On wavelet transform based convolutional neural network and twin support vector regression for wind power ramp event prediction

Dhiman, Harsh S.; Deb, Dipankar; Guerrero, Josep M.

Published in:
Sustainable Computing: Informatics and Systems

DOI (link to publication from Publisher):
10.1016/j.suscom.2022.100795

Publication date:
2022

Document Version
Early version, also known as pre-print

On Wavelet Transform based Convolutional Neural Network and Twin Support Vector Regression for Wind Power Ramp Event Prediction

Harsh S. Dhiman*[a], Dipankar Deb[b], Josep M. Guerrero[c]

[a]*Department of Electrical Engineering, Adani Institute of Infrastructure Engineering, Ahmedabad, India-382421*
[b]*Department of Electrical and Computer Science Engineering, Institute of Infrastructure Technology Research and Management (IITRAM), Ahmedabad, India 380026.*
[c]*Department of Energy Technology, Aalborg University, 9220 Aalborg East, Denmark*

Abstract

Power produced from renewable energy sources carbon negative and promises an increased reliability for grid integration. Wind energy sector globally has an installed capacity of over 650 GW and promises to grow substantially. In this work, wind power ramp events that arise from sudden change in wind power are studied. Forecasting wind power ramp events is an important problem statement in the current wind power industry. Wind power integration to utility grid is impacted by the forecast accuracy. To improve the reliability and security, wind speed and power forecasts are extensively studied. Ramp events are predicted in this manuscript using hybrid techniques employing wavelet decomposition transform in tandem with convolutional neural network and twin support vector machines. Wind speed datasets from Spain, Germany and Argentina are considered and error metrics are computed. It is observed that CNN based method is 28.76% and 26.43% superior from wavelet based random forest and TSVR method.

Keywords: Convolutional neural network, Log-energy entropy, Wind power, Ramp event, Random forest, Wavelet transform

*Corresponding author
Email addresses: harsh.dhiman@aii.ac.in (Harsh S. Dhiman*), dipankardeb@iitram.ac.in (Dipankar Deb), joz@et.aau.dk (Josep M. Guerrero)

1. Introduction

Rapid development in modern infrastructure has resulted in increased energy demands [1, 2]. Coal-based power plants have resulted in a severe environmental deterioration in the form of greenhouse emissions. Wind energy has received genuine interest from investors looking into the benefits of negligible fuel cost among renewables. Globally, wind energy installations have crossed more than 650 GW [3]. Although wind energy systems offer several advantages, the inherent randomness in its flow results in an increased generation cost and marginal power system reliability [4]. The randomness in wind speed can further introduce complexities in optimal power dispatch and lead to congestion problems in interconnected systems. Therefore, it is essential to study and incorporate accurate wind speed forecasting for enhanced power system stability. Forecasting studies in wind energy systems pertain to three broad aspects: wind power (and ramp) forecasting, and uncertainty analysis. Wind forecasting's essence is in the critical weather information available via numerical weather prediction (NWP) data. The advancements in statistical and data-driven models have become significantly handy for power forecasting at different time horizons.

Among statistical models, the prominent ones in usage include (i) autoregressive (AR) models [5], (ii) autoregressive integrated moving average (ARIMA), and (iii) autoregressive conditional heteroscedastic (ARCH) models. Data-driven models essentially rely on the availability of a large amount of historical data where machine learning methods like Support vector regression (SVR) [6, 7], Artificial neural networks [8, 9]. Extreme learning machine [10] are actively used. Neural networks also find usage in wind forecasting with diverse configurations explored in terms of the layers [11, 12]. Further, Long short-term memory (LSTM) and Elman neural network (ENN) based deep learning models are useful with signal processing methods like Fourier transform, wavelet transform, and empirical mode decomposition [13]. Overall, a broad spectrum of forecasting is available, but the current trends are focused on increased short-term accuracy [14].

In this manuscript, we focus on the prediction of wind power ramps using wavelet transform (WT) based convolutional neural network and twin support vector regression. The ability of WT to decompose the time-series results in fine-tuned features, helping the training process. The ramp events modeled for a wind farm using a classification and regression-based approach. In the classification-based method, we predict the label corresponding to a class of ramp events. In a regression-based system, we forecast the wind time-series corresponding to a wind farm site and the ramp event occurrence based on the set threshold. Wind power ramps are defined as sudden changes in wind power production, and mathematically, expressed as

$$|P_{t+\Delta t} - P_t| > P_{\text{threshold}}, \tag{1}$$

where Δt is the time interval considered for ramp events and P_t denotes the power production at time t. Ramp event definition is also affected by the threshold level ($\mathrm{P}_{threshold}$). Usually in large wind farms this threshold varies from 5%-15% of the installed capacity. Ramp-up events occur when there is a sudden increase in wind power and ramp-down occurs when there is sudden fall in the wind power produced. The fundamental concept of time-series is applicable to wind power as well where the non-stationary wind speed ensure that the standard statistical methods like ARMA and ARIMA fail. In [15], various time-series based multivariate and univariate methods are discussed for wind power ramp events. A data-mining based approach is used where SVM outperformed MLP, random forest [16, 17, 18] and decision tree. To improve the prediction performance, various decomposition algorithms such as empirical mode decomposition (EMD) [19], variational mode decomposition (VMD) [20] are actively used. In [21], authors have explored the effect of training data on machine learning and deep learning algorithms. The wind speed and ramp events are forecasted based on random forest technique and compared with TSVR. In [22], authors use EMD and VMD to decompose wind power time-series to various intrinsic mode functions (IMFs). Convolutional neural network and Long short-term memory networks are used to extract features and the predicted sub-series

is aggregated to arrive at final predictions. In [23], detection and statistical analysis of ramp events is discussed. For optimal detection of ramp events, algorithms like dynamic programming, trend fitting, ramp score function, and sliding window approaches are discussed. In [24], statistical models are wind power ramp events are discussed where the probabilistic approach is leveraged to model the ramp scenarios. Wind power data from Bonneville Power Administration (BPA) is collected for 2 years (Jan 01, 2005-Dec 31, 2006) sampled at 30 second interval. In [25], auto-regressive hidden Markov model based short-term wind speed and ramp event forecasts for time horizon 0.5-6 hours is carried out. In [26], a regime-switching neural network (RS-NN) is proposed that decides the neural network architecture based on the wind power regime under consideration. During ramp-down events, it is observed that RS-NN outperformed standard neural network by a percent.

Wang et al. developed a short-term wind speed forecasting using CNN model which is based on data collected by 3 neighbouring wind farms in China. Transfer learning is also applied to resolve the issue of data shortage of newly constructed wind farms and in medical image analysis [27]. Transfer CNN has given better results compared to KRR, SVR and simple CNN [28]. Zhen et al. proposed a hybrid deep learning wind forecasting model based on temporal-Spatial feature selection. Bi-LSTM-CNN has a best accuracy of 0.9929 even with frequently changing wind power, and CNN has the highest average computation speed of 0.0741 [29]. Jaseena et al. developed a decomposition-based forecasting model using bidirectional LSTM networks which outperforms other models. However, the model works on univariate LSTM networks, while excluding the correlated features in the input features [30]. Shahid et al. have proposed a novel genetic long short term memory (GLSTM) framework which includes long short term memory and genetic algorithm (GA). Results shows that comparison of GSLSTM give comparable more accuracy as there is less computational complexity [31]. Wu et al. have proposed a novel correlation model for ultra-short-term prediction (STCM) based on convolutional neural networks-long short-term memory (CNN-LSTM) evaluated using MAE, MAPE, RMSE

and NRMSE. Results indicate more reliability over traditional structure models [32]. In [33], the authors discuss an advanced deep learning framework based on characterizing the probability distribution function (PDF) of a wind farm site. The proposed method is compared with CNN+GRU, CNN-GRU+KDE for wind farm sites in England and Iran. Results reveal a median MAPE of approximately 8% for both wind farm sites. In [34], authors discuss an ensemble of deep learning LSTM architecture on the input wind speed time-series. This ensemble architecture is followed by a final layer of SVR where the hyperparameters of SVR are optimized via external optimization, and the forecasting results are validated for a site in China.

The main objective of this manuscript is to understand and analyze the ramp event prediction based on hybrid wavelet-CNN and wavelet-TSVR based regression techniques. The size of the training data is varied to check the best possible regressor among the tested ones. Consider the significance of historical data in wind speed forecasting. Henceforth in wind power ramp events, the length of the training data is varied to assess the best regression technique. Since a significant autocorrelation exists in the wind speed time series, the dependency on the historical data is vital for formulating the forecasting task. Further, for two signal pre-processing techniques, the inherent randomness is checked by determining the log-energy entropy.

The remaining sections of this work are as follows: Section 2 presents machine-learning and deep-learning based regression. In Section 3, the datasets and methodology is explained followed by results and discussion in Section 4. Conclusions are presented in Section 5.

2. Methods for Wind Power Ramp Prediction

For a training set $T = \{(x_i, y_i) : x_i \in \mathbb{R}^n, y_i \in \mathbb{R},\ i = 1, 2..., l\}$, SVR models minimize a linear combination of the loss function and regularization

term, $f(x) : w^T x + b,\ w \in \mathbb{R}^n, b \in R$. Nonlinear function estimation uses $f(x) : K(x^T, A^T)u + b$ with the Kernel K fulfilling the Mercer criterion [35].

The ϵ-SVR model minimizes a C-insensitive loss function ignoring an error up to ϵ, along with the $\frac{1}{2}w^T w$ regularization to solve the optimization problem

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{l} |y_i - f(x_i)|_\epsilon, \tag{2}$$

where $|y_i - f(x_i)|_\epsilon = \max(0, |y_i - f(x_i)| - \epsilon)$ is the loss function. Solving the ϵ-SVR problem (2) with slack variables κ_i and κ_i^* for $i = 1, 2, \ldots, l$, as

$$\begin{aligned} &\min_{w,b,\kappa,\kappa^*} \quad \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{l} (\kappa_i + \kappa_i^*) \\ &\text{subject to} \\ &y_i - (A_i w + b) \leq \epsilon + \kappa_i, \quad (A_i w + b) - y_i \leq \epsilon + \kappa_i^*,\ \kappa_i, \kappa_i^* \geq 0. \end{aligned} \tag{3}$$

Twin Support Vector Regression (TSVR) [36] derives two non-parallel hyperplanes close to one of the classes and farthest as possible from each other. TSVR solves two QPPs with a reduced computation time and complexity compared to quadratic programming problems (QPP). The mathematical formulation is

$$\begin{aligned} \min\ \frac{1}{2}(Y - e\varepsilon_i - (Aw_i + eb_i))^T (Y - e\varepsilon_i - (Aw_i + eb_i)) + C_i e^T v_i \\ \textbf{s.t. } Y - (Aw_i + eb_i) \geq e\varepsilon_i - v_i, v_i \geq 0,\ i = 1, 2, \end{aligned} \tag{4}$$

for parameters $C_i, \varepsilon_i \geq 0$ and slack variables v_i. More details of the dual QPPs and KKT conditions are available [36]. In non-linear regressions, kernel technique helps solve the QPP by determining the average of $K(x^T, A^T)w_1 + b_1$ and $K(x^T, A^T)w_2 + b_2$.

Long Short-Term Memory Recurrent Neural Networks contains cyclic connections that make them more proficient than an ordinary feed-forward neural network [37]. The LSTM contains a special block known as the memory blocking the recurrent hidden layer. These blocks contain memory cells capable of storing the time-related condition of the system in addition to a special unit called gates to control the progression of data. Each memory block has an input and an output gate. The input gate runs the information stream enactment into the

memory cell. The output gate runs the output information stream enactment towards the remainder of the LSTM. The forget gate entryway scales the inside condition of the cell before contributing it as a contribution to the cell through the self-intermittent association of the cell, thereby adaptively overlooking or resetting the cells' memory. LSTM networks are appropriate for classification, processing, and generating empirical results based on time-series data since there can be slacks of the obscure term between significant occasions in time-arrangement information. Deep LSTM RNN's are built by stacking numerous LSTM layers. This explicitly implies that the input of the network at a given timestamp propagates through many LSTM layers. Deep LSTM RNNs have an edge over the normal LSTM RNN, they can improve the use of boundaries by circulating them over the space through various layers. Instead of expanding the memory size of the standard model by a factor of two, one can have four layers around a similar number of boundaries. These outcomes in input experiencing more non-straight activities per time steps.

3. Datasets & Framework of prediction models

The terrain is a significant factor in wind speed variation in wind farms. The wind farms typically facilitates a logarithmic wind speed profile in terms of height. In this study, all the turbines have a hub height of 90 meters and a rotor diameter of 120 meters. The surface roughness length(z_0) for onshore and offshore and hilly terrains are 0.005m, 0.0002m, and 1m, respectively. A combination of wavelet transform technique (for time-series defragmentation into appropriate sub-signals) and machine learning/deep learning algorithms fortify the existing hybrid prediction model. The collection of wind speed datasets happens at three different periods, i.e., three months, 12 months, and 36 months at the height of 10m over 10 minutes. Use of logarithmic law [38] for transforming the available wind speed to a 90m hub height would be desirable to get accurate wind power ramp events. Table 1 depicts the site coordinates and statistical parameters for datasets of Sotavento (onshore), Veja Mate (offshore),

and Madryn. The choice of datasets for wind speed is based on the different topological features of the farm site. It is well established fact that wind speed changes with terrain and so as to validate the proposed wavelet-based CNN and TSVR model, the wind farm sites are therefore from onshore, offshore and hilly terrain.

Table 1: Wind speed statistics for farm sites

Dataset	Mean value	Standard deviation
Sotavento, Spain (**A**)	3.245	1.781
Veja Mate, Germany (**B**)	3.433	2.182
Madryn, Argentina (**C**)	3.651	2.238

Denoting a ramp threshold $\beta_{th}\%$ of the rated wind power (P_{rated}), we get

$$\Delta P_w^{ramp} = \begin{cases} +\beta_{th}\% \text{ of } P_{rated} = P^u, \\ -\beta_{th}\% \text{ of } P_{rated} = P^d, \end{cases} \tag{5}$$

where P^u and P^d threshold levels depict up and down ramp events respectively. With a threshold level of 10%, we consider rated (12m/sec) wind speed. We find the wind power for the given datasets for a suitable ramp forecast when P^d or P^u get violated. The performance metrics for the testing samples are

$$\text{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i - p_i)^2\right]^{1/2} \tag{6}$$

$$\text{U1} = \frac{\sqrt{\frac{1}{n}\times\sum_{i=1}^{n}\left(\hat{p}_i - p_i\right)^2}}{\left(\sqrt{\frac{1}{n}\times\sum_{i=1}^{n}p_i^2} + \sqrt{\frac{1}{n}\times\sum_{i=1}^{n}\hat{p}_i^2}\right)} \tag{7}$$

$$\text{U2} = \frac{\sqrt{\frac{1}{n}\times\sum_{i=1}^{n}\left(\left(p_{i+1} - \hat{p}_{i+1}\right)/p_i\right)^2}}{\sqrt{\frac{1}{n}\times\sum_{i=1}^{n}\left(\left(p_{i+1} - \hat{p}_i\right)/p_i\right)^2}}, \tag{8}$$

where $\hat{p}_i, p_i, \bar{p}$ are the forecasted, measured and average values. Fourier transform, Wavelet transforms (WT), and Wavelet packet decomposition methods find widespread usage. A wavelet filter named Daubechies 4 decomposes wind speed time-series into low and high frequency components using Wavelet trans-

form [1]. Out of these, all the five detail signals along with 5^{th} approximate signal is input.

Literature suggests that hybrid methods yield superior forecasting performance compared to benchmark models [39]. Signal processing techniques, along with machine learning methods, improve forecast accuracy by eliminating stochastic variations. When used in wind speed decomposition, signal transforms like Fourier transform cause loss of information concerning time scale and gets overcome by WT that captures signal information both in frequency and time scale. Figure 1 shows the WT process involving five levels of decomposed approximation signals (low-frequency parts). The other signals at each decomposition stage are detail signals (D1, D2, ..., D5) with high-frequencies) constituting the feature vector for each of sampled wind speed shown in Figure 2 as the target value for the forecasting algorithms (SVR and its variants).

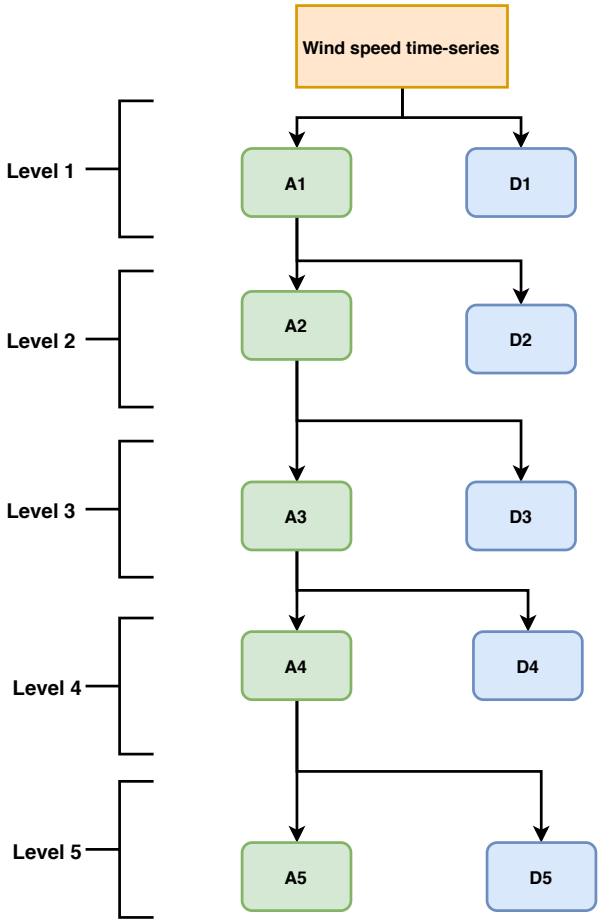

Figure 1: DWT representation [39]

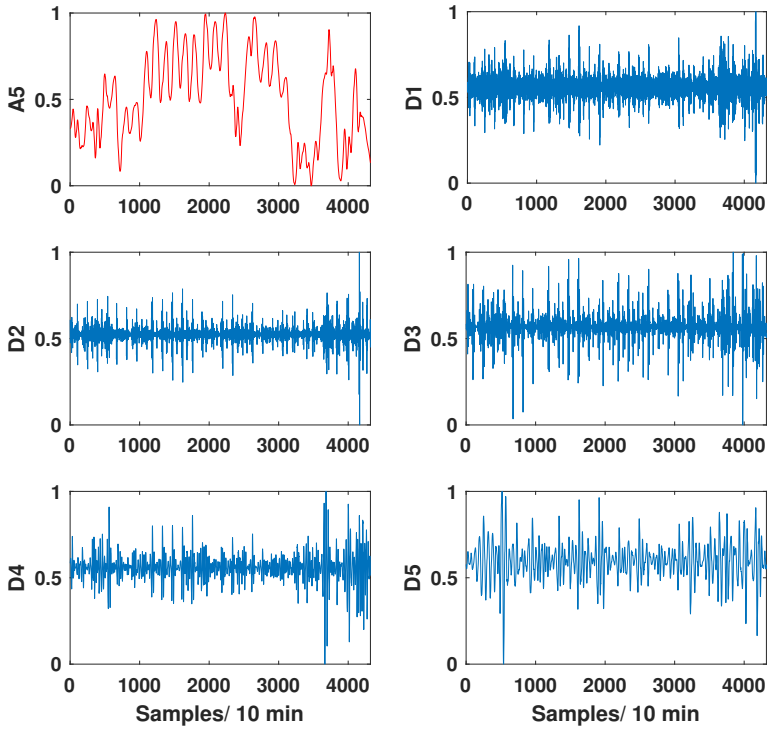

Figure 2: Normalized features for dataset A

Figure 3 illustrates WT-based ramp event prediction and depicts the procedure to quantify wind ramp randomness using log-energy entropy. The wind speed data is collected in form of a time-series and is checked thoroughly for any potential missing value. This stage is known as data cleaning where all the feature variables are also normalized in the range [0,1]. After this, the wind power is computed using the available wind speed and a ramp threshold of $\beta_0 = 10\%$ is set. The wind power is then compared with the set thresholds (upper and lower) as per equation (5). If the condition as mentioned in flowchart is satisfied, the instance is counted as a ramp event. The sign of ΔP_w^{Ramp} indicates the type of ramp event where the positive sign means ramp-up event and negative sign means ramp-down. The event prediction is carried out for two sampling intervals, that is, Δt=10 minutes and Δt=1 hour. The errors for the respective ramp events are calculated based on different types of forecasting methods

as discussed in next section. Common error metrics used in wind speed forecasting include Root mean squared error (RMSE), Mean squared error (MSE). However, for predicting ramp-up and ramp-down events, we have used R^{up} and R^{down} which are basically the absolute errors at the respective ramp events. The individual error metrics are used to establish a common ground for comparing different forecasting models. Algorithm 1 refers to a psuedocode of the proposed algorithm employing wavelet-CNN method.

Data: Detect and predict ramp events in wind farms
Result: Prediction of ramp-up and ramp-down events
initialization set ramp threshold β, calculate wind power $P = \frac{1}{2}\rho A v^3$;
while *Scan for each wind speed sample in series* $X_i = [x_1, x_2, \ldots, x_N]$ *where* $i = 1, 2, \ldots, N$ **do**
    instructions;
    if $\Delta P = P(i+1) - P(i) \geq P_{threshold}$ **then**
        Ramp-up event detected;
    end
    if $\Delta P = P(i+1) - P(i) \leq P_{threshold}$ **then**
        Ramp-down event detected;
    end
    Apply wavelet transform with Daubechies 4 filter and 5-level decomposition;
    Pass the approximate and detail signals to the CNN algorithm for prediction;
end

Algorithm 1: Wavelet-CNN algorithm for ramp event prediction

Further, the inherent stochasticity or randomness in wind power time-series is also computed using log-energy entropy. Log-energy entropy is often used to determine the randomness in a signal. In this manuscript, we have used the concept of log-energy entropy to address the decomposition effectiveness of wavelet transform and empirical mode decomposition. Both of the signal decomposition techniques are widely used in data pre-processing for building

machine learning and deep learning models.

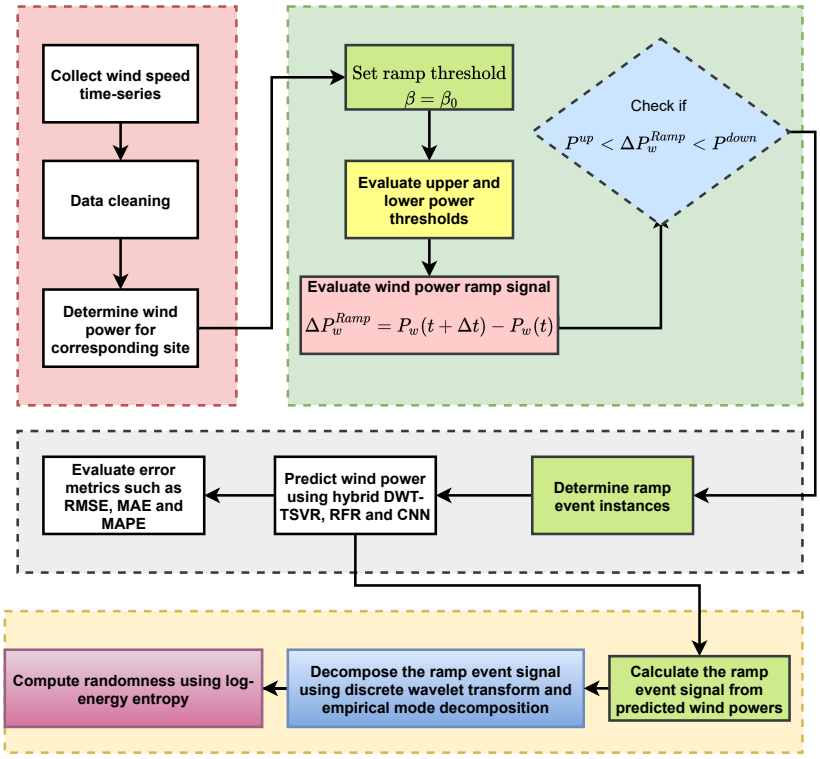

Figure 3: Block diagram for wind power ramp event prediction using wavelet transform

4. Results and Discussions

The prediction performance for methods like convolutional neural networks, random forest, and twin support vector machines is discussed. Wind speed datasets from Spain, Germany, and Argentina are considered. First, the time-series is converted into wind power series. The ramp events (up and down) are deciphered with a 10% threshold of the turbine-rated power. Wind power is calculated considering a wind turbine diameter of 112 meters at 90 meters of hub height. Wind power prediction is carried out by dividing the available data into two sets: training (80%) and testing (20%). Further, to avoid data-leakage, the training data is subjected to 10-fold cross-validation. The hyper-parameter tuning enhances the prediction performance. For this purpose, the bandwidth

of kernel function and regularization parameter C of TSVR is tuned in the set $[2^{-10}, 2^{-9}, ...2^{9}, 2^{10}]$ using a grid-search method. Similarly, for the random forest, we perform the tuning in a range of $(1 - 1000)$ number of trees. **In case of wavelet-CNN, after wavelet decomposition of the wind power ramp signal, two convolution layers with 64 filters each and a kernel size of 3 is used followed by one max-pooling layer and one drop-out layer. This means that convolutional layers helps in extracting meaningful and new features. While the pooling layer reduces the dimensions of incoming feature vector. In case of our work, we have tuned the hyper-parameters such as epochs and batch size in the range (0-300) and (2-30) respectively.**

In this analysis, the wind speed datasets are sampled at 10-minute intervals. The prediction performance is depicted in Table 2 for training data of 12 months and 36 months for wind farm sites in Spain and Germany. The inputs are the decomposed signals obtained from WT. Discrete WT with Daubechies filter (db4) is used, and the resultant approximate and detail signals are used as feature variables in the prediction process. The features are the column vectors in this analysis. The benefit of using discrete WT is its ability to capture the information in frequency and time scale. It is observed that wavelet-CNN outperforms wavelet-RF and wavelet-TSVR in terms of error metrics such as RMSE. Further, the ramp events for each wind farm site are analyzed separately.

For each type of event, the wavelet-CNN-based method provides higher accuracy when the length of training data is 36 months. Figure 4 illustrates the forecasting diagram for wind farm site in Spain. Wavelet-CNN based hybrid method outperforms the SVR and TSVR based hybrid method. In terms of the prediction performance, TSVR based method outperforms RF and CNN for all the datasets when the training data length is 3 months. In Figure 5, the ramp-event signal for the wind farm site in Germany is visualized. The ramp signal shows the variation in wind power produced for three different time frame of data. It is worthwhile to note that the variation is minimal for a 3-month data and is maximum for 36-month data. The input given to the prediction model

when given for a larger training data is seen evidently in the prediction performance for wavelet-CNN model. Further, with increase in the training data size to 36 months, CNN based method outperforms TSVR and RF methods.

Table 2: Performance metrics for wind farm sites

Dataset	Months	Model	RMSE (%)	MSE	R^{up}	R^{down}	U1	U2	CPU time (sec)
	3		0.12625	0.0159	0.09532	0.11112	0.01605	0.00081	152.5
	12	TSVR	0.1375	0.01890	0.23726	0.14012	0.00027	0.0.002325	266
	36		0.09832	0.0096	0.14063	0.10375	0.000065	0.0009	864.75
	3		0.2825	0.07980	0.21632	0.24552	0.0000525	0.000325	280.25
A	12	RFR	0.2925	0.08555	0.2625	0.2955	0.01827	0.25	517.75
	36		0.3025	0.09150	0.28077	0.329	0.011325	0.3451	801.75
	3		0.9725	0.9457	0.3287	0.35052	0.000525	0.00575	257.5
	12	CNN	0.5975	0.3570	0.35262	0.3977	0.000075	0.00115	475.25
	36		0.275	0.0756	0.2578	0.27877	0.001	0.01587	751.75
	3		0.1925	0.0370	0.2227	0.2303	0.01835	0.2692	307.75
	12	TSVR	0.2275	0.05175	0.2481	0.2754	0.00575	0.2187	557.75
	36		0.23	0.0529	0.25795	0.2990	0.003225	0.1713	938.5
	3		0.2425	0.0588	0.2403	0.2480	0.001175	0.00635	282.75
B	12	RFR	0.2925	0.0855	0.2528	0.2755	0.01467	0.16667	504.25
	36		0.32	0.1024	0.2907	0.29842	0.007425	0.1316	775
	3		1.995	3.9800	0.3051	0.3290	0.000775	0.036	258.25
	12	CNN	0.1525	0.0232	0.24335	0.2279	0.00035	0.01222	480.25
	36		0.1375	0.01890	0.2479	0.268025	0.000005	0.0021	776.25
	3		0.1775	0.03150	0.2227	0.2303	0.000235	0.03427	287.75
	12	TSVR	0.185	0.03422	0.2418	0.2542	0.001025	0.1604	525.25
	36		0.155	0.02402	0.25795	0.2990	0.00065	0.1291	778.75
	3		0.2775	0.0770	0.2381	0.24575	0.00025	0.00121	298
C	12	RFR	0.3625	0.1314	0.2803	0.3278	0.004125	0.0031	502.25
	36		0.405	0.1640	0.2684	0.27065	0.00392	0.00285	825.25
	3		0.2125	0.04515	0.2369	0.2390	0.0000325	0.02227	246.75
	12	CNN	0.1575	0.0248	0.24677	0.24462	0.0000275	0.02065	519.25
	36		0.085	0.0072	0.2053	0.2295	0.0000025	0.01962	751.5

The CPU time is calculated for three algorithms for variety of dataset sizes. From this analysis, it is observed that TSVR, Random Forest and Wavelet-CNN take approximately comparable time on CPU for training data of size 12 months and 24 months. However, when the training data for 36 months is considered, wavelet-CNN outperforms TSVR and RFR in terms of CPU time. Although the training time for CNN based prediction model is higher in case of 12 months

and 24 months of training data seize, but the trade-off obtained in terms of a low RMSE is commendable. These algorithms can also be used in tandem with the NWP models to improve the prediction accuracy. The analysis is also carried out for sampling interval Δt =1 hour.

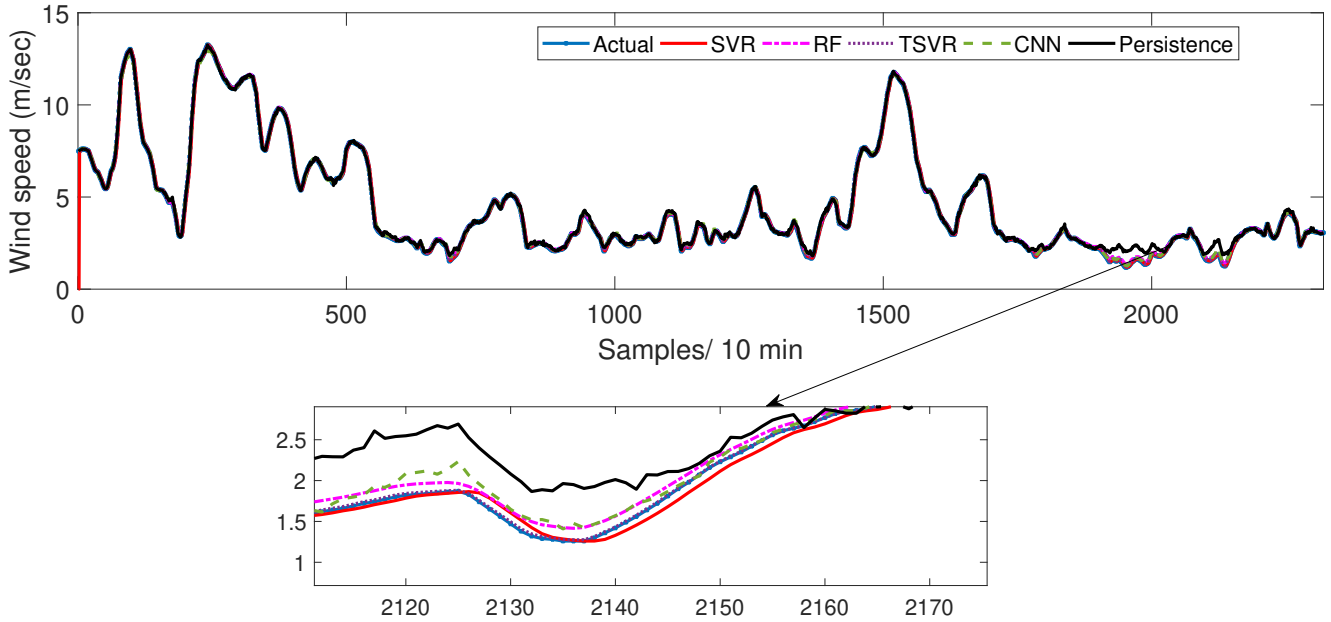

Figure 4: Ramp event forecasting using TSVR, CNN and RF for sampling interval Δt=10 minutes for wind farm site in Spain

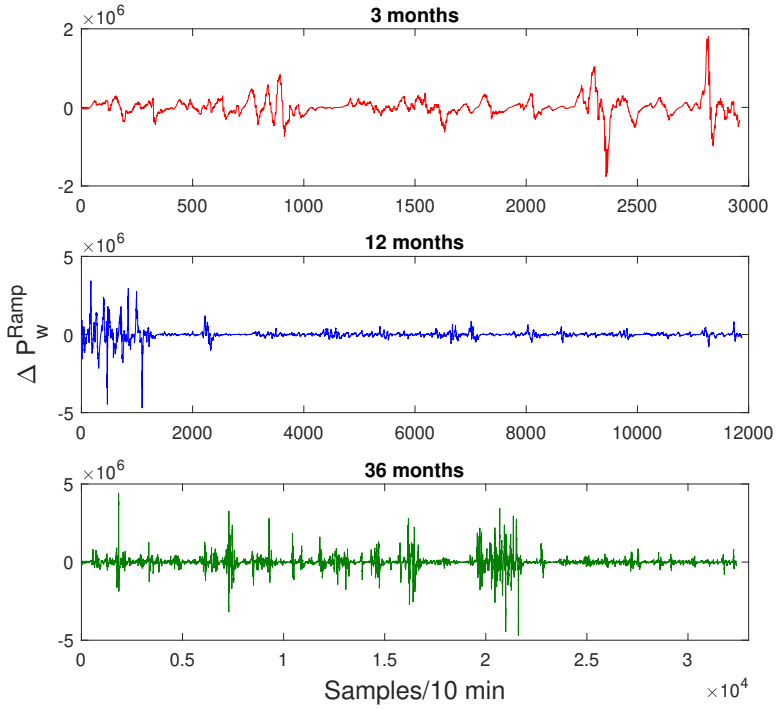

Figure 5: Ramp signal for Veja Mate wind site

For an Argentinian site, the proposed wavelet-based techniques are compared with Persistence and classical SVR methods. The error metrics RMSE and MAE are computed with 10-fold CV and results are within bounds. Wavelet-CNN is 28.76% and 26.43% effective than wavelet-RF and wavelet-TSVR methods in terms of RMSE. For ramp-up event, CNN yields the minimum error followed by TSVR and RF as depicted in Table 3. It is worthwhile to note that for wind speed forecasting as well as for ramp event prediction, the amount of training data available governs the prediction performance. With a large training data, the superiority of deep-learning models like CNN and LSTM can be examined. The ability of these algorithms to extract meaningful features from the input data makes the entire prediction process efficient and thus eliminates the need for explicit feature extraction and reduction techniques.

Table 3: Performance metrics for Δt=1 hour for wind farm site in Argentina

Method	RMSE	MAE	R^{up}	R^{down}	CPU (secs)
TSVR	2.141 ± 0.11	1.314 ± 0.12	0.7921	0.8988	1243
RF	2.211 ± 0.37	1.511 ± 2.01	0.8121	0.9138	1250
CNN	1.575 ± 0.11	1.016 ± 0.25	0.4911	0.8101	1513
SVR	2.071 ± 1.89	1.126 ± 2.91	0.8961	0.9502	1346
Persistence	3.041 ± 2.21	2.017 ± 0.31	0.9015	0.9929	1411
RNN [40]	1.141 ± 1.11	1.214 ± 1.12	1.7921	1.8988	1495
LSTM [41]	**1.121 ± 1.07**	1.114 ± 1.01	0.4121	0.8138	1205
Bi-LSTM	1.275 ± 1.09	1.116 ± 1.05	1.0911	0.8101	1513

Further, we have compered deep learning-based methods like RNN, LSTM, and Bi-LSTM with the proposed wavelet-CNN. Results reveal that LSTM yields a minimum RMSE and MAE for ramp event forecasts. The ability of LSTM to extract deep and meaningful features from the underlying time-series makes it an excellent candidate for forecasting wind speed ramp event. The parameters of LSTM network include a learning rate of 0.0001 and a batch size of 128.

For RNN, the RMSE is reported is 1.141 compared to 1,275 with Bi-LSTM model. The improvement in LSTM and Bi-LSTM over RNN is due to the presence of time-dependent states which solve the problem of vanishing gradient as experienced in RNNs. Further, the prediction performance of wavelet-LSTM is found 28.8% better than wavelet-CNN model in terms of RMSE. Compared to TSVR model, LSTM performs 47.64% better in terms of RMSE. A common observation that error in ramp-down events is more compared to ramp-up events.

The prediction framework of ramp events is further extended by testing the residuals of wavelet-CNN model. The test for heteroscedasticity is performed by plotting the residuals with the fitted values.

Heteroscedasticity refers to a situation where the residuals of a forecasting model have a non-constant variance. Figure 6 illustrates the test for heteroscedasticity where the normalized residuals are plotted against the normalized forecasted values (based on wavelet-CNN) of dataset A. Since the plot does not take a funnel shape, we can confirm that the given residuals do not suffer from heteroscedasticity. Adverse effect of heteroscedasticity include large error in estimation of confidence and prediction intervals.

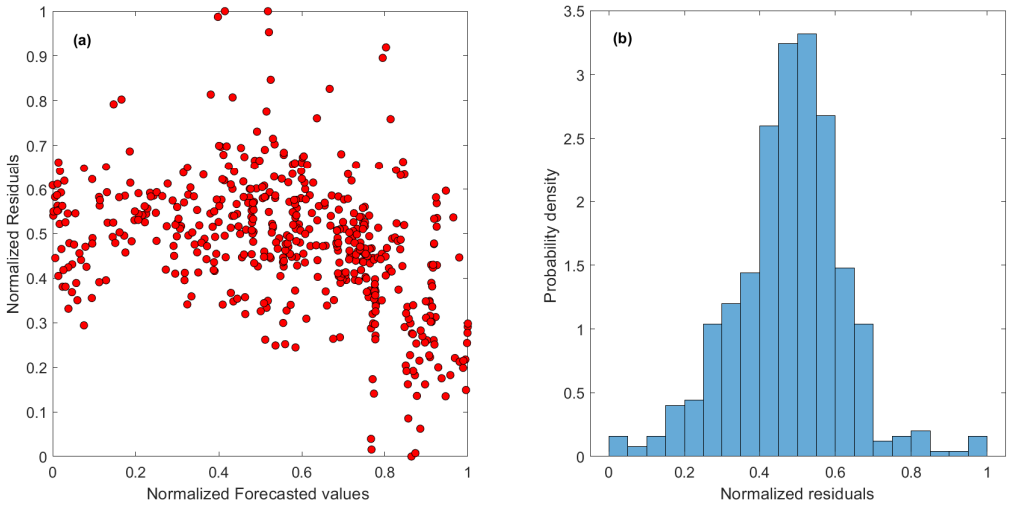

Figure 6: Residual analysis for ramp event forecasting. (a) Test for heteroscedasticity of residuals, and (b) distribution of normalized residuals as probability density function.

In this work, ramp prediction with a 10 minute sampling interval is considered. By using advanced algorithms like an SVR variant, random forest regression, and deep learning models like CNN, we extend the ramp event prediction study. In cases of ramp events for 3 and 12 months, TSVR results in a minimum error. Log-energy entropy based randomness for a time-series $f(t)$ is

$$E\{f(t)\} = \sum_{t=0}^{T} \log(f(t)^2). \tag{9}$$

The concept of entropy in general is used to describe the information content in

a given signal. Some of the popular entropy types like Shannon and Log-energy are used to estimate the randomness. For this purpose, Log-energy entropy can also be used as a feature vector in classification task. With respect to the wind energy domain, the same can also be utilized for classification of ramp events into their distinct states. In the current context, we have leveraged Log-energy entropy to estimate randomness in the ramp signal. Table 4 depicts the log-energy entropy to assess the randomness.

Table 4: Log-energy entropy in wind power ramps

Site	Technique	Randomness	
		WT	EMD
A	TSVR	2.2795×10^{3}	3.6546×10^{3}
	RFR	2.6210×10^{3}	3.2311×10^{3}
	CNN	2.6212×10^{3}	3.4645×10^{3}
B	TSVR	4.1295×10^{3}	3.6396×10^{3}
	RFR	4.6120×10^{3}	3.2287×10^{3}
	CNN	4.3112×10^{3}	3.4721×10^{3}
C	TSVR	1.4397×10^{3}	2.8913×10^{3}
	RFR	1.3764×10^{3}	2.1321×10^{3}
	CNN	1.7204×10^{3}	2.5127×10^{3}

For the considered wind farm sites, the log-energy entropy calculated using WT and EMD indicates that randomness is lower when the time-series is decomposed via WT than EMD for all the datasets. The industrial importance of this research problem is based on the ability of machine learning and deep learning techniques to forecast ramp events. Significant ramp events cause power reversal which hamper wind power integration in terms of reliability and security. Furthermore, these large power reversals cannot be stored in energy storage systems due to SoC constraints. Hence, it helps predict such significant events to plan power dispatch optimally.

5. Conclusions

In this manuscript, ramp events are studied from a regression point of view. Wind speed datasets from Spain, Germany, and Argentina are considered with 10 minute and 1 hour sampling duration. Wavelet transforms as a signal pre-processing technique helps improve the ramp event prediction efficacy. Moreover, the ramp event prediction is also made considering a larger size of training data in the form of 3, 12, and 36 months. Wavelet filter named Daubechies 4 decomposes the time-series into low and high-frequencies. Machine learning (ML) techniques such as SVR, RF, and TSVR are tested and compared with deep learning (DL) techniques such as the CNN. Results reveal that the wavelet-CNN-based prediction method for ramp events is superior to ML techniques when the training data length is 36 months. The randomness in wind speed is assessed by measuring the log-energy entropy. Quantitatively, wavelet-CNN is 28.76% and 26.43% effective than wavelet-RF and wavelet-TSVR methods in terms of RMSE.

Acknowledgements

This work was supported by VILLIUM FONDEN under VILLIUM Investigator Grant (no 25920): Center for Research on Microgrids (CROM); www.crom.et.aau.dk

References

[1] Dhiman HS, Deb D, Guerrero JM. Hybrid machine intelligent SVR variants for wind forecasting and ramp events. Renewable and Sustainable Energy Reviews 2019;108:369–79.

[2] Denić N, Petković D, Spasić B. Global economy increasing by enterprise resource planning. In: Encyclopedia of Renewable and Sustainable Materials. Elsevier; 2020, p. 331–7. doi:10.1016/b978-0-12-803581-8.11590-5.

[3] GWEC . Global wind report 2019 — global wind energy council. https://gwec.net/global-wind-report-2019/; 2020. (Accessed on 02/06/2021).

[4] Dhiman HS, Deb D, Muyeen SM, Kamwa I. Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines. IEEE Trans on Energy Conversion 2021;doi:10.1109/tec.2021.3075897.

[5] Huang Z, Chalabi Z. Use of time-series analysis to model and forecast wind speed. Journal of Wind Engineering and Industrial Aerodynamics 1995;56(2-3):311–22. doi:10.1016/0167-6105(94)00093-s.

[6] Dhiman HS, Deb D, Carroll J, Muresan V, Unguresan ML. Wind turbine gearbox condition monitoring based on class of support vector regression models and residual analysis. Sensors 2020;20(23):6742. doi:10.3390/s20236742.

[7] Shamshirband S, Petković D, Amini A, Anuar NB, Nikolić V, Ćojbašić Ž, et al. Support vector regression methodology for wind turbine reaction torque prediction with power-split hydrostatic continuous variable transmission. Energy 2014;67:623–30. doi:10.1016/j.energy.2014.01.111.

[8] Azad HB, Mekhilef S, Ganapathy VG. Long-term wind speed forecasting and general pattern recognition using neural networks. IEEE Transactions on Sustainable Energy 2014;5(2):546–53. doi:10.1109/tste.2014.2300150.

[9] Petković D, Pavlović NT, Ćojbašić Ž. Wind farm efficiency by adaptive neuro-fuzzy strategy. International Journal of Electrical Power & Energy Systems 2016;81:215–21. doi:10.1016/j.ijepes.2016.02.020.

[10] Lazarevska E. Wind speed prediction with extreme learning machine. In: 2016 IEEE 8th International Conference on Intelligent Systems (IS). IEEE; 2016,doi:10.1109/is.2016.7737415.

[11] Cheng L, Zang H, Ding T, Sun R, Wang M, Wei Z, et al. Ensemble recurrent neural network based probabilistic wind speed forecasting approach. Energies 2018;11(8):1958. doi:10.3390/en11081958.

[12] Petković D, Ćojbašič Ž, Nikolić V. Adaptive neuro-fuzzy approach for wind turbine power coefficient estimation. Renewable and Sustainable Energy Reviews 2013;28:191–5. doi:10.1016/j.rser.2013.07.049.

[13] Liu L, Ji T, Li M, Chen Z, Wu Q. Short-term local prediction of wind speed and wind power based on singular spectrum analysis and locality-sensitive hashing. Journal of Modern Power Systems and Clean Energy 2018;6(2):317–29. doi:10.1007/s40565-018-0398-0.

[14] Zuluaga CD, Álvarez MA, Giraldo E. Short-term wind speed prediction based on robust kalman filtering: An experimental comparison. Applied Energy 2015;156:321–30. doi:10.1016/j.apenergy.2015.07.043.

[15] Zheng H, Kusiak A. Prediction of wind farm power ramp rates: A data-mining approach. Journal of Solar Energy Engineering 2009;131(3). doi:10.1115/1.3142727.

[16] Muzammal M, Talat R, Sodhro AH, Pirbhulal S. A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. Information Fusion 2020;53:155–64. URL: https://doi.org/10.1016/j.inffus.2019.06.021. doi:10.1016/j.inffus.2019.06.021.

[17] Lin Y, Jin X, Chen J, Sodhro AH, Pan Z. An analytic computation-driven algorithm for decentralized multicore systems. Future Generation Computer Systems 2019;96:101–10. URL: https://doi.org/10.1016/j.future.2019.01.031. doi:10.1016/j.future.2019.01.031.

[18] Zahid N, , Sodhro AH, Kamboh UR, Alkhayyat A, Wang L, et al. AI-driven adaptive reliable and sustainable approach for internet of things enabled healthcare system. Mathematical Biosciences and Engineer-

ing 2022;19(4):3953–71. URL: https://doi.org/10.3934/mbe.2022182. doi:10.3934/mbe.2022182.

[19] Ren Y, Suganthan PN, Srikanth N. A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods. IEEE Transactions on Sustainable Energy 2015;6(1):236–44. doi:10.1109/tste.2014.2365580.

[20] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Transactions on Signal Processing 2014;62(3):531–44. doi:10.1109/tsp.2013.2288675.

[21] Dhiman HS, Deb D. Machine intelligent and deep learning techniques for large training data in short-term wind speed and ramp event forecasting. International Transactions on Electrical Energy Systems 2021;31(9). URL: https://doi.org/10.1002/2050-7038.12818. doi:10.1002/2050-7038.12818.

[22] Yin H, Ou Z, Huang S, Meng A. A cascaded deep learning wind power prediction approach based on a two-layer of mode decomposition. Energy 2019;189:116316. doi:10.1016/j.energy.2019.116316.

[23] Sevlian R, Rajagopal R. Detection and statistics of wind power ramps. IEEE Transactions on Power Systems 2013;28(4):3610–20. doi:10.1109/tpwrs.2013.2266378.

[24] Cui M, Ke D, Sun Y, Gan D, Zhang J, Hodge BM. Wind power ramp event forecasting using a stochastic scenario generation method. IEEE Transactions on Sustainable Energy 2015;6(2):422–33. doi:10.1109/tste.2014.2386870.

[25] Barber C, Bockhorst J, Roebber P. Auto-regressive hmm inference with incomplete data for short-horizon wind forecasting. In: Lafferty J, Williams C, Shawe-Taylor J, Zemel R, Culotta A, editors. Advances in Neural Information Processing Systems; vol. 23. Curran Associates, Inc.; 2010,.

[26] Gallego C, Costa A, Cuerva A. Improving short-term forecasting during ramp events by means of Regime-Switching Artificial Neural Networks. In: 10th EMS Annual Meeting. 2010, p. EMS2010–367.

[27] Al-Rakhami MS, Islam MM, Islam MZ, Asraf A, Sodhro AH, Ding W. Diagnosis of COVID-19 from x-rays using combined CNN-RNN architecture with transfer learning 2020;URL: https://doi.org/10.1101/2020.08.24.20181339. doi:10.1101/2020.08.24.20181339.

[28] Wang Z, Zhang J, Zhang Y, Huang C, Wang L. Short-term wind speed forecasting based on information of neighboring wind farms. IEEE Access 2020;8:16760–70. doi:10.1109/access.2020.2966268.

[29] Zhen H, Niu D, Yu M, Wang K, Liang Y, Xu X. A hybrid deep learning model and comparison for wind power forecasting considering temporal-spatial feature extraction. Sustainability 2020;12(22):9490. doi:10.3390/su12229490.

[30] Jaseena K, Kovoor BC. Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks. Energy Conversion and Management 2021;234:113944. doi:10.1016/j.enconman.2021.113944.

[31] Shahid F, Zameer A, Muneeb M. A novel genetic LSTM model for wind power forecast. Energy 2021;223:120069. doi:10.1016/j.energy.2021.120069.

[32] Wu Q, Guan F, Lv C, Huang Y. Ultra-short-term multi-step wind power forecasting based on CNN-LSTM. IET Renewable Power Generation 2021;15(5):1019–29. doi:10.1049/rpg2.12085.

[33] Afrasiabi M, Mohammadi M, Rastegar M, Afrasiabi S. Advanced deep learning approach for probabilistic wind speed forecasting. IEEE Transactions on Industrial Informatics 2021;17(1):720–7. doi:10.1109/tii.2020.3004436.

[34] Chen FC, Jahanshahi MR. NB-CNN: Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion. IEEE Transactions on Industrial Electronics 2018;65(5):4392–400. doi:10.1109/tie.2017.2764844.

[35] Vapnik VN. The Nature of Statistical Learning Theory. Springer New York; 2000.

[36] Peng X. TSVR: An efficient twin support vector machine for regression. Neural Networks 2010;23(3):365–72.

[37] Malhotra P, Ramakrishnan A, Anand G, Vig L, Agarwal P, Shroff GM. Lstm-based encoder-decoder for multi-sensor anomaly detection. CoRR 2016;abs/1607.00148. arXiv:1607.00148.

[38] Irwin JS. A theoretical variation of the wind profile power-law exponent as a function of surface roughness and stability. Atmospheric Environment (1967) 1979;13(1):191–4.

[39] Dhiman H, Deb D, Balas VE. Supervised Machine Learning in Wind Forecasting and Ramp Event Prediction. Elsevier; 2020. doi:10.1016/c2019-0-03735-1.

[40] Shabbir N, Kutt L, Jawad M, Amadiahanger R, Iqbal MN, Rosin A. Wind energy forecasting using recurrent neural networks. In: 2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE). IEEE; 2019,URL: https://doi.org/10.1109/bdkcse48644.2019.9010593. doi:10.1109/bdkcse48644.2019.9010593.

[41] Araya IA, Valle C, Allende H. LSTM-based multi-scale model for wind speed forecasting. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Springer International Publishing; 2019, p. 38–45. URL: https://doi.org/10.1007/978-3-030-13469-3_5. doi:10.1007/978-3-030-13469-3_5.