A deep learning framework for probabilistic dynamic cable rating in offshore HVDC systems

Yan, S.; Ghaffarian Niasar, M.; Popov, M.

DOI
10.1016/j.ijepes.2025.110802

Publication date
2025

Document Version
Final published version

Published in
International Journal of Electrical Power & Energy Systems

Citation (APA)
Yan, S., Ghaffarian Niasar, M., & Popov, M. (2025). A deep learning framework for probabilistic dynamic cable rating in offshore HVDC systems. *International Journal of Electrical Power & Energy Systems*, *170*, Article 110802. https://doi.org/10.1016/j.ijepes.2025.110802

Electrical Power and Energy Systems 170 (2025) 110802

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

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

A deep learning framework for probabilistic dynamic cable rating in offshore HVDC systems

Shen Yan *, Mohamad Ghaffarian Niasar, Marjan Popov

Delft University of Technology, Faculty of EEMCS, Mekelweg 4, 2628CD, Delft, The Netherlands

ARTICLE INFO

Keywords:
Deep learning
DCR
xLSTM
BNN
HVDC

ABSTRACT

The effective prediction of dynamic cable ratings (DCR) in the HVDC cable is pivotal for enhancing transmission efficiency and maximizing electricity sales in offshore wind farms. Due to complex wind conditions, traditional machine learning methods, such as support vector machines, struggle to provide accurate long-term DCR predictions and express prediction uncertainties. To address these challenges, this article proposes a novel deep learning framework for dynamic cable rating prediction based on encoder–decoder networks, in which the encoder utilizes Bidirectional extended-long Short-Term Memory networks to encode contextual information from the input data. The decoder introduces an additive attention mechanism, which allows the network to focus on relevant features in the input sequence. In addition, to capture the uncertainty for DCR prediction, a Bayesian neural network approximation method based on the Monte Carlo dropout method is introduced. Finally, this paper introduces a thermal risk estimation method by considering both the maximum conductor temperature limit and the temperature gradient limit. Results demonstrate that the proposed method not only improves electric field distribution but also achieves superior economic benefits.

1. Introduction

The integration of offshore wind farms into electrical grids has grown significantly due to global renewable energy initiatives. A crucial aspect of effectively harnessing this energy is efficient electricity transmission [1]. Therefore, underground cables are vital, particularly in high-voltage direct current (HVDC) transmission systems widely used to connect offshore wind farms to the power grid. The conventional rated capacity of HVDC underground cables often based on worst-case assumptions, is no longer sufficient to support the continuous growth of wind farms [2]. To address this challenge, there is a growing emphasis on better utilization of cable capacities and optimal use of the cable connection to ensure maximum efficiency and reliability in power transmission.

A promising solution is the dynamic thermal ratings (DTR) method, which is used to fully exploit the available capacity of a device without sacrificing safety and reliability. DTR technology was initially applied to equipment such as transmission lines [3,4] and transformers [5]. Daminov et al. [5] conducted an in-depth study on the DTR of transformers, pointing out that while traditional rating methods generally use fixed limit temperatures or typical load curves for estimation, DTR can combine real-time and predictive load and environmental information to accurately portray the load-bearing capacity of the equipment. The work emphasizes the importance of considering multiple thermal limitations of the equipment when determining the DTR, and balancing the temperature and current limitations in a feasible region, and maximizing the equipment utilization while meeting the thermal safety boundaries. Meanwhile, Dynamic Cable Rating (DCR) has been widely researched in recent years as a method to evaluate the real-time current-carrying capacity of cables [6]. So far, research has focused on the DCR of AC cables [7], exploration of the DCR of HVDC cables remains relatively uncharted. Juan et al. introduced a method for optimal sizing of high voltage AC export cables in offshore wind farms by integrating wind power generation and seabed temperature data into a thermo-electrical model, coupled with a probabilistic lifetime analysis. However, applying similar strategies to HVDC cables remains an open area of research. Calculating the DCR of HVDC cable systems is not as simple as that of AC cables. The reason for this is that AC cables are always subject to a maximum temperature limit, whereas the DCR of HVDC cables should combine the maximum temperature limit and the temperature gradient limit [8] as the temperature gradient will increase hetero-charge density, which will cause local electric field distortion. Therefore, it is necessary to detect the internal temperature of insulation to prevent the temperature from being too high [9].

Moreover, DCR is not sufficient to optimize curtailment decisions in offshore wind farms as information on future load current scenarios

* Corresponding author.
E-mail addresses: S.Yan-1@tudelft.nl (S. Yan), m.ghaffarianniasar@tudelft.nl (M. Ghaffarian Niasar), m.popov@tudelft.nl (M. Popov).

https://doi.org/10.1016/j.ijepes.2025.110802
Received 8 October 2024; Received in revised form 27 April 2025; Accepted 1 June 2025

and conductor temperatures is needed hours in advance. Therefore, the implementation of advanced predictive modeling techniques becomes crucial. Currently, the existing DCR prediction methods are mainly conventional probabilistic methodologies such as Markov Chain (MC) model [10,11], regression analysis models such as the Support Vector Machine (SVM) algorithm [12–14], and the Autonomous Integrated Moving Average (ARIMA) model [15]. In these models, the MC method is heavily based on state transition probabilities to predict future states. For long-term forecasting, this approach may struggle to accurately capture complex dynamic behavior. Although the application of the regression analysis algorithm in the prediction of DCR has demonstrated considerable predictive effectiveness in offshore distribution grids, for offshore wind farms, it is difficult to accurately capture and predict cable load demands. Therefore, existing models cannot properly interrelate input and output data by extracting deep-level features without specialized feature engineering. In this regard, deep learning models are capable of modeling a non-linear relationship between input and output data with flexible network structures to provide data-driven solutions.

Recently, machine learning techniques have been extensively applied to different engineering fields [16–19], especially Long Short-Term Memory Networks (LSTMs), have made significant progress in the field of wind power prediction [20–23], but a single LSTM model usually does not adequately account for the uncertainty of wind energy when dealing with wind power prediction [21,22]. To overcome this limitation, a model that can generate probability distributions is needed. Liu et al. [23] proposed a data-driven energy storage management strategy by combining a deep reinforcement learning framework with an LSTM-LUBE-based interval prediction method. This approach unifies wind power interval forecasting with dynamic decision-making to optimize storage charge/discharge operations under fluctuating wind power and load conditions. Beside, the use of Bayesian methods can incorporate uncertainty considerations into the forecasts, resulting in confidence intervals for the forecasts [24–26]. Although Bayesian neural network (BNN) models have been previously reported for wind energy forecasting, most of them are based on conventional fully connected NN models, such as single-layer perceptron [24], multi-layer perceptron (MLP) in [25] or convolutional gated repeat units (GRU) [26]. These models often struggle with the complexity of wind energy data, inadequately model long-term temporal dynamics, and may oversimplify the inherent uncertainties in forecasting.

The above literature provides valuable information for studying cable rating and wind power generation prediction. However, for offshore wind farms, wind speed variation is quite uncertain, so the accuracy of existing methods of predicting cable rating values is not high enough, and traditional models cannot balance the need to account for wind energy uncertainty while capturing the long-term dependence of wind energy. Therefore, this paper proposes a novel multi-feature and multistep prediction model to predict the DCR of HVDC cables in offshore wind farms, and our main contributions are summarized as follows.

- A novel prediction model (Attention-BNN-S2S) is proposed based on an encoder–decoder structure, in which the encoder uses the Bi-sLSTM networks to capture long-term dependence, while the decoder introduces the additive attention mechanism to focus on relevant features in the input sequence. In addition, to consider DCR prediction uncertainty and obtain probabilistic predictions, we also utilize the MC-dropout method to approximate BNN. The superiority of the Attention-BNN-S2S prediction model is demonstrated by comparing various machine learning methods.
- The cable temperature calculation model is built based on the thermoelectric equivalent (TEE) method, and the fourth-order Runge–Kutta method (RK4) was used to solve the difference equation to obtain the cable temperature. Then it was compared with the FEM model built in COMSOL in terms of steady-state and transient to verify the TEE model. The modified DCR prediction model is established by combining the Attention-BNN-S2S prediction model with the TEE model.

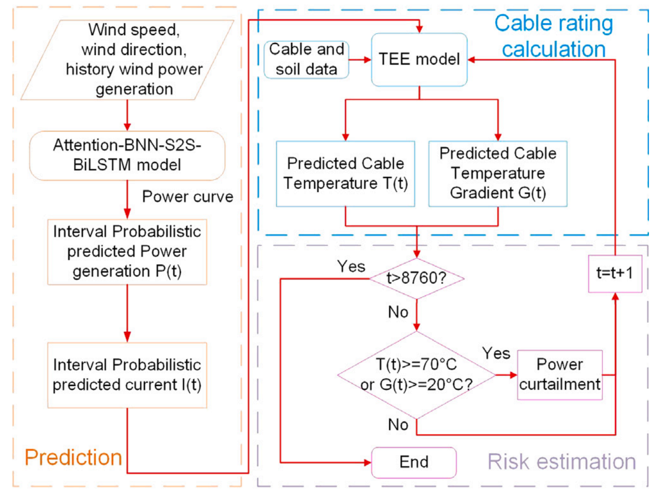


Fig. 1. Introduction to Attention-BNN-S2S-DCR prediction model.

- Considering the maximum conductor temperature limit and the temperature gradient limit, this paper proposes a thermal risk assessment method for HVDC cables. The thermal overload risk of HVDC cables in offshore wind farms is evaluated by comparing different ML methods, reflecting the superiority of the DCR prediction model in this paper.

The paper is structured as follows. Section 2 describes the proposed methodology followed by parameter setting and verification. Section 3 describes the definition of the study case. Section 4 presents and discusses the experimental results. Finally, Section 5 provides meaningful conclusions.

2. Proposed model

2.1. Overview of the proposed DCR prediction framework

The schematic of the Attention-BNN-S2S-DCR model is shown in Fig. 1. The DCR prediction process is as follows:

2.1.1. Feature engineering and wind power generation prediction

Feature Engineering is a crucial step in machine learning and data mining, as it transforms raw data into meaningful inputs for predictive models. Drawing on domain knowledge about wind power generation, this study refines both feature construction and selection to capture the temporal and physical dynamics of the system. In this work, historical wind power generation is used as a primary input feature, while future wind power generation is designated as the main output to be predicted. As wind power generation exhibits strong autocorrelation, meaning that present output is often closely related to recent past values. By including historical wind power in the input set, the model exploits these temporal patterns, leading to more accurate forecasts of short-term power fluctuations.

To accommodate directional variability, the wind speed is decomposed into two perpendicular components. The horizontal component $X_{\text{component}}$ is derived by multiplying the measured wind speed by the cosine of the wind direction (in radians), while the vertical component, $Y_{\text{component}}$, is found by multiplying the wind speed by the sine of the wind direction:

$$X_{\text{component}} = \text{Windspeed} \times \cos\left(\frac{\text{Wind Direction} \times \pi}{180}\right), \tag{1}$$

$$Y_{\text{component}} = \text{Windspeed} \times \sin\left(\frac{\text{Wind Direction} \times \pi}{180}\right). \tag{2}$$

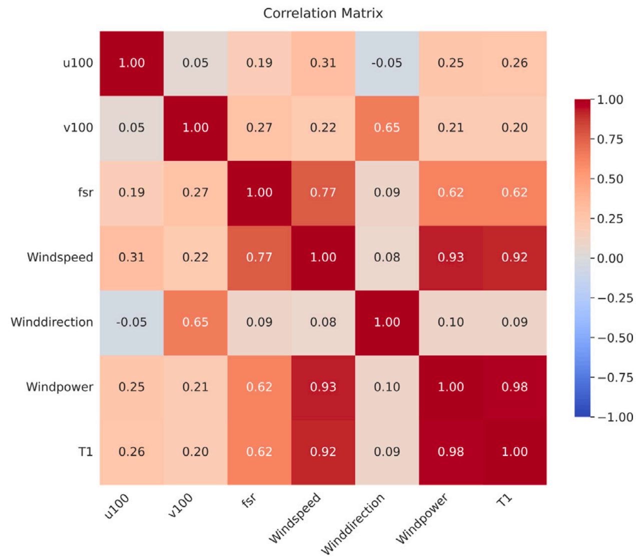


Fig. 2. Correlation matrix.

These features provide a clearer physical interpretation of wind flow, where different orientations may affect turbine performance in distinct ways.

Since wind power often exhibits significant temporal dependence, one-period lag wind power (denoted by $T1$) is introduced to model autocorrelative effects. Analysis of the autocorrelation function (ACF) and partial autocorrelation function (PACF) reveals that a single-step lag is the most influential, although additional lags may also be relevant if extended dependence is evident. For a stationary time series $\{y_t\}$ with mean μ and variance σ^2, the ACF at lag k is defined as:

$$\rho(k) = \frac{\mathrm{Cov}(y_t, y_{t-k})}{\mathrm{Var}(y_t)} = \frac{E[(y_t-\mu)(y_{t-k}-\mu)]}{\sigma^2}. \tag{3}$$

The PACF measures the correlation between y_t and y_{t-k} after removing the effects of the intermediate lags $1, 2, \ldots, k-1$. This can be obtained by fitting an autoregressive model of order k,

$$y_t = \phi_{1,k}\, y_{t-1} + \phi_{2,k}\, y_{t-2} + \cdots + \phi_{k,k}\, y_{t-k} + \epsilon_t. \tag{4}$$

The PACF at lag k is then given by the last coefficient in this model:

$$\alpha(k) = \phi_{k,k}. \tag{5}$$

Fig. 2 shows that windspeed maintains a strong positive correlation (0.93) with Power, underscoring the well-known physical relationship between wind velocity and energy output. Meanwhile, the lagged power feature (T_1) demonstrates an even higher correlation (0.98) with Power, highlighting the importance of temporal dependence in wind power generation. By contrast, Wind Direction remains largely independent of other variables apart from moderate correlation with the vertical wind-speed component v100 (0.65). The forecast surface roughness (fsr) feature also exhibits moderate to high correlation with windspeed (0.77) and Power (0.62), indicating a non-negligible sea level roughness influence on overall system conditions.

Fig. 3 illustrates the time-series properties of wind power. The ACF exhibits a gradual decline, suggesting that current power output is provided by multiple preceding time steps, while the PACF emphasizes the dominant contribution of the first lag. This pattern corroborates the decision to include T_1 in the feature set, as it directly captures the primary autoregressive effect. After evaluating correlations, stationarity, and predictive importance, the following features are retained in the final model: $X_{\text{component}}$, $Y_{\text{component}}$, Lagged wind power: T_1, historical real wind power and fsr.

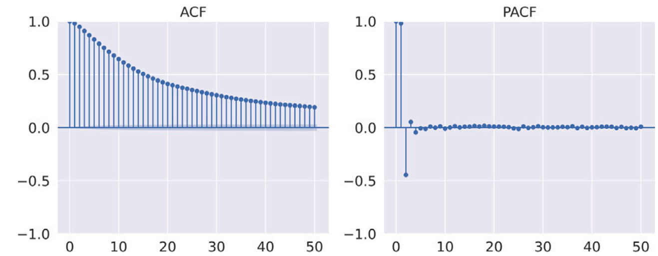


Fig. 3. Determining derivative indicators via ACF/PACF.

The wind power is converted to a current time series by using:

$$I(t) = \frac{P(t)}{V_{ref}} \tag{6}$$

where $I(t)$ is the load current (A). $P(t)$ is the power output (W). In this paper, the bipolar DC transmission system is used. V_{ref} represents the reference voltage (V) of each exported cable. Consequently, the total voltage of the system is V_{ref}.

Fig. 4 shows the core functionality of the Attention-BNN-S2S model. The model first encodes the input sequence into a series of hidden states using Bi-sLSTM, then decodes the sequence into an output sequence, with each step of the decoder utilizing an attention mechanism to focus on different parts of the input sequence. Finally, it passes through the MC-Dropout layer, where multiple forward passes are made during the inference process, each time randomly shutting down neurons to generate distributions of possible outcomes.

2.1.2. *Cable rating calculation*

The predicted interval current loads and, the cable and soil parameters are entered into the thermoelectric equivalent (TEE) model to calculate conductor temperatures and temperature gradients.

2.1.3. *Thermal risk estimation and power curtailment*

The thermal risk estimation is calculated at each time step t considering the probabilistic distribution of the obtained conductor temperature. The realistic thermal risk r_{risk} and predicted thermal p_{risk} are calculated by:

$$r_{\text{risk}} = \begin{cases} \frac{\sum_{i=1}^{h} \mathbb{I}(T(t+i)\geq T_{\text{limit}})}{h} \\ \frac{\sum_{i=1}^{h} \mathbb{I}(G(t+i)\geq G_{\text{limit}})}{h} \end{cases} \tag{7}$$

$$p_{\text{risk}} = \begin{cases} \frac{\sum_{i=1}^{h} \mathbb{I}(T'(t+i)\geq T_{\text{limit}})}{h} \\ \frac{\sum_{i=1}^{h} \mathbb{I}(G'(t+i)\geq G_{\text{limit}})}{h} \end{cases} \tag{8}$$

where:

- $T(t+i)$: Actual conductor temperature at future time step $t+i$,
- T_{limit}: Maximum allowable conductor temperature,
- $G(t+i)$: Actual power generation at time step $t+i$,
- G_{limit}: Maximum allowable power generation,
- $T'(t+i)$, $G'(t+i)$: Predicted temperature and power generation, respectively,
- h: Prediction horizon,
- $\mathbb{I}$: Indicator function, returning 1 if the condition inside is true, otherwise 0.

The Eqs. (7), (8) reflect the number of times the cable temperature exceeded 70 °C or the temperature gradient exceeded 20 °C. Power curtailment is activated when one of two conditions is met. The whole process continues in a loop until the end of the entire test set.

To assess the thermal aging of the cable insulation material, the Arrhenius equation is applied to estimate the fraction of loss of life over

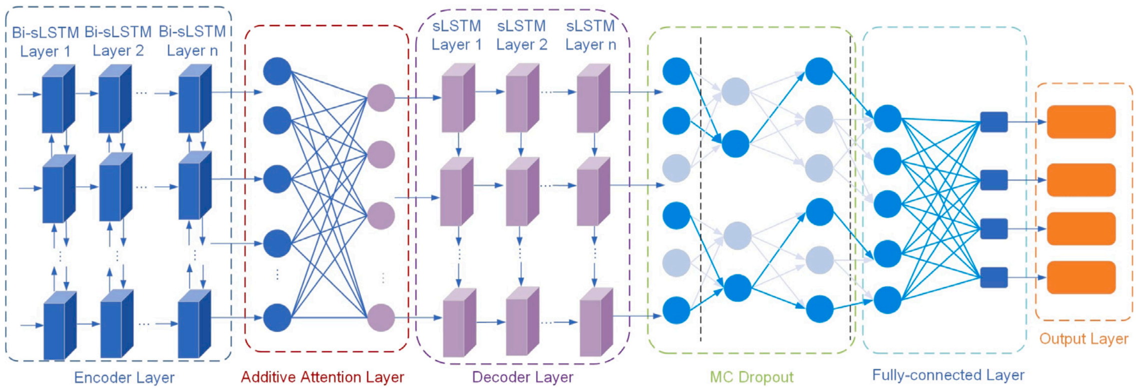


Fig. 4. Structure of the Attention-BNN-S2S Model.

time. The index of thermal cable aging ($A_{thermal}$) at a temperature of 70 °C is described by Eqs. (9)–(11). The loss-of-life fraction is defined as:

$$L_C = \frac{\Delta t}{L_0 \cdot \exp\left(\frac{-\Delta W}{k_B} cT\right)\left(\frac{E}{E_0}\right)^{-\left(n_0 - b\,cT\right)}}, \tag{9}$$

$$cT = \frac{1}{\theta_o} - \frac{1}{\theta_C}, \tag{10}$$

$$A_{thermal}(t) = \frac{L_C}{L_{70}} = \exp\left(\frac{\Delta W}{k_B}\left(\frac{1}{343} - \frac{1}{\theta_C}\right)\right)\left(\frac{E}{E_0}\right)^{-\left(b\left(\frac{1}{343}-\frac{1}{\theta_C}\right)\right)} \tag{11}$$

Here θ_o is the reference operation temperature, ΔW is the insulation material's activation energy, k_B is Boltzmann's constant, and L_o is the expected cable lifetime when it operates at θ_o. This cable circuit is designed with a nominal life of 30 years [15]. In this study, the DC electric field is calculated by [27]:

$$E(r) = \frac{U_0}{r_o}\,\delta\left[1 - \left(\frac{r_i}{r_o}\right)^{\delta}\right]\left(\frac{r}{r_o}\right)^{\delta-1}, \tag{12}$$

where r_i and r_o are the inner and outer insulation radii, respectively, r is a generic radius, and δ is the field inversion coefficient, calculated as follows:

$$\delta = \frac{a\,w_e + b\,E_m}{1 + b\,E_m} = \frac{a\,\Delta T_i\left(\frac{r_o}{r_i}\right) + b\,U_0\left(r_o - r_i\right)}{1 + b\,U_0\left(r_o - r_i\right)}, \tag{13}$$

Here, ΔT_i is the temperature drop between the inner insulation (r_i) and outer insulation (r_o) (in K), and E_m is the mean value of the electric field (in kV/mm). The parameters a and b are the temperature and field coefficients of the electrical conductivity σ, given by:

$$\sigma(r,t) = \sigma_0 \exp\left[a\,T(r,t) + b\,E(r,t)\right]. \tag{14}$$

The Weibull distribution is applied to model the probability of cable failure due to thermal aging. The cumulative thermal aging A_{cum} is computed annually as the sum of the equivalent cable aging over the operational year:

$$A_{cum} = \sum_{y=1}^{n} \frac{A_{thermal}(y)}{8760} \tag{15}$$

where n is the cable's operational year, and 8760 represents the total hours in a year. The failure probability at a given cumulative aging, $P_{fail}(A_{cum})$, is defined as:

$$P_{fail}(A_{cum}) = 1 - \exp\left(-\left(\frac{A_{cum}}{T_{thresh}}\right)^{\beta}\right) \tag{16}$$

Here, β is the shape parameter of the Weibull distribution, T_{thresh} is the aging threshold. The hazard rate $H_{rate}(A_{cum})$ which represents the instantaneous failure risk, is derived as follows:

$$H_{rate}(A_{cum}) = \frac{d(P_{fail}(A_{cum}))}{d(A_{cum})} \cdot \frac{1}{1 - P_{fail}(A_{cum})} \tag{17}$$

The annual economic evaluation accounts for maintenance, replacement, interruption costs, and opportunity costs from lost generation. The total economic benefit (B_{total}) is calculated as:

$$B_{\text{total}} = \sum_{y=1}^{n}\Big(R_{\text{op}} - \big(C_{\text{maint}} + C_{\text{replace}} + C_{\text{interrupt}} + LR_{\text{curtail}}\big)\Big) \tag{18}$$

where C_{maint}, $C_{replace}$, and $C_{interrupt}$ represent the maintenance, replacement, and interruption costs, respectively, while R_{op} is the operating revenue, and $LR_{curtail}$ is the lost revenue due to curtailed energy generation. The values for these costs are based on Ref. [15]. The energy is priced at €95/MWh for offshore energy. The capital cost of the cable is assumed to be 1 million € per kilometer [28], with a total cable length of 135 km.

2.2. Bidirectional extended long short-term memory network

This paper introduces the extended long short-term memory network (xLSTM) to address the limitations of standard LSTM networks in capturing the complex temporal dependencies and long-range patterns required for DCR predictions. The xLSTM architecture incorporates a softmax-based gating mechanism to balance the contributions of input and forget gates at each time step. This adjustment helps to control the data transfer over time, further refining the prediction of conductor temperature and overload risk. Additionally, Softplus activation at the output layer ensures non-negative predictions, favoring slight overestimation, which is critical for safety in thermal overload scenarios. By adopting this method, the model is not only better equipped to handle noise and interference in time series signals, and to enhance its resilient ability. One of the xLSTM model, called scalar lstm (slstm), adds a scalar update mechanism to the traditional LSTM [29], which is shown in Fig. 5(a). This design optimizes the gating mechanism by providing fine-grained control of the internal memory cells, making it more suitable for processing sequence data with small time variations. xLSTM usually uses exponential gating and normalization techniques to improve the stability and accuracy of the model in processing long sequence data:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot z_t \tag{19}$$

$$n_t = f_t \cdot n_{t-1} + i_t \tag{20}$$

$$h_t = o_t \tilde{h}_t, \tilde{h}_t = c_t / n_t \tag{21}$$

$$z_t = tanh(\tilde{z}_t), \tilde{z}_t = w_z^{\top} x_t + r_z h_{t-1} + b_z \tag{22}$$

in which $tanh$ and σ are the nonlinear activation functions. $\odot$ denotes pointwise multiplication. The gating mechanisms are adapted as follows, with the option to stabilize using additional states:

$$i_t = \exp(\tilde{i}_t), \tilde{i}_t = w_i^{\top} x_t + r_i h_{t-1} + b_i \tag{23}$$

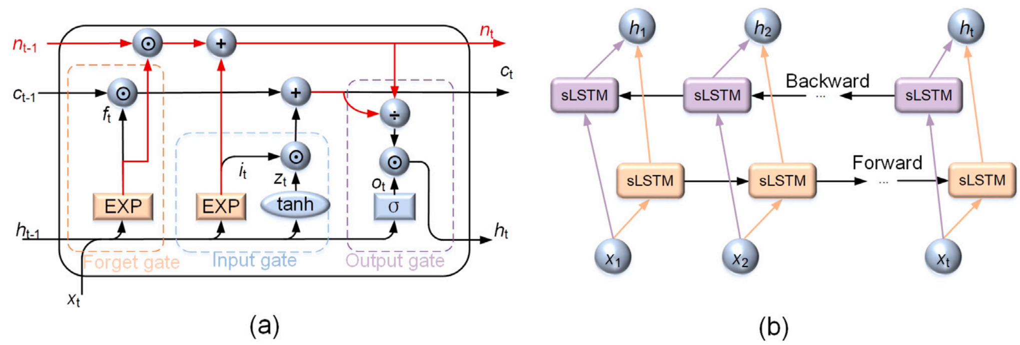


Fig. 5. Bi-sLSTM model. (a) sLSTM unit structure. (b) Bi-sLSTM flowchart.

$$f_t = \sigma(\tilde{f}_t)\text{or}\exp(\tilde{f}_t), \tilde{f}_t = w_f^{\top} x_t + r_f h_{t-1} + b_f \quad (24)$$

$$o_t = \sigma(\tilde{o}_t), \tilde{o}_t = w_o^{\top} x_t + r_o h_{t-1} + b_o \quad (25)$$

The variables are defined as follows: i_t, f_t, o_t are the input gate, forget gate, and output gate respectively. c_t, h_t, and z_t are the cell state, hidden state, and intermediate state respectively. Stabilize gates with an additional state m_t:

$$m_t = \max(\log(f_t) + m_{t-1}, \log(i_t)) \quad (26)$$

$$i_t' = \exp(\log(i_t) - m_t) = \exp(\tilde{i}_t - m_t) \quad (27)$$

$$f_t' = \exp(\log(f_t) + m_{t-1} - m_t) \quad (28)$$

Inspired by the outstanding performance of Bidirectional LSTM (Bi-LSTM) to encode the necessary information in a sequence [30], it is selected as an encoder for taking the temporal relation of wind power generation into account. Traditional unidirectional LSTM-based models process temporal data in a single forward direction, which can overlook important contextual information from future time steps. In contrast, Bi-LSTM introduces a backward pass—processing the data sequence in reverse. By leveraging both forward and backward dependencies, Bi-sLSTM captures more comprehensive temporal patterns, including subtle correlations and recurring trends that may appear before and after each time step. The principle of Bi-sLSTM is to split the neurons of a regular sLSTM into two directions, one for positive time direction (forward states), and the other for negative time direction (backward states). By utilizing two-time directions, the sequential information from the past and future of the current frame can be used. The flowchart of Bi-sLSTM is shown in the encoder part of Fig. 5(b).

2.3. Additive attention mechanism

The additive attention mechanism is introduced, and at each decoding step, the attention score e_t^i is computed to represent the importance of each element in the input sequence to the current decoder output. The attention score can be expressed as:

$$e_t^i = W \cdot \tanh(V_t h_t + V_i h_i + b) \quad (29)$$

where W, V_t, V_i and b are the weights and biases for linear layers. All four tensors are initialized with Xavier-uniform (Glorot) initialization and are updated together with the rest of the network by back-propagation using the Adam optimizer. h_t is the hidden state of the decoder. h_i is the hidden state of the encoder. The scores were normalized using the softmax function to obtain the attention weights α_t^i:

$$\alpha_t^i = \frac{\exp(e_t^i)}{\sum_{k=1}^{T} \exp(e_t^k)} \quad (30)$$

in which indicates the degree of attention of the decoder to the encoder output h_i at time step t. T is the time step of the encoder output. The computed attentional weights are used to generate a weighted average context vector c_t, which is a weighted representation of the input data with weights determined by the current decoder:

$$c_t = \sum_{i=1}^{T} \alpha_t^i h_i \quad (31)$$

2.4. Encoder and decoder

The input sequence is fed into the encoder, which encodes it into a context vector by learning the input representation. This vector is then passed to the decoder, which learns to generate the output sequence. The advantage of the Seq2Seq (S2S) architecture is its ability to handle variable input and output sequence lengths [30]. In this paper, the encoder processes the input sequence $X = x_1, x_2, \ldots, x_t$ into a representation vector $v = v_1, v_2, \ldots, v_t$ using a Bi-sLSTM:

$$\begin{bmatrix} v_t \\ H_t \end{bmatrix} = \varphi(x_t) \quad (32)$$

where $H_t \in \mathbb{R}^n$ is the hidden state at time t, which is shown in Fig. 5(b). ϕ represents the architecture of an encoder. The decoder ψ generates the output sequence $Y = \{y_1, \ldots, y_m\}$ based on the representation vector v by using a unidirectional sLSTM:

$$\begin{bmatrix} p(y_t | \{y_t | i < t\}, v) \\ s_t \end{bmatrix} = \psi(s_{t-1}, y_{t-1}, v) \quad (33)$$

in which $p(y_t | \{y_t | i < t\}, v)$ represents the probability of predicting the output y_t at the current time step t, given the known historical output $\{y_t \mid i < t\}$ and the information provided by the encoder v. To focus more on those historical data that are most critical to the current prediction, this paper introduces an additive attention mechanism, thus Eq. (33) can be changed as:

$$\begin{bmatrix} p(y_t | \{y_t | i < t\}, c_t) \\ s_t \end{bmatrix} = \psi(s_{t-1}, y_{t-1}, c_t) \quad (34)$$

The overall process of the encoder–decoder is shown in Fig. 6. The decoder output y_t at each time step is determined by calculating a relevance score e_t between the encoder output v_i and the decoder's last hidden state s_{t-1}. This score is then normalized to form attention weights α_t^i to create a context vector through a weighted sum of encoder outputs.

2.5. MC dropout-based Bayesian neural network

The Dropout variational inference, known as the Monte Carlo (MC) Dropout method, was introduced to perform approximate Bayesian inference without requiring significant changes to the standard architecture of artificial neural networks (ANNs) [31]. Specifically, this

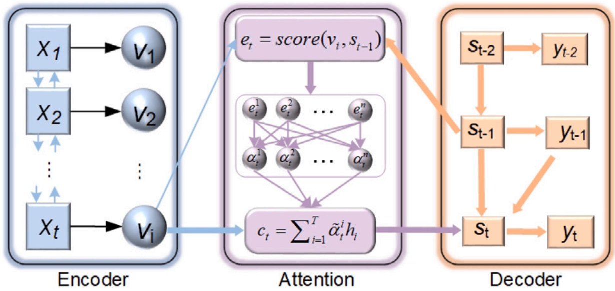


Fig. 6. The proposed attention mechanism in the sLSTM.

approach treats dropout masks as a means of sampling from an approximate posterior over the network weights. By doing so, MC Dropout can be used to estimate both the predictive mean and the predictive uncertainty of a model's output, effectively approximating a Bayesian Neural Network (BNN).

The Bayes' theorem, which states that:

$$P(\omega \mid X, Y) = \frac{P(Y \mid X, \omega)\, P(\omega)}{P(Y \mid X)}, \tag{35}$$

where $P(\omega \mid X, Y)$ is the posterior distribution of weights ω given the training data X and labels Y. The term $P(\omega)$ is the prior distribution of the weights, while $P(Y \mid X, \omega)$ represents the likelihood function, i.e., the probability of observing Y given ω and X. The denominator $P(Y \mid X)$ is the marginal likelihood (or evidence) that serves as a normalizing constant. In practice, directly computing $P(\omega \mid X, Y)$ is intractable for large-scale neural networks. Consequently, various approximation strategies have been proposed to tackle this problem.

MC Dropout [31] leverages the standard dropout technique—often used to reduce overfitting—by keeping the dropout layer active during training and testing. For conventional training, dropout randomly "drops" units (neurons) in each layer according to a Bernoulli mask with probability p. In MC Dropout, we continue to apply the same random dropout masks at test time, effectively generating different "thinned" versions of the network. Mathematically, one can interpret these thinned networks as samples from an approximate posterior $q(\omega \mid \theta)$, where θ encapsulates the dropout parameters.

When performing inference with MC Dropout, each forward pass corresponds to a sample ω_t drawn from this approximate posterior distribution of weights. By running T stochastic forward passes, we obtain:

$$P(\omega \mid D) \approx \frac{1}{T} \sum_{t=1}^{T} P\left(Y \mid X, \omega_t\right), \tag{36}$$

where $D = \{(X, Y)\}$ is the training dataset, and $\{\omega_t\}_{t=1}^{T}$ are drawn through different dropout masks. Instead of a single deterministic output, the network's prediction is then given by averaging the outputs of these sampled networks:

$$\hat{Y} = \frac{1}{T} \sum_{t=1}^{T} f\left(X, \omega_t\right), \tag{37}$$

where $f(X, \omega_t)$ denotes the output of the neural network at the tth forward pass. Hence, multiple runs with different dropout masks yield both a mean prediction and an empirical variance that can serve as an estimate of predictive uncertainty.

In addition to the predictive mean in Eq. (37), the variance across the T sampled outputs can quantify the epistemic uncertainty:

$$\mathrm{Var}[\hat{Y}] = \frac{1}{T} \sum_{t=1}^{T} \left(f(X, \omega_t) - \hat{Y}\right)^2. \tag{38}$$

A higher variance indicates that the model is less certain about its prediction. This insight is particularly important in applications requiring robust decision-making under uncertainty.

Among various Bayesian methods for uncertainty quantification, such as Markov Chain Monte Carlo (MCMC) and variational inference, MC Dropout remains particularly attractive due to its practical advantages. First, MC Dropout offers a simple and scalable approach: unlike MCMC, which often incurs significant computational overhead, MC Dropout requires only multiple stochastic forward passes at inference time. Second, it can be easily integrated into existing deep learning architectures with minimal modifications. Moreover, the theoretical connection between dropout and approximate variational inference in deep Gaussian processes, established by Gal and Ghahramani [31], provides a solid Bayesian justification for its use. Finally, MC Dropout effectively captures epistemic uncertainty, making it highly valuable in applications where prediction confidence is critical.

Therefore, MC Dropout-based BNNs strike a favorable balance between computational cost, ease of deployment, and theoretical rigor, making them a common choice for uncertainty quantification in deep neural networks.

2.6. Cable thermo-electric equivalent model

The finite difference method (FDM) was selected to solve the thermal model of the cable as it allows the inclusion of variable load current and ambient parameters at every time step during the evaluation process. The schematic diagram of the thermal equivalent is shown in Fig. 7 [31,32].

Specifically, $C_1 = Q_c$, $C_2 = pQ_{ins}$, $C_3 = (1-p)Q_{ins}$, $C_4 = Q_{screen}$, $C_5 = Q_{sheath}$, $C_6 = p_2 Q_o$, $C_7 = (1-p_2)Q_o$, $C_8 = Q_{soil}$ represent the thermal capacity of conductor, insulation, insulation screen, metallic sheath, outer sheath, soil respectively. T_1, T_2, and T_3 represent the thermal resistance between one conductor and the metallic sheath, and the thermal resistance of the outer sheath. Because the region between the metallic and the outer sheath is usually quite thin, and the materials used often have relatively high thermal conductivity, the value of T_2 is much smaller compared to T_1 and T_3. Consequently, it is often negligible in practice and can be omitted to simplify the thermal model. The set of partial differential equations obtained from the cable circuit for the calculation of cable temperatures at nodes θ_c, θ_{in}, θ_w, θ_s, θ_{soil}, θ_{amb} where the subscripts c, in, w, s, soil, amb correspond to the conductor, insulation, water-blocking tape, sheath, soil, and external environment are shown below:

$$\frac{d}{dt}\theta_c (C_1 + C_2) = W_C(t) - \frac{\theta_c - \theta_{in}}{T_1} \tag{39}$$

$$\frac{d}{dt}\theta_{in} (C_3 + C_4 + C_5 + C_6 + C_7) = \frac{\theta_c - \theta_{in}}{T_1} - \frac{\theta_{in} - \theta_s}{T_3} \tag{40}$$

$$\frac{d}{dt}\theta_s C_8 = \frac{\theta_{in} - \theta_s}{T_3} - \frac{\theta_s - \theta_{soil}}{T_{4a}} \tag{41}$$

$$\frac{d}{dt}\theta_{soil} C_9 = \frac{\theta_s - \theta_{soil}}{T_{4a}} - \frac{\theta_{soil} - \theta_{amb}}{T_{4b}} \tag{42}$$

in which T_1 and T_3 are defined as:

$$T_1 = \frac{\rho_{ins}}{2\pi} \ln\left(\frac{D_{\text{screen}}}{D_{\text{cond}}}\right) \tag{43}$$

$$T_3 = \frac{\rho_o}{2\pi} \ln\left(1 + \frac{2 \cdot E_o}{D_o}\right) \tag{44}$$

where ρ_{ins} is the thermal resistivity of the insulation, D_{screen} is the mean diameter of the screen, D_{cond} is the conductor diameter, ρ_o is the thermal resistivity of the outer sheath, E_o is the thickness of the outer sheath, D_o is the outer diameter of the metallic sheath. The IEC 60287 thermal resistance model for the thermal environment of underground cables divides the total thermal resistance into the cladding impedance T4a and the external environment impedance T4b, defined as follows:

$$T_{4a} = \frac{\rho_{\text{soil}}}{2\pi} \ln\left(1 + \frac{0.5\, L}{D_{cond}}\right), \tag{45}$$

$$T_{4b} = \frac{\rho_{\text{soil}}}{2\pi} \ln\left(u + \sqrt{u^2 - 1}\right), \tag{46}$$

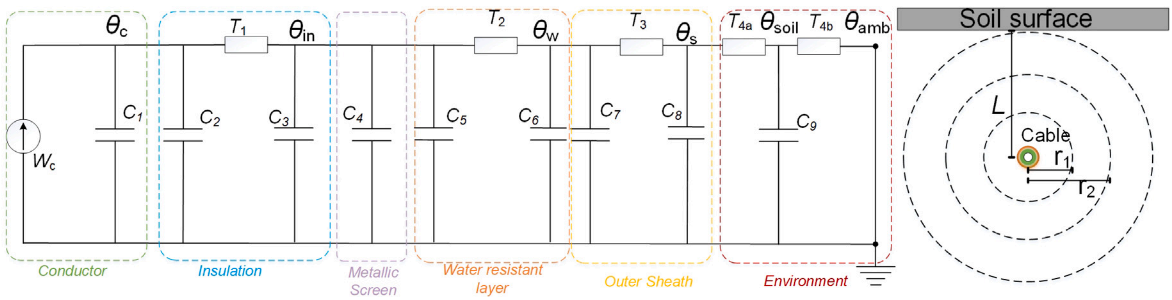


Fig. 7. Thermal equivalent of the HVDC power cable and soil layered thermal resistance modeling.

$$u = \frac{2L}{D_{cond} + 0.5L} \tag{47}$$

where ρ_{soil} is the thermal resistivity of the soil ($\Omega \cdot$ m), L is the burial depth of the cable centerline below the ground surface (m), u is a dimensionless geometry factor used in the external resistance calculation. Define the matrix form of differential equations:

$$A \cdot \frac{d}{dt}\Theta(t) = B(t) \tag{48}$$

The coefficient matrix A and $\Theta(t)$:

$$A = \begin{bmatrix} C_1 + C_2 & 0 & 0 & 0 \\ 0 & C_3 + C_4 + C_5 + C_6 + C_7 & 0 & 0 \\ 0 & 0 & C_8 & 0 \\ 0 & 0 & 0 & C_9 \end{bmatrix} \tag{49}$$

$$\Theta(t) = \begin{bmatrix} \theta_c(t) \\ \theta_{in}(t) \\ \theta_s(t) \\ \theta_{soil}(t) \end{bmatrix} \tag{50}$$

Forcing term vector $B(t)$:

$$B(t) = \begin{bmatrix} W_C(t) - \dfrac{\theta_c(t) - \theta_{in}(t)}{T_1} \\ \dfrac{\theta_c(t) - \theta_{in}(t)}{T_1} - \dfrac{\theta_{in}(t) - \theta_s(t)}{T_3} \\ \dfrac{\theta_{in}(t) - \theta_s(t)}{T_3} - \dfrac{\theta_s(t) - \theta_{soil}(t)}{T_{4a}} \\ \dfrac{\theta_s(t) - \theta_{soil}(t)}{T_{4a}} - \dfrac{\theta_{soil}(t) - \theta_{amb}(t)}{T_{4b}} \end{bmatrix} \tag{51}$$

In the steady state case, the cable temperature θ_c can be expressed as:

$$\theta_c(t) \approx \theta_{\text{soil}}(t) + W_c T_1 + T_{\text{soil}}(W_c + W_s) + \frac{T_3(\theta_{\text{soil}}(t) - \theta_{amb}(t))}{T_{\text{soil}}} \tag{52}$$

where $W_c(t)$ represent the conductor loss:

$$W_c(t) = R_{DC} \cdot I(t)^2 \tag{53}$$

At the reference temperature of 20 °C, the DC resistance $R_{\text{DC},20}$ is known. Because the resistance varies with conductor temperature, it must be corrected using the aluminum temperature coefficient α_{Al}. The general expression reads

$$R_{\text{DC}} = R_{\text{DC},20} \times \left[1 + \alpha_{\text{Al}}\,(T_c - 20)\right]. \tag{54}$$

The above equation are solved by using the 4th Runge–Kutta (RK4) method for cable temperature. Transform this equation to solve for $\Theta(t)$ in standard form:

$$\frac{d}{dt}\Theta(t) = A^{-1}B(t) \tag{55}$$

Let:

$$f(t, \Theta(t)) = A^{-1}B(t) \tag{56}$$

Four slopes k_1, k_2, k_3, k_4 are computed as:

$$\begin{cases} k_1 = \Delta t \cdot f(t, \Theta(t)) = \Delta t \cdot A^{-1}B(t) \\ k_2 = \Delta t \cdot f\left(t + \frac{\Delta t}{2}, \Theta(t) + \frac{k_1}{2}\right) = \Delta t \cdot A^{-1}B\left(t + \frac{\Delta t}{2}\right) \\ k_3 = \Delta t \cdot f\left(t + \frac{\Delta t}{2}, \Theta(t) + \frac{k_2}{2}\right) = \Delta t \cdot A^{-1}B\left(t + \frac{\Delta t}{2}\right) \\ k_4 = \Delta t \cdot f(t + \Delta t, \Theta(t) + k_3) = \Delta t \cdot A^{-1}B(t + \Delta t) \end{cases} \tag{57}$$

Finally, the conductor temperature is updated as the value of the next time step:

$$\Theta(t + \Delta t) = \Theta(t) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \tag{58}$$

3. Datasets and parameters setting

3.1. Datasets and feature engineering

The proposed methodology is applied to an actual 914 MW DolWin2 project in the North Sea, which contains Gode (1&2) wind farm (582 MW) and Nordsee1 wind farm (332 MW). The illustration of the test network is shown in Fig. 8. For simplicity, we model the connection between the offshore and onshore converter stations as a single 90 km ± 320 kV HVDC underground cable circuit. The distance between the two poles of the cable is 0.5 m. The cables are installed in ducts at 1.0 m depth.

The hourly wind speeds and synthetic hourly power generation of the Gode (1&2) wind farm comes from the publicly available dataset [33]. For the Nordsee1 wind farm, we extract ERA5 data for wind farm location from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [34], and convert the wind speed data into synthetic power output using the turbines (Senvion 6M126)' power curves [35]. The combined dataset is divided into training, validation, and test sets, with 60% of the data used for training, 10% for validation, and the remaining 30% for testing. The pseudocode for the overall training and testing process of Attention-BNN-S2S-DCR prediction is shown as Algorithm 1 and Algorithm 2, respectively.

3.2. Evaluation metrics

The methodology is evaluated considering wind farm over-planting (WFO) scenarios. Four key performance metrics were assessed: root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2), which are calculated by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2} \tag{59}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - p_i| \tag{60}$$

$$MAPE = \frac{1}{n}\left|\frac{y_i - p_i}{y_i}\right| \tag{61}$$

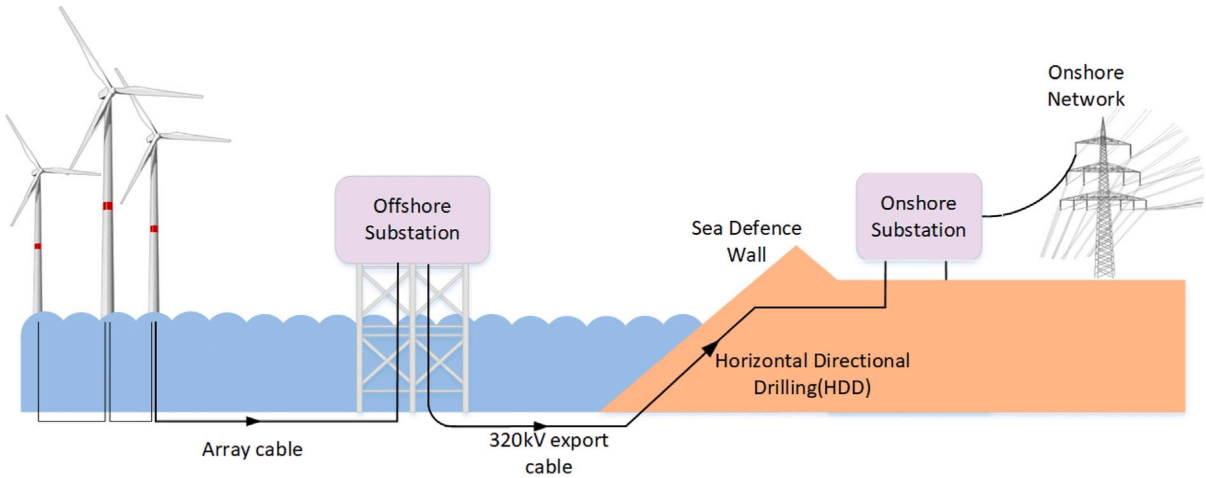


Fig. 8. Diagram for the test network.

Algorithm 1 Prediction Process

Input: Training dataset (X_{tr}, Y_{tr}), validation dataset (X_{val}, Y_{val}), test dataset (X_{test}, Y_{test})
Output: Predictions for test dataset, performance metrics
for each training epoch **do**
  Encoder output O_t = Bi-sLSTM based encoder(X_{tr})
  Initialize decoder input I_t, decoder hidden state s_t from encoder's final hidden state H_t
  for each prediction step **do**
    Compute attention weights e from s_t and O_t
    Generate context vector v_t using e and O_t
    Decoder output y_t = Decoder(I_t, v_t)
    Update I_t with last step y_{t-1}
  end for
  Compute loss and backpropagate
  Evaluate model performance on (X_{val}, Y_{val})
end for
Perform multiple predictions using (X_{test}, Y_{test}) to generate a distribution of outcomes
Return: Predictions, evaluation metrics

Algorithm 2 Thermal Risk Estimation

Input: Basic wind farm current $I_{BWF}(t)$, wind farm overplanting current $I_{WFO}(t)$, soil parameters
Output: Cable temperature $T_{cable}(t)$, temperature gradient $G_{cable}(t)$, predicted overload probability $p_{risk}(t)$, and realistic overload probability $r_{risk}(t)$
for $t = 0$ to 8760 **do**
  Calculate $T_{cable}(t)$ and $G_{cable}(t)$
  if $T_{cable}(t) \geq 70°C$ or $G_{cable}(t) \geq 20°C$ **then**
    Trigger power curtailment: $I(t) = I_{BWF}(t)$
  else
    $I(t) = I_{WFO}(t)$
  end if
  Calculate $p_{risk}(t)$ and $r_{risk}(t)$
end for
Return: $T_{cable}(t)$, $G_{cable}(t)$, $p_{risk}(t)$, $r_{risk}(t)$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - p_i)^2}{\sum_{i=1}^{n}(\bar{y}_i - p_i)^2} \tag{62}$$

in which y_i and p_i are the normalized original value and predicted value, and n is the total length of the time series. The model's performance in predicting cable temperature exceeding 70 °C is evaluated using standard classification metrics. These metrics are defined as follows: The thermal overload performance of the model is evaluated using standard classification metrics. These metrics are defined as follows: True Positive (TP):both the real and predicted temperatures exceed 70 °C. False Positive (FP): the predicted temperature exceeds 70 °C, but the real temperature does not. True Negative (TN):neither the real nor predicted temperatures exceed 70 °C. False Negative (FN):the real temperature exceeds 70 °C, but the predicted temperature does not.

Accuracy measures the general correctness of the model, including positive and negative classifications. It is defined:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{63}$$

The true positive accuracy (ATP) measures the proportion of true positives relative to all predictions (both positive and negative). It is calculated as:

$$ATP = \frac{TP}{TP + TN + FP + FN} \tag{64}$$

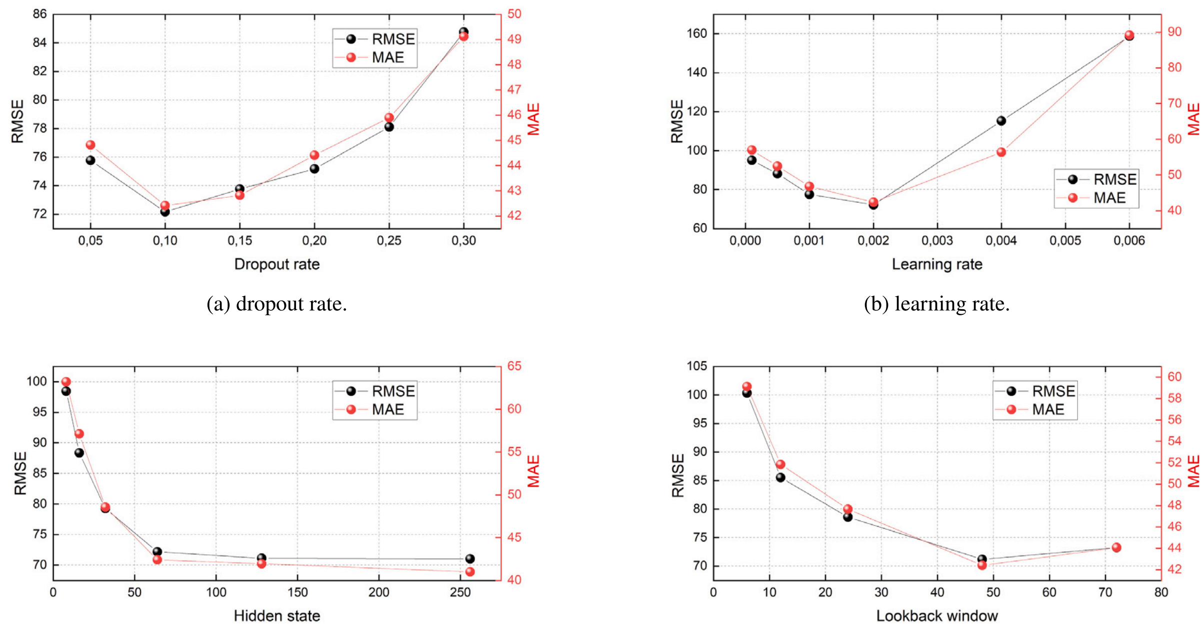
The true negative accuracy (ATN) represents the proportion of true negatives relative to all predictions:

$$ATN = \frac{TN}{TP + TN + FP + FN} \tag{65}$$

The false positive rate (FPR) represents the proportion of negative instances that are incorrectly classified as positive:

$$FPR = \frac{FP}{FP + TN} \tag{66}$$

We have introduced the following quantitative indicators to measure the predictive uncertainty of our model. Specifically, we focus on the Coverage and Mean Interval Width (MIW). These two metrics are straightforward however, yet informative for interval-based uncertainty estimation. Coverage measures the proportion (or probability) of each true value within the predicted intervals across all samples. It shows how frequently the model's predicted intervals actually contain the ground-truth values. Let N be the total number of data points. For the ith sample, let Lower$_i$ and Upper$_i$ be the lower and upper bounds of



(a) dropout rate.

(b) learning rate.

(c) hidden state.

(d) lookback window.

Fig. 9. Sensitive analysis of hyperparameters.

the predictive interval, respectively, and let Real_i be the true observed value. Then the Coverage is defined as:

$$\mathrm{Coverage} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{\mathrm{Lower}_i \le \mathrm{Real}_i \le \mathrm{Upper}_i\}, \tag{67}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function, which takes value 1 if the condition inside is satisfied, and 0 otherwise.

MIW quantifies the average width of the predictive intervals. It directly measures how "wide" these intervals are on average. Using the same notation as above, define:

$$\mathrm{MIW} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{Upper}_i - \mathrm{Lower}_i\right). \tag{68}$$

3.3. Parameters setting and hyperparameter optimization

The 320 kV HVDC offshore cable is used, and the geometry and thermal properties of the cable components are presented in Table 1. This article develops the model of finite differences of the cable in MATLAB R2021b and the model of finite elements of the cable in COMSOL Multiphysics 5.6. The prediction algorithm was developed using PyCharm software in the SURF research cloud server with two NVIDIA A10 GPUs, 22 Cores and 176 GB RAM. The code is written in Python 3.9, and the framework is PyTorch. Many researchers have demonstrated the importance of selecting hyperparameters for deep learning models. By adjusting the appropriate hyperparameters, the performance of the models can be improved. Furthermore, overfitting can be prevented, and the training process can be accelerated. To validate the impact of hyperparameters in neural networks on the results, this research adopts the same dataset and uses different combinations of hyperparameters for wind power prediction. The modified hyperparameters include learning rate, hidden state, dropout rate, and lookback window while keeping other hyperparameters consistent. The results of the hyperparameter tuning are presented in Fig. 9.

(1) Dropout Rate. Adjusting the dropout rate helps to decide between overfitting and underfitting. Lower dropout (0.05) may fail to regularize sufficiently (RMSE = 75.77 MW), while higher dropout (0.30) can overly diminish the network's capacity (RMSE = 84.76 MW). An intermediate value (e.g., 0.10 or 0.15) yields smaller RMSE/MAE and thus better performance. The best balance in our dataset appears at 0.10 with RMSE = 72.18 MW and $R^2 = 0.97$.

(2) Learning Rate. Choosing an appropriate learning rate is crucial for stable and efficient training. Extremely small rates (e.g., 0.0001) converge too slowly, often resulting in higher final errors (RMSE = 95.03 MW), while excessively large rates (0.004 or 0.006) can lead to unstable updates or divergence (RMSE = 115.26/158.74 MW). A moderate rate (e.g., 0.002) achieves the lowest RMSE = 72.18 MW and highest $R^2 = 0.97$.

(3) Hidden State. The number of hidden units dictates the model's representational power. Fewer units (e.g., 8 or 16) may underfit and yield high RMSE (98.46/88.37 MW), whereas excessive units (e.g., 128 or 256) can risk overfitting or increased complexity (RMSE around 71.12–73.90 MW). A moderately sized hidden state (64) produced the best trade-off (RMSE = 72.18 MW, $R^2 = 0.97$).

(4) Lookback Window. This parameter controls how many past time steps are fed into the model. A short window (e.g., 6 or 12) may miss essential seasonal or diurnal patterns (RMSE > 85 MW), whereas an overly long window (e.g., 96) might introduce noise or redundant information (RMSE = 76.11 MW). Our experiments show a sweet spot at 48 time steps, resulting in RMSE = 71.18 MW and $R^2 = 0.97$.

Generally, these results underscore the importance of hyperparameter tuning in neural network-based forecasting. Each hyperparameter (dropout rate, learning rate, hidden state, and lookback window) exerts a tangible impact on predictive accuracy (RMSE, MAE, MAPE) and model fit (R^2). Thus, the hyperparameters are determined according to the improved whale optimization algorithm, the IWOA optimization algorithm. The core idea of the algorithm is to adjust the search paths of solutions to find the global optimum in the search space. This paper introduces adaptive parameters and multiple search strategies to improve global exploration and local exploitation capabilities. The steps of the IWOA are summarized as follows:

4. Results and analysis

4.1. TEE model validation

To verify the accuracy of the Thermal–Electrical Equivalent (TEE) model, we first note that its core principle is to approximate the cable's

Table 1
Cable parameters [36–40].

Structures	Diameter [mm]	Thermal conductivity [W/m K]	Heat capacity [J/kg K]	Density [kg/m^3]
Copper conductor	50	450	393	8700
Conductor screen	55.1	0.23	2603	922
XLPE insulation	107.1	0.286	2603	922
Insulation screen	109.5	0.286	2603	922
Water-blocking tape	121.5	0.167	2182	1100
Aluminum sheath	138.5	236	900	2700
PE sheath	148	0.286	2532	948
Soil	–	1.0	855	1500

Algorithm 3 IWOA for Hyperparameter Optimization

Input: Training dataset (X_{tr}, Y_{tr}), validation dataset (X_{val}, Y_{val}), test dataset (X_{test}, Y_{test})
Output: Optimized hyperparameters
Initialize population P with random hyperparameters
Evaluate fitness for each individual in P
Select the leader L from P with the best fitness
while stopping criteria not met **do**
  for each individual i in P **do**
    Generate a random probability p
    if $p < 0.5$ **then**
      Update i using humpback whale's approach behavior towards L
    else
      Update i using spiral path mimicking whale's hunting behavior
    end if
    Evaluate fitness of the updated position of i
  end for
  Update the leader L with the best fitness from the updated population
  Perform roulette wheel selection to select parents
  Generate offspring through crossover and mutation
  Replace the worst individuals with the new offspring
end while
Return the optimized hyperparameters from the final L

heat transfer paths using a network of thermal resistances and capacitances, thereby reducing the computational complexity significantly. In this framework, complex 3D or 2D domains are often reduced to a simpler thermal circuit where material properties such as thermal conductivity, density, and specific heat are treated as constants over the temperature range of interest. The external environment is represented by equivalent boundary conditions, which can be modeled as isothermal or convective boundaries, depending on the specific setup. Heat capacity terms are included to capture the transient temperature rise in both the conductor and its surroundings. Consequently, the TEE model allows for fast calculation of conductor temperature under various loading scenarios, making it well-suited for real-time or online monitoring applications.

A two-dimensional finite element (FEM) model was then built in COMSOL Multiphysics to validate the TEE predictions. Fig. 10 illustrates the 2D temperature distribution from the FEM simulation, where the cable's cross-section is explicitly defined with the conductor, insulation, outer jacket, and surrounding medium. The material parameters (thermal conductivity, density, and specific heat) match those of the TEE model, ensuring comparable conditions. Sufficient mesh refinement was performed to consider steep temperature gradients near the conductor region, and mesh independence tests confirmed that further refinement did not noticeably alter the temperature field. For a steady-state current of 1470 A, the conductor temperature reached 70 °C, as shown by the contour plot. To replicate the same conditions in the TEE model, a uniform volumetric heat generation ($Q = I^2R$) was imposed in the conductor region, and the outer boundaries were defined either as fixed or convective/radiative conditions identical to those assumed by the TEE approach.

Fig. 11 shows the comparison of the results for which TEE model predictions are plotted against FEM solutions for transient and steady-state conditions. The agreement is notable: for the final steady-state condition, the difference in conductor temperature remains within ±1 °C, while throughout the transient heating phase, the maximum deviation also stays within the same range. This close match underscores the TEE model's capability to effectively capture heat conduction and thermal inertia effects that govern the temperature rise. Furthermore, the TEE method runs in approximately 3 s, making it about 600 times faster than the FEM simulation under the same computational environment. Such a significant reduction in computation time is critical for practical cable monitoring scenarios or real-time temperature regulation, where an online tool must rapidly process and deliver accurate temperature estimates.

Despite these advantages, a few simplifying assumptions in the TEE framework should be noted. Thermal properties are assumed constant, so changes in soil moisture or complex cable-layer compositions can affect accuracy if not updated accordingly. Furthermore, the TEE approach mainly relies on a one-dimensional or lumped approximation of radial heat conduction, which may not capture lateral or asymmetric effects in more intricate installations. Real-world conditions such as wind, rainfall, or intermittent solar heating are continuously modeled as fixed ambient or convective boundaries, introducing further approximation if these external factors vary significantly over time. Nevertheless, for many engineering applications requiring near real-time or online temperature estimation, the TEE model offers a good balance between computational efficiency and accuracy, as evidenced by the close alignment with the COMSOL results.

4.2. Prediction results and evaluation

A comparative study is conducted to analyze the wind-power forecast performance of five models—SVM, RNN, GRU, LSTM, and the proposed Attention-BNN-S2S over four forecast horizons: 6 h, 12 h, 24 h, and 48 h. In Fig. 12, the forecast horizon (6–48 h ahead) refers to how far in the future each curve predicts, and the horizontal axis shows a 250-h display window (t = 8500–8750h) that is excerpted from the 8760 h test set purely for visual clarity. The same performance trend holds across the entire test span. In the short term (6 h), all models, including the conventional SVM, RNN, GRU, and LSTM, achieve comparable accuracy. The forecast horizon extends to 12 h and beyond, however, these baselines fail to capture the rapid peaks and troughs of wind-power output, and their root-mean-square error rises sharply. By contrast, Attention-BNN-S2S remains closely aligned with the measured data even for 24 h and 48 h. This robustness is attributed to its bi-sLSTM encoder with an attention mechanism, which extracts and preserves salient temporal features, and to its unidirectional sLSTM decoder, which excels at generating coherent long-range trajectories. Moreover, the 95% predictive interval of Attention-BNN-S2S is consistently narrow, highlighting its accuracy and well-calibrated uncertainty, especially on long horizons where conventional models struggle.

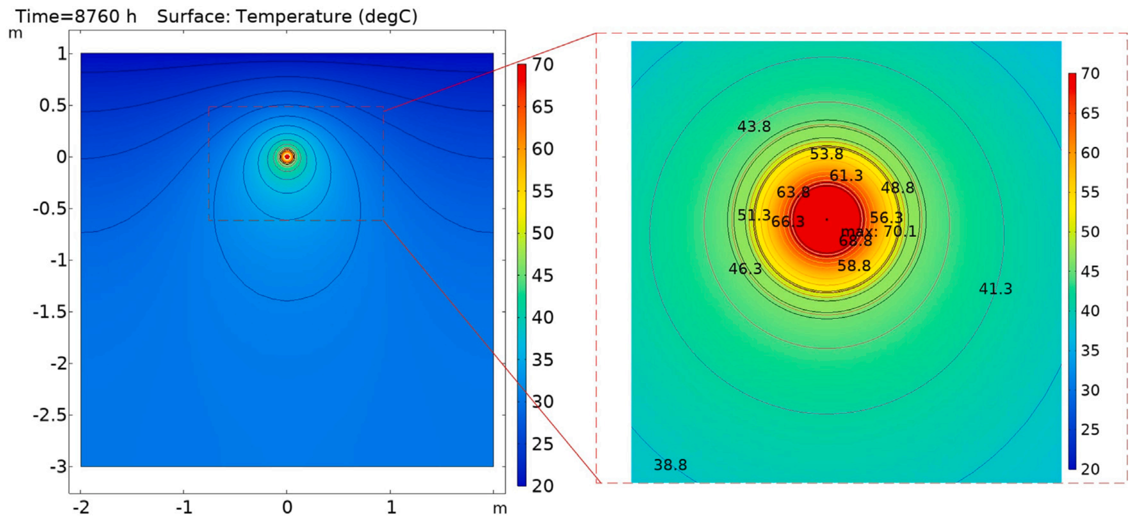


Fig. 10. Conductor temperature results of the COMSOL model.

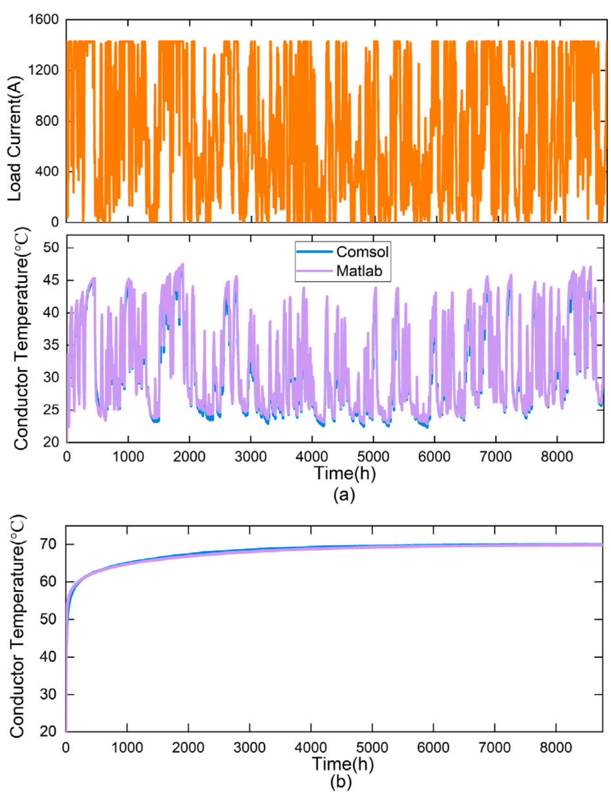


Fig. 11. Comparison of conductor temperature between the TEE and COMSOL model for transient and steady-state situations.

Converting the predicted wind power into current and then embedding it into the thermal model of the cable gives the predicted cable temperature as in Fig. 13. After considering the power curtailment, the improved model predicts both the conductor temperature and the temperature gradient to be more in line with the actual results. It should be noted that when t = 8500–8550, there is a lot of variation in the load current, the cable temperature, and the temperature gradient. This is due to the high temperature of the cable in this interval, resulting in a power cut. The 1-hour load current curtailment period considered in this paper is conservative, thus, the load current is curtailed hourly and varies widely.

In addition, a comprehensive evaluation was conducted on these five distinct predictive models. Fig. 14 shows the comparison of four error metrics for different models in wind power generation prediction. Attention-BNN-S2S consistently outperformed other models in all four key performance metrics, indicating a superior capability for short- and long-term predictions.

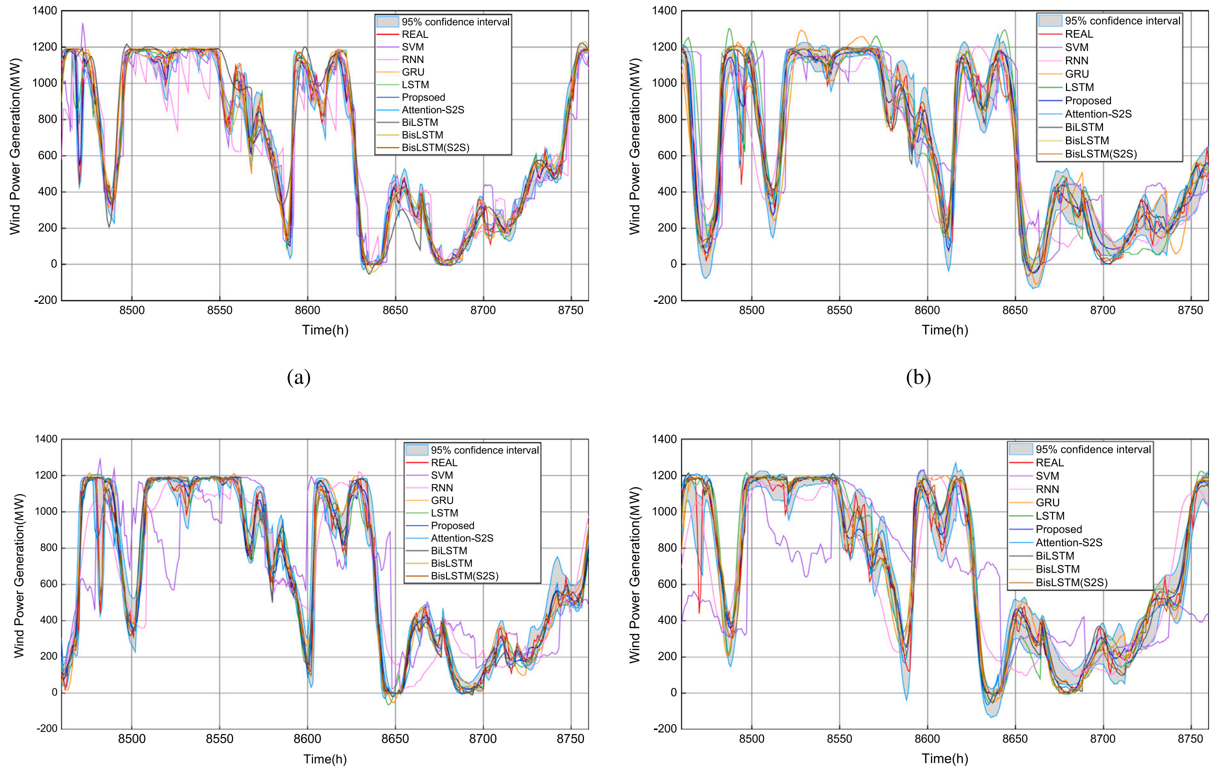
Table 2 presents a comprehensive view of predictive uncertainty for three targets: wind power generation (P_w), conductor temperature (T_{cable}), and temperature gradient (∇T) across four prediction horizons. A higher coverage (close to 1.0) indicates that most real values are within the predicted intervals, providing a reliable capture of uncertainty. Notably, shorter horizons (6 h, 12 h) often have higher coverage (e.g., 0.88 for P_w and 0.895 for T_{cable}), while longer horizons may drop in coverage as forecast uncertainty increases (e.g., 24 h and 48 h for P_w). A narrow interval width with high coverage typically reflects strong confidence in the model's predictions. We see that T_{cable} has relatively small widths (ranging from 6.92 to 11.538) compared to those of P_w, which can go as high as 317.326 at 48 h. This indicates that wind power forecasting carries more inherent variability (larger intervals) over extended horizons. Meanwhile, ∇T (temperature gradient) exhibits the smallest absolute MIWs, though it also shows an increase in interval width for longer horizons (from 1.21 at 6 h to 3.415 at 48 h).

The results of the performance evaluation of different models for conductor temperature and temperature gradient are shown in Tables 3 and 4, respectively. The proposed model achieves the lowest RMSE, indicating the smallest average prediction error, which remains impressively low even at 48 h (1.79 compared to the next best, GRU at 3.18). In the context of temperature-gradient predictions, the proposed model again demonstrates its robustness. It consistently shows the lowest RMSE and MAE across all forecast horizons, outshining other models.

Fig. 15 demonstrates the scatterplot of the predicted- and actual conductor temperature for all five comparison models. R2 measures the number of variations in the data, a higher value close to 1 indicates the perfect fit. It is observed that the best-performing forecasting model based on R2 values is a modified model.

4.3. Comparison of computational efficiency

To assess the computational overhead of our proposed method, we compare the training times of all benchmark and ablation models at various forecast horizons (6 h, 12 h, 24 h, 48 h), as detailed in Table 5. The results indicate that although the proposed model achieves the highest accuracy, it also requires the longest training time, ranging from 49 m 42 s to 56 m 26 s. In contrast, simpler models such as



(a) (b) (c) (d)

Fig. 12. Probabilistic results for wind power prediction on different predict horizons. (a) 6 h, (b) 12 h, (c) 24 h, (d) 48 h.

Table 2
Predictive interval uncertainty quantification.

Target	Coverage				Mean interval width			
	6 h	12 h	24 h	48 h	6 h	12 h	24 h	48 h
P_w	0.88	0.86	0.74	0.824	230.69	255.84	282.72	317.326
T_{cable}	0.869	0.895	0.904	0.968	6.92	8.14	9.49	11.538
∇T	0.75	0.78	0.791	0.918	1.21	1.51	1.939	3.415

Table 3
Conductor temperature predict performance evaluation.

Models	RMSE				MAE				MAPE				R^2			
	6 h	12 h	24 h	48 h	6 h	12 h	24 h	48 h	6 h	12 h	24 h	48 h	6 h	12 h	24 h	48 h
SVM [12–14]	3.84	6.63	10.60	14.11	2.52	4.35	7.43	10.54	0.07	0.13	0.24	0.37	0.94	0.81	0.42	0.37
RNN	3.64	5.90	5.24	5.54	2.44	4.15	3.67	3.86	0.07	0.13	0.11	0.12	0.94	0.82	0.87	0.87
GRU	1.67	1.96	2.37	3.18	0.94	1.12	1.32	1.88	0.02	0.03	0.03	0.05	0.99	0.99	0.98	0.96
LSTM	2.10	2.06	3.25	3.15	1.06	1.16	1.64	1.85	0.03	0.03	0.05	0.05	0.98	0.98	0.96	0.96
BiLSTM	1.98	2.00	3.10	3.05	1.04	1.12	1.55	1.83	0.02	0.03	0.04	0.05	0.98	0.98	0.96	0.96
BisLSTM	1.92	1.95	3.05	3.00	1.02	1.08	1.48	1.80	0.02	0.02	0.04	0.05	0.98	0.98	0.96	0.96
BisLSTM(S2S)	1.65	1.79	2.65	2.80	0.92	1.00	1.42	1.60	0.02	0.03	0.04	0.04	0.99	0.99	0.97	0.97
Attention-BisLSTM(S2S)	1.15	1.40	1.90	2.25	0.65	0.85	1.10	1.40	0.01	0.02	0.03	0.03	0.99	0.99	0.98	0.98
Proposed	0.74	0.96	1.35	1.79	0.42	0.57	0.77	1.05	0.01	0.02	0.02	0.03	0.99	0.99	0.99	0.99

Table 4
Temperature gradient predict performance evaluation.

Models	RMSE				MAE				MAPE				R^2			
	6 h	12 h	24 h	48 h	6 h	12 h	24 h	48 h	6 h	12 h	24 h	48 h	6 h	12 h	24 h	48 h
SVM [12–14]	2.33	3.80	5.57	7.10	1.43	2.37	3.78	5.22	0.43	0.86	1.63	3.33	0.89	0.70	0.18	0.10
RNN	2.11	3.11	2.81	3.00	1.33	2.12	1.90	2.01	0.35	0.66	0.62	0.86	0.90	0.74	0.81	0.80
GRU	1.01	1.18	1.39	1.83	0.52	0.62	0.72	1.00	0.15	0.17	0.21	0.32	0.98	0.98	0.97	0.94
LSTM	1.19	1.17	1.81	1.81	0.55	0.62	0.88	0.99	0.14	0.19	0.26	0.32	0.98	0.98	0.94	0.94
BiLSTM	1.15	1.14	1.75	1.76	0.54	0.59	0.84	0.93	0.13	0.18	0.24	0.31	0.98	0.98	0.95	0.95
BisLSTM	1.08	1.10	1.68	1.72	0.53	0.58	0.80	0.92	0.12	0.17	0.23	0.30	0.98	0.98	0.95	0.95
BisLSTM(S2S)	0.95	1.00	1.45	1.57	0.47	0.54	0.76	0.86	0.10	0.14	0.20	0.27	0.99	0.99	0.96	0.95
Attention-BisLSTM(S2S)	0.79	0.88	1.25	1.36	0.42	0.50	0.68	0.81	0.08	0.12	0.19	0.25	0.99	0.99	0.97	0.96
Proposed	0.47	0.60	0.84	1.10	0.24	0.33	0.44	0.60	0.06	0.08	0.12	0.16	0.99	0.99	0.98	0.98

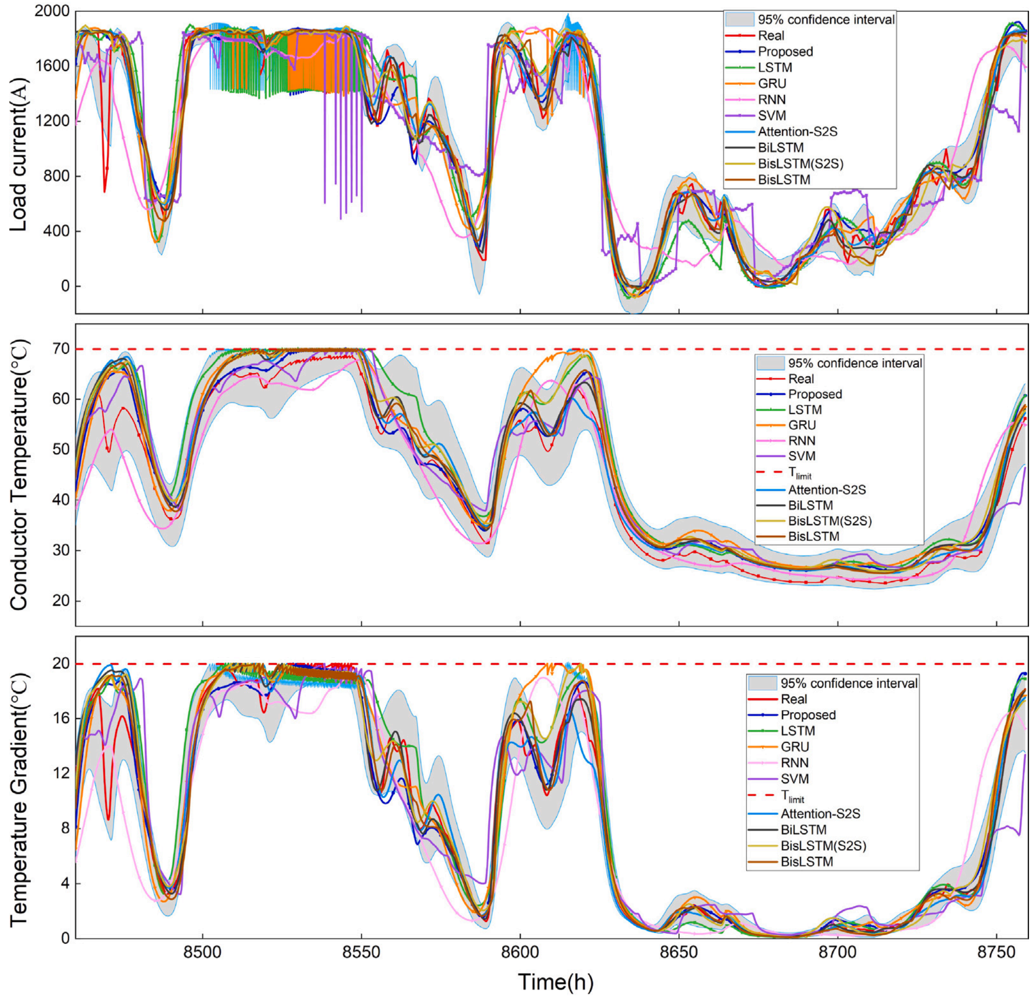


Fig. 13. Load Current, cable temperature, and temperature gradient prediction results under different models in the test data comprising 300 h.

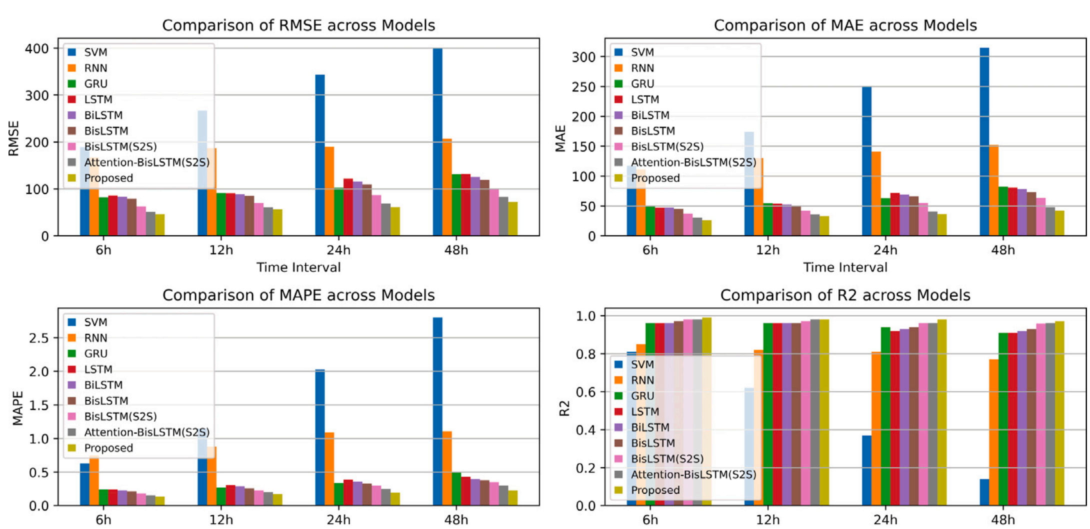


Fig. 14. Comparison of four error metrics for different models.

SVM can be trained in as little as 10–14 min, and RNN requires around 15–20 min. This additional time consumption mainly stems from the integrated multi-stage feature extraction and more complex network components, which substantially improve predictive performance at the cost of increased computational effort. These training durations remain practical compared to the relatively large forecast horizons (6–48 h), ensuring that the approach remains feasible for real-world wind power applications. Future work will focus on optimizing the proposed architecture to reduce training time while maintaining high predictive accuracy.

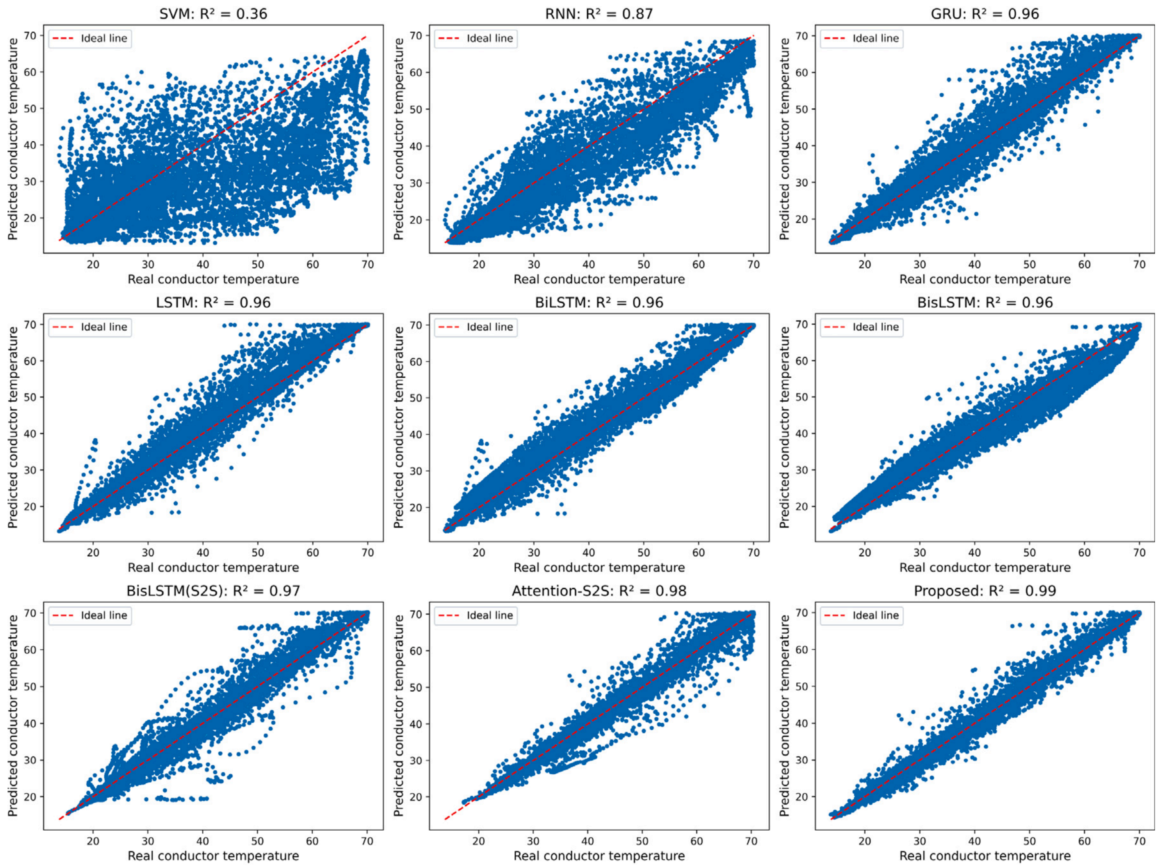


Fig. 15. 48 h ahead Scatterplot of forecast versus actual cable temperature for different models.

Table 5
Training time comparison.

Models	Training time			
	6 h	12 h	24 h	48 h
SVM	14 m 21 s	12 m 58 s	11 m 47 s	10 m 33 s
RNN	20 m 08 s	17 m 45 s	16 m 44 s	15 m 51 s
GRU	24 m 01 s	21 m 35 s	22 m 10 s	20 m 20 s
LSTM	23 m 35 s	21 m 17 s	21 m 00 s	19 m 37 s
BiLSTM	28 m 05 s	26 m 14 s	25 m 36 s	24 m 55 s
BisLSTM	31 m 43 s	29 m 12 s	28 m 08 s	27 m 16 s
BisLSTM(S2S)	35 m 21 s	33 m 02 s	32 m 10 s	30 m 49 s
Attention-BisLSTM(S2S)	41 m 12 s	38 m 55 s	37 m 33 s	36 m 21 s
Proposed	49 m 42 s	50 m 52 s	51 m 34 s	56 m 26 s

We observe a clear trade-off between accuracy and computational overhead: our proposed method significantly reduces errors (in RMSE, MAE) compared to its simpler counterparts. Given that an exact wind power forecast can substantially aid operational decision-making, we consider this extra computational investment justified. In practical settings where shorter training cycles are imperative, one could employ techniques such as model pruning, early stopping, or lighter architectures. However, even with the more demanding nature, the proposed model remains feasible for real-world deployment because it trains well within time frames typically allocated for operational wind power forecasts. Hence, it can be concluded that the proposed framework strikes a worthwhile balance: higher computational cost in exchange for significantly improved predictive performance.

4.4. Thermal overload risk estimation test

Finally, the thermal overload risk (TOR) estimate of different models is evaluated and summarized in Table 6. The models tested include SVM, RNN, GRU, LSTM, and the proposed method, with prediction horizons ranging from 6 to 48 h.

The analysis of the results reveals that, as the prediction horizon increases, a slight decrease in accuracy is observed across all models. However, the proposed model consistently outperforms traditional methods (SVM, RNN, GRU, and LSTM) in terms of accuracy and FPR. For the 6-hour prediction horizon, the proposed model achieves an accuracy of 0.99, outperforming the GRU (0.98) and the LSTM (0.97). Furthermore, the false positive rate (FPR) of the proposed model is the lowest across all horizons, reaching 0.0046 for the 6-hour prediction horizon, which is substantially better than the 0.031 of LSTM and the 0.020 of GRU for the same horizon. This suggests that the proposed method is particularly effective in reducing false alarms in short-term risk prediction.

As the horizon extends to 48 h, the performance of all models slightly degrades, as expected due to the increasing uncertainty in long-term predictions. However, the proposed model maintains superior performance with an accuracy of 0.98 and an FPR of 0.012, compared to the other models, such as LSTM, which show a decline to 0.968 accuracy and 0.04 FPR. The ATP values, which represent the average ability of the models to identify the overload events correctly, also indicate the robustness of the proposed method. Across all horizons, the proposed model achieves consistently high ATP values of around 0.19, comparable to other deep learning models like LSTM and GRU, which also achieve ATP values of around 0.19 to 0.20. However, the SVM model shows significantly lower ATP values, particularly at shorter horizons, indicating its weaker performance in capturing overload conditions accurately.

In conclusion, the proposed model demonstrates significant improvements in both accuracy and false positive rates across all time horizons compared to traditional machine learning models (SVM) and

Table 6
Thermal overload risk estimation of different models.

Models	Horizon	Acc.	ATP	ATN	FPR
SVM [12–14]	6 h	0.92	0.10	0.82	0.008
	12 h	0.91	0.11	0.81	0.009
	24 h	0.90	0.12	0.80	0.011
	48 h	0.89	0.13	0.79	0.015
RNN	6 h	0.95	0.14	0.81	0.006
	12 h	0.91	0.086	0.82	0.007
	24 h	0.91	0.096	0.81	0.010
	48 h	0.93	0.13	0.80	0.020
GRU	6 h	0.98	0.20	0.79	0.020
	12 h	0.978	0.195	0.78	0.020
	24 h	0.98	0.20	0.78	0.030
	48 h	0.97	0.19	0.77	0.030
LSTM	6 h	0.97	0.020	0.77	0.031
	12 h	0.978	0.198	0.78	0.027
	24 h	0.97	0.19	0.78	0.030
	48 h	0.968	0.20	0.77	0.040
BiLSTM	6 h	0.975	0.021	0.78	0.028
	12 h	0.980	0.19	0.79	0.026
	24 h	0.975	0.19	0.78	0.029
	48 h	0.970	0.20	0.78	0.035
BisLSTM	6 h	0.978	0.022	0.79	0.026
	12 h	0.982	0.193	0.79	0.024
	24 h	0.978	0.192	0.79	0.028
	48 h	0.972	0.20	0.79	0.033
BisLSTM(S2S)	6 h	0.982	0.023	0.79	0.024
	12 h	0.985	0.195	0.80	0.022
	24 h	0.982	0.195	0.79	0.026
	48 h	0.975	0.20	0.79	0.030
Attention-BisLSTM(S2S)	6 h	0.988	0.024	0.80	0.015
	12 h	0.989	0.198	0.80	0.012
	24 h	0.985	0.196	0.80	0.018
	48 h	0.980	0.20	0.80	0.025
Proposed	6 h	0.99	0.19	0.80	0.005
	12 h	0.99	0.19	0.80	0.006
	24 h	0.98	0.19	0.79	0.010
	48 h	0.98	0.18	0.80	0.012

advanced deep learning models (RNN, GRU, LSTM). These results validate the effectiveness of the proposed method for estimating thermal overload risk, particularly in scenarios that require high precision in both short- and long-term predictions.

4.5. DCR economic beneficial analysis considering temperature and temperature gradient limitation

This section deals with the necessity of the introduced temperature gradient limitation and elaborates on the economic benefits and aging implications of two DTR power curtailment strategies: one considering only the temperature limitation (DTR1), and the other considering both temperature and temperature gradient limitations (DTR2). Firstly, the 'electric field distortion rate' κ is defined:

$$\kappa(r,t) = \frac{E_{\max}(r,t) - E_{\min}(r)}{E_{\max}(r)}. \tag{69}$$

In the absence of high temperature or conductivity gradients, κ remains close to zero; large positive or negative swings indicate a strong distortion or even a local polarity reversal of the DC field.

Comparing the blue (DTR1) and green (DTR2) curves in Fig. 16 shows that including the temperature gradient constraint cuts the peak-to-peak spread of κ by roughly a factor of two, thus more than halving the range of field distortion. DTR2 also reduces the incidence of negative excursions (which may correspond to localized polarity reversal). This improvement arises directly from suppressing large temperature gradients over the cable's cross-section, which in return limits large local changes in conductivity $\sigma(r,t)$ and prevents sharp electric field intensification.

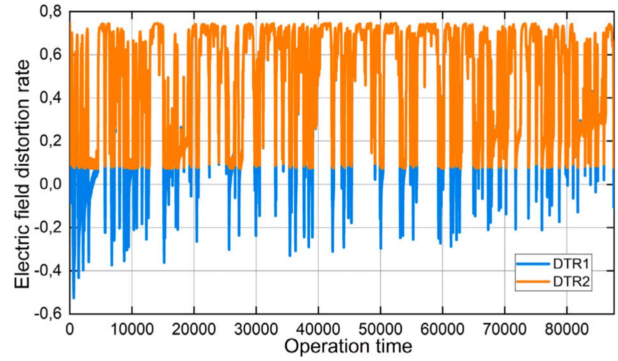


Fig. 16. Electric field distortion rate variation.

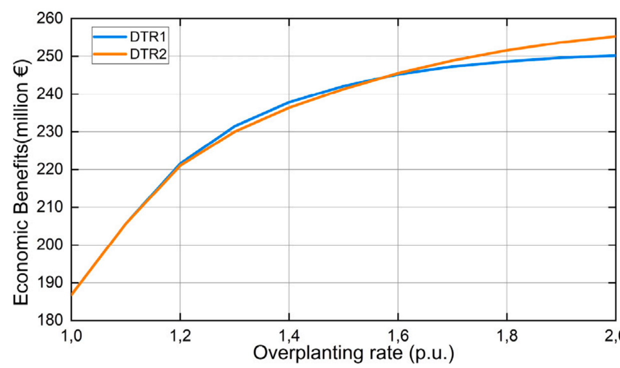


Fig. 17. Economic beneficial analysis of different overplanting rates.

From a modeling standpoint, the key parameter is the field inversion coefficient δ in Eqs. (12) and (13), which depends on both the temperature drop ΔT_i and the mean-field E_m. By limiting ΔT_i in real time, DTR2 effectively constrains the growth of δ in localized regions, by mitigating the "field inversion" phenomenon that triggers very high local electric fields. Furthermore, the exponential dependence of σ on temperature [see Eq. (14)] is less pronounced under the smoother thermal distributions enforced by DTR2. The uniform conductivity profile results in fewer electric field spikes and better insulation reliability bolstering overall system stability, prolonging cable lifespan, and reducing the likelihood of thermally driven polarity reversals in the DC field.

Then, the economic benefits were analyzed, as shown in Fig. 17. As illustrated in Fig. 17, both DTR1 and DTR2 yield the same economic benefits (approximately 180–230 M€) when the overplanting rate is below 1.2 p.u., as no power curtailment thresholds are triggered. In the intermediate range of 1.2–1.6 p.u., the extra thermal-gradient constraints in DTR2 lead to minor power curtailments, reducing economic returns compared to DTR1 (e.g., at 1.5 p.u. the DTR2 revenue is about 3 M€ lower). However, when the overplanting rate exceeds 1.6 p.u., the benefits of mitigating extreme temperature gradients and consequent insulation aging begin to outweigh these curtailment losses. As a result, at 2.0 p.u. DTR2 yields approximately 2 M€ higher revenue than DTR1, reflecting improved long-term reliability and reduced electric-field distortion.

From a practical engineering perspective, selecting an optimal overplanting rate requires balancing short-term gains against potential thermal stress on the cable. For moderate overplanting (below 1.6 p.u.), DTR1 offers slightly higher revenue without additional complexity.

However, for higher rates, adopting DTR2 provides more robust temperature distribution control, and reduces the risk of insulation degradation and outage costs. Generally, operators should estimate the lifetime impact of thermal gradients particularly at the higher end of overplanting rates. Regular monitoring of conductor temperature and insulation integrity is needed for the full leverage of benefits of advanced DTR strategies.

4.6. Discussion

4.6.1. Advantage of the model

While conventional recurrent architectures (e.g., GRU, standard LSTM) are widely used for time series forecasting, they face limitations when dealing with highly volatile and long-range dependent signals, as commonly observed in offshore wind farms. In particular, GRUs and basic LSTMs struggle to capture intricate temporal dependencies over extended horizons, leading to suboptimal multi-step predictions. Our proposed Attention-BNN-S2S framework, incorporating the xLSTM design, addresses these shortcomings through the following features:

(1) Enhanced Gating via xLSTM. Unlike standard LSTM cells that rely on static sigmoid or tanh gates, the xLSTM module introduces exponential or softmax-based gating and scalar states. This refinement promotes robust control over long-range information flow, mitigating vanishing or exploding gradients. As a result, the model takes into account the nuanced diurnal and seasonal patterns characterizing offshore wind.

(2) Seq2Seq Architecture with Additive Attention. Traditional GRU/LSTM models often rely on a fixed-length internal representation, which can cause information loss over long sequences. By contrast, the Sequence-to-Sequence (S2S) framework encodes variable-length input into a context vector, and the additive attention mechanism selectively “highlights” the most relevant historical steps for each decoding moment. This targeted focus is especially crucial for offshore wind datasets, which can exhibit rapid and nonlinear fluctuations under changing weather fronts.

(3) Bayesian Inference via MC Dropout. One key advantage over deterministic GRU/LSTM models is our approximate Bayesian component, which leverages Monte Carlo Dropout. Beyond providing a single mean prediction, the model quantifies epistemic uncertainty through repeated stochastic forward passes. This is invaluable in offshore settings, where prediction confidence can be as important as the point forecast.

(4) Scalability and Ease of Integration. Unlike more complex Bayesian methods (full Markov chain Monte Carlo), MC Dropout requires minimal changes to training procedures, making it practical to deploy. Combined with additive attention and sLSTM cells, the overall structure remains modular, simplifying adaptation to large-scale offshore wind datasets.

4.6.2. Limitation of the model

While the proposed Attention-BNN-S2S-DCR model provides a systematic framework for predicting wind power generation, determining cable ratings, and estimating thermal risk, it also has certain limitations that must be acknowledged.

(1) The accuracy of the wind power generation predictions is highly reliant on the availability of high-quality historical data for features such as wind speed components, wind direction, and sea surface roughness. For scenarios where these measurements are noisy, sparse, or unavailable, the model’s performance may significantly deteriorate.

(2) Power curtailment is triggered when the predicted conductor temperature or temperature gradient exceeds predefined thresholds. Real-world operation might involve additional constraints, for instance, market demands, contractual obligations, or dynamic ramp-rate limits that this simplified curtailment strategy does not address.

(3) The thermoelectric equivalent model that is used for cable rating and temperature estimation simplifies real-world thermal dynamics such as uneven soil layers and transient ambient conditions. In reality, these factors can vary over time and space, potentially causing discrepancies between the predicted and actual conductor temperatures. By design, the Attention-BNN-S2S framework employs MC Dropout to approximate Bayesian inference. While this approach quantifies uncertainty, it requires multiple stochastic forward passes during inference, which can increase computational costs—particularly for large-scale deployments or systems demanding real-time responses.

4.6.3. Potential future applications for other renewable energy systems

The proposed framework lays a foundation for dynamic cable rating in offshore wind farms by proactively predicting power generation, estimating thermal loads, and enabling timely power curtailment when thermal risks are high. Its underlying principles, particularly the sequence-to-sequence structure with the probabilistic inference layer can be adapted for other renewable energy domains:

(1) Solar power integration: The sequence-to-sequence approach could be extended to forecast solar irradiation and power output in photovoltaic (PV) systems. Just like wind power forecast, solar power forecasts will enable dynamic cable rating and proactive risk mitigation, especially in regions where overcurrent or conductor overheating may be a concern.

(2) Hybrid renewable grids: In multi-source grids combining wind, solar, and other renewables, an enhanced version of this model could simultaneously learn from multiple environmental variables such as cloud cover, humidity, and wind conditions to produce a comprehensive load forecast. The TEE model would then incorporate varying load profiles from these sources to better estimate thermal risks and coordinate power curtailment across different generation assets.

(3) Battery storage and demand response: Future work could explore how to integrate energy storage systems or demand response strategies into the cable rating decision loop. By accurate generation and load prediction, operators can schedule charging/discharging or load-shifting measures to keep conductor temperatures within secure limits.

5. Conclusion

This paper proposed a novel algorithm: Attention-BNN-S2S-DCR model to predict DCR of HVDC cables. The analysis and contribution of this paper draw the following conclusions:

(1) The Attention-BNN-S2S model utilizes an encoder–decoder structure with a Bi-sLSTM network. This model incorporates BNN to produce probabilistic forecasts and an additive attention mechanism to capture long-term dependencies. The superiority of the proposed model is established through comparative analyses with various machine learning techniques.

(2) A computational model of cable temperature is developed based on the TEE method. Then, the accuracy of the TEE model was validated by applying the steady-state and transient conditions of the COMSOL model.

(3) Finally, this paper proposes a novel method for estimating the thermal overload risk for HVDC cables, which considers not only the limits of the conductor temperature but also the limits of the temperature gradient. Simulation results using real wind profiles and varying overplanting rates (from 1.2 p.u. to 2.0 p.u.) show that stricter controls of the temperature gradient, initially cause slight revenue decreases due to curtailments. However, for higher overplanting rates (above 1.6 p.u.), the long-term benefits, such as mitigating insulation degradation and reducing outage risk, ultimately provide improved reliability and higher net income compared with strategies without gradient constraints.

CRediT authorship contribution statement

Shen Yan: Writing – original draft, Conceptualization. **Mohamad Ghaffarian Niasar:** Supervision. **Marjan Popov:** Supervision.

Declaration of competing interest

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

Acknowledgments

This work was supported by the Chinese Scholarship Council (CSC) (grant no. 202206660001).

Data availability

Data will be made available on request.

References

[1] Eslahi M, Vahidi B, Siano P. A flexible risk-averse strategy considering uncertainties of demand and multiple wind farms in electrical grids. IEEE Trans Ind Informat 2022;18(4):2255–63.
[2] Kazmi SH, et al. Offshore windfarm design optimization using dynamic rating for transmission components. IEEE Trans Power Syst 2022;37(3):1820–30.
[3] Moradzadeh A, et al. Deep learning-based cyber resilient dynamic line rating forecasting. Int J Electr Power Energy Syst 2022;142:108257.
[4] Aizpurua JI, et al. Probabilistic machine learning aided transformer lifetime prediction framework for wind energy systems. Int J Electr Power Energy Syst 2023;153:109352.
[5] Daminov I, et al. Assessment of dynamic transformer rating, considering current and temperature limitations. Int J Electr Power Energy Syst 2021;129:106886.
[6] Yu Z, Liu X, Li Z. A secure dispatch strategy subject to the uncertainties of DLR and wind power. IEEE Trans Ind Informat. 2022;18(7):4499–509.
[7] Pérez-Rúa J-A, Das K, Cutululis NA. Optimum sizing of offshore wind farm export cables. Int J Electr Power Energy Syst 2019;113:982–90.
[8] Brakelmann H, Anders GJ. Current rating considerations in designing HVDC cable installations. IEEE Trans Power Deliv 2018;33(5):2315–23.
[9] Yan S, et al. Effect of multi-factors on heterocharges for oil-impregnated paper in converter transformer using modified charge transport model. IET Gener Transm Distrib 2021;15(21):3048–57.
[10] Colin MAH, Pilgrim JA. Offshore cable optimization by probabilistic thermal risk estimation. In: Proc. IEEE int. conf. probabilistic methods appl. to power syst.. 2018, p. 1–6.
[11] Colin MAH, Pilgrim JA. Cable thermal risk estimation for overplanted wind farms. IEEE Trans Power Deliv 2020;35(2):609–17.
[12] Huang R, Pilgrim JA, Lewin PL, Payne D. Dynamic cable ratings for smarter grids. In: Proc. IEEE PES ISGT Europe. 2013, p. 1–5.
[13] Huang R, et al. Managing cable thermal stress through predictive ratings. In: Proc. IEEE elect. insul. conf.. 2015, p. 110–3.
[14] Huang R, et al. Use of day-ahead load forecasting for predicted cable rating. In: Proc. IEEE PES innovative smart grid technol.. 2014, p. 1–6.
[15] Liu S, Kopsidas K. Risk-based underground cable circuit ratings for flexible wind power integration. IEEE Trans Power Deliv 2021;36(1):145–55.
[16] Sun Z, et al. Predicting bridge longitudinal displacement from monitored operational loads with hierarchical CNN for condition assessment. Mech Syst Signal Process 2023;200:110623.
[17] Lei X, et al. Interpretable machine learning methods for clarification of load-displacement effects on cable-stayed bridge. Measurement 2023;220:113390.
[18] Lei X, et al. Displacement response estimation of a cable-stayed bridge subjected to various loading conditions with one-dimensional residual convolutional autoencoder method. Struct Heal Monit 2023;22(3):1790–806.
[19] Chen H, et al. Low-carbon economic dispatch of integrated energy system containing electric hydrogen production based on VMD-GRU short-term wind power prediction. Int J Electr Power Energy Syst 2023;154:109420.
[20] Hossain MA, et al. Optimized forecasting model to improve the accuracy of very short-term wind power prediction. IEEE Trans Ind Informat 2023;19(10):10145–59.
[21] Arora P, et al. Probabilistic wind power forecasting using optimized deep auto-regressive recurrent neural networks. IEEE Trans Ind Informat 2023;19(3):2814–25.
[22] Lv L, et al. A VMD and LSTM Based Hybrid Model of Load Forecasting for Power Grid Security. IEEE Trans Ind Inf 2022;18(9):6474–82.
[23] Liu F, et al. Deep reinforcement learning based energy storage management strategy considering prediction intervals of wind power. Int J Electr Power Energy Syst 2023;145:108608.
[24] Blonbou R. Very short-term wind power forecasting with neural networks and adaptive Bayesian learning. Renew Energy 2011;36(3):1118–24.
[25] Sahay KB, Shukla P. One hour-ahead wind speed forecasting of sotavento galicia SA wind farm by using ANN. In: Proc. int. elect. eng. congr.. 2018, p. 1–4.
[26] Liu Y, et al. Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model. Appl Energy 2020;260:114259.
[27] Diban B, et al. Life-based geometric design of HVDC cables—Part 2: Effect of electrical and thermal transients. IEEE Trans Dielectr Electr Insul 2023;30:97–105.
[28] Kapsali M, Kaldellis JK. 2.14 - offshore wind power basics. Compr Renew Energy 2012;431–68.
[29] Beck M, et al. xLSTM: Extended long short-term memory. 2024, ArXiv Preprint 2405 (04517).
[30] Ji Z, Xiong K, Pang Y, Li X. Video summarization with attention-based encoder–decoder networks. IEEE Trans Circuits Syst Video Technol 2020;30(6):1709–17.
[31] Li X, Zou C, Peng C, Wu C. Uncertainty quantification in intelligent-based electrical resistivity tomography image reconstruction with Monte Carlo dropout strategy. IEEE Trans Geosci Remote Sens 2023;61:1–6.
[32] Olsen RS, Holbøll J, Gudmundsdottir US. Dynamic temperature estimation and real-time emergency rating of transmission cables. In: Proc. IEEE power and energy soc. gen. meet.. 2012, p. 1–5.
[33] Grothe O, Kächele F, Watermeyer M. Analyzing Europe's biggest offshore wind farms: A data set with 40 years of hourly wind speeds and electricity production. Energies 2022;15(5):1700.
[34] Hersbach H, et al. ERA5 hourly data on single levels from 1979 to present. 2018, [Online]. Available: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview. [Accessed: April 2023].
[35] Archer CL, Jacobson MZ. Supplying baseload power and reducing transmission requirements by interconnecting wind farms. J Appl Meteorol Clim 2007;46(11):1701–17.
[36] Liu Y, et al. Simulation of electric field distribution in the XLPE insulation of a 320 kV DC cable under steady and time-varying states. IEEE Trans Dielectr Electr Insul 2018;25(3):954–64.
[37] Rizvi Z, Wuttke F, Sembdner K. Thermal analysis of submarine power cable considering natural convection. In: Proc. 1st international conference on geo-energy and geo-environment. 2015, p. 1–4.
[38] Shi Z, et al. Simulation study on ±320kV DC power cable steady-state temperature field distribution and its influencing factors. In: IOP conf. ser.: earth environ. sci.. 2020, 012001.
[39] Xie S, et al. Triple jumps of XLPE insulated HVDC cable development in China: from 160 kV, 200 kV to 320 kV. In: Proc. 9th int. conf. insulated power cables (jicable'15). 2015, p. 1–6.
[40] EU Copernicus Marine Service Information. 2015. [Online]. Available: https://doi.org/10.48670/moi-00059. [Accessed: April 2023].