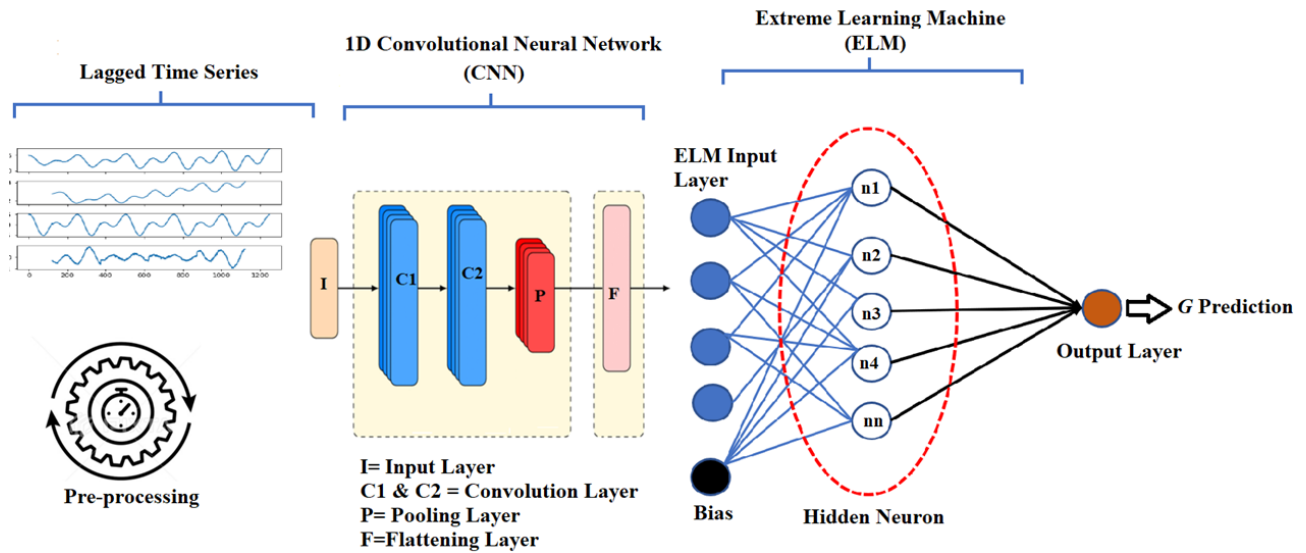


Graphical Abstract

Probabilistic-based Electricity Demand Forecasting with Hybrid Convolutional Neural Network-Extreme Learning Machine Model

Sujan Ghimire, Ravinesh C. Deo, David Casillas-Pérez, Sancho Salcedo-Sanz, S. Ali Pourmousavi, U. Rajendra Acharya



Highlights

Probabilistic-based Electricity Demand Forecasting with Hybrid Convolutional Neural Network-Extreme Learning Machine Model

Sujan Ghimire, Ravinesh C. Deo, David Casillas-Pérez, Sancho Salcedo-Sanz, S. Ali Pourmousavi, U. Rajendra Acharya

- A hybrid deep learning model for daily electricity demand prediction is proposed.
- A Kernel Density Estimation is applied to derive probabilistic prediction intervals.
- The algorithm quantifies uncertainties in point-based electricity demand predictions.
- The proposed model outperforms deep learning models using statistical uncertainty indicators.

Probabilistic-based Electricity Demand Forecasting with Hybrid Convolutional Neural Network-Extreme Learning Machine Model

Sujan Ghimire^a, Ravinesh C. Deo^{a,*}, David Casillas-Pérez^b, Sancho Salcedo-Sanz^{a,1}, S. Ali Pourmousavi^d, U. Rajendra Acharya^{a,1}

^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*

^d*The University of Adelaide, School of Electrical and Electronic Engineering, Australia*

^e*International Research Organization for Advanced Science and Technology, (IROAST), Kumamoto University, Kumamoto 860-8555, Japan*

Abstract

Implementing key engineering solutions to optimise the operation of energy industries requires daily electricity demand forecasting and including uncertainty, to promote markets insight analysis as part of their strategic planning, regulating and supplying electricity to consumers. This paper proposes hybrid artificial intelligence models combining convolutional neural networks (CNN) as a feature extraction algorithm with extreme learning machines (ELM) as a framework to predict electricity demand with confidence intervals generated by Kernel Density Estimation (KDE) approaches. In order to develop CELM-KDE model, time-lagged series of daily electricity demand with local climate variables based on the air temperature, atmospheric vapour pressure, evaporation, solar radiation, humidity and sea level pressures are used to train the proposed CELM-KDE hybrid model. In order to fully evaluate the newly developed model from a point-based, as well as a probabilistic prediction strategy, the observed and predicted electricity demand as well as the probability distribution of errors are analysed using KDE method that operates without prior data distribution assumptions. Based on observed and predicted electricity demand and the relevant probabilistic confidence intervals generated by the CELM-KDE model, the final results show that the proposed method attains significantly better probability interval predictions than traditionally-used point-based models. The proposed CELM-KDE model is demonstrated to be highly effective in providing a comprehensive coverage of predicted errors, as well as providing greater insights into the average bandwidth and detailed predicted electricity demand in the testing stage. The results also indicate that the proposed hybrid model is a reliable decision support tool to develop engineering solutions in area of energy modelling, monitoring and forecasting, which could potentially be useful to the industry policymakers. We show that the point-and probabilistic-based electricity demand predictive models can be employed as an effective tool to improve accuracy of forecasting and provision of insights for national electricity markets and key energy industry stakeholder application tools.

Keywords: Time-series forecasts, Deep Learning, Extreme Learning Machine, Convolutional Neural Network, Kernel Density Estimation, Prediction Interval

*Corresponding author: (Prof Ravinesh C Deo) ravinesh.deo@usq.edu.au

Email addresses: sujan.ghimire@usq.edu.au (Sujan Ghimire), ravinesh.deo@usq.edu.au (Ravinesh C. Deo), david.casillas@urjc.es (David Casillas-Pérez), sancho.salcedo@uah.es (Sancho Salcedo-Sanz), a.pourm@adelaide.edu.au (S. Ali Pourmousavi), rajendra.acharya@usq.edu.au (U. Rajendra Acharya)

1. Introduction

Electricity power system operators, utilities, and private investors alike have a strong interest in short-term accurate electricity demand forecasting which enables grid operation, bidding in the wholesale market, and planning maintenance [1]. Forecasts that are inaccurate or are over-predict electricity demand, as well as those that under-predict it, may cause blackouts and loss of loads in extreme cases. Predicting electricity demand (G) is therefore crucial for government officials, power producers, and distributors [2]. Because of the operational, tactical, and strategic consequences of G prediction, accurate demand prediction became even more crucial for electricity suppliers. The short-and medium-term effects are increased operating costs, management difficulties, security and stability issues [3], and a lack of sustainability in the power supply system [4, 5, 6]. Similarly, the long-term effects include a loss of market share and weakened competitiveness in the energy market [7]. This is because extra costs of installing, connecting, and maintaining distributed energy resources (DERs), such as solar panels and wind turbines, are significant, and the lack of stability and power security makes it difficult to manage the power supply systems efficiently. Additionally, when market share is lost, companies become less competitive in the energy market, leading to decreased profits. Therefore reliable predictive models for electricity demand are crucial to the successful operation of national electricity markets and profitability of the energy sector.

In general, there are three main methods for predicting the G : (i) econometrics and statistical techniques; (ii) traditional artificial intelligence (AI); and (iii) hybrid AI models, which incorporate models from econometrics, statistics and AI fields [8]. In the first category, well-known methods for modelling G include Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) and Autoregressive Integrated Moving Average (ARIMA) models [9, 10, 11]. Regarding traditional AI techniques, there are several types of machine learning (ML) frameworks devoted to this prediction problem, such as Artificial Neural Networks (ANN) [12, 13, 14, 15], Support Vector Regression (SVR) [16, 17, 18, 11, 19] and Decision Trees (DT) models [20] have been extensively adopted for related problems. Alternative to ML frameworks, recent advancements in deep learning (DL), an advanced form of ML developed to deliberately explore and model complex datasets, have found good applications in the G prediction problem. These DL models are therefore revealing promising results in respect to their capabilities and robustness to predict G [21, 22, 23]. Finally, hybrid approaches mixing different types of ML/DL and alternative approaches have shown excellent performance in the prediction of G . Through the integration of different algorithms, hybrid models can be created to capitalize on each algorithm's merits to improve the overall predictive performance. In the case of G time-series, spatial (station-based) as well as temporal information must be considered in order to obtain accurate results [24, 25].

In addition to the problem of predicting the spatially and temporally diverse, as well as chaotic G data series described above, a major limitation of existing studies pertains to the large focus on single 'point-based' electricity demand predictions, rather than the probability and confidence interval of such predictions, which is now required for day-to-day management. The past studies reviewed to date have yielded deterministic predictions without much information, if any, on the probability interval over which a predicted G value would lie, the spread of such predictions, and the level of uncertainty in the predicted value [26]. Due to the stochastic nature of G data time-series, point-based predictions become more uncertain and erratic as the level of G uncertainty increases, and this limitation of existing electricity demand models could become a critical challenge in managing and planning electricity supply. It is essential

to quantify these uncertainties in a probabilistic manner in order to reduce uncertainties in G estimations or at least better understand their sources or nature. As an alternative to point-based forecasting, the overall prediction interval (PI) presented in this paper could yield valuable insights regarding the likelihood of uncertainty in point predictions [27]. In general, the evaluation of PIs could consist of calculating the lower and the upper bounds of errors of predicted G and represented as a pre-determined probability, let's say $(1 - \alpha)\%$ denoted as the confidence level [28, 29]. As compared to only a point-based value, the PI associated with any predicted G is significantly beneficial to decision-makers in the energy industry. This is attributable to the fact that it not only provides a range within which a predicted electricity demand is likely to fall but also provides an $\alpha\%$ value representing a probability of its precision, which is often referred to as the confidence level.

In this study, we propose a hybrid approach formed by a CNN as a feature extraction tool, and an Extreme Learning Machine (ELM) prediction model to explore the historical patterns in G as well as climate datasets. The final hybrid approach is known as CELM model for daily prediction of point-based G data. We will then integrate the KDE approach to evaluate prediction errors in G , in order to quantify the likely uncertainties in point-based electricity demand predictions. The approach in this paper is also informed by an earlier study of Yu et al. [30] whereby a KDE-based model was developed with a fixed bandwidth to evaluate the distribution of wind power prediction errors although no prior study has integrated the KDE method with deep learning models for electricity demand prediction. A KDE approach, however, was utilised with the volatility forecast model in a study by Trapero et al. [31] where the results showed improved efficacy of their solar irradiation forecast model with relevant prediction intervals. The study of Zhou et al. [32] has adopted an LSTM network with the KDE model to compute wind power using a probabilistic interval prediction method. In addition, this paper also provides explainability of the proposed hybrid CELM-KDE model by adopting the SHapley Additive exPlanations (SHAP) [33] approach, aiming to provide model interpretability in respect to which of the features from the climate variables as well as the lagged demand data could be considered as the influential contributor towards the prediction. Finally, we offer a practical application of the proposed CELM-KDE model for the first time to generate prediction intervals over which reliable G values can be attained at four geographically diverse study sites in Southeast Queensland. According to results documented later, the proposed CELM-KDE model outperforms five other competing deep learning models used for both point-based predictions and prediction interval of electricity demand.

The remainder of the paper has been structured as follows: next section details some specific related work on electricity demand prediction, specifically highlighting previous AI-based and hybrid approaches. Section 3 details the proposed hybrid CELM-KDE model proposed for daily prediction of G . Section 4 shows the experiments carried out to proof the effectiveness of the proposed hybrid CELM-KDE approach. Section 5 closes the paper with some conclusions and final remarks on the research carried out.

2. Related work

This section reviews the most important previous work on electricity demand prediction. Table 1 summarizes the research works reviewed here. As previously mentioned, the first models to predict G were based on statistical and econometrics models. Some researchers started to apply techniques based on the concept of persistence or historical behaviour to predict the future values [34]. Also known as analogue-based method [35], these techniques use methods like Exponential Smoothing that outperform some of the more sophisticated optimisation techniques such as Particle Swarm Optimisation [36]. Other statistical approaches such as ARIMA-GARCH models were applied to the prediction of G [34]. However, these traditional statistical techniques are neither adequate in capturing nonlinear properties of time series nor

flexible enough to accommodate for the abrupt changes in G that can occur due to sudden changes in consumer electricity use attributable to factors extreme weather, social, or other kinds of events [37].

Therefore, in recent years, research has largely focused on different AI models for improving G prediction and drawing greater insights into the distribution of predicted errors for better decision-making. These methods include both Deep Learning (DL) and Machine Learning (ML) [38, 39] as well as their hybrid counterparts aimed to capitalise on the merits of individual algorithms for improved G prediction. For example, an ANN model proposed by the work of Hamzaçebi et al. [40] has predicted the monthly G in Turkey to demonstrate its efficacy over a Seasonal ARIMA model. Conventional methods like ANN, SVR and DT however have several drawbacks including their poor generalisation and anti-noise performance, local optimisation issues, parameter modification and computationally inefficiencies [41]. Therefore, an Extreme Learning Machine (ELM) [42, 43] has been rapidly adopted because of its generalisation capabilities, the speed of model convergence without compromising the quality of predictions and the easiness at which the model's parameters can be set. In order to improve G prediction problems, the study of Ertugrul [44] developed a recurrent ELM model with their simulations showing that this model performs very well in predicting unsteady or chaotic datasets such as those encountered in the proposed study's G time-series.

Modern DL models like Recurrent Neural Network (RNN) [45, 46], Gated Recurrent Unit (GRU) [47], Convolutional Neural Network (CNN) [48, 49] and Long Short-Term Memory (LSTM) network [50, 51, 52] has become popular to tackled prediction problems related to electricity demand. Through additional mapping layers to explore historical patterns, RNN models can improve the accuracy of G predictions. In particular, an LSTM and a GRU model, which is a variation of an RNN model, can resolve vanishing gradient issues that are commonly encountered in RNN models [53]. In fact, Zheng et al. [54] has demonstrated that an LSTM model can outperform conventional models by exploring long-term dependencies in daily G data. In one study, an LSTM model was trained using multivariate input like temperature, day characteristics, date and time [55]. Compared with a more established model like SVR and Random Forest Regressor, the results showed improved accuracy of predicted G . Furthermore, researchers have used a CNN model to overcome an algorithm's limitation in predicting time-series data and as a result of local connections and feature-sharing properties of a CNN model, the training time and parameterization can also be reduced. This concurs with the study of Amarasinghe et.al. [56] where a CNN model has been adopted to accurately predict G values. In fact, their results also showed that a CNN model can outperform an SVR, as well as an ANN, and some of the other DL techniques.

Several recent research studies are experimenting with hybrid models to solve the problem of predicting G . For example, the study of Kim et al integrated CNN with an LSTM algorithms to build a hybrid predictive model for short-term G predictions, capturing its improved performance with a GRU, attention-based LSTM, LSTM and bidirectional LSTM model. Further, the study of [57] has created a hybrid predictive model by integrating a CNN model with a multi-layer bidirectional LSTM model to show its superiority over bidirectional LSTM, standard LSTM, and CNN-LSTM model. Unlike previous studies that did not, this study has blended an LSTM model with an CNN model to evaluate the efficacy of CNN-LSTM, CNN-LSTM, and CNN-LSTM auto-encoder models. Similarly, the study of Sajjad et al. [25] as well as Afrasiabi et al. [58] found that the CNN-GRU hybrid model is able to successfully predict the G series, reaffirming the better predictive capability of a hybrid deep learning over a standard (or standalone) model.

Table 1: Summary of the related works on electricity demand prediction analyzed in this paper.

Method	Articles
Persistence-based	[34]
Analogue-based	[35]
PSO	[36]
ANN	[40, 41]
SVR and DT	[41]
ELM	[42, 43, 44]
RNN	[45, 46]
GRU	[47]
CNN	[48, 49]
LSTM	[50, 51, 52, 54, 55]
CNN	[56]
CNN+LSTM	[57]
CNN+GRU	[25, 58]

3. Materials and Methods

This section provides details of the proposed hybrid CELM-KDE model employed in the daily G and associated confidence interval predictions at four study sites in southeast Queensland. As mathematical description of convolutional neural network, extreme learning machine and benchmark models e.g., RFR, SVR, KNNR, LSTM and DNN are available elsewhere [59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69], we only provide here details of CELM, followed by the proposed hybrid CELM-KDE model.

3.1. Hybrid Extreme Learning Machine-Convolutional Neural Networks Model

To construct a feature extraction system for the final predictive model, this study has first designed the hybrid CELM model before embedding the Kernel Density Estimation method to study the confidence intervals of predicted electricity demand. In principle, the CNN layers are deployed to extract predictive features related to electricity demand G whereas the ELM layer is finally used to integrate all of these features to build a low-latency predictive framework for $G(MW)$ data series.

Figure 1 shows a topological architecture of the hybrid CELM model.

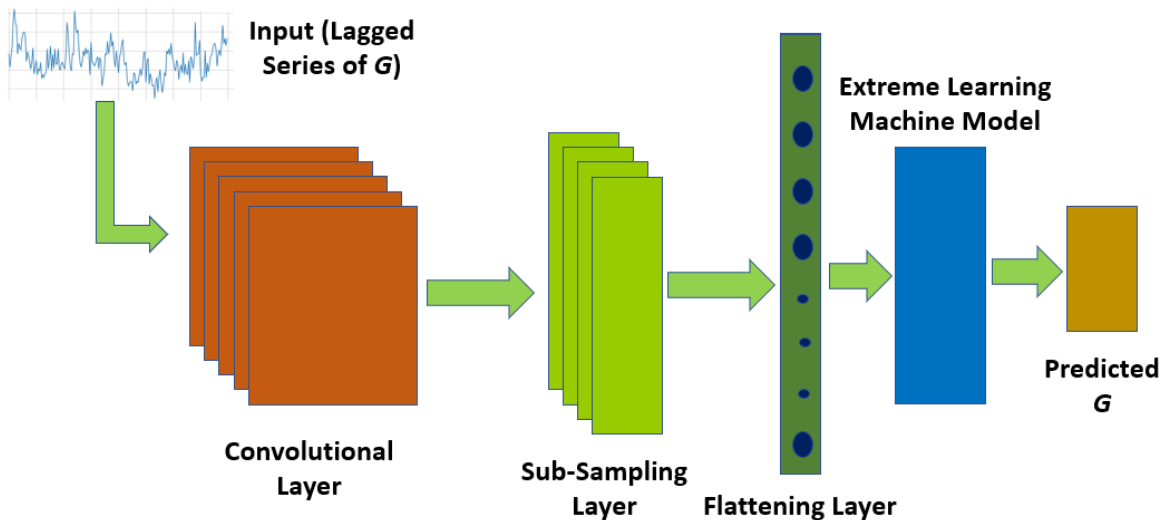


Figure 1: Framework of the the proposed hybrid CELM model architecture.

In general, two stages make up the proposed hybrid CELM model: the first is the feature extraction stage that uses CNN layers and the second is used as the layer for prediction of daily G values. In accordance with earlier works, time-series data can be successfully extracted with a CNN layer whereas a massive volume of sequential data can be handled relatively quickly using an ELM model to feed in crucial predictive feature. Therefore in this study, the hybrid CELM model is fed with the pre-processed historical G data features firstly into the CNN model which extracts feature information for perceived output. The CNN model's output is then used as the input for the final ELM model, after its training phase to predict the G in the testing phase.

3.2. Kernel Density Estimation: Confidence Interval Predictions of Electricity Demand

In this study we adopt the prior work of Rosenblatt and Emanuel Parzen [70] that devised a non-parametric density estimation technique, known as KDE, in order to provide our hybrid CELM model a new capability to generate confidence intervals of predicted G values. In contrast to conventional parameter estimations, the KDE method is able to solve the problem by using the data distribution's actual properties rather than making prior assumptions about the dataset [71]. Additionally, the proposed KDE method is able to achieve a greater superiority, accuracy and smoothness throughout the estimation intervals compared with other non-parametric estimations like a histogram approach, nearest neighbour method, and Rosenblatt method [72].

Suppose we have a dataset comprised of electricity demand (G) prediction error $p = (p_1, p_2, p_3, \dots, p_n)$ where n is the number of electricity demand error samples. Based on non-parametric KDE, the probability density function (PDF) of error in G can be estimated as follows (1).

$$\hat{f}(p, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{p - p_i}{h}\right) \quad (1)$$

where $\hat{f}(p, h)$ is the KDE of predicted G error, h is the bandwidth that determines the interval division of error data distribution, p_i is the i^{th} G prediction error sample point and the function $K(p, h)$ denotes the kernel function whereas electricity demand error and the bandwidth are independent variables of that function.

The choice of kernel function and bandwidth h are the main variables impacting the non-parametric KDE and this involves a significant amount of selectivity in the kernel functions. The most often used kernel functions are: Epanechnikov [73], gamma, and uniform kernels. It is noteworthy that the choice of a kernel function has a lesser influence on the KDE compared to the bandwidth. Due to its smooth properties, a Gaussian kernel is frequently regarded as the best option and is suited for a wide range of applications. In this study, we adopt the Gaussian kernel expressed as follows:

$$K(p) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-p^2}{2}\right) \quad (2)$$

Therefore, the non-parametric KDE with a Gaussian kernel is expressed as Equation (3):

$$\hat{f}(p, h) = \frac{1}{\sqrt{2\pi}nh} \sum_{i=1}^n \exp\left(-\frac{1}{2}\left(\frac{p - p_i}{h}\right)^2\right) \quad (3)$$

The bandwidth directly impacts the PDF of the G prediction error in the non-parametric KDE. Therefore the distribution of the estimated model is impacted by local fluctuations, which becomes severe when the bandwidth is too tiny and is affected by special sample points.

chosen to assess the efficacy of the proposed hybrid CELM-KDE model. These historical data at a 30-min interval from 01/07/2011 to 30/06/2021 were obtained from Energex, the federal government-owned electricity distribution company based in South East Queensland (<https://www.energex.com.au>) and ground-based climate variables were extracted from Scientific Information for Land Owners (SILO) [75]. The 30-min electricity demand data (G) were converted to cumulative daily demand. For interpretation of model performance in terms of electricity demand normally measured in MWh in this paper, the respective timescale over which results are presented should be applied. The statistical characteristics of data are shown in Table 2.

Table 2: Descriptive statistics of daily electricity demand (G) at the tested substations in South-east Queensland where the proposed hybrid CELM model has been implemented.

Statistical Parameters	Coolum	Cornubia	Caloundra	Lawnton
Median (MW)	469.79	206.05	1045.44	613.69
Mean (MW)	474.21	208.60	1057.63	624.90
Standard deviation (MW)	58.30	35.75	130.08	103.45
Variance	3399.04	1277.78	16920.13	10701.25
Maximum (MW)	724.77	357.91	1622.60	1146.17
Minimum (MW)	16.87	29.25	516.45	83.30
Range	707.90	328.66	1106.15	1062.87
Interquartile range	73.63	44.28	153.97	117.71
Skewness	-0.01	0.45	0.37	0.65
Kurtosis	5.95	4.12	4.56	4.98

From Table 2, the standard deviation of G for Cornubia, Coolum, Lawnton and Caloundra sub-stations are 35.75, 58.30, 103.45 and 130.08 MW, respectively, indicating the high fluctuation in electricity demand. Furthermore, to determine times of peak electricity demand, we were also curious about the seasonality of our data. In that regard, a box plot (Figure 3(a)) was also used, and the results show that the summer months of December, January and February have the highest electricity use. Also, the Weekends have lower G consumption than the weekdays (Figure 3(b)) for all four sub-stations.

Table 3 shows the details of the SILO data, the climate input time series, were recorded every 24h (daily), and there are no missing values, so the imputation of time series is not required.

In this work, the aim was to predict the accumulated daily electricity demand using the hybrid CELM-KDE model. Therefore, statistical studies were carried out by computing partial autocorrelation function (PACF) and Mutual Information (MIF) to find the time-varying patterns in the G time series. The PACF for G time series for 2020 are displayed in Figure 4.

Notably, the presence of repeating patterns in PACF suggests that there is seasonality in G time series. The PACF plots for all sub-stations show that there is a strong correlation at $lag1$. It should be noted that the PACF plots only show linear relationships between electricity demand; however, the time series may also include non-linear relationships. These non-linear interactions can be captured by the models utilised in this study. Furthermore, the curve indicating the values of MIF as the delay duration (τ) rises is shown in Figure 5. It is observable that as the delay time increased, the MIF deteriorated. When the value of MIF reaches the first local minimum, the ultimate delay duration is typically established. As a result, the G delay time is chosen to be $\tau = 6, 5, 6$ and 5 for the Coolum, Cornubia, Caloundra and Lawnton, respectively. It means that the closest six datapoints, i.e. $G_{(t)}, G_{(t-1)}, \dots, G_{(t-6)}$, have similar information, namely, dependent relation for Coolum and Caloundra. Similarly, for Cornubia and Lawnton, the closest five datapoints, i.e. $G_{(t)}, G_{(t-1)}, \dots, G_{(t-5)}$, are correlated.

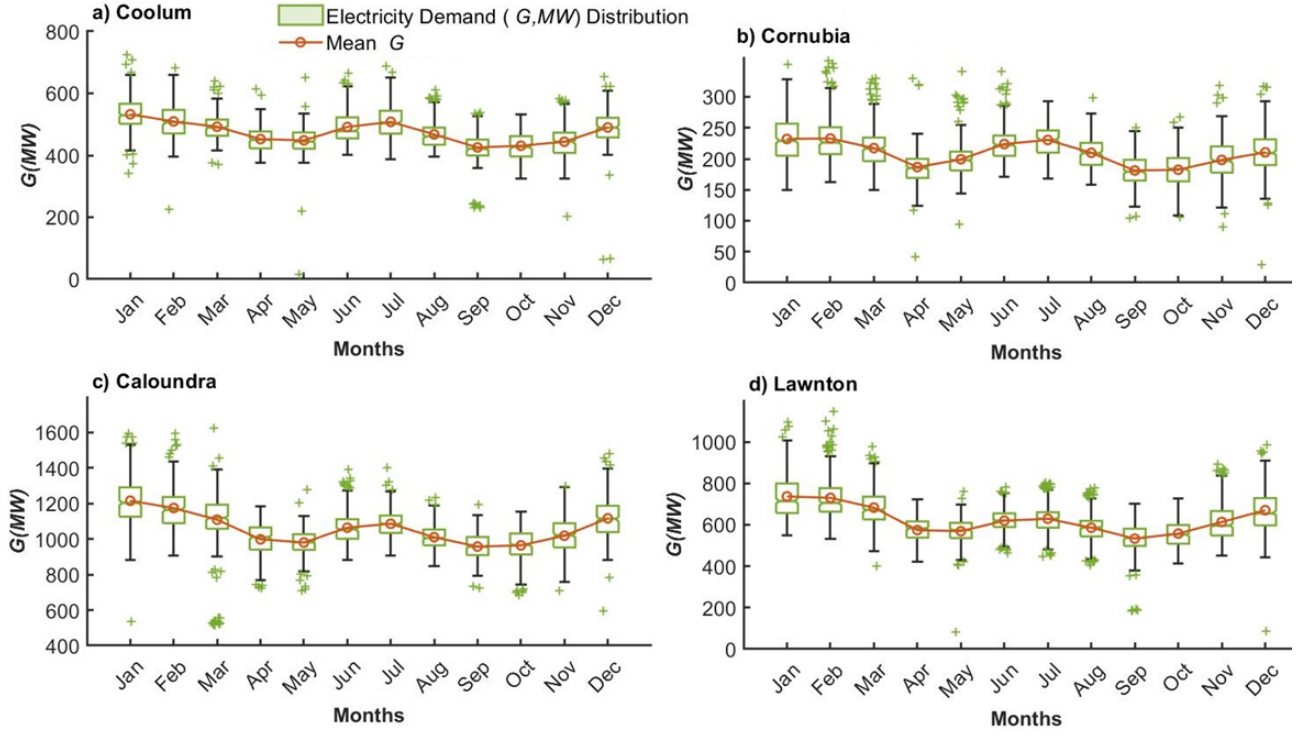


Figure 3(a): Box plot of the electricity demand by month at the four substations in Southeast Queensland.

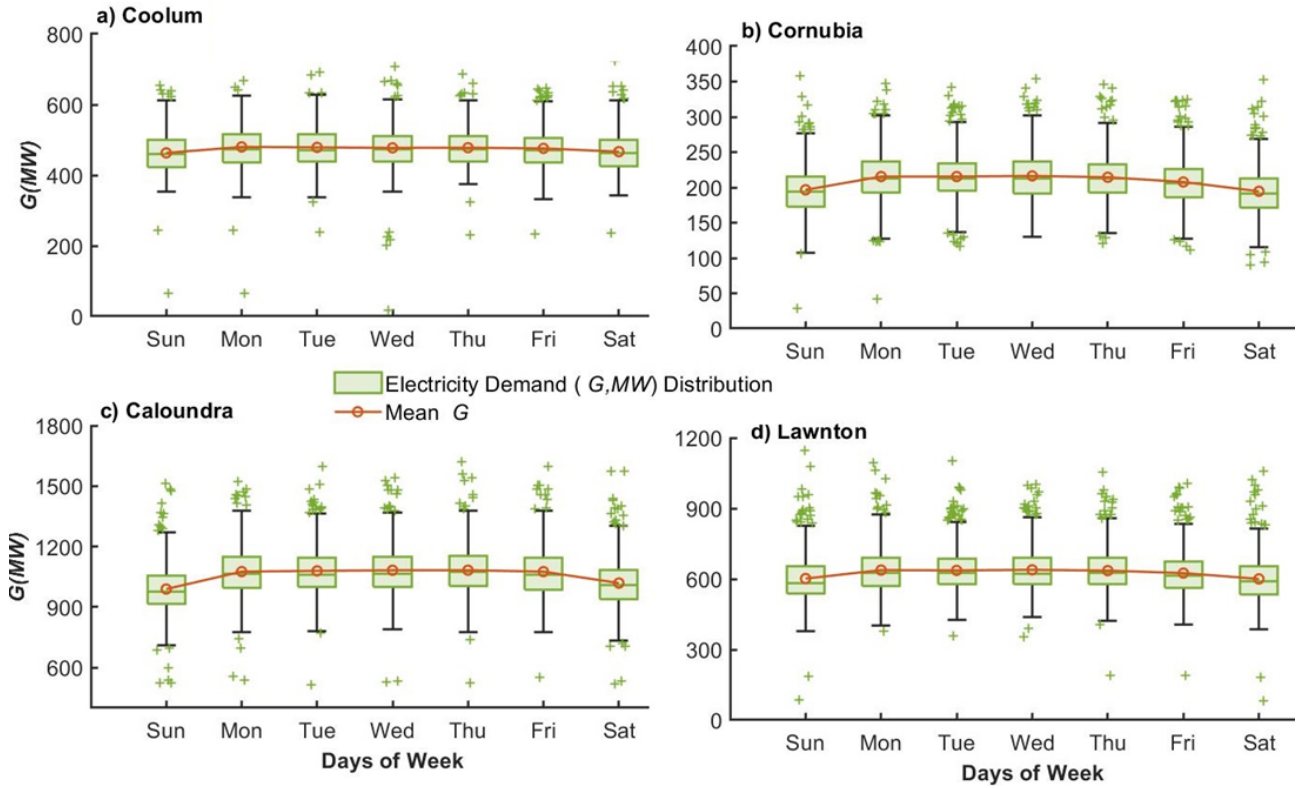


Figure 3(b): Box plot of the electricity demand by the days of the week at the four substations in Southeast Queensland.

Table 3: Description of the extensive pool of predictor variables from the Scientific Information for Landowners (SILO) database used for the prediction of daily electricity demand G (MW) at four substations in southeast Queensland.

Climate-based Predictor Variables	Acronym
Maximum temperature ($^{\circ}\text{C}$)	Tmax
Minimum temperature ($^{\circ}\text{C}$)	Tmin
Vapour pressure (hPa)	VP
Vapour pressure deficit (hPa)	VPd
Evaporation - synthetic estimate (mm)	Esyn
Solar radiation - total incoming downward shortwave radiation on a horizontal surface (MJ/m^2)	GSR
Relative humidity at the time of maximum temperature (%)	RHmax
Relative humidity at the time of minimum temperature (%)	RHmin
Evapotranspiration - Morton's areal actual evapotranspiration (mm)	Etm
Mean sea level pressure (hPa)	MSLP

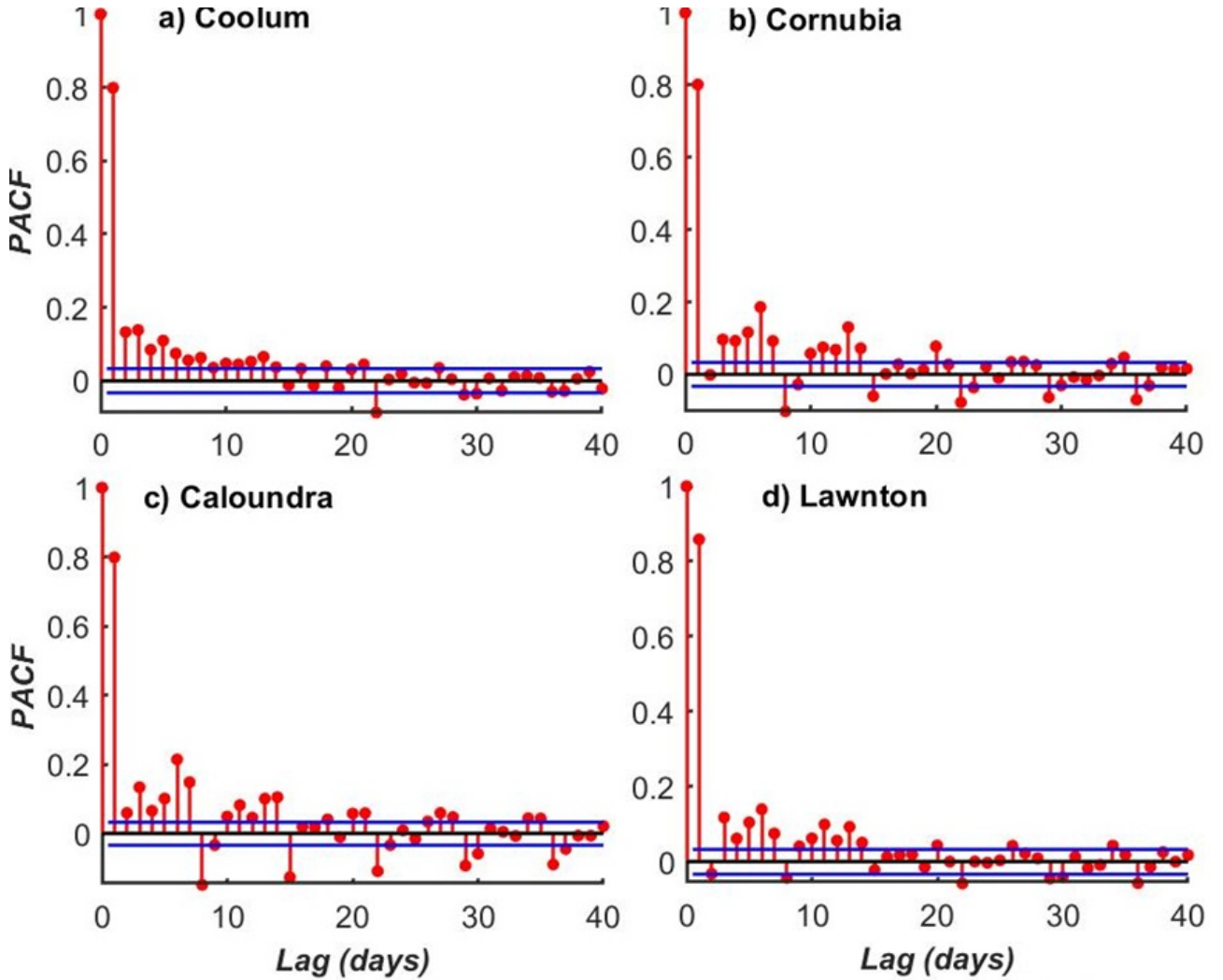


Figure 4: The partial autocorrelation function (PACF) of the daily time series of G data used in training the proposed hybrid CELM-KDE model for Coolum, Cornubia, Caloundra and Lawnton study substations. The blue lines denote the statistically significant boundary at the 95% confidence interval

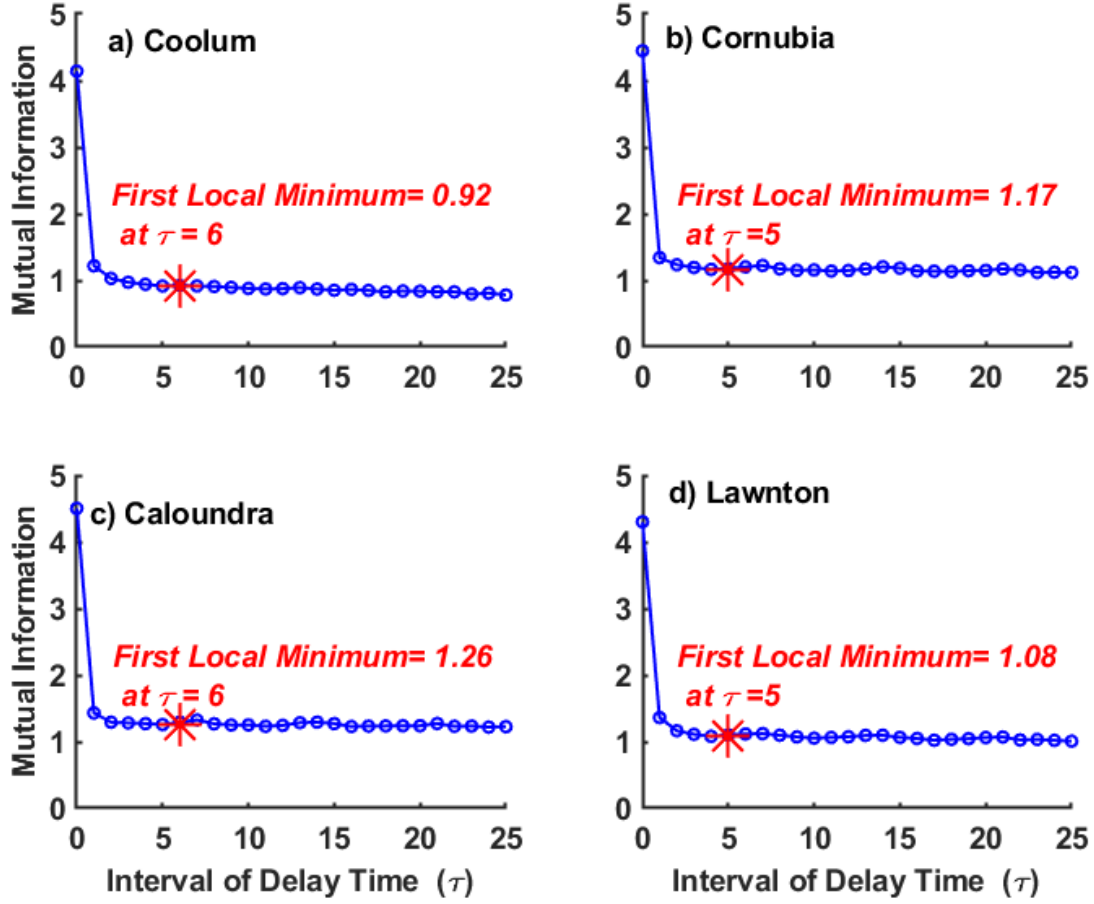


Figure 5: The mutual information function (MIF) test performed to validate the PACF for the input matrix of antecedent lagged daily electricity demand (G, MW) when designing the proposed hybrid CELM-KDE model.

The ELM model was used to validate the *PACF* and *MIF* criterion for choosing the order of lag variable. Figure 6 displays the ELM model (Transfer function= *Sigmoid* and number of hidden neuron= 50) performance in terms of Root Mean Square Error ($RMSE, MW$), Relative RMSE ($RRMSE, \%$) and correlation coefficient (r) with varying lag (1-40) for the 4 substations at South East Queensland (SEQ). It is important to note that from Figure 6, the optimal lagged input combination was 6, 5, 6, and 5 for daily G data for Coolum, Cornubia, Caloundra and Lawnton, respectively, with the higher magnitude of Correlation Coefficient (r) and lower magnitude of $RMSE$ and $RRMSE$.

3.4. Explainability of the Proposed Hybrid CELM-KDE Model

In order to build a robust CELM-KDE model, it is vital to firstly normalise the G data as well as the predictor variables to prevent the non-convergence of the model that can be brought on by different magnitudes of numerical values. In this study, a linear normalisation has been considered appropriate as follows:

$$X_{\text{norm}} = \frac{X_t - X_{\min}}{X_{\max} - X_{\min}} \quad (7)$$

where X_{norm} = normalised X_t , $X_{\max}$ and $X_{\min}$ are, respectively, the maximum and minimum values in the dataset.

The normalised G and climate variables is then divided into a training set and testing set for training and testing of the proposed CELM-KDE model. The training (80%) comprises a dataset from 01/07/2011 to 30/06/2020 (3,652 data points), and the testing (20%) set comprises

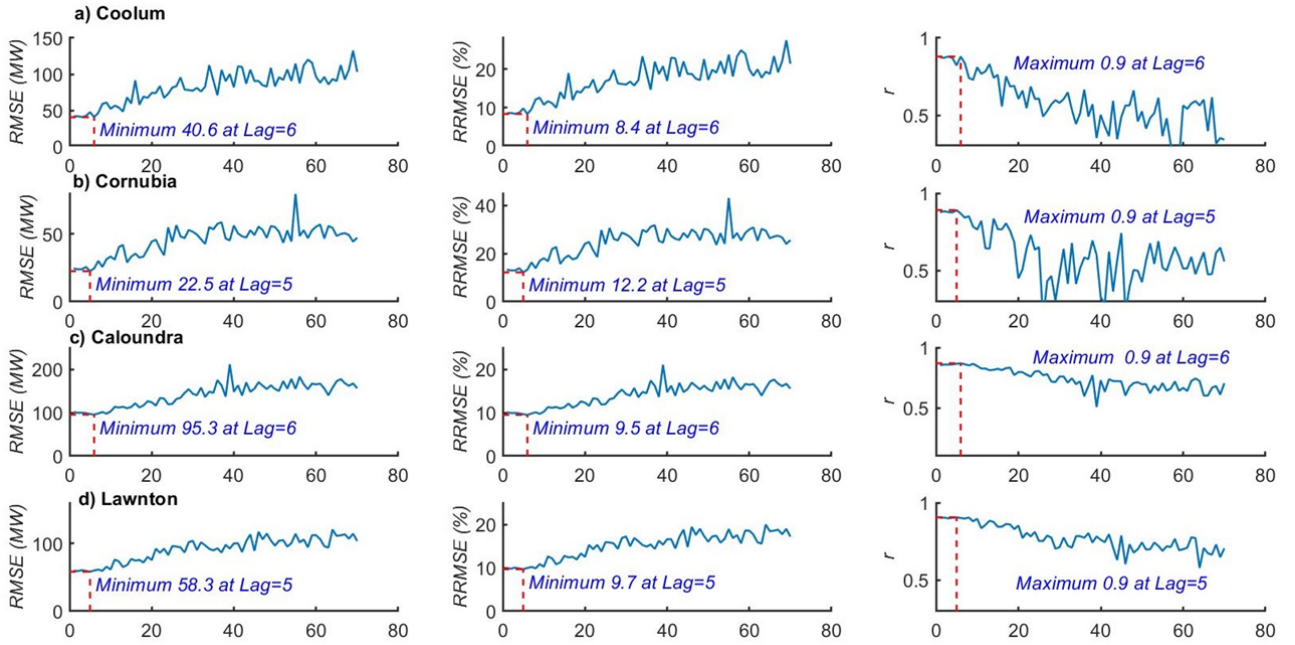


Figure 6: Validation of the antecedent lagged electricity demand (G) using the extreme learning machine (ELM) Network.

the dataset from 01/07/2020 to 30/06/2021 (365 data points). The training set is further divided into training (80%) and validation (20%) set. With an overall aim to reduce the prediction errors, the training set is used to fit the optimal model parameters, and the validation set is used to determine the optimal setting for the hyperparameters.

In this paper, we provide greater explainability of the proposed hybrid CELM-KDE model by adopting the SHapley Additive exPlanations (SHAP) [76, 33] approach. This method is becoming very prominent for model interpretability aiming to demonstrate which of the features from the climate variables as well as the lagged G input could be considered as the main contributor towards the overall prediction of daily G dataset.

As revealed in Figure 7, the SHAP values generated by the CatBoost Regressor model clearly illustrates each predictor variable's effect on the model's output while also providing the rank of the variable attributes according to the total of those values across all of the samples. The colour indicates the value of the characteristic (blue: for low, red: for high). It is noticeable that the climate-related model input features such as **Tmax** and **Lag1** (i.e., the first lag of G) are both found to have a high degree of importance of the G predictions generated by the hybrid CELM-KDE model for all of the four sub-stations. There also appears to be a positive correlation between the input defined as **Lag1** and the model output, as well as between **Tmax** and the model output as indicated by the commensurate increase between the feature values and the SHAP values.

3.4.1. Development of the Proposed Hybrid CELM-KDE Model

The proposed hybrid CELM-KDE model in this study integrates the individual merits of the CNN (as a feature extraction tool) and ELM (as a predictive modelling framework), as well as the KDE approach to generate confidence intervals of predicted daily G values. The convolutional layer was the core of the proposed CELM model. In our proposed framework, the CNN model extracts features among variables (Lagged G and climate variables from SILO) that affect G . Each CNN layer uses the ReLU activation function. Then, the CNN layer's output is flattened to be fed to the complementary ELM model to predict the daily G .

Figure 8 shows a block diagram of the proposed model. A summarised CELM algorithm involving the following eight steps, is shown also in Figure 9:

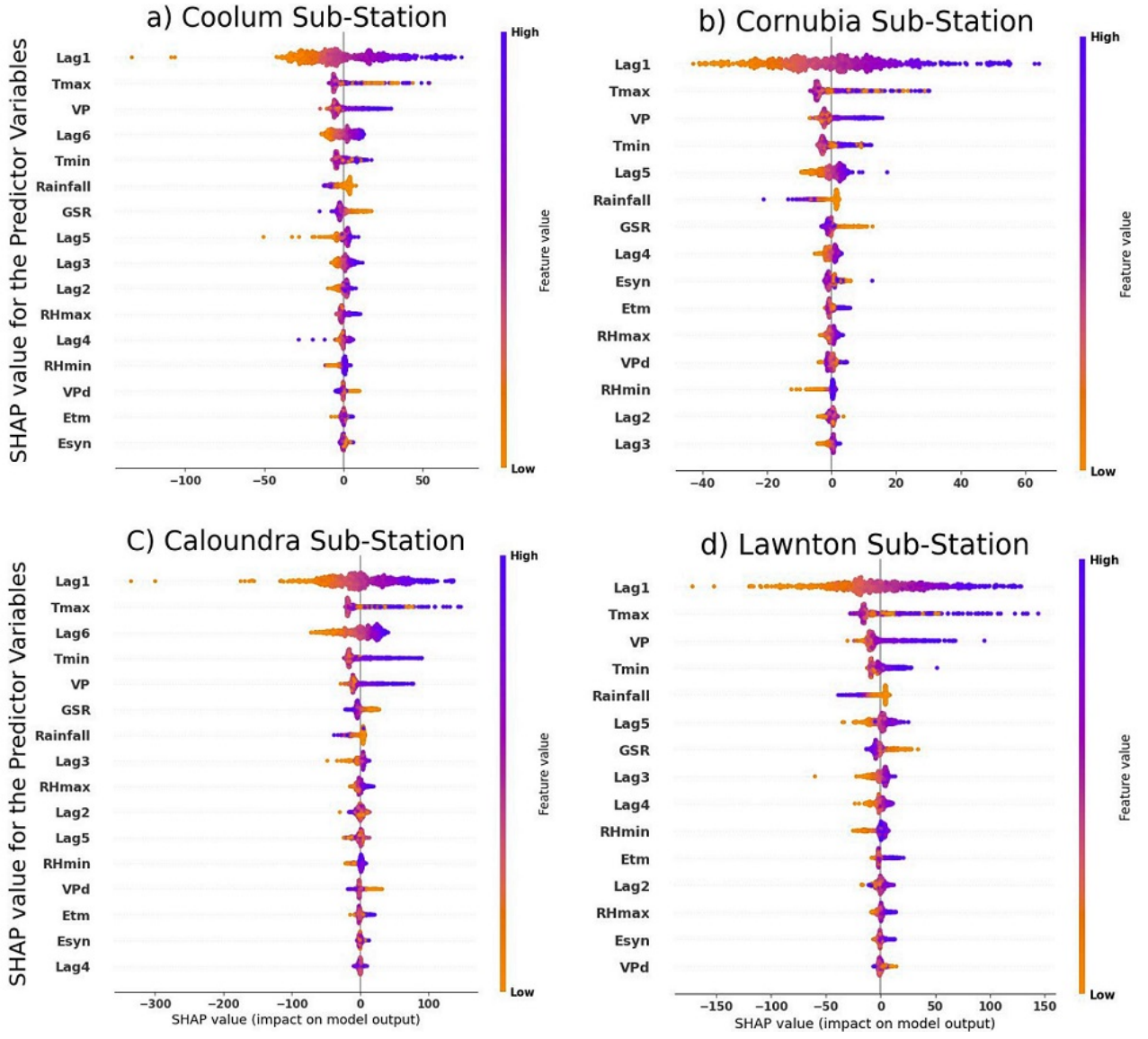


Figure 7: The explainability of the input features using the SHapley Additive exPlanations (SHAP) method used in the design of the proposed hybrid CELM-KDE model. These features were ranked in descending order of importance (e.g., **Lag1**, which is the historical electricity demand to capture the highest feature importance). SHAP values on horizontal axis denote the influence of individual features on the model output (e.g., positive SHAP = improvement of model output). The vertical axis denotes the value of a given feature (i.e., the closer the colours are to red, the higher the value of the feature).

- Convert climatic variables and the lagged G dataset into signal vectors for the CNN model's input.
- Initialise the CNN parameters, such as the maximum epoch of iteration, the learning rate, the optimiser and the number of filters.
- Implement convolution and pooling.
- Fine-tune the stacked model using the Optuna hyperparameter optimization framework, as shown in Figure A.24
- Finish the optimisation and acquire the features of the dataset
- Use the extracted features for regression by ELM.

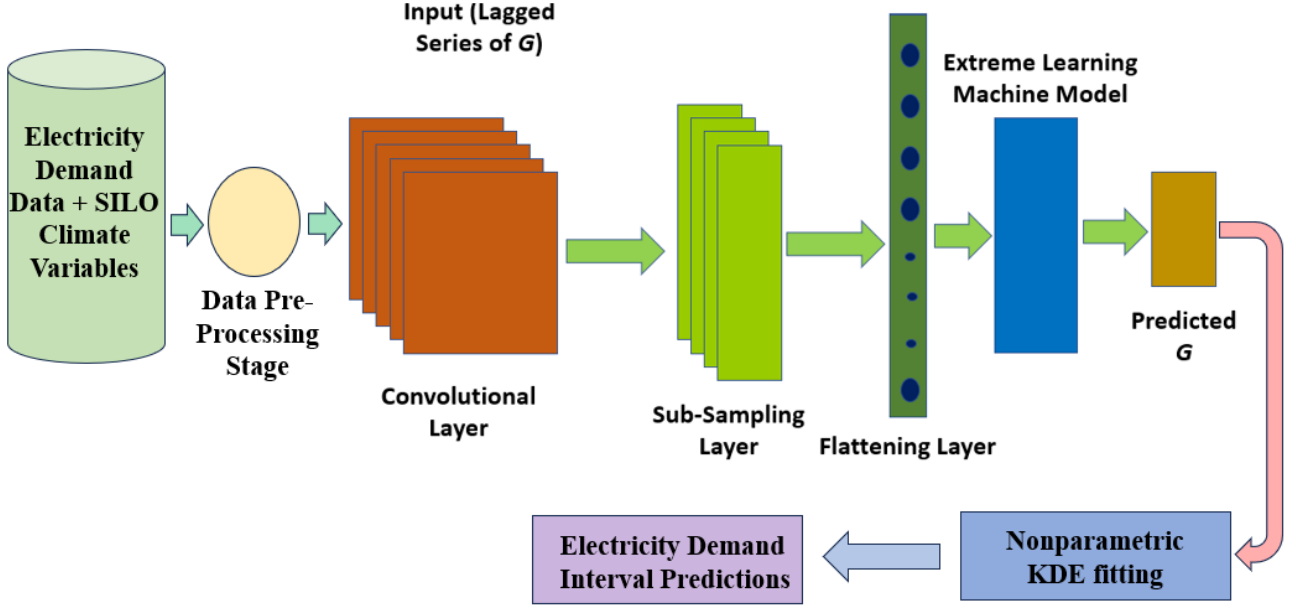


Figure 8: Block diagram of modeling framework of the proposed hybrid CELM-KDE model used for electricity demand predictions.

- Initialise the ELM parameters, such as the total number of neurons in the reservoir and the activation function.
- Get the final optimised CELM model and predict on the test dataset.

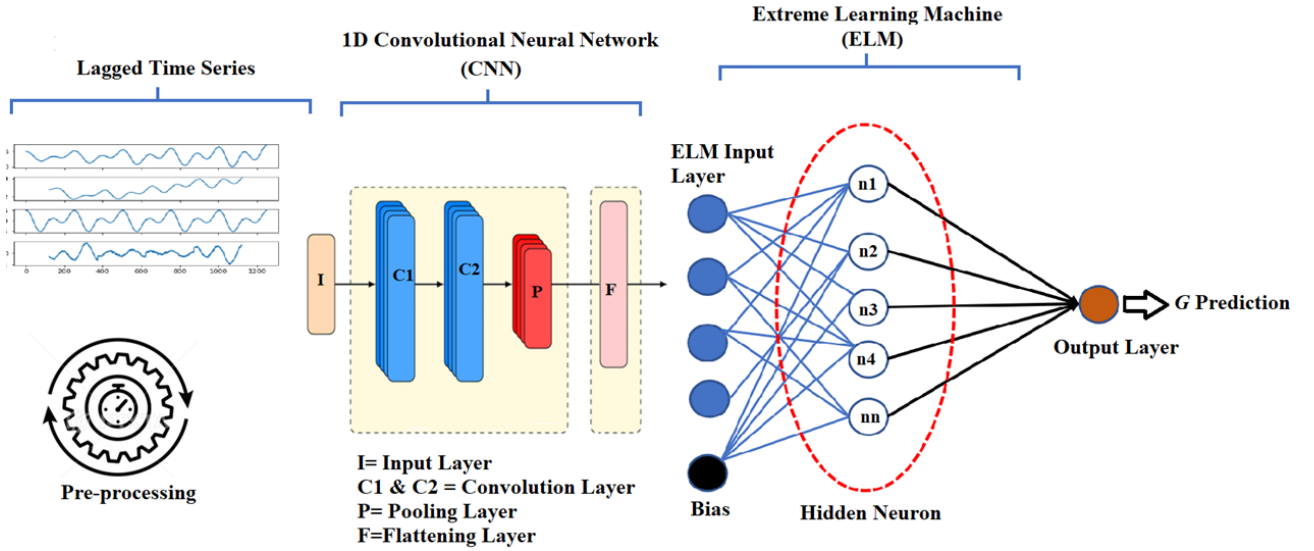


Figure 9: Topological structure of the hybrid CNN model integrated with an ELM layer used in daily electricity demand (G, MW) prediction.

We then implemented the proposed KDE method used for daily electricity demand G interval predictions described as follows:

- Step 1: The proposed CELM and benchmark models are used to generate point predictions $Y = [y_1, \dots, y_n]$ of electricity demand G where $n = \text{testing dataset}$.
- Step 2: Calculating Prediction Error (PE) between the actual $[x_1, \dots, x_n]$ (test dataset) and predicted Y . This PE can be expressed as $PE = [e_1, \dots, e_n]$, where $e_i = y_i - x_i, i = [1, \dots, n]$.

- Step 3: Normalising prediction errors to improve the KDE fitting performance using $PE' = \frac{PE - \mu}{\sigma}$, where PE' , μ and σ are normalised PE , mean and standard deviation respectively of the actual PE .
- Step 4: Establishing a non-parametric KDE model for normalised PE (PE') to obtain cumulative PE distribution function as per Equation (3) and determining the KDE bandwidth using Silverman's rule of thumb.
- Step 5: Obtaining the optimum interval $[PE'_{lo}, PE'_{up}]$ of the PE at 95% confidence using Equations (4) and (5)
- Step 6: Calculating the prediction interval of electricity demand at 95% confidence level by transforming PE intervals reversely.

The inverse transforming formula of PE intervals is expressed as:

$$Y_{up} = Y + [PE'_{up} \cdot \sigma + \mu] \cdot Y \quad (8)$$

and

$$Y_{lo} = Y + [PE'_{lo} \cdot \sigma + \mu] \cdot Y. \quad (9)$$

where PE'_{lo} and PE'_{up} = lower bounds and upper bounds of the PE , μ and σ = mean and standard deviation of PE respectively, Y = the point prediction of electricity demand using CELM and benchmark models and Y_{up} and Y_{lo} = predictive bounds of electricity demand G .

3.4.2. Benchmark Models

In order to appraise the merits of the proposed hybrid CELM-KDE model, this study has developed five competing deep learning (DNN, SVR, RFR, KNNR, LSTM) models. All models were constructed using Keras 2.2.4 [77] library on TensorFlow 1.13.1 [78] back-end in Python 3.6 programming platform. The training of all models was conducted on Intel Core i7 CPU with 32 GB RAM.

Selecting the most optimal hyperparameters is crucial so in this research, we adopted the Optuna [79] method that provides an open-source program to tune the model hyperparameters. The Optuna package provides benefit of assessing hyperparameters using a define-by-run API, which estimates the parameters based on subsequent evaluations. It also determines an objective function score for each collection of hyperparameters (minimised or maximised) and after that creates a probabilistic model favouring efficient hyperparameters in the search. Additionally, this method also uses an early stopping or pruning process to eliminate unfavourable possibilities without first testing them. Furthermore, it has a parallelisation capability allowing many scenario executed simultaneously on various CPUs. In this study, we trained all models by minimising ($RMSE$) with a set of hyperparameters. Table 4 shows the hyperparameters search space for the CELM as well as benchmark models with the optimal hyperparameters selected via Optuna highlighted in boldfaced.

3.5. Rationale for Performance Metrics and Explanations for Model Evaluation

This study acknowledges that minimal prediction error is a key goal in point-based G predictions. While minimising such as error is feasible through a robust predictive model, it becomes more difficult in case of probabilistic predictions if we prefer a PI that minimises this error for a quantile or percentile as an upper and lower bound PI . A PI should be ideally as sharp as feasible as possible while still encapsulating a real point-based predicted values as a

Table 4: The architecture of the hybrid CELM as well as DNN, RFR, KNNR, and LSTM models used for daily electricity demand prediction. Note: - ReLU = Rectified Linear Units; Adam = Adaptive Moment Estimation, gbdt= traditional Gradient Boosting Decision Tree, rbf=Radial Basis Function, logistic= Logistic Sigmoid Function, tanh= Hyperbolic Tangent Activation Function, loguniform= In the log-uniform distribution, points are sampled uniformly between $\log(a)$ and $\log(b)$, where $\log$ is most frequently the logarithm with base 10.

Predictive Models	Model Hyperparameters	Hyperparameter Selection	Coolum	Cornubia	Caloundra	Lawnton
Convolution Neural Network Integrated with Extreme Learning Machine (CELM)	Filter1 (CNN)	[60, 80,120,250]	80	60	80	120
	Filter 2 (CNN)	[40,50,60,70,80]	80	40	50	60
	Filter 3 (CNN)	[20,10,30,5]	20	30	10	20
	Number of neurons (ELM)	[20,30,40,50,60,70,80]	50	40	60	20
	Activation Function (ELM)	Sigmoid				
Deep Neural Network (DNN)	Epochs (CNN)	[1000]				
	Solver (CNN)	['adam']				
	Batch Size (CNN)	[5,10,15,20,25,30]	10	5	10	5
	Hiddenneuron 1	[60,100,200,250,300,500]	200	60	100	50
	Hiddenneuron 2	[20,30,40,50,60,70]	50	30	20	20
Random Forest Regression (RFR)	Hiddenneuron 3	[10,20,30,40,50]	30	40	40	30
	Hiddenneuron 4	[5,6,7,8,12,15,18]	12	15	12	18
	Batch Size	[5,10,15,20,25,30]	5	10	10	5
	Solver	['adam']				
	Epochs	[1000]				
K Nearest Neighbor (KNNR)	The maximum depth of the tree.	('max_depth', range (1,110,1))	27	18	17	21
	The number of trees in the forest.	('n_estimators', range (5,800,2))	80	74	62	78
	Minimum number of samples to split an internal node	('min_samples_split', range (2,100,1))	8	10	8	10
	The number of features to consider when looking for the best split.	['auto', 'sqrt', 'log2']	auto	auto	auto	auto
	Number of neighbors to use by default for neighbors queries.	('n_neighbors', range (2,25))	14	10	8	11
Support Vector Regression (SVR)	Algorithm	['auto', 'ball_tree', 'kd_tree', 'brute']	auto	auto	auto	auto
	Kernel	'rbf'				
	Epsilon	0.0001				
	Cost Function (C)	Loguniform (-40,10)	19487.12	17458.12	11254.25	19457.21
	Penalty function (Gamma)	Loguniform (-40,10)	0.065	0.114	0.158	0.035
Long Short Term Memory Network (LSTM)	LSTM cell 1	[50, 60,100,200]	50	60	100	60
	LSTM cell 2	[40,50,60,70,130]	40	50	70	40
	LSTM cell 3	[20,10,30,5]	20	10	10	30
	LSTM Cell 4,5 and 6	[Fixed as 30,20,10]				
	Activation function	relu				
	Epochs	[1000]				
	Drop rate	[0,0.1,0.2]	0	0.1	0.1	0
	Batch Size	[5,10,15,20,25,30]	5	10	5	15

PI that is too widely distributed is unlikely to generate much useful information for accurate decision-making.

In Table 4(a) [80, 81, 82, 83], 4(b) [66, 84, 75], 4(c) [85, 86, 87, 88], and 4(d) [89, 90, 91], we describe these metrics used to evaluate the skill of both point-based and probabilistic predictions of daily G dataset in the hybrid CELM-KDE model's testing phase.

Table 4(a): Deterministic performance measures denoted as Class A used to evaluate the proposed hybrid CELM-KDE model. *Note: G^m and G^p are the observed and predicted value of G , $\langle G^m \rangle$ and $\langle G^p \rangle$ are the observed and predicted mean of G , p stands for the model prediction, x for the observation, pr for perfect prediction (persistence), and r for the reference prediction, VAR is the variance, SD is the standard deviation, and n corresponds to the size (number) of predictions.*

Deterministic Performance Measure (Class A)	Definition
Correlation Coefficient	$r = \frac{\sum_{i=1}^n (G^m - \langle G^m \rangle)(G^p - \langle G^p \rangle)}{\sqrt{\sum_{i=1}^n (G^m - \langle G^m \rangle)^2} \sqrt{\sum_{i=1}^n (G^p - \langle G^p \rangle)^2}} \quad (10)$
Root Mean Square Error (MW)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (G^p - G^m)^2} \quad (11)$
Mean Absolute Error (MW)	$MAE = \frac{1}{n} \sum_{i=1}^n G^p - G^m \quad (12)$
Relative Root Mean Square (%)	$RRMSPE = \frac{RMSE}{G^m} \times 100\% \quad (13)$
Relative Mean Absolute (%)	$RMAPE = \frac{MAE}{G^m} \times 100\% \quad (14)$
Uncertainty at 95%	$1.96(SD^2 - RMSE^2)^{0.5} \quad (15)$
t-statistic	$TS = \sqrt{\frac{(n-1) \times MBE^2}{RMSE^2 - MBE^2}} \quad (16)$
Mean Bias Error (MW)	$MBE = (100 / \langle G^m \rangle) \sum_{i=1}^n (G^p_i - G^m_i) \quad (17)$
Standard deviation of the Relative Error	$STDRE = \left(\frac{1}{n-1} \sum_{i=1}^n \left(\frac{G^p - G^m}{G^m} \right)^2 \right)^{1/2} \quad (18)$
Explained Variance Score	$E_{var} = 1 - \frac{\text{Var}(G^m - G^p)}{\text{Var}(G^m)} \quad (19)$
Absolute Percentage Bias (%)	$APB = \frac{\sum_{i=1}^n (G^m - G^p) * 100}{\sum_{i=1}^n G^m} \quad (20)$
Skill Score	$SS = 1 - \frac{RMSE(p, x)}{RMSE(pr, x)} \quad (21)$

In principle, we devised four classes of evaluation metrics in testing phase where observed and predicted G are compared. It is noteworthy that this comprises of Class A measures (Table 4a) as an indicators of dispersion (or “error”) of individual points (with a trivial value should be attained for an idea model). According to Li et al. [95], different ranges of $RRMSPE$ and $RMAPE$ can be defined to show a models’ capability. A model precision could be considered as excellent for $0 \leq RRMSPE$ or $RMAPE \leq 10\%$, good for $10\% \leq RRMSPE$ or $RMAPE \leq 20\%$, fair for $20\% \leq RRMSPE$ or $RMAPE \leq 30\%$ and poor for $RRMSPE$ or $RMAPE \geq 30\%$.

In terms of Class B measures (Table 4b), the overall performance indicators take a value of unity for a perfect model. As these metrics are normalized between $[0, 1]$, they provide respective comparison of one model against another. However, Class C metrics (Table 4c) are used to study the distribution similitude where the goal is to compare one or more cumulative frequency distributions of modelled G data to that of a reference data. We also adopted the study of Gueymard that has proposed the Combined Performance Index (CPI) which combines $RMSE$, KSI , and $OVER$ into one statistical indication where the smallest value would indicate the best model from a frequency distribution [96].

Table 4(b): Deterministic performance measures denoted as Class B used to evaluate the proposed hybrid CELM-KDE model. *Note: G^m and G^p are the observed and predicted value of G , $\langle G^m \rangle$ and $\langle G^p \rangle$ are the observed and predicted mean of G , n is the size (number) of predictions, and CV_m and CV_p are the observed and predicted coefficient of variation.*

Deterministic Performance Measure (Class B)	Definition
Willmot's Index	$E_{WI} = 1 - \frac{\sum_{i=1}^n (G^m - G^p)^2}{\sum_{i=1}^n (G^p - \langle G^m \rangle + G^m - \langle G^m \rangle)^2} \quad (22)$
Nash-Sutcliffe	$E_{NS} = 1 - \frac{\sum_{i=1}^n (G^m - G^p)^2}{\sum_{i=1}^n (G^m - \langle G^m \rangle)^2} \quad (23)$
Legates and McCabe's Index	$E_{LM} = 1 - \frac{\sum_{i=1}^n G^m - G^p }{\sum_{i=1}^n G^m - \langle G^m \rangle } \quad (24)$
Theil's Inequality Coefficient	$TIC = \frac{\sqrt{\frac{1}{n} \times \sum_{i=1}^n (G^p - G^m)^2}}{\left(\sqrt{\frac{1}{n} \times \sum_{i=1}^n (G^m)^2} + \sqrt{\frac{1}{n} \times \sum_{i=1}^n (G^p)^2} \right)} \quad (25)$
Kling-Gupta Efficiency	$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\langle G^p \rangle}{\langle G^m \rangle} - 1 \right)^2 + \left(\frac{CV_p}{CV_m} \right)^2} \quad (26)$

Table 4(c): Deterministic performance measures denoted as Class C used to evaluate the proposed hybrid CELM-KDE model. *Note: D_n is the absolute difference between the calculated and measured CDF, X_{min} and X_{max} are the minimum and maximum values of D_n , A_c is the critical area, D_c is a statistical characteristic of the reference distribution or critical value, N is number of points, and $\Phi(N)$ is a pure function of N [92, 93]*

Deterministic Performance Measure (Class C)	Definition
Kolmogorov-Smirnov Index	$KSI = \frac{100}{A_c} \int_{X_{min}}^{X_{max}} D_n dx \quad (27)$
OVER Index	$OVER = \frac{100}{A_c} \int_{X_0}^{X_1} \max(D_n - D_c, 0) dx \quad (28)$
	$\text{where } A_c = D_c(X_{max} - X_{min}) \quad (29)$
	$\text{and } D_c = \Phi(N)/N^{1/2} \quad (30)$
Combined Performance index	$CPI = \frac{KSI + OVER + 2RMSE}{4} \quad (31)$

We also adopted the Class D measures (Table 4d) as an interval forecasting evaluation metric. Here, a trustworthy prediction interval is one with a bigger $PICP$ and a lower $PINAW$ [97]. Furthermore, the value of $FValue$ is meant to incorporate two contrasting indices such as $PICP$ and $PINAW$ serving as a thorough indicator to choose the best PI [98] value. In general, a higher value of F is likely to indicate a better performance of an interval prediction model. As prediction interval and coverage widths are isomorphic we have analysed the trade-off between coverage and prediction interval width using Winkler score (WS) [99] since the coverage can be readily attained by having a bigger prediction interval width. More generally, a better-performing prediction interval is characterised by a lower $ARIL$ and WS and finally we use $CRPS$ where a skewed negative value with a lower number can indicate a higher prediction accuracy and the projected CDF is closely aligned with actual CDF .

In order to find the best performing model, it is difficult to compare different statistical indicators at the same time since each has its own merits and shortcomings. In this study, the overall model performance was ranked using Global Performance Indicator (GPI) [100]

Table 4(d): The probabilistic performance measures denoted as Class D used to evaluate the proposed hybrid CELM-KDE model. *Note:* N denotes the number of test samples, y_i is the i^{th} observation, $L(G_i)$ and $U(G_i)$ represent lower and upper bounds of the i^{th} G Prediction Interval, respectively. G^m is the observed value of G. R is the Range. [94]. In CRPS metrics, $1(\cdot)$ is the Heaviside function, it takes the value of 1 when $t > y$ and equals 0 otherwise.

Deterministic Performance Measure (Class D)	Definition
Prediction Interval Coverage Probability	$PICP = \frac{1}{N} \sum_{i=1}^N c_i$ (32)
	where $c_i = \begin{cases} 1 & \text{if } y_i \in (U(G_i), L(G_i)) \\ 0 & \text{otherwise} \end{cases}$ (33)
Mean Prediction Interval Width	$MPIW = \frac{1}{N} \sum_{i=1}^N (U(G_i) - L(G_i))$ (34)
F Value	$F = \frac{PICP \times 2 \times \frac{1}{MPIW}}{PICP + \frac{1}{MPIW}}$ (35)
Average Relative Interval Length	$ARIL = \frac{1}{N} \sum_{i=1}^N \frac{(U(G_i) - L(G_i))}{G^m_{,i}}$ (36)
Winkler Score	$WS = \begin{cases} \Delta_i & L(G_i) \leq y_i \leq U(G_i) \\ \Delta_i + 2(L(G_i) - y_i)/\alpha & y_i < L(G_i) \\ \Delta_i + 2(y_i - U(G_i))/\alpha & y_i > U(G_i) \end{cases}$ (37)
	where $\Delta_i = U(G_i) - L(G_i)$ (38)
Normalized Mean Prediction Interval Width	$PINAW = \frac{1}{N * R} \left(\sum_{i=1}^N (U(G_i) - L(G_i)) \right)$ (39)
Continuous Rank Probability Score (MW)	$CRPS = \frac{1}{N} \sum_{i=1}^N crps(F_i, y_i)$ (40)
	where $crps(F, y) = \int_{-\infty}^{\infty} (F(t) - 1(t - y))^2 dy$ (41)

calculated using six different metrics.

$$GPI_i = \sum_{j=1}^6 \alpha_j (g_j - y_{ij}) \quad (42)$$

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

The model performance was further evaluated using the direction of movement measured by a Directional Symmetry (DS). The DS provides a statistical measure of a model's ability to predict the direction of change, positive or negative, of a time series from one value to the next value:

$$DS = \frac{1}{n} \sum_{t=2}^n d_t \times 100\% \quad (43)$$

where,

$$d_t = \begin{cases} 1 & \text{if } (G_t^m - G_{t-1}^m)(G_t^p - G_{t-1}^m) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (44)$$

We also adopt the Promoting Percentage of Absolute Percentage Bias λ_{APB} , Mean Absolute Error λ_{MAE} , and RMSE λ_{RMSE} [101] as follows:

$$\lambda_{APB} = \left| \frac{(APB_1 - APB_2)}{APB_1} \right| \quad (45)$$

$$\lambda_{KGE} = \left| \frac{(KGE_1 - KGE_2)}{KGE_1} \right| \quad (46)$$

$$\lambda_{RMSE} = \left| \frac{(RMSE_1 - RMSE_2)}{RMSE_1} \right| \quad (47)$$

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.

We employed Diebold-Mariano (DM) as well as Harvey, Leybourne and Newbold (HLN) test to demonstrate superiority of the CELM-KDE against all benchmark models following a statistical approach. Note that both statistic should exceed zero for the best model [102, 103, 104].

4. Results and Discussion

In this section we appraise the efficacy of the proposed hybrid CELM-KDE model for its ability to predict point-based daily electricity demand, as well as the confidence intervals of G predictions. The proposed model is also evaluated across geographically diverse study sites in southeast Queensland, Australia. Various statistical performance measured and visual analysis of observed and predicted G values are compared for the hybrid CELM-KDE as well as the benchmark model.

4.1. Point Electricity Demand Prediction Results

In this section, we compare the proposed hybrid CELM-KDE model with the benchmark models using diagnostic plots and performance metrics to assess its efficacy for daily G prediction. Tables 5 shows the values of r , $RMSE(MW)$, $MAE(MW)$, $STDRE(\%)$, and E_{var} for CELM-KDE model in comparison with DNN, RFR, SVR, LSTM and KNNR models for four stations in southeast Queensland. In terms of metrics in Table 5, the proposed CELM-KDE model outperforms the alternative models for all sub-stations. For instance, the CELM-KDE model obtains an $r \approx 0.946$ and a $RMSE \approx 16.422$ for Cornubia sub-stations which are substantially better values than the ones obtained by the method in the second-position (the LSTM) with $r \approx 0.929$ and a $RMSE \approx 19.120$. Thus, it is proven that the proposed CELM-KDE model could be used as a well-designed prediction tool for daily G prediction.

Even though the hybrid CELM-KDE models used in daily electricity demand (G) prediction seems to perform dramatically better than the alternative methods, it is important to note that a direct comparison is deemed irrational due to geographic differences and subsequently different G characteristics at these study sites. These differences can prevent an inter-site comparison in the absence of normalised error metrics. We have therefore used the relative errors, i.e., $RRMSPE$ and $RMAPE$, as shown in Tables 6, to overcome this problem and enable comparison of the model performance at geographically disparate sites.

We now revert to the Nash-Sutcliffe coefficient (E_{NS}) and Legates and McCabe's Index (E_{LM}) to evaluate the proposed CELM-KDE model as these metrics are considered superior to Willmott's Index (E_{WI}) especially when reasonably large values are anticipated even for a poorly fitted model (Table 7). In this study, we have used (E_{WI}) as a modified E_{WI} where errors and discrepancies receive proper weighting instead of being exaggerated by square values. Compared to the DNN, RFR, SVR, LSTM and KNNR models for all four substations, the comparative performance in Table 6 reveals that the hybrid CELM-KDE model exhibits the lowest $RRMSPE$ and $RMAPE$. For instance, the CELM-KDE model for the study site Cornubia has produced a lower $RRMSPE$ ($\approx 8.896\%$) and $RMAPE$ ($\approx 7.258\%$) compared to other models ($RRMSPE$ and $RMAPE \geq 10\%$). Furthermore, the hybrid CELM-KDE model performs well at all four substations with high magnitudes of (e.g. Lawnton substation) $E_{WI} \approx 0.909$, $E_{NS} \approx 0.857$, and $E_{LM} \approx 0.627$. These metrics for the other models are of lower magnitude.

Table 5: The testing performance of the proposed hybrid CELM-KDE model *vs.* the DNN, RFR, SVR, LSTM, and KNNR, as benchmark models measured by the correlation coefficient (r), root mean square error ($RMSE$) and the mean absolute error (MAE) between observed and predicted daily G values.

Sub-Stat.	Model	Model Performance Metrics				
		r	$RMSE(MW)$	$MAE(MW)$	$STDRE$	E_{var}
Coolum	CELM	0.935	31.069	24.347	3.87%	0.764
	DNN	0.922	34.621	26.535	4.24%	0.720
	RFR	0.926	33.974	25.787	4.25%	0.730
	SVR	0.925	33.896	26.265	4.17%	0.727
	LSTM	0.928	32.809	25.179	4.07%	0.742
	KNNR	0.910	36.412	28.366	4.36%	0.679
Cornubia	CELM	0.946	16.422	12.927	5.93%	0.798
	DNN	0.882	24.681	19.831	10.75%	0.597
	RFR	0.933	19.162	15.214	7.77%	0.750
	SVR	0.943	21.286	17.099	9.72%	0.734
	LSTM	0.929	19.12	15.117	7.14%	0.743
	KNNR	0.917	22.885	18.675	9.49%	0.689
Caloundra	CELM	0.944	67.215	53.402	4.34%	0.788
	DNN	0.917	81.117	65.092	5.26%	0.705
	RFR	0.935	74.562	60.29	4.86%	0.749
	SVR	0.939	72.71	58.769	4.73%	0.772
	LSTM	0.938	71.529	56.682	4.71%	0.771
	KNNR	0.912	92.393	75.88	6.19%	0.641
Lawnton	CELM	0.962	38.412	29.845	3.87%	0.857
	DNN	0.954	43.218	33.702	4.76%	0.829
	RFR	0.948	45.886	36.021	4.85%	0.806
	SVR	0.929	53.573	42.241	5.73%	0.745
	LSTM	0.952	42.967	33.634	4.62%	0.822
	KNNR	0.924	53.974	44.743	5.70%	0.728

Most significantly, $RRMSPE$ and E_{NS} values for all four substations are less than 10% and ≥ 0.8 , respectively, using the hybrid CELM-KDE which indicates a high-performance model. In light of this, it was clear that the hybrid CELM-KDE model outperformed the DNN, LSTM, and all other alternative models.

The three model performance metrics, i.e., $U95$, TS , and TIC , are shown by their respective values in Table 8. As it can be observed, the hybrid CELM-KDE model encountered lesser uncertainty compared to the other prediction models with ≈ 86.1167 and ≈ 90.513 for the CELM-KDE and LSTM, respectively indicating that the former model is more generalisable compared with benchmark models. Additionally, the value of TS shows significant disparities between the models' performance for all four substations, e.g. for Cornubia: the value of TS was ≈ 1.217 , ≈ 7.110 , ≈ 6.220 , ≈ 10.127 , ≈ 5.133 , ≈ 9.893 for CELM, DNN, RFR, SVR, LSTM, and KNNR models, respectively. Accordingly, Theil's Inequality Coefficient (TIC) is another criterion to measure the prediction accuracy used in this study. Note that the TIC values nearer to 0 are expected to show a higher predictive capability. The hybrid CELM-KDE model seems to excel at all of the substations based on the value of TIC , which are much lower for the CELM-KDE compared with other models. Additionally, using the DM and HLN tests, we note that our model predicted G is hypothetically evaluated against DNN, RFR, SVR, LSTM and KNNR models. The DM and HLN values, as listed in Table 9, support the claim that the proposed CELM-KDE model is more accurate than all alternative models. These analysis show that the CELM model's capacity is acceptable in respect to producing more accurate G predictions to concur with previous results evaluated using deterministic performance indicators.

Table 6: Comparison of the accuracy of proposed hybrid CELM-KDE model *vs.* the benchmark models using relative errors ($RRMSPE$, %) and ($RMAPE$, %) for geographically diverse tested sites. The optimal model is boldfaced.

Sub-Stations	Predictive Model	Model Performance Metrics	
		$RRMSPE(\%)$	$RMAPE(\%)$
Coolum	CELM	6.428%	5.000%
	DNN	7.162%	5.360%
	RFR	7.028%	5.206%
	SVR	7.012%	5.356%
	LSTM	6.788%	5.119%
	KNNR	7.533%	5.768%
Cornubia	CELM	8.896%	7.258%
	DNN	13.370%	11.954%
	RFR	10.380%	8.962%
	SVR	11.531%	10.508%
	LSTM	10.357%	8.606%
	KNNR	12.397%	11.132%
Caloundra	CELM	6.688%	5.435%
	DNN	8.071%	6.704%
	RFR	7.419%	6.228%
	SVR	7.235%	6.080%
	LSTM	7.117%	5.844%
	KNNR	9.193%	7.954%
Lawnton	CELM	6.371%	4.965%
	DNN	7.168%	5.755%
	RFR	7.610%	6.087%
	SVR	8.885%	7.200%
	LSTM	7.126%	5.708%
	KNNR	8.952%	7.716%

Figure 10 presents a thorough evaluation of the proposed CEM-KDE model by showing the frequency distribution of absolute prediction error ($|PE|$) generated by the CELM-KDE model *vs.* the alternative models. For all four substations, the CELM-KDE model achieved $|PE|$ that were below the lowest range ($\pm 25MW$, $\pm 15MW$, $\pm 60MW$, and $\pm 25MW$ for Coolum, Cornubia, Caloundra and Lawnton, respectively).

Figure 11 shows a box plot of $|PE|$ for the proposed CELM-KDE as well as benchmark models. Importantly, we note a smaller PE for the case of CELM-KDE. Furthermore, to acquire a clearer view of the residual distributions, the Kernel Smoothing Density Function ($KSDF$) plots of standardised residuals are presented (Figure 12). Notably, the hybrid CELM-KDE model's standardised residual $KSDF$ plot is quite similar to the conventional norms. We also show the Empirical Cumulative Distribution Function ($ECDF$) in Figure 13 to acquire a transparent representation of PE distributions. Evidently, the $ECDF$ line plots of the proposed CELM-KDE model has a very close profile for all substations compared with benchmark models. The superior efficacy of the hybrid CELM model in predicting daily G is further determined for all four substations in respect to the frequency plot (Figure 10), box-plot (Figure 11), $KSDF$ and $ECDF$ plot (Figures 12 and 13), all suggesting the CELM-KDE model exceeds the performance of the benchmark models.

Table 7: Evaluation of the proposed hybrid CELM-KDE model using the Willmott’s Index (E_{WI}), Nash–Sutcliffe Coefficient (E_{NS}) and the Legates & McCabe’s (E_{LM}) Index. The best model is boldfaced.

Sub-Stations	Predictive Model	Model Performance Metrics		
		E_{WI}	E_{NS}	E_{LM}
Coolum	CELM	0.855	0.764	0.538
	DNN	0.833	0.710	0.496
	RFR	0.839	0.721	0.510
	SVR	0.836	0.721	0.501
	LSTM	0.851	0.738	0.522
	KNNR	0.802	0.677	0.461
Cornubia	CELM	0.877	0.797	0.562
	DNN	0.553	0.565	0.328
	RFR	0.788	0.730	0.485
	SVR	0.652	0.682	0.421
	LSTM	0.818	0.729	0.488
	KNNR	0.653	0.636	0.367
Caloundra	CELM	0.860	0.787	0.552
	DNN	0.802	0.693	0.454
	RFR	0.801	0.740	0.494
	SVR	0.817	0.755	0.507
	LSTM	0.831	0.761	0.525
	KNNR	0.618	0.613	0.363
Lawnton	CELM	0.909	0.857	0.627
	DNN	0.875	0.820	0.579
	RFR	0.852	0.798	0.550
	SVR	0.787	0.728	0.472
	LSTM	0.875	0.821	0.580
	KNNR	0.779	0.720	0.441

Table 8: Evaluation of the proposed hybrid CELM-KDE model in terms of the Uncertainty at 95% confidence ($U95$), t-Statistics (TS) and the Thiel's Inequality Coefficient (TIC)

Sub-Stations	Predictive Model	Model Performance Metrics		
		$U95$	TS	TIC
Coolum	CELM	86.167	-0.421	0.032
	DNN	94.952	-4.121	0.036
	RFR	93.168	-4.139	0.035
	SVR	93.301	-3.377	0.035
	LSTM	90.513	-2.830	0.034
	KNNR	100.700	-2.092	0.038
Cornubia	CELM	45.505	1.217	0.044
	DNN	66.332	7.110	0.064
	RFR	51.854	6.220	0.050
	SVR	55.691	10.127	0.055
	LSTM	52.128	5.133	0.050
	KNNR	60.008	9.893	0.059
Caloundra	CELM	185.940	1.978	0.033
	DNN	222.010	4.497	0.040
	RFR	204.480	4.131	0.037
	SVR	197.050	6.017	0.035
	LSTM	195.700	4.559	0.035
	KNNR	249.000	6.759	0.045
Lawnton	CELM	106.480	-0.959	0.031
	DNN	118.090	4.776	0.035
	RFR	125.570	4.517	0.037
	SVR	145.420	5.802	0.043
	LSTM	118.970	1.621	0.035
	KNNR	148.250	3.840	0.044

Table 9: Evaluation of the proposed hybrid CELM-KDE model in terms of: (a) Diebold–Mariano (DM) test statistic, (b) Harvey–Leybourne–Newbold (HLN) test statistic. The column is compared with rows so and if the result is positive, the model in the rows outperforms the model in the column.

a)

	CELM	DNN	RFR	SVR	LSTM	KNNR
CELM		4.12	4.784	3.194	5.59	5.833
DNN			-4.529	-5.376	-3.558	-1.555
RFR				1.974	-0.075	5.007
SVR					-1.475	1.717
LSTM						3.736

b)

	CELM	DNN	RFR	SVR	LSTM	KNNR
CELM		4.326	5.023	3.353	5.869	6.124
DNN			-4.755	-5.644	-3.735	-1.633
RFR				2.072	-0.079	5.257
SVR					-1.548	1.802
LSTM						3.922

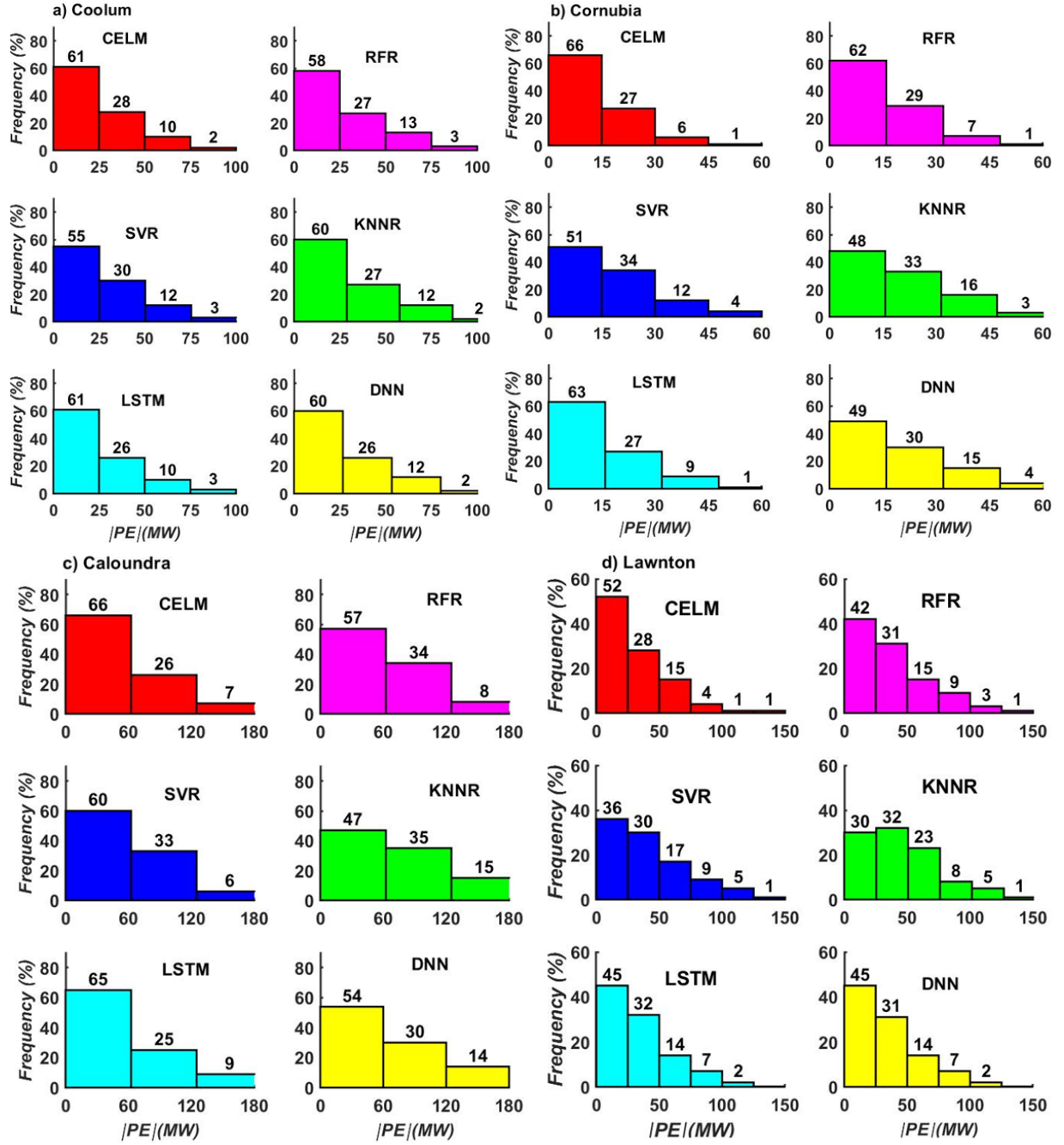


Figure 10: Cumulative frequency of absolute prediction error ($|PE|$) generated by hybrid CELM-KDE compared with benchmark models.

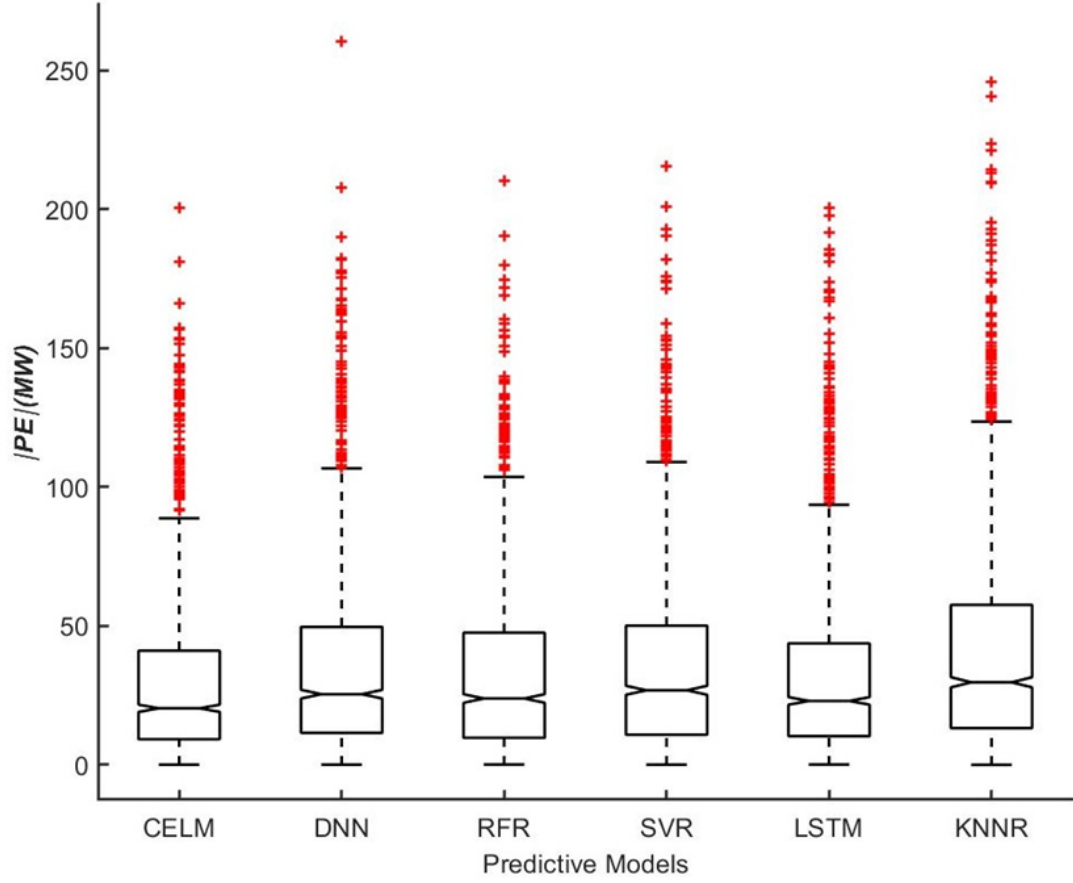


Figure 11: Box plot of the absolute prediction error ($|PE|$) of daily G generated by the hybrid CELM-KDE compared with benchmark models.

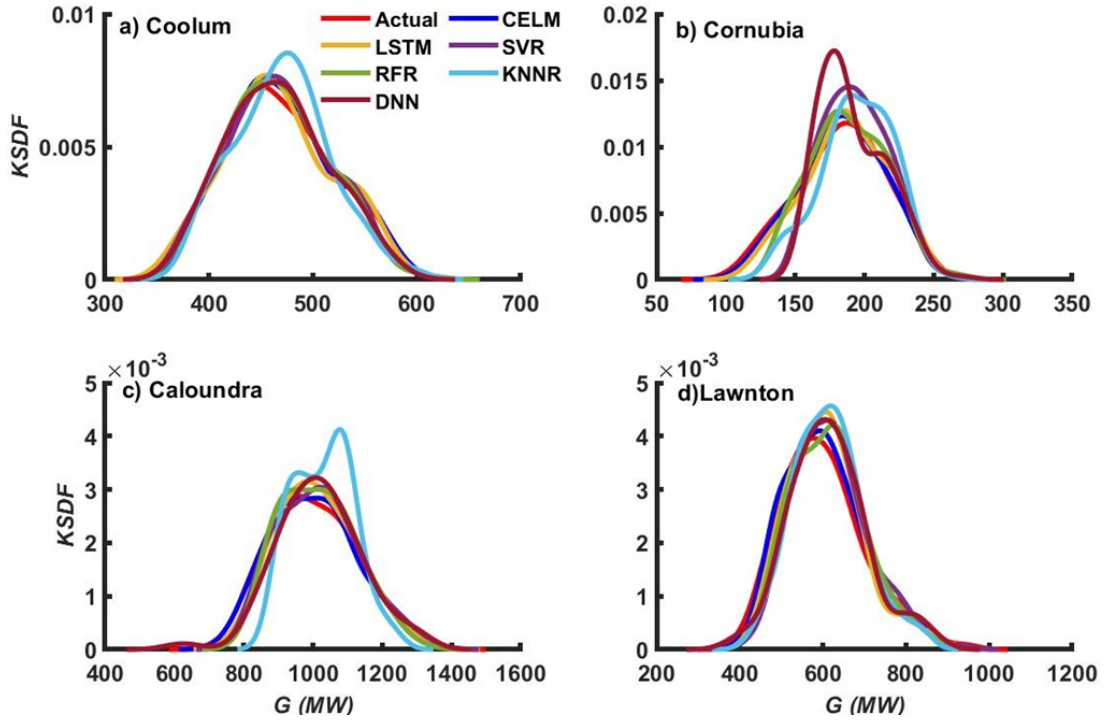


Figure 12: Kernel Smoothing Density Function ($KSDF$) of predicted and actual G (MW), generated by the proposed the proposed hybrid CELM-KDE *vs.* the benchmark models.

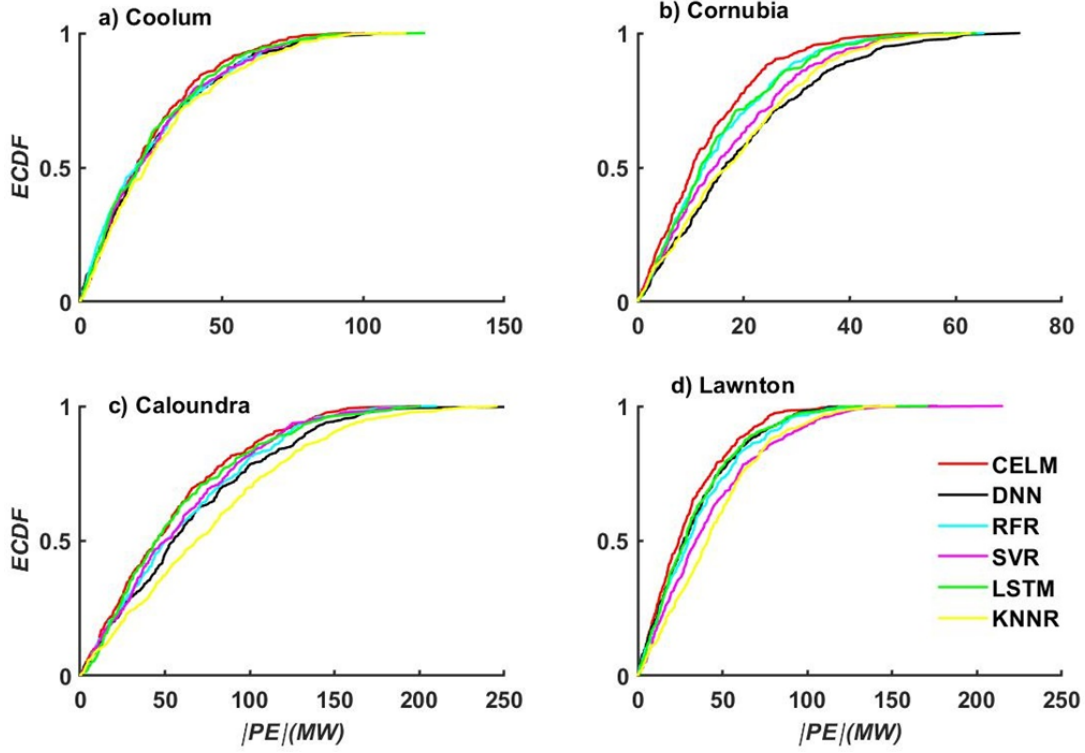


Figure 13: Empirical Cumulative Distribution Function ($ECDF$) of predicted and actual $G(MW)$, generated by the proposed hybrid CELM-KDE *vs.* the benchmark models.

Using KGE , APB , GPI , and CPI , we now reaffirm the accuracy of the the hybrid CELM model. In particular, the KGE metric aims to tackle several issues found in case of E_{NS} by a breakdown of E_{NS} into its component pieces such as correlation, variability bias, and mean bias error. These findings demonstrate that the hybrid CELM-KDE model majorly outperforms the counterpart models, see Figure 14 where a reasonably high KGE and a relatively low APB are evident. Furthermore, as each of the deterministic indicators mentioned so far have benefits and constraints, it is perhaps challenging to assess the accuracy of several models with a single metric.

We now revert to GPI as a metric proposed by Despotovic et al. [100] to address the above issues. In general, GPI is adopted to rank the overall models' performance whereby a high value is expected to indicate a greater model accuracy. Figure 15 shows this metric where the proposed hybrid CELM-KDE model is seen to outperform all of the benchmark models, indicated by a high GPI at all sub-stations. As additional measure of the hybrid CELM-KDE model accuracy, we show CPI as a metric that can discriminate various models more strictly by combining traditional information on dispersion and bias as $RMSE$) and providing information on distribution resemblance through KSI and $OVER$. With the lowest CPI at all sub-stations tested, Figure 16 reaffirm that the hybrid CELM-KDE model outperforms all benchmark models quite significantly.

Table 10 shows the Promotion Percentages (λ) computed to assess the proposed CELM-KDE model whereby λ provides the relative % improvement of this objective model against all benchmark models. Similar to previously reported findings (Tables 5-9), the hybrid CELM-KDE models exhibits a superior performance at all sub-stations. It should be noted that DNN, RFR, SVR, LSTM and KNNR, for example, is considerably worse in terms of RMSE for the case of Cornubia by $\approx 50\%$, $\approx 16\%$, $\approx 29\%$, $\approx 16\%$, and $\approx 39\%$, respectively, when compared to the hybrid CELM-KDE model (see Table 10). Consequently, with positive promoting percentage errors, we note that the hybrid CELM-KDE model seems to have a better predictive skill.

Figure 17 shows the directional similarity DS of the proposed hybrid CELM model where

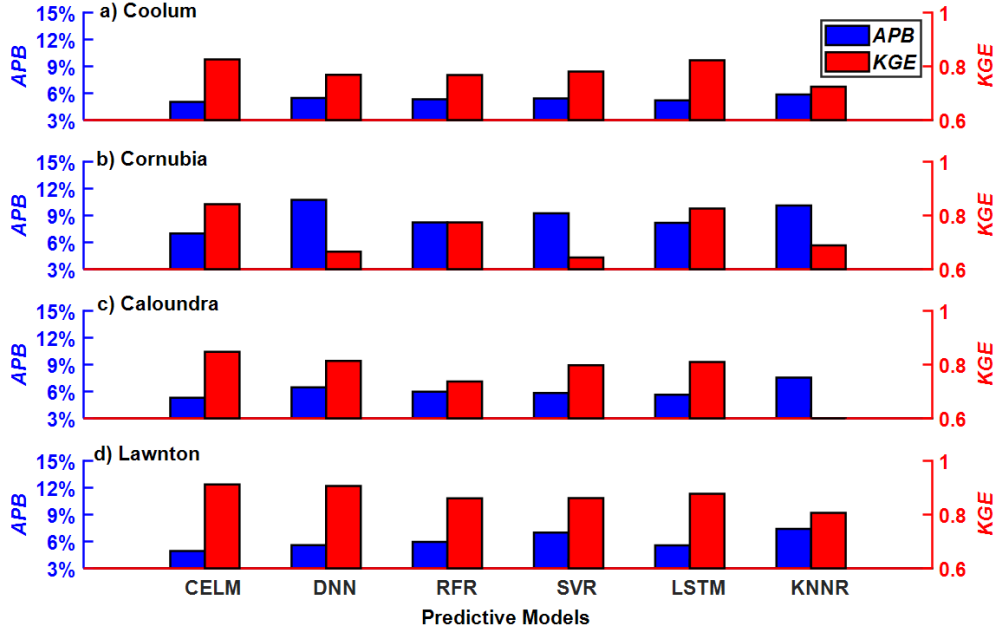


Figure 14: Bar chart in terms of absolute percentage bias APB , % and Kling-Gupta Efficiency KGE used to evaluate the proposed hybrid CELM-KDE model.

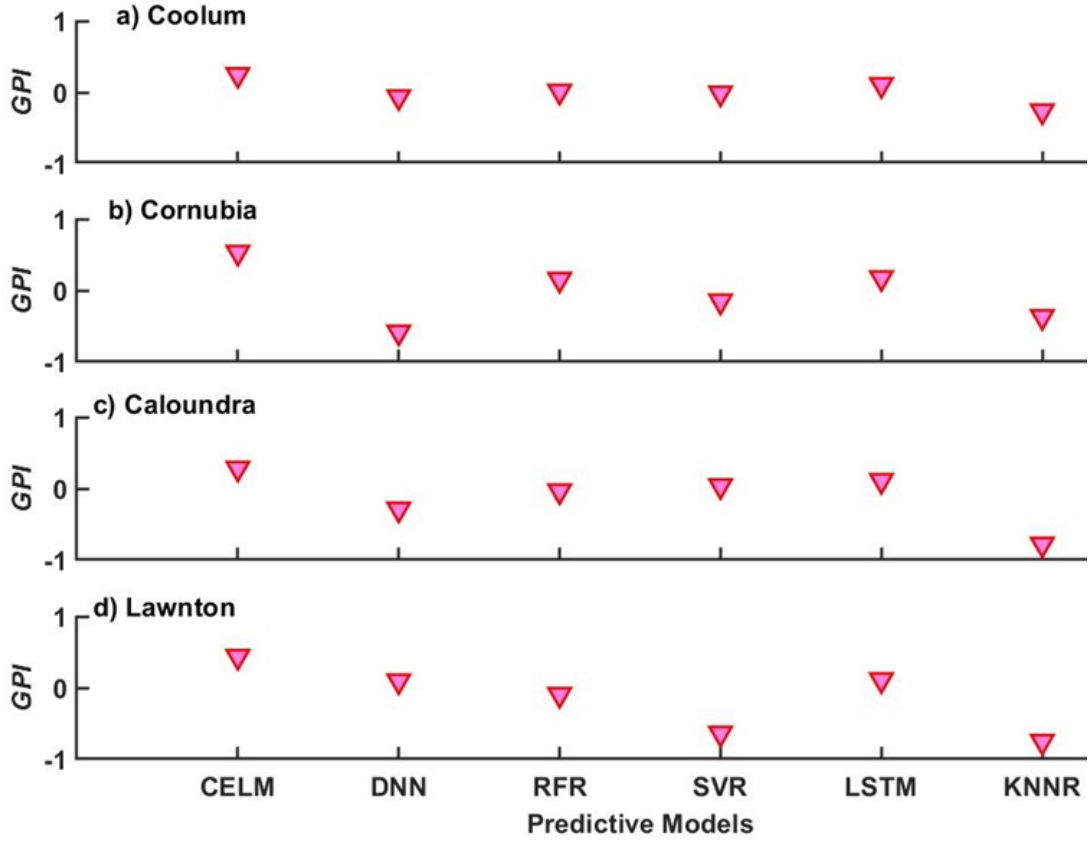


Figure 15: Global Performance Indicator GPI comparing the proposed hybrid CELM-KDE with the benchmark models.

the magnitude of this metric is greater than other models with an average of $\approx 82.5\%$. To concur with this finding, the Taylor diagram shown in Figure 18 that maps the correlation (r) enables us a complementary assessment of the model performances. In terms of standard deviation

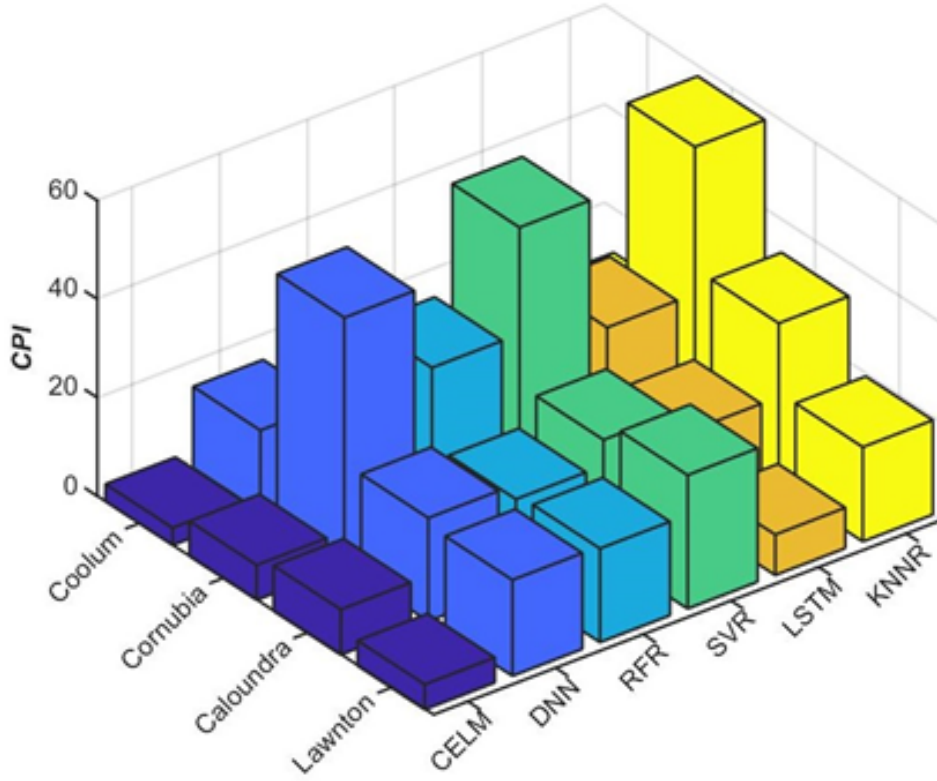


Figure 16: The Combined Performance Index CPI comparing the hybrid CELM-KDE with the benchmark models.

(SD), the Taylor diagram provides a tangible and compelling statistical comparison between predicted and observed G based on r . In accordance with earlier findings, and considering the positive λ , Taylor plot and DS metrics, we can assert with confidence that the hybrid CELM-KDE model is significantly capable in predicting daily G data.

4.2. Electricity Demand Interval Predictions

Although the point predictions can help us to predict the recorded G value as a definitive (or deterministic) test point, there are several factors that can contribute to their underlying predicted uncertainties. We have therefore used the interval prediction approach by selecting the prediction error (PE) through a Kernel Density Estimation (KDE) distribution test on all PE values. The purpose of this was to better understand the prediction intervals at the 95% confidence level where a Gaussian distribution was used with bandwidth determined using the Sliverman's rule of thumb.

In Figure 19, we now show a bar chart representing the $PICP$ and $PINAW$ which are probabilistic measures of prediction intervals providing a range of values that are likely to contain the deterministically predicted value or a single observation given the specified settings of the predictors. To interpret this result, consider for example, a 95% prediction interval of $[5,10]$ is expected to show that the model is 95 % confident that the next new observation will fall within this range of electricity demand.

Evidently, for the data tested at the Coolum substation examined at the 95% confidence level, the magnitude of the $PICP$ generated by the hybrid CELM-KDE, as well as DNN, RFR,

Table 10: The promoting percentage metric, for the comparison models against objective (i.e., CELM) model in the testing phase, showing the improvement of the proposed method in respect to benchmark models. Note that λ_{RMSE} Promoting Percentage 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	Coolum			Cornubia			Caloundra			Lawnton		
	λ_{RMSE}	λ_{APB}	λ_{KGE}	λ_{RMSE}	λ_{APB}	λ_{KGE}	λ_{RMSE}	λ_{APB}	λ_{KGE}	λ_{RMSE}	λ_{APB}	λ_{KGE}
DNN	11.43%	8.99%	6.89%	50.29%	53.41%	21.02%	20.68%	21.89%	4.02%	12.51%	12.92%	0.60%
RFR	9.35%	5.92%	7.06%	16.68%	17.69%	8.06%	10.93%	12.90%	13.06%	19.46%	20.69%	5.65%
SVR	9.10%	7.88%	5.43%	29.62%	32.27%	23.57%	8.18%	10.05%	5.95%	39.47%	41.53%	5.57%
LSTM	5.60%	3.42%	0.37%	16.43%	16.94%	1.98%	6.42%	6.14%	4.48%	11.86%	12.69%	3.81%
KNNR	17.20%	16.51%	12.27%	39.36%	44.46%	18.18%	37.46%	42.09%	29.60%	40.51%	49.92%	11.56%

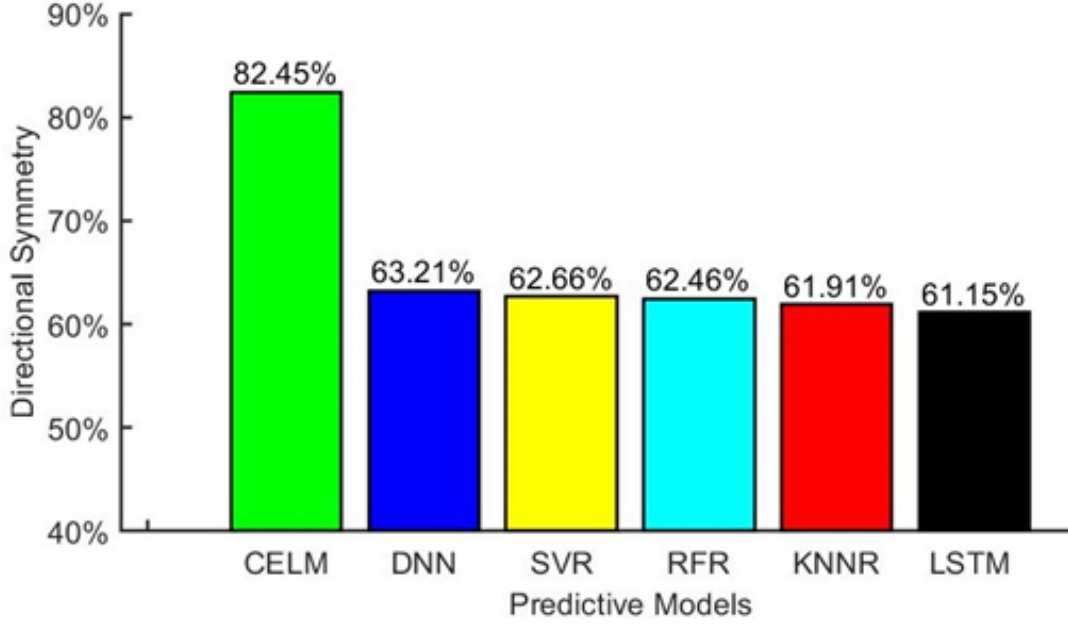


Figure 17: Directional Symmetry DS comparing proposed hybrid CELM-KDE with the benchmark models.

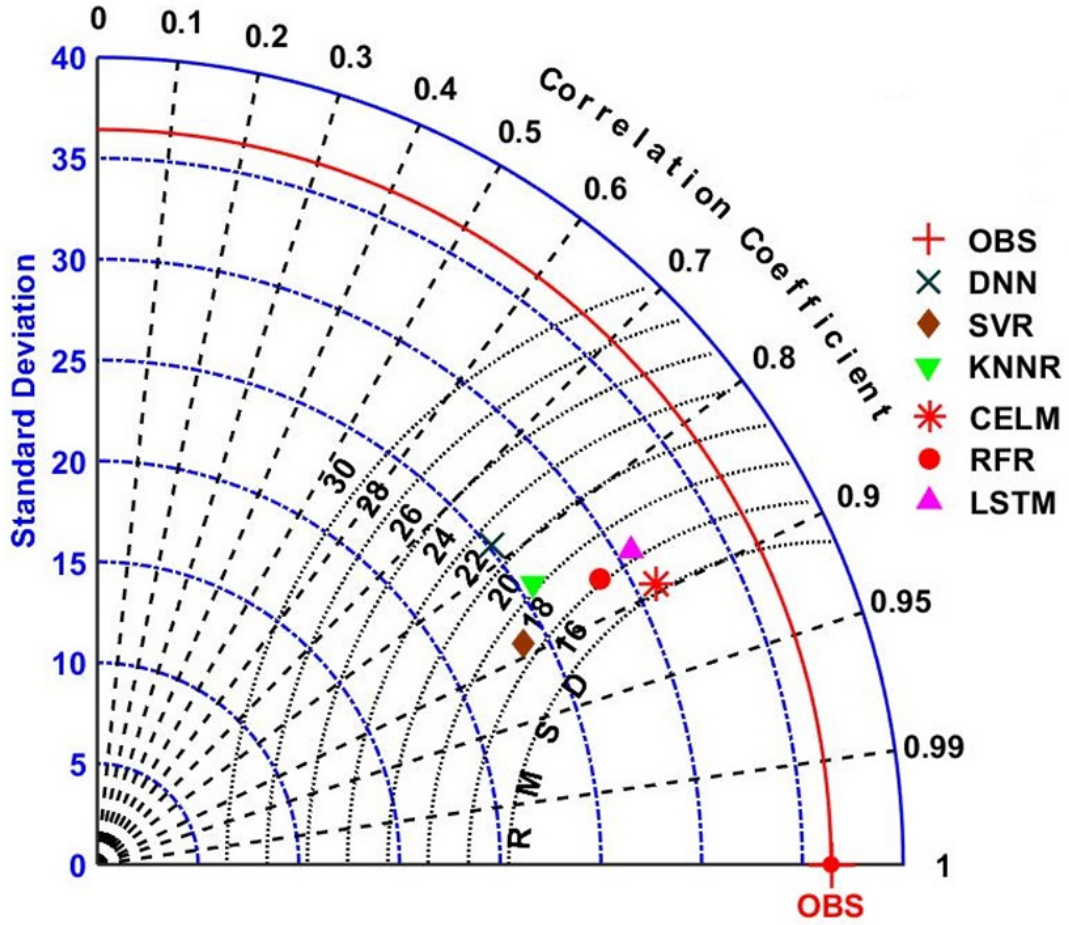


Figure 18: Taylor diagram depicting correlation coefficient r , standard deviation SD and root mean square $RMSE$ deviation comparing the proposed hybrid CELM-KDE with the benchmark models.

SVR, LSTM and KNNR are $\approx 95.6\%$, $\approx 96.4\%$, $\approx 95.9\%$, $\approx 95.6\%$, $\approx 95.6\%$, and $\approx 95.9\%$, respectively.

Since the present *PICP* for all models appear to exceed the 95% confidence interval, we can ascertain that all of the developed models tend to satisfy the requirement of the confidence level with $PICP \geq 95\%$. Notably, the DNN model has the highest *PICP* with $\approx 96.4\%$, for Coolum whereas *PINAW* generated by the hybrid CELM-KDE, as well as the DNN, RFR, SVR, LSTM and KNNR models are ≈ 0.392 , ≈ 0.419 , ≈ 0.413 , ≈ 0.414 , ≈ 0.403 , and ≈ 0.449 , respectively. Even though *PICP* for DNN is high, the *PINAW* value is somewhat higher at ≈ 0.419 . Based on *PINAW* criteria, we can conclude that the newly developed CELM-KDE model is the best one with $PICP \geq 95\%$ and a smaller interval width ($PINAW \approx 0.392$).

To interpret the *PI* values that have an exceptional quality of predictions, one should also expect a wide coverage probability with a shorter interval width and this could also mean an excellent interval prediction results with a large value of *PICP* and a small value of the resulting *PINAW*.

To fulfil the above, we now assess coverage probability and width of prediction interval thoroughly by means of the *F*-value that integrates both the *PICP* and the *PINAW* for all present models. It is imperative to note that the *F* – value generated by the proposed hybrid CELM-KDE model far exceeds that of the DNN, RFR, SVR, LSTM and KNNR models with ≈ 1.391 , ≈ 1.373 , ≈ 1.374 , ≈ 1.370 , ≈ 1.381 , and ≈ 1.341 , respectively as shown in Figure 20. With the highest value of *F* – value, we can ascertain that the hybrid CELM-KDE model has outperformed the benchmark models and this occurs for all of the present substations. We can therefore state with confidence that the proposed hybrid CELM-KDE model is able to generate daily electricity demand prediction intervals with a much greater quality.

Finally, we adopt the Winkler score (*WS*) and the Average Relative Interval Length (*ARIL*) of the PIs generated by all models at the four substations, as reported in Table 11.

At the 95% confidence level, the proposed hybrid CELM-KDE model registered the lowest value of *WS* and *ARIL*. For individual stations, we note these results: Coolum: *WS* ≈ 147.33 & *ARIL* ≈ 0.27 , Cornubia: *WS* ≈ 80.36 & *ARIL* ≈ 0.38 , Caloundra: *WS* ≈ 302.49 & *ARIL* ≈ 0.28 , and Lawnton: *WS* ≈ 185.34 & *ARIL* ≈ 0.27), and this demonstrates that the objective model certainly outperforms the comparative prediction models. In fact, the PIs for the daily *G* under 95% confidence level for different models are shown in Figure 22 where the hybrid CELM-KDE model is seen to be relatively superior to that of other models with a low value of *CRPS* and *PINAW*. Compared with alternative benchmark models, the proposed hybrid CELM-KDE model performs better at the 95% confidence levels and therefore can deliver more detailed prediction information on electricity demand for the present study sites.

In respect to load prediction models developed using various sources of data that can perhaps apply the proposed hybrid CELM-KDE model, Figure 21 shows a schematic that outlines different variables like plug-in electric vehicles, roof-top solar, electricity end-use trends, weather conditions, days of the week as well as economic trends. Forecasting electricity demand is essential to ensure a reliable power supply for present and future consumers so including such diverse datasets could further increase the practicality of the proposed hybrid CELM-KDE model. In general, load forecasting assists in power system operation for efficiently managing the bulk electric system, strategizing the necessary upgrades, and delivering dependable service at optimal costs. These predictions also support operators in making informed decisions for short-term market strategies and the long-term infrastructure investments. To predict consumer electricity needs, energy planners should analyze data from various sources, as illustrated in this figure. While we have currently focused on weather parameters, a future study can also plan to incorporate the remaining five parameters in future studies.

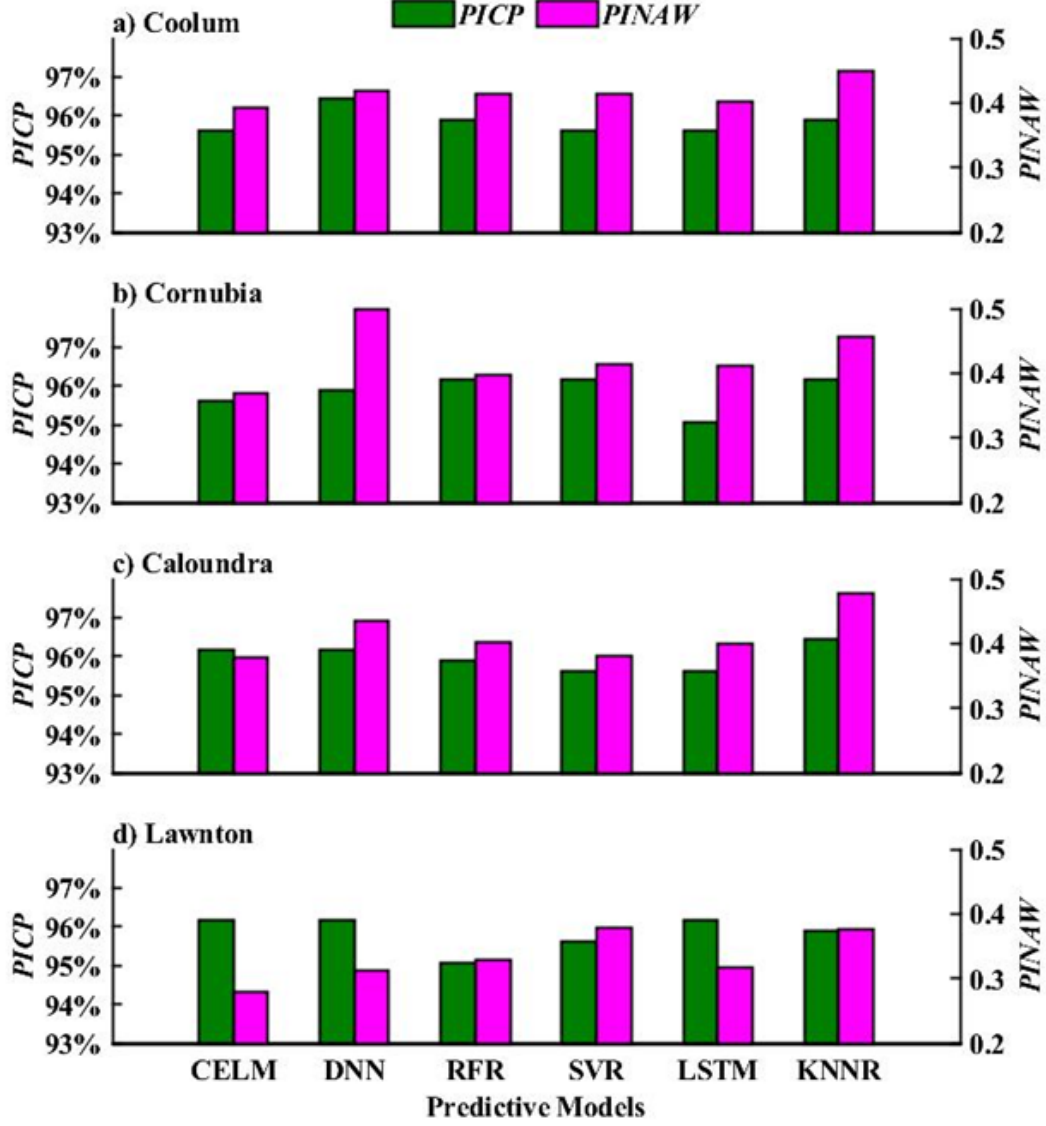


Figure 19: Interval prediction performance based on the $PICP$ and the $PINAW$ criteria applied to the proposed hybrid CELM-KDE *vs.* the benchmark models.

5. Conclusions and Research Future Work

This article has presented a new deep learning hybrid framework using Convolution Neural Network and successfully integrated with Extreme Learning Machines and Kernel Density Estimation approaches to create a hybrid CELM-KDE model that was able to predict daily electricity demand for stations in southeast Queensland. In the modeling of electricity demand, prediction intervals were calculated to measure the uncertainty in predicted values, so that power system operators, electricity energy utilities, and wholesale electricity markets could gain a better understanding of future predicted electricity demand risks. A number of well-known deep learning methods, such as DNN, RFR, SVR, LSTM, and KNNR, were evaluated against multiple years of electricity demand profiles in order to verify the overall superiority of the objective model.

This study has specifically generated a robust CELM-KDE model to estimate deterministic (point-based) electricity demand predictions. A combination of climate variables (temperature, humidity, solar radiation, evaporation, etc.) was added to the CELM-KDE model's inputs for

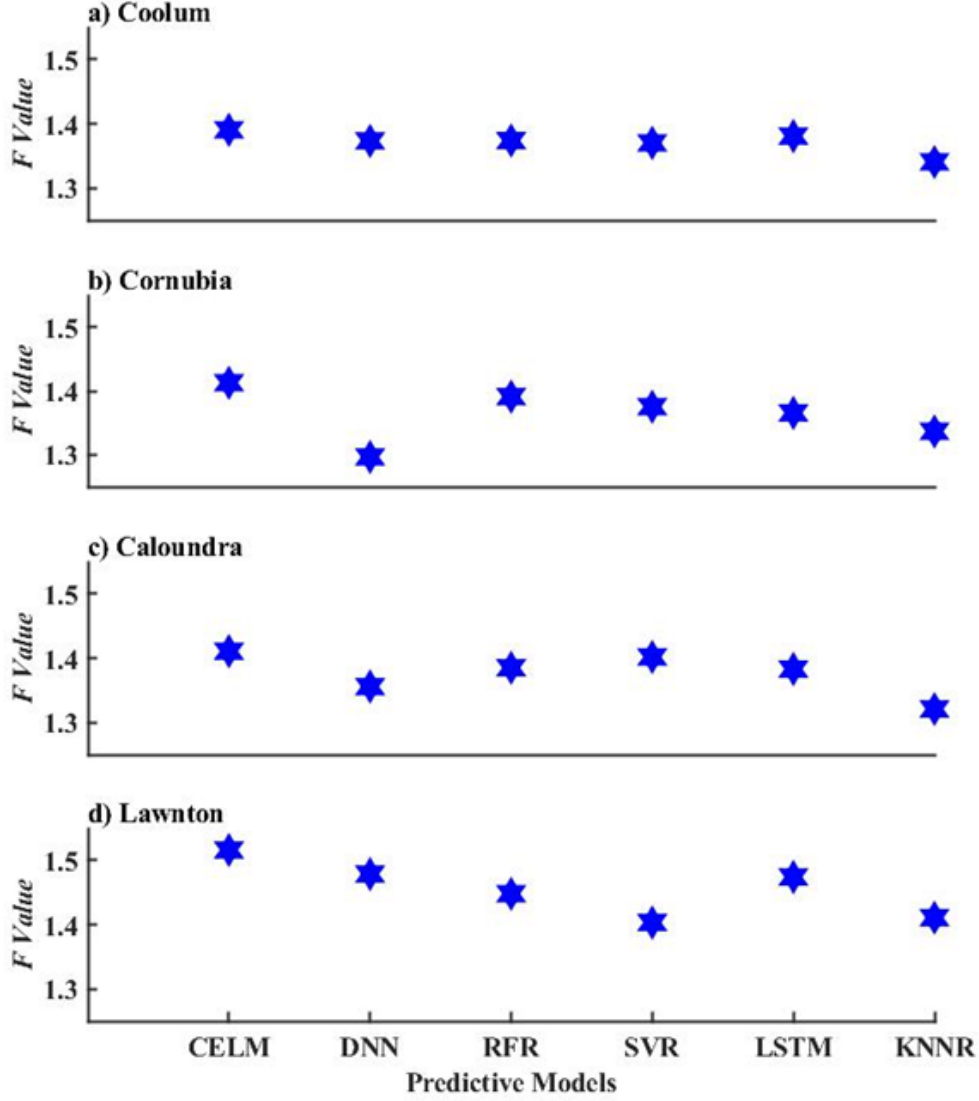


Figure 20: Interval prediction performance based on $F - Value$ generated by the proposed hybrid CELM-KDE *vs.* the benchmark models.

the analysis of sequential demand data, and partial autocorrelation and mutual information function tests were used to determine the length (lagged) of electricity demand sequence. Also, to build a two-stage framework to reliably predict daily electricity demand, the prediction errors were fitted using the Kernel Density Estimation method in order to acquire Prediction Intervals. Through the use of a variety of assessment criteria and diagnostic plots, we were able to thoroughly evaluate the performance of the hybrid CELM-KDE model over time. The results of this study are summarised as follows:

- Based on the low relative forecasting errors and high-performance metrics, the proposed hybrid CELM's accuracy is proven to be superior to competing models. The hybrid CELM-KDE model generated the lowest absolute and relative errors in its testing phase for all four substations, as well as the lowest uncertainty at 95% confidence t-Statistics and Thiel's Inequality Coefficient, while achieving the highest correlation coefficient, Willmott's, Nash and Legates Index, and Kling-Gupta efficiency.
- The proposed hybrid CELM-KDE model outperformed all benchmark models, as mea-

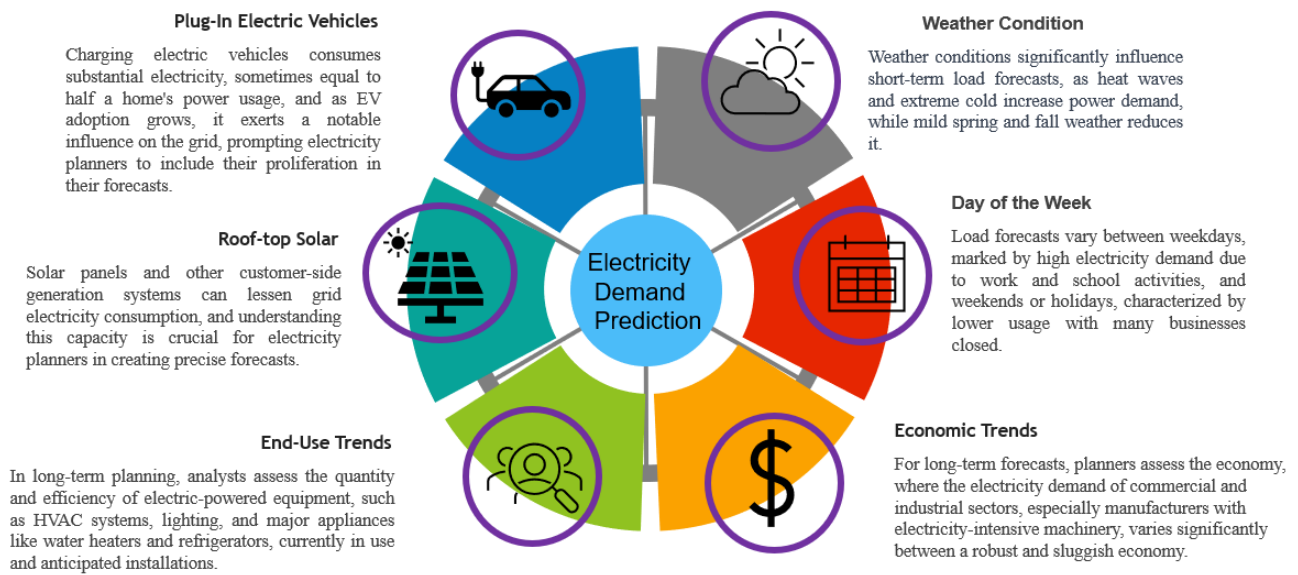


Figure 21: Schematic diagram of the load prediction model that can be developed using various sources of data.

Table 11: Probabilistic prediction at 95% confidence level using Winkler Score (WS) and Average Relative Interval Length ($ARIL$) computed in the testing phase of the proposed CELM-KDE and the benchmark models.

Sub-Stations	Predictive Model	Model Performance Metrics	
		WS	$ARIL$
Coolum	CELM	147.33	0.27
	DNN	159.13	0.29
	RFR	157.10	0.29
	SVR	154.36	0.29
	LSTM	157.30	0.28
	KNNR	163.83	0.31
Cornubia	CELM	80.36	0.38
	DNN	110.14	0.52
	RFR	89.59	0.41
	SVR	88.19	0.43
	LSTM	88.73	0.43
	KNNR	94.03	0.48
Caloundra	CELM	302.49	0.28
	DNN	358.73	0.32
	RFR	323.14	0.30
	SVR	314.99	0.28
	LSTM	333.71	0.30
	KNNR	386.46	0.35
Lawnton	CELM	185.34	0.27
	DNN	204.47	0.30
	RFR	215.03	0.31
	SVR	248.32	0.36
	LSTM	210.22	0.30
	KNNR	231.93	0.36

sured by the Promoting Percentage Error, Diebold-Mariano, and Harvey–Leybourne–Newbo statistics. Moreover, histograms and box plots of prediction error as well as their Kernel Smoothing Density Function and Empirical Cumulative Distribution Function plots in the testing phase highlighted the superiority of the proposed hybrid CELM-KDE model.

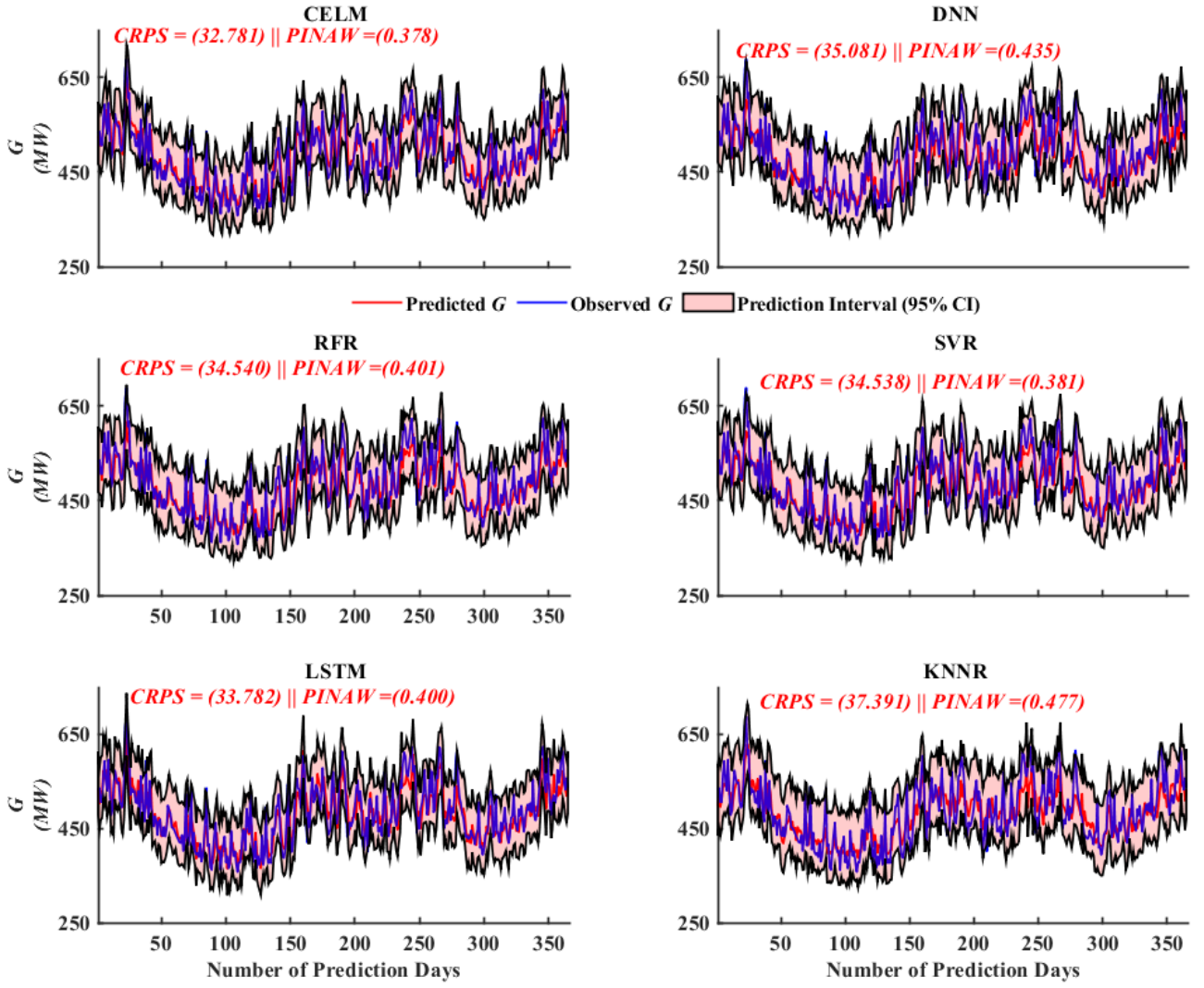


Figure 22: Electricity demand prediction interval under 95% confidence level generated by the proposed hybrid CELM-KDE model for different models (For Coolum sub-station).

- Thorough analysis of metrics such as global performance indicator, directional similarity, or other diagnostic plots compared predicted and observed electricity demand showed that the the proposed hybrid CELM-KDE model outperformed the comparable DNN, RFR, SVR, LSTM and KNNR models for daily prediction horizons.

To fully evaluate the proposed hybrid CELM-KDE model, an analysis of the quality of electricity demand predictions generated in respect to the comparison models, which in fact also assessed the prediction intervals, used the Winkler Score, Normalized Mean Prediction Interval Width, Prediction Interval Coverage Probability, and F-values. These probability-based indicators attained the the lowest value to ascertain the efficacy of the proposed CELM-KDE model. For example, the Winkler Score and the Average Relative Interval Length registered the lowest value for the proposed CELM-KDE model compared with five other benchmark models for all four study sites.

Among the probability-based indices that evaluated the quality of electricity predictions, the F – value also assessed the two mutually constrained factors including the Prediction Interval Coverage Probability and the Normalised Mean Prediction Interval Width. According to our findings, the objective CELM-KDE achieved better overall results proving its skill in enhancing the accuracy of electricity demand interval predictions. The proposed CELM-KDE model also has a superior ability to measure the underlying non-determinacy character of

electricity demand due to lower magnitude of Winkler Score and Average Relative Interval Length and relatively high magnitude of the F -value for all tested substations.

According to the analysis and results presented, the proposed hybrid CELM-KDE model appears to have a strong potential to accurately predict the daily electricity demand, at least in the context of four sub-stations (i.e., Coolum, Cornubia, Caloundra, Lawnton) in south-east Queensland. This framework was also able to handle short-and long-term relationships in electricity demand data as well as analysing the nonlinear associations and complexities with weather/climate variables that affect the electricity demand patterns. However, some limitations can be also find in this research: The present study has utilized a limited set of 11 climate-based predictors (see Table 2) to provide features impacting daily electricity demand. These variables were not monitored currently with electricity demand at tested sub-stations but generated later by gridding methods used in Scientific Information for Landowners (SILO) database, which could add uncertainties or errors in hybrid CELM-KDE model. A future study could therefore add a greater pool of variables from real-time weather observations made at the tested sites to improve the model's accuracy, as well as adding other variables to retrain the hybrid CELM-KDE model. Another limitation of the present study is the daily time-step used, which is a medium term prediction but is certainly limited in terms of providing more real-time, short-term predictions to manage the increased demand, black-outs or power failures. Therefore, a future study could retrain the hybrid CELM-KDE model at shorter-term time intervals such as hourly, or sub-hourly to construct a real-time predictive system for management of electricity demand. This could be achieved by using real-time weather observations such as European Centre for Medium-Range Weather Forecasts (hourly, 3-hourly, 6-hourly) as well as medium range (up to 15 days ahead), extended range (up to 46 days ahead) and long range (up to one year ahead) timescales [105].

Regarding the limitations in the proposed hybrid CELM-KDE method, we acknowledge that predictor (climate) and target (electricity demand) datasets used could have non-linear and non-stationary stochastic patterns, including jumps, periodicity, trends etc, which affects the overall ability of the CELM-KDE model to predict accurately. In a future study, this method could improved by incorporating a data decomposition method such as wavelet transforms and improved ensemble empirical mode decomposition [106] to split non-linear and non-stationary stochastic patterns prior to feeding them into the hybrid CELM-KDE model, which has been shown to improve the overall accuracy of machine learning models.

Since the proposed hybrid CELM-KDE model relies solely on the historical data, future studies may readily adopt the method to predict the electricity demand for a variety of demographic groups where different patterns of electricity use are noted. However, such future studies should retrain the model with more diverse datasets from those demographic regions to provide a robust predictive framework. In future, we will also examine the trends in electricity demand using a variety of nonlinear exogenous characteristics such as climatic and economic variables. Finally, to improve interval prediction's performance on both one-step and multi-step predictions, our future studies can also combine interval prediction with multi-step electricity demand prediction.

As a final note in respect to further applications, the present study has trained the hybrid CELM-KDE model only for electricity demand prediction by integrating convolution neural networks with extreme learning machine approaches and adopting a probabilistic prediction approach that can support energy companies in their regular operations. As we have used the Kernel Density Estimation method to enable probabilistic predictions with no prior assumptions to estimate the prediction interval (PI) and provide probability density functions for the predicted errors, the proposed method can be used in other areas such as electricity price predictions, solar, or wind energy simulations, including other time-series predictions where: (i) historical datasets of these target variables have repeating patterns that can be captured by re-

training the present CELM-KDE model and (ii) including exogenous (or related variable) data to provide added features such as real-time weather variables affecting electricity and other energy parameters.

In respect to wider applications, there is also the possibility of retraining the hybrid CELM-KDE model using other kinds of time-series variables such as stock market or financial analysis and predictions, hydrological applications like drought or flood prediction, and even radio communications areas where time-based message sequences can be transmitted. These wider applications will no doubt require robust retraining of the model to provide enough features for the model to predict accurately and resulting predictive errors to be minimised. These applications will be beneficial to society, especially in decision-making sectors where accurate models are needed by industry, policymakers, governments and other organizations.

Acknowledgement

This study was partially supported through the project PID2020-115454GB-C21 of the Spanish Ministry of Science and Innovation (MICINN).

Data Availability

Data were acquired from ENERGEX (<https://www.energex.com.au>). The developed code can be downloaded from: <https://github.com/sujanme25/CNN-ELM>.

References

- [1] D. Li, C. Xiao, X. Zeng, Q. Shi, Short-mid term electricity consumption prediction using non-intrusive attention-augmented deep learning model, *Energy Reports* 8 (2022) 10570–10581. doi:<https://doi.org/10.1016/j.egyr.2022.08.195>.
- [2] Y. Chen, M. Guo, Z. Chen, Z. Chen, Y. Ji, Physical energy and data-driven models in building energy prediction: A review, *Energy Reports* 8 (2022) 2656–2671. doi:<https://doi.org/10.1016/j.egyr.2022.01.162>.
- [3] Z. Zhuang, H. Tao, Y. Chen, V. Stojanovic, W. Paszke, An optimal iterative learning control approach for linear systems with nonuniform trial lengths under input constraints, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 53 (6) (2023) 3461–3473. doi:[10.1109/TSMC.2022.3225381](https://doi.org/10.1109/TSMC.2022.3225381).
- [4] R. Hu, S. Wen, Z. Zeng, T. Huang, A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm, *Neurocomputing* 221 (2017) 24–31. doi:<https://doi.org/10.1016/j.neucom.2016.09.027>.
- [5] N. Zeng, H. Zhang, W. Liu, J. Liang, F. E. Alsaadi, A switching delayed PSO optimized extreme learning machine for short-term load forecasting, *Neurocomputing* 240 (2017) 175–182. doi:<https://doi.org/10.1016/j.neucom.2017.01.090>.
- [6] H. Tao, J. Qiu, Y. Chen, V. Stojanovic, L. Cheng, Unsupervised cross-domain rolling bearing fault diagnosis based on time-frequency information fusion, *Journal of the Franklin Institute* 360 (2) (2023) 1454–1477. doi:<https://doi.org/10.1016/j.jfranklin.2022.11.004>.
- [7] E. Ceperic, V. Ceperic, A. Baric, A strategy for short-term load forecasting by support vector regression machines, *IEEE Transactions on Power Systems* 28 (4) (2013) 4356–4364. doi:[10.1109/TPWRS.2013.2269803](https://doi.org/10.1109/TPWRS.2013.2269803).

- [8] M. Zolfaghari, B. Sahabi, A hybrid approach to model and forecast the electricity consumption by neurowavelet and ARIMAX-GARCH models, *Energy Efficiency* 12 (8) (2019) 2099–2122. doi:<https://doi.org/10.1007/s12053-019-09800-3>.
- [9] J. M. Vilar, R. Cao, G. Aneiros, Forecasting next-day electricity demand and price using nonparametric functional methods, *International Journal of Electrical Power & Energy Systems* 39 (1) (2012) 48–55. doi:<https://doi.org/10.1016/j.ijepes.2012.01.004>.
- [10] H. Liu, J. Shi, Applying arma-garch approaches to forecasting short-term electricity prices, *Energy Economics* 37 (2013) 152–166. doi:<https://doi.org/10.1016/j.eneco.2013.02.006>.
- [11] M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, Y. Li, Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia, *Advanced Engineering Informatics* 35 (2018) 1–16. doi:<https://doi.org/10.1016/j.aei.2017.11.002>.
- [12] C. Sekhar, R. Dahiya, Robust framework based on hybrid deep learning approach for short term load forecasting of building electricity demand, *Energy* 268 (2023) 126660. doi:<https://doi.org/10.1016/j.energy.2023.126660>.
- [13] M. R. Baker, K. H. Jihad, H. Al-Bayaty, A. Ghareeb, H. Ali, J.-K. Choi, Q. Sun, Uncertainty management in electricity demand forecasting with machine learning and ensemble learning: Case studies of COVID-19 in the US metropolitans, *Engineering Applications of Artificial Intelligence* 123 (2023) 106350. doi:<https://doi.org/10.1016/j.engappai.2023.106350>.
- [14] G. Işık, H. Ögüt, M. Mutlu, Deep learning based electricity demand forecasting to minimize the cost of energy imbalance: A real case application with some fortune 500 companies in türkiye, *Engineering Applications of Artificial Intelligence* 118 (2023) 105664. doi:<https://doi.org/10.1016/j.engappai.2022.105664>.
- [15] X. Song, N. Wu, S. Song, V. Stojanovic, Switching-like event-triggered state estimation for reaction-diffusion neural networks against dos attacks, *Neural Processing Letters* (2023) 1–22doi:<https://doi.org/10.1007/s11063-023-11189-1>.
- [16] L. Ghelardoni, A. Ghio, D. Anguita, Energy load forecasting using empirical mode decomposition and support vector regression, *IEEE Transactions on Smart Grid* 4 (1) (2013) 549–556. doi:<https://doi.org/10.1109/TSG.2012.2235089>.
- [17] P. Pelka, Analysis and forecasting of monthly electricity demand time series using pattern-based statistical methods, *Energies* 16 (2) (2023) 827. doi:<https://doi.org/10.3390/en16020827>.
- [18] M. S. Al-Musaylh, R. C. Deo, Y. Li, J. F. Adamowski, 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, *Applied energy* 217 (2018) 422–439. doi:<https://doi.org/10.1016/j.apenergy.2018.02.140>.
- [19] A. Yang, W. Li, X. Yang, Short-term electricity load forecasting based on feature selection and least squares support vector machines, *Knowledge-Based Systems* 163 (2019) 159–173. doi:<https://doi.org/10.1016/j.knosys.2018.08.027>.

- [20] T. Ahmad, H. Chen, Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches, *Energy and Buildings* 166 (2018) 460–476. doi:<https://doi.org/10.1016/j.enbuild.2018.01.066>.
- [21] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, Y. Zhang, Short-term residential load forecasting based on LSTM recurrent neural network, *IEEE Transactions on Smart Grid* 10 (1) (2017) 841–851. doi:<https://doi.org/10.1109/TSG.2017.2753802>.
- [22] W. L. P. Jayasinghe, R. C. Deo, A. Ghahramani, S. Ghimire, N. Raj, Deep multi-stage reference evapotranspiration forecasting model: Multivariate empirical mode decomposition integrated with the Boruta-Random forest algorithm, *IEEE Access* 9 (2021) 166695–166708. doi:<https://doi.org/10.1109/ACCESS.2021.3135362>.
- [23] S. Ghimire, T. Nguyen-Huy, M. S. sekharusaylh, R. C. Deo, D. Casillas-Pérez, S. Salcedo-Sanz, A novel approach based on integration of convolutional neural networks and echo state network for daily electricity demand prediction, *Energy* 275 (2023) 127430. doi:<https://doi.org/10.1016/j.energy.2023.127430>.
- [24] S. Ghimire, T. Nguyen-Huy, R. Prasad, R. C. Deo, D. Casillas-Perez, S. Salcedo-Sanz, B. Bhandari, Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction, *Cognitive Computation* 15 (2) (2023) 645–671. doi:<https://doi.org/10.1007/s12559-022-10070-y>.
- [25] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, S. W. Baik, A novel CNN-GRU-based hybrid approach for short-term residential load forecasting, *IEEE Access* 8 (2020) 143759–143768. doi:<https://doi.org/10.1109/ACCESS.2020.3009537>.
- [26] A. Khosravi, S. Nahavandi, D. Creighton, A. F. Atiya, Comprehensive review of neural network-based prediction intervals and new advances, *IEEE Transactions on neural networks* 22 (9) (2011) 1341–1356. doi:<https://doi.org/10.1109/TNN.2011.2162110>.
- [27] G. Tang, Y. Wu, C. Li, P. K. Wong, Z. Xiao, X. An, A novel wind speed interval prediction based on error prediction method, *IEEE transactions on industrial informatics* 16 (11) (2020) 6806–6815. doi:<https://doi.org/10.1109/TII.2020.2973413>.
- [28] Y. Wang, H. Tang, T. Wen, J. Ma, Direct interval prediction of landslide displacements using least squares support vector machines, *Complexity* 7082594 (2020). doi:<https://doi.org/10.1155/2020/7082594>.
- [29] H. Quan, D. Srinivasan, A. Khosravi, Particle swarm optimization for construction of neural network-based prediction intervals, *Neurocomputing* 127 (2014) 172–180. doi:<https://doi.org/10.1016/j.neucom.2013.08.020>.
- [30] X. Yu, W. Zhang, H. Zang, H. Yang, Wind power interval forecasting based on confidence interval optimization, *Energies* 11 (12) (2018) 3336. doi:<https://doi.org/10.3390/en11123336>.
- [31] J. R. Trapero, Calculation of solar irradiation prediction intervals combining volatility and kernel density estimates, *Energy* 114 (2016) 266–274. doi:<https://doi.org/10.1016/j.energy.2016.07.167>.
- [32] B. Zhou, X. Ma, Y. Luo, D. Yang, Wind power prediction based on LSTM networks and nonparametric kernel density estimation, *IEEE Access* 7 (2019) 165279–165292. doi:<https://doi.org/10.1109/ACCESS.2019.2952555>.

- [33] S. S. Prasad, R. C. Deo, S. Salcedo-Sanz, N. J. Downs, D. Casillas-Pérez, A. V. Parisi, Enhanced joint hybrid deep neural network explainable artificial intelligence model for 1-hr ahead solar ultraviolet index prediction, *Computer Methods and Programs in Biomedicine* 241 (2023) 107737. doi:<https://doi.org/10.1016/j.cmpb.2023.107737>.
- [34] S. Salcedo-Sanz, D. Casillas-Pérez, J. Del Ser, C. Casanova-Mateo, L. Cuadra, M. Piles, G. Camps-Valls, Persistence in complex systems, *Physics Reports* 957 (2022) 1–73. doi:<https://doi.org/10.1016/j.physrep.2022.02.002>.
- [35] F. Wang, K. Li, L. Zhou, H. Ren, J. Contreras, M. Shafie-Khah, J. P. Catalão, Daily pattern prediction based classification modeling approach for day-ahead electricity price forecasting, *International Journal of Electrical Power & Energy Systems* 105 (2019) 529–540. doi:<https://doi.org/10.1016/j.ijepes.2018.08.039>.
- [36] M. S. Kiran, E. Özceylan, M. Gündüz, T. Paksoy, Swarm intelligence approaches to estimate electricity energy demand in Turkey, *Knowledge-Based Systems* 36 (2012) 93–103. doi:<https://doi.org/10.1016/j.knosys.2012.06.009>.
- [37] N. Mbuli, M. Mathonsi, M. Seitshiro, J.-H. C. Pretorius, Decomposition forecasting methods: A review of applications in power systems, *Energy Reports* 6 (2020) 298–306. doi:<https://doi.org/10.1016/j.egyr.2020.11.238>.
- [38] E. U. Haq, X. Lyu, Y. Jia, M. Hua, F. Ahmad, Forecasting household electric appliances consumption and peak demand based on hybrid machine learning approach, *Energy Reports* 6 (2020) 1099–1105. doi:<https://doi.org/10.1016/j.egyr.2020.11.071>.
- [39] Z. Cui, J. Wu, W. Lian, Y.-G. Wang, A novel deep learning framework with a COVID-19 adjustment for electricity demand forecasting, *Energy Reports* 9 (2023) 1887–1895. doi:<https://doi.org/10.1016/j.egyr.2023.01.019>.
- [40] C. Hamzaçebi, H. A. Es, R. Çakmak, Forecasting of turkey’s monthly electricity demand by seasonal artificial neural network, *Neural Computing and Applications* 31 (7) (2019) 2217–2231. doi:<https://doi.org/10.1007/s00521-017-3183-5>.
- [41] L. Tang, Y. Wu, L. Yu, A non-iterative decomposition-ensemble learning paradigm using RVFL network for crude oil price forecasting, *Applied Soft Computing* 70 (2018) 1097–1108. doi:<https://doi.org/10.1016/j.asoc.2017.02.013>.
- [42] F. Gao, H. Chi, X. Shao, Forecasting residential electricity consumption using a hybrid machine learning model with online search data, *Applied Energy* 300 (2021) 117393. doi:<https://doi.org/10.1016/j.apenergy.2021.117393>.
- [43] C. Liu, B. Sun, C. Zhang, F. Li, A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine, *Applied energy* 275 (2020) 115383. doi:<https://doi.org/10.1016/j.apenergy.2020.115383>.
- [44] Ö. F. Ertugrul, Forecasting electricity load by a novel recurrent extreme learning machines approach, *International Journal of Electrical Power & Energy Systems* 78 (2016) 429–435. doi:<https://doi.org/10.1016/j.ijepes.2015.12.006>.
- [45] M. Cai, M. Pipattanasomporn, S. Rahman, Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques, *Applied energy* 236 (2019) 1078–1088. doi:<https://doi.org/10.1016/j.apenergy.2018.12.042>.

- [46] T. G. Grandon, J. Schwenzer, T. Steens, J. Breuing, Electricity demand forecasting with hybrid statistical and machine learning algorithms: Case study of Ukraine, arXiv preprint arXiv:2304.05174 (2023). doi:<https://doi.org/10.48550/arXiv.2304.05174>.
- [47] Y. Wang, M. Liu, Z. Bao, S. Zhang, Short-term load forecasting with multi-source data using gated recurrent unit neural networks, *Energies* 11 (5) (2018) 1138. doi:<https://doi.org/10.3390/en11051138>.
- [48] P.-H. Kuo, C.-J. Huang, A high precision artificial neural networks model for short-term energy load forecasting, *Energies* 11 (1) (2018) 213. doi:<https://doi.org/10.3390/en11010213>.
- [49] S. Ghimire, T. Nguyen-Huy, M. S. AL-Musaylh, R. C. Deo, D. Casillas-Pérez, S. Salcedo-Sanz, Integrated multi-head self-attention transformer model for electricity demand prediction incorporating local climate variables, *Energy and AI* (2023) 100302doi:<https://doi.org/10.1016/j.egyai.2023.100302>.
- [50] Y. Cheng, C. Xu, D. Mashima, V. L. Thing, Y. Wu, Powerlstm: power demand forecasting using long short-term memory neural network, in: *International Conference on Advanced Data Mining and Applications*, Springer, 2017, pp. 727–740. doi:https://doi.org/10.1007/978-3-319-69179-4_51.
- [51] O. Laib, M. T. Khadir, L. Mihaylova, Toward efficient energy systems based on natural gas consumption prediction with LSTM recurrent neural networks, *Energy* 177 (2019) 530–542. doi:<https://doi.org/10.1016/j.energy.2019.04.075>.
- [52] S. Reddy, S. Akashdeep, R. Harshvardhan, S. Kamath, Stacking deep learning and machine learning models for short-term energy consumption forecasting, *Advanced Engineering Informatics* 52 (2022) 101542. doi:<https://doi.org/10.1016/j.aei.2022.101542>.
- [53] S. Hochreiter, The vanishing gradient problem during learning recurrent neural nets and problem solutions, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6 (02) (1998) 107–116. doi:<https://doi.org/10.1142/S0218488598000094>.
- [54] J. Zheng, C. Xu, Z. Zhang, X. Li, Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network, in: *2017 51st Annual conference on information sciences and systems (CISS)*, IEEE, 2017, pp. 1–6. doi:[10.1109/CISS.2017.7926112](https://doi.org/10.1109/CISS.2017.7926112).
- [55] S. Bouktif, A. Fiaz, A. Ouni, M. A. Serhani, Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches, *Energies* 11 (7) (2018) 1636. doi:<https://doi.org/10.3390/en11071636>.
- [56] K. Amarasinghe, D. L. Marino, M. Manic, Deep neural networks for energy load forecasting, in: *2017 IEEE 26th international symposium on industrial electronics (ISIE)*, IEEE, 2017, pp. 1483–1488. doi:[10.1109/ISIE.2017.8001465](https://doi.org/10.1109/ISIE.2017.8001465).
- [57] F. U. M. Ullah, A. Ullah, I. U. Haq, S. Rho, S. W. Baik, Short-term prediction of residential power energy consumption via CNN and multi-layer bi-directional LSTM networks, *IEEE Access* 8 (2019) 123369–123380. doi:[10.1109/ACCESS.2019.2963045](https://doi.org/10.1109/ACCESS.2019.2963045).
- [58] M. Afrasiabi, M. Mohammadi, M. Rastegar, L. Stankovic, S. Afrasiabi, M. Khazaei, Deep-based conditional probability density function forecasting of residential loads, *IEEE Transactions on Smart Grid* 11 (4) (2020) 3646–3657. doi:[10.1109/TSG.2020.2972513](https://doi.org/10.1109/TSG.2020.2972513).

- [59] J. A. Fill, D. E. Fishkind, The moore–penrose generalized inverse for sums of matrices, *SIAM Journal on Matrix Analysis and Applications* 21 (2) (2000) 629–635. doi:<https://doi.org/10.1137/S0895479897329692>.
- [60] S. He, J. Wu, D. Wang, X. He, Predictive modeling of groundwater nitrate pollution and evaluating its main impact factors using random forest, *Chemosphere* 290 (2022) 133388. doi:<https://doi.org/10.1016/j.chemosphere.2021.133388>.
- [61] M. A. Khan, M. I. Shah, M. F. Javed, M. I. Khan, S. Rasheed, M. El-Shorbagy, E. R. El-Zahar, M. Malik, Application of random forest for modelling of surface water salinity, *Ain Shams Engineering Journal* 13 (4) (2022) 101635. doi:<https://doi.org/10.1016/j.asej.2021.11.004>.
- [62] H. Kantz, T. Schreiber, *Nonlinear time series analysis*, Vol. 7, Cambridge university press, 2004. doi:<https://doi.org/10.1017/CB09780511755798>.
- [63] V. Vapnik, *The nature of statistical learning theory*, Springer science & business media, 1999. doi:<https://doi.org/10.1007/978-1-4757-3264-1>.
- [64] W. Guan, H. Chen, G. Li, Y. He, X. Hu, Z. Ling, Power line engineering computer investment prediction model based on SVR-PCA, in: *2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, IEEE, 2022, pp. 110–115. doi:[10.1109/EEBDA53927.2022.9744747](https://doi.org/10.1109/EEBDA53927.2022.9744747).
- [65] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural computation* 9 (8) (1997) 1735–1780. doi:[10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [66] S. Ghimire, R. C. Deo, D. Casillas-Pérez, S. Salcedo-Sanz, Improved complete ensemble empirical mode decomposition with adaptive noise deep residual model for short-term multi-step solar radiation prediction, *Renewable Energy* 190 (2022) 408–424. doi:<https://doi.org/10.1016/j.renene.2022.03.120>.
- [67] Y. Cao, L. Liu, Y. Dong, Convolutional long short-term memory two-dimensional bidirectional graph convolutional network for taxi demand prediction, *Sustainability* 15 (10) (2023) 7903. doi:<https://doi.org/10.3390/su15107903>.
- [68] H. Kim, S. Park, S. Kim, Time-series clustering and forecasting household electricity demand using smart meter data, *Energy Reports* 9 (2023) 4111–4121. doi:<https://doi.org/10.1016/j.egyr.2023.03.042>.
- [69] J. B. Marin, L. M. Marulanda, F. V. Duque, Analyzing electricity demand in colombia: A functional time series approach, *International Journal of Energy Economics and Policy* 13 (1) (2023) 75. doi:<https://doi.org/10.32479/ijeep.13728>.
- [70] M. Sha, Y. Xie, The study of different types of kernel density estimators, in: *2nd International Conference on Electronics, Network and Computer Engineering (ICENCE 2016)*, Atlantis Press, 2016, pp. 336–340. doi:[10.2991/icence-16.2016.67](https://doi.org/10.2991/icence-16.2016.67).
- [71] A. Gramacki, *Nonparametric kernel density estimation and its computational aspects*, Vol. 37, Springer, 2018. doi:<https://doi.org/10.1007/978-3-319-71688-6>.
- [72] Z. Ji, D. Niu, M. Li, W. Li, L. Sun, Y. Zhu, A three-stage framework for vertical carbon price interval forecast based on decomposition–integration method, *Applied Soft Computing* 116 (2022) 108204. doi:<https://doi.org/10.1016/j.asoc.2021.108204>.

- [73] H. He, J. Pan, N. Lu, B. Chen, R. Jiao, Short-term load probabilistic forecasting based on quantile regression convolutional neural network and Epanechnikov kernel density estimation, *Energy Reports* 6 (2020) 1550–1556. doi:<https://doi.org/10.1016/j.egy.2020.10.053>.
- [74] B. W. Silverman, *Density estimation for statistics and data analysis*, Routledge, 2018. doi:<https://doi.org/10.1201/9781315140919>.
- [75] S. Ghimire, R. C. Deo, D. Casillas-Pérez, S. Salcedo-Sanz, E. Sharma, M. Ali, Deep learning CNN-LSTM-MLP hybrid fusion model for feature optimizations and daily solar radiation prediction, *Measurement* (2022) 111759doi:<https://doi.org/10.1016/j.measurement.2022.111759>.
- [76] S. Lundberg, S. Lee, A unified approach to interpreting model predictions. arxiv 2017, arXiv preprint arXiv:1705.07874 (2022).
- [77] F. Chollet, et al., Keras (2015) (2017).
- [78] P. Goldsborough, A tour of tensorflow, arXiv preprint arXiv:1610.01178 (2016). doi:<https://doi.org/10.48550/arXiv.1610.01178>.
- [79] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631. doi:<https://doi.org/10.1145/3292500.3330701>.
- [80] S. Ghimire, R. C. Deo, D. Casillas-Pérez, S. Salcedo-Sanz, Boosting solar radiation predictions with global climate models, observational predictors and hybrid deep-machine learning algorithms, *Applied Energy* 316 (2022) 119063. doi:<https://doi.org/10.1016/j.apenergy.2022.119063>.
- [81] S. Ghimire, B. Bhandari, D. Casillas-Pérez, R. C. Deo, S. Salcedo-Sanz, Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia, *Engineering Applications of Artificial Intelligence* 112 (2022) 104860. doi:<https://doi.org/10.1016/j.engappai.2022.104860>.
- [82] S. Ghimire, T. Nguyen-Huy, R. C. Deo, D. Casillas-Perez, S. Salcedo-Sanz, 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, *Sustainable Materials and Technologies* 32 (2022) e00429. doi:<https://doi.org/10.1016/j.susmat.2022.e00429>.
- [83] C. Castillo-Botón, D. Casillas-Pérez, C. Casanova-Mateo, S. Ghimire, E. Cerro-Prada, P. Gutierrez, R. Deo, S. Salcedo-Sanz, Machine learning regression and classification methods for fog events prediction, *Atmospheric Research* 272 (2022) 106157. doi:<https://doi.org/10.1016/j.atmosres.2022.106157>.
- [84] S. Ghimire, R. C. Deo, H. Wang, M. S. Al-Musaylh, D. Casillas-Pérez, S. Salcedo-Sanz, Stacked LSTM sequence-to-sequence autoencoder with feature selection for daily solar radiation prediction: A review and new modeling results, *Energies* 15 (3) (2022) 1061. doi:<https://doi.org/10.3390/en15031061>.
- [85] N. Tagasovska, D. Lopez-Paz, Single-model uncertainties for deep learning, *Advances in Neural Information Processing Systems* 32 (2019).

- [86] H. Yamamoto, J. Kondoh, D. Kodaira, Assessing the impact of features on probabilistic modeling of photovoltaic power generation, *Energies* 15 (15) (2022) 5337. doi:<https://doi.org/10.3390/en15155337>.
- [87] T. Pearce, A. Brintrup, M. Zaki, A. Neely, High-quality prediction intervals for deep learning: A distribution-free, ensembled approach, in: *International conference on machine learning*, PMLR, 2018, pp. 4075–4084.
- [88] T. S. Salem, H. Langseth, H. Ramampiaro, Prediction intervals: Split normal mixture from quality-driven deep ensembles, in: *Conference on Uncertainty in Artificial Intelligence*, PMLR, 2020, pp. 1179–1187.
- [89] C. M. Fernández-Peruchena, J. Polo, L. Martín, L. Mazorra, Site-adaptation of modeled solar radiation data: The siteadapt procedure, *Remote Sensing* 12 (13) (2020) 2127. doi:<https://doi.org/10.3390/rs12132127>.
- [90] C. A. Gueymard, A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects, *Renewable and Sustainable Energy Reviews* 39 (2014) 1024–1034. doi:<https://doi.org/10.1016/j.rser.2014.07.117>.
- [91] O. M. Kam, S. Noël, H. Ramenah, P. Kasser, C. Tanougast, Comparative weibull distribution methods for reliable global solar irradiance assessment in France areas, *Renewable Energy* 165 (2021) 194–210. doi:<https://doi.org/10.1016/j.renene.2020.10.151>.
- [92] G. Marsaglia, W. W. Tsang, J. Wang, Evaluating Kolmogorov’s distribution, *Journal of statistical software* 8 (2003) 1–4. doi:<https://doi.org/10.18637/jss.v008.i18>.
- [93] T. Alothman, S. A. Alsaif, A. Alfakhri, A. Alfadda, Performance assessment of 25 global horizontal irradiance clear sky models in riyadh, in: *2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, IEEE, 2022, pp. 1–6.
- [94] J. Bottieau, Y. Wang, Z. De Grève, F. Vallée, J.-F. Toubreau, Interpretable transformer model for capturing regime switching effects of real-time electricity prices, *IEEE Transactions on Power Systems* (2022). doi:<https://doi.org/10.1109/TPWRS.2022.3195970>.
- [95] M.-F. Li, X.-P. Tang, W. Wu, H.-B. Liu, General models for estimating daily global solar radiation for different solar radiation zones in mainland China, *Energy conversion and management* 70 (2013) 139–148. doi:<https://doi.org/10.1016/j.enconman.2013.03.004>.
- [96] C. A. Gueymard, Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models, *Solar Energy* 86 (8) (2012) 2145–2169. doi:<https://doi.org/10.1016/j.solener.2011.11.011>.
- [97] F. Al Amer, L. Lin, Empirical assessment of prediction intervals in Cochrane meta-analyses, *European Journal of Clinical Investigation* (2021) e13524–e13524doi:<https://doi.org/10.3102/00346543231216463>.
- [98] B. Du, S. Huang, J. Guo, H. Tang, L. Wang, S. Zhou, Interval forecasting for urban water demand using PSO optimized KDE distribution and LSTM neural networks, *Applied Soft Computing* 122 (2022) 108875. doi:<https://doi.org/10.1016/j.asoc.2022.108875>.

- [99] R. L. Winkler, A decision-theoretic approach to interval estimation, *Journal of the American Statistical Association* 67 (337) (1972) 187–191. doi:<https://doi.org/10.2307/2284720>.
- [100] M. Despotovic, V. Nedic, D. Despotovic, S. Cvetanovic, Review and statistical analysis of different global solar radiation sunshine models, *Renewable and Sustainable Energy Reviews* 52 (2015) 1869–1880. doi:<https://doi.org/10.1016/j.rser.2015.08.035>.
- [101] H. Liu, X. Mi, Y. Li, Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network, *Energy Conversion and Management* 166 (2018) 120–131. doi:<https://doi.org/10.1016/j.enconman.2018.04.021>.
- [102] S. Sun, H. Qiao, Y. Wei, S. Wang, A new dynamic integrated approach for wind speed forecasting, *Applied Energy* 197 (2017) 151–162. doi:<https://doi.org/10.1016/j.apenergy.2017.04.008>.
- [103] F. X. Diebold, R. S. Mariano, Comparing predictive accuracy, *Journal of Business & Economic Statistics* 20 (1) (2002) 134–144. doi:<https://doi.org/10.1198/073500102753410444>.
- [104] M. Costantini, C. Pappalardo, Combination of forecast methods using encompassing tests: An algorithm-based procedure, Tech. rep., Reihe Ökonomie/Economics Series (2008).
- [105] M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, Y. Li, Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF reanalysis atmospheric predictors in southeast Queensland, Australia, *Renewable and Sustainable Energy Reviews* 113 (2019) 109293. doi:<https://doi.org/10.1016/j.rser.2019.109293>.
- [106] M. S. Al-Musaylh, R. C. Deo, Y. Li, Electrical energy demand forecasting model development and evaluation with maximum overlap discrete wavelet transform-online sequential extreme learning machines algorithms, *Energies* 13 (9) (2020) 2307. doi:<https://doi.org/10.3390/en13092307>.

Appendix A. Acronyms

Appendix A.1. Acronyms

Table A.16 provides the acronyms used in this paper.

Table A.16: Acronyms

Term	Acronyms
Artificial Intelligence	AI
Artificial Neural Network	ANN
Autoregressive Integrated Moving Average	ARIMA
CNN integrated with ELM	CELM
Convolutional Neural Network	CNN
Decision Trees	DT
Deep Learning	DL
Empirical Risk Minimisation	ERM
Extreme Learning Machine	ELM
Gated Recurrent Unit	GRU
Generalized AutoRegressive Conditional Heteroskedasticity	GARCH
K-Nearest Neighbors	KNN
K-Nearest Neighbors Regression	KNNR
Kernel Density Estimate	KDE
Long Short-Term Memory	LSTM
Machine Learning	ML
Mutual Information Function	MIF
Partial Autocorrelation Function	PACF
Prediction Interval	PI
Probability Density Function	PDF
Random Forest Regressor	RFR
Recurrent Neural Networks	RNN
SHapley Additive exPlanations	SHAP
Scientific Information for Land Owners	SILO
Structural Risk Minimisation	SRM
Support Vector Regression	SVR
South-East Queensland	SEQ

Appendix A.2. Hyperparameter Tuning

Figure A.24 shows the schematic of the Optuna hyperparameter optimisation method. In this study, we harnessed the power of Optuna, a versatile and efficient hyperparameter optimization framework. Optuna simplifies the intricate process of fine-tuning machine learning models, offering a notable advantage. With its user-friendly Application Programming Interface (API) and adaptive search algorithms, Optuna efficiently navigates hyperparameter search spaces, resulting in quicker convergence and more optimal model configurations. Its seamless integration with popular machine learning libraries, support for parallel and distributed optimization, and provision of visualization tools for straightforward analysis enhance its utility. Optuna’s capacity to adaptively focus on promising hyperparameter regions, complemented by its active community and ongoing development, positions it as a valuable tool for enhancing model performance while conserving time and computational resources. Furthermore, considering our use of a cutting-edge hyperparameter selection algorithm, one could argue that an ablation study may appear redundant or unwarranted. This advanced method has demonstrated its

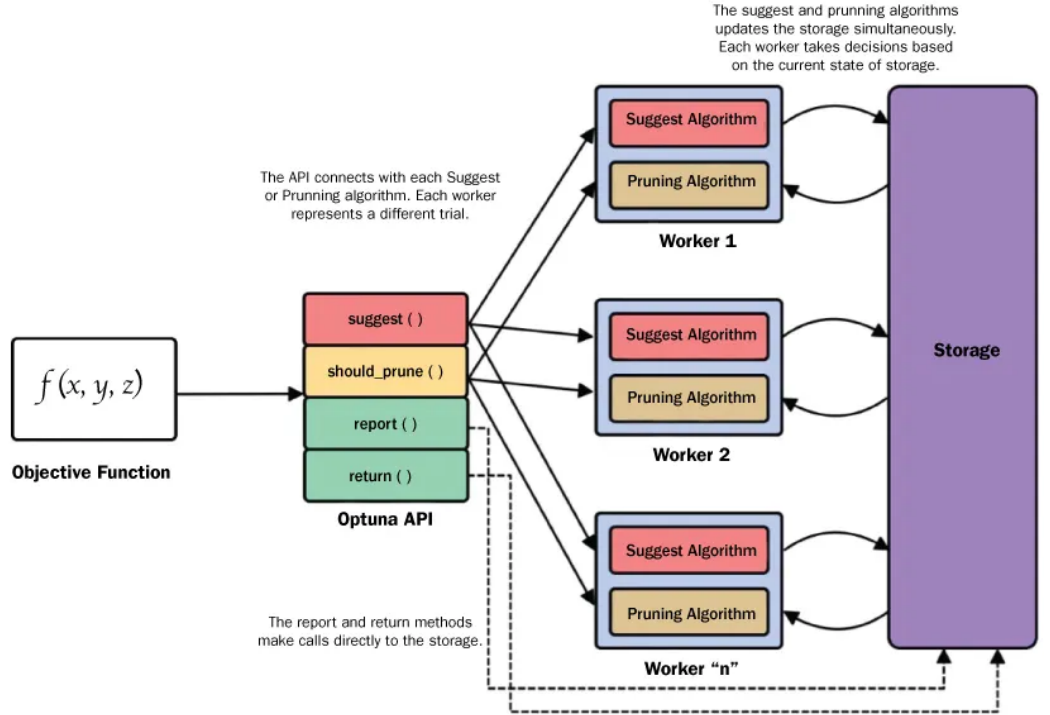


Figure A.24: Optuna Hyperparameter optimization framework.

exceptional ability to automatically pinpoint the most optimal hyperparameter configurations for our machine learning models. It conducts systematic, efficient exploration of the hyperparameter search space, yielding high-performing setups while substantially reducing the manual effort typically associated with hyperparameter tuning. This might suggest that conducting an ablation study, which traditionally involves the labor-intensive process of individually modifying or removing specific hyperparameters or components to assess their impact, offers limited benefits. Nonetheless, it is crucial to acknowledge that although advanced hyperparameter optimization techniques excel at identifying robust configurations, they may not reveal insights into the intricate workings or nuanced interactions among individual hyperparameters. Thus, considering the potential value of an ablation study for gaining a more granular understanding, we may contemplate its inclusion in future research endeavors.