IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 August 2023, accepted 4 September 2023, date of publication 11 September 2023,
date of current version 19 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3313632

RESEARCH ARTICLE

An Enhanced Performance Evaluation Metrics for Wind Power Ramp Event Forecasting

SOLUI YU, (Student Member, IEEE), AND JIN HUR, (Senior Member, IEEE)
Department of Climate and Energy Systems Engineering, College of Engineering, Ewha Womans University, Seoul 03760, South Korea

Corresponding author: Jin Hur (jhur@ewha.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government through MSIT under Grant 2022R1F1A1074397, and in part by the Korea Electric Power Corporation under Grant R21XO01-1.

ABSTRACT As wind power is volatile and intermittent, an increase in the proportion of wind power generation may affect the security of power systems. Accurate wind power generation and ramp forecasting are crucial for reducing operating costs and maintaining the power balance. This study proposed an enhanced performance evaluation metric for wind power ramp event forecasting, and analyzed the forecasting results using this metric. Experimental ramp event forecasting was conducted on wind farms located in Jeju, South Korea by employing an indirect forecasting method for wind power output forecasting. Considering the evaluation of forecasting performance, the accuracy (ACC) was observed to be 0.86 for Case #1 and 0.90 for Case #2. The results highlight the advantages of the proposed metric over the widely used confusion matrix for performance evaluation, which has more detailed and visually based analytical capabilities. The simulation demonstrates that the proposed approach is poised to make a noteworthy contribution in the advancement of tool development, enabling real-time curtailment forecasting and facilitating well-informed decision-making in high wind power scenarios.

INDEX TERMS Performance evaluation metric, power system security, ramp event forecasting, wind power.

I. INTRODUCTION

Owing to international cooperation to realize carbon neutrality, the energy transition from fossil fuels to renewable energy is accelerating, and the penetration of renewable sources into the energy mix is gradually increasing. To meet the climate target, the International Renewable Energy Agency (IRENA) suggested that the total share of renewable energy should be increased to approximately two-thirds of the total primary energy supply by 2050, and that 60% or more of the total final energy consumption should be covered by renewable energy by each country [1]. According to the 5th Energy Master Plan of South Korea, the target for renewable sources based on power generation is set at 22.2% by 2034, of which wind and solar will account for approximately 19% [2]. As the proportion of renewable energy with intermittency and uncertainty increases, imbalances in energy supply and demand may occur when integrating power systems, and system reliability and acceptability are key issues [3]. Wind resources are in the spotlight in the renewable energy sector because of their large-scale power generation capacity and high space utilization, and various studies are being conducted on the power forecasting of long-to short-term wind power output to ensure the security of the power system [4].

As system instability and curtailment owing to the real-time injection of wind power occur, forecasting research on wind fluctuations and ramp characteristics is required [5]. The ramp rate is an indicator of fluctuations, which means a change in wind power output per unit time with direction; the increase and decrease in wind power output are referred to as up ramps and down ramps, respectively.

Ramp-related forecasting studies can be classified into direct and indirect forecasting methods [6]. Direct methods forecast ramp information directly using the historical ramp rate or event data. Indirect methods detect ramp events based on the results of wind power output and wind speed forecasting.

Furthermore, similar to wind power output forecasting, ramp-event forecasting uses physical, statistical, and machine learning approaches [7]. Physical approaches are

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaodong Liang.

implemented based on weather variables such as wind speed and direction, using physical formulas [8]. A complex forecasting process, including mathematical calculations to deal with meteorological and spatial information, is inevitable; a Numerical Weather Prediction (NWP)-based study is a typical example.

Statistical approaches correspond to methods for identifying and forecasting the statistical relationships between ramp events, power output, and other variables [9]. A statistical model is suitable for quantifying the relationship between variables when there is a certain pattern, and case studies are performed using Monte Carlo [10] and Principal Component Analysis (PCA) [11].

Machine learning approaches imitate the neural network activity of the brain and derive forecasts by learning past data [12]. In these approaches, the use of ramp forecasting is increasing because the relationship between the input variables (wind power output, ramp rate, etc.) are nonlinear, or there is no constant pattern [13]. Furthermore, because the wind power output is derived only by learning calculations, errors owing to multiple forecasting can be reduced [14]. Han et al. performed wind power ramp event forecasting using a Support Vector Machine (SVM), Long and Short-Term Memory (LSTM) [15], Optimized Swinging Door algorithm (OpSDA) [16], which are typical deep learning models.

TABLE 1. Method for ramp event forecasting.

Approaches	Interpretation
Physical Approaches	Approaches that utilize physical or mathematical expression through meteorological and spatial information
Statistical Approaches	Approaches to find the relation between single/multiple variables based on historical data of wind power
Machine Learning Approaches	Approaches using a forecasting model based on machine learning models

The recent literature on ramp event prediction has been compared, analyzed, and organized, as summarized in Table 2. Initially, the utilization of NWP was dominant in ramp event detection studies conducted until 2015, with a limited time series model-based approach [4]. NWP models are developed through numerical simulations of atmospheric dynamics and various physical processes [17]. These models begin with initial predictions based on analysis that capture real-time atmospheric conditions using a three-dimensional grid. However, spanning from 2017 to 2022, there has been noticeable shift towards machine learning based forecasting approach. For instance, Zhang et. al. performed ramp prediction by improving the short-term wind power output prediction performance as part of the Wind Forecast Improvement Project [16]. To extract the characteristics of the ramp event, ramp prediction in multiple spatial and temporary scales was performed using the OpSDA algorithm. Cornejo-Bueno et al. extensively explored machine learning regression techniques such as Support Vector Regression (SVR) and Artificial Neural Network (ANN), and performed ramp forecasting by applying reanalysis data of weather variables [17]. Taylor used autoregressive logit models to implement a model that predicts the probability of ramp event generation for multiple thresholds [18]. Y. Fujimoto et. al enhanced the accuracy of ramp forecasting through subsystem for wind power forecasting based on NWP and Swing Door Algorithm (SDA) [8]. Han et. al. successfully identifying ramp events by training time series data with Convolutional Neural Network (CNN) and LSTM algorithms using wind power output data and ramp feature as input [14]. Cornejo-Bueno et.al predicted the occurrence of ramp events through a hybrid model that combines extreme learning machine (ELM) and SVM algorithms [19], and Y. Zhao et. al employed a statistical approach, specifically the Bayesian network, to detect ramp events [20]. Furthermore, Okada et. al proposed a real-time ramp prediction model integrating an optimal NWP model with PCA model [21].

The contemporary literature emphasizes that ramp event prediction through deep learning models or hybrid models is dominant. In particular, OpSDA was used frequently, the SDA is a data compression technique that employs the parallelogram rule to filter samples [14]. Furthermore, there is an optimized version of the algorithm tailored for ramp prediction and other forecasting tasks, known as OpSDA [16]. OpSDA has gained significant popularity for its application in predicting ramp event of wind and solar power. The evaluation mechanisms of ramp event encompass confusion matrix and ACC, CSI, and POD that can be obtained through it. The primary parameters defining ramp events, such as duration and threshold, vary from the literature analysis, common benchmark designates instances where wind power fluctuations surpass 10-30% of installed capacity within a 6-hour interval.

In this study, a short-term power generation forecasting method using an LSTM model was proposed, which has excellent performance in forecasting time-series data with nonlinear characteristics, such as wind power output. After learning the LSTM model using the output data of the Jeju wind farm as the input, the wind power forecasting value for the period of the test data was derived. Subsequently, based on the error indicators for the forecasted and measured values, the forecasting accuracy was identified, and the performance of the forecasting model was verified. Thereafter, ramp-event detection is derived by calculating the ramp rate based on the output forecasting value. The proposed method for ramp forecasting can be used in various ways to integrate power systems by forecasting both wind power output and ramp information. By reducing the operating cost of the power system owing to the intermittent characteristics of wind power generation and taking preemptive responses against sudden power fluctuations, a power balance can be achieved through more efficient power system operation.

TABLE 2. Recent literatures of ramp event forecasting.

Reference	Ramp Definition		Forecasting Method		Performance Evaluation Metric
	Duration (Δt)	Threshold ($P_{threshold}$)	Approach Type	Model Type	
J. Zhang et. al. (2017) [16]	< 4 h	10-30%	Machine learning	NWP, OpSDA	Confusion matrix
L. Cornejo-Bueno et al. (2017) [17]	6 h	20 MW	Physical, Machine learning	NWP, Regression ML	RMSE, MAE
J. W. Tailor (2017) [18]	1-2 h	10-30%	Statistical	Autoregressive logit models	Brier skill score
Y. Fujimoto et. al (2019) [10]	0.5-48 h	5-30%	Physical, Machine learning	NWP, SDA	Confusion matrix
J. Li et.al (2020) [8]	15 min	30%	Machine learning	TECI	MAPE
L. Han et.al (2020) [14]	15 min	3%	Machine learning	OpSDA, LSTM, CNN	MAPE, Confusion matrix
L. Cornejo-Bueno et al. (2020) [19]	6 h	2%	Physical, Statistical, Machine learning	SVR, ELM	Confusion matrix, ROC
Y. Zhao et. al (2021) [20]	30 min	10-11%	Statistical	Bayesian network	Score
M. Okada et. al (2022) [21]	6-12 h	30%	Statistical, Physical	NWP, PCA, GCM	Confusion matrix

The remainder of this study is organized as follows: Section II describes the methodology of LSTM-based wind power output forecasting and ramp-event forecasting, and Section III proposes a ramp-event forecasting modeling method and process. Section IV presents the forecasting results of a case study of a wind farm located on Jeju Island and evaluates the performance of the model using existing and advanced methods. Finally, section V concludes the study.

II. PROPOSED PERFORMANCE VALIDATION METRIC

This section presents the methodology used in this research, encompassing the sequential explanation of LSTM as a deep learning model, definition of ramp events, and performance evaluation metrics for ramp event forecasting. This section discusses the confusion matrix, a well-established performance evaluation metric for ramp-event forecasting, and provides detailed and in-depth information on heatmap-based metrics.

A. LSTM (LONG AND SHORT-TERM MEMORY) MODEL

In this study, forecasting was implemented by utilizing the LSTM model among ANN models with an excellent ability to map the nonlinearity of wind resources. LSTM model is a type of Recurrent Neural Network (RNN) model suitable for modeling time series data [22]. An RNN model has the advantage of reflecting the impact of historical data in forecasting; however, the longer the data, the less suitable it is for long-term forecasting because of the gradient vanishing problem [23]. The LSTM model complements the long-term dependencies of the RNN model, is suitable for forecasting that considers past and present data, and demonstrates excellent performance [24].

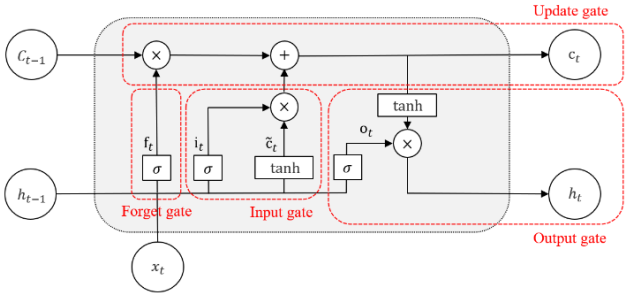

FIGURE 1. Architecture of LSTM model.

TABLE 3. Mathematical equation of LSTM model.

Gate Name	Equation
Input Gate	$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$ $\tilde{c_t} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$
Forget Gate	$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
Update Gate	$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c_t}$
Output Gate	$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$ $h_t = o_t \odot \tanh(c_t)$

As shown as Figure 1, LSTM model has an architecture that sequentially passes through four gates, input, forget, update, and output gate to learn past data and derive output values, and Table 3 shows the mathematical expression of each gate [25]. First, the input gate determines whether new information can be stored and is composed of i_t in charge

of the input and $\tilde{c}_t$ in charge of the update. Next, the forget gate is used to determine the storage of past information and is determined by f_t. The update gate updates the existing information based on the values derived from the input gate and update gate and stores the value in c_t. Finally, the output gate determines the final output value and is composed of o_t and h_t.

The variables and symbols for the equations are as follows. x_t represents the input at time t and y_t depicts the output at time t. h_{t-1} is a variable that stores the hidden state at time $(t-1)$. W corresponds to the weighted matrix, b represents the bias variables, and each subscript has an initial gate name. σ means a sigmoid functions and *tanh* is a hyperbolic tangent. $\odot$ is an arithmetic symbol representing the Hadamard product.

B. RAMP EVENT DEFINITION

The ramp rate is the ratio of the change in wind power output during the time interval (Δt) [7], [16] and is expressed by Equation (1). The magnitude of the ramp rate indicates that the value is obtained by taking the absolute value of the ramp rate; the larger the magnitude, the greater the change in the output. The ramp rate has both positive and negative directions. A positive sign represents increasing wind power output, whereas a negative sign refers to decreasing wind power output.

$$Ramp\ Rate = P(t + \Delta t) - P(t) \tag{1}$$

A ramp event is defined as a very high fluctuation over a short time and indicates that the ramp rate is greater than or equal to $P_{threshold}$ [26]. The definition of the ramp rate is usually based on its magnitude, duration, and direction [16]. In this study, a ramp event is defined in Table 4 and has the following four parameters:

- Magnitude ($P(t)$): Variation in wind output at time t.
- Duration (Δt): Time interval considered identifying a ramp event.
- Direction ($+/-$): Increase ($+$) or decrease ($-$) in wind power output.
- Threshold ($P_{threshold}$): Reference value to determine a ramp event; $P^u_{threshold}$ and $P^d_{threshold}$ refer to the thresholds of up- and down-ramp events, respectively. Here, $P_{threshold}$ is dependent on the installed capacity of the wind farm and is calculated as a range spanning from 10% to 30% of the total installed capacity.

TABLE 4. The definition of ramp event.

Ramp Events	Equation
Up Ramp Event	$P(t+\Delta t) - P(t) > P^u_{threshold}$
Down Ramp Event	$P(t+\Delta t) - P(t) < -P^d_{threshold}$

C. PERFORMANCE VALIDATION METRIC FOR RAMP EVENT DETECTION

1) CONFUSION MATRIX

The confusion matrix defined in Table 5 was used to demonstrate the performance of the ramp-event forecasting. A confusion matrix is suitable to show the relationship between each case or variable when an event is classified as an independent binary case or variable [27]. Hence, this study enables the categorization of cases involving ramp events, making the confusion matrix a suitable metric for illustrating the outcomes.

TABLE 5. The component of confusion matrix.

Index		Ramp Forecasting		Total
		True	False	
Ramp Observation	True	TP	FN	TP+FN
	False	FP	TN	FN+TN
Total		TP+FP	FN+TN	N=TP+FP+FN+TN

"True" means that the prediction and the measurement match, and "False" represents that the prediction and the measurement do not match. "Positive" indicates when the ramp event is forecasted to occur, and "Negative" refers to when the ramp event is forecasted to not occur. By combining T or F and P or N according to whether the forecasting and the actual measurement match, the forecasting result is represented as a total of four cases: TP, FN, FP, and TN [28].

- True Positives or Hits (TP)
  : Ramp event and ramp event occurrences were forecasted.
- False Negatives or Misses (FN)
  : A ramp event occurred but was not forecasted.
- False Positives or False Alarms (FP)
  : A ramp event did not occur but was forecasted.
- True Negative (TN)
  : Ramp event did not occur, and no ramp event was forecasted.

TABLE 6. Evaluators for ramp event forecasting.

Evaluator	Equation
ACC (Accuracy)	$\frac{TP+TN}{TP+FN+FP+TN}$
CSI (Critical Success Index)	$\frac{TP}{TP+FN+FP}$
POD (Probability of Detection)	$\frac{TP}{TP+FN}$
FA (Forecast Accuracy)	$\frac{TP}{TP+FP}$

The evaluators can be derived by combining the components of the confusion matrix and utilized for forecasting

performance evaluation. Typical evaluators are summarized in Table 6. ACC represents the proportion of correct forecasting points among all events [14], and CSI refers to the fraction of true positive events and all events except true negatives [4]. POD represents the rate of properly predicted ramp events among the total ramp events [16], and FA represents the percentage of actual ramp events among the forecasted ramp events [29]. Using the above values, the ramp event detection performance can be determined and compared.

2) PROPOSED PERFORMANCE EVALUATION METRIC

The proposed performance evaluation metric was presented to examine the forecasting performance by considering the direction of the ramp rate and the occurrence of ramp events. As shown in Table 7, the confusion matrix could only distinguish dichotomously whether a ramp event occurred; however, in this study, the category was classified into four categories: (1) up-ramp events, (2) up-ramp rates, (3) down-ramp rates, and (4) down-ramp event as shown in Table 7. The ''Ramp Observation - True'' in the confusion matrix is divided into ''up ramp event'' and ''down ramp event'', and the ''Ramp Observation - False'' is separated into ''up ramp rate'' and ''down ramp rate'', as a result proposed metric is specified by direction.

TABLE 7. The category of proposed performance evaluation metric for ramp event forecasting.

Category	Description	Range [MW]
Up Ramp Event	Occurrence or detection of up ramp event	[6, ∞]
Up Ramp Rate	Occurrence or detection of up ramp rate but not up ramp event	[0, 6)
Down Ramp Rate	Occurrence or detection of down ramp rate but not down ramp event	(-6, 0]
Down Ramp Event	Occurrence or detection of down ramp event	[-∞, -6]

The existing evaluation method, which is a confusion matrix, has a 2 × 2 format, whereas the proposed method has a 4 × 4 format, which has the advantage of being able to distinguish in detail. The existing method represents only four areas, but the proposed metric can display 16 areas. The detailed area classification is shown in Figure 2 by subdividing the parts marked with the same color in the proposed method. In addition, the number of points corresponding to each area of the proposed method is described, and the brightness is differentiated to facilitate visual comparison. Moreover, visual elements are maximized using color gradation techniques by displaying a large number representing dark colors and a small number representing light colors.

The schematic analysis method of the proposed performance evaluation metric is as follows: the darker the color of the right-down diagonal (A, F, K, and P) and the lighter the color of the right-up diagonal (D, G, J, and N), the better the performance of the ramp rate and ramp event. This is because the right-down diagonal represents the point at which the forecasting is accurate, and the right-up diagonal represents the opposite forecasting. The proposed matrix has symmetric characteristics based on a diagonal, which makes it easy to check the bias for the up and down directions.

Confusion Matrix		Ramp Forecasting	
		True	False
Ramp Observation	True	TP	FN
	False	FP	TN

(a) Original Evaluation Metric: Confusion Matrix

Proposed Metric		Ramp Forecasting			
		Up Ramp Event	Up Ramp Rate	Down Ramp Rate	Down Ramp Event
Ramp Observation	Up Ramp Event	A	B	C	D
	Up Ramp Rate	E	F	G	H
	Down Ramp Rate	I	J	K	L
	Down Ramp Event	N	M	O	P

(b) Proposed Evaluation Metric

FIGURE 2. Comparison of confusion matrix and proposed metric.

III. RAMP EVENT FORECASTING MODELING

The following section describes the forecasting models of wind power and ramp events, including details of the input data, theoretical background of each algorithm, and overall process of ramp event predictions.

A. INPUT DATA DESCRIPTION

This study utilizes empirical data from wind farms in Jeju, Korea. The characteristics of the wind power output data are as follows:

- Output period: 2021.1.1 00:00:00 – 2021.12.31 23:45:00
- Data time interval: 15 min
- Installed capacity: 60 MW
- Average output: 21.7 MW

To compare the monthly variability in the wind power output, fluctuations were evaluated according to the standard deviation (SD) and interquartile range (IQR), which are representative dispersion indicators. The larger the values of the two indicators, the greater the output fluctuation. The two values of each indicator were averaged monthly to determine the degree of output variability. Monthly rankings were derived independently for SD and IQR, and the final rank

of variability was calculated by assuming that the smaller the sum of each rank, the greater the variance. As shown in Table 8 and Figure 3, the output dispersion was the highest in November and December and the lowest in July.

TABLE 8. Monthly variability analysis and ranking of wind farm data.

Month	SD	IQR	Rank of Volatility
January	13.9	23.2	4
February	14.1	22.7	5
March	14.8	23.1	3
April	13.4	19.2	7
May	11.6	15.6	8
June	9.5	7.8	11
July	7.6	8.9	12
August	9.4	9.9	10
September	15.6	15.9	5
October	12.3	11.2	8
November	16.9	23.3	1
December	16.7	29.5	1

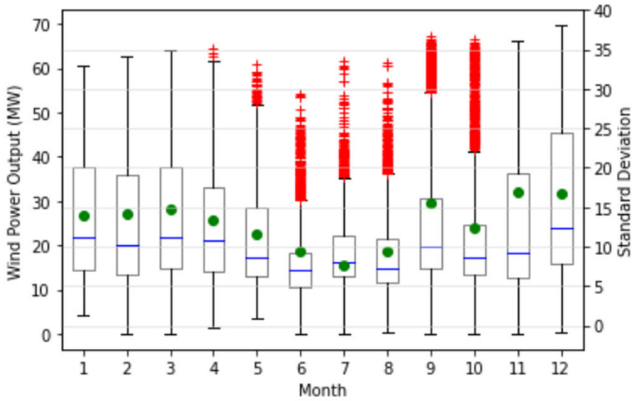

FIGURE 3. Variability analysis of wind power output.

The training and test sets for wind power output and ramp event forecasting are presented in Table 9. Two cases were considered: Case #1 was a low-volatility case that predicted wind power in July, and Case #2 was a highly volatile case that predicted wind power in December; the time points were the same. The period of the test data for Cases #1 and #2 was set to July and December, and the period of the training sets for each case was set to June and November, which immediately preceded the test data period.

B. RAMP EVENT FORECASTING PROCESS

To forecast the ramp rate and wind power generation events, the following three steps were performed, as shown in Figure 4.

- Step 1: Wind power output data pre-processing
- Step 2: LSTM-based short-term wind-power output forecasting

TABLE 9. Classification of data sets by case.

Case Number	Period of Training Data	Period of Test Data
Case #1 : Low Volatile Case	2021-6-1 00:00 ~ 2021-6-30 23:45	2021-7-1 00:00 ~ 2021-7-31 23:45
Case #2 : High Volatile Case	2021-11-1 00:00 ~ 2021-11-30 23:45	2021-12-1 00:00 ~ 2021-12-31 23:45

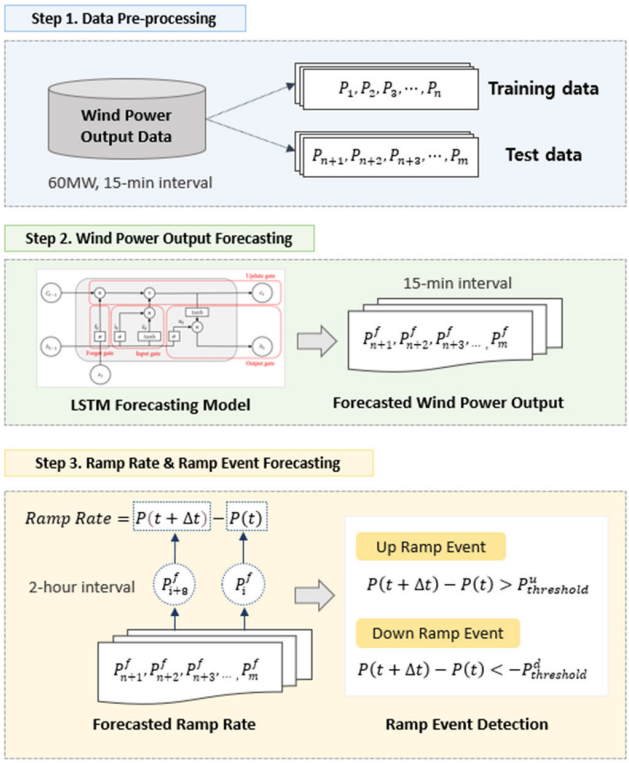

FIGURE 4. Wind power output and ramp forecasting process.

- Step 3: Wind power ramp rate and ramp-event forecasting

1) DATA PRE-PROCESSING

Step 1 includes all processes for refining input values before entering the forecasting model, such as input data preprocessing and dataset splitting. The LSTM model-based wind power output forecasting model uses wind power data in 15-minute time interval as input data. In this model, only the output of wind power generation was set as the input variable, except for weather variables such as wind speed, humidity, and temperature. After the input data were categorized into training and test sets, data preprocessing, such as missing value processing, was performed. Because the model forecasts the difference in wind output for the time interval rather than directly forecasting wind power output, the time series data of the output difference are utilized for training and test data. Subsequently, because the magnitude of the wind power output varied depending on the season, time, and

wind generator, the data were normalized through min-max scaling. Min-max normalization is a method of converting all values from 0 to 1 by corresponding the minimum value of each element to 0 and the maximum value to 1.

Algorithm 1 Data Pre-Processing

Software & tools:R program
Input: Training and test data of Wind Power Output Data in Jeju Island's Wind Farm (.csv)
Define training and test data
$\{P_{training}(t)\} = P_1, P_2, P_3, \cdots, P_n$
$\{P_{test}(t)\} = P_{n+1}, P_{n+2}, P_{n+3}, \cdots, P_m$
Define the difference of wind power output for time interval
$\{D_{training}(t)\} = D_1, D_2, D_3, \cdots, D_n$
$\{D_{test}(t)\} = D_{n+1}, D_{n+2}, D_{n+3}, \cdots, D_m$
For data scaling **do**
Min-max normalization
Until $i = m$
Output: $\{D_{training}(t)\}$,$\{D_{test}(t)\}$

2) WIND POWER OUTPUT FORECASTING

After the data input is completed, Step 2 follows, and Figure 5 shows the flow chart of the overall wind power output forecasting. During the training stage of the LSTM model, the LSTM-based forecasting model was implemented using the *Keras* library in R. An LSTM-based forecasting model performs optimization by repeating a process of sequentially passing through input, forget, update, and output gates. During the learning process of this model, optimization was performed using an Adaptive Moment Estimation (Adam) optimizer. The Adam optimizer is a general optimization technique, an algorithm that combines the strengths of Momentum and RMS Prop, and is most commonly used in deep learning with techniques that improve both the direction and size of learning. Consequently, the difference in the wind power output for the time interval was obtained, and the final wind power output was calculated as the sum of the difference in the wind power output over the time interval and the output 15 min prior.

3) RAMP EVENT FORECASTING

Finally, in Step 3, based on the forecasting results of the wind power output, the 2-hour ahead ramp rate is calculated. Ramp events were detected and categorized based on their definition. Error verification was performed using a confusion matrix, and the proposed metric used a heatmap to determine the forecasting accuracy for the size, direction, and timing of ramp events.

IV. SIMULATION RESULTS

This section analyzes the forecasting performance based on the results of the 2-hour ahead ramp forecasting. An LSTM-based model was implemented to forecast the wind power

Algorithm 2 Wind Power Forecasting using LSTM Model

Software & tools:R program
Input: $\{P_{test}(t)\}$, $D_{training}(t)\}$,$\{D_{test}(t)\}$
LSTM Modeling and Forecasting
Model Training
For building model and forecasting D_i^f $(i = n+1)$ **do**
$\{D_{forecasted}(t)\} = D_{n+1}^f, D_{n+2}^f, D_{n+3}^f, \cdots, D_m^f$
Inverse transform of $\{D_{forecasted}(t)\}$
Until $i = m$
For forecasting wind power output $(i = n+1)$ **do**
$P_{i+1}^f = P_i + D_{i+1}^f$
$\{P_{forecasted}(t)\} = P_{n+1}^f, P_{n+2}^f, P_{n+3}^f, \cdots, P_m^f$
Until $i = m$
Calculation of forecasting error (NMAE)
installed capacity = 60 (MW)
For calculating NMAE $(i = n+1)$ **do**
$Error_i = \left|P_{i+1}^f - P_{i+1}\right| \div installed\ capacity \times 100(\%)$
$\{NMAE\} = Error_{n+1}, Error_{n+2}, Error_{n+3}, \cdots, Error_m$
Output:
(1) $\{P_{forecasted}\}$: Forecasted Wind Power Output (.xlsx)
(2) $\{NMAE\}$: NMAE for LSTM Forecasting Model

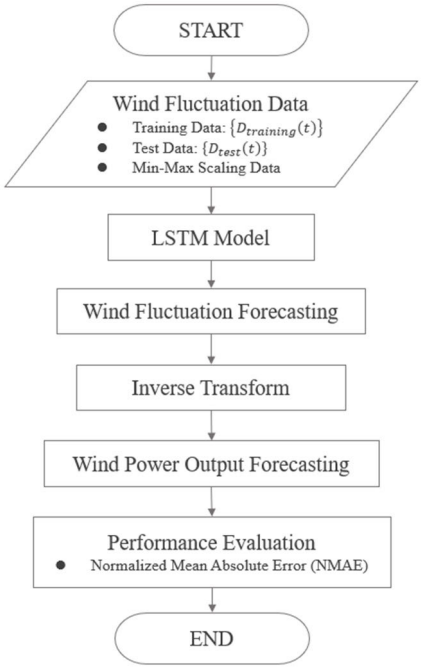

FIGURE 5. Flow chart of wind power output forecasting.

using empirical data from a wind farm in Jeju, Korea. Consequently, an indirect method for forecasting the ramp rate and ramp events was implemented by applying LSTM wind power forecasts. Finally, a confusion matrix and heatmap were utilized to verify the performance of the ramp-event detection. This section describes the results of the wind power and ramp event forecasting.

Algorithm 3 Ramp rate and ramp event forecasting and performance evaluation

Software & tools: Excel, Python
Input: $\{P_{test}(t)\}$, $\{P_{forecasted}(t)\}$ (.xlsx)
For $i = n + 1$ **do**
Actual ramp rate calculation
$R_i = P_{i+8} - P_i$
$\{R_{measured}(t)\} = R_{n+1}, R_{n+2}, R_{n+3}, \cdots, R_m$
Forecasted ramp rate calculation
$R_i^f = P_{i+8}^f - P_i^f$
$\{R_{forecasted}(t)\} = R_{n+1}^f, R_{n+2}^f, R_{n+3}^f, \cdots, R_m^f$
Performance evaluation using confusion matrix
Distinguish TP, FN, FP, TN by criteria
Count the number of points corresponding to each category
Performance evaluation using proposed matrix tool
Distinguish 16 categories by criteria
Count the number of points corresponding to each category
Until $i = m$
Tabulate confusion matrix
Visualize the proposed metric
Output:
(1) confusion matrix (format: table)
(2) proposed tool (format: heatmap)

A. WIND POWER OUTPUT FORECASTING

This section presents the wind power output forecasting results for Step 2 in Figure 5. Figures 6 and 7 show the forecasting results for July and December, which are the test data of Cases #1 and #2, respectively, as listed in Table 6, where the solid blue line is the actual measurement value, and the dotted red line is the forecasted value.

Figures 6 and 7 show the overall wind power output variability for the entire test dataset for each scenario. First, in July, the average output was approximately 18 MW, low output below 30 MW was dominant, and the overall trend was stable. Moreover, in December, the average output was approximately 30 MW, with a trend of repeating a low output power of 10 MW and a high output power of 50 MW. July, with low-power variability, and December, with high-power variability, followed the overall pattern of the measured values.

The histogram in Figure 8 displays the distribution of absolute error of two scenarios. In the Case #1 (labeled as "July" and depicted in red), there is a noticeable concentration of small forecasting errors, evident in both the bar and line graphs. On the other hand, the Case #2 (labeled as "December" and depicted in blue) shows an increasing frequency of occurrences as the forecasting error grows, beginning at an absolute error of approximately 2 MW. To summarize, both scenarios exhibit a predominant concentration of absolute errors within the 0 to 2 MW range. However, the forecasting for the July scenario (Case #1) appears to be superior.

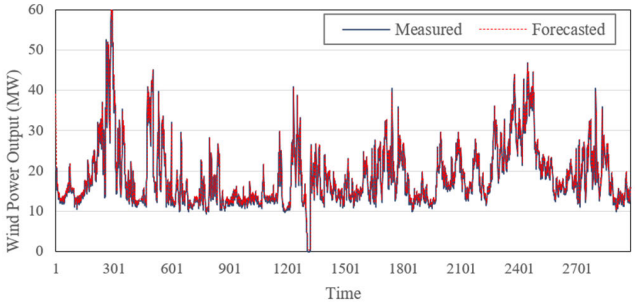

FIGURE 6. Wind power output forecasting in July.

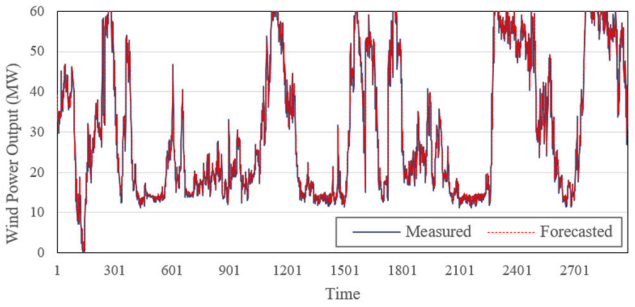

FIGURE 7. Wind power output forecasting in December.

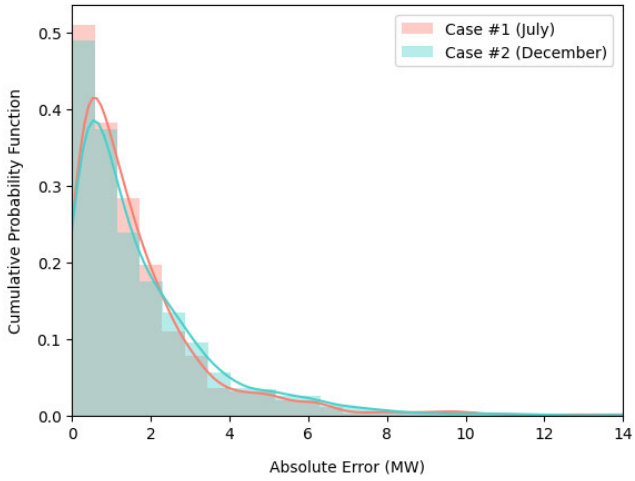

FIGURE 8. Histogram of wind power output forecasting error.

To verify the performance of the proposed model, the normalized mean absolute error (NMAE) was used to analyze the accuracy of wind power output forecasting. NMAE is the normalized value of the mean absolute error (MAE) and is suitable as an error indicator for datasets of different scales. NMAE is expressed in Equation (2) by dividing the percentage of MAE by the installed capacity [30].

$$NMAE(\%) = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\left(P_{forecasted}\right)_i - \left(P_{measured}\right)_i\right|}{Installed\ capacity} \times 100 \tag{2}$$

The explanation for the variables in the NMAE formula is as follows: N represents the total number of periods to be forecast, where is the number of points of the test data. $P_{forecasted}$ represents the forecasted value, $P_{measured}$ refers to the measured value, and the *installed capacity* is the installed wind power capacity of 60 MW.

This study forecasts the wind power output after 15 min, and the NMAE for individual forecasting in each case is summarized in Table 10. The forecast for July was 2.74% and that for December was 2.98%. For additional forecasting accuracy analysis, the number of points at which the difference between the actual and forecasted values exceeded 6 MW, which is 10% of the installed capacity of the wind farm, was investigated. In July (Case #1), there were 95 points with a difference between the forecasted and measured values of 6 MW or more for a total of 2,976 predicted points, and 96.81% of the total data showed an error of less than 6 MW. In December (Case #2), 123 points had absolute errors of over 6 MW, for a total of 2,976 points. Therefore, for all the points in the test data, 95.87% exhibited an error of less than 6 MW. The experimental results clearly indicate that in Case #1, the NMAE was lower and the proportion of the absolute error rate lower than 6 MW for wind power forecasting was higher than that in Case #2. These findings provide strong evidence that the predictions perform better in July.

TABLE 10. 15-minute ahead wind power forecasting.

Case Number	Training Data	Test Data	NMAE (%)
Case #1 : Low Volatile Case	June	July	2.74
Case #2 : High Volatile Case	November	December	2.98

B. RAMP RATE AND RAMP EVENT FORECASTING

In this section, a ramp event is detected and compared with a measured event using the forecasted wind power output verified in the previous section. Based on the wind power output value of 15 minutes, a 2-hour ahead ramp rate was calculated, and a ramp event was detected.

According to previous studies, the duration of ramp event is 30 minutes to 6 hours, and the range of thresholds is 10-30% in Table 2. In Table 4, the duration was set to 2 h, and the threshold was set to 6 MW, which is 10% of the installed capacity of a wind farm located on Jeju Island.

The results of ramp event forecasting after 2 h for each case are shown in the confusion matrices in Tables 11 and 12. In both scenarios, the total forecasting point was 2,968. For December, the numbers of TP and TN were higher, totaling 2,657, compared to July's counts of 2,549. In other words, the proposed method accurately forecasts whether a ramp event will occur in December. Larger values of ACC, CSI, POD, and FA can be judged to be better; therefore, even according to the evaluator's calculation results summarized in Table 13, the value of the evaluator in Case #2 is larger than Case #1, so the ramp event forecasting for December seems to be better.

TABLE 11. Confusion matrix of 2-hour ahead ramp forecasting in July.

Index		Ramp Forecasting		Total
		True	False	
Ramp Observation	True	433	219	652
	False	200	2,116	2,316
Total		633	2,335	2,968

TABLE 12. Confusion matrix of 2-hour ahead ramp forecasting in December.

Index		Ramp Forecasting		Total
		True	False	
Ramp Observation	True	598	139	737
	False	172	2,059	2,231
Total		770	2,198	2,968

TABLE 13. Evaluators of ramp forecasting of each case.

Evaluator	Case #1 (July) : Low Volatile Case	Case #2 (December) : High Volatile Case
ACC	0.86	0.90
CSI	0.51	0.66
POD	0.66	0.81
FA	0.67	0.77

Figures 9 and 10 show visualizations of the actual measurements and forecasting values of the ramp rate for all points, excluding true positives and true negatives in the confusion matrix. In Figures 9 and 10, (a) shows the wind power output, and (b) and (c) show that the forecasting results for the ramp rate are positive and negative, respectively. In graphs (b) and (c), green circles represent the measured ramp rate, red marks represent the forecasted ramp rate, and the blue dotted line indicates the threshold of the ramp event. There is a limitation in that it is difficult to determine whether the forecasting was successful on the graph. This study presents a metric to subdivide the categories in the existing confusion matrix and quantitatively and visually show the forecasting results effectively.

C. PROPOSED PERFORMANCE EVALUATION TOOL FOR RAMP EVENT FORECASTING

In this study, because the time points are forecasted for July and December, the ramp event forecasting performance can be compared using the proposed method without scaling. If the time points differ, the forecasting results can be compared by adjusting the timescale.

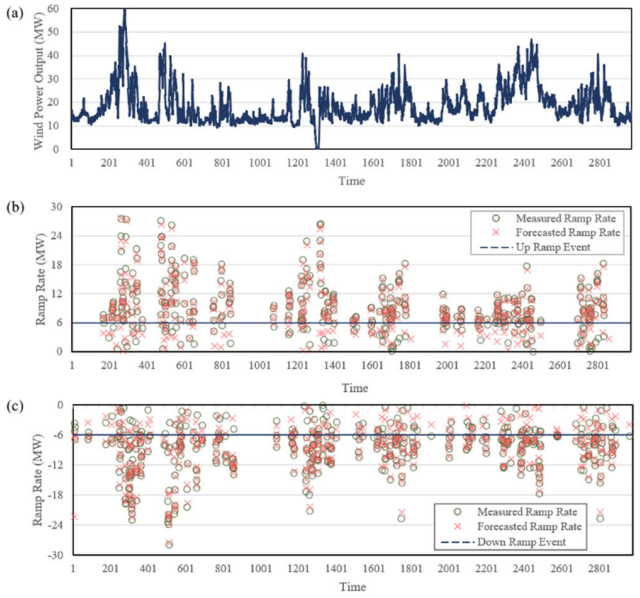

FIGURE 9. The result of 2-hour ahead ramp rate for July.

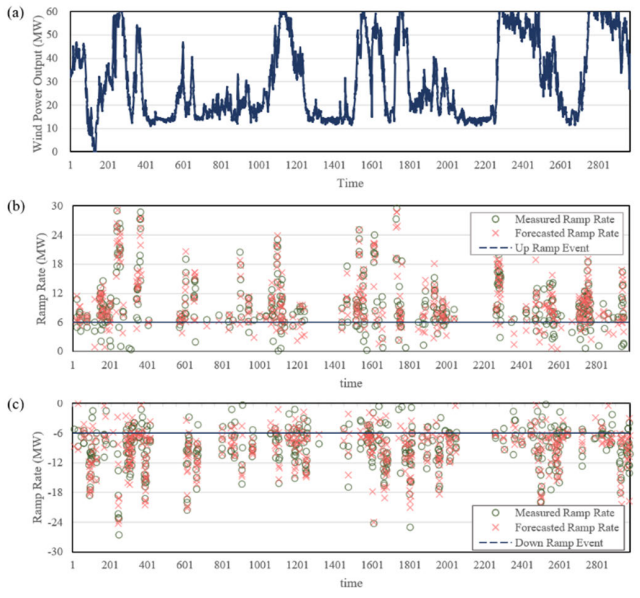

FIGURE 10. The result of 2-hour ahead ramp rate for December.

Figures 11 and 12 show the results of ramp-event forecasting for each scenario using the proposed method. The proposed metric distinguishes between the detection of up-ramp and down-ramp events among true positives, which indicates a forecasting ramp event. Figure 11 shows that the forecasting result of July corresponding to Case #1 has a lower frequency of detection of up-ramp and down-ramp events than that of December, corresponding to Case #2. Among the true negatives, July had many points at which a ramp event did not occur, but the proportion of up ramp rate was slightly high. In addition, false negatives and false positives were divided into four categories in each,

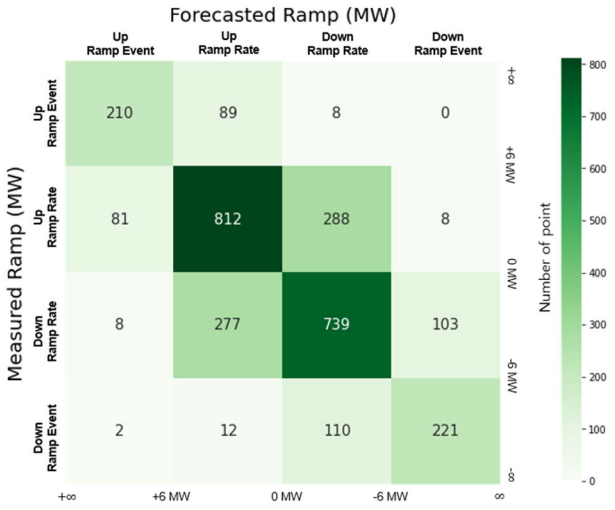

FIGURE 11. Proposed ramp forecasting evaluation metric for july.

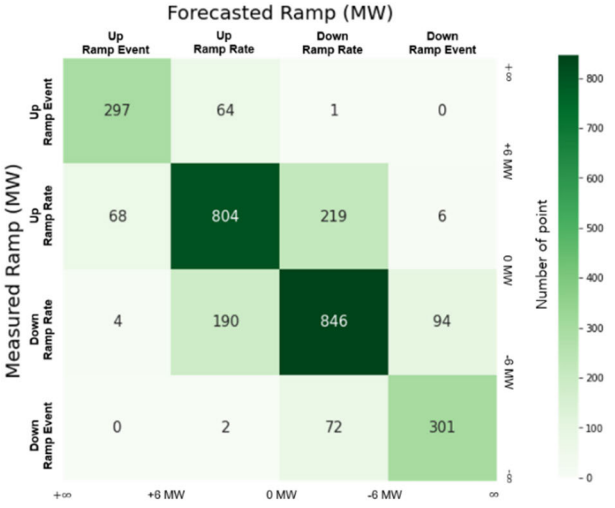

FIGURE 12. Proposed ramp forecasting evaluation metric for December.

and the number in all categories was higher in July, which means relatively bad forecasting comparing to December. In terms of color, the color of the right-down diagonal is dark in December, which can be judged to have a highly accurate prediction frequency. The color of the right-down diagonal is relatively bright in July, and it seems that there are many aggregations in areas other than the diagonal.

V. CONCLUSION

In this study, short-term wind power output forecasting was conducted based on an LSTM model, and a 2-hour ahead ramp event detection method was proposed. This study forecasted ramp events using an indirect method and compared the forecasting results for each case using a confusion matrix and representative evaluators. In addition, the forecasting results were analyzed using a performance evaluation

metric that visualized the results in color while considering the direction of the ramp rate and ramp events as well as the detection of ramp events. In the case study based on empirical data of wind farm located in Jeju Island, ramp event detection was performed in July with the lowest output volatility and December with the highest output volatility. Based on the wind power output forecasting, NMAE-based forecasting error for July (Case #1) was 2.74% and for December (Case #2) was 2.98%. In ramp event forecasting, the accurate detection of ramp events out of a total of 2,968 data points revealed 2,549 for Case #1 and 2,657 for December Case #2. Notably, in the ramp event forecasting assessment for Case #2, all evaluation metrics - ACC at 0.90, CSI at 0.66, POD at 0.81, and FA at 0.77 - surpass those of Case #1. Contrary to the wind power forecast, Case #2 exhibited a superior forecasting accuracy in detecting ramp events.

The proposed ramp-event forecasting is a short-term forecasting method that can be applied to various time scales and thresholds and can function as a method for wind power generation resources to be flexible in the power market, such as real-time curtailment and ancillary service. Specifically, the frequency of curtailment is experiencing a rapid escalation attributed to the sudden up-ramp event and transmission congestion in power systems with a high proportion of renewable energy, as observed on Jeju Island. Therefore, it is expected that the ramp event detection approach proposed in this paper can be useful in the preemptive response of curtailment. Furthermore, we expand the confusion matrix, a simple performance evaluation method, and propose a new performance evaluator using a heatmap, which is implemented to make it more granular and visually recognizable so that it can be appropriately used to compare various time scales or forecasting models. The proposed method will play a pivotal role to control the output variability of wind resources in advance and support power system operators and power generation companies in making appropriate decisions.

In the future, we will implement an analysis comparing the forecasting performance of NWP and OpSDA models, both of which are renowned for ramp event forecasting. For the case study on Jeju Island, we will proceed by training and fine-tuning these models. Subsequently, we will assess their performance using evaluation methods such as a confusion matrix and a heatmap based metric that we propose. Furthermore, we will utilize into the forecast spot prices and penalties/incentives for wind power energy in real-time renewable trading or auction markets. Consequently, the final goal is to improve the performance of wind forecasting by identifying trends or characteristics by season and time, conducting power flow analysis, and N-1 contingency analysis using PSS/E with the forecasted value as input. The main purpose of forecasting and power system analysis is to detect the probability of outages and blackouts and provide insight to enable an appropriate response in advance.

REFERENCES

[1] *Global Energy Transformation: A Roadmap to 2050*, IRENA, International Renewable Energy Agency, Abu Dhabi, 2018, pp. 8–15.

[2] Korea Energy Economics Institute, "Energy focus," *Spring Issue Energy Focus*, vol. 83, pp. 28–35, May 2021.

[3] S. Impram, S. V. Nese, and B. Oral, "Challenges of renewable energy penetration on power system flexibility: A survey," *Energy Strategy Rev.*, vol. 31, pp. 100539–100550, Aug. 2020.

[4] C. Gallego-Castillo, A. Cuerva-Tejero, and O. Lopez-Garcia, "A review on the recent history of wind power ramp forecasting," *Renew. Sustain. Energy Rev.*, vol. 52, pp. 1148–1157, Dec. 2015.

[5] Y. Cao, W. Wei, C. Wang, S. Mei, S. Huang, and X. Zhang, "Probabilistic estimation of wind power ramp events: A data-driven optimization approach," *IEEE Access*, vol. 7, pp. 23261–23269, 2019.

[6] D. Zhang, H. Zhang, X. Zhang, X. Li, K. Ren, Y. Zhang, and Y. Guo, "Research on AGC performance during wind power ramping based on deep reinforcement learning," *IEEE Access*, vol. 8, pp. 107409–107418, 2020.

[7] G. D. Amico, F. Petroni, and S. Vergine, "Ramp rate limitation of wind power: An overview," *Energies*, vol. 15, no. 16, pp. 4–6, Aug. 2022.

[8] J. Li, T. Song, B. Liu, H. Ma, J. Chen, and Y. Cheng, "Forecasting of wind capacity ramp events using typical event clustering identification," *IEEE Access*, vol. 8, pp. 176530–176539, 2020.

[9] G. W. Chang, H. J. Lu, Y. R. Chang, and Y. D. Lee, "An improved neural network-based approach for short-term wind speed and power forecast," *Renew. Energy*, vol. 105, pp. 301–311, May 2017.

[10] Y. Fujimoto, Y. Takahashi, and Y. Hayashi, "Alerting to rare large-scale ramp events in wind power generation," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 55–65, Jan. 2019.

[11] C. Gallego-Castillo, E. Garcia-Bustamante, A. Cuerva, and J. Navarro, "Identifying wind power ramp causes from multivariate datasets: A methodological proposal and its application to reanalysis data," *IET Renew. Power Gener.*, vol. 9, no. 8, pp. 867–875, Nov. 2015.

[12] R. C. Fong, W. J. Scheirer, and D. D. Cox, "Using human brain activity to guide machine learning," *Sci. Rep.*, vol. 8, no. 1, pp. 5386–5397, Mar. 2018.

[13] Y. Cui, Z. Chen, Y. He, X. Xiong, and F. Li, "An algorithm for forecasting day-ahead wind power via novel long short-term memory and wind power ramp events," *Energy*, vol. 263, pp. 1–12, Jan. 2023.

[14] L. Han, Y. Qiao, M. Li, and L. Shi, "Wind power ramp event forecasting based on feature extraction and deep learning," *Energies*, vol. 13, no. 23, pp. 6449–6466, Dec. 2020.

[15] U. Cali and V. Sharma, "Short-term wind power forecasting using long-short term memory based recurrent neural network model and variable selection," *Int. J. Smart Grid Clean Energy*, vol. 8, no. 2, pp. 103–110, 2019.

[16] J. Zhang, M. Cui, B.-M. Hodge, A. Florita, and J. Freedman, "Ramp forecasting performance from improved short-term wind power forecasting over multiple spatial and temporal scales," *Energy*, vol. 122, pp. 528–541, Mar. 2017.

[17] L. Cornejo-Bueno, L. Cuadra, S. Jiménez-Fernández, J. Acevedo-Rodríguez, L. Prieto, and S. Salcedo-Sanz, "Wind power ramp events prediction with hybrid machine learning regression techniques and reanalysis data," *Energies*, vol. 10, no. 11, pp. 1784–1811, Nov. 2017.

[18] J. W. Taylor, "Probabilistic forecasting of wind power ramp events using autoregressive logit models," *Eur. J. Oper. Res.*, vol. 259, no. 2, pp. 703–712, Jun. 2017.

[19] L. Cornejo-Bueno, C. Camacho-Gómez, A. Aybar-Ruiz, L. Prieto, A. Barea-Ropero, and S. Salcedo-Sanz, "Wind power ramp event detection with a hybrid neuro-evolutionary approach," *Neural Comput. Appl.*, vol. 32, no. 2, pp. 391–402, Jan. 2020.

[20] D. Lyners, H. Vermeulen, and M. Groch, "Wind power ramp event detection using a multi-parameter segmentation algorithm," *Energy Rep.*, vol. 7, pp. 5536–5548, Nov. 2021.

[21] M. Okada, K. Yamaguchi, R. Kodama, N. Ogasawara, H. Kato, V. Q. Doan, N. N. Ishizaki, and H. Kusaka, "Development of a wind power ramp forecasting system via meteorological pattern analysis," *Wind Energy*, vol. 25, no. 11, pp. 1900–1916, Aug. 2022.

[22] F. Shahid, A. Zameer, and M. Muneeb, "A novel genetic LSTM model for wind power forecast," *Energy*, vol. 223, pp. 1–10, May 2021.

[23] Y. Liu, L. Guan, C. Hou, H. Han, Z. Liu, Y. Sun, and M. Zheng, "Wind power short-term prediction based on LSTM and discrete wavelet transform," *Appl. Sci.*, vol. 9, pp. 1–15, Mar. 2019.

[24] B. Zhou, X. Ma, Y. Luo, and D. Yang, "Wind power prediction based on LSTM networks and nonparametric kernel density estimation," *IEEE Access*, vol. 7, pp. 165279–165292, 2019.
[25] B. Shao, D. Song, G. Bian, and Y. Zhao, "Wind speed forecast based on the LSTM neural network optimized by the firework algorithm," *Adv. Mater. Sci. Eng.*, vol. 2021, pp. 1–13, Sep. 2021.
[26] L. Bianco, I. V. Djalalova, J. M. Wilczak, J. Cline, S. Calvert, E. Konopleva-Akish, C. Finley, and J. Freedman, "A wind energy ramp tool and metric for measuring the skill of numerical weather prediction models," *Weather Forecasting*, vol. 31, no. 4, pp. 1137–1156, Aug. 2016.
[27] M. Heydarian, T. E. Doyle, and R. Samavi, "MLCM: Multi-label confusion matrix," *IEEE Access*, vol. 10, pp. 19083–19095, 2022.
[28] A. Vanacore, M. S. Pellegrino, and A. Ciardiello, "Fair evaluation of classifier predictive performance based on binary confusion matrix," *Comput. Statist.*, pp. 1–21, Nov. 2022.
[29] Y. Liu, Y. Sun, D. Infield, Y. Zhao, S. Han, and J. Yan, "A hybrid forecasting method for wind power ramp based on orthogonal test and support vector machine (OT-SVM)," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 451–457, Apr. 2017.
[30] G. G. Platero, M. G. Casado, M. P. Garcìa, P. J. Madero, and D. A. Baeza, "CECRE: Supervision and control of Spanish renewable energies in the last 15 years," *J. Modern Power Syst. Clean Energy*, vol. 10, no. 2, pp. 269–276, Mar. 2022.

SOLUI YU (Student Member, IEEE) received the B.S. degree in climate and energy systems engineering from Ewha Womans University, South Korea, in 2023, where she is currently pursuing the master's degree with the Department of Climate and Energy Systems Engineering. Her current research interests include short-term wind power forecasting and power economics.

JIN HUR (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, South Korea, in 1997 and 1999, respectively, and the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin, in 2012. He is currently an Associate Professor with the Department of Climate and Energy Systems Engineering, Ewha Womans University. His current research interest includes integrating high levels of variable generating resources into electric power systems.