論文 / 著書情報
Article / Book Information

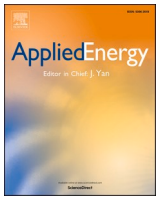
Title	A novel decision support system for enhancing long-term forecast accuracy in virtual power plants using bidirectional long short-term memory networks
Authors	Reza Nadimi, Mika Goto
Citation	Applied Energy, Volume 382, ,
Pub. date	2025, 1
DOI	https://dx.doi.org/10.1016/j.apenergy.2025.125273
Creative Commons	Information is in the article.

Applied Energy 382 (2025) 125273

Contents lists available at ScienceDirect

Applied Energy

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



A novel decision support system for enhancing long-term forecast accuracy in virtual power plants using bidirectional long short-term memory networks

Reza Nadimi *, Mika Goto

Department of Innovation Science, School of Environment and Society, Institute of Science Tokyo, 3-3-6, Shibaura, Minato-ku, Tokyo 108-0023, Japan

HIGHLIGHTS

- Proposed a novel decision support system (DSS) to generate unknown future inputs.
- Integrated a bidirectional long short-term memory (BiLSTM) with the DSS for long-term prediction.
- Incorporated a variable lookback period within the BiLSTM model.

ARTICLE INFO

Keywords:
Virtual power plant
Univariate time series forecasting
Long short-term memory network
Day ahead power market
Decision support system

ABSTRACT

Accurate forecasting of power generation is a serious challenge of virtual power plant (VPP) in day ahead (DA) market because of the volatility and uncertainty of renewables. The recursive prediction technique used in bidirectional long short-term memory (BiLSTM) network often struggles with long-term accuracy. This study proposes a novel decision support system (DSS) to generate unknown future inputs, called "DSS test data", in the recursive prediction technique and tackle the long-term forecasts limitation. The proposed DSS integrates the K-means clustering algorithm and the least squared optimization method. The K-means clustering algorithm classifies historical data into five distinct day types—rainy, overcast, partly cloudy, cloudy, and sunny—based on maximum daily power generation. The DSS employs least squared optimization method to refine the DSS test data for the BiLSTM model, utilizing the most recent seven days of data. Additionally, this study incorporates a variable lookback period within the BiLSTM model to enhance the accuracy of the forecasting model. The DSS-BiLSTM model forecasts VPP power generation 38 h ahead in the Japanese DA power market. Compared to BiLSTM, LSTM, transformer network, attention-based network, gated recurrent unit, and five statistical time series models, the proposed model demonstrates superior accuracy and reduced dispersion in long-term forecasts. The daily mean absolute error for the DSS-BiLSTM, BiLSTM, LSTM, transformer network, attention-based network, and gated recurrent unit models, for a 38-h forecast horizon, are 0.26 GW, 0.48 GW, 0.45 GW, 0.69 GW, 0.66 GW, and 0.62 GW, respectively. This pattern is consistent across the three other error metrics and various forecasting time horizons, indicating that the DSS-BiLSTM model consistently outperforms the other models evaluated in this study in terms of prediction accuracy. The main advantages of the proposed model include ease of implementation, low dispersion, and high forecasting accuracy across various settlement periods, as evidenced by multiple accuracy metrics.

1. Introduction

The Japanese government is actively pursuing "Green Transformation" (GX) to achieve decarbonization, secure a stable energy supply, and foster economic growth [1]. This initiative emphasizes the maximization of decarbonized power sources and investment in GX economy transition bonds. Accurate energy consumption forecasting is crucial to the GX plan, as it ensures energy security, and enhances economic efficiency [2]. Effective forecasting enables proactive responses to fluctuations in energy supply and demand [3].

As part of the GX plan, Japan aims to increase its reliance on

* Corresponding author.
E-mail address: rn.nadimi@gmail.com (R. Nadimi).

https://doi.org/10.1016/j.apenergy.2025.125273
Received 8 December 2024; Accepted 2 January 2025
Available online 13 January 2025

Nomenclature

ACF	Autocorrelation function
Adam	Adaptive moment estimation
AIC	Akaike information criterion
ANN	Artificial neuron network
AR	Autoregressive
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
BIC	Bayesian information criterion
BiLSTM	Bidirectional Long Short-Term Memory
CEEMDAN	Complete ensemble empirical mode decomposition with adaptive noise
CNN	Convolutional neural network
CS algorithm	Cuckoo search algorithm
DA	Day-ahead power market
DERs	Distributed energy resources
DSS	Decision support system
DSS-BiLSTM	DSS-based BiLSTM
EMD	Empirical mode decomposition
EMS	Energy Management System
ESS	Energy storage system
GRU	Gated recurrent unit
ID	Intraday power market
IELM	Incremental extreme learning machine algorithm
IWOA	Improved whale optimization algorithm
JEPX	Japan Electric Power Exchange
KF	Kalman filter
KHM	K-harmonic mean clustering
LMD	Local mean decomposition
LSTM	Long short-term memory
LTF	Load Tracking Feature
MAE	Mean absolute error
MIMO	Multi-input and multi-output
ML	Machine learning
MLP	Multi-layer perceptron
MSE	Mean squared error
O&M	Operation and maintenance
PACF	Partial autocorrelation function
RF	Rrelieff algorithm
RLMD	Robust local mean decomposition
RMSE	Root mean squared error
RNN	Recurrent neural network
SARIMA	Seasonal autoregressive integrated Moving average
SARIMAX	SARIMA + exogenous variables
SCNs	Stochastic configuration networks
SLFN	Single-hidden-layer feedforward network
sMAPE	Symmetric mean absolute percentage error
SOC	Battery state of charge
SVM	Support vector machine
TEPCO	Tokyo electric power company
VPP	Virtual power plant
WRFM	Weather research and forecast mode
WT	Weather type

renewable energy sources, necessitating the forecasting of potential wind and solar power generation [4,5]. Additionally, to improve economic efficiency across the energy sector, utilities in Japan are optimizing resource allocation through various demand response programs [6]. These programs incentivize consumers to adjust their energy usage based on forecasted demand loads, ensuring that energy generation aligns with anticipated consumption [7,8]. One well-established method for predicting power generation and consumption is time series forecasting.

Observations recorded at regular intervals are called a time series. Time series forecasting refers to employ techniques to predict future observation based on previously collected data [9]. This approach is particularly crucial in the fields of energy generation forecasting, power demand prediction, and grid management. The accuracy of time series forecasting significantly influences effective planning, stable operation, and the integration of renewable energy sources [10]. Traditional statistical methods (e.g. AR, ARMA, ARIMA), deep learning approaches (e.g. transformed-based, CNN-based, and MLP-based models), and hybrid models are widely used in time series forecasting [11].

Time series forecasting techniques can be classified in terms of variable, prediction step, model, and forecasting horizon. Similar to the regression analysis [12], a multivariate time series forecasting seeks to identify relationships between multiple variables. In contrast, univariate forecasting aims to figure out dependencies between a single variable and its lagged values. Both univariate or multivariate methods can be applied to predict either single-step or multi-step VPP power generation using statistical and/or ANN models.

Accurate forecasting of VPP power generation increases VPP profitability [13], improves grid stability [14], and maximizes the share of renewable energy sources in the power generation portfolio [15]. However, a large portion of VPP resources belongs to solar and wind power with variability of renewable energy sources, so it is challenging to accurately forecast VPP power generation for the long term [16], especially within the power market context. As shown in Table 1, most studies have predominantly concentrated on short-term prediction horizons. Additionally, forecasting renewables for DA bidding is infrequent due to the challenges associated with requirements on the long-term horizon.

Despite the inherent challenges of renewable energy sources, several approaches exist to improve long-term forecasting of VPP power generation. These include the application of advanced machine learning techniques and hybrid models that integrate various statistical and machine learning methods [17,18,19]. Such approaches enhance prediction accuracy and facilitate more efficient long-term planning, although achieving perfect accuracy remains elusive. The challenge in long-term prediction lies in the reliance on multi-step forecasting approaches, which necessitate the use of a sequence of unknown future input data over a predetermined horizon. The generation of this unknown future input data for each forecasting step represents a complex and intricate problem. The direct and recursive prediction are well-known techniques to generate unknown future input.

The direct prediction technique trains a separate model for each forecasting step while utilizing a constant dataset. Although this approach effectively eliminates cumulative error propagation from previous time points by forecasting each step independently, it overlooks the temporal dependencies between successive outputs (unknown future inputs) [20]. Unlike the direct method, the recursive prediction technique trains a model for different dataset [21,22]. Previous forecasts (unknown future inputs) are recursively added into the head of dataset while same length of added data is dropped down from the tail of dataset. The model's errors accumulate over time since each forecast is based on previous predictions. The error propagation causes the recursive technique becomes sensitive to false forecast data and leads to bias in parameters' estimation [23]. As a result, this technique to be less accurate for long-term forecasts.

To balance error propagation with forecast dependency, the rectify prediction method [24,25] and the multi-input/multi-output prediction method [26,27] have been proposed. The rectify method mitigates recursive forecast bias, reducing mean squared error. In contrast, the multi-input/multi-output approach provides a multivariate estimation

Table 1
Summary of time series forecasting studies in the energy field within recent years.

Model type	Prediction step	horizon	Variable type	Input/output variables list	Purposes	Application field	Ref.
ANN (Attention-based transformer)	Multi-step	Short-term	Univariate	Solar irradiance	Time dependency improvement	Generation (China)	[29]
Attention mechanism and quantile regression	Multi-step	Short-term and Long-term	Univariate	Global horizontal index	Enhancing the prediction model's performance	United States and India	[35]
Transformer networks	Multi-step	Short-term	Multivariate	historical solar power generation, weather data (actual and forecast), solar geometry /Hourly solar power generation	Improving accuracy of solar power forecasting	Generation (South Korea)	[36]
Hybrid model (bidirectional GRU, autoregressive, and attention mechanism)	Single-step	short-term	Multivariate/	1-Wind power generation forecasting, 2-Demand load forecasting	Robust estimation in the presence of outliers	Australia	[37]
BiLSTM, LSTM, and statistics time series models	Single-step and Multi-step	short-term	Univariate	Solar power output	verifying the effectiveness of different models	Generation (Solar farm in China)	[17]
Encoder-decoder LSTM and BiLSTM	Multi-step	Short-term	Multivariate	solar irradiation, temperature/ solar irradiation	Accuracy improvement	Generation (Tokyo)	[18]
Hybrid model (conditional generative adversarial network and BiLSTM)	Single-step	Short-term	Multivariate	global horizontal radiation, temperature, humidity, historical solar power/solar power	Improving prediction accuracy	Generation (Desert Knowledge Australia Solar Centre)	[38]
EMD-GRU-Attention (EMD, GRU, KF)	Multi-step	Short-term	Univariate	Wind speed, temperature, solar radiation, relative humidity, solar zenith angle/ solar radiation	Improving the forecasting performance (best performance for 3 h ahead)	heat load management (Los Angeles, Tianjin, Harare, Santiago)	[39]
Hybrid model*	Multi-step	Short-term	Multivariate	irradiance, wind direction, wind speed, temperature, and relative humidity/PV power	Improvement of the forecasting performance	Market (Three sites in the USA)	[19]
Hybrid model (LTF- RF-MLP)	Single-step	Short-term	Univariate	electrical load	Improving forecasting performance and accuracy	Demand (ISO New England, USA)	[40]
Functional network	Multi-step	short-term	Univariate	wind speed	forecasting accuracy	Generation (Dodge City, KS, USA)	[41]
AR model combined with WRFM	Multi-step	short-term	Univariate	wind speed	Accurately predicting steady wind power than the unsteady wind power	Generation (Fuzhou, China)	[42]
Hybrid (CEEMDAN, RLMD, IWOA, LSTM)	Multi-step	short-term	Univariate	wind speed	Forecasting performance' improvement	Generation (two wind farms in China)	[43]
Hybrid (EMD, LMD, SCNs, SVM)	Multi-step	short-term	Univariate	wind speed	Assisting O&M of wind farms, and improving prediction accuracy	Generation (China)	[44]
Hybrid (EMD, KHM, CS algorithm, SLFN)	Multi-step	Very-short-term	Univariate	wind speed	Improving model performance	Generation (four weather stations in China)	[45]
LSTM and ensemble EMD (MIMO technique)	Multi-step	short-term	Univariate	leakage flow of Nuclear Power Plants Reactor Coolant Pump	minimizing unexpected shutdowns	Generation (Italian MAC4PRO- DS2 projects)	[28]
LSTM-gate recurrent unit (LSTM-GRU)	Multi-step	Long-term	Multivariate	voltage, current, brake pedal depth, weather data, and driving behavior/Battery state-of-charge	Improving SOC prediction accuracy in electric vehicle	Demand (annual operation data of an electric taxi in China)	[46]
Encoder-decoder LSTM	Multi-step	short-term	Multivariate	PV generation and household load forecasting	Efficient EMS architecture and ESS sizing	Generation and demand load consumption (Barcelona, Spain)	[47]

* Combination of improved sparrow search algorithm, fuzzy c-means, grey relation algorithm, improved complete ensemble EMD with adaptive noise, sample entropy, and conditional time-series generative adversarial networks.

of the conditional probability distribution for future values based on past observations. This method addresses dependencies between forecast values, reduces parameter bias, and minimizes error accumulation in long-term predictions [28]. However, both methods face significant challenges in terms of complexity and computational cost [29].

This study proposes a decision support system to generate unknown future inputs in the recursive prediction technique which is called "DSS test data". The paper's originality relies on the K-Means clustering algorithm and the least squared optimization to construct the DSS test data. The DSS comprises classification of weather types (e.g. rainy, overcast, cloudy, partly cloudy, and sunny) based on maximum daily VPP power generation. To address the dependencies between forecast values, the proposed model maintains the most recent seven days data belongs to each weather type. To reduce the parameter bias, the least squared optimization model is used to minimize the difference between actual VPP generation and the average value of the stored last seven days data. The output of the least squared optimization is the adjusting coefficients through which the bias of the prediction parameters is adjusted. Furthermore, in the context of long-term forecasting, this study incorporates a variable lookback period within the LSTM and BiLSTM models. The length of the lookback period is dynamically determined based on the prediction horizon, thereby enhancing the accuracy and adaptability of the forecasting models in electric power market.

Day-ahead is a profitable short-term power market in JEPX for VPP owners [13], but the problem is the accurate prediction of the VPP

power generation. The minimum tradable amount of power in the DA market of JEPX is 0.1 MW for each settlement period [30], while the last moment to submit the bidding data is at 10:00 a.m. in the present day. Generally, a VPP owner should submit a set of 48 bidding data,[1] seven hours earlier than dispatch time. In other words, forecasting model should predict for the next 38 h starting from 10:00 a.m. in the existing day, and lasting to the end of next day. In this case, the forecasting model is a multi-step prediction with 38 h ahead (or 76 settlement periods). While, the accuracy of forecasting models decreases remarkably by keeping distance from the last moment of bid submission. The worst issue is that accurate prediction data mostly occurs in early hours ahead from 10:00 a.m., while the VPP profit is calculated based on prediction data from the first dispatch time onward (00:00–24:00). In other words, the first 14 h of forecasting data do not get involved into profit calculation, while their values affect the desired prediction data (00:00–24:00). Therefore, developing a precise long-term forecasting model is essential for enhancing VPP profitability.

The purpose of this study is to develop a precise long-term forecasting model to enhance VPP power generation forecasting. The proposed model integrates using the BiLSTM and DSS, called DSS-BiLSTM. The DSS is a computer-based system which is designed to aid decision makers in assessing different scenarios [31]. The system relies on input data, knowledge base, inference engine, and models to analyze and provide useful information (output) in cases of uncertainty and the lack of information [32,33]. The proposed model distinguishes itself from previous multi-step forecasting models through its ease of implementation, reduced dispersion, and enhanced accuracy.

The scope of this study is a VPP system with solar and wind power generation in Tokyo, Japan. The VPP accounted for 9.1 % of whole demand load in Tokyo Metropolitan area which was 15.8 GW on average per 30-min settlement period from April 2022 to October 2023 [34].

The paper structures as follows: Section 2 describes the research's methodology including statistical and artificial intelligent models along with the DSS-BiLSTM model. Section 3 presents the results of the methodology section. Section 4 discusses the findings of the current research and unrolls the main results. Section 5 summarizes the main results and future work.

2. Methodology

A critical challenge for participants in the DA power market is to accurately predict power generation over a long period with acceptable stability. As shown in Fig. 1, the DA power market in JEPX requires submitting bid data, which includes forecasting information, at least 14 h prior to the commencement of the first dispatch product and covering a 24-h period. Thus, the total prediction horizon extends to 38 h. In contrast, bids can be adjusted in the ID market through 12 scheduled auctions, which occur between 7 and 1 h before each dispatch product. The gate closure interval is one hour before each dispatch product. Therefore, this study represents the prediction results for three key time horizons: 38 h, 7 h, and 1 h ahead. Additionally, predictions for 30 min and 2 h ahead are provided. The 30-min prediction data is included to facilitate comparison with TEPCO's forecasting data. The 2-h forecast is applied to assess prediction accuracy just one hour before the gate closure.

The initial prediction is the first step for VPP owner in the DA power market. This initial prediction is adjusted by approaching to the gate closure, which occurs one hour before the power dispatch time. Ideally, the VPP owner should predict at least 1.5 h earlier (equivalent to 3 settlement periods) to adjust the initial bid data for a given dispatch period.[2] The significant disparity between the 38-h forecast horizon (76 settlement periods) and the 1.5-h adjustment window can diminish the accuracy of the initial predictions. Conversely, frequent adjustments to the initial submitted bid, which are derived from forecasting models, can increase the risk of incurring balancing costs and negatively impact the VPP's credibility in the DA power market. To address these challenges, this study proposes an accurate time series forecasting model to reduce the frequency and magnitude of adjustments to the initial bid, as illustrated in Fig. 2.

The upper part of Fig. 2 indicates the forecasting module designed to predict VPP generation data for power market applications through training and testing datasets. The main purpose of the training data is to identify or adjust the model's parameters and to validate the fitted model. In contrast, the testing data is applied to evaluate the model's performance, as future data is unavailable.

The accuracy of the forecasting model decreases over longer prediction because of the constant model coefficients over time. In the recursive forecasting model, time series coefficients are estimated through the walk-forward optimization method applied to a rolling window of the dataset, which advances incrementally as time progresses. The walk-forward method facilitates single-step predictions. However, for multi-step forecasting, the process of updating the rolling window is hampered by unavailability of future data.

The lower part of Fig. 2 outlines a method for generating future data to be used as testing data via the DSS module. This module transforms an initial rough weather estimation (weather types) into a set of test data. This process is facilitated by the DSS module's knowledge base and inference engine, which will be further elaborated upon in subsequent section.

2.1. Direct prediction techniques: LSTM and BiLSTM

Generally, ANNs are classified into two fundamental neural networks called CNNs and RNNs. RNNs are particularly effective in dealing with sequential data, such as time series, due to their internal state feature, which functions as a form of memory. The LSTM network, a specialized class of RNNs, is designed to enhance the ability of RNNs to retain information over long periods. The LSTM and BiLSTM networks belong to RNN class, aimed to provide a short-term memory for RNN over long period. The main distinction between LSTM and BiLSTM networks lies in their information processing approach. Standard LSTMs process information in a single direction, utilizing only past data to inform their predictions. In contrast, BiLSTMs process data in both forward and backward directions, allowing them to incorporate both past and future information at each time step [48].

Fig. 3 displays an LSTM unit at time t with the hidden state (h_t) and cell state (c_t). These states are used to maintain values across time intervals, facilitated by three gates: the forget gate, the input gate, and the output gate. Both the hidden state and cell state are represented as vectors of the same size, corresponding to the number of nodes in the LSTM network. The input and output of the LSTM unit are as follows:

$$\{X_i^{Inp}\} = (x_{t-n+i+1}, \ldots, x_{t+i}) = \begin{bmatrix} x_{t-n+1} & x_{t-n+2} & \cdots & x_{t-n+r} \\ \vdots & \vdots & \vdots & \vdots \\ x_t & x_{t+1} & \cdots & x_{t+r-1} \end{bmatrix} \quad (1)$$

$$\{X_i^{Out}\} = (x_{t+1+i}, \ldots, x_{t+m+i}) = \begin{bmatrix} x_{t+1} & x_{t+2} & \cdots & x_{t+r} \\ \vdots & \vdots & \vdots & \vdots \\ x_{t+m} & x_{t+m+1} & \cdots & x_{t+m+r-1} \end{bmatrix} \quad (2)$$

here, n and m indicate the lookback period and the number of prediction units ahead, respectively. The term r, in conjunction with n and m, specifies the period or day for which the prediction is being planned. The term i ranges from zero to (r-1).

The mathematical relationships between the input data, hidden state, cell state, and output data in the LSTM are as follows:

[1] VPP power generation for each 30-min from 00:00 to 24:00.

[2] Because it is mostly difficult to buy/sell power in 30-min before gate closure.

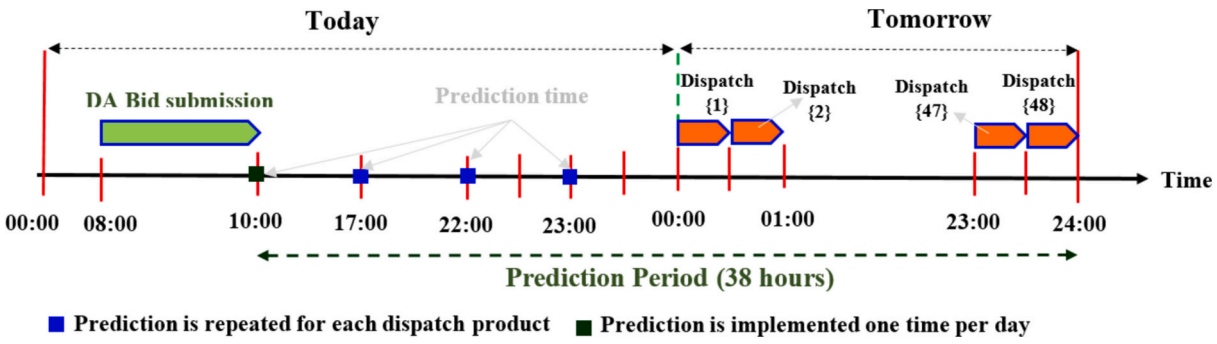


Fig. 1. Bid prediction and dispatch time for DA market in JEPX.

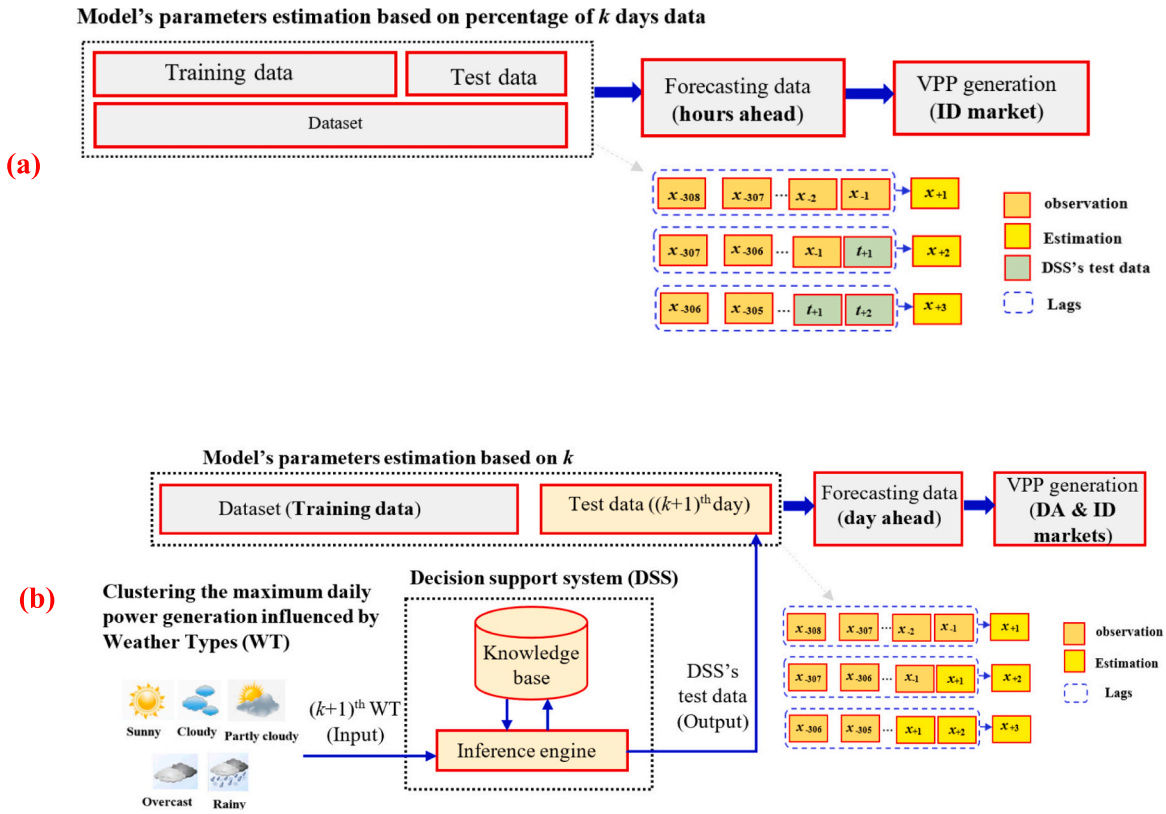


Fig. 2. Conventional time series forecasting model (a) vs the proposed DSS-BiLSTM model (b).

$$f_t = \sigma\left(W_h^f h_{t-1} + W_{XIn}^f X_t^{Inp}\right) \quad (3)$$

$$i_t = \sigma\left(W_h^i h_{t-1} + W_{XIn}^i X_t^{Inp}\right) \quad (4)$$

$$o_t = \sigma\left(W_h^o h_{t-1} + W_X^o X_t^{Inp}\right) \quad (5)$$

$$\tilde{c}_t = tanh\left(W_h^{\tilde{c}} h_{t-1} + W_{XIn}^{\tilde{c}} X_t^{Inp}\right) \quad (6)$$

$$c_t = \sigma\left(f_t \odot c_{t-1} + i_t \odot \tilde{c}_t\right) \quad (7)$$

$$h_t = tanh(c_t) \odot o_t \quad (8)$$

$$X_t^{Out} = softmax\left(W_h^{Xout} h_t\right) \quad (9)$$

where W_{XInp}^f represents the weight matrix of the current input data for the forget gate. The rest of the weights are interpreted in the same way.

Three activation functions are defined based on the following equations:

$$\sigma(inp) = {}^{1}/_{\left(1 + e^{-inp}\right)} \quad (10)$$

$$tanh(inp) = 1 - e^{-2inp}/\left(1 + e^{-2inp}\right) \quad (11)$$

$$softmax(inp_i) = e^{inp_i} \Big/ \sum_{j=1}^{n} e^{inp_j} \quad (12)$$

If an estimation model obtains a negative VPP power generation for a specific settlement period, the negative value will be replaced with the average value of the last seven days[3] for the corresponding settlement period. The BiLSTM module applies a single-step forecasting approach using a recursive technique to predict VPP power generation data. This technique utilizes the last 308 settlement period, encompassing the most

[3] Mostly it occurs in [00:00–7:00] and [18:00–24:00]

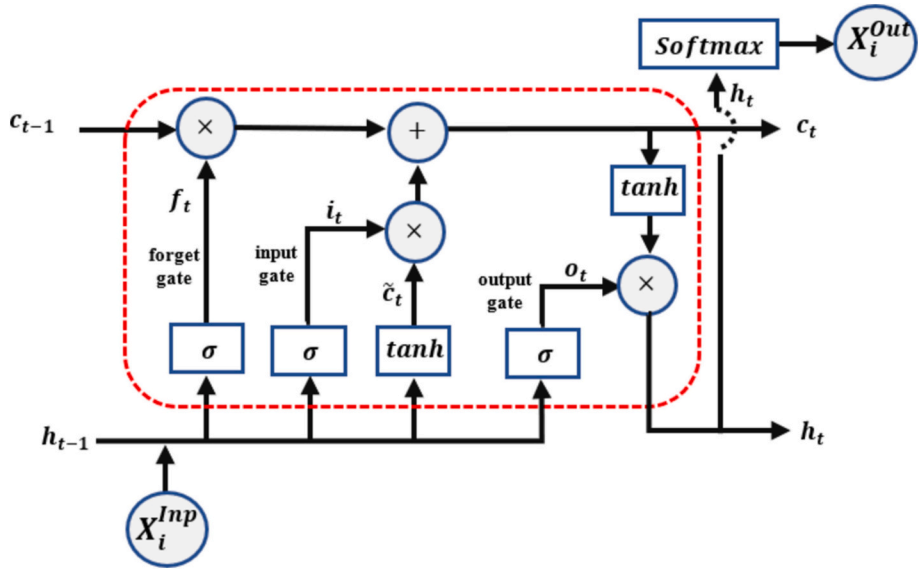


Fig. 3. Single LSTM unit at time *t*.

recent 6 days plus the current day up to 10:00 AM (the deadline for day-ahead bidding).

2.2. Recursive prediction technique: statistical models and proposed model

The study applies a sliding time window for time series analysis, focusing on the most recent 308 settlement periods, which approximately corresponds to a seven-day period. This approach is utilized for both the statistical models and the proposed model to ensure a consistent and relevant dataset for analysis.

2.2.1. Statistical models

This study employs the following statistical time series models [49] to forecast VPP power generation data: AR (p), ARMA (p, q), ARIMA (p, d, q), SARIMA (p, d, q) (P, D, Q)$_m$, and SARIMAX (p, d, q) (P, D, Q)$_m$. The parameters (p, d, q) and (P, D, Q)$_m$ represent non-seasonal and seasonal parts of the statistical model, respectively. The terms p and q imply the number of lags, and the size of the moving average window, respectively. The term d indicates the differencing step to eliminate the non-stationary in the observed data. The term m specifies the periods' number in each season, while the uppercase P, D, and Q have the same meaning as p, d, q just for the seasonal case.

The optimal model is selected based on the parameters q and p, which are determined through visual inspection of the ACF and PACF. However, extracting the statistical parameters based on ACF and PACF is a time intensive procedure. Akaike and Bayesian information criteria (AIC and BIC) are two popular methods to obtain the best model among a set of candidate models for a given data [50]. Both methods rely on prediction error's estimator and model's information lost in choosing the best model.

$$AIC_{\xi} = 2K - 2ln\left(\mathscr{L}\left(\widehat{\theta}|X_i^{Inp}\right)\right) \tag{13}$$

$$BIC_{\xi} = Klog(n) - 2ln\left(\mathscr{L}\left(\widehat{\theta}|X_i^{Inp}\right)\right) \tag{14}$$

where $\mathscr{L}(*)$ term indicates the likelihood function over the parameter $\theta = (\theta_1, \ldots, \theta_v)$ in the ξ^{th} model for given data X_i^{Inp}. The best model has the lowest AIC (BIC) value. This study chooses a model with minimum AIC and BIC values.

If time series data, $\left\{X_i^{Inp}\right\}$ shows non-stationary behavior then the first order ($\left\{X_i^{Inp} - X_{i-1}^{Inp}\right\}$ or $d = 1$) or second order ($\left\{X_i^{Inp} - X_{i-1}^{Inp}\right\}$-$\left\{X_{i-1}^{Inp} - X_{i-2}^{Inp}\right\}$ or $d = 2$) differencing is carried to establish stationary data. The average and variance of a stationary time series do not vary over time. Augmented Dickey-Fuller test [51] is applied to test if a time series process is stationary. The VPP power generation series in this study is a stationary process, thus its differencing step is zero. Eq. (15) represents the SARIMAX (p, 0, q) (P, 0, Q)$_s$ model for time series $X_{i=1}^{Inp}$ by which the rest of statistical models mentioned in this research are drawn out.

$$\begin{aligned} x_t = c + \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{o=1}^{P} \Phi_o \left[x_{t-os} + \sum_{i=1}^{p} \varphi_i x_{t-(os+i)} \right] \\ + \sum_{l=1}^{Q} \psi_l \left[\varepsilon_{t-ls} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-(ls+j)} \right] + \sum_{u=1}^{r} \delta_u E x_{u_t} + \varepsilon_t \end{aligned} \tag{15}$$

where ε_t is an independent normal random variable with mean zero and variance σ^2. The red term in Eq. (15) models the exogenous factors, and the term c indicates a constant value.

2.2.2. DSS-BiLSTM model

The DSS-BiLSTM model combines the BiLSTM model with the proposed decision support system. The key aspect of the DSS-BiLSTM model is its use of the BiLSTM model within a recursive technique as detailed by the training data specified in Eq. (1). The DSS module is constructed based on extensive time series data, ensuring that its knowledge base consists of five WTs: sunny, partly cloudy, cloudy, overcast, and rainy days. The following steps are carried out to establish the proposed DSS module:

Step1) Calculating of maximum VPP power generation:

$$y_i = max(x_1, \ldots, x_{48}), i = 1, 2, \ldots, m, \tag{16}$$

where y_i represents the maximum power generation for each day, derived from the settlement period data, $x_1, \ldots, x_{48}$. This maximum daily power generation is used as a feature for subsequent clustering.

Step2) Clustering of maximum daily power generation data: The y_i

data is clustered into the five WTs by the K-means clustering[4] method [52]. The boundaries of these clusters are determined and stored. The variance ratio criterion was used to choose the number of clusters [53]. Notably, the flexibility of the proposed model is enhanced through frequent updates to the DSS knowledge base. In Step 1, the DSS knowledge base updates its cluster data at the end of each day. However, by reducing this update interval to hourly or every 30 min, while computational time may increase, the model's flexibility is significantly improved.

Step3) Adjusting of coefficients for each WT type: The clustering results from "Step2" are used to find the adjustment coefficients, $\alpha_{t,WT}$, based on the following least squares method:

$$minerror_t^{WT} = \sum_{l=1}^{WT} \left(x_{t,WT}^{l} - \alpha_{t,WT} \times \overline{x}_{t,WT}^{l} \right)^2 ; WT = \{Rainy, \ldots, Sunny\} \tag{17}$$

$$\overline{x}_{t,WT}^{l} = \sum_{k=l-7}^{l-1} \frac{x_{t,WT}^{k}}{7} \tag{18}$$

where $x_{t,WT}^{l}$ indicates the actual VPP power generation at settlement period t on the l^{th} day corresponding to the WT type. The adjustment coefficients are represented as a matrix with dimension corresponding to the number of WT types as rows and 48 columns, reflecting the 48 settlement periods within a day.

Step4) Creating of DSS test data: The DSS test data are generated based on the results from previous steps through the following phases:

Phase1) Creation of average WT curve for the next day: for each WT, the knowledge base stores data from the last seven days. The average value for each settlement period is calculated as follows:

$$\overline{x}_{t,WT}^{k+1} = \sum_{j \in last\ 7\ WT\ day} x_t^j/7, t = [1,2,\ldots,48]; k = \{7,8,\ldots\} \tag{19}$$

For example, the knowledge base will store data from the last 336 (7 × 48) settlement periods for sunny day. The average over the last seven sunny days provides a curve with 48 settlement period values for sunny days.

Phase2) Creation of average WT curve for the current day: The squared Euclidean distance for each WT, denoted as d_t^{WT}, is applied to find minimum distance between the current day's data (all data before 10:00 a.m. or 19 data) and corresponding data from the last seven days as follows:

$$d_t^{WT} = \sum_{t=1}^{20} \left(x_t - \overline{x}_{t,WT}^{l} \right)^2 ; WT = \{Rainy, \ldots, Sunny\} \tag{20}$$

For example, if d_t^{WT}indicates the minimum distance, the current day is categorized as sunny:

$$\overline{x}_{t,Sunny}^{k} = \sum_{j \in last\ 7\ Sunny\ day} x_t^j/7, t = [21,22,\ldots,48] \tag{21}$$

Phase3) Creation of DSS test data: The unadjusted DSS test data consist of a vector with 76 elements, obtained from the previous phases, which may correspond to different or the same WTs. The following formula is used create the DSS test data for the k^{th} and $(k+1)^{\text{th}}$ days for a given WT:

$$DSS = \left[\left\{ \overline{x}_{t,WT}^{k} \times \alpha_{t,WT}; t = 21,22,\ldots,48 \right\}, \left\{ \overline{x}_{t,WT}^{k+1} \times \alpha_{t,WT}; t = 1,2,\ldots,48 \right\} \right] \tag{22}$$

[4] This study explored Bisecting K-Means, Gaussian Mixture Models, and Hierarchical clustering using three linkage criteria: Ward, Complete, and Average. However, none of these methods yielded better results than the K-Means clustering method.

It is clear that the adjusting coefficients depend on the WT.

Phase4) Prediction of VPP power generation via the BiLSTM model: The DSS test data, consists of a 76-element vector adjusted by the coefficients, are used for prediction. The BiLSTM model, trained with 308-element training data using a single-step time series forecasting approach (walk-forward approach), is applied to predict VPP power generation for the DSS test data.

Step5) Updating the last seven data after receiving the actual data: At the end of next day, the actual data of the next day as well as 14 h data of the previous day is available. Both days data are categorized by their daily maximum values and cluster boundaries.

Steps 1, 4, and 5 are implemented daily to update the WT types and to adjust coefficients. Updating the clustering data and adjusting coefficients depend on a noticeable shook in climate change which changes the VPP power generation. This study utilized data from the past two years for clustering and maintained the boundary data for 81 days.

2.3. Measuring of prediction accuracy

Various properties are compared to choose a proper prediction method such as scale invariance of data, accuracy, sensitivity, consistency, symmetry, robustness, and interpretability [54]. This study applies several well-known measures to quantify the forecasting accuracy such as sMAPE, MAE, MSE, and RMSE which their equations are given in the below.

$$sMAPE = \frac{100}{N} \times \sum_{i=1}^{N} e_i; e_i = \left| VPP_i^A - VPP_i^F \right| / \left(\left| VPP_i^A \right| + \left| VPP_i^F \right| \right) / 2 \tag{23}$$

$$MAE = \frac{\sum_{i=1}^{N} e_i}{N}; e_i = \left| VPP_i^A - VPP_i^F \right| \tag{24}$$

$$MSE = \frac{\sum_{i=1}^{N} e_i}{N}; e_i = \left(VPP_i^A - VPP_i^F \right)^2 \tag{25}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} e_i}{N}}; e_i = \left(VPP_i^A - VPP_i^F \right)^2 \tag{26}$$

where VPP_i^A and VPP_i^F indicate actual and forecasted VPP power generation at i^{th} settlement period, respectively.

The Diebold-Mariano test [55] is conducted to statistically compare the forecasting accuracy of two prediction models. The test evaluates if Model 1 (M1) provides a significantly better forecast than Model 2 (M2) by testing the following hypothesis:

$$\text{Diebold}-\text{Mariano test} \begin{cases} Null\ hypothesis: E\left(e_i^{M1} - e_i^{M2}\right) \geq 0 & i = 1,\ldots,N \\ Alternative\ hypothesis: E\left(e_i^{M1} - e_i^{M2}\right) < 0 & i = 1,\ldots,N \end{cases} \tag{27}$$

This study calculates the p-value based on the statistical normal distribution and compares it against a significance level of 5 %. To implement the DM test, the DM test statistic is computed as:

$$DM = \frac{\overline{d}}{\sqrt{\widehat{\sigma}_{\overline{d}}/N}} \tag{28}$$

where $\overline{d}$ measures the average value of the difference between the error values of the two models, defined as $d_i = \left(e_i^{M1} - e_i^{M2}\right)$. Additionally, $\widehat{\sigma}_{\overline{d}}$ is the estimated variance of the loss differentials which is defined as:

$$\widehat{\sigma}_{\overline{d}} = \gamma_0 + 2\sum_{k=1}^{j-1}\gamma_k; \gamma_k = \frac{1}{N}\sum_{i=k+1}^{N}(d_i - \overline{d})(d_{i-k} - \overline{d}) \quad (29)$$

where γ_k denotes the autocovariance at lag k. The dispersion of prediction errors for a specific model across various forecasting horizons is quantified using the coefficient of variation, which is calculated as follows:

$$CV = \frac{Standard\ deviation\ of\ error\ data}{Mean\ of\ error\ data} \times 100 \quad (30)$$

here, the error data are evaluated across five distinct forecasting horizons for both the prediction method and the accuracy measure. For instance, in the case of the sMAPE indicator applied to the AR prediction model, the error data include measurements at 30 min, 1 h, 2 h, 7 h, and 38 h.

2.4. Data acquisition and models' parameters

All publicly available VPP power generation data were downloaded from the TEPCO website [34], encompassing both forecasted and actual VPP generation data, as illustrated in Fig. 4. TEPCO provides single-step forecast data for 30-min intervals. This study is used the actual data for prediction. In addition to multi-step prediction, this study conducts a single-step prediction to compare the results of the proposed model with TEPCO single-step forecasts. This study downloaded TEPCO data from 04/01/2022 to 03/21/2024 with a resolution of 30 min. The VPP power generation sources include wind and solar power. The average daily power generation over the two-year period is 1.5 GW, with minimum and maximum values of 0.037 GW and 4.92 GW per settlement period, respectively. Data from 04/01/2022 to 12/31/2023 were used for clustering, while the period from 01/01/2023 to 12/31/2024 was utilized for model training in the BiLSTM and LSTM models. The period from 01/01/2024 to 03/21/2024 was applied for model prediction.

Table 2 summarizes the parameters and configurations for the LSTM, BiLSTM and statistical models. Additionally, this study evaluates the prediction results for the transformer network,[5] attention-based network,[6] and GRU.[7] It is noteworthy that the number of units in the attention-based network was initially set according to Table 2; however, the results were unsatisfactory. As a result, the study increased the number of LSTM units in the attention-based model to 100 in order to achieve more accurate prediction results. Additionally, the number of epochs for all three models was set to 500, consistent with the LSTM and BiLSTM models.

For all prediction horizons, identical parameters were utilized, with the only variation being the number of lags. Various trials and error analyses were conducted to determine an appropriate number of lags for each horizon, based on the sMAPE measure. Ultimately, the number of lags was set to $2 \times m$ to balance the trade-off between computational time and model accuracy. This setting is consistent with the principle that the extent of interaction coverage[8] is directly proportional to the size of the look-back window [56]. A larger look-back window allows for the incorporation of a broader range of historical data, which enables the capture of more relevant interactions and dependencies within the time series. Conversely, maintaining a fixed value for the look-back window while increasing the prediction horizon restricts the modeling of variations across prediction horizons to a limited and static set of historical data points.

The configuration of the DSS-BiLSTM model mirrors that of the LSTM and BiLSTM models given in Table 2. These configurations were determined through extensive trials and error analysis, as shown in Fig. 5, which examines the hidden nodes extracted from [17]. The DSS-BiLSTM model employs data from the last seven days for training and utilizes a recursive prediction technique for forecasting.

3. Results

This section presents the forecasting results for eight models across various prediction horizons, including 30 min, 1 h, 2 h, 7 h, and 38 h ahead. The performance of these models is evaluated using four specified prediction metrics All models were implemented using Python (version 3.12), and computations were executed on a computer equipped with a 12-core central processing unit and 128 gigabyte of random access memory (Table 4 summarizes the computational time for each model).

3.1. Time series forecasting performance for 30-min horizon

Table 3 presents the forecasting performance for 30-min intervals, evaluated using four prediction metrics. The prediction period spans from January 1, 2024, to March 21, 2024. The row immediately following the table heading shows the accuracy of TEPCO forecasts compared to the actual VPP power generation data. The subsequent rows in Table 3 compare the forecasting results of various models against the actual VPP power generation data.

According to Table 3, the DSS-BiLSTM model generally outperforms all other models across the four metrics of prediction, including the TEPCO forecast data. Although the sMAPE value for the proposed model is slightly higher than that of the TEPCO predictions, the other prediction metrics for the DSS-BiLSTM model are significantly lower. It is important to evaluate model performance using multiple metrics, as the sMAPE may remain low for statistical models with longer time horizons, while other metrics, particularly MSE), may increase substantially.

3.2. Time series forecasting performance for all horizons

Fig. 6 depicts the performance of all eight models in terms of sMAPE, MAE, MSE, and RMSE indicators, calculated for 81 days (1st January 2024 to 21st March 2024). The selection of an appropriate metric depends on the characteristics of the data and the priorities of decision-makers. In this study, sMAPE, as a relative metric, exhibited superior performance for VPP power generation data, which spans a wide range of values, with minimum and maximum values of 5000 kWh and 8,229,000 kWh, respectively.

According to the sMAPE indicator, the DSS-BiLSTM model achieved the best results for all time steps. Furthermore, the proposed DSS-BiLSTM model overcomes the limitations associated with long-term predictions (for a prediction horizon of 38 h), where traditional LSTM and BiLSTM models failed to generate prediction for unknown future input data at each forecasting horizon. For the 38-h prediction horizon, the DSS-BiLSTM model exhibited the lowest error values across all assessed metrics.

However, the DSS-BiLSTM model demonstrated slightly higher error values for 1-h and 2-h forecasts when evaluated using MAE, MSE, and RMSE metrics. This discrepancy can be attributed to two primary factors:

1- The DSS test data is generated once daily, based on the maximum daily VPP power generation. Consequently, the DSS-BiLSTM model

[5] This study specifies the following hyperparameters for the transformer network in the PyTorch package (version 2.5.1): Input dimension = Output dimension = 1, Sequence length = 48, Number of layers = 2, Number of heads = 2, and Feedforward dimension = 10.

[6] This study sets the following values for the hyperparameters of the attention-based network in Python: LSTM layer with 100 units, dimension of the input data 48 × 1, and activation layer = 'Softmax'.

[7] This study defines the following hyperparameters for the GRU in the TensorFlow library (version 2.17.0) of Python: Input dimension = Output dimension = 1, Sequence length = 48, Number of layers = 2, Number of heads = 2, and dimension feedforward = 10.

[8] Between two look-back datasets

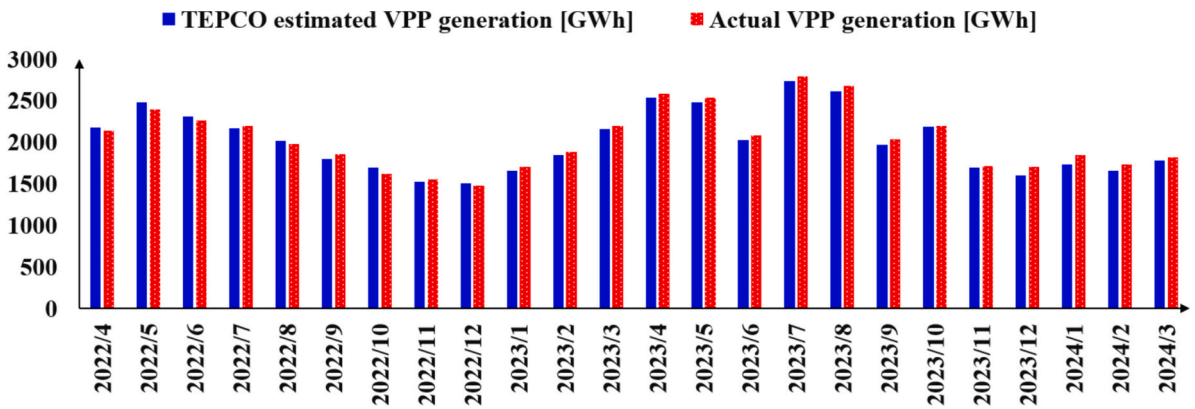


Fig. 4. Total monthly VPP power generation.

Table 2
Hyperparameters setting and configuration for statistical and ANN models.

LSTM & BiLSTM	Statistic models
1) Number of neurons (hidden nodes): 5, 2) Number of epochs: 500, 3) Batch size: 48 (The value of 48 corresponds to a total of 48 settlement periods), 4) Optimizer: Adam with learning rate of 0.001 (default value), 5) Direct prediction technique, 6) Training data: all data in 2023 with 30-min resolution, 7) Number of layers: 3 (LSTM), and 2 (BiLSTM)	1) Model selection based on AIC and BIC values. 2) Recursive multi-step prediction 3) Training data: last seven days

uses this static test data to predict VPP power generation across all 48 settlement periods. As a result, the only source of variability in the forecasts is the unsystematic (random) error [57], which is the difference between the actual historical data and the DSS test data. This leads to small variation in forecast data across different time horizons. To enhance prediction accuracy, it would be beneficial to update the DSS test data multiple times throughout the day.

2- The LSTM and BiLSTM models tend to outperform the DSS-BiLSTM model in short-term forecasts (e.g., 1-h and 2-h intervals) because they jointly model both systematic (biased) and unsystematic errors. This integrated approach, while advantageous for short-term

Table 3
30-min ahead forecasting accuracy measures for 81 days.

Prediction Model	sMAPE [%]	MAE [GW]	MSE [Unitless × 10^{12}]	RMSE [GW]
TEPCO	21	0.12	0.05	0.23
AR	40	0.06	0.01	0.11
ARMA	35	0.07	0.03	0.16
ARIMA	36	0.06	0.01	0.11
SARIMA	21	0.06	0.03	0.16
SARIMAX	22	0.06	0.03	0.16
LSTM	30	0.21	0.15	0.38
BiLSTM	30	0.22	0.15	0.39
DSS-BiLSTM	22	0.01	0.00	0.02

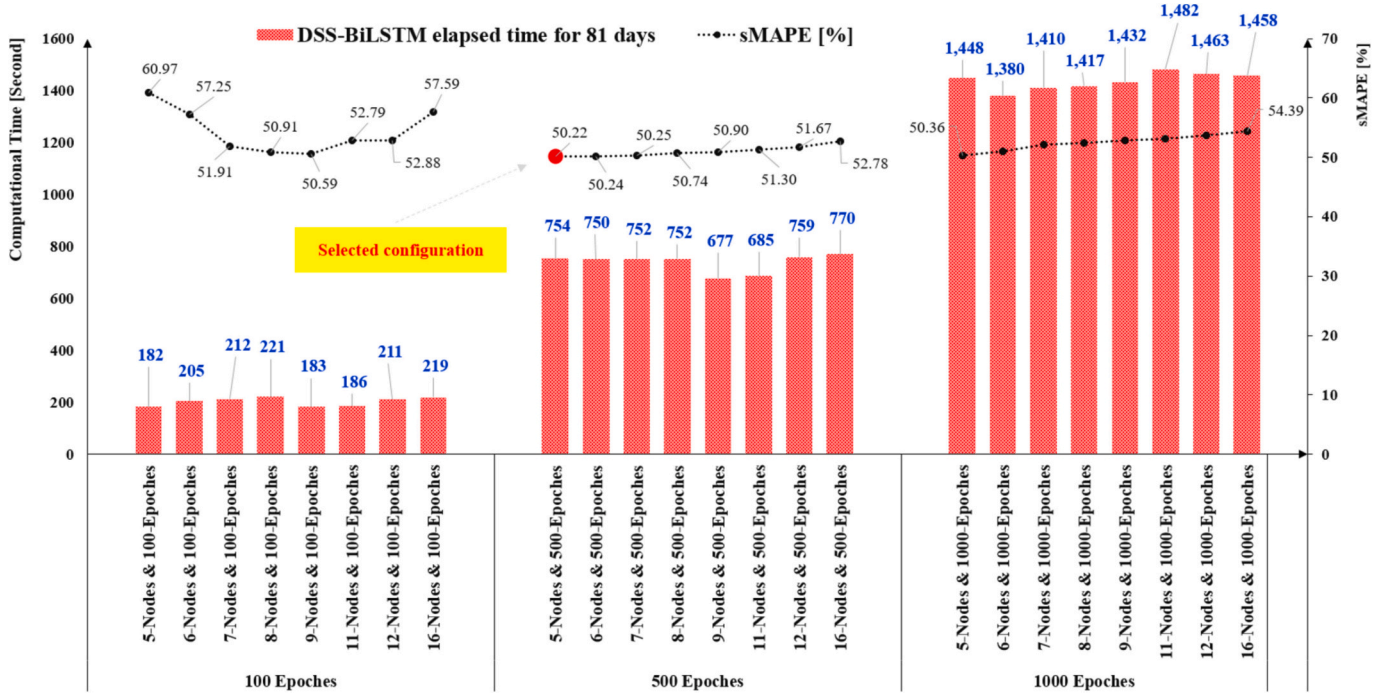


Fig. 5. Hyper parameters configuration of DSS-BiLSTM model for 38-h ahead prediction.

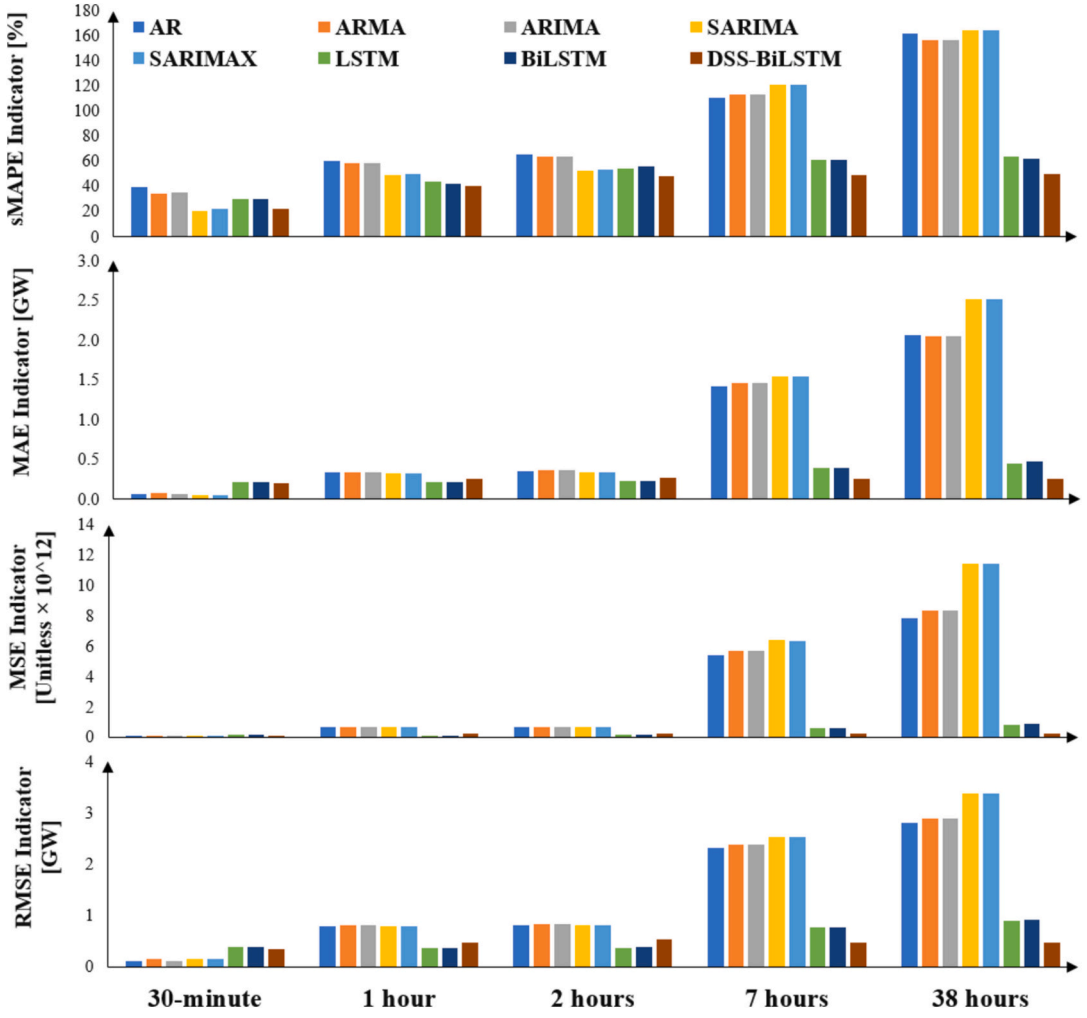


Fig. 6. Model performance based on four prediction measures.

Table 4
Correlation coefficient and computational time.

Algorithm	correlation coefficient [%]				average computational time [second]			
	1-h	2 h	7 h	38 h	1-h	2 h	7 h	38 h
AR	94 %	93 %	18 %	−44 %	0.75	0.37	0.11	0.08
ARMA	94 %	93 %	13 %	−35 %	0.75	0.37	0.11	0.08
ARIMA	94 %	93 %	13 %	−35 %	5.88	2.97	0.72	0.13
SARIMA	94 %	93 %		−38 %	4.81	2.42	0.64	0.12
SARIMAX	94 %	93 %	3 %	−39 %	4.81	2.42	0.64	0.12
LSTM	99 %	98 %	94 %	91 %	11.24	9.17	12.68	16.99
BiLSTM	99 %	98 %	93 %	90 %	12.54	12.86	12.7	16.36
DSS-BiLSTM	98 %	97 %	98 %	98 %	8.71	8.78	8.64	8.84

predictions, can result in higher error values for longer time horizons (7 h and 38 h) when using the LSTM and BiLSTM models.

Table 4 presents the results of the correlation analysis between actual VPP data and forecasted data. The analysis reveals that the forecast data generated by the DSS-BiLSTM model exhibits a consistently high correlation with the actual data across various prediction horizons. In contrast, the correlation magnitude for the other ANN models decreases notably from approximately 99 % in the 1-h ahead prediction to around 90 % in the 38-h ahead prediction. The correlation magnitude reaches to negative values in the case of statistical prediction data. This trend underscores the limitations of statistical models in achieving accurate long-term predictions, despite their relatively low computational times.

It is generally anticipated that the computational time of the DSS-BiLSTM model would be greater than or equal to that of the standard BiLSTM model. However, the DSS-BiLSTM model demonstrates a lower computational cost. This reduction in computational expense can be attributed to its implementation of a recursive approach, which utilizes single-step predictions based on test data generated by the DSS system. In contrast, LSTM and BiLSTM models perform multi-step predictions, resulting in increased computational time as the prediction horizon lengthens. This highlights a key advantage of the DSS-BiLSTM model in managing computational resources effectively while providing robust long-term forecasts.

A comparative analysis of various ANN models was conducted in [17], which concluded that the LSTM and BiLSTM models achieved optimal forecasting performance for a 20 MW grid-connected photovoltaic station in China. The primary distinctions between the models discussed in [17] and those evaluated in this study are the number of neurons, layers, and the lookback period. Specifically, the LSTM and BiLSTM models in [17] were configured with 11 and 7&8 neurons,

respectively, and utilized 1 and 2 layers. The lookback periods employed in [17] were 4 and 8 time steps, whereas this research adopts a lookback period of assumes $2 \times m$ where the amount of m is determined in Eq. (2). In this study, the LSTM and BiLSTM models proposed in [17], referred to as LSTM_TX and BiLSTM_TX, were implemented for VPP data with lookback period of 8 steps.

Fig. 7 shows the performance of ANN models in terms of four error metrics. The results indicate that a fixed lookback period (e.g. 8) correlates with an increase in prediction error as the forecast horizon extends. In contrast, the lookback period varies in this study and avoids of error increment. Furthermore, the proposed DSS-BiLSTM model demonstrates superior prediction accuracy relative to both the BiLSTM_TX and LSTM_TX models. This suggests that the DSS-BiLSTM model is more effective at mitigating error across varying prediction horizons. Furthermore, Fig. 7 shows that for short-term forecasts (e.g., one hour), the predictions from all models are relatively similar. According to the sMAPE metric, the attention-based model outperforms the LSTM model for one-hour predictions, while the other error metrics do not exhibit a consistent trend. However, as the prediction horizon increases (from one hour to 38 h), the DSS-BiLSTM model consistently outperforms the other models.

Table 5 presents the dispersion of accuracy error metrics for each prediction model, illustrating how the error metrics vary with an increasing prediction time horizon. The DSS-BiLSTM model is distinguished by a significantly lower rate of escalation in prediction error compared to the other models evaluated. This observation suggests that the DSS-BiLSTM model demonstrates superior stability and consistency in forecasting performance over extended time horizons.

Table 5
Percentage of dispersion for prediction models.

Prediction Model	sMAPE	MAE	MSE	RMSE
AR	50 %	90 %	108 %	74 %
ARMA	51 %	89 %	108 %	73 %
ARIMA	51 %	90 %	108 %	75 %
SARIMA	65 %	98 %	116 %	79 %
SARIMAX	64 %	98 %	116 %	79 %
LSTM	25 %	33 %	77 %	41 %
BiLSTM	24 %	36 %	79 %	42 %
DSS-BiLSTM	25 %	10 %	51 %	13 %

To ensure clarity in the visual representation of results, the study selects a subset of models characterized by lower sMAPE values. For these models, forecasting values are plotted against actual values for the first week of January 2024, with predictions made at 38-h forecast horizon as shown in Fig. 8.

The study chooses the SARIMA model deliberately because its

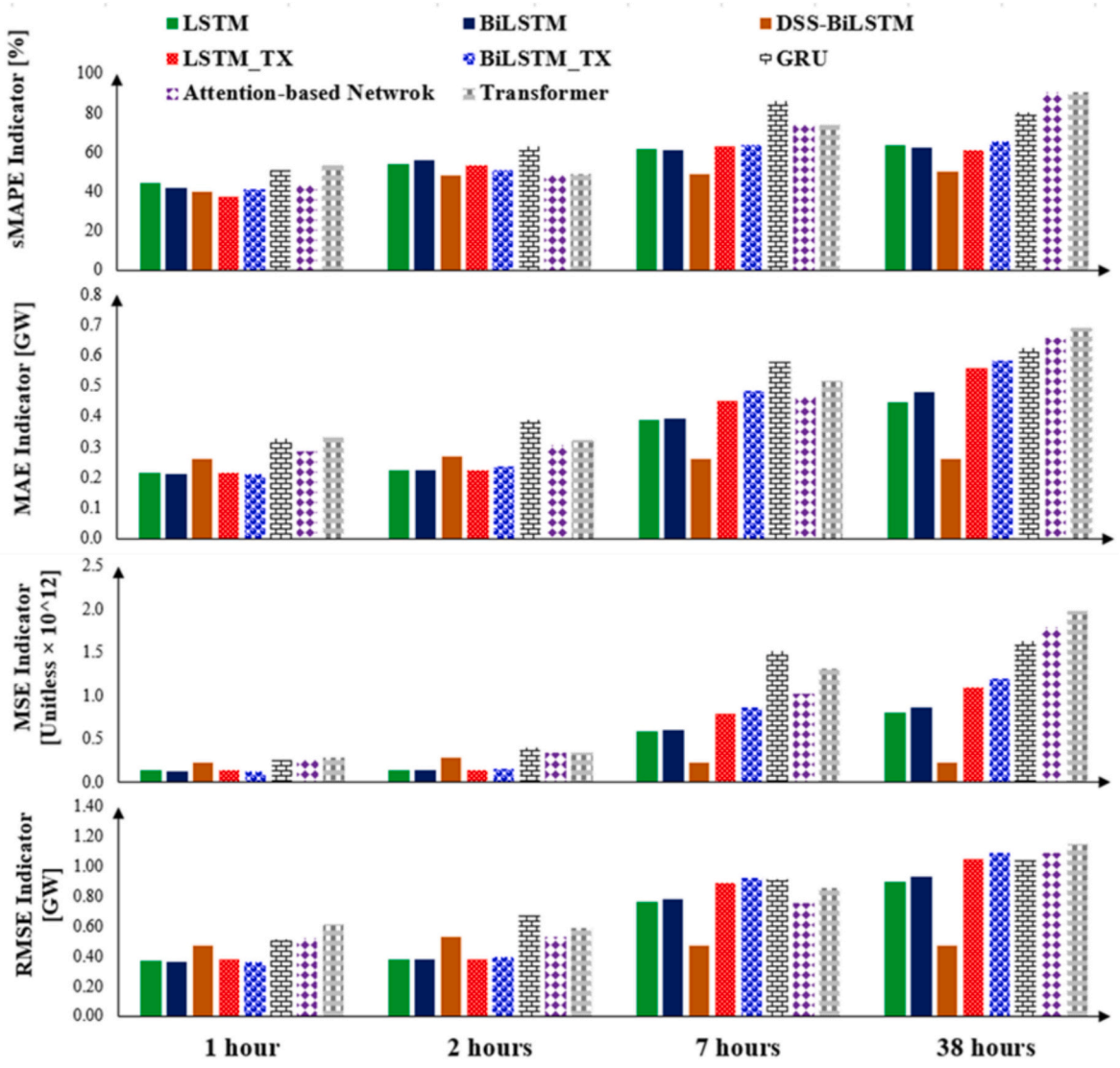


Fig. 7. ANN model performance based on four prediction measures.

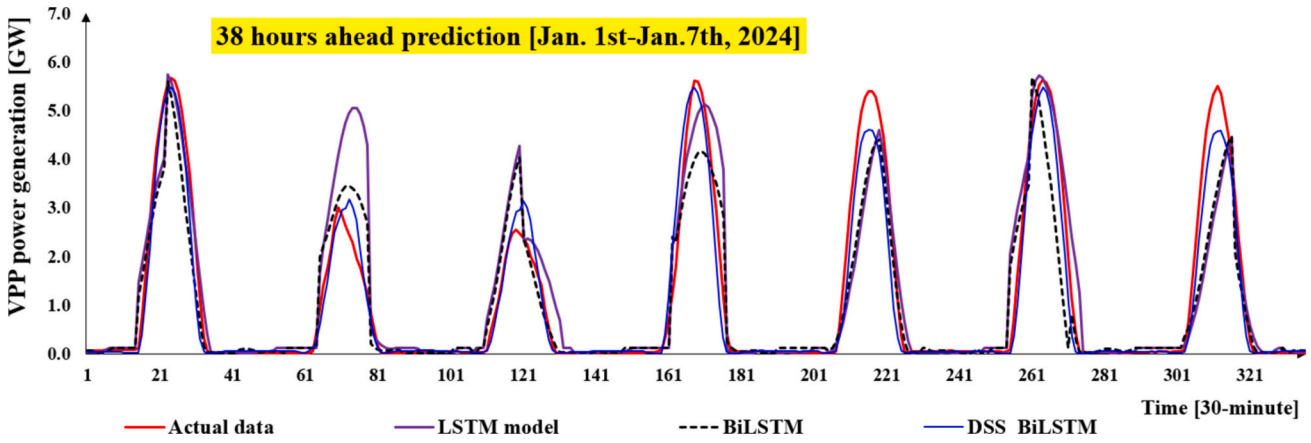


Fig. 8. Comparison of actual VPP generation and 38 h ahead prediction for a few techniques.

sMAPE value is close to the sMAPE value of the LSTM model. While, Fig. 8 illustrates that in all hours, the SARIMA results were almost flat curve. In other words, the forecasting values of the SARMIA model were much lower than the actual data that caused a lower sMAPE value. Moreover, Fig. 8 indicates a lower dispersion for the DSS-BiLSTM model compared to the LSTM and Bi-LSTM models for 38 h ahead prediction. The results emphasize that the DSS-BiLSTM model can be used for bidding purpose in the JEPX market with a slight adjusting before gate closure. While, the rest of the model are either unreliable or require a significant adjustment before gate closure.

Table 5 and Fig. 6 indicate that the LSTM, BiLSTM, and DSS-BiLSTM models exhibit superior prediction performance relative to the other models evaluated. Consequently, the Diebold-Mariano test was conducted to compare the forecasting accuracy of these three models. As shown in Table 6, the null hypothesis is rejected for the DSS-BiLSTM model across all time horizons, signifying that its forecasts are statistically significantly better than those of the LSTM and BiLSTM models. Additionally, Table 6 reveals that for the 30-min and 7-h time horizons, no significant difference is observed between the LSTM and BiLSTM models when assessed using the sMAPE indicator. Furthermore, based on the MAE and MSE indicators, the Diebold-Mariano test also indicates no significant difference between the LSTM and BiLSTM models at the 2-h and 7-h time horizons.

3.3. Weather types impact on DSS-BiLSTM results

To assess the impact of weather types on prediction error, the MSE indicator is utilized due to its alignment with the unit of electric power, while other error indicators showed a similar pattern. Fig. 9 illustrates both the frequency of weather types based on clustering (right side) and the distribution of weather types according to the MSE metric for the DSS-BiLSTM model (left side). The frequency histogram suggests that the highest MSE error is expected for partly cloudy conditions. However, the MSE-based distribution indicates that sunny days contribute the most to the prediction error. The correlation coefficient between the two datasets is 60 %, indicating a need for potential improvements in the clustering methodology.

4. Discussion

As outlined in the Introduction section and detailed in Table 1, numerous sophisticated hybrid models have demonstrated high accuracy in forecasting power generation across multiple time steps. Most of these models are designed for short-term power generation or demand forecasting. In contrast, the DSS-BiLSTM model proposed in this paper is tailored for forecasting bidding data in the power market. The DSS-BiLSTM model is particularly noteworthy for the following reasons:

- Simplicity: The DSS-BiLSTM model offers a significant advantage in terms of ease of implementation and computational efficiency compared to the complex hybrid models listed in Table 1. This model integrates the BiLSTM framework with a DSS comprising clustering of data and maintaining the most recent seven data points based on five distinct day types (rainy, overcast, cloudy, partly cloudy, and sunny). Generating test data for BiLSTM model was the main idea of this paper which was carried out by the DSS system based on clustering data and adjustment coefficients. The adjustment coefficients were calculated through the least squared optimization method. After generating DSS test data, the DSS-BiLSTM model employs a single-step forecasting approach to prepare multi-step forecasts.

Table 6
Diebold-Mariano test results.

Horizon	sMAPE			MAE			MSE		
	(M1 = LSTM, M2 = BiLSTM)	(M1 = LSTM, M2 = DSS-BiLSTM)	(M1 = BiLSTM, M2 = DSS-BiLSTM)	(M1 = LSTM, M2 = BiLSTM)	(M1 = LSTM, M2 = DSS-BiLSTM)	(M1 = BiLSTM, M2 = DSS-BiLSTM)	(M1 = LSTM, M2 = BiLSTM)	(M1 = LSTM, M2 = DSS-BiLSTM)	(M1 = BiLSTM, M2 = DSS-BiLSTM)
30 min	−1.71 (0.087)*	22.15 (0.000)	21.42 (0.000)	−12.66 (0.000)	39.91 (0.000)	40.11 (0.000)	−13.13 (0.000)	31.91 (0.000)	31.77 (0.000)
1 h	4.60 (0.000)+	−5.31 (0.000)	−9.25 (0.000)	8.18 (0.000)	−9.94 (0.000)	−10.91 (0.000)	−11.58 (0.000)	−10.90 (0.000)	−11.65 (0.000)
2 h	−2.98 (0.003)	6.94 (0.000)	8.64 (0.000)	−0.63 (0.53)*	−7.57 (0.000)	−7.24 (0.000)	−1.48 (0.14)*	−11.97 (0.000)	−11.55 (0.000)
7 h	1.14 (0.25)*	14.92 (0.000)	13.90 (0.000)	−0.87 (0.382)*	13.91 (0.000)	14.04 (0.000)	−1.82 (0.068)*	12.79 (0.000)	12.68 (0.000)
38 h	1.974 (0.048)	15.74 (0.000)	14.57 (0.000)	−4.79 (0.000)	15.49 (0.000)	18.93 (0.000)	−2.97 (0.003)	15.2 (0.000)	17.33 (0.000)

* : Null hypothesis is not rejected; + A (B): A represents normal test statistic, and B indicates the p-value

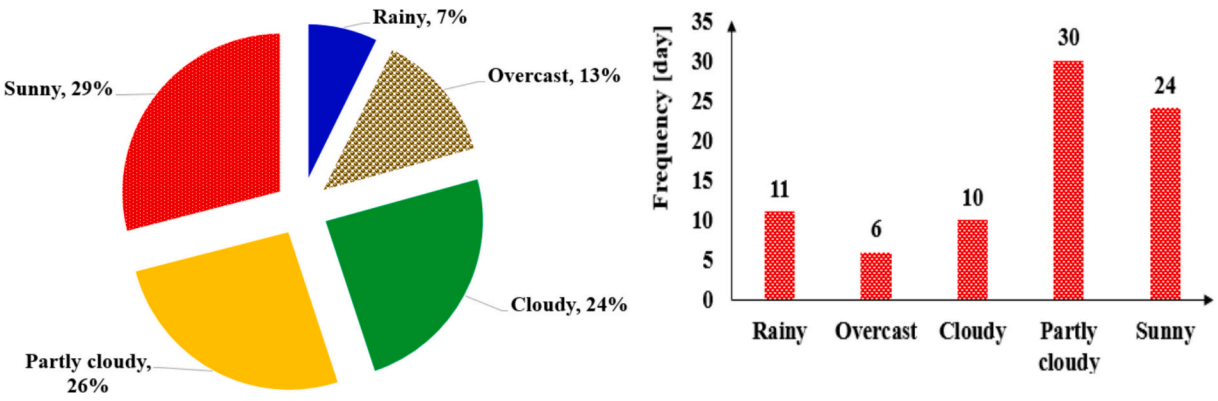


Fig. 9. Distribution of weather types based on MSE indicator for DSS-BiLSTM model (left), Frequency of weather types (right).

- Model Performance: The performance of the DSS-BiLSTM model was assessed using sMAPE, MAE, MSE, and RMSE metrics, alongside several other ANN models with varying parameters. Comparison with the ANN models detailed in [17] revealed that maintaining a fixed lookback period for LSTM and BiLSTM models led to increased prediction errors as the forecast horizon lengthened. In contrast, the DSS-BiLSTM model demonstrated more accurate results by varying prediction horizons compared to other ANN models. The DSS-BiLSTM model also exhibited lower error dispersion for long-term forecasts, as depicted in Fig. 6. Moreover, the DSS-BiLSTM model represented a consistently high correlation with the actual VPP data and a stable computational efficiency across various prediction horizons as given in Table 4.
- Power Market Application: Unlike many studies focused on power generation or demand load forecasting, this study addressed multi-step forecasting specifically within the power market domain. Accurate bidding data requires a robust and precise model for long-term prediction. The DSS-BiLSTM model enables VPP owners to submit bids with minimal adjustments before gate closure.

5. Conclusion

This study combined bidirectional long short-term memory (BiLSTM) network with decision support system (DSS) to forecast power generation of a virtual power plant (VPP) with 1.5 [GW] generation capacity in Tokyo Metropolitan area, Japan. Instead of using prediction data in the training dataset, the proposed model generated the sequence of data, called DSS test data, by maintaining dependency between forecast values and reducing the bias of the estimated model's parameters. To generate the sequence of data, the K-means clustering algorithm and the least squared optimization model were used in the DSS. The DSS test data was applied as a testing data into the BiLSTM model to forecast VPP power generation for 38- h ahead as a bidding data to use in Japan electric power exchange (JEPX) market. The DSS-BiLSTM model's performance was evaluated and compared with BiLSTM, LSTM, Transformer, Attention-based network, GRU, and five statistical models (AR, ARMA, ARIMA, SARIMA, and SARIMAX).

The results of forecasting models showed that the DSS-BiLSTM model more accurately forecasted the VPP power generation than TEPCO prediction in 30-min ahead in terms of all four prediction measures (Table 3). For the rest of steps ahead forecasting (1, 2, 7, and 38 h ahead), the DSS-BiLSTM model had consistently outperformed the rest of models in terms of sMAPE, MAE, MSE, and RMSE metrics (Fig. 6 and Table 6). For a 38-h forecasting horizon, the sMAPE metric for the DSS-BiLSTM, BiLSTM, LSTM, transformer network, attention-based network, and GRU models were 50 %, 62 %, 64 %, 91 %, 90 %, and 80 %, respectively. This pattern is consistently observed across the three additional error metrics—MAE, MSE, and RMSE— and across various forecasting time horizons. These results demonstrate that the DSS-BiLSTM model consistently outperforms both the BiLSTM and LSTM models in terms of prediction accuracy.

Unlike other studies which typically utilize a fixed lookback period for time series forecasting for the LSTM and BiLSTM models, this study explored the impact of employing a variable lookback period to enhance long-term prediction accuracy. The results, as shown in Fig. 7, indicate that LSTM and BiLSTM models incorporating a variable lookback period achieved lower errors across the MAE, MSE, and RMSE metrics compared to models with a fixed lookback period, particularly as the forecasting horizon extended. The main advantages of the DSS-BiLSTM model was its simplicity, low dispersion, and prediction accuracy. In contrast to various hybrid models proposed for multi-step prediction, the DSS-BiLSTM model generated test data and utilized single-step technique to forecast VPP power generation. The results of forecasting did diverge for long-term prediction (38 h ahead). In other words, the growth rate of the prediction error was very low over long-term prediction and was more reliable than the rest of models mentioned in this research.

Future research will involve a comparative analysis of the DSS-BiLSTM model against Transformers, attention-based networks, and graph neural networks to identify the most suitable model for VPP power generation data by evaluating their performance across various forecasting metrics. Moreover, we will explore the impact of more frequent updates to the DSS test data (more than once per day) on enhancing the accuracy of the DSS-BiLSTM model, particularly for short-term predictions, such as 1-h and 2-h forecasts. Ultimately, the results from this study will be utilized to identify an optimal confidence level for bid data by accounting for market power price variations as the gate closure approaches.

CRediT authorship contribution statement

Reza Nadimi: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mika Goto:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

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 will be made available on request.

Acknowledgements

This work was supported by Council for Science, Technology andInnovation (CSTI), Cross-ministerial Strategic Innovation PromotionProgram (SIP), the 3rd period of SIP “Smart energy management system”,Grant Number JPJ012207 (Funding agency: JST).

References

[1] The Electric Power Industry in Japan 2024. Japan Electric Power Information Center Inc; 2024.
[2] Kitamura T, Managi S. Energy security and potential supply disruption: a case study in Japan. Energy Policy 2017;vol. 110:90–104.
[3] Washizu A, Ju Y, Yoshida A, Tayama M, Amano Y. Modeling the distributed energy resource aggregator services in a macroeconomic framework: the application to Japan. Energy 2024;vol. 312.
[4] Mizuno E. Overview of wind energy policy and development in Japan. Renewable and Sustainable Energy Reviews 2014;vol. 40:999–1018.
[5] Ohtake H, Shimose K-I, Fonseca JGDS, Takashima T, Oozeki T, Yamada Y. Accuracy of the solar irradiance forecasts of the Japan meteorological agency mesoscale model for the Kanto region, Japan. Solar Energy 2013;Vols. 98:138–52. Part B.
[6] Malehmirchegini L, Farzaneh H. Incentive-based demand response modeling in a day-ahead wholesale electricity market in Japan, considering the impact of customer satisfaction on social welfare and profitability. Sustainable Energy, Grids and Networks 2024;vol. 34:101044.
[7] Kermanshahi B, Iwamiya H. Up to year 2020 load forecasting using neural nets. International Journal of Electrical Power & Energy Systems 2002;vol. 24(9): 789–97. no.
[8] Nepal B, Yamaha M, Yokoe A, Yamaji T. Electricity load forecasting using clustering and ARIMA model for energy management in buildings. Japan Architectural Review 2020;vol. 3(1):62–76. no.
[9] Taieb SB. Machine learning strategies for multi-step-ahead time series forecasting. Université Libre de Bruxelles; 2014.
[10] Wang C, Zhao H, Liu Y, Fan G. Minute-level ultra-short-term power load forecasting based on time series data features. Applied Energy 2024;vol. 372: 123801.
[11] Gulay E, Mustafa S, Akgun OB. Forecasting electricity production from various energy sources in Türkiye: a predictive analysis of time series, deep learning, and hybrid models. Energy 2024;vol. 286:129566.
[12] Shakouri HG, Nadimi R, Ghaderi SF. Investigation on objective function and assessment rule in fuzzy regressions based on equality possibility, fuzzy union and intersection concepts. Computers & Industrial Engineering 2017;vol. 110:207–15.
[13] Nadimi R, Takahashi M, Tokimatsu K, Goto M. The reliability and profitability of virtual power plant with short-term power market trading and non-spinning reserve diesel generator. Energies 2024;vol. 17(9). no.
[14] Zapata J, Vandewalle J, D’haeseleer W. A comparative study of imbalance reduction strategies for virtual power plant operation. Applied Thermal Engineering 2014;vol. 71(2):847–57. no.
[15] Yu S, Zhou S, Zheng S, Li Z, Liu L. Developing an optimal renewable electricity generation mix for China using a fuzzy multi-objective approach. Renewable Energy 2019;vol. 139:1086–98.
[16] Rozon F, McGregor C, Owen M. Long-term forecasting framework for renewable energy technologies’ installed capacity and costs for 2050. Energies 2023;vol. 16 (19). no.
[17] Sharadga H, Hajimirza S, Balog RS. Time series forecasting of solar power generation for large-scale photovoltaic plants. Renewable Energy 2020;vol. 150: 797–807.
[18] Gao Y, Miyata S, Akashi Y. Multi-step solar irradiation prediction based on weather forecast and generative deep learning model. Renewable Energy 2022;vol. 188: 637–50.
[19] Li F, Zheng H, Li X. A novel hybrid model for multi-step ahead photovoltaic power prediction based on conditional time series generative adversarial networks. Renewable Energy 2022;vol. 199:560–86.
[20] Lu Y, Tian Z, Zhou R, Liu W. Multi-step-ahead prediction of thermal load in regional energy system using deep learning method. Energy and Buildings 2021; vol. 233.
[21] In Y, Jung J-Y. Simple averaging of direct and recursive forecasts via partial pooling using machine learning. International Journal of Forecasting 2022;vol. 38 (4):1386–99. no.
[22] Xue P, Jiang Y, Zhou Z, Chen X, Fang X, Liu J. Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms. Energy 2019; vol. 188.
[23] Koesdwiady A, Khatib AE, Karray F. Methods to Improve multi-step time series prediction. In: in IEEE International Joint Conference on Neural Networks (IJCNN). Brazil: Rio de Janeiro; 2018.
[24] Marcellino M, Stock JH, Watson MW. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. Journal of Econometrics 2006;vol. 135(1–2):499–526. no.
[25] Taieb SB, Hyndman RJ. Recursive and direct multi-step forecasting: the best of both worlds. 2012.
[26] Bontempi G. Long term time series prediction with multi-input multi-output local learning. Université Libre de Bruxelles; 2008.
[27] Taieb SB, Sorjamaa A, Bontempi G. Multiple-output modeling for multi-step-ahead time series forecasting. Neurocomputing 2010;vol. 73:1950–7.
[28] Nguyen H-P, Baraldi P, Zio E. Ensemble empirical mode decomposition and long short-term memory neural network for multi-step predictions of time series signals in nuclear power plants. Applied Energy 2021;vol. 283.
[29] Zhou Y, Li Y, Wang D, Liu Y. A multi-step ahead global solar radiation prediction method using an attention-based transformer model with an interpretable mechanism. International Journal of Hydrogen Energy 2023;vol. 48(40): 15317–15,330.
[30] Exchange JEP. Japan wholesale electric power exchange trading guide. JEPX; 2019.
[31] Elkady S, Hernantes J, Labaka L. Decision-making for community resilience: a review of decision support systems and their applications. Heliyon 2024;vol. 10 (12).
[32] Jiang D, Zhu W, Muthu BA, Seetharam TG. Importance of implementing smart renewable energy system using heuristic neural decision support system. Sustainable Energy Technologies and Assessments 2021;vol. 45.
[33] Shakouri HG, Nadimi R, Ghaderi SF. A hybrid TSK-FR model to study short-term variations of the electricity demand versus the temperature changes. Expert Systems with Applications 2009;vol. 36(2):1765–72. no.
[34] "TEPCO Power Grid TEPCO. April 1, 2015 [Online]. Available: https://www.tepco.co.jp/en/forecast/html/download-e.html. [Accessed 01 112,023].
[35] Sharda S, Singh M, Sharma K. RSAM: robust self-attention based multi-horizon model for solar irradiance forecasting. IEEE Transactions on Sustainable Energy 2021;vol. 12(2):1394–405. no.
[36] Kim J, Obregon J, Park H, Jung J-Y. Multi-step photovoltaic power forecasting using transformer and recurrent neural networks. Renewable and Sustainable Energy Reviews 2024;vol. 200.
[37] Yang Y, Wang Z, Zhao S, Zhou H, Wu J. Robust autoregressive bidirectional gated recurrent units model for short-term power forecasting. Engineering Applications of Artificial Intelligence 2024;vol. 138:109453.
[38] Huang X, Li Q, Tai Y, Chen Z, Liu J, Shi J, et al. Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM. Energy 2022;vol. 246:123403.
[39] Kong X, Du X, Xue G, Xu Z. Multi-step short-term solar radiation prediction based on empirical mode decomposition and gated recurrent unit optimized via an attention mechanism. Energy 2023;vol. 282.
[40] Rafati A, Joorabian M, Mashhour E. An efficient hour-ahead electrical load forecasting method based on innovative features. Energy 2020;vol. 201.
[41] Ahmed A, Khalid M. An intelligent framework for short-term multi-step wind speed forecasting based on Functional Networks. Applied Energy 2018;vol. 225:902–11.
[42] Lu S. Multi-step ahead ultra-short-term wind power forecasting based on time series analysis. In: In IEEE international conference on computer information and big data applications (CIBDA), Guiyang, China; 2020.
[43] Wang H, Xiong M, Chen H, Liu S. Multi-step ahead wind speed prediction based on a two-step decomposition technique and prediction model parameter optimization. Energy Reports 2022;vol. 8:6086–100.
[44] Tian Z, Chen H. Multi-step short-term wind speed prediction based on integrated multi-model fusion. Applied Energy 2021;vol. 298.
[45] Chen X-J, Zhao J, Jia X-Z, Li Z-L. Multi-step wind speed forecast based on sample clustering and an optimized hybrid system. Renewable Energy 2021;vol. 165: 595–611.
[46] Hong J, Liang F, Yang H, Zhang C, Zhang X, Zhang H, et al. Multi- forword-step state of charge prediction for real-world electric vehicles battery systems using a novel LSTM-GRU hybrid neural network. eTransportation 2024;vol. 20.
[47] Oscar P, Garc'ıa P, Gonz'alez L, Villa G. Multi-step machine learning forecasting of power consumption and PV generation for distributed energy management applications. In: In IEEE Conference Proceedings (IEEE Conf Proc), USA; 2023.
[48] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks 2005;vol. 18(5–6): 602–10. no.
[49] Hamilton JD. Time series analysis. Princeton University Press; 1994.
[50] Burnham KP, Anderson DR. Multimodel inference: understanding aic and bic in model selection. Sociological Methods & Research 2004;vol. 33(2):261–304. no.
[51] Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 1979;vol. 74(366): 427–31. no.
[52] Nadimi R, Tokimatsu K. Modeling of quality of life in terms of energy and electricity consumption. Applied Energy 2018;vol. 212:1282–94.
[53] Schubert E. Stop using the elbow criterion for k-means and how to choose the number of clusters instead. 2023 [Online].
[54] Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. International Journal of Forecasting 2006;vol. 22(4):679–88.
[55] Diebold FX, Mariano RS. Comparing predictive accuracy. Journal of Business & Economic Statistics 1995;vol. 13(3):253–63. no.
[56] Li ZL, Zhang GW, Yu J, Xu LY. Dynamic graph structure learning for multivariate time series forecasting. Pattern Recognition 2023;vol. 138:109423.
[57] Robeson SM, Willmott CJ. Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. Plos One 2023;vol. 18(2):e0279774. no.