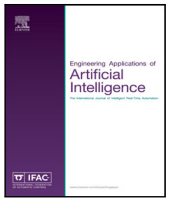
Engineering Applications of Artificial Intelligence 135 (2024) 108702

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



Research paper

Point-based and probabilistic electricity demand prediction with a Neural Facebook Prophet and Kernel Density Estimation model

Sujan Ghimire [a], Ravinesh C. Deo [a], S. Ali Pourmousavi [b], David Casillas-Pérez [c,*], Sancho Salcedo-Sanz [a,d]

[a] School of Mathematics, Physics, and Computing, University of Southern Queensland, Springfield, QLD, 4300, Australia
[b] The University of Adelaide, School of Electrical and Electronic Engineering, Australia
[c] Department of Signal Processing and Communications, Universidad Rey Juan Carlos, Fuenlabrada, 28942, Madrid, Spain
[d] Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, 28805, Madrid, Spain

ARTICLE INFO

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

MSC:
68T05
68T20

Keywords:
Time-series prediction
Deep Learning
Kernel Density Estimation
Gated Recurrent Unit
Convolutional Neural Networks
Neural Facebook Prophet

ABSTRACT

Electricity demand prediction is crucial to ensure the operational safety and cost-efficient operation of the power system. Electricity demand has predominantly been predicted deterministically, while uncertainty analysis has been usually overlooked. To address this research gap, an integrated Neural Facebook Prophet (NFBP) model and Gaussian Kernel Density Estimation (KDE) model is proposed in this paper, as a way to obtain point and interval predictions of electricity demand, quantifying this way the uncertainty in the predictions. First, historical lagged data, created by utilizing the Partial Auto-correlation Function and Mutual Information Test, is applied to train a prediction model based on NFBP, Deep Learning (DL) as well as Statistical Models. Second, the model Prediction Errors (PE) are derived from the difference between actual and predicted values. A splitting strategy based on the mean and standard deviation of PE is proposed. Finally, electricity demand prediction intervals are obtained by applying Gaussian KDE on split PE. To verify the effectiveness of the proposed model, simulation studies are carried out for three prediction horizons on freely available datasets for the Bulimba sub-station in Southeast Queensland, Australia. Compared with DL models (Long-Short Term Memory Network and Deep Neural Network), the Root Mean Square Error of the NFBP model was reduced by 6.1% and 11.3% for 0.5-hr ahead, 22.7% and 26.3% for 6-hr ahead, and 31.8% and 29.9% for daily prediction. In addition, the Prediction Interval normalized Interval width is smaller in magnitude for the proposed NFBP-KDE model compared to other DL and Statistical models.

1. Introduction

As a form of secondary energy, electricity is highly flexible and contributes significantly to socioeconomic development (Yu et al., 2014). An adequate energy supply is necessary for improved economic development and public health (Ismail et al., 2008). Since electricity cannot be stored at large scales, we need an accurate estimate of demand (G) at different time scales to maintain supply and demand balance for power system expansion planning and daily operation, policy-making, market and algorithmic designs, etc. Furthermore, erroneous G predictions could have detrimental societal and economic implications. Underestimating G could result in higher electricity prices, forced outages and power disruptions, which would lower productivity and impede economic progress. On the other hand, if G is overestimated, extra capacity might be planned or scheduled at the supply, transmission, and distribution sides, resulting in costly oversized grid infrastructure and, ultimately, higher electricity prices that would burden consumers (Charles et al., 2022). Moreover, as part of the United Nations Sustainable Development Goal 7 (SDG7) roadmap, all people should have affordable, reliable, sustainable, and modern energy services by 2030. Therefore, modelling and prediction of G is a crucial task for policymakers for both developed and developing countries. In summary, a reliable and accurate estimate of G will prevent costly mistakes and will be essential in the development of energy-efficient techno-economic planning of the energy infrastructure to meet SDG7.

Widespread interest has been shown in G prediction, and significant resources have been invested in developing new prediction tools and methodologies. The G prediction model can be classified into (a) Persistence, (b) Physical, (c) Statistical, and (d) Artificial Intelligence

* Corresponding author.
E-mail addresses: sujan.ghimire@usq.edu.au (S. Ghimire), ravinesh.deo@usq.edu.au (R.C. Deo), a.pourm@adelaide.edu.au (S. Ali Pourmousavi), david.casillas@urjc.es (D. Casillas-Pérez), sancho.salcedo@uah.es (S. Salcedo-Sanz).

https://doi.org/10.1016/j.engappai.2024.108702
Received 2 December 2022; Received in revised form 22 February 2024; Accepted 27 May 2024
Available online 10 June 2024

(AI) approaches (Afrasiabi et al., 2020; Ghimire et al., 2023a, 2024). The persistence-based approach (Salcedo-Sanz et al., 2022; Ghimire et al., 2023b) is one of the simplest methods for time-series prediction. However, these models do not perform well, in general, for medium- and long-term predictions, which is usually needed in G prediction. Physical models are based on mathematical formulas that use historical and meteorological data. These physical models are computationally inefficient (Khodayar et al., 2017). Therefore, such models are not reliable for G prediction. Statistical models are often based on autoregressive approaches, such as Auto Regressive Integrated Moving Average (ARIMA) (Fan and McDonald, 1994), Generalized AutoRegressive Conditional Heteroskedasticity (Bikcora et al., 2018), and Multiple Linear Regression (Farzana et al., 2014) techniques, and are computationally less expensive than physical models (Shi et al., 2017; Fu et al., 2023). Statistical models can be used in conjunction with other models to reduce uncertainty even though they cannot directly capture uncertainty patterns from G time-series. Furthermore, statistical models perform well for time-series with linear features. Still, they are ineffective for longer time-series prediction because they cannot learn the complex non-linear structure of G (Khodayar and Wang, 2018). Contrarily, AI-based models are based on shallow and deep structured techniques capable of learning complex and nonlinear patterns. However, shallow-based methods such as Artificial Neural Networks (ANN) (Lertpalangsunti and Chan, 1998; Mamun and Nagasaka, 2004), Wavelet Networks (WN) (Khoa et al., 2004), Fuzzy Logic's (FL) (Farahat, 2004), Expert System (ES) (Kandil et al., 2001), Support Vector Machines (SVM) (Pai and Hong, 2005), Extreme Learning Machine (ELM) (Li et al., 2015) or Random Forest (RF) (Matrenin et al., 2022) have performed poorly in feature mining. Hence, to improve the accuracy of these models, feature selection and extraction techniques are needed, which poses a challenging problem (Salcedo-Sanz et al., 2018).

Deep Learning (DL) models and Conventional Machine Learning (ML) models are the two subgroups under which the AI-based approach can be further subdivided (Almalaq and Edwards, 2017; Ghimire et al., 2023b). By learning multi-layer hierarchical features from the input data, DL can address the issues with shallow-based methods (Chen et al., 2011). The most effective architectures at the moment for the analysis of time-series data are Long Short-Term Memory (LSTM) (Masood et al., 2022; Kong et al., 2017), and Convolutional Neural Networks (CNN) (Raza et al., 2018). While G contains both spatial and temporal data, LSTM can only learn the temporal aspects of the series. The temporal characteristics of electricity demand data cannot be learned by CNN either, but it can learn the spatial features. Therefore, many hybrid models have been created to predict G precisely to overcome this issue (Alhussein et al., 2020). Kim and Cho (2019) proposed the CNN-LSTM neural network for electricity consumption prediction. They showed that the proposed model outperforms the LSTM, GRU, Echo State Networks (Liu et al., 2023), Bidirectional LSTM (BILSTM), and Attention LSTM models. Similarly, Ullah et al. (2019) developed CNN integrated with BILSTM to find that the proposed model outperforms LSTM, CNN-LSTM and BILSTM model with the smallest value of the Mean Square Error (MSE) and Root Mean Square Error (RMSE). Additionally, the CNN integrated with LSTM Autoencoder was proposed in Khan et al. (2020), CNN integrated with Gated Recurrent Unit (GRU) was proposed in Khan et al. (2021a) and Afrasiabi et al. (2020) and shown that the proposed model CNN-GRU outperforms the most of the standalone as well as hybrid DL models. In Hybrid DL models, Multivariate Empirical Mode Decomposition (EMD) (Huang et al., 2022), Prophet with LSTM (Bashir et al., 2022), Dynamic-Harmonic Regression (Permata et al., 2022), Dilated CNN with LSTM (Khan et al., 2021b), Hybrid ANN (Liao et al., 2022), and EMD-ELM (Sulaiman et al., 2022) are some recent examples of popular models developed for G prediction. Although these models overcome the drawbacks of shallow-based approaches, they are unreliable for implementation and have poor prediction accuracy (Khan et al., 2022). Moreover, the choice of one hybrid prediction model over another is frequently based on the expert's choice because there is no universal agreement on any one prediction model.

All of these aforementioned AI-based, Statistical, and Hybrid models come close to obtaining efficient *point* prediction of G with some degree of predictive errors. Contrary to conventional point G prediction, which typically only yields a single value and makes it impossible to precisely determine the probability and fluctuation range of the expected value, G probabilistic prediction can reveal a great deal of information about uncertainty. The probabilistic models offer more useful information for decision-makers because they can provide a measure of risk in prediction accuracy as a lower and an upper bound or a Prediction Interval (PI) in which G output lies with a given probability. The probabilistic prediction (Mayer et al., 2023; Fatema et al., 2023) has been the topic of a lot of research lately (Chevallier et al., 2007; Yang et al., 2018) in the field of solar radiation and wind speed prediction. One example is the work of Zhang et al. (2014) that contains an in-depth analysis of the Probabilistic prediction models.

The Lower–Upper Bound Estimation (LUBE) (Kavousi-Fard et al., 2015), Mean-Variance (MV) (Dewolf et al., 2022), Persistence (Beyaztas and Shang, 2022), Quantile Regression (QR) (Haque et al., 2014; Bremnes, 2004), Hybrid Intelligent Algorithm (HIA) (Wan et al., 2013), and Conditional Density Estimation based on Nadaraya–Watson estimator (Bessa et al., 2012) are examples of conventional probabilistic approaches. These probabilistic approaches discussed above either involve a heuristic optimization algorithm that may converge to local minima or assume a parametric distribution for PE, such as Gaussian, t-student, Beta, etc. and have certain limitations. For instance, with the MV method, Neural Networks (NNs) models are used to calculate the conditional distribution of a target with additional Gaussian noise and variable variance. It produces misleading PIs because it underestimates the data variance. Similarly, LUBE and HIA methods are computationally inefficient because of the slow learning process and overtraining (Kavousi-Fard et al., 2015). Furthermore, in several applications, QR is used to estimate various quantiles of a predictive distribution; nonetheless, it requires a laborious optimization method to minimize its specified cost function and is unsuitable for the large data (Khorramdel et al., 2018a). Thus, to overcome these limitations, nonparametric approaches, e.g., Quantile Regression Forests (Almeida et al., 2015), Gaussian Processes (Van der Meer et al., 2018), Bootstrap (Grantham et al., 2016), Kernel Density Estimation (KDE) (Yamazaki et al., 2015; Chai et al., 2016; Yamazaki et al., 2016), have been developed and been extensively used in the literature. It should be noted that nonparametric approaches hardly ever make assumptions regarding the type of distribution.

Among the nonparametric approaches, the KDE technique, which is gaining popularity, stands out among other methods because of its advantages of easy calculation and fast implementation (Trapero, 2016). In Juban et al. (2008), KDE and QR are compared for wind power prediction. The outcomes demonstrated that the predicted probability density distribution derived using the KDE technique had enhanced sharpness and reliability compared to the QR method. Similarly, in Du et al. (2022), water demand interval prediction was made using KDE, and the result shows that the PI metrics of KDE are better than that of beta and normal distribution. Moreover, Pan et al. (2021) implemented KDE-GRU for the interval prediction for solar generation, and the results demonstrated that the KDE-GRU generates high quality PI compared to other benchmark models. In Han et al. (2019), the ELM model was first used to calculate the point prediction, and KDE was fitted to PE to generate the PI, the generated PI has the lowest prediction width and high coverage probability.

In this study, the non-parametric KDE method was utilized to model the probability of G. Non-parametric KDE generates the probabilistic density function of point PE, and the cumulative G error distribution function is calculated to obtain the upper and lower bounds of interval prediction. To derive the probabilistic density function of the G error

distribution, a high-accuracy G prediction model must first be developed. Therefore, this study proposes a novel point and PIs prediction method for electricity demand based on the Neural Facebook Prophet (NFBP) and KDE. The NFBP is a hybrid prediction framework based on PyTorch and trained with standard DL methods (Shohan et al., 2022). The NFBP combines conventional FbProphet components and NN to quickly make very accurate time-series prediction (ChikkaKrishna et al., 2022). More specifically, the NFBP is first utilized to generate the point prediction. Then, the Prediction Errors (PE) of G obtained from the point prediction are fitted by the KDE method. Due to the highly-variable patterns of G, we divide G prediction error into three parts based on mean and standard deviation and use the corresponding KDE to obtain the PIs.

In context of literature review presented so far, the motivations for this study are as follows:

- Incorrectly predicting electricity demand can have detrimental societal and economic effects, especially by underestimating the G that can result in higher electricity prices, power outage or disruptions, all of which can lower the economic productivity. As a result of overestimated G predictions, a greater capacity needs to be planned at the supply, transmission, and the distribution side, resulting in more expensive infrastructure and higher prices. Therefore, this study has been motivated by a requirement that reliable, accurate estimates of G be provided to the energy companies and policymakers.
- Previous research has largely used persistence-based modelling approaches, e.g., Salcedo-Sanz et al. (2022). Such models, however, are not as robust good for medium- and long-term G predictions. Furthermore, physical models, which are mathematically governed, can be computationally inefficient whereas statistical models may not be able to capture the uncertainty patterns directly found in the G time series. Although statistical models do work well for linearly-featured datasets, they are relatively ineffective for long-term predictions of electricity demand due to their nonlinear structures. In order to overcome these deficiencies, this study has been motivated by the need to learn the complex and nonlinear patterns in models used for estimating electricity demand.
- A review of the literature suggests shallow-based methods e.g., ANN, WN, FL, ES, SVM, ELM, or RF are not effective at feature mining. Feature selection and pattern extraction techniques must be used, where possible, to improve the accuracy of these models. This research is therefore motivated by the fact that conventional models like LSTMs can only learn the temporal aspects of a data series, while the G is likely to contain both spatial and temporal information. Furthermore, conventional models like CNN can learn the spatial characteristics but not the temporal characteristics found in electricity demand data. This deficiency in deep learning models is contributing to an ongoing motivation to develop hybrid deep learning models to predict the G more precisely.
- Modelling electricity demand also requires the capability to generate confidence intervals over which the predictions are considered reliable by decision-makers. Because most previous studies have yielded only a single predicted value and cannot precisely determine the probability and the fluctuation range of estimated G data, our study is motivated by the need to create a robust model that can reveal significant information about model uncertainties. In electricity demand modelling area, probabilistic models can provide useful information as a measure of risk, as well as a lower or an upper bound or a (PI) within which the G output lies, including a certain probability of estimated demand.
- Finally, our study is motivated by a growing recognition that among several nonparametric approaches, the KDE is an easy and fast method for studying the distribution of model predictions. As compared to the traditional methods such as QR, the KDE-derived probability density distributions can have greater sharpness and reliability. To obtain the upper and the lower bounds of the interval predictions, we therefore need to adopt a non-parametric method to determine the probabilistic density function and cumulative error distributions. This certainly requires a new method of predicting the electricity demand based on points data and their PIs. Therefore one potential opportunity is to integrate the NFBP with the KDE method so that the NFBP can integrate conventional FbProphet components and neural networks to make more accurate G prediction.

The primary contributions of this paper are as follows:

- To the author's knowledge, this study is the first attempt using the hybrid NFBP model for point and interval prediction of electricity demand. The KDE hyperparameter optimized using grid search by minimizing Mean Integrated Squared Error ($MISE$) is proposed to estimate the probability density function (PDF) curves of PEs that are obtained by the difference between the real values of G and the predicted values generated from the NFBP model.
- Considering only a single PDF curve is difficult to fit all points PE computed by the NFBP model. Therefore, a new splitting strategy is proposed to divide the point PE into different levels based on the mean and standard deviation.
- The proposed NFBP model is trained in such a way that it has the capabilities of point-based prediction of electricity demand as well as the prediction of PI at three different time horizons to show a multi-step capability at 0.5-hr, 6-hr and day ahead predictions.

To pursue the above, a case study at Bulimba sub-station, Queensland, Australia is carried out to verify the performance of the proposed model for point-based and PI of electricity demand. A comparative analysis with DL models (LSTM, DNN, BILSTM), Hybrid (CBILSTM, LSTMCNN, CGRU, CBILSTM) models and a Statistical model (SARIMAX) shows the NFBP model achieves the best performance for point as well as interval prediction. The remainder of this paper is organized as follows. In Section 2, the method proposed in this paper is presented, including the Prophet model (FbProphet), NFBP model and KDE. Data available and predictive model development are presented in Section 3. Experimental results are presented and discussed in Section 4. Finally, the conclusion of this paper is given in Section 5. Besides, Appendix A shows the list of acronyms used along this work.

2. Methods

As mentioned in the literature review, various approaches have been proposed for predicting G, each with unique features and applicability to different datasets. It is difficult for generic models to correctly represent the patterns concealed in large and densely sampled datasets. DL models such as CNN, LSTM, GRU, BILSTM, DNN, statistical models such as ARIMA, Seasonal ARIMA (SARIMA), Facebook Prophet (Prophet) and hybrid models such as CNN-LSTM, LSTM-CNN, CNN-BILSTM, CNN-GRU are widely used prediction approaches. However, many real-world applications still find DL and hybrid models unsuitable because of their high computational cost, huge data requirements, and time-consuming hyperparameter tuning procedures.

Therefore, in this study, to predict G on 0.5hr interval, 6-hr interval and daily interval, a DL variation of prophet time-series models called as Neural Facebook Prophet (NFBP) is utilized and tested for one of the electricity sub-station at Southeast Queensland. This section does not address the rationale for the comparative models used in this study because the reasons are explained in more detail in other sources. The CNN-LSTM theory's specifics have been introduced in Chen et al. (2021), Gao et al. (2020) and Zhang et al. (2022) for effectively predicting a variety of datasets, including solar radiation (Ghimire

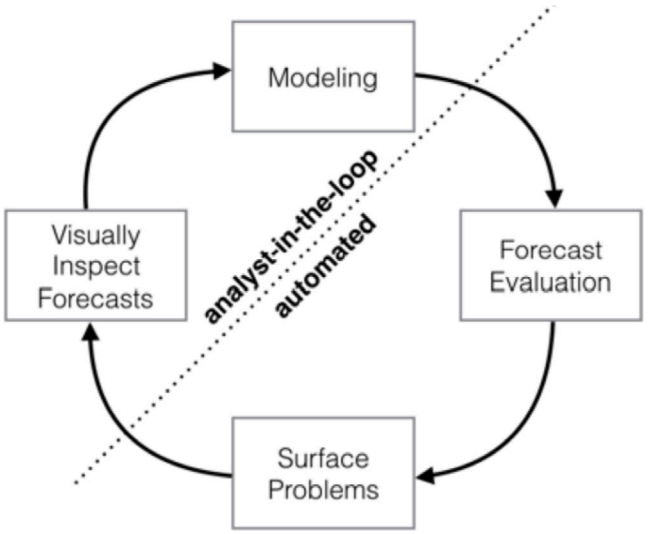


Fig. 1. Prediction process in FbProphet using Analyst-in-the-loop model.

et al., 2022d,g), photovoltaic power (Agga et al., 2022), residential energy consumption (Kim and Cho, 2019), photosynthetic photon flux density (Deo et al., 2022) and wind speed prediction (Xu and Wei, 2022). Additionally, the works in Ghimire et al. (2022a,f) and Ghimire et al. (2022b) offer the mathematical equations for the additional DL models (e.g., LSTM, CNN, DNN, and GRU) that are employed in this study. On the other hand, Ghimire (2019), Ghimire et al. (2022e,c) provides a detailed description of the theoretical framework of hybrid models.

2.1. Fundamental concepts of facebook prophet model

In 2018, Facebook pioneered the development of the Prophet model (FbProphet), which has now become a well-known time-series data processing algorithm (Taylor and Letham, 2018). In general, the FbProphet approach aims to decompose any given time-series data into four components known as the trend, seasonal, holiday, and the residual component. This method is not sensitive to missing data and the outliers (Hyndman and Athanasopoulos, 2018). The three components of trend, seasonality and holidays are given by Eq. (1):

$$y_t = g_t + s_t + h_t + \epsilon_t \tag{1}$$

where y_t is time-series, g_t is the trend term, s_t is the seasonal term, h_t is the holiday term and ϵ_t is the residual term. The g_t and s_t can be expressed by Eqs. (2) and (3), respectively:

$$g_t = \frac{C}{1 + e^{-k(t-b)}} \tag{2}$$

$$s_t = \sum_{n=1}^{N} a_n \cos\left(\frac{2\pi nt}{T}\right) + b_n \sin\left(\frac{2\pi nt}{T}\right) \tag{3}$$

where C is the upper limit of the trend, k and b are the growth rate and offset, respectively, t is the time, T is the period ($T = 364.25$ for yearly seasonality or $T = 7$ for weekly seasonality) and n is half the number of periods.

The Prophet prediction model uses the Analyst-in-the-Loop model, as seen in Fig. 1 because it is common for analysts with the domain expertise to fill the knowledge gaps of the statistical aspect of the issue. Additionally, there are several places in the model specifications where analysts can apply their outside knowledge and expertise without a thorough understanding of the underlying statistics.

2.2. Fundamental concepts of neural facebook prophet model

Neural Facebook Prophet (NFBP) model is a FbProphet successor that has not been used in the field of G prediction. NFBP considers three additional regression components in addition to the trend, seasonality, and holiday components: the auto-regression effect on historical observations, the regression effect of exogenous variables, and the regression effect of lagged observations of exogenous variables. Activation functions may learn the nonlinear behaviours in the time-series by employing neural networks to learn each component. Triebe et al. demonstrated the model's ability to provide more qualitative, comprehensible prediction components and prove NFBP efficacy (Triebe et al., 2021). Moreover, NFBP can represent local context with auto-regression and covariate regression, making it a viable tool for time-series prediction. Furthermore, NFBP is user-friendly, and model hyperparameters can be automatically tweaked for optimum performance.

The NFBP model is composed of multiple additive components, which include trend, seasonality, special events, future regression, auto-regression, and lagged regression. Each component can be adjusted and configured separately, and the final prediction is achieved by combining the individual results from each component. The modules of auto-regression and lagged regression are modelled using neural networks. The various components of the model are illustrated in below in Eq. (4):

$$y_t = g_t + s_t + h_t + \epsilon_t + \alpha_t + f_t + l_t, \tag{4}$$

where g_t, s_t, h_t, and ϵ_t components are the same in the classical Prophet algorithm (Eqs. (1)). Additionally, NFBP introduces three types of regressors: auto-regressors α_t, future regressors f_t, and lagged regressors l_t. In general, NFBP incorporates the idea of generating future predictions utilized in AR models. These models predict by combining past values or lags of a variable through a linear combination. But, NFBP uses AR-Net to predict within a specific horizon using a single model. This method includes external variables that have known future values and lagged regressors that are past values not known in the future.

2.3. Interval prediction method based on kernel density estimation

In G interval prediction, the upper and the lower bounds of G prediction points are predicted, and the PI is calculated under the specified confidence level. The concept of PI is extremely useful in electricity demand modelling as interval predictions give decision-makers additional prediction information. With accurate and acceptable PI, the risk associated with decision-making in the power system can be mitigated effectively. More broadly, the interval prediction is related to short-term and medium-term demand forecasting as it provides a measure of uncertainty around the point forecast for a given period. For example, if a short-term forecast predicts that demand for a product will be 100 units next week, the corresponding PI may be 90 to 110 units with 95% confidence, indicating a 95% probability that the actual demand will fall within the range of 90 to 110 units. However, it is important to note that interval prediction differs from short-term or medium-term demand forecasting. The latter provides a single estimate of the expected demand for a particular period, while interval prediction provides a range of values within which the actual demand is likely to occur. Interval prediction is useful in assessing the level of uncertainty around the point forecast and can aid in decision-making processes.

Hence, in this study, point PEs of G are statistically investigated and quantified based on the KDE approach. Then, using a combination of the findings from point prediction and the interval estimate results of point PE, the upper and lower bounds of G interval prediction are determined for various confidence levels. In KDE, the probability density estimate result is entirely determined by the distribution of the sample data to arrive at a more accurate estimation result since

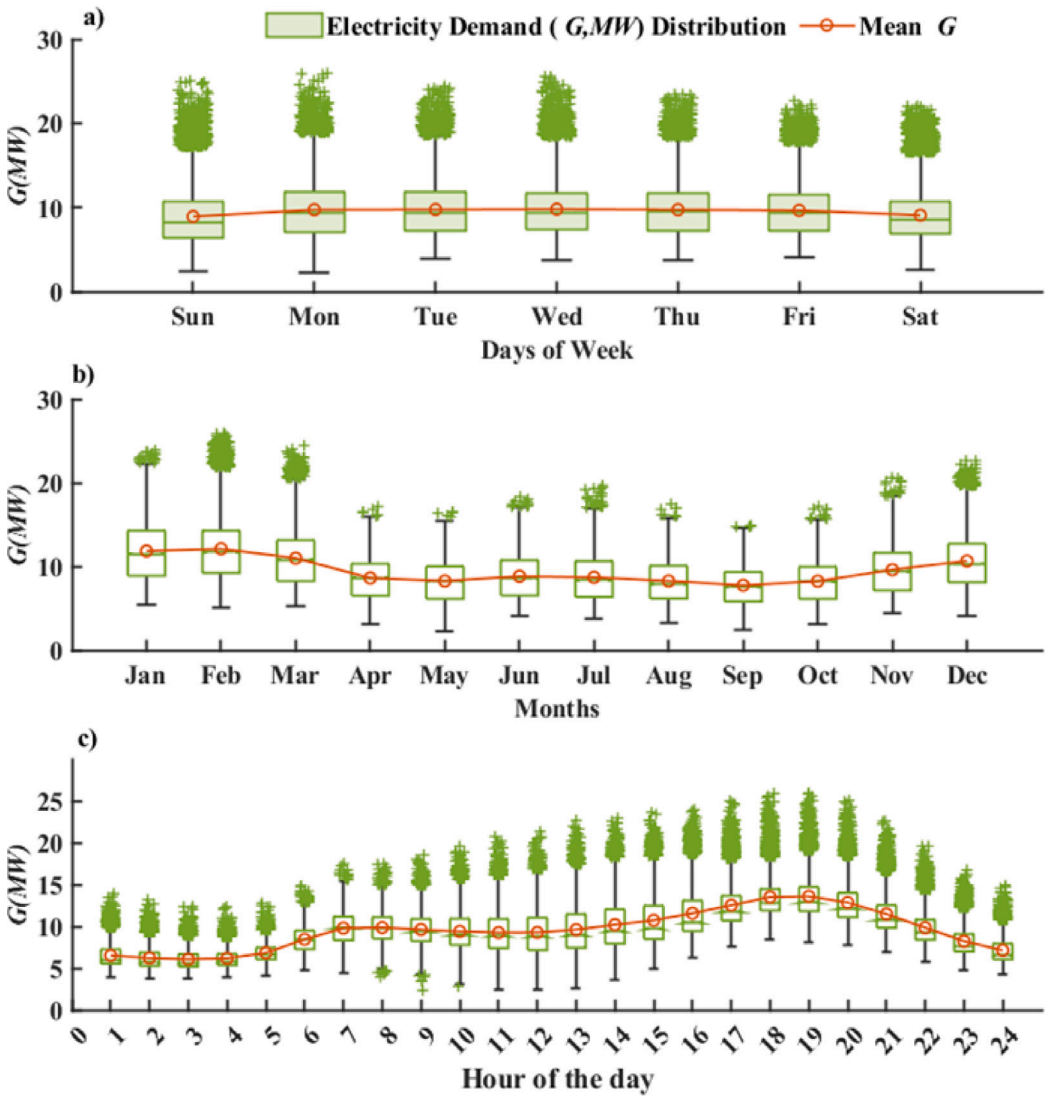


Fig. 2. Box plot of electricity demand at Bulimba Substation (a) Days of the week, (b) Month of the Year, and (c) hours of the day.

it does not need to presume the particular shape of the distribution function (Kang et al., 2018; Khorramdel et al., 2018b). Moreover, KDE offers the benefits of simplicity and is computationally efficient (Zhao et al., 2018). Assuming that $er = \left[G_{er_1}, G_{er_2}, \ldots, G_{er_n}\right]$ is the prediction errors of G, the PDF $\hat{f}_h\left(G_{er}\right)$ of samples G_{er} based on KDE method can be defined as in Eq. (5):

$$\hat{f}_h\left(G_{er}\right) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{G_{er} - G_{er_i}}{h}\right), \tag{5}$$

where N indicates the number of samples, h represents the bandwidth coefficient, $K(\cdot)$ represents the kernel function, which has various types, such as Uniform, Gaussian, Triangular, Epanechnikov, Quartic, and so on. Among them, the Gaussian kernel function is more commonly used, and is expressed in Eq. (6):

$$K\left(\eta\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}}, \tag{6}$$

where $\eta = \frac{G_{er} - G_{er_i}}{h}$.

The smoothness of the density function estimate and the trade-off between bias and variance are controlled by the bandwidth parameter h. A large value of h produces an extremely smooth (low variance), high bias density distribution. An unsmooth (high variance) yet low bias density distribution results from a modest value of h. The KDE approximation is far more influenced by the value of h than by the real kernel function. The Mean Integrated Square Error (*MISE*) may be optimized to find h:

$$MISE\left(h\right) = \mathrm{E}\left[\int \left(\hat{f}\left(x\right) - f\left(x\right)\right)^2 dx\right]. \tag{7}$$

Due to the unidentified density function f, the *MISE* formula cannot be used directly. As a result, various additional techniques have been created to ascertain the ideal value of h. Rule-of-thumb methods and cross-validation are typically used to choose the bandwidth value. Under specific presumptions on the underlying density function f and its estimate $\hat{f}$, the rule of thumb approaches get close to the ideal value of h. Silverman's rule-of-thumb, see Eq. (8):

$$h = \hat{\sigma}^{-\frac{2}{5}}, \tag{8}$$

and Scott's general rule-of-thumb, see Eq. (9):

$$h = n^{-\frac{1}{5}} \cdot \hat{\sigma}, \tag{9}$$

are popular strategies to find the optimal value of h, where $\hat{\sigma}$ is the standard deviation of the data, and n is the number of sample points. In this study, the ideal bandwidth value was computed using cross-validation. A grid search approach is used to get the value of h that

Table 1
Descriptive statistics of daily electricity demand G (MW) obtained at four substations in South-east Queensland where the integrated hybrid Neural Facebook Prophet (NFBP) with a Kernel Density Estimation (KDE) model has been implemented.

Statistical parameters	0.5-hr interval	6-hr interval	Daily interval
Median	9.11	112.35	429.45
Mean	9.52	114.19	456.76
Standard deviation	3.21	34.58	93.55
Variance	10.30	1195.84	8750.81
Maximum	25.93	274.92	832.96
Minimum	0.00	27.92	269.68
Range	53.93	247.00	563.28
Interquartile range	4.44	43.51	109.89
Skewness	0.82	0.78	1.16
Kurtosis	4.15	3.84	4.19

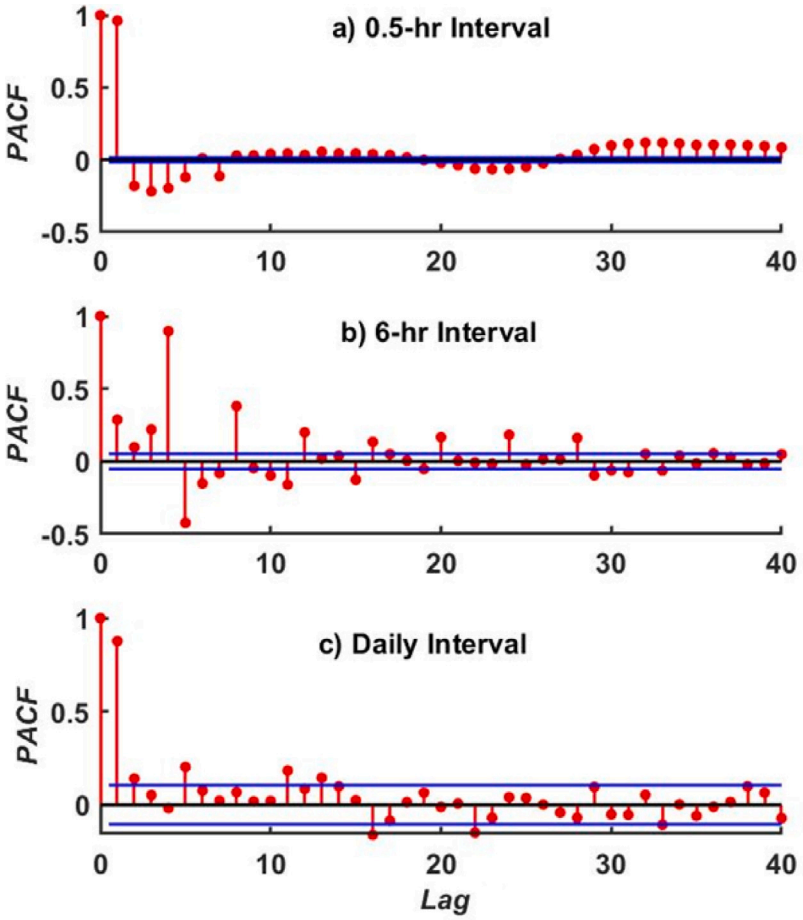


Fig. 3a. PACF of historical G time-series in the model's training phase for Bulimba substation used in this study. The blue lines denote the statistically significant boundary at the 95% confidence interval.

minimizes the sample *MISE*:

$$\mathrm{MISEE}_n(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}\left(x_i\right) - f\left(x_i\right)\right)^2. \tag{10}$$

The cumulative distribution function is calculated using an integration approach once the optimal value of h has been obtained. For a given significance level α, the probability that the predicted value of G falls within the PI range is defined as prediction interval nominal confidence (PINC), given in Eq. (11).

$$PINC = 100(1-\alpha)\% \tag{11}$$

At the significance level α, the PI $\hat{I}_i^{\alpha}$ for sample is expressed as in Eq. (12):

$$\hat{I}_i^{\alpha} = \hat{U}_i^{\alpha} - \hat{L}_i^{\alpha}, \tag{12}$$

where $\hat{U}_i^{\alpha}$ and $\hat{L}_i^{\alpha}$ refer to the upper and lower boundaries of the PI, respectively.

3. Material and methods

3.1. Electricity demand data

Electricity demand data (originally measured in MW) from a substation in Bulimba (27.45° S, 153.07° N), Queensland, was used as a case study to build the model and verify the performance of the proposed point and interval G prediction model under ultra-short-term and medium-term paradigm. Historical G from 01/01/2015 to 30/06/2021 were selected for analysis where the time interval is originally 30 min (113904 samples). Furthermore, for the 6-hourly (9493 samples) and daily interval (2374 samples), the 30-min G time-series

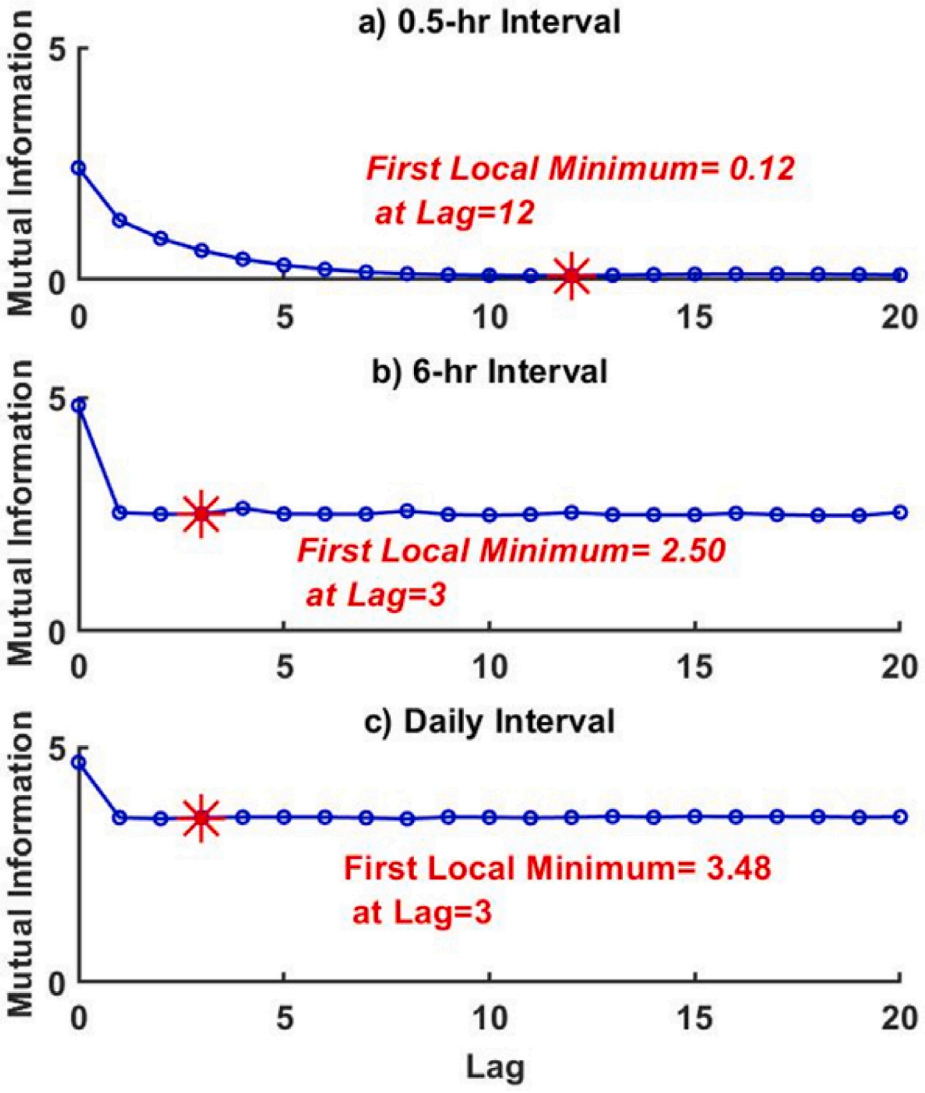


Fig. 3b. Mutual information test used to validate PACF for the input matrix of antecedent lagged electricity demand (G, MW) to design NFBP model.

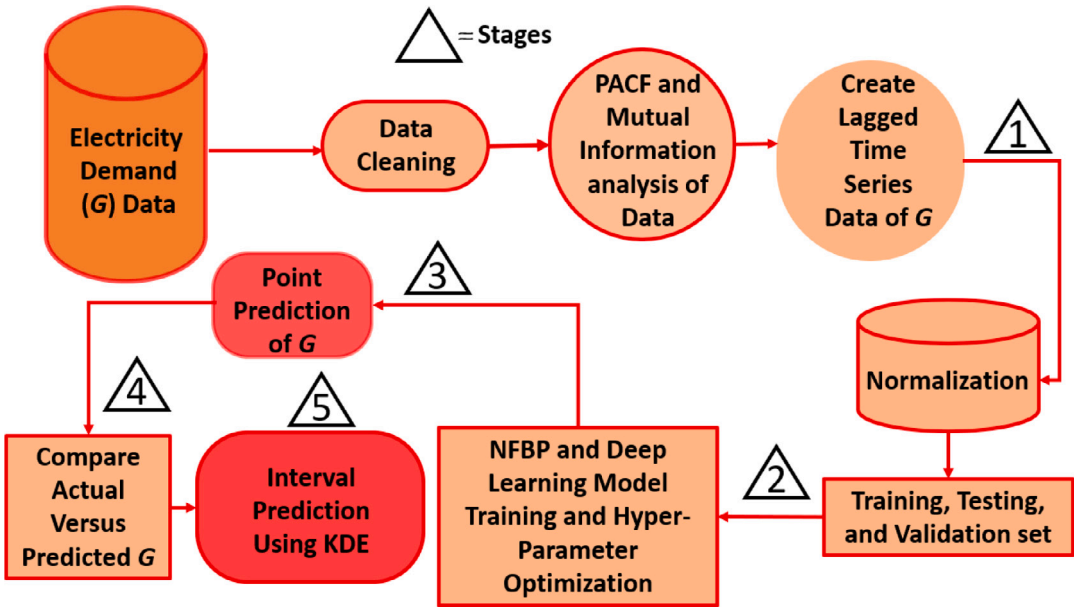


Fig. 4. Research methodology adopted in this study for prediction of *G, MW* for Bulimba Sub-station.

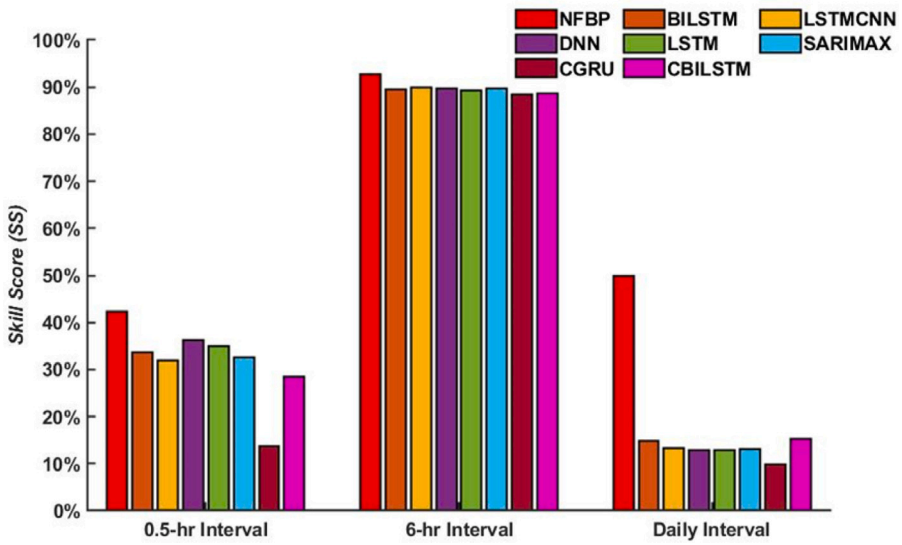


Fig. 5. Bar chart comparing the efficacy of the proposed NFBP model in terms of the tested Skill Score (SS) for 0.5-hr time, 6 h-time and daily prediction horizons.

Table 2
Architecture of the Neural Facebook Prophet (NFBP) model vs. Long short-term memory (LSTM), Bidirectional LSTM (BILSTM), CNN integrated with BILSTM (CBILSTM), LSTM integrated with CNN (LSTMCNN), CNN integrated with Gated Recurrent Unit (CGRU) and Seasonal Autoregressive Integrated Moving Average Exogenous (SARIMAX) models developed for electricity demand prediction at Bulimba Substation.

Predictive models	Model hyperparameters	Hyperparameter selection	0.5-hr interval	6-hr interval	Daily interval
NFBP	Size of the forecast horizon(n_forecast)	1			
	The number of hidden layers of the FFNNsused in the overall model (num_hidden_layers)	[1, 2, 3]	**2**	**1**	**1**
	Number of units in the hidden layers(d_hidden)	[10, 20, 30, 40, 50, 60, 80, 100, 120, 150, 200]	**80**	**100**	**50**
	Epochs (n_epochs)	[1000]			
	Learning rate (learning_rate)	[0.01, 0.001, 0.002, 0.004, 0.005, 0.006]	**0.002**	**0.001**	**0.001**
	Solver (optimizer)	['AdamW']			
LSTMCNN	LSTM layer	[1, 2, 3, 4]	**2**	**1**	**2**
	LSTM cell (units)	[40-200]	**43, 97**	**84**	**112, 89**
	CNN Layer	[1, 2, 3, 4]	**1**	**1**	**1**
	CNN Filter	[10-150]	**32**	**27**	**38**
BILSTM	BILSTM Layer	[1, 2, 3, 4]	**3**	**2**	**2**
	BILSTM cell (units)	[10-200]	**121, 77, 82**	**141, 49**	**151, 65**
DNN	DNN Layer	[1, 2, 3, 4]	**2**	**3**	**2**
	Hiddenneuron	[10-200]	**83, 75**	**77, 51, 43**	**82, 41**
CBILSTM	CNN Layer	[1, 2, 3, 4]	**2**	**2**	**2**
	CNN Filter	[10-150]	**77, 46**	**68, 48**	**83, 42**
	Bi-LSTM Layer	[1, 2, 3, 4]	**2**	**1**	**2**
	Bi-LSTM cell(Units)	[10-100]	**75, 46**	**84**	**94, 44**
CGRU	CNN Layer	[1, 2, 3, 4]	**2**	**2**	**1**
	CNN Filter	[10-170]	**91, 43**	**123, 84**	**74**
	GRU Layer	[1, 2, 3, 4]	**3**	**2**	**2**
	GRU Cell	[10-180]	**52, 21**	**67, 43**	**113, 22**
LSTM	LSTM Layer	[1, 2, 3, 4]	**4**	**3**	**3**
	LSTM cell	[10-170]	**109, 45, 54, 43**	**143, 85, 41**	**95, 62, 87**
SARIMAX	Order (p, d, q)	range (0, 6)	(1, 1, 1)	(1, 1, 2)	(1, 1, 2)
	Seasonal (P, D, Q)	range (0, 4)	(1, 0, 1, 12)	(1, 1, 2, 12)	(1, 1, 2, 12)

data was converted by calculating the sum of every 12 values and 24 values starting from 01/07/2011 at 12:00 midnight until 30/06/2021 at midnight.

As a result, four points, constructed at 06:00 a.m. (from 12:00 Midnight. to 6:00 a.m.), 12:00 Noon (from 6:00 a.m. to 12:00 Noon), 18:00 p.m. (from 12:00 Noon to 06:00 p.m.), and 12:00 Midnight (from 06:00 p.m. to 12:00 Midnight), were obtained for each day. These four points were summed up to get the daily G values. For interpretation of the model performance in terms of the electricity demand in MWh, the respective timescale over which results are presented should be applied to the respective G values in MW.

Fig. 2 shows the original G data at the Bulimba Substation. As expected, the magnitude of G during the week is relatively larger than at the weekend and its distribution is also very different with 4:00 p.m. to 8:00 p.m. as the peak period, and excessive electricity consumption observed during the summer period (December, January and February).

Table 1 provides descriptive statistics of G (MW) for 0.5-hr, 6-hr and daily prediction horizons. It is also imperative to note that a data imputation stage is not required for this study as there are no missing data for this sub-station.

3.2. Predictive model development

3.2.1. Preprocessing and data segregation

To prevent prediction models from being impacted by the various ranges, data standardization is required before the prediction process. The G values are normalized through the z-score normalization method described by Eq. (13):

$$G' = \frac{G - \mu}{\sigma}, \tag{13}$$

where G' is the normalized value of G, μ represents the mean and σ corresponds to standard deviation of G. Lag selection is a crucial step before computing time-series after normalization. To do this, the partial autocorrelation function (PACF) is used to acquire time-lagged statistics from the time-series data of G. This time-lagged information was used to evaluate the temporal dependencies between G for a current time and G_t values at a certain point in an earlier period (i.e., a time lag of G_{t-1}, G_{t-2}, G_{t-3}, G_{t-4}, G_{t-5}, etc.). This temporal reliance in the G time-series was evaluated for 40 lags (i.e., from G_{t-1} to G_{t-40}) as depicted in Fig. 3a (the blue lines indicate the 95% confidence band).

The information derived from PACF, however, illustrates collinearity between the most recent and previous G. In a nonlinear process, the approach may thus fail to extract effective inputs (Mehr et al., 2018).

Table 3a
Deterministic performance measure, denoted as Class A metrics.

Deterministic performance measure (Class A)	Definition	
Correlation Coefficient	$r = \frac{\sum_{i=1}^{n}(G^m - \langle G^m\rangle)(G^p - \langle G^p\rangle)}{\sqrt{\sum_{i=1}^{n}(G^m - \langle G^m\rangle)^2}\sqrt{\sum_{i=1}^{n}(G^p - \langle G^p\rangle)^2}}$	(18)
Root Mean Square Error (MW)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(G^p - G^m)^2}$	(19)
Mean Absolute Error (MW)	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert G^p - G^m\rvert$	(20)
Relative Root Mean Square percentage Error (%)	$RRMSPE = \frac{RMSE}{G^m} \times 100\%$	(21)
Relative Mean Absolute Percentage Error (%)	$RMAPE = \frac{MAE}{G^m} \times 100\%$	(22)
Uncertainty at 95%	$U_{95} = 1.96(SD^2 - RMSE^2)^{0.5}$	(23)
t-statistic	$TS = \sqrt{\frac{(n-1)\times MBE^2}{RMSE^2 - MBE^2}}$	(24)
Standard deviation of the Relative Error	$STDRE = \left(\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{G^p - G^m}{G^m}\right)^2\right)^{1/2}$	(25)
Explained Variance Score	$E_{var} = 1 - \frac{\mathrm{Var}(G^m - G^p)}{\mathrm{Var}(G^m)}$	(26)
Absolute Percentage Bias (%)	$APB = \frac{\sum_{i=1}^{n}(G^m - G^p) * 100}{\sum_{i=1}^{n} G^m}$	(27)
Skill Score	$SS = 1 - \frac{RMSE(p,x)}{RMSE(pr,x)}$	(28)
Mean Bias Error (MW)	$MBE = (100/\langle G^m\rangle)\sum_{i=1}^{i=N}\left(G^p{}_i - G^m{}_i\right)$	(29)

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.*

Table 3b
Deterministic performance measure, denoted as Class B metrics.

Deterministic performance measure (Class B)	Definition	
Willmot's Index	$E_{WI} = 1 - \frac{\sum_{i=n}^{n}(G^m - G^p)^2}{\sum_{i=n}^{n}(\lvert G^p - \langle G^m\rangle\rvert + \lvert G^m - \langle G^m\rangle\rvert)^2}$	(30)
Nash–Sutcliffe Equation	$E_{NS} = 1 - \frac{\sum_{i=1}^{n}(G^m - G^p)^2}{\sum_{i=1}^{n}(G^m - \langle G^m\rangle)^2}$	(31)
Legates and McCabe's Index	$E_{LM} = 1 - \frac{\sum_{i=1}^{n}\lvert G^m - G^p\rvert}{\sum_{i=1}^{n}\lvert G^m - \langle G^m\rangle\rvert}$	(32)
Theil's Inequality Coefficient	$TIC = \frac{\sqrt{\frac{1}{n}\times\sum_{i=1}^{n}(G^p - G^m)^2}}{\left(\sqrt{\frac{1}{n}\times\sum_{i=1}^{n}(G^m)^2} + \sqrt{\frac{1}{n}\times\sum_{i=1}^{n}(G^p)^2}\right)}$	(33)
Kling–Gupta Efficiency	$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\langle G^p\rangle}{\langle G^m\rangle} - 1\right)^2 + \left(\frac{CV_p}{CV_m}\right)^2}$	(34)

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 *size (number) of predictions,* CV *Coefficient of Variation.*

The Mutual Information (MI), regarded as a non-linear generalization of the PACF, was also considered in this study to address this problem. The MI criteria, see Eq. (14) is often used to identify time-delayed coordinates that are as independent of one another as is practical.

$$G_{yy}(\tau) = \sum_{i=1}^{N} pr(y(i), y(i+\tau)) \log \frac{pr(y(i), y(i+\tau))}{pr(y(i)) \cdot pr(y(i+\tau))} \tag{14}$$

where N the sample size, pr is a measure of probability, $pr(x, y)$ and $pr(x)$ is the joint and marginal probability. $G_{yy}(\tau)$ represents the amount of information that one can know about G time-series that are separated by the time lag τ. The optimal τ is where $G_{yy}(\tau)$ reaches its first local minimum. In Fig. 3b, the calculated mutual information function for the G time-series is displayed. It is clear that with increasing delays, mutual function rapidly declines. However, the mutual information decreases

Table 3c
Deterministic performance measure, denoted as Class C metrics.

Deterministic performance measure (Class C)	Definition	
Kolmogorov–Smirnov Index	$KSI = \frac{100}{A_c}\int_{X_{min}}^{X_{max}} D_n dx$	(35)
Critical Limit Overestimation Index	$OVER = \frac{100}{A_c}\int_{X_0}^{X_1} \max(D_n - D_c, 0) dx$	(36)
	$where \quad A_c = D_c(X_{max} X_{min})$	(37)
	$where \quad D_c = \Phi(N)/N^{1/2}$	(38)
Combined Performance index	$CPI = \frac{KSI + OVER + 2RMSE}{4}$	(39)

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 (Marsaglia et al., 2003; Alothman et al., 2022).

Table 3d
Probabilistic performance measure, denoted as Class D metrics.

Deterministic performance measure (Class D)	Definition	
Prediction Interval Coverage Probability	$PICP = \frac{1}{N}\sum_{i=1}^{N} c_i$	(45)
	$where\ c_i = \begin{cases} 1 & \text{if } y_i \in (U(G_i), L(G_i)) \\ 0 & \text{otherwise} \end{cases}$	(46)
Mean Prediction Interval Width	$MPIW = \frac{1}{N}\sum_{i=1}^{N} (U(G_i) - L(G_i))$	(47)
F Value	$F = \frac{PICP \times 2 \times \frac{1}{MPIW}}{PICP + \frac{1}{MPIW}}$	(48)
Average Relative Interval Length	$ARIL = \frac{1}{N}\sum_{i=1}^{N} \frac{(U(G_i) - L(G_i))}{G^m{}_{,i}}$	(49)
Winkler Score	$WS = \begin{cases} \Delta_i & L(G_i) \le y_i \le U(G_i) \\ \Delta_i + 2(L(G_i) - y_i)/\alpha & y_i < L(G_i) \\ \Delta_i + 2(y_i - U(G_i))/\alpha & y_t > U(G_i) \end{cases}$	(50)
	$where \quad \Delta_i = U(G_i) - L(G_i)$	(51)
Normalized Mean Prediction Interval Width	$PINAW = \frac{1}{N * R}\left(\sum_{i=1}^{N} (U(G_i) - L(G_i))\right)$	(52)
Continuous Rank Probability Score (MW)	$CRPS = \frac{1}{N}\sum_{i=1}^{N} crps(F_i, y_i)$	(53)
	$where \quad crps(F, y) = \int_{-\infty}^{\infty} (F(t) - 1(t - y))^2 dy$	(54)

Note: N *denotes the number of test samples.* y_i is the ith *observation,* $L(G_i)$ *and* $U(G_i)$ *represent lower bound and upper bound of the* i^{th} G *Prediction Interval respectively.* G^m *is the observed value of* G. R *is the Range* (Bottieau et al., 2022). *In* $CRPS$ *metrics,* $1(\cdot)$ *is the Heaviside function, it takes the value of 1 when* $t > y$ *and equals 0 otherwise.*

slowly after 12 lags for 0.5-hr intervals, three lags for 6-hr intervals and daily intervals. Therefore, according to Eq. (14), the optimum delay time (lag) is selected as 12, 3 and 3 for 0.5-hr interval, 6-hr interval and daily interval, respectively. Based on the MI criteria in Fig. 3b, the most effective inputs for G prediction can be mathematically expressed as Eq. (14) for 0.5-hr interval ($G_{0.5\text{-hr}}$), Eq. (15) for 6-hr ($G_{6\text{-hr}}$)

Table 4
The testing performance of the NFBP model vs. LSTMCNN, BILSTM, DNN, CBILSTM, CNN and SARIMAX models as measured by Correlation Coefficient (r), Root Mean Square Error ($RMSE, MW$), Mean Absolute Error (MAE, MW) by Standard Deviation of Relative Error ($STDRE$), Explained Variance (E_{var}), Uncertainty at 95% ($U95$) and t-statistics (TS).

Time interval	Predictive model	Model performance metrics						
		r	$RMSE$	MAE	$STDRE$	E_{var}	U_{95}	TS
0.5-hr	**NFBP**	**0.989**	**0.664**	**0.340**	**5.983**	**0.957**	**1.840**	**0.002**
	CBLSTM	0.987	0.712	0.370	6.694	0.950	1.973	1.947
	LSTMCNN	0.987	0.721	0.376	6.751	0.949	1.998	1.727
	CGRU	0.988	0.698	0.362	6.504	0.952	1.934	1.474
	LSTM	0.988	0.704	0.366	6.784	0.951	1.952	4.813
	BILSTM	0.987	0.717	0.371	6.805	0.950	1.989	1.550
	SARIMAX	0.984	0.811	0.389	9.734	0.936	2.249	0.164
	DNN	0.986	0.739	0.385	8.375	0.947	2.046	6.839
6-hr	**NFBP**	**0.974**	**10.276**	**7.175**	**7.792**	**0.902**	**28.487**	**0.017**
	CBLSTM	0.962	12.459	8.480	9.920	0.858	34.421	4.546
	LSTMCNN	0.964	12.157	8.351	9.509	0.863	33.688	1.681
	CGRU	0.963	12.318	8.517	9.401	0.859	34.136	1.496
	LSTM	0.961	12.610	8.704	9.772	0.852	34.952	1.129
	BILSTM	0.963	12.284	8.536	9.354	0.860	34.053	0.623
	SARIMAX	0.958	13.104	9.813	9.141	0.840	36.330	0.060
	DNN	0.959	12.982	8.980	10.656	0.846	35.848	4.833
Daily	**NFBP**	**0.968**	**29.976**	**22.001**	**4.246**	**0.879**	**83.145**	**0.177**
	CBLSTM	0.945	39.054	29.519	5.406	0.796	108.310	0.421
	LSTMCNN	0.944	39.416	29.520	5.556	0.792	109.310	0.566
	CGRU	0.943	39.496	29.629	5.612	0.791	109.530	0.488
	LSTM	0.943	39.524	29.650	5.583	0.791	109.600	0.622
	BILSTM	0.944	39.441	29.635	5.584	0.792	109.370	0.607
	SARIMAX	0.942	40.169	30.271	5.697	0.784	111.420	−0.190
	DNN	0.945	38.943	29.336	5.557	0.798	107.780	1.818

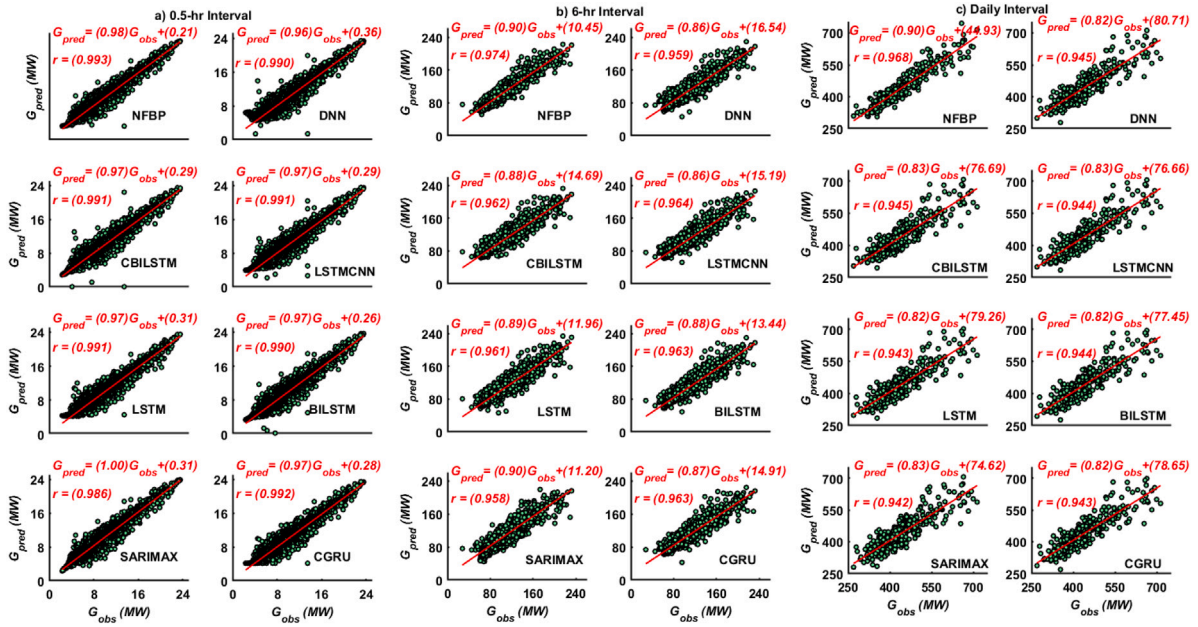


Fig. 6. Scatter plots of the actual electricity demand (G_{obs}, MW) and the predicted electricity demand (G_{pred}, MW) at Bulimba substation for (a) 0.5-hr, (b) 6h-time and (c) daily prediction horizons. Note: Red line shows least-square regression $y = mx + c$ where y is the G_{pred}, x is the G_{obs} and r is the correlation coefficient. The name of each model is provided in Table 2.

and Eq. (16) for daily interval (G_{daily}).

$$G_{0.5\text{-hr}} = f\left(G_{t-1}, G_{t-2}, G_{t-3}, G_{t-4}, G_{t-5}, G_{t-6}, G_{t-7}, G_{t-8}, G_{t-9}, G_{t-10}, G_{t-11}, G_{t-12}\right) \quad (15)$$

$$G_{6\text{-hr}} = f\left(G_{t-1}, G_{t-2}, G_{t-3}\right) \quad (16)$$

$$G_{daily} = f\left(G_{t-1}, G_{t-2}, G_{t-3}\right) \quad (17)$$

To train and test the proposed models, the normalized time-series of lagged G values are split into training and test sets. The dataset from 01/07/2015 to 30/06/2020 is used for training, while dataset from 01/07/2020 to 30/06/2021 is utilized for testing. The training

Table 5
The performance of the NFBP model vs. LSTMCNN, BILSTM, DNN, CBILSTM, CNN and SARIMAX models using the Willmott's Index (E_{WI}), Nash–SutcliffeCoefficient (E_{NS}), the Legates & McCabe's (E_{LM}) Index of Agreement and Theil's inequality coefficient (TIC). Note that the best model is boldfaced (blue).

Time interval	Predictive model	Model performance metrics			
		E_{WI}	E_{NS}	E_{LM}	TIC
0.5-hr	**NFBP**	**0.971**	**0.957**	**0.868**	**0.035**
	CBLSTM	0.967	0.950	0.856	0.037
	LSTMCNN	0.966	0.949	0.854	0.038
	CGRU	0.968	0.952	0.859	0.036
	LSTM	0.967	0.951	0.858	0.037
	BILSTM	0.967	0.950	0.856	0.037
	SARIMAX	0.959	0.936	0.849	0.042
	DNN	0.964	0.947	0.850	0.038
6-hr	**NFBP**	**0.937**	**0.902**	**0.727**	**0.045**
	CBLSTM	0.903	0.856	0.677	0.055
	LSTMCNN	0.908	0.863	0.682	0.054
	CGRU	0.906	0.859	0.676	0.054
	LSTM	0.905	0.852	0.669	0.056
	BILSTM	0.909	0.860	0.675	0.054
	SARIMAX	0.903	0.840	0.627	0.058
	DNN	0.893	0.844	0.658	0.057
Daily	**NFBP**	**0.912**	**0.879**	**0.675**	**0.034**
	CBLSTM	0.848	0.795	0.564	0.044
	LSTMCNN	0.846	0.792	0.564	0.045
	CGRU	0.846	0.791	0.563	0.045
	LSTM	0.845	0.791	0.562	0.045
	BILSTM	0.847	0.791	0.562	0.045
	SARIMAX	0.845	0.784	0.553	0.045
	DNN	0.842	0.797	0.567	0.044

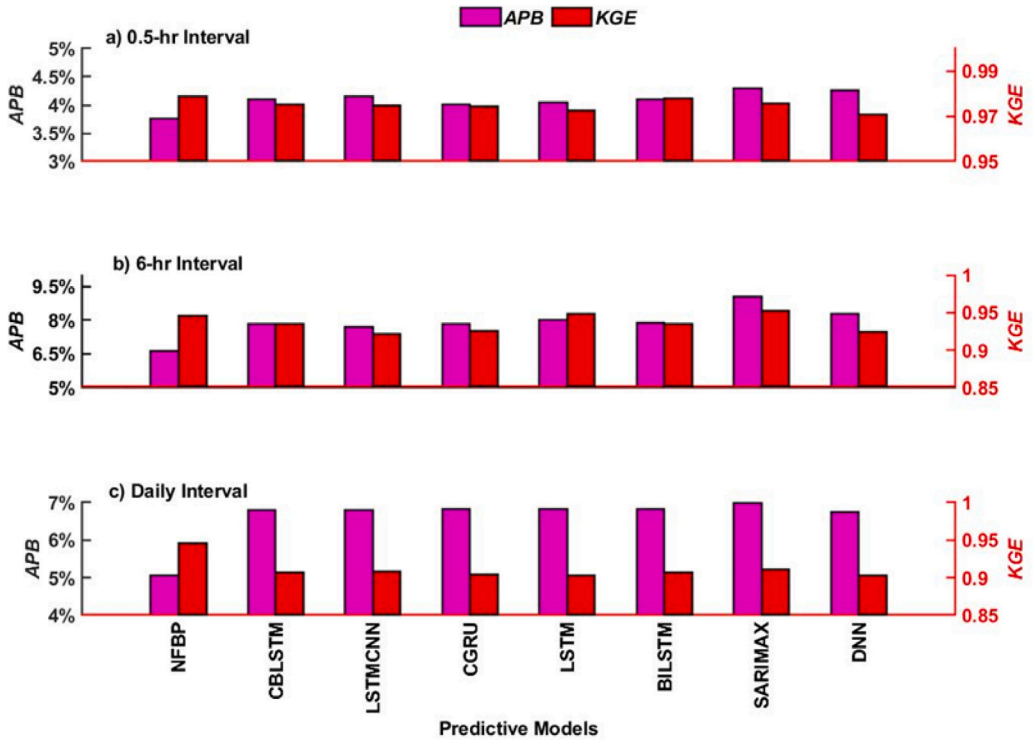


Fig. 7. Bar chart comparing the efficacy of the proposed NFBP model in terms of the tested Absolute Percentage Bias ($APB, \%$) and Kling–Gupta Efficiency (KGE). Note: The name of each model are provided in Table 2.

set is split into training (80%) and validation (20%). The validation set is used to find the best setting for the model hyperparameters, while the training set is used to fit the model parameters by lowering the prediction loss.

3.2.2. NFBP and benchmark models development

The five-stage research methodology used in this study is depicted in Fig. 4. Data preprocessing and the data division into training, validation, and testing sets, as described in earlier sections, make up the first two stages. In the third stage, SARIMAX, LSTM LSTMCNN, BILSTM, DNN, CGRU, CBILSTM and NFBP models are developed. All models development and hyperparameter tuning are implemented in Python 3.8 with numpy, pandas, statsmodels, fbProphet, and Keras libraries (Oliphant, 2006; Dirckx, 1980; Triebe et al., 2021). The computational experiment is an MS Windows machine with an Intel(R) Core(TM) i7-1135G7 @ 2.40 GHz and 32.0 GB. This study utilizes the "pmdarima" library to choose the ideal SARIMAX model parameters, i.e., p, d, q, P, D, Q, and s for point prediction of G as per the smallest Bayesian Information Criteria scores (Chaturvedi et al., 2022).

The other DL benchmark models (LSTM, LSTMCNN, BILSTM, DNN, CGRU, CBILSTM) as well as the proposed model's (NFBP) optimal parameters, which were determined using Hyperopt (Komer et al., 2019), are summarized in Table 2. The Huber loss is utilized as the loss function for creating the NFBP model, in contrast to the MSE, which is widely used in other ML models. In general, Huber Loss is less susceptible to outliers and assists in preventing exploding gradients (Girshick, 2015). Similarly, the Adaptive Moment Estimation with weight decay (*AdamW*) (Loshchilov and Hutter, 2017) is used as the optimization function. *AdamW* was initialized with the configured learning rate (Table 2), the exponential decay rate for the first moment β_1 and second moment β_2 as 0.9 and 0.999 respectively, the weight decay (δ) and epsilon (ε) are set to 1×10^{-4} and 1×10^{-8}, respectively. Unlike *Adam*, which is widely used in ML and DL models, the *AdamW* optimization function used in NFBP has better generalization performance. Additionally, the *Batch Size* of 25 was chosen for NFBP model training.

Similarly, the DL (LSTM and DNN) and Hybrid models (LSTMCNN, BILSTM, LSTMCNN, CGRU and CBILSTM) for point prediction of G were trained using the Adaptive Moment Estimation (*Adam*) optimizer, with Batch Size of 25 and loss function set to the RMSE. This study has adopted the Rectified Linear Unit (*ReLU*) as an activation function for each layer of the DL benchmark models (LSTM LSTMCNN, BILSTM, DNN, CGRU, and CBILSTM). Furthermore, the dropout layer (*DOL*) with dropout rate of 0.2 was used after each LSTM, BILSTM, GRU and CNN layers to prevent overfitting. The *DOL* is a regularization technique that modifies the network's configuration and enhances generalization while training by randomly removing some neurons. Additionally, Early Stopping (*es*) and *ReduceLROnPlateau* was employed in this study to prevent overfitting. In *es*, during model training, if an RMSE on validation data stops improving, the model's training will stop (Zhao et al., 2019). Similarly, *ReduceLROnPlateau* is employed during the training phase of our model to optimize weights more precisely, which enhances the performance of our model.

In our study, the *ReduceLROnPlateau* was used to monitor the validation loss and reduce the learning rate of the *Adam* Optimizer linearly by a factor of 0.1 (empirically selected) if the validation loss does not decrease for 20 epochs (empirically selected) (Arshad et al., 2021). In the fourth stage of the model development, SARIMAX, LSTM LSTMCNN, BILSTM, DNN, CGRU, CBILSTM and NFBP models' predictive abilities are assessed at the model evaluation stage by assessing their test set predictions with an actual test dataset. The prediction accuracy of the proposed model (NFBP), as well as benchmark models, are compared using deterministic metrics (Table 3a, Table 3b and Table 3c).

Using KDE, the PIs for point PEs at a 95% confidence level were estimated in the fifth stage of the model development. The general procedure for implementing the KDE approach for interval prediction of G can be summarized as follows.

Step 1: For each model, the total point predicted G on the testing dataset is separated into three levels based on the average and standard deviation of point prediction results to discern the distribution of prediction errors. The dividing details are as follows;

$$Y_{lev1} < \mu - \sigma \tag{40}$$

$$\mu - \sigma \le Y_{lev2} \le \mu + \sigma \tag{41}$$

$$Y_{lev3} \ge \mu + \sigma \tag{42}$$

where μ and σ represent the average and standard deviation of the entire point prediction results, respectively. Y_{lev1}, Y_{lev2}, and Y_{lev3} are the divided three levels of predicted G on testing set.

Step 2: PEs between the actual $[x_1, x_2, x_3, \ldots, x_n]$ (test dataset) and predicted Y_i at each level was calculated. This PE can be expressed as $PE = [e_1, e_2, e_3, \ldots, e_n]$, where $e_i = y_i - x_i$, $i = [1, 2, 3, \ldots, n]$;

Step 3: For each level, PE values are normalized using $PE' = \frac{PE-\mu}{\sigma}$, where PE', μ and σ are normalized PE, mean and standard deviation of the actual PE, respectively.

Step 4: A nonparametric KDE model is established for the normalized PE (PE') to obtain the cumulative PE distribution function as per Eq. (4). In this step, the grid search approach is used to get the KDE bandwidth that minimizes the sample MISE (Eq. (10)).

Step 5: The optimum interval $[PE'_{lo}, PE'_{up}]$ of the PE is obtained at 95% confidence level using Eqs. (12).

Step 6: After that, the PIs of electricity demand at 95% confidence levels are calculated by transforming the obtained PE intervals reversely. The inverse transforming formula of obtained PE intervals can be expressed as:

$$Y_{up} = Y + \left[PE'_{up} * \sigma + \mu\right] * Y \tag{43}$$

$$Y_{lo} = Y + \left[PE'_{lo} * \sigma + \mu\right] * Y \tag{44}$$

where PE'_{lo} and PE'_{up} refer to the obtained lower bounds and upper bounds of the PE, respectively. μ and σ are the mean and standard deviation of the PE, respectively. G refers to the point predictions of electricity demand using the proposed model (NFBP) and benchmark models. Y_{up} and Y_{lo} are the predictive bounds of the electricity demand G.

Step 7: Finally, the interval prediction accuracy of the proposed model (NFBP) is compared with benchmark models using probabilistic metrics (Table 3d).

Furthermore, given that each of the aforementioned statistical indicators (Table 3a, Table 3b and Table 3c) has benefits and drawbacks of its own, it could be challenging to assess the accuracy of various models using a single statistical indicator. Therefore, in this study, the overall model performance was ranked using the Global Performance Indicator (GPI) (Despotovic et al., 2015). GPI was calculated using six metrics.

$$\mathrm{GPI}_i = \sum_{j=1}^{6} \alpha_j (g_j - y_{ij}) \tag{55}$$

where α_j is the median of scaled values of statistical indicator, $j = 1$ for $RMSE$, MAE, $MAPE$, $RRMSE$ and MBE ($j = 1, 2, 3, 4, 5$), -1 for r; g_j is the scaled value of the statistical indicator j for model i. A high value of GPI indicates the better accuracy of the model.

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

$$DS = \frac{1}{n} \sum_{t=2}^{n} d_t \times 100\% \tag{56}$$

where,

$$d_t = \begin{cases} 1 & \text{if } (G_t^m - G_{t-1}^m)(G_t^p - G_{t-1}^m) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{57}$$

Table 6
The comparison of the accuracy of the NFBP model vs. LSTMCNN, BILSTM, DNN, CBILSTM, CNN and SARIMAX models in terms of the relative errors ($RRMSPE, \%$) and ($RMAPE, \%$) computed within the test sites. Note that the best model is boldfaced (blue).

Time interval	Predictive model	Model performance metrics	
		RRMSPE	*RMAPE*
0.5-hr	**NFBP**	**7.34%**	**3.93%**
	CBLSTM	7.87%	4.21%
	LSTMCNN	7.97%	4.31%
	CGRU	7.71%	4.15%
	LSTM	7.79%	4.25%
	BILSTM	7.93%	4.24%
	SARIMAX	8.97%	4.58%
	DNN	8.17%	4.61%
6-hr	**NFBP**	**9.47%**	**6.71%**
	CBLSTM	11.48%	7.97%
	LSTMCNN	11.20%	7.80%
	CGRU	11.35%	7.98%
	LSTM	11.62%	8.15%
	BILSTM	11.32%	7.93%
	SARIMAX	12.08%	9.37%
	DNN	11.96%	8.60%
Daily	**NFBP**	**6.91%**	**4.97%**
	CBLSTM	9.00%	6.68%
	LSTMCNN	9.08%	6.69%
	CGRU	9.10%	6.73%
	LSTM	9.11%	6.72%
	BILSTM	9.09%	6.72%
	SARIMAX	9.25%	6.87%
	DNN	8.97%	6.69%

Table 7
The promoting percentage for benchmark models against the proposed NFBP model in the testing phase. λ_{RMSE} = Promoting Percentage of Root Mean Square Error, λ_{APB} = Promoting Percentages of Absolute Percentage Bias and λ_{RRMSE} = Promoting Percentages of Relative Root Mean Square Error.

Predictive models	0.5 hr interval			6 hr interval			Daily interval		
	λ_{RMSE}	λ_{APB}	λ_{RRMSE}	λ_{RMSE}	λ_{APB}	λ_{RRMSE}	λ_{RMSE}	λ_{APB}	λ_{RRMSE}
CBLSTM	7.21%	9.03%	7.21%	21.24%	18.18%	21.25%	30.28%	34.18%	30.28%
LSTMCNN	8.58%	10.63%	8.57%	18.30%	16.39%	18.32%	31.49%	34.18%	31.49%
CGRU	5.11%	6.48%	5.11%	19.87%	18.70%	19.87%	31.76%	34.67%	31.76%
LSTM	6.12%	7.72%	6.12%	22.71%	21.31%	22.72%	31.85%	34.77%	31.85%
BILSTM	8.08%	9.09%	8.08%	19.54%	18.97%	19.54%	31.58%	34.70%	31.57%
SARIMAX	22.25%	14.55%	22.25%	27.52%	36.76%	27.53%	34.00%	37.59%	34.00%
DNN	11.27%	13.44%	11.27%	26.33%	25.16%	26.33%	29.91%	33.34%	29.91%

4. Experimental results and discussion

This section details the results obtained from the proposed hybrid NFBP model with a Kernel Density Estimation method for: (a) 0.5-hr ahead (30 min), (b) 6-hr ahead, and (c) one day ahead electricity demand prediction at Bulimba substation. The study also comprises of two unique aspects in the application of the proposed model for: (i) deterministic (point-based) and (ii) and interval-based electricity demand predictions.

4.1. Results of deterministic prediction using class a metrics

To assess the performance of the prediction methods, the deterministic prediction component of the proposed model (NFBP) is compared with DL models (LSTM LSTMCNN, BILSTM, DNN, CGRU, CBILSTM) and statistical model (SARIMAX) using the class A Metrics. In Class A metrics, the Correlation Coefficient (r) measures the degree of collinearity between observed and predicted G, where the value always lies between $-1 \leq r \leq 1$ (Deo et al., 2021). The r value closest to 1 shows a higher correlation between the observed and predicted values (Jayasinghe et al., 2022). Furthermore, the RMSE is the square root of the Mean Squared Error between the predicted and observed values, whereas the Mean Absolute Error (*MAE*) is the average absolute error between Observed and predicted values. Both *RMSE* and *MAE* can range from $0 \leq RMSE \; or \; MAE \leq \infty$ (Castillo-Botón et al., 2022). In terms of physical interpretations, the value of *RMSE* and *MAE* are ideally zero for a perfect prediction OF G (Jayasinghe et al., 2021). Similarly, the best model is likely to attain a low value of the Standard Deviation of Relative Error (*STDRE*) as well as the Explained Variance Score (E_{var}), ideally close to unity.

As shown in Table 4, the r, *RMSE*, *MAE*, *STDRE* and E_{var} of the NFBP model for 0.5-hr prediction horizon are $\approx$ 0.989, $\approx$ 0.664, $\approx$ 0.340, $\approx$ 5.983 and $\approx$ 0.957, respectively, with lower magnitude of *RMSE*, *MAE*, *STDRE* and higher magnitude of r and E_{var}. The NFBP model performance is better than other benchmarked models (LSTM LSTMCNN, BILSTM, DNN, CGRU, CBILSTM and SARIMAX) for all prediction horizons (0.5-hr, 6-hr and daily interval).

Similarly, the skill score, a metric that compares the NFBP model results with a simple persistence-based model, demonstrates that the NFBP model is more accurate than the persistence model (Fig. 5) as well as the other benchmarked models. It should be noted that the persistence model utilized here as a naïve reference model is defined

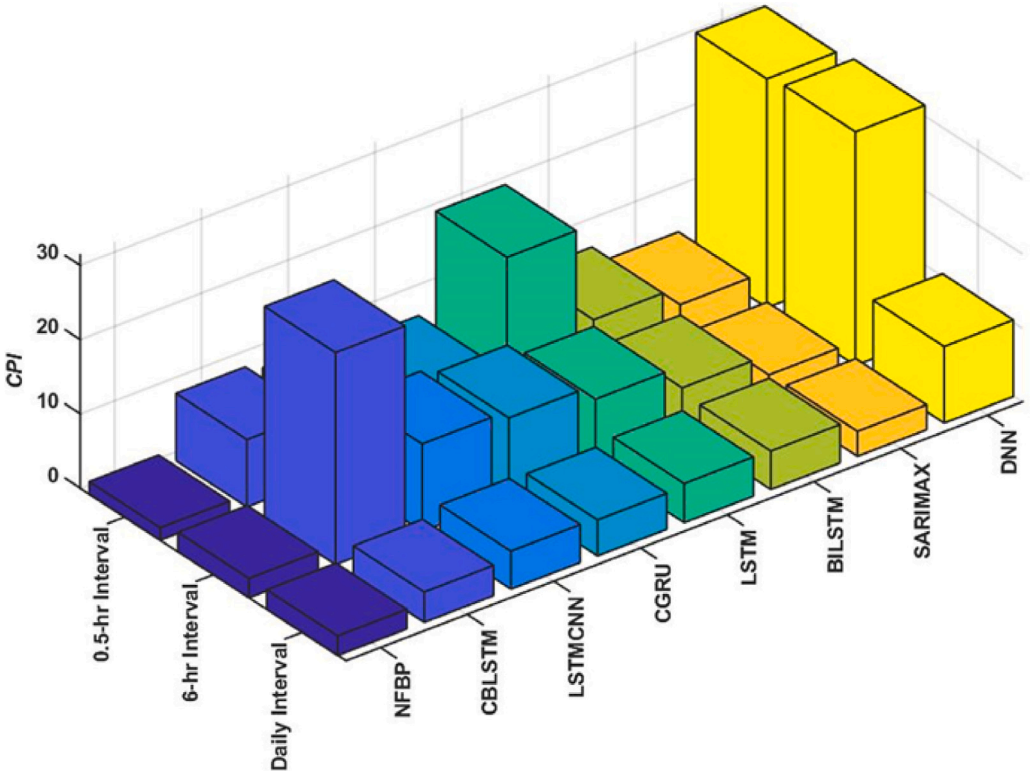


Fig. 8. Combined Performance Indicator (CPI) used to evaluate the proposed NFBP model relative to five other benchmarked models. Note: The name of each model are provided in Table 2.

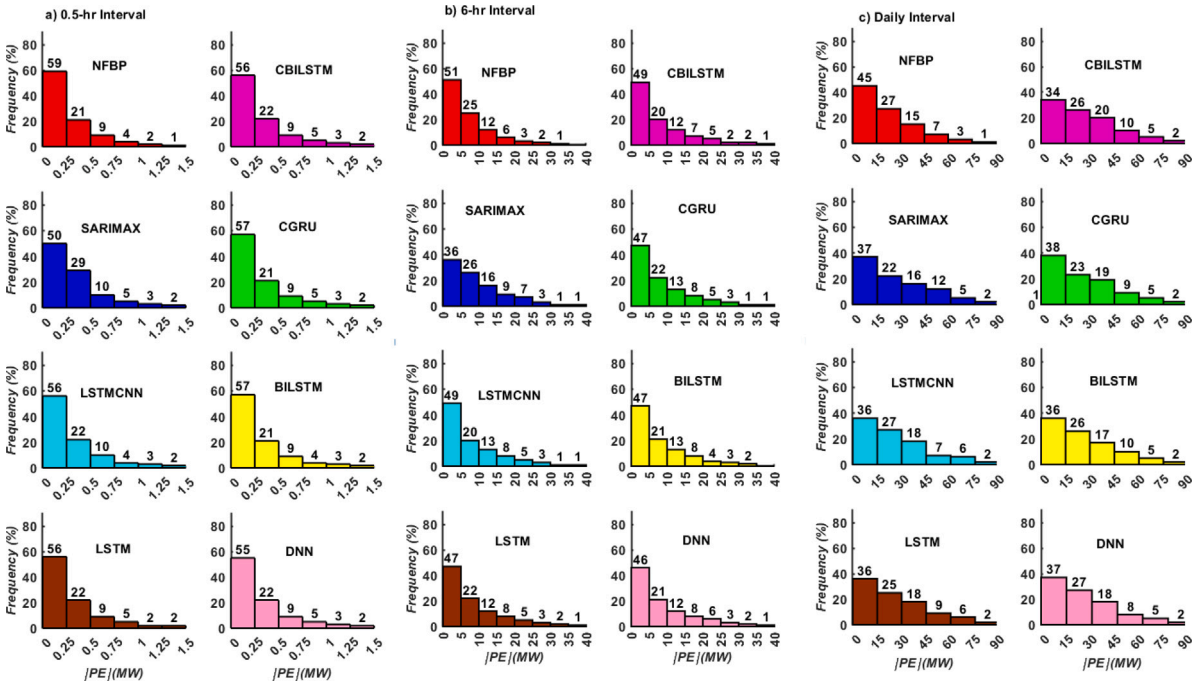


Fig. 9. Cumulative frequency of the PE generated by the NFBP compared to other benchmarked models at Bulimba substation for (a) 0.5-hr time interval, (b) 6h-time interval and (c) daily interval. Note: The name of each model is provided in Table 2.

by a prediction equal to the time-series' most recent value (i.e., from time t to time $t+1$). Additionally, the uncertainty at 95% U_{95} and t-Statistic TS metrics are shown in Table 4. The TS indicator is the ratio of the departure of estimated value of a parameter from its hypothesized value to its standard error, while the U_{95} is Uncertainty at 95% confidence level. A value of 0 is considered to be the perfect

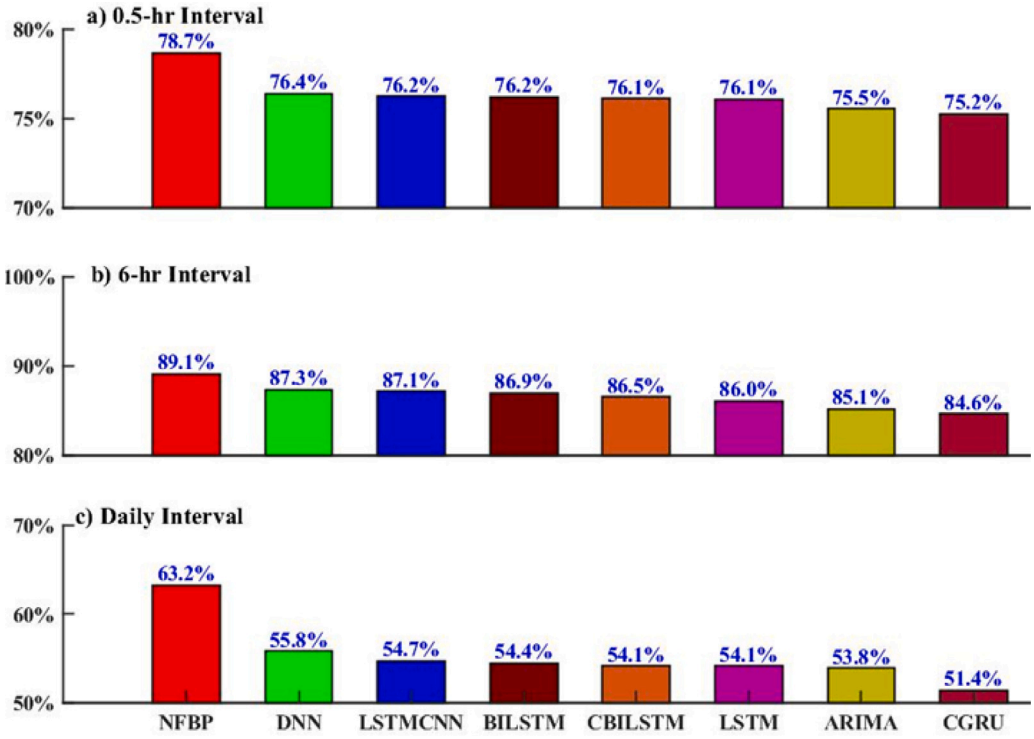


Fig. 10. Performance comparison of NFBP model compared to the other models in terms of the Directional Symmetry ($DS, \%$) measures within the testing phase at Bulimba substation for (a) 0.5-hr, (b) 6-hr and (c) day ahead prediction horizons. Note: The name of each model is provided in Table 2.

performance for these metrics. As evidenced in Table 5, the magnitude of U_{95} and TS are lower ($U_{95} \approx 1.840(0.5-\text{hr ahead}), U_{95} \approx 28.487(6-\text{hr ahead}), U_{95} \approx 83.145(day\ ahead), TS \approx 0.035(0.5-\text{hr ahead}), TS \approx 0.045(6-\text{hr ahead}), TS \approx 0.034(day\ ahead)$) compared to other benchmarked models. In general, the use of a diverse set of metrics enables one to comprehensively evaluate the efficacy of the proposed NFBP model.

Additionally, by examining the agreement between the predicted and observed G, further insights into the models are revealed. The predicted value (G_{pred}) vs the observed electricity demand value (G_{obs}) for three prediction horizons, namely 0.5 h ahead (30 min), 6-hr ahead and one day ahead) are plotted in Fig. 6. Each sub-panel consists of the least square regression line with the equation $y = mx + C$, where m (Gradient of best fit) and C (y-intercept) are constants used to quantify the model's accuracy. The optimum values of m and C are 1 and 0, respectively, for a perfect $(1:1)$ match between the predicted and observed G. For all three prediction horizons (Fig. 6), despite the significant scatter between the observed and the predicted G values, a reasonable linear fit is certainly apparent, albeit with varying degrees of precision for all the models under consideration. The NFBP model, in particular, has a relatively high magnitude of the gradient of the best-fit line (m) and a low magnitude of C. This indicates that the proposed NFBP model may be a more effective data-intelligent tool than other benchmark models.

We further evaluated the prediction models in Table 6 using normalized error metrics expressed in terms of the relative (%) $RMSE$ and MAE values. To explain this metric in its physical sense, a model's precision level is considered to be excellent if the *RRMSE* or *RMAE* $\leq$ 10%, good if 10% $\leq RMAE\ or\ RRMSE \leq$ 20%, fair if 20% $\leq RMAE\ or\ RRMSE \leq$ 30% and poor if the *RRMSE* or *RMAE* $\geq$ 30% (Deo et al., 2016). It is evident from Table 6 that the relative $RMSE/MAE$ is lowest over the ranges of 6.91–7.34% and 3.93–6.71%, respectively, for the NFBP model compared to other benchmarked models. By contrast, the SARIMAX model yielded an error of relative $RMSE \approx 12.08\%$ while that of the DNN model is $\approx$11.96% for 6-hr ahead prediction. This demonstrates that the DNN model's prediction is severely out of phase with the observations because the relative error exceeded the 10% threshold for a good model. Overall, the present analysis based on class A metrics provides persuasive evidence that the NFBP model has a significant ability to predict G at different prediction horizons, and its performance exceeds that of the alternative models.

4.2. Results of deterministic prediction using class b metrics

In addition to the Class A metrics mentioned so far, we use the Willmott's Index (E_{WI}), Nash–SutcliffeCoefficient (E_{NS}), and the Legates-Index McCabe's (E_{LM}) index to further assess the accuracy of the proposed NFBP model. It is imperative to note that these Class B metrics are structurally different from Class A metrics providing a normalized measure of the model performance. Notably, (E_{NS}) metric can determine the relative magnitude of the residual variance whereas (E_{WI}) can detect the additive and proportional differences in the observed and predicted means and variances and has the advantage over the r, $RMSE$ and MAE used as Class A metrics. Furthermore, (E_{WI}) represents the ratio of the MSE and the potential error. The potential error in the denominator represents the largest value that the squared difference of each pair (observed and predicted) can attain. The range of (E_{WI}) lies between 0 (no correlation) and 1 (perfect fit). Both (E_{LM}) and (E_{NS}) values vary from $-\infty \leq E_{NS}/E_{LM} \leq 1$.

Table 8
Evaluation of the NFBP model vs. LSTMCNN, BILSTM, DNN, CBILSTM, CNN and SARIMAX models in terms of the Diebold–Mariano (DM) test statistic.

(a) 0.5-hr Interval

	NFBP	CBLSTM	LSTMCNN	CGRU	LSTM	BILSTM	SARIMAX	DNN
NFBP		**5.412**	**7.967**	**8.233**	**5.234**	**3.854**	**1.556**	**9.865**
CBLSTM			1.690	−1.287	−0.524	1.096	1.158	2.702
LSTMCNN				−3.050	−1.696	−0.356	1.061	2.618
CGRU					1.679	1.248	1.227	6.297
LSTM						0.701	1.161	4.458
BILSTM							1.135	1.528
SARIMAX								−0.830

(b) 6-hr Interval

	NFBP	CBLSTM	LSTMCNN	CGRU	LSTM	BILSTM	SARIMAX	DNN
NFBP		**9.627**	**9.496**	**8.954**	**8.102**	**9.047**	**8.897**	**9.903**
CBLSTM			−2.510	−1.028	0.944	−1.156	1.991	3.192
LSTMCNN				1.931	3.293	1.202	3.034	6.035
CGRU					1.862	−0.337	2.447	4.908
LSTM						−2.127	1.397	2.441
BILSTM							2.661	3.780
SARIMAX								−0.319

(c) Daily Interval

	NFBP	CBLSTM	LSTMCNN	CGRU	LSTM	BILSTM	SARIMAX	DNN
NFBP		**5.271**	**5.455**	**5.826**	**5.601**	**5.596**	**6.486**	**5.843**
CBLSTM			0.783	0.832	1.074	0.800	1.804	−0.271
LSTMCNN				0.317	0.898	0.099	1.218	−1.322
CGRU					0.152	−0.315	1.192	−1.630
LSTM						−0.445	1.134	−1.630
BILSTM							1.188	−1.354
SARIMAX								−1.941

Note: The column of the table is compared with the rows, and if the result is positive, the model in the rows outperforms the one in the column; on the contrary, if it is negative, then the one in the column is superior. The best model is boldfaced (blue).

Table 9
Evaluation of the NFBP model vs. LSTMCNN, BILSTM, DNN, CBILSTM, CNN and SARIMAX models in terms of the Harvey–Leybourne–Newbold (HLN) test statistic.

a) 0.5-hr Interval

	NFBP	CBLSTM	LSTMCNN	CGRU	LSTM	BILSTM	SARIMAX	DNN
NFBP		**5.4256**	**7.9868**	**8.2532**	**5.2469**	**3.8636**	**1.5603**	**9.8898**
CBLSTM			1.6945	−1.2899	−0.5257	1.0988	1.1606	2.7084
LSTMCNN				−3.0574	−1.7002	−0.3572	1.0635	2.6245
CGRU					1.6829	1.2506	1.2301	6.3122
LSTM						0.7028	1.1642	4.4693
BILSTM							1.1379	1.5313
SARIMAX								−0.8322

b) 6-hr Interval

	NFBP	CBLSTM	LSTMCNN	CGRU	LSTM	BILSTM	SARIMAX	DNN
NFBP		**9.780**	**9.647**	**9.097**	**8.230**	**9.191**	**9.039**	**10.061**
CBLSTM			−2.550	−1.044	0.959	−1.174	2.023	3.242
LSTMCNN				1.962	3.345	1.221	3.082	6.131
CGRU					1.891	−0.342	2.486	4.986
LSTM						−2.161	1.419	2.480
BILSTM							2.704	3.840
SARIMAX								−0.324

c) Daily Interval

	NFBP	CBLSTM	LSTMCNN	CGRU	LSTM	BILSTM	SARIMAX	DNN
NFBP		**5.533**	**5.726**	**6.115**	**5.879**	**5.874**	**6.809**	**6.133**
CBLSTM			0.822	0.873	1.128	0.839	1.894	−0.284
LSTMCNN				0.333	0.943	0.104	1.278	−1.388
CGRU					0.160	−0.331	1.252	−1.711
LSTM						−0.467	1.191	−1.711
BILSTM							1.247	−1.421
SARIMAX								−2.037

Note: The column of the table is compared with the rows, and if the result is positive, the model in the rows outperforms the one in the column; on the contrary, if it is negative, then the one in the column is superior. The best model is boldfaced (blue).

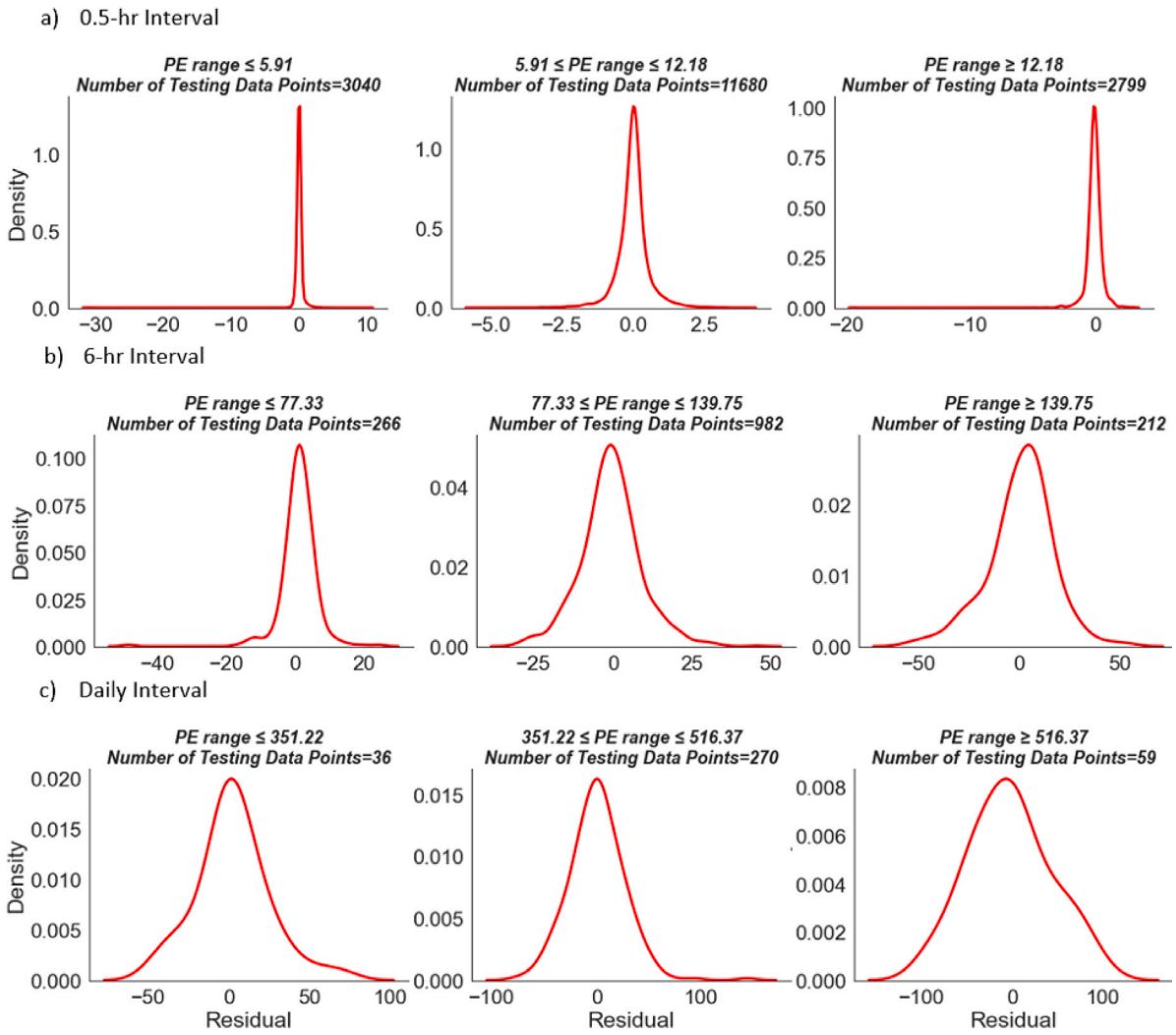


Fig. 11. Probability density function curves of prediction errors for three PE ranges (Eq. (40)–(42)) fitted using the KDE method in (a) 0.5-hr, (b) 6-hr, and (c) day ahead prediction horizons).

Table 5 summarizes the average performance of seven benchmark models against the proposed NFBP model in respect to four normalized metrics based on Willmott's Index, Legates & McCabe's Index and the Theil's inequality coefficient.

As per Table 5, the proposed NFBP model's results for these performance criteria are closer to unity and higher in magnitude ($E_{WI} \approx 0.971$ (0.5-hr ahead), $E_{WI} \approx 0.937$ (6-hr ahead), $E_{WI} \approx 0.912$ (day ahead), $E_{NS} \approx 0.957$ (0.5-hr ahead), $E_{NS} \approx 0.902$ (6-hr ahead), $E_{NS} \approx 0.879$ (day ahead), $E_{LM} \approx 0.868$ (0.5-hr ahead), $E_{LM} \approx 0.727$ (6-hr ahead), $E_{LM} \approx 0.675$ (day ahead)) than any of the other benchmark models. The magnitude of E_{WI}, E_{NS} and E_{LM} ranges from 0.959 – 0.968, 0.936 – 0.952 and 0.849 – 0.858, respectively, for 0.5-hr ahead prediction, 0.893 – 0.909, 0.840 – 0.859 and 0.627 – 0.682, respectively, for 6-hr ahead prediction, and 0.845 – 0.848, 0.784 – 0.797 and 0.553 – 0.567, respectively, for day ahead prediction.

It should also be noted that both (E_{WI} and E_{NS} are standardized measures of the degree of model prediction error bounded by [0, 1] with 1 as the ideal model's value, whereas TIC aims to measure of how well the modelled values of the electricity demand time series data compare to the corresponding time series of observed values. Notably, the TIC value of zero is expected to represent an equal/identical distribution of measured and modelled electricity demand in the testing phase whereas higher values are expected to represent a higher level of inequality. In accordance with Table 5, TIC, representing a model's perfect fit when equal to zero for the proposed NFBP model, are closer to 0 and relatively smaller in magnitude than the comparative models. This reaffirms the suitability of the NFBP model for the prediction of G at different prediction horizons.

We now revert to two complementary metrics, the Kling–Gupta Efficiency and the Absolute Percentage Bias to assess the performance of the proposed NFBP model. In its physical sense, KGE is based on the decomposition of the E_{NS} used earlier into its correlation, bias, and the ratio of variances or coefficients of variation in a more balanced way to tackle the drawbacks with E_{NS}. The results demonstrate that the proposed NFBP predictive model greatly outperforms the counterpart models with a relatively high KGE and a relatively low APB, to accord with Fig. 7. For instance, the KGE and APB value when comparing the NFBP model with a subsequent best prediction model (i.e., CBILSTM, LSTMCNN, CGRU, LSTM, BILSTM, SARIMAX) and DNN in the grouping [NFBP: CBILSTM: LSTMCNN: CGRU: LSTM: BILSTM: SARIMAX: DNN] were [0.945,5.06% : 0.906,6.80% : 0.908,6.80% : 0.904,6.82% : 0.903,6.830% : 0.906,6.82% : 0.911,6.973% : 0.902,6.758%] for the day ahead prediction. In tandem with class A Metrics, the proposed NFBP model's performance based on Class B Metrics also confirms its greater reliability to predict electricity demand at different prediction

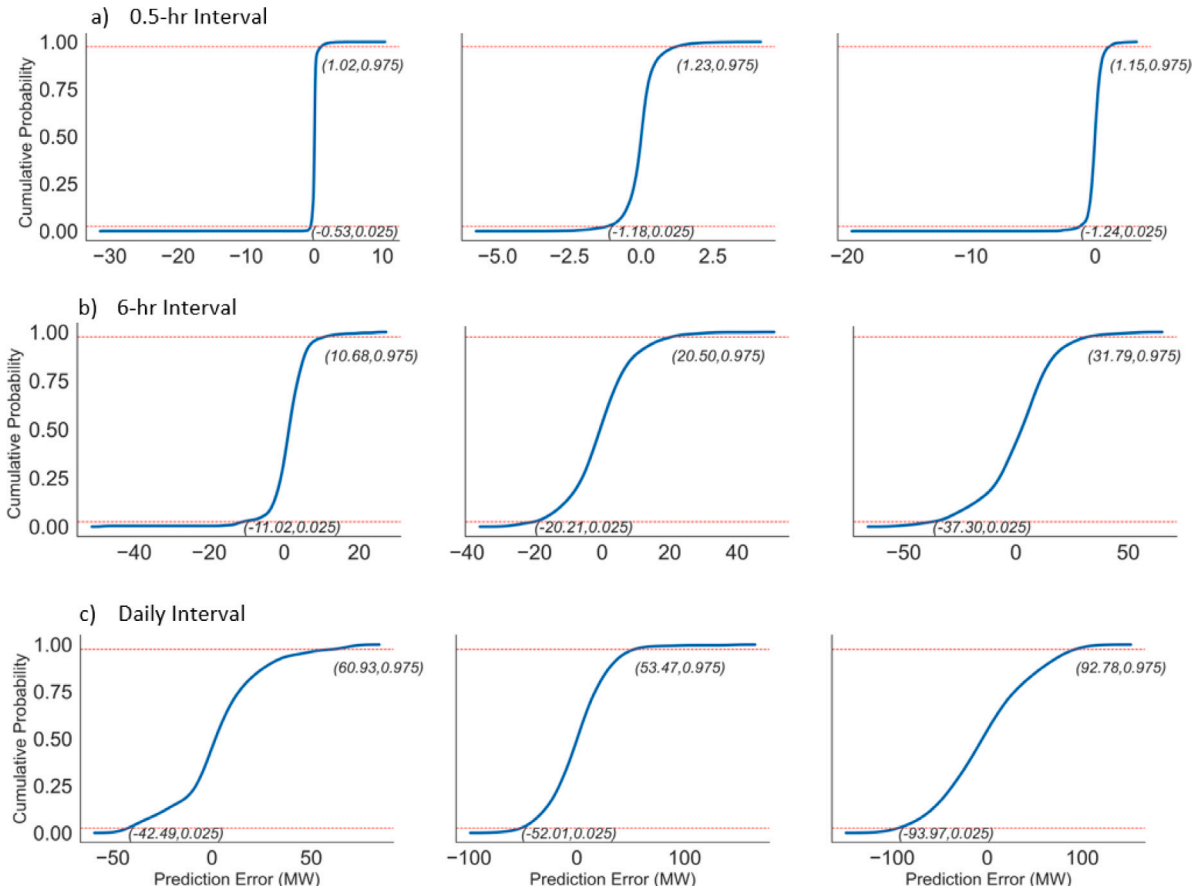


Fig. 12. Cumulative distribution function of PE ranges (Eq. (40)–(42)) in (a) 0.5-hr, (b) 6-hr, and (c) day ahead prediction horizons. Note: The bottom and top red lines show the confidence level (95%). The upper and lower prediction errors are also shown.

horizons for the specific case of Bulimba substations, with similar outcomes for the other tested sites.

4.3. Results of deterministic prediction using class c metrics

A further evaluation of the NFBP model is conducted using Class C metrics (indicators of cumulative distribution function (CDF) similitude). The Combined Performance Index (CPI) incorporates two measures of CDF similarity (KSI and OVER), as well as standard statistical measurements ($RMSE$), and calculates the average of the three results. As demonstrated in Espinar et al. (2009), the CPI primary benefit is its high discrimination between different models. The 3D bar plot for $CPI(\%)$ is depicted in Fig. 8, it can be seen that the CPI magnitude is lower (≈1.663%, ≈2.619% and ≈3.359% for 0.5-hr, 6-hr and day ahead prediction horizons, respectively) than that of comparative models. It is thus clear that the proposed NFBP model was able to produce more accurate results for G prediction compared to all of the benchmark DL methods (i.e., CBILSTM, LSTMCNN, CGRU, LSTM, BILSTM, and DNN), and other statistical methods (i.e., SARIMAX).

Besides Class A, B and C metrics, further evaluation was done by comparing the frequency distribution of absolute PE produced by NFBP and other benchmark methods, shown in Fig. 9. Each error bin contains the percentage of all tested points, shown at the top of each error bar. Each error bin has a size of 0.25 MW in 0.5-hr ahead prediction, 5 MW in 6-hr ahead prediction, and 15 MW for the day ahead prediction. Interestingly, G predictions using the NFBP model were found to have the greatest frequency of errors within the smallest error bracket, encompassing 59%, 51% and 45% of the test data. On the other hand, the CBILSTM, LSTMCNN, CGRU, LSTM, BILSTM, SARIMAX, and DNN accumulated 56%, 56%, 56%, 57%, 56%, 57%, 50% and 55%, respectively, for 0.5-hr ahead prediction. Following Figs. 5, 7, and 8, and Tables 4–5, the NFBP model generates most of its prediction errors within the lowest magnitude band, making it more accurate in predicting G at different time horizons.

In this study, Promoting Percentages (λ_{RMSE}, λ_{RRMSE} and λ_{APB}) are also evaluated to determine the improved performance of the models. The promoting percentage are defined by, $\lambda_{APB} = \frac{(APB_1 - APB_2)}{APB_1}$, $\lambda_{RMSE} = \frac{(RMSE_1 - RMSE_2)}{RMSE_1}$ and $\lambda_{RRMSE} = \frac{(RRMSE_1 - RRMSE_2)}{RRMSE_1}$. In these equations, APB_1, $RMSE_1$ and KGE_1 are objective model (NFBP) performance metrics, and APB_2, $RMSE_2$ and KGE_2 are benchmark model performance. The NFBP model for all prediction horizons displays improved performance, consistent with past findings. For instance, at 0.5-hr ahead prediction, the NFBP considerably outperformed the CBILSTM, LSTMCNN, CGRU, LSTM, BILSTM, SARIMAX, and DNN models by λ_{RMSE} ≈7%, ≈8%, ≈5%, ≈6%, ≈8%, ≈22%, and ≈11%, respectively. Similarly, for 6-hr and day ahead prediction, λ_{RMSE}, λ_{RRMSE} and λ_{APB} are >15% for the CBILSTM, LSTMCNN, CGRU, LSTM, BILSTM, SARIMAX, and DNN models. As a result, with positive promoting percentage errors, the NFBP model seems to have highly developed predictive abilities that enabled it to accurately predict G (see Table 7).

Although the proposed NFBP model on an overall basis showed the best performance, the results from the hybrid DL models used as comparisons (i.e., CBILSTM, LSTMCNN, CGRU, BILSTM) occasionally varied in respect to the other statistical performance metrics. We therefore present the point-based prediction results for the proposed NFBP and the benchmark models based on Directional Symmetry (DS). In fact, the DS is a percent of the prediction where the movement of the prediction is the same as the movement of the target variable and calculated using Eq. (57). The DS values are shown as a bar chart for all the models in Fig. 10. It can be seen that the DS value up to

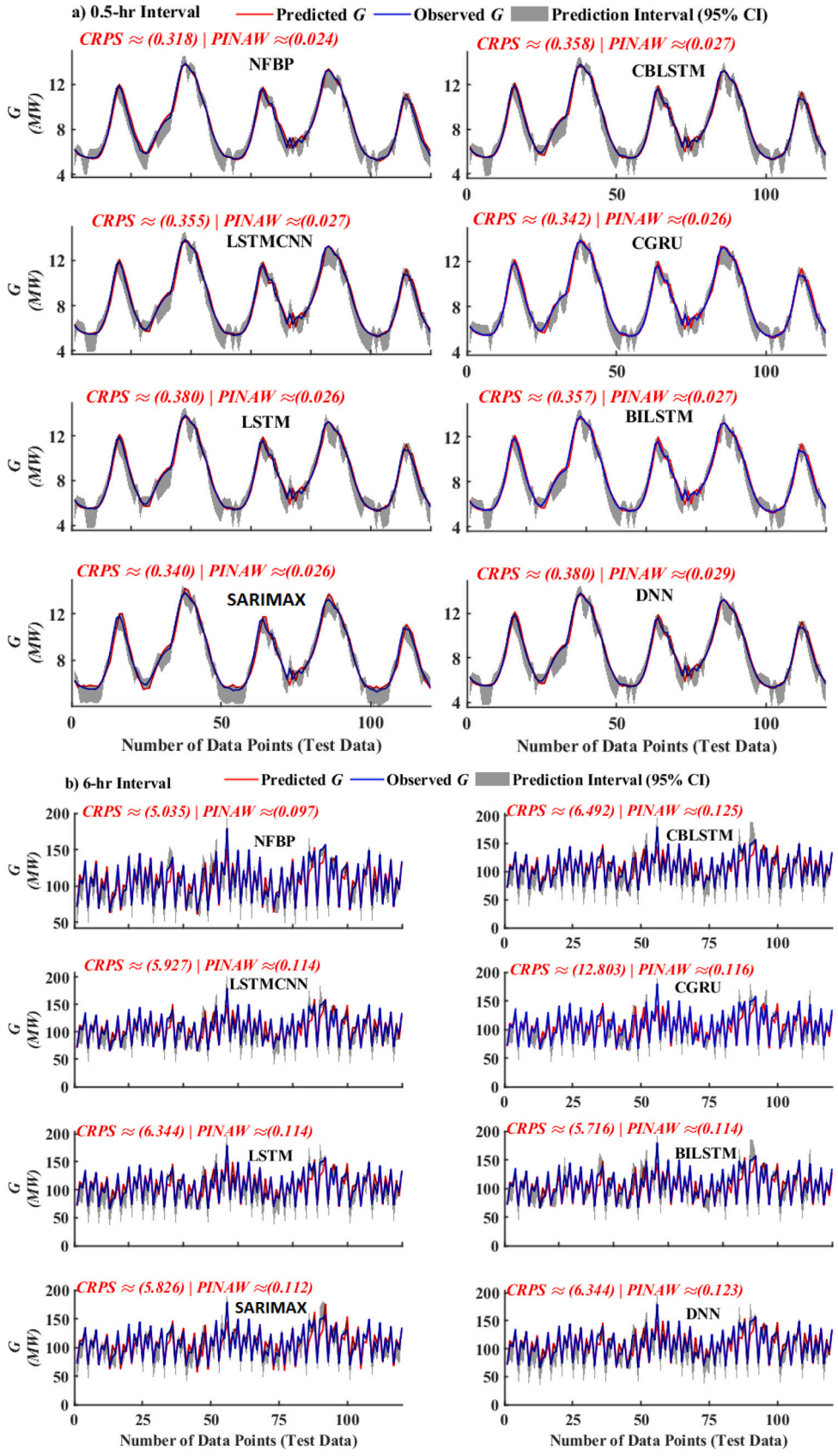


Fig. 13. PIs constructed by KDE method for the proposed model (NFBP) and other benchmark models at the confidence level of 95% for (a) 0.5-hr, (b) 6-hr and (c) day ahead prediction horizons. Note: $CRPS$ and $PINAW$ values are shown in the figure.

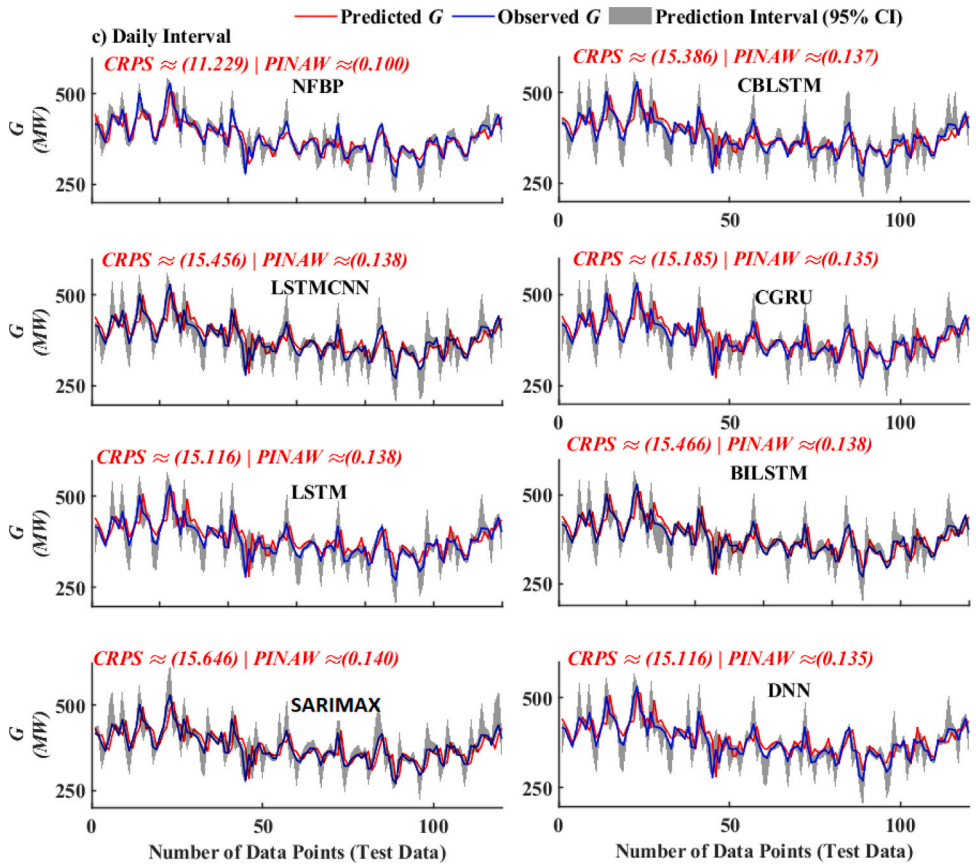


Fig. 13. (*continued*).

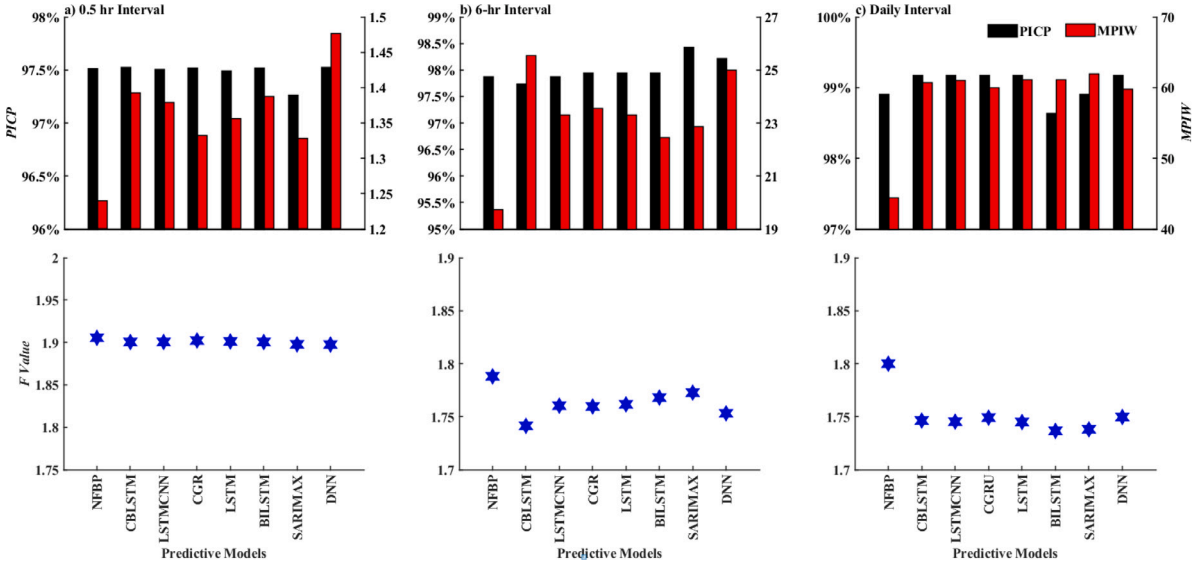


Fig. 14. Prediction Interval Coverage Probability ($PICP$,%) and Mean Prediction Interval Width ($MPIW$,MWh) for (a) 0.5-hr, (b) 6-hr and (c) day ahead prediction horizons. The bottom chart shows the value of comprehensive index (F).

≈79%, ≈89% and ≈63% are achieved with the proposed NFBP model for 0.5-hr, 6-hr and day ahead prediction horizons, respectively.

Lastly, to discuss the efficacy of the suggested NFBP model specifically for point prediction, the Diebold–Mariano (DM) test is used to evaluate the equal predictive ability of two competing models. Further, the modified version of DM for large samples, known as the Harvey, Leybourne and Newbold (HLN) test, is also utilized. Table 8 and Table 9 display the findings of the DM and HLN test using the $RMSE$ loss

Table 10
Probabilistic prediction results for 95% confidence level in terms of Winkler Score (WS) and Average relative interval length (ARIL) for four substations. The optimal model is boldfaced in blue.

Time interval	Predictive model	Model performance metrics	
		WS	ARILV
0.5-hr Interval	**NFBP**	**1.767**	**0.154**
	CBLSTM	1.994	0.174
	LSTMCNN	1.997	0.173
	CGRU	1.905	0.166
	LSTM	1.957	0.170
	BILSTM	1.984	0.173
	SARIMAX	1.872	0.163
	DNN	2.114	0.185
6-hr Interval	**NFBP**	**23.764**	**0.200**
	CBLSTM	29.750	0.262
	LSTMCNN	27.302	0.239
	CGRU	27.223	0.242
	LSTM	26.876	0.232
	BILSTM	26.566	0.226
	SARIMAX	27.295	0.230
	DNN	28.854	0.256
Daily Interval	**NFBP**	**49.780**	**0.104**
	CBLSTM	68.732	0.141
	LSTMCNN	68.847	0.142
	CGRU	67.927	0.138
	LSTM	69.120	0.142
	BILSTM	68.718	0.141
	SARIMAX	68.571	0.144
	DNN	66.772	0.139

function for 0.5-he, 6-hr and day ahead prediction of G at Bulimba substation. Both DM and HLN test results are positive and >1 for NFBP compared with other benchmark models, which further proves the superiority of the proposed NFBP model for the prediction of G.

4.4. Results of interval prediction using class d metrics

After obtaining the deterministic point prediction, the probability density is computed using the KDE approach. Electricity demand G exhibits sporadic patterns. Therefore, the point PE are split into three categories to enhance the quality of the PI. The suggested interval prediction model is trained using the three degrees of prediction errors, which were determined based on the average and standard deviation of the point prediction outcomes in the testing set (Eq. (40)–(42)). These three parts are $Y_{lev1} < \mu - \sigma$, $\mu - \sigma \leq Y_{lev2} \leq \mu + \sigma$ and $Y_{lev3} \geq \mu + \sigma$, where Y_{lev1}, Y_{lev2}, and Y_{lev3} are the three parts obtained by splitting, and the corresponding three error parts (PE_{range} in Fig. 11) are obtained according to them. The Gaussian KDE estimation was performed on the split PE to calculate the PI at 95% confidence level. The probability density distribution of PE for three split levels at three distinct prediction horizons (0.5-hr, 6-hr and day ahead) is depicted in Fig. 11. The KDE fitting effect is significantly impacted by the kernel density estimation's bandwidth (h). In our work, the ideal bandwidth for the three PE levels is established following a grid search approach that minimizes the $MISE$ (Eq. (9). For 0.5-hr ahead prediction, the optimal bandwidth (h) was 0.027, 0.044 and 0.067. For 6-hr ahead prediction, optimal h was 0.927, 1.765 and 3.750 and 5.53, 7.06 and 17.92 for the day ahead prediction. In this study, an integration method is employed to get the upper and lower bounds of prediction errors at a target confidence level.

Fig. 12 displays the Cumulative Probability Distributions for three PE levels (ranges) at different prediction horizons. As can be seen from Fig. 12, the CDF of PE vary depending on the PE ranges, this demonstrates that the PE splitting strategy implemented in this study can better characterize the probability distribution of PE. The intervals of the PE at 90% confidence level are also shown in Fig. 12. According to Fig. 11, Fig. 12 and Eq. (42) and (43), the PIs of electricity demand with the confidence level of 95% can be obtained (see Fig. 13).

As shown in Fig. 13, the interval prediction results of the NFBP model can promise to follow the variation of the electricity demand G time-series while covering more observations with a narrower PI than other models. The Continuous Ranked Probability Score ($CRPS$) and Prediction Interval Normalized Average Width (PINAW) score are shown in Fig. 13, the discrepancy between the expected and observed CDF is measured by the $CRPS$ and quantifies both sharpness and reliability. Whereas the $PINAW$ quantifies the prediction sharpness. A sharper PI is not always better, as it may come at the expense of reliability. Therefore, by computing the $CRPS$, we can assess if the prediction from the NFBP model is closer to the true distribution than the benchmark models. The $CRPS$ and $PINAW$ values are indeed lower for the NFBP model for any prediction horizons (for 0.5-hr ahead: $CRPS \approx 0.318$, $PINAW \approx 0.024$, for 6-hr ahead: $CRPS \approx 5.035$, $PINAW \approx 0.097$ and for the day ahead: $CRPS \approx 11.229$, $PINAW \approx 0.100$), which proves that the NFBP model integrated with KDE can provide more comprehensive prediction information.

In addition, Prediction Interval Coverage Probability ($PICP$) and Mean Prediction Interval Width ($MPIW$) are utilized to evaluate the effectiveness of the benchmark models and the proposed model. The proportion of measured values that fall within the range of PIs is assessed by the index $PICP$. In contrast, $MPIW$ represents the mean interval width and is a crucial parameter for evaluating PIs. Even if all measured values are contained inside the interval, and the $PICP$ is 100%, a too-wide $MPIW$ signifies excessive uncertainty and is useless as a prediction. To address this issue, a comprehensive index F (Eq. (38)) is used to assess the effectiveness of the PI. To determine the best PI, the index F integrates (Eq. (38)) the two opposing indices. The interval prediction model is better for the greater value of F. The interval predicting results of the proposed model and benchmark models are shown in Fig. 14. The NFBP model's $PICP$ is not significantly different from the other benchmark models at the confidence level of 95%, but its $MPIW$ values are lower (for the day ahead prediction; NFBP: ≈ 44.37, CBILSTM: ≈ 60.74, LSTMCNN:

Table 11
Prediction Interval performance based on splitting strategies applied in this study.

(a) 0.5-hr Interval						
	The applied splitting strategy			Without Splitting		
	PICP	*PINAW*	*F*	*PICP*	*PINAW*	*F*
NFBP	**0.975**	**0.024**	**1.906**	0.950	0.044	1.824
CBLSTM	**0.975**	**0.027**	**1.900**	0.950	0.051	1.813
LSTMCNN	**0.975**	**0.027**	**1.901**	0.950	0.050	1.813
CGRU	**0.975**	**0.026**	**1.902**	0.950	0.049	1.816
LSTM	**0.975**	**0.026**	**1.901**	0.950	0.050	1.815
BILSTM	**0.975**	**0.027**	**1.900**	0.950	0.051	1.812
SARIMAX	**0.973**	**0.026**	**1.898**	0.950	0.048	1.818
DNN	**0.975**	**0.029**	**1.898**	0.950	0.053	1.809
(b) 6-hr Interval						
	The applied splitting strategy			Without Splitting		
	PICP	*PINAW*	*F*	*PICP*	*PINAW*	*F*
NFBP	**0.979**	**0.097**	**1.788**	0.952	0.218	1.577
CBLSTM	**0.977**	**0.125**	**1.741**	0.950	0.261	1.522
LSTMCNN	**0.979**	**0.114**	**1.760**	0.952	0.257	1.530
CGRU	**0.979**	**0.116**	**1.760**	0.951	0.257	1.529
LSTM	**0.979**	**0.114**	**1.762**	0.951	0.266	1.519
BILSTM	**0.979**	**0.110**	**1.768**	0.951	0.258	1.527
SARIMAX	**0.984**	**0.112**	**1.773**	0.951	0.263	1.522
DNN	**0.982**	**0.123**	**1.753**	0.951	0.269	1.515
(c) Daily Interval						
	The applied splitting strategy			Without Splitting		
	PICP	*PINAW*	*F*	*PICP*	*PINAW*	*F*
NFBP	**0.989**	**0.100**	**1.800**	0.953	0.281	1.503
CBLSTM	**0.992**	**0.137**	**1.746**	0.951	0.367	1.410
LSTMCNN	**0.992**	**0.138**	**1.745**	0.953	0.376	1.404
CGRU	**0.992**	**0.135**	**1.749**	0.953	0.378	1.402
LSTM	**0.992**	**0.138**	**1.745**	0.953	0.374	1.406
BILSTM	**0.986**	**0.138**	**1.737**	0.956	0.372	1.411
SARIMAX	**0.989**	**0.140**	**1.738**	0.956	0.379	1.404
DNN	**0.992**	**0.135**	**1.750**	0.953	0.375	1.405

Table A.1
List of acronyms.

Term	Acronyms
Artificial Intelligence	AI
Artificial Neural Network	ANN
Autoregressive Integrated Moving Average	ARIMA
Bidirectional LSTM	BILSTM
Convolutional Neural Network	CNN
Deep Learning	DL
Deep Neural Network	DNN
Empirical Mode Decomposition	EMD
Expert System	ES
Explained Variance Score	E_{var}
Extreme Learning Machine	ELM
Fuzzy Logic's	FL
Gated Recurrent Unit	GRU
Hybrid Intelligent Algorithm	HIA
Kernel Density Estimation	KDE
Long Short-Term Memory	LSTM
Lower–Upper Bound Estimation	LUBE
Machine Learning	ML
Mean Absolute Error	MAE
Mean Square Error	MSE
Mean-Variance	MV
Mutual Information	MI
Neural Facebook Prophet	NFBP
Neural Networks	NNs

(*continued on next page*)

≈ 61.04, CGRU: ≈ 59.94, LSTM: ≈ 61.15, BILSTM: ≈ 6.10, SARIMAX: ≈ 61.91, and DNN: ≈ 59.76) and $F-Value$ are higher (for the day ahead prediction; NFBP: ≈ 1.8, CBILSTM: ≈ 1.74, LSTMCNN: ≈ 1.74, CGRU: ≈ 1.74, LSTM: ≈ .74, BILSTM: ≈ 1.73, SARIMAX: ≈ 1.73, and DNN: ≈ 1.75) than those of benchmark models. Furthermore, the Average Relative Interval Length $ARIL$ and the Winkler score WS value of the

Table A.1 (*continued*).

Partial Autocorrelation Function	PACF
Prediction Errors	PE
Prediction Interval	PI
prediction interval nominal confidence	PINC
Quantile Regression	QR
Random Forest	RF
Rectified Linear Unit	ReLU
Root Mean Square Error	RMSE
Standard Deviation of Relative Error	STDRE
Support Vector Machines	SVM
Sustainable Development Goal 7	SDG7
Wavelet Networks	WN

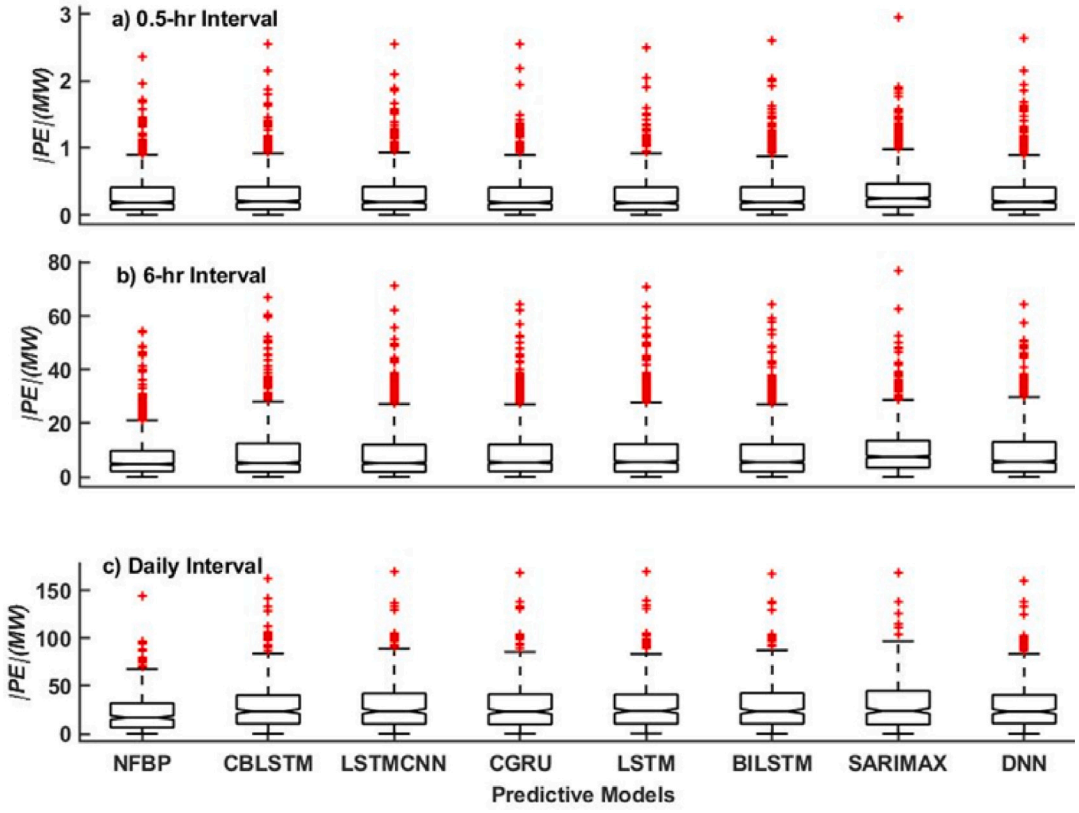


Fig. B.1. Box plot of the absolute prediction error of daily G generated by the proposed NFBP model in respect to seven other comparative models within the testing phase at Bulimba substation for (a) 0.5-hr time interval, (b) 6h-time interval and (c) daily interval. Note: The name of each model is provided in Table 2.

PI are reported in Table 10. The $ARIL$ is comparable to the $MPIW$ and considers the average width of uncertainty bounds with respect to the observed value; the PI performs best when the $ARIL$ is the smallest. For a particular confidence level, the WS thoroughly assess the PIs' reliability and accuracy. When the WS has a lower absolute value, the PIs are of higher quality. From Table 10, it can be seen that the $ARIL$ and WS values are lower for the NFBP-KDE model compared to the benchmark models. Congruence with earlier findings (e.g., in Fig. 13), the PI of NFBP integrated with KDE has a greater quality to assure reliability with a narrower interval, providing decision-makers with more precise information on electricity demand.

Lastly, to validate the efficacy of the splitting strategy utilized in this study, Table 11 was created to show the $PICP$, $PINAW$ and F value of interval prediction. It can be seen that with three parts of splitting strategy, the performance metrics are improved comparatively for 95% confidence level.

5. Conclusions, limitation and future research directions

A reliable point and interval electricity demand (G) prediction methodology is essential for ensuring the reliability of energy systems planning and operation. In this study, a point and interval prediction model based on Neural Facebook prophet (NFBP) is proposed. The NFBP is a hybrid model in which some model components are configured as neural networks to further improve the performance of time-series prediction. For the empirical analysis, G data from the Bulimba substation was used to evaluate the validity and applicability of NFBP for three different prediction horizons, i.e., 0.5-hr ahead, 6-hr ahead and Day ahead. Moreover, to identify the sequence length of G, which is used as input for NFBP, the PACM is used. As a first step, the point prediction error (PE) is calculated by comparing the actual G with the model's predicted values. Afterwards, a splitting strategy based on the mean and standard deviation of PE is presented to solve the problem of fitting the PE of the G with one PDF. Next, the Gaussian KDE method is utilized for fitting the PDF curves of different levels PE. The hyperparameter (Bandwidth) of KDE is optimized using the

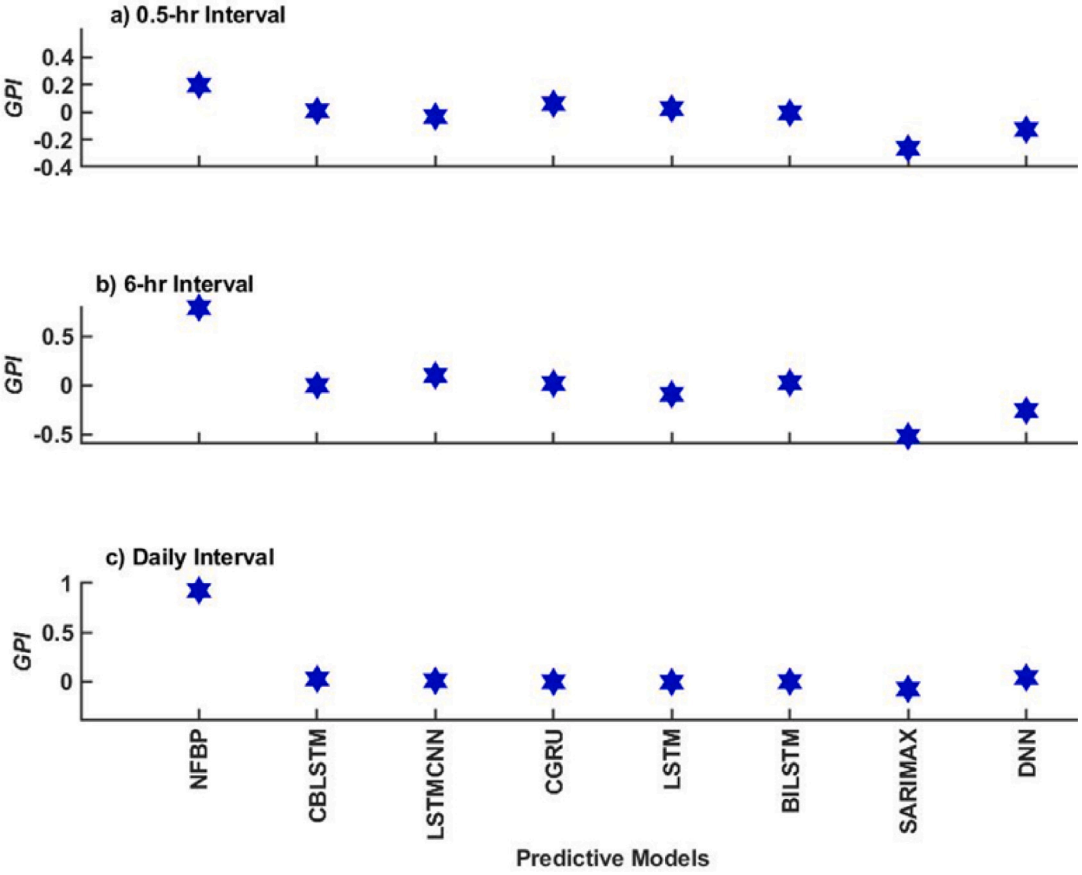


Fig. B.2. Global Performance Indicator (GPI) used to evaluate the proposed NFBP model relative to five other benchmarked models. Note: The name of each model is provided in Table 2.

grid search approach by minimizing the mean Integrated Square Error. Then, the upper bounds and the lower bounds of the best intervals of PE are added to the point predictions to attain the final PI of G.

To verify the superiorities of the proposed model (NFBP), this study compares the proposed NFBP with CBILSTM, LSTMCNN, CGRU, LSTM, BILSTM, SARIMAX, and DNN models and evaluates the performances of predicted G in terms of the deterministic and probabilistic metrics. The obtained conclusions are as follows:

- Point prediction results on three different prediction horizons demonstrate that $RMSE$, MAE and $RMAE$ and $RRMSE$ of hybrid NFBP model are lower than benchmark models (CBILSTM, LSTMCNN, CGRU, LSTM, BILSTM, SARIMAX, and DNN), indicating that hybrid NFBP model has a higher prediction accuracy.
- The magnitude of normalized metrics (Willmott's Index (E_{WI}), Nash–SutcliffeCoefficient (E_{NS}) and the Legates & McCabe's (E_{LM}) Index of Agreement) for NFBP model are higher than that of the benchmark models, which shows a higher prediction capability of the NFBP.
- The positive value of promoting percentage error (λ_{RMSE}, λ_{RRMSE} and λ_{APB}), DM and HLN Statistic further indicate that the NFBP model has higher prediction accuracy and stability than the other seven benchmark models.
- In terms of interval prediction, at 95% confidence level, PINAW values of NFBP, CBILSTM, LSTMCNN, CGRU, LSTM, BILSTM, SARIMAX, and DNN in 0.5-hr ahead prediction are ≈0.024, ≈0.027, ≈0.027, ≈0.026, ≈0.026, ≈0.027, ≈0.026, and ≈0.029, respectively, indicating that the proposed NFBP model shows superiority in interval prediction, whose PI has higher interval coverage and narrower interval width.

In conclusion, the proposed NFBP integrated with KDE is a promising prediction tool demonstrating a significant advantage in precise point prediction and reliable interval prediction than the benchmark models. These results indicate that the proposed model (NFBP) can also offer a new research idea for the other datasets with seasonal fluctuations in many other fields, including monthly/seasonal air pollutants concentrations, economic indicators, hydrology, solar, wind and so on. While the proposed model exhibits commendable performance, further enhancement is possible through time-series decomposition methods like variation mode decomposition or empirical mode decomposition. Hybridization can be pursued by combining the proposed NFBP with other deep learning models (e.g., LSTM, DNN, CNN) for more effective feature mapping. Integrating weather-based parameters as exogenous variables holds potential to augment the model's performance. Notably, this study employed grid search for Mean Integrated Squared Error ($MISE$) to estimate KDE hyperparameters; however, $MISE$'s sensitivity to outliers, uniform error treatment, limitations in capturing asymmetry, and potential trade-offs warrant careful consideration, including the risk of sacrificing interval forecast accuracy. Future work will explore decomposition methods, exogenous variable integration, and the utilization of Quantile Loss, Kullback–Leibler Divergence, and Probability Integral Transform for probabilistic prediction.

CRediT authorship contribution statement

Sujan Ghimire: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software. **Ravinesh C. Deo:** Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Validation, Writing – review & editing. **S. Ali Pourmousavi:** Writing – review & editing. **David Casillas-Pérez:** Formal analysis,

Investigation, Visualization, Writing – original draft, Writing – review & editing, Supervision. **Sancho Salcedo-Sanz:** Formal analysis, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

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

Acknowledgements

The authors thank the data providers, reviewers and the Editor for their thoughtful comments, suggestions and the review process. This study has been partially supported through a project PID2020-115454GB-C21 funded by the Spanish Ministry of Science and Innovation (MICINN).

Appendix A. Acronyms

Table A.1 provides the acronyms used in this paper.

Appendix B. Further results and discussions

Box plots of the model error, represented as $|PE|$, for the proposed NFBP and the benchmark models are presented in Fig. B.1. Evidently, the absolute Prediction Error ($|PE|$) distribution generated by the proposed NFBP model for all prediction time horizons is relatively smaller (when few outliers are disregarded) compared to CBILSTM, LSTMCNN, CGRU, LSTM, BILSTM, SARIMAX, and DNN models. The proposed NFBP model, therefore, achieved the best accuracy at three distinct prediction horizons, as shown by the frequency distribution plots and the box plots.

Each model's GPI was examined and the models were categorized as a result. The model with the highest GPI value was deemed the best in the selected case. Fig. B.2 shows the GPI values for each model. NFBP model has the best performance for prediction of G for three prediction horizons with $GPI \approx 0.2$, ≈ 0.8, and ≈ 0.9 for 0.5-hr ahead, 6-hr ahead and day-ahead prediction. For 0.5-hr ahead prediction, CGRU (≈0.06) is second best model followed by LSTM (≈0.02), CBILSTM (≈0.009), BILSTM (≈−0.05), LSTMCNN (≈−0.03), DNN (≈−0.12), and SARIMAX (≈−0.26) models. For 6-hr ahead prediction, LSTMCNN (≈0.09) is the second best model, followed by BILSTM (≈0.023), CGRU (≈0.015), CBLSTM (≈ −0.004), LSTM (≈−0.09), DNN (≈−0.26), and SARIMAX (≈−0.52) models. Similarly, for the day ahead prediction, DNN (≈ 0.03) is the second best model, followed by CBLSTM (≈0.02), LSTMCNN (≈0.004), BILSTM (≈−0.004), CGRU (≈−0.006), LSTM (≈−0.009), SARIMAX (≈−0.07) models. Similar to the findings from previous comparative studies, we see that NFBP models offered better accuracy results than the other models for three different prediction horizons.

References

Afrasiabi, M., Mohammadi, M., Rastegar, M., Stankovic, L., Afrasiabi, S., Khazaei, M., 2020. Deep-based conditional probability density function forecasting of residential loads. IEEE Trans. Smart Grid 11 (4), 3646–3657.

Agga, A., Abbou, A., Labbadi, M., El Houm, Y., Ali, I.H.O., 2022. CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. Electr. Power Syst. Res. 208, 107908.

Alhussein, M., Aurangzeb, K., Haider, S.I., 2020. Hybrid CNN-LSTM model for short-term individual household load forecasting. IEEE Access 8, 180544–180557.

Almalaq, A., Edwards, G., 2017. A review of deep learning methods applied on load forecasting. In: 2017 16th IEEE International Conference on Machine Learning and Applications. ICMLA, IEEE, pp. 511–516.

Almeida, M.P., Perpinan, O., Narvarte, L., 2015. PV power forecast using a nonparametric PV model. Sol. Energy 115, 354–368.

Alothman, T., Alsaif, S.A., Alfakhri, A., Alfadda, A., 2022. 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, pp. 1–6.

Arshad, M., Qureshi, M., Inam, O., Omer, H., 2021. Transfer learning in deep neural network-based receiver coil sensitivity map estimation. Magn. Reson. Mater. Phys. Biol. Med. 34 (5), 717–728.

Bashir, T., Haoyong, C., Tahir, M.F., Liqiang, Z., 2022. Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. Energy Rep. 8, 1678–1686.

Bessa, R.J., Miranda, V., Botterud, A., Wang, J., Constantinescu, E.M., 2012. Time adaptive conditional kernel density estimation for wind power forecasting. IEEE Trans. Sustain. Energy 3 (4), 660–669.

Beyaztas, U., Shang, H.L., 2022. Robust bootstrap prediction intervals for univariate and multivariate autoregressive time series models. J. Appl. Stat. 49 (5), 1179–1202.

Bikcora, C., Verheijen, L., Weiland, S., 2018. Density forecasting of daily electricity demand with ARMA-GARCH, CAViaR, and CARE econometric models. Sustain. Energy Grids Netw. 13, 148–156.

Bottieau, J., Wang, Y., De Grève, Z., Vallée, F., Toubeau, J.-F., 2022. Interpretable transformer model for capturing regime switching effects of real-time electricity prices. IEEE Trans. Power Syst..

Bremnes, J.B., 2004. Probabilistic wind power forecasts using local quantile regression. Wind Energy 7 (1), 47–54.

Castillo-Botón, C., Casillas-Pérez, D., Casanova-Mateo, C., Ghimire, S., Cerro-Prada, E., Gutierrez, P., Deo, R., Salcedo-Sanz, S., 2022. Machine learning regression and classification methods for fog events prediction. Atmos. Res. 272, 106157.

Chai, S., Niu, M., Xu, Z., Lai, L.L., Wong, K.P., 2016. Nonparametric conditional interval forecasts for PV power generation considering the temporal dependence. In: 2016 IEEE Power and Energy Society General Meeting. PESGM, IEEE, pp. 1–5.

Charles, J.K., Munyoro, N.O., Moses, P.M., Mbuthia, J.M., 2022. Improved expectations-augmented model for short & medium term demand forecasting in Kenya. In: 2022 IEEE PES/IAS PowerAfrica. IEEE, pp. 1–5.

Chaturvedi, S., Rajasekar, E., Natarajan, S., McCullen, N., 2022. A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. Energy Policy 168, 113097.

Chen, C., Duan, S., Cai, T., Liu, B., 2011. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Sol. Energy 85 (11), 2856–2870.

Chen, Y., Wang, Y., Dong, Z., Su, J., Han, Z., Zhou, D., Zhao, Y., Bao, Y., 2021. 2-D regional short-term wind speed forecast based on CNN-LSTM deep learning model. Energy Convers. Manage. 244, 114451.

Chevallier, C., Pinson, P., Kariniotakis, G., 2007. Trading wind generation from short term probabilistic forecasts of wind power. IEEE Trans. Power Syst. 22 (3), 1148–1156.

ChikkaKrishna, N.K., Rachakonda, P., Tallam, T., 2022. Short-term traffic prediction using Fb-PROPHET and neural-PROPHET. In: 2022 IEEE Delhi Section Conference. DELCON, IEEE, pp. 1–4.

Deo, R.C., Ghimire, S., Downs, N.J., Raj, N., 2021. Optimization of windspeed prediction using an artificial neural network compared with a genetic programming model. In: Research Anthology on Multi-Industry Uses of Genetic Programming and Algorithms. IGI Global, pp. 116–147.

Deo, R.C., Grant, R.H., Webb, A., Ghimire, S., Igoe, D.P., Downs, N.J., Al-Musaylh, M.S., Parisi, A.V., Soar, J., 2022. Forecasting solar photosynthetic photon flux density under cloud cover effects: novel predictive model using convolutional neural network integrated with long short-term memory network. In: Stochastic Environmental Research and Risk Assessment. Springer, pp. 1–38.

Deo, R.C., Wen, X., Qi, F., 2016. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Appl. Energy 168, 568–593.

Despotovic, M., Nedic, V., Despotovic, D., Cvetanovic, S., 2015. Review and statistical analysis of different global solar radiation sunshine models. Renew. Sustain. Energy Rev. 52, 1869–1880.

Dewolf, N., Baets, B.D., Waegeman, W., 2022. Valid prediction intervals for regression problems. Artif. Intell. Rev. 1–37.

Dirckx, J.H., 1980. Armand j. Quick: pioneer and prophet of coagulation research. Ann. Intern. Med. 92 (4), 553–558.

Du, B., Huang, S., Guo, J., Tang, H., Wang, L., Zhou, S., 2022. Interval forecasting for urban water demand using PSO optimized KDE distribution and LSTM neural networks. Appl. Soft Comput. 122, 108875.

Espinar, B., Ramírez, L., Drews, A., Beyer, H.G., Zarzalejo, L.F., Polo, J., Martín, L., 2009. Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. Sol. Energy 83 (1), 118–125.

Fan, J., McDonald, J., 1994. A real-time implementation of short-term load forecasting for distribution power systems. IEEE Trans. Power Syst. 9 (2), 988–994.

Farahat, M., 2004. Long-term industrial load forecasting and planning using neural networks technique and fuzzy inference method. In: 39th International Universities Power Engineering Conference, 2004, Vol. 1. UPEC 2004, IEEE, pp. 368–372.

Farzana, S., Liu, M., Baldwin, A., Hossain, M.U., 2014. Multi-model prediction and simulation of residential building energy in urban areas of Chongqing, South West China. Energy Build. 81, 161–169.

Fatema, I., Lei, G., Kong, X., 2023. Probabilistic forecasting of electricity demand incorporating mobility data. Appl. Sci. 13 (11), 6520.

Fu, Y., Virani, N., Wang, H., 2023. Masked multi-step probabilistic forecasting for short-to-mid-term electricity demand. arXiv preprint arXiv:2302.06818.

Gao, B., Huang, X., Shi, J., Tai, Y., Zhang, J., 2020. Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. Renew. Energy 162, 1665–1683.

Ghimire, S., 2019. Predictive modelling of global solar radiation with artificial intelligence approaches using MODIS satellites and atmospheric reanalysis data for Australia (Ph.D. thesis). University of Southern Queensland.

Ghimire, S., Bhandari, B., Casillas-Pérez, D., Deo, R.C., Salcedo-Sanz, S., 2022a. Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia. Eng. Appl. Artif. Intell. 112, 104860.

Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., 2022b. Boosting solar radiation predictions with global climate models, observational predictors and hybrid deep-machine learning algorithms. Appl. Energy 316, 119063.

Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., 2022c. Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Deep Residual model for short-term multi-step solar radiation prediction. Renew. Energy 190, 408–424.

Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., 2023a. Efficient daily electricity demand prediction with hybrid deep-learning multi-algorithm approach. Energy Convers. Manage. 297, 117707.

Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., 2024. Electricity demand error corrections with attention bi-directional neural networks. Energy 291, 129938.

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

Ghimire, S., Deo, R.C., Wang, H., Al-Musaylh, M.S., Casillas-Pérez, D., Salcedo-Sanz, S., 2022e. Stacked LSTM sequence-to-sequence autoencoder with feature selection for daily solar radiation prediction: A review and new modeling results. Energies 15 (3), 1061.

Ghimire, S., Nguyen-Huy, T., AL-Musaylh, M.S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., 2023b. A novel approach based on integration of convolutional neural networks and echo state network for daily electricity demand prediction. Energy 127430. http://dx.doi.org/10.1016/j.energy.2023.127430, URL https://www.sciencedirect.com/science/article/pii/S0360544223008241.

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

Ghimire, S., Nguyen-Huy, T., Prasad, R., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., Bhandari, B., 2022g. Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction. Cogn. Comput. 1–27.

Girshick, R., 2015. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1440–1448.

Grantham, A., Gel, Y.R., Boland, J., 2016. Nonparametric short-term probabilistic forecasting for solar radiation. Sol. Energy 133, 465–475.

Han, Y., Wang, N., Ma, M., Zhou, H., Dai, S., Zhu, H., 2019. A PV power interval forecasting based on seasonal model and nonparametric estimation algorithm. Sol. Energy 184, 515–526.

Haque, A.U., Nehrir, M.H., Mandal, P., 2014. A hybrid intelligent model for deterministic and quantile regression approach for probabilistic wind power forecasting. IEEE Trans. Power Syst. 29 (4), 1663–1672.

Huang, Y., Hasan, N., Deng, C., Bao, Y., 2022. Multivariate empirical mode decomposition based hybrid model for day-ahead peak load forecasting. Energy 239, 122245.

Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: Principles and Practice. OTexts.

Ismail, Z., Jamaluddin, F., Jamaludin, F., 2008. Time series regression model for forecasting Malaysian electricity load demand. Asian J. Math. Stat. 1 (3), 139–149.

Jayasinghe, W.L.P., Deo, R.C., Ghahramani, A., Ghimire, S., Raj, N., 2021. Deep multi-stage reference evapotranspiration forecasting model: Multivariate empirical mode decomposition integrated with the boruta-random forest algorithm. IEEE Access 9, 166695–166708.

Jayasinghe, W.L.P., Deo, R.C., Ghahramani, A., Ghimire, S., Raj, N., 2022. Development and evaluation of hybrid deep learning long short-term memory network model for pan evaporation estimation trained with satellite and ground-based data. J. Hydrol. 607, 127534.

Juban, J., Fugon, L., Kariniotakis, G., 2008. Uncertainty estimation of wind power forecasts: Comparison of probabilistic modelling approaches. In: European Wind Energy Conference & Exhibition. EWEC 2008, EWEC, pp. 10–pages.

Kandil, M., El-Debeiky, S., Hasanien, N., 2001. The implementation of long-term forecasting strategies using a knowledge-based expert system: part-II. Electr. Power Syst. Res. 58 (1), 19–25.

Kang, Y.-J., Noh, Y., Lim, O., et al., 2018. Development of a kernel density estimation with hybrid estimated bounded data. J. Mech. Sci. Technol. 32 (12), 5807–5815.

Kavousi-Fard, A., Khosravi, A., Nahavandi, S., 2015. A new fuzzy-based combined prediction interval for wind power forecasting. IEEE Trans. Power Syst. 31 (1), 18–26.

Khan, S.U., Haq, I.U., Khan, Z.A., Khan, N., Lee, M.Y., Baik, S.W., 2021a. Atrous convolutions and residual GRU based architecture for matching power demand with supply. Sensors 21 (21), 7191.

Khan, N., Haq, I.U., Khan, S.U., Rho, S., Lee, M.Y., Baik, S.W., 2021b. DB-Net: A novel dilated CNN based multi-step forecasting model for power consumption in integrated local energy systems. Int. J. Electr. Power Energy Syst. 133, 107023.

Khan, Z.A., Hussain, T., Ullah, A., Rho, S., Lee, M., Baik, S.W., 2020. Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with a LSTM-AE based framework. Sensors 20 (5), 1399.

Khan, Z.A., Ullah, A., Haq, I.U., Hamdy, M., Maurod, G.M., Muhammad, K., Hijji, M., Baik, S.W., 2022. Efficient short-term electricity load forecasting for effective energy management. Sustain. Energy Technol. Assess. 53, 102337.

Khoa, T., Phuong, L., Binh, P., Lien, N.T., 2004. Application of wavelet and neural network to long-term load forecasting. In: 2004 International Conference on Power System Technology, 2004, Vol. 1. PowerCon 2004, IEEE, pp. 840–844.

Khodayar, M., Kaynak, O., Khodayar, M.E., 2017. Rough deep neural architecture for short-term wind speed forecasting. IEEE Trans. Ind. Inform. 13 (6), 2770–2779.

Khodayar, M., Wang, J., 2018. Spatio-temporal graph deep neural network for short-term wind speed forecasting. IEEE Trans. Sustain. Energy 10 (2), 670–681.

Khorramdel, B., Azizi, M., Safari, N., Chung, C., Mazhari, S., 2018a. A hybrid probabilistic wind power prediction based on an improved decomposition technique and kernel density estimation. In: 2018 IEEE Power & Energy Society General Meeting. PESGM, IEEE, pp. 1–5.

Khorramdel, B., Chung, C., Safari, N., Price, G., 2018b. A fuzzy adaptive probabilistic wind power prediction framework using diffusion kernel density estimators. IEEE Trans. Power Syst. 33 (6), 7109–7121.

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

Komer, B., Bergstra, J., Eliasmith, C., 2019. Hyperopt-sklearn. In: Automated Machine Learning. Springer, Cham, pp. 97–111.

Kong, W., Dong, Z.Y., Hill, D.J., Luo, F., Xu, Y., 2017. Short-term residential load forecasting based on resident behaviour learning. IEEE Trans. Power Syst. 33 (1), 1087–1088.

Lertpalangsunti, N., Chan, C., 1998. An architectural framework for the construction of hybrid intelligent forecasting systems: application for electricity demand prediction. Eng. Appl. Artif. Intell. 11 (4), 549–565.

Li, S., Wang, P., Goel, L., 2015. A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection. IEEE Trans. Power Syst. 31 (3), 1788–1798.

Liao, Z., Huang, J., Cheng, Y., Li, C., Liu, P.X., 2022. A novel decomposition-based ensemble model for short-term load forecasting using hybrid artificial neural networks. Appl. Intell. 1–15.

Liu, Z., Tahir, G.A., Masuyama, N., Kakudi, H.A., Fu, Z., Pasupa, K., 2023. Error-output recurrent multi-layer Kernel Reservoir Network for electricity load time series forecasting. Eng. Appl. Artif. Intell. 117, 105611.

Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

Mamun, M.A., Nagasaka, K., 2004. Artificial neural networks applied to long-term electricity demand forecasting. In: Fourth International Conference on Hybrid Intelligent Systems. HIS'04, IEEE, pp. 204–209.

Marsaglia, G., Tsang, W.W., Wang, J., 2003. Evaluating Kolmogorov's distribution. J. Stat. Softw. 8, 1–4.

Masood, Z., Gantassi, R., Choi, Y., 2022. A multi-step time-series clustering-based Seq2Seq LSTM learning for a single household electricity load forecasting. Energies 15 (7), 2623.

Matrenin, P., Safaraliev, M., Dmitriev, S., Kokin, S., Ghulomzoda, A., Mitrofanov, S., 2022. Medium-term load forecasting in isolated power systems based on ensemble machine learning models. Energy Rep. 8, 612–618.

Mayer, M.J., Biró, B., Szücs, B., Aszódi, A., 2023. Probabilistic modeling of future electricity systems with high renewable energy penetration using machine learning. Appl. Energy 336, 120801.

Mehr, A.D., Nourani, V., Kahya, E., Hrnjica, B., Sattar, A.M., Yaseen, Z.M., 2018. Genetic programming in water resources engineering: A state-of-the-art review. J. Hydrol. 566, 643–667.

Oliphant, T.E., 2006. A Guide to NumPy, vol. 1, Trelgol Publishing USA.

Pai, P.-F., Hong, W.-C., 2005. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. Electr. Power Syst. Res. 74 (3), 417–425.

Pan, C., Tan, J., Feng, D., 2021. Prediction intervals estimation of solar generation based on gated recurrent unit and kernel density estimation. Neurocomputing 453, 552–562.

Permata, R.P., Prastyo, D.D., et al., 2022. Hybrid dynamic harmonic regression with calendar variation for Turkey short-term electricity load forecasting. Procedia Comput. Sci. 197, 25–33.

Raza, M.Q., Mithulananthan, N., Li, J., Lee, K.Y., 2018. Multivariate ensemble forecast framework for demand prediction of anomalous days. IEEE Trans. Sustain. Energy 11 (1), 27–36.

Salcedo-Sanz, S., Casillas-Pérez, D., Del Ser, J., Casanova-Mateo, C., Cuadra, L., Piles, M., Camps-Valls, G., 2022. Persistence in complex systems. Phys. Rep. 957, 1–73.

Salcedo-Sanz, S., Cornejo-Bueno, L., Prieto, L., Paredes, D., García-Herrera, R., 2018. Feature selection in machine learning prediction systems for renewable energy applications. Renew. Sustain. Energy Rev. 90, 728–741.

Shi, H., Xu, M., Li, R., 2017. Deep learning for household load forecasting—A novel pooling deep RNN. IEEE Trans. Smart Grid 9 (5), 5271–5280.

Shohan, M.J.A., Faruque, M.O., Foo, S.Y., 2022. Forecasting of electric load using a hybrid LSTM-Neural prophet model. Energies 15 (6), 2158.

Sulaiman, S., Jeyanthy, P.A., Devaraj, D., Shihabudheen, K., 2022. A novel hybrid short-term electricity forecasting technique for residential loads using Empirical Mode Decomposition and Extreme Learning Machines. Comput. Electr. Eng. 98, 107663.

Taylor, S.J., Letham, B., 2018. Forecasting at scale. Amer. Statist. 72 (1), 37–45.

Trapero, J.R., 2016. Calculation of solar irradiation prediction intervals combining volatility and kernel density estimates. Energy 114, 266–274.

Triebe, O., Hewamalage, H., Pilyugina, P., Laptev, N., Bergmeir, C., Rajagopal, R., 2021. NeuralProphet: Explainable forecasting at scale. arXiv:2111.15397.

Ullah, F.U.M., Ullah, A., Haq, I.U., Rho, S., Baik, S.W., 2019. Short-term prediction of residential power energy consumption via CNN and multi-layer bi-directional LSTM networks. IEEE Access 8, 123369–123380.

Van der Meer, D.W., Shepero, M., Svensson, A., Widén, J., Munkhammar, J., 2018. Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian processes. Appl. Energy 213, 195–207.

Wan, C., Xu, Z., Pinson, P., Dong, Z.Y., Wong, K.P., 2013. Optimal prediction intervals of wind power generation. IEEE Trans. Power Syst. 29 (3), 1166–1174.

Xu, X., Wei, Y., 2022. An ultra-short-term wind speed prediction model using LSTM and CNN. Multimedia Tools Appl. 81 (8), 10819–10837.

Yamazaki, T., Homma, H., Wakao, S., Fujimoto, Y., Hayashi, Y., 2016. Estimation prediction interval of solar irradiance based on just-in-time modeling for photovoltaic output prediction. Electr. Eng. Japan 195 (3), 1–10.

Yamazaki, T., Wakao, S., Fujimoto, Y., Hayashi, Y., 2015. Improvement of prediction interval estimation algorithm with just-in-time modeling for PV system operation. In: 2015 IEEE 42nd Photovoltaic Specialist Conference. PVSC, IEEE, pp. 1–6.

Yang, X., Ma, X., Kang, N., Maihemuti, M., 2018. Probability interval prediction of wind power based on KDE method with rough sets and weighted Markov chain. IEEE Access 6, 51556–51565.

Yu, S., Wei, Y.-M., Guo, H., Ding, L., 2014. Carbon emission coefficient measurement of the coal-to-power energy chain in China. Appl. Energy 114, 290–300.

Zhang, C., Ma, H., Hua, L., Sun, W., Nazir, M.S., Peng, T., 2022. An evolutionary deep learning model based on TVFEMD, improved sine cosine algorithm, CNN and BiLSTM for wind speed prediction. Energy 124250.

Zhang, Y., Wang, J., Wang, X., 2014. Review on probabilistic forecasting of wind power generation. Renew. Sustain. Energy Rev. 32, 255–270.

Zhao, X., Liu, J., Yu, D., Chang, J., 2018. One-day-ahead probabilistic wind speed forecast based on optimized numerical weather prediction data. Energy Convers. Manage. 164, 560–569.

Zhao, J., Mao, X., Chen, L., 2019. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomed. Signal Process. Control 47, 312–323.