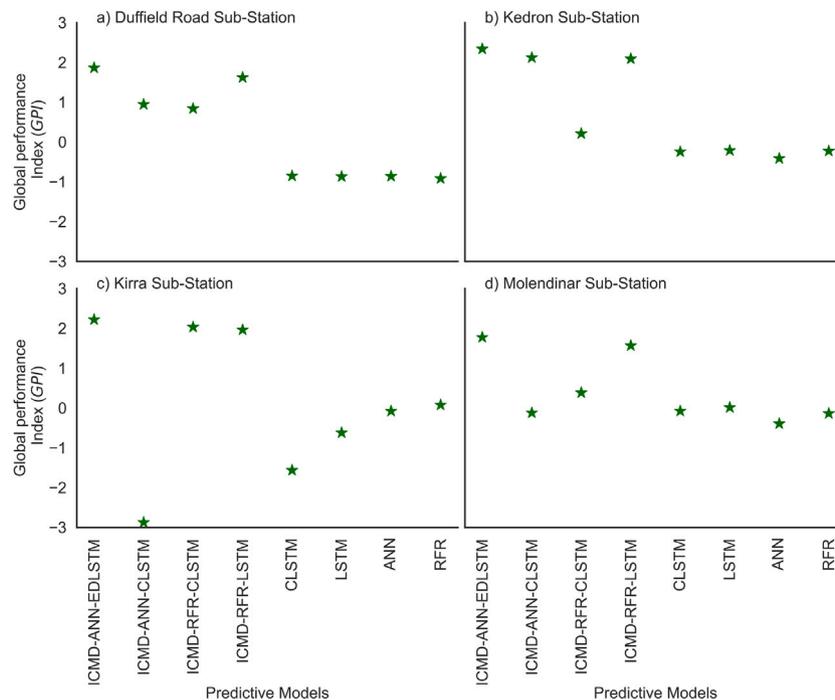


Fig. 9. Global Performance Indicator (GPI) used to evaluate the proposed ICMD-ANN-EDLSTM model relative to seven other benchmarked models.

Furthermore, condensing the benchmarking and comparing findings into one unique and correct parameter may be problematic; yet, a unique metric termed Combined Performance Index ($CPI\%$) was also employed in this study to demonstrate the prediction ability of the proposed model. CPI is defined as a weighted sum of numerous measures that integrate dispersion and distribution function similitude information ($CPI = (KSI + OVER + 2RMSE)/4$) where, KSI is Kolmogorov-Smirnov test Integral and $OVER$ is the relative frequency of exceedance. It should be also noted that the lower the value of

CPI the better the model is. The more information regarding this CPI metrics can be found in Ref. [107].

Fig. 10 shows the bar chart of the CPI value for each model at four sub-stations. The CPI bar chart shows that the proposed decomposition based deep learning hybrid model:ICMD-ANN-EDLSTM has yielded lowest CPI values for all four substation. For instance, at Molendinar sub-station the $CPI \approx 3.39\%$, $\approx 6.20\%$, $\approx 5.62\%$, $\approx 4.08\%$, $\approx 6.73\%$, $\approx 6.90\%$, and $\approx 6.13\%$ for ICMD-ANN-EDLSTM, ICMD-ANN-CLSTM,

Table 7

The geographic comparison of the Deep Hybrid ICMD-ANN-EDLSTM model vs. other comparative models in terms of the relative errors (*RRMSE*, %) and (*RMAE*, %) computed within the test sites. Note that the best model is boldfaced (blue).

Sub-stations	Predictive model	Model performance metrics	
		<i>RRMSE</i>	<i>RMAE</i>
Duffield Road	ICMD-ANN-EDLSTM	3.89%	2.82%
	ICMD-ANN-CLSTM	5.41%	4.01%
	ICMD-RFR-CLSTM	5.58%	4.15%
	ICMD-RFR-LSTM	4.26%	3.17%
	CLSTM	8.45%	6.51%
	LSTM	8.46%	6.60%
	ANN	8.43%	6.49%
	RFR	8.57%	6.60%
Kedron	ICMD-ANN-EDLSTM	3.42%	2.32%
	ICMD-ANN-CLSTM	3.79%	2.70%
	ICMD-RFR-CLSTM	7.70%	5.28%
	ICMD-RFR-LSTM	3.88%	2.69%
	CLSTM	8.42%	6.23%
	LSTM	8.41%	6.04%
	ANN	8.97%	6.06%
	RFR	8.46%	6.05%
Kirra	ICMD-ANN-EDLSTM	1.66%	1.07%
	ICMD-ANN-CLSTM	7.92%	6.68%
	ICMD-RFR-CLSTM	1.96%	1.21%
	ICMD-RFR-LSTM	1.96%	1.36%
	CLSTM	6.65%	5.11%
	LSTM	5.35%	4.05%
	ANN	4.71%	3.35%
	RFR	4.50%	3.21%
Molendinar	ICMD-ANN-EDLSTM	3.81%	2.53%
	ICMD-ANN-CLSTM	6.83%	5.28%
	ICMD-RFR-CLSTM	6.18%	4.33%
	ICMD-RFR-LSTM	4.16%	2.83%
	CLSTM	7.03%	4.95%
	LSTM	6.96%	4.74%
	ANN	7.58%	5.28%
	RFR	7.19%	4.96%

ICMD-RFR-CLSTM, ICMD-RFR-LSTM, CLSTM, LSTM, ANN, and RFR respectively. According to the results of the preceding *GPI* and *CPI* metric analysis, the objective mode: ICMD-ANN-EDLSTM outperforms the decomposition based deep learning hybrid benchmark models (ICMD-ANN-CLSTM, ICMD-RFR-LSTM, and ICMD-RFR-CLSTM), deep learning hybrid model (CLSTM) and the standalone models (LSTM, RFR, and ANN) in addressing nonlinear issues, and it is more suited for daily *G* prediction. To estimate the performance of the constructed hybrid forecasting system, a more scientific and thorough evaluation is carried out. The Taylor diagram, Diebold–Mariano (*DM*) and Harvey, Leybourne and Newbold (*HLN*) testing is used to examine the prediction error for each model.

In Fig. 11, the Taylor diagram, which calculates the angular location of the inverse cosine of the correlation coefficient, is shown to demonstrate the model that is closest to the actual data during the testing period. In Fig. 11 it can be seen that, the correlation coefficient (*r*) on the radial axis and the Standard Deviation (*SD*) on the polar axis are used combined to adapt the model with the best fit to the predictors. For all prediction scenarios, the decomposition based deep learning hybrid model: ICMD-ANN-EDLSTM with the highest *r* value produced the closest prediction to the actual *G* data for all our sub-stations. Again, the modelled data generated by the other decomposition based deep learning hybrid models (ICMD-ANN-CLSTM, ICMD-RFR-LSTM, and ICMD-RFR-CLSTM), deep learning hybrid model (CLSTM) and standalone (LSTM, RFR, and ANN) differed significantly from that generated by the ICMD-ANN-EDLSTM model.

Furthermore, Table 10 shows the *DM* and *HLN* statistical index table of the proposed model: ICMD-ANN-EDLSTM with each comparative model in order to evaluate the efficacy of the proposed model. From Table 10, the proposed model: ICMD-ANN-EDLSTM, performs better

Table 8

The testing performance of the Deep Hybrid ICMD-ANN-EDLSTM model vs. benchmark models as measured by Kling Gupta Efficiency (*KGE*), and Absolute Percentage Bias (*APB*). Note that the best model is boldfaced (blue).

Sub-stations	Predictive model	Model performance metrics	
		<i>KGE</i>	<i>APB</i>
Duffield Road	ICMD-ANN-EDLSTM	0.972	2.80%
	ICMD-ANN-CLSTM	0.9011	4.04%
	ICMD-RFR-CLSTM	0.8735	4.20%
	ICMD-RFR-LSTM	0.9324	3.16%
	CLSTM	0.7394	6.38%
	LSTM	0.7068	6.42%
	ANN	0.7293	6.45%
	RFR	0.7414	6.45%
Kedron	ICMD-ANN-EDLSTM	0.9727	2.35%
	ICMD-ANN-CLSTM	0.9711	2.71%
	ICMD-RFR-CLSTM	0.8312	5.45%
	ICMD-RFR-LSTM	0.9686	2.74%
	CLSTM	0.7747	6.25%
	LSTM	0.848	6.16%
	ANN	0.708	6.40%
	RFR	0.8121	6.18%
Kirra	ICMD-ANN-EDLSTM	0.9908	1.10%
	ICMD-ANN-CLSTM	0.7551	6.57%
	ICMD-RFR-CLSTM	0.9833	1.26%
	ICMD-RFR-LSTM	0.9727	1.39%
	CLSTM	0.6665	4.92%
	LSTM	0.8208	4.02%
	ANN	0.9064	3.43%
	RFR	0.9023	3.26%
Molendinar	ICMD-ANN-EDLSTM	0.9603	2.52%
	ICMD-ANN-CLSTM	0.8035	5.13%
	ICMD-RFR-CLSTM	0.878	4.31%
	ICMD-RFR-LSTM	0.9232	2.78%
	CLSTM	0.8651	4.84%
	LSTM	0.8614	4.66%
	ANN	0.8205	5.25%
	RFR	0.8323	4.87%

since the *DM* and *HLN* test statistics are positive. When compared to benchmark models, the proposed model: ICMD-ANN-EDLSTM had higher prediction accuracy. As a result, our proposed method is a useful tool for daily *G* prediction.

5. Discussion, limitations and future research work

Based on the results presented so far, it is apparent that the proposed ICMD-ANN-EDLSTM model is considerably superior in respect to its competing counterpart models used in daily electricity demand predictions. The practical implications of this method is also clear. Energy-company decision-makers and environmental professionals can benefit from employing the ICMD-ANN-EDLSTM model as a solid framework for performing accurate predictions, which can aid in the effective management and planning of energy and environmental resources. For diverse prediction applications, the developed models can simulate the non-stationary and nonlinear properties of solar radiation, wind speed, streamflow, and air pollutants.

Despite the rational performances provided by the ICMD-ANN-EDLSTM model and the benchmark models, there are certain limitation worth addressing. As a result, additional research is required to assess the generalisation capabilities of the ICMD-ANN-EDLSTM model for the *G* prediction. However, the current study only included data from four sub-stations located at SEQ, Australia, and the prediction capacity of ICMD-ANN-EDLSTM models for additional locations was not evaluated. Consideration of electricity demand data from other sub-stations located at different state of Australia may be beneficial in validating the proposed models' broad-scale applicability.

Additionally, this study focused on the *G* time-series on a daily scale. The discussion can be broadened in the future to include the

Table 9

The promoting percentage metric, λ for the comparison models against objective (i.e., ICMD-ANN-EDLSTM) model in the testing phase. Note that λ_{RMSE} = Promoting Percentages of the Root Mean Square Error, λ_{KGE} = Promoting Percentages of Kling Gupta Efficiency, and λ_{APB} = Promoting Percentages of Absolute Percentage Bias.

Predictive models	Duffield Road sub-station			Kedron sub-station			Kirra sub-station			Molendinar sub-station		
	λ_{RMSE}	λ_{KGE}	λ_{APB}	λ_{RMSE}	λ_{KGE}	λ_{APB}	λ_{RMSE}	λ_{KGE}	λ_{APB}	λ_{RMSE}	λ_{KGE}	λ_{APB}
ICMD-ANN-CLSTM	39.3%	7.3%	44.2%	10.9%	0.2%	15.4%	376.3%	23.8%	496.9%	79.3%	16.3%	103.8%
ICMD-RFR-CLSTM	43.6%	10.1%	49.9%	125.5%	14.5%	132.3%	17.8%	0.8%	14.1%	62.4%	8.6%	71.2%
ICMD-RFR-LSTM	9.6%	4.1%	12.7%	13.4%	0.4%	16.6%	17.7%	1.8%	26.6%	9.4%	3.9%	10.5%
CLSTM	117.6%	23.9%	127.8%	146.4%	20.4%	166.3%	299.6%	32.7%	347.4%	84.6%	9.9%	92.2%
LSTM	117.7%	27.3%	129.1%	146.2%	12.8%	162.3%	221.5%	17.2%	265.5%	82.6%	10.3%	85.0%
ANN	116.8%	25.0%	130.2%	162.5%	27.2%	172.6%	182.8%	8.5%	211.7%	99.1%	14.6%	108.6%
RFR	120.5%	23.7%	130.4%	147.5%	16.5%	163.1%	170.7%	8.9%	196.7%	88.9%	13.3%	93.6%

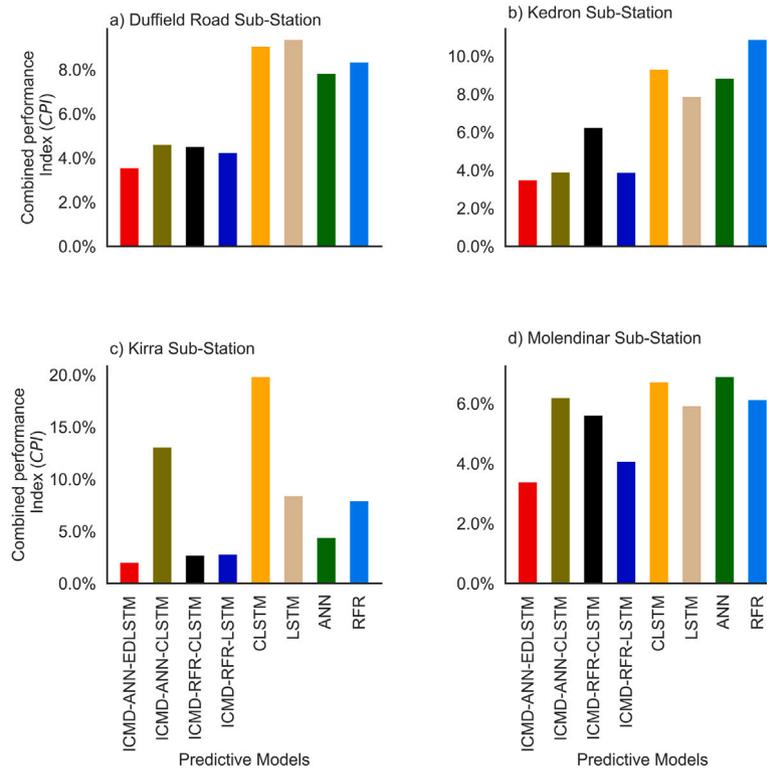


Fig. 10. Bar chart showing the Combined Performance Index of the proposed deep learning hybrid ICMD-ANN-EDLSTM model vs. seven other benchmark models.

Table 10

Evaluation of the Deep Hybrid Fused Network (FNET) model against comparison models in terms of: (a) The Diebold–Mariano (*DM*) test statistic, (b) The Harvey–Leybourne–Newbold (*HLN*) test statistic. Note: - 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. Note that the best model is boldfaced (blue).

(a)	ICMD-ANN-EDLSTM	ICMD-ANN-CLSTM	ICMD-RFR-CLSTM	ICMD-RFR-LSTM	CLSTM	LSTM	ANN	RFR
ICMD-ANN-EDLSTM		5.4272	10.9405	5.3796	14.5428	13.307	11.5596	12.8061
ICMD-ANN-CLSTM			6.5283	-4.6149	10.6105	9.8496	9.477	10.8175
ICMD-RFR-CLSTM				-10.5297	2.9215	1.9781	3.3929	2.3794
ICMD-RFR-LSTM					14.802	13.3312	11.591	12.8219
CLSTM						-2.0486	1.7196	-0.8564
LSTM							3.6848	1.0498
ANN								-2.5625
(b)	ICMD-ANN-EDLSTM	ICMD-ANN-CLSTM	ICMD-RFR-CLSTM	ICMD-RFR-LSTM	CLSTM	LSTM	ANN	RFR
ICMD-ANN-EDLSTM		5.4531	10.9928	5.4053	14.6122	13.3706	11.6148	12.8672
ICMD-ANN-CLSTM			6.5595	-4.637	10.6611	9.8967	9.5223	10.8691
ICMD-RFR-CLSTM				-10.58	2.9354	1.9875	3.4091	2.3907
ICMD-RFR-LSTM					14.8727	13.3949	11.6463	12.8831
CLSTM						-2.0584	1.7278	-0.8605
LSTM							3.7024	1.0548
ANN								-2.5748

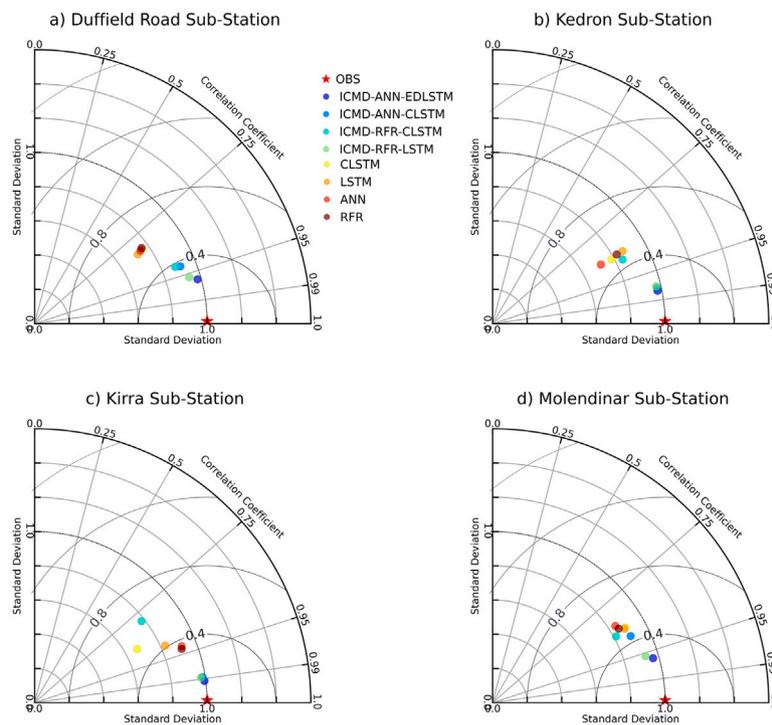


Fig. 11. Performance assessment of the deep learning hybrid ICMD-ANN-EDLSTM model as well as benchmark models for daily prediction of G in form of Taylor diagrams during the testing phase.

minute level time scale for strategic management and planning of renewable energy resources. Furthermore, one of the limitation of the decomposition based deep learning hybrid models is that it cannot properly balance prediction accuracy and computation time. The training time of standalone models (LSTM, RFR, and ANN) in this study is rather low, however the prediction accuracy is not good. Therefore, the model with a complex structure and numerous hyperparameters will definitely take a long time to train. Lastly, the prediction of electricity demand in this study was based on its previous demand values (univariate prediction method). In future study, we will incorporate some external elements into the prediction, such as weather conditions, and time index (such as day of the week and hour of the day).

In comparison with literature on decomposition methods for artificial intelligence and machine learning models, our study provides important advantages over previous studies (see Table 11). For example, The study of [108] developed ANN model with climatic variables for 6-h and daily G forecasting using Energex data for 8 stations in southeast Queensland, Australia. Although the study sites were different to the present study, we noted a relative $RMSE$ between 3.88–10.26% generated by their best (ANN) model compared to a significantly lower value of 1.66–3.89% for the present ICMD-ANN-EDLSTM model. This could perhaps be attributable to use of the more advanced EDLSTM method for predicting IMFs 6–10 and ANN method for predicting IMFs 1–5 whereas in [108], authors only used an ANN model without any decomposition of data.

In another recent study [18], authors used an integration of convolutional neural networks and echo state network (CESN) for daily electricity demand prediction at four sites (Roderick, Rocklea, Hemmant, Carpendale), in Southeast Queensland, Australia. The study found a relative $RMSE$ of 5.86–14.56%, which is also higher than the relative errors encountered in the present study. In this study, the CESN did not incorporate a data decomposition method, and therefore, this could have contributed to a lower performance of their hybrid model.

In other studies [109], authors develop a Maximum Overlap Discrete Wavelet Transform-Online Sequential Extreme Learning Machines Algorithm whereby wavelets was used to decompose daily

G data at three regional campuses (i.e., Toowoomba, Ipswich, and Springfield) at the University of Southern Queensland, Australia. The proposed MODWT-PACF-OS-ELM model was tested against the non-wavelet equivalent OS-ELM model to show that for all of the three datasets, a significantly greater accuracy was achieved with in a relative error of 4.31% vs. 11.31%, for the case of the Toowoomba as well as a similarly high performance for the other sites. This performance, however, remains lower than the proposed ICMD-ANN-EDLSTM model.

Finally, the study of [31] developed a hybrid two-phase particle swarm optimisation-support vector regression (PSO-SVR) model integrated with the improved version of empirical mode decomposition with adaptive noise multi-resolution tool for demand forecasting to show a significantly lower error of 2.01–4.65% (weekend forecast horizon) and 1.22–4.92% (whole week forecast horizon) for entire Queensland's state's aggregated demand dataset. It should be noted that this data was not station-based, and therefore, could have contained much less instabilities, patterns and trends due to the summing up of the entire state's electricity use compared with station-based data that are usually very diverse in its features. Therefore, these comparisons demonstrate the efficacy of the proposed hybrid ICMD-ANN-EDLSTM model for daily electricity demand modelling at a sub-station level.

The primary advantage of our method lies in its capability to better understand the characteristics of each of the time-series components in the G dataset using the ICMD method so that one can also tailor the prediction algorithm to better fit the model data prior to developing the predictive model. A combination of the ANN and the EDLSTM algorithms in this study has also allowed for better predictions since each algorithm has its own strengths and weaknesses. Furthermore, the combination of the two approaches with a data decomposition method, and perhaps, with the others deep learning algorithms, can also lead to a more robust predictive model that can handle more complex data sets and provide more accurate predictions of electricity demand.

Despite significant advantages in respect to performance efficacy, there still remains some room for improvement. This study relies solely on the G data, ignoring other factors such as weather, season, and socio-economic factors. Therefore in future, researchers could resolve this

Table 11

Comparison of the proposed hybrid ICMD-ANN-EDLSTM model with literature studies focusing on Queensland's electricity demand data. The relative error *RMSE* is shown only for the objective model in each case study.

Reference	Forecast horizon	Type of model	Error (% <i>RMSE</i>)
Present study	Daily; 2 stations	ICMD-ANN-EDLSTM vs. ICMD-ANN-CLSTM ICMD-RFR-CLSTM ICMD-RFR-LSTM CLSTM LSTM ANN RFR	1.66–8.57%
[108]	Daily; 8 stations	Hybrid ANN (bootstrap) vs. ANN MARS MLR ARIMA	7.38–11.43%
[18]	Daily; 4 stations	CNN with echo state network vs. SVR MLR XGB DNN LGB	5.86–11.77%
[109]	Daily; 3 stations	MODWT-PACF-OS-ELM (MPOE) vs. PACF-OS-ELM (POE)	5.47–7.49%
[31]	Various	ICEEMDAN-PSO-SVR vs. ICEEMDAN-M5 ICEEMDAN-MARS PSO-SVR M5 model tree MARS	Weekend: 2.01–5.03% Work days: 1.13–3.26% Whole week: 1.22–4.95% Public holiday: 2.97–8.62%

limitation of the current model by using several of these datasets to retrain the hybrid ICMD-ANN-EDLSTM model. In doing so, they need to adopt multivariate empirical mode decomposition (MEMD) used in studies other than electricity demand area [58,110]. The MEMD will provide an edge over the ICMD used given that ICMD operates on univariate data only. In terms of signal decomposition, the MEMD can decompose multivariate signal (e.g. *G*, weather variables, social presence, etc datasets) into several IMF groups, with each IMF group having the same length and components containing the same frequency distribution, in the same order of the group. This will enable the ANN and EDLSTM methods to be applied in a similar manner while capturing the patterns in weather as well as demand and other factors to retrain the model.

Furthermore, the present study considered only the point-based electricity demand prediction while ignoring the uncertainties and confidence intervals of these predictions. Since electricity industries are required to make important decisions, a probabilistic forecasting framework of the hybrid ICMD-ANN-EDLSTM model could be more appealing for confident decisions for the energy sector. Finally, the present hybrid ICMD-ANN-EDLSTM model still adopts a black-box approach, so it can be re-developed using multivariate data such as weather- or seasonal variables and improved using explainable and interpretable AI methods such as Shapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME) [111], as well as DeepSHAP, Deep Learning Important Features (DeepLIFT) [112] and Causal Explanations for Model Interpretation under Uncertainty (CXplain) [113] to help investigate the importance/contribution of each feature in predicting the *G* dataset.

Another important future study could use the proposed method with weather datasets in context of a real demand side management system to predict the electricity demand at local scales not tested in the present work, and especially for the different seasons where the inclusion of cooling and heating requirements could be expected to provide a more responsive model subject to demand of electricity. In such a circumstance, the proposed hybrid ICMD-ANN-EDLSTM

could use weather observations as well as modelled weather variables from European Centre for Medium Range Weather Forecasts (at multi-hourly scales) or other sources like the Global Forecast System (GFS), Global Ensemble Forecast System (GEFS), or other sources that provide near real-time predictions to aid in more accurate electricity demand predictions. Finally, as an independent study, researchers may adopt consumer electricity usage and socio-economic driven variables such as social and recreational events as lurking variables causing changes in daily electricity demand in different localities in re-training of the hybrid ICMD-ANN-EDLSTM models for accurate estimation of electricity demand in a real demand side management system.

6. Conclusions

Forecasting electricity demand (*G*, MWh) can assist in power generation planning and electricity power systems development. To estimate daily *G*, we have proposed a hybrid Artificial Neural Network (ANN)-Encoder-Decoder Long Short-Term Memory (EDLSTM) based on Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICMD), where the ICMD was adopted as a feature identification method to reveal the patterns in *G* dataset. The main contribution of this work is the proposed ICMD-ANN-EDLSTM model, which integrates components from both the ICMD method and the ANN-EDLSTM model. This approach allows a better capturing of the complex dynamics of daily *G* data. This way, the proposed model is able to make more accurate predictions of the daily *G* than either of the standalone models. The scientific contributions of this study are therefore highly relevant to energy industries currently facing significant challenges in predicting their electricity use, as well as spot selling price of electricity in face of renewables pumped into the grid as well as weather- or other socio-economic related factors.

In the first stage of the proposed approach, the original *G* time-series are decomposed into numerous Intrinsic Mode Functions (IMFs), which represent distinct patterns in *G* data. To create the input matrix for the proposed: ICMD-ANN-EDLSTM model, the Partial Autocorrelation

Function has been used to find the significant lagged variables using the decomposed IMF series of G data. In addition, the ANN model is used to predict the highest frequency components whereas the EDLSTM model was used for lower frequency IMFs and to calculate the optimal hyperparameters. To generate the final prediction result, the prediction findings of each IMF component are aggregated to attain the predicted G values.

Based on the daily electricity demand data from four substations in South-east Queensland, Australia, the proposed ICMD-ANN-EDLSTM model was evaluated. Additionally, seven other competing models were compared with the proposed model, which comprised of ICMD-ANN-CLSTM, ICMD-RFR-LSTM, ICMD-RFR-CLSTM, CLSTM, LSTM, ANN, and RFR. Regarding the numerical results obtained, the proposed hybrid ICMD-ANN-EDLSTM model yielded high accuracy and prediction stability based on the point prediction assessment metric, as well as the promoting percentages, Diebold–Mariano test, Harvey, Leybourne, and Newbold test, box plot, scatter plot, and Taylor diagram, which provided exhaustive analysis of the predicted G values. As an example, considering the comparison of the hybrid ICMD-ANN-EDLSTM with the hybrid ICMD-RFR-LSTM, ICMD-RFR-CLSTM, ICMD-ANN-CLSTM, LSTM, CLSTM, RFR, and ANN, the Promoting Percentage for Root Mean Square Error $RMSE$ were $\approx 9.4\%$, $\approx 62.4\%$, $\approx 79.3\%$, $\approx 82.6\%$, $\approx 84.6\%$, $\approx 88.9\%$, $\approx 88.9\%$, and $\approx 99.1\%$. Likewise, the Relative Mean Absolute Error ($RMSE$) were $\approx 2.82\%$, $\approx 4.01\%$, $\approx 4.15\%$, $\approx 3.17\%$, $\approx 6.51\%$, $\approx 6.60\%$, $\approx 6.49\%$, and $\approx 6.602\%$ for ICMD-ANN-EDLSTM, ICMD-RFR-LSTM, ICMD-RFR-CLSTM, ICMD-ANN-CLSTM, LSTM, CLSTM, RFR, and ANN, respectively (at Duffield Road sub-station).

In respect to the relative percentage error, the results demonstrated significantly smaller error values for the proposed hybrid ICMD-ANN-EDLSTM model for all four study sites (see Table 7). These ranged from $RRMSE$ of 1.66–3.81% and $RMSE$ of 1.07–2.82%. This compares with larger values for the other benchmark models as well as those in literature for Queensland-based datasets (see Table 11). Likewise, the other normalised metrics such as KGE and APB (Table 8) and WI , E_{NS} and E_{LM} are all higher than benchmark model for the proposed ICMD-ANN-EDLSTM model to further ascertain its efficacy in predicting the daily electricity demand for the four study sites.

Our proposed hybrid ICMD-ANN-EDLSTM model is significantly new as earlier studies did not employ two methods (i.e., an ANN followed by an EDLSTM model) to predict the patterns in electricity demand (see Fig. 6). The use of two-stage method of an ANN followed by an EDLSTM model to predict the low- and high frequency IMF signals generated from the G data is a novel contribution enabling the more subtle, yet important predictive features in electricity demand to be used for accurate modelling. In order to improve the prediction results, consideration of the characteristics of each component (IMFs) and the use of different prediction algorithms (ANN and EDLSTM) has provided a significant advantage over earlier studies.

CRedit authorship contribution statement

Sujan Ghimire: Writing – original draft, Conceptualization, Methodology, Software, Editing, Proofreading, Model development, Application. **Ravinesh C. Deo:** Writing, Editing, Proofreading. **David Casillas-Pérez:** Writing, Editing, Proofreading. **Sancho Salcedo-Sanz:** Writing, Editing, Proofreading.

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>).

Acknowledgement

This study has received partial support through the project PID2020-115454GB-C21, funded by the Spanish Ministry of Science and Innovation (MICINN).

References

- [1] Yong B, Xu Z, Shen J, Chen H, Tian Y, Zhou Q. Neural network model with Monte Carlo algorithm for electricity demand forecasting in Queensland. In: Proceedings of the Australasian computer science week multiconference. 2017, p. 1–7.
- [2] Jin Y, Gao X, Wang M. The financing efficiency of listed energy conservation and environmental protection firms: evidence and implications for green finance in China. *Energy Policy* 2021;153:112254.
- [3] Kavaklioglu K. Modeling and prediction of Turkey's electricity consumption using support vector regression. *Appl Energy* 2011;88(1):368–75.
- [4] Wang D, Yue C, ElAmraoui A. Multi-step-ahead electricity load forecasting using a novel hybrid architecture with decomposition-based error correction strategy. *Chaos Solitons Fractals* 2021;152:111453.
- [5] Zhao H, Han X, Guo S. DGM (1, 1) model optimized by MVO (multi-verse optimizer) for annual peak load forecasting. *Neural Comput Appl* 2018;30(6):1811–25.
- [6] Santhosh M, Venkaiah C, Kumar DV. Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction. *Energy Convers Manage* 2018;168:482–93.
- [7] Zhang X, Wang J, Gao Y. A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM. *Energy Econ* 2019;81:899–913.
- [8] Pappas SS, Ekonomou L, Karampelas P, Karamousantas D, Katsikas S, Chatzarakis G, Skafidas P. Electricity demand load forecasting of the Hellenic power system using an ARMA model. *Electr Power Syst Res* 2010;80(3):256–64.
- [9] Nepal B, Yamaha M, Yokoe A, Yamaji T. Electricity load forecasting using clustering and ARIMA model for energy management in buildings. *Japan Archit Rev* 2020;3(1):62–76.
- [10] Wang J, Zhang L, Li Z. Interval forecasting system for electricity load based on data pre-processing strategy and multi-objective optimization algorithm. *Appl Energy* 2022;305:117911.
- [11] Ghosh S. Univariate time-series forecasting of monthly peak demand of electricity in northern India. *Int J Indian Cult Bus Manage* 2008;1(4):466–74.
- [12] Liu C, Wu W-Z, Xie W, Zhang J. Application of a novel fractional grey prediction model with time power term to predict the electricity consumption of India and China. *Chaos Solitons Fractals* 2020;141:110429.
- [13] Takeda H, Tamura Y, Sato S. Using the ensemble Kalman filter for electricity load forecasting and analysis. *Energy* 2016;104:184–98.
- [14] Chen Y, Dong Z, Wang Y, Su J, Han Z, Zhou D, Zhang K, Zhao Y, Bao Y. Short-term wind speed predicting framework based on EEMD-GA-LSTM method under large scaled wind history. *Energy Convers Manage* 2021;227:113559.
- [15] Alobaidi MH, Chebana F, Meguid MA. Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. *Appl Energy* 2018;212:997–1012.
- [16] Behm C, Nolting L, Praktinjo A. How to model European electricity load profiles using artificial neural networks. *Appl Energy* 2020;277:115564.
- [17] Wang B, Zhang L, Ma H, Wang H, Wan S. Parallel LSTM-based regional integrated energy system multienergy source-load information interactive energy prediction. *Complexity* 2019;2019.
- [18] Ghimire S, Nguyen-Huy T, Al-Musaylh MS, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. A novel approach based on integration of convolutional neural networks and echo state network for daily electricity demand prediction. *Energy* 2023;127430.
- [19] Wang K, Huang SB, Ding YL. Application of GRNN neural network in short term load forecasting. In: *Advanced materials research*, Vol. 971. Trans Tech Publ; 2014, p. 2242–7.
- [20] Bin H, Zu YX, Zhang C. A forecasting method of short-term electric power load based on BP neural network. In: *Applied mechanics and materials*, Vol. 538. Trans Tech Publ; 2014, p. 247–50.
- [21] Qin M, Li Z, Du Z. Red tide time series forecasting by combining ARIMA and deep belief network. *Knowl-Based Syst* 2017;125:39–52.
- [22] Mughees N, Mohsin SA, Mughees A. Deep sequence to sequence bi-LSTM neural networks for day-ahead peak load forecasting. *Expert Syst Appl* 2021;175:114844.
- [23] Lee JY, Kim S. Forecasting daily peak load by time series model with temperature and special days effect. *Korean J Appl Statist* 2019;32(1):161–71.
- [24] El Desouky A, El Kateb M. Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA. *IEE Proc, Gener Transm Distrib* 2000;147(4):213–7.
- [25] Cai M, Pipattanasomporn M, Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Appl Energy* 2019;236:1078–88.

- [26] Muzaffar S, Afshari A. Short-term load forecasts using LSTM networks. *Energy Procedia* 2019;158:2922–7.
- [27] Cui Z, Wu J, Lian W, Wang Y-G. A novel deep learning framework with a COVID-19 adjustment for electricity demand forecasting. *Energy Rep* 2023;9:1887–95.
- [28] Tak H, Kim T, Cho H-G, Kim H. A new prediction model for power consumption with local weather information. *J Korea Contents Assoc* 2016;16(11):488–98.
- [29] Chen Y, Xu P, Chu Y, Li W, Wu Y, Ni L, Bao Y, Wang K. Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings. *Appl Energy* 2017;195:659–70.
- [30] Liu D, Yang Q, Yang F. Predicting building energy consumption by time series model based on machine learning and empirical mode decomposition. In: 2020 5th IEEE international conference on big data analytics (ICBDA). IEEE; 2020, p. 145–50.
- [31] 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.
- [32] 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.
- [33] Ghimire S, Nguyen-Huy T, Prasad R, Deo RC, Casillas-Pérez D, Salcedo-Sanz S, Bhandari B. Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction. *Cogn Comput* 2022;1–27.
- [34] Lai CS, Mo Z, Wang T, Yuan H, Ng WW, Lai LL. Load forecasting based on deep neural network and historical data augmentation. *IET Gener Transm Distrib* 2020;14(24):5927–34.
- [35] Liu C, Sun B, Zhang C, Li F. A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine. *Appl Energy* 2020;275:115383.
- [36] Wu C, Li J, Liu W, He Y, Nourmohammadi S. Short-term electricity demand forecasting using a hybrid ANFIS–ELM network optimised by an improved parasitism–predation algorithm. *Appl Energy* 2023;345:121316.
- [37] Li A, Xiao F, Zhang C, Fan C. Attention-based interpretable neural network for building cooling load prediction. *Appl Energy* 2021;299:117238.
- [38] Wang Z, Chen Z, Yang Y, Liu C, Li X, Wu J. A hybrid autoformer framework for electricity demand forecasting. *Energy Rep* 2023;9:3800–12.
- [39] Li L, Meinrenken CJ, Modi V, Culligan PJ. Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features. *Appl Energy* 2021;287:116509.
- [40] Kim J, Oh S, Kim H, Choi W. Tutorial on time series prediction using 1D-CNN and BiLSTM: A case example of peak electricity demand and system marginal price prediction. *Eng Appl Artif Intell* 2023;126:106817.
- [41] An N, Zhao W, Wang J, Shang D, Zhao E. Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. *Energy* 2013;49:279–88.
- [42] 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.
- [43] Zheng H, Yuan J, Chen L. Short-term load forecasting using EMD-LSTM neural networks with a xgboost algorithm for feature importance evaluation. *Energies* 2017;10(8):1168.
- [44] Ni K, Wang J, Tang G, Wei D. Research and application of a novel hybrid model based on a deep neural network for electricity load forecasting: a case study in Australia. *Energies* 2019;12(13):2467.
- [45] Jin Y, Guo H, Wang J, Song A. A hybrid system based on LSTM for short-term power load forecasting. *Energies* 2020;13(23):6241.
- [46] Yu S, Wang K, Wei Y-M. A hybrid self-adaptive particle swarm optimization–genetic algorithm–radial basis function model for annual electricity demand prediction. *Energy Convers Manage* 2015;91:176–85.
- [47] Mostafavi ES, Mostafavi SI, Jaafari A, Hosseinpour F. A novel machine learning approach for estimation of electricity demand: An empirical evidence from thailand. *Energy Convers Manage* 2013;74:548–55.
- [48] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen N-C, Tung CC, Liu HH. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond Ser A Math Phys Eng Sci* 1998;454(1971):903–95.
- [49] Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal* 2009;1(01):1–41.
- [50] Yeh J-R, Shieh J-S, Huang NE. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Adv Adapt Data Anal* 2010;2(02):135–56.
- [51] Li X, Li C. Improved CEEMDAN and PSO-SVR modeling for near-infrared noninvasive glucose detection. *Comput Math Methods Med* 2016;2016.
- [52] Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2011, p. 4144–7.
- [53] Colominas MA, Schlotthauer G, Torres ME. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed Signal Process Control* 2014;14:19–29.
- [54] 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.
- [55] Ghimire S, Nguyen-Huy T, Deo RC, Casillas-Pérez 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.
- [56] 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.
- [57] Breiman L. Random forests. *Mach Learn* 2001;45(1):5–32.
- [58] 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.
- [59] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Routledge; 2017.
- [60] Guo W, Che L, Shahidehpour M, Wan X. Machine-learning based methods in short-term load forecasting. *Electr J* 2021;34(1):106884.
- [61] Ilseve E, Gol M. A comparative study on feature selection based improvement of medium-term demand forecast accuracy. In: 2019 IEEE milan powertech. IEEE; 2019, p. 1–6.
- [62] Krishnadas G, Kiprakis A. A machine learning pipeline for demand response capacity scheduling. *Energies* 2020;13(7):1848.
- [63] Yang A, Li W, Yang X. Short-term electricity load forecasting based on feature selection and least squares support vector machines. *Knowl-Based Syst* 2019;163:159–73.
- [64] Tiyasha T, Tung TM, Bhagat SK, Tan ML, Jawad AH, Mohtar WHMW, Yaseen ZM. Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models. *Mar Pollut Bull* 2021;170:112639.
- [65] Sharafati A, Khosravi K, Khosravinia P, Ahmed K, Salman SA, Yaseen ZM, Shahid S. The potential of novel data mining models for global solar radiation prediction. *Int J Environ Sci Technol* 2019;16(11):7147–64.
- [66] Bhagat SK, Tiyasha T, Tung TM, Mostafa RR, Yaseen ZM. Manganese (Mn) removal prediction using extreme gradient model. *Ecotoxicol Environ Saf* 2020;204:111059.
- [67] Ali M, Prasad R, Xiang Y, Yaseen ZM. Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. *J Hydrol* 2020;584:124647.
- [68] 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.
- [69] Fang Z, Yu X, Zeng Q. Random forest algorithm-based accurate prediction of chemical toxicity to tetrahymena pyriformis. *Toxicology* 2022;480:153325.
- [70] Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput* 2019;31(7):1235–70.
- [71] Lin Y, Yan Y, Xu J, Liao Y, Ma F. Forecasting stock index price using the CEEMDAN-LSTM model. *North Am J Econom Finance* 2021;57:101421.
- [72] Wang B, Wang J. Deep multi-hybrid forecasting system with random EWT extraction and variational learning rate algorithm for crude oil futures. *Expert Syst Appl* 2020;161:113686.
- [73] Bian C, He H, Yang S, Huang T. State-of-charge sequence estimation of lithium-ion battery based on bidirectional long short-term memory encoder-decoder architecture. *J Power Sources* 2020;449:227558.
- [74] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. 2014, arXiv preprint arXiv:1406.1078.
- [75] Lu K, Sun WX, Wang X, Meng XR, Zhai Y, Li HH, Zhang RG. Short-term wind power prediction model based on encoder-decoder LSTM. In: IOP conference series: earth and environmental science, Vol. 186. IOP Publishing; 2018, 012020.
- [76] Kosana V, Madasthu S, Teeparthi K. A novel hybrid framework for wind speed forecasting using autoencoder-based convolutional long short-term memory network. *Int Trans Electr Energy Syst* 2021;31(11):e13072.
- [77] Kosana V, Teeparthi K, Madasthu S. Hybrid convolutional Bi-LSTM autoencoder framework for short-term wind speed prediction. *Neural Comput Appl* 2022;1–10.
- [78] Tong J, Xie L, Fang S, Yang W, Zhang K. Hourly solar irradiance forecasting based on encoder–decoder model using series decomposition and dynamic error compensation. *Energy Convers Manage* 2022;270:116049.
- [79] Zhu K, Li Y, Mao W, Li F, Yan J. LSTM enhanced by dual-attention-based encoder-decoder for daily peak load forecasting. *Electr Power Syst Res* 2022;208:107860.
- [80] Lin J, Ma J, Zhu J, Cui Y. Short-term load forecasting based on LSTM networks considering attention mechanism. *Int J Electr Power Energy Syst* 2022;137:107818.

- [81] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, et al. Recent advances in convolutional neural networks. *Pattern Recognit* 2018;77:354–77.
- [82] Ghimire S, Nguyen-Huy T, Prasad R, Deo RC, Casillas-Pérez D, Salcedo-Sanz S, Bhandari B. Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction. *Cogn Comput* 2022;1–27.
- [83] 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.
- [84] Diehl PU, Neil D, Binas J, Cook M, Liu S-C, Pfeiffer M. Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In: 2015 international joint conference on neural networks (IJCNN). IEEE; 2015, p. 1–8.
- [85] Rahman MM, Shakeri M, Tiong SK, Khatun F, Amin N, Pasupuleti J, Hasan MK. Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks. *Sustainability* 2021;13(4):2393.
- [86] Ferrero Bermejo J, Gómez Fernández JF, Olivencia Polo F, Crespo Márquez A. A review of the use of artificial neural network models for energy and reliability prediction. A study of the solar PV, hydraulic and wind energy sources. *Appl Sci* 2019;9(9):1844.
- [87] Wu Z, Luo G, Yang Z, Guo Y, Li K, Xue Y. A comprehensive review on deep learning approaches in wind forecasting applications. *CAAI Trans Intell Technol* 2022;7(2):129–43.
- [88] Zhang Y, Li Y, Zhang G. Short-term wind power forecasting approach based on Seq2Seq model using NWP data. *Energy* 2020;213:118371.
- [89] Lai J-P, Chang Y-M, Chen C-H, Pai P-F. A survey of machine learning models in renewable energy predictions. *Appl Sci* 2020;10(17):5975.
- [90] Sanner MF, et al. Python: a programming language for software integration and development. *J Mol Graph Model* 1999;17(1):57–61.
- [91] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: Machine learning in python. *J Mach Learn Res* 2011;12:2825–30.
- [92] Ketkar N. Introduction to keras. In: *Deep learning with python*. Springer; 2017, p. 97–111.
- [93] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, et al. {TensorFlow}: a system for {large-scale} machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI 16). 2016, p. 265–83.
- [94] Oreshkin BN, Carпов D, Chapados N, Bengio Y. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. 2019, arXiv preprint arXiv:1905.10437.
- [95] Bergstra J, Komer B, Eliasmith C, Yamins D, Cox DD. Hyperopt: a python library for model selection and hyperparameter optimization. *Comput Sci Discov* 2015;8(1):014008.
- [96] Wu J, Chen X-Y, Zhang H, Xiong L-D, Lei H, Deng S-H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J Electron Sci Technol* 2019;17(1):26–40.
- [97] Castillo-Botón C, Casillas-Pérez D, Casanova-Mateo C, Ghimire S, Cerro-Prada E, Gutierrez P, Deo R, Salcedo-Sanz S. Machine learning regression and classification methods for fog events prediction. *Atmos Res* 2022;272:106157.
- [98] Deo RC, Grant RH, Webb A, Ghimire S, Igoe DP, Downs NJ, Al-Musaylh MS, Parisi AV, Soar J. Forecasting solar photosynthetic photon flux density under cloud cover effects: novel predictive model using convolutional neural network integrated with long short-term memory network. *Stoch Environ Res Risk Assess* 2022;1–38.
- [99] 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.
- [100] 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.
- [101] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I—A discussion of principles. *J Hydrol* 1970;10(3):282–90.
- [102] 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.
- [103] 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.
- [104] McKenzie J. Mean absolute percentage error and bias in economic forecasting. *Econom Lett* 2011;113(3):259–62.
- [105] 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.
- [106] Diebold FX, Mariano RS. Comparing predictive accuracy. *J Bus Econom Statist* 2002;20(1):134–44.
- [107] Gueymard CA. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. *Renew Sustain Energy Rev* 2014;39:1024–34.
- [108] 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.
- [109] 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.
- [110] Ali M, Deo RC, Maraseni T, Downs NJ. Improving SPI-derived drought forecasts incorporating synoptic-scale climate indices in multi-phase multivariate empirical mode decomposition model hybridized with simulated annealing and kernel ridge regression algorithms. *J Hydrol* 2019;576:164–84.
- [111] Pokharel S, Sah P, Ganta D. Improved prediction of total energy consumption and feature analysis in electric vehicles using machine learning and shapley additive explanations method. *World Electr Veh J* 2021;12(3):94.
- [112] Li J, Zhang C, Zhou JT, Fu H, Xia S, Hu Q. Deep-lift: deep label-specific feature learning for image annotation. *IEEE Trans Cybern* 2021;52(8):7732–41.
- [113] Schwab P, Karlen W. Cxplain: Causal explanations for model interpretation under uncertainty. *Adv Neural Inf Process Syst* 2019;32.