Secure Probabilistic Prediction of Dynamic Thermal Line Rating

N. Safari, *Member*, *IEEE*, S. M. Mazhari, *Member*, *IEEE*, C. Y. Chung, *Fellow*, *IEEE*, and S. B. Ko, *Senior Member*, *IEEE*

***Abstract*—Accurate short-term prediction of overhead line (OHL) transmission ampacity can directly affect the efficiency of power system operation and planning. Any overestimation of the dynamic thermal line rating (DTLR) can lead to the lifetime degradation and failure of OHLs, safety hazards, etc. This paper presents a secure yet sharp probabilistic model for the hour-ahead prediction of the DTLR. The security of the proposed DTLR limits the frequency of DTLR prediction exceeding the actual DTLR. The model is based on an augmented deep learning architecture that makes use of a wide range of predictors, including historical climatology data and latent variables obtained during DTLR calculation. Furthermore, by introducing a customized cost function, the deep neural network is trained to consider the DTLR security based on the required probability of exceedance while minimizing the deviations of the predicted DTLRs from the actual values. The proposed probabilistic DTLR is developed and verified using recorded experimental data. The simulation results validate the superiority of the proposed DTLR compared with the state-of-the-art prediction models using well-known evaluation metrics.**

***Index Terms*—Deep neural network, dynamic thermal line rating, overhead line, prediction, recurrent neural network.**

I. Introduction

THERMAL line rating (TLR) is the primary culprit limiting the current carrying capability of the overhead line (OHL) [1]. Both IEEE [2] and CIGRE [3] have put forward TLR calculation methods. Although these methods provide very similar results, the one proposed by IEEE is simpler and easier to use [4]. The TLR of OHL is a weather-dependent variable and conventionally calculated using the heat balance equation in the worst-case weather scenarios [5]. Despite its conservativeness in some cases, this static TLR (STLR) might exceed the real OHL thermal constraint. Consequently, the OHL might be exposed to damage due to the lack of TLR monitoring.

To overcome the shortcomings of the STLR, dynamic TLR (DTLR) is proposed in which the thermal condition of OHL can be monitored. Thus, DTLR unlocks the additional capacity headroom of current OHLs in a secure way, thereby addressing network congestion and postponing/eliminating the need for transmission expansion [6]. As a significant additional benefit, DTLR facilitates the delivery of highly variable and uncertain power from renewable energy systems (RESs) to the end-users due to perceptible correlation between RES generation and the additional capacity provided by the DTLR. For such reasons, DTLR has recently grasped the attention of governments and transmission companies and is considered as an enabling tool for enhancing the penetration of RESs [7]-[9]. DTLR applications and technologies are comprehensively reviewed in [10], [11].

DTLR is a function of several climatology variables such as wind speed, wind direction, etc. [2]. Therefore, its values for upcoming hours need to be predicted. The DTLR prediction can be employed in various power system problems such as unit commitment, economic dispatch, optimal power flow, etc. [5], [12]. In this respect, specialized research communities have devoted momentous efforts to develop DTLR monitoring and prediction models [9] - [22]. As the DTLR monitoring and prediction may not be feasible through the entire OHL, critical spans are identified for this purpose. In [23] - [25], heuristic methods are brought forward to identify the number and locations of the monitoring stations required to make the OHL fully observable from the TLR perspective.

Relying on the literature, the DTLR prediction has been interpreted with two different viewpoints. Some researchers predict the maximum allowable current at the OHL thermal limit [12]-[16]. In contrast, others consider DTLR prediction as estimating the future value of an OHL temperature provided that it carries a certain amount of current [19] - [20]. While both perspectives bring about intriguing advancements to the field and provide insightful information about OHL thermal constraints, this paper focuses on DTLR prediction based on the first definition.

DTLR can be broadly divided into direct [12] - [18], indirect [19], and hybrid [20] - [22] methods. In the direct method, the required values are calculated based on the straightforward computation of the conductor's maximum allowable

Manuscript received: August 28, 2020; revised: October 26, 2020; accepted: November 16, 2020. Date of CrossCheck: November 16, 2020. Date of online publication: January 25, 2021.

The views expressed in this paper are those of the authors and do not represent the view of SaskPower. This paper was prepared when N. Safari was with the Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK, S7N 5A9, Canada. The work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Saskatchewan Power Corporation (SaskPower).

N. Safari (corresponding anthor) is with Grid Operations Support, SaskPower, Regina, SK, S4P 0S1, Canada (e-mail: n.safari@usask.ca).

S. M. Mazhari, C. Y. Chung, and S. B. Ko are with the Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK, S7N 5A9, Canada (e-mail: s.m.mazhari@usask.ca; c.y.chung@usask.ca; seokbum.ko@usask.ca).

DOI: 10.35833/MPCE.2020.000641

current considering the maximum permissible temperature. In this respect, the process is accomplished by employing weather data and a heat balance equation. In contrast, the measured values such as sag position, mechanical tension, etc., are used in indirect DTLR calculation. In the hybrid method, the indirect DTLR calculation along with weather-related factors is employed. Direct methods are currently of notable interest as they are less dependent on external equipment and lead to low-cost outcomes compared with their counterparts [7].

From the prediction output, DTLR prediction models can be divided into deterministic [18]-[22] and probabilistic [12]-[17] methods. In the deterministic method, a single quantity associated with the most likely value of DTLR in the next sample point is considered as the output. In the probabilistic method, the information of uncertainty in the DTLR prediction can be acquired. Since any overestimation of DTLR may lead to unprecedented issues, the probabilistic prediction is the most welcome by power system operators [5]. Accordingly, this is the focus of this paper. The recommendations of IEEE and CIGRE joint task force (JTF) [4] further emphasizes the importance of probabilistic prediction of DTLR. According to JTF's recommendation, the conductor temperature of OHL should be less than the maximum allowable conductor temperature for 99% of the time [4]. Therefore, it is critical to assess the performance of the DTLR prediction models for extreme probability of exceedance (POE) values, e.g., 99%. However, this is not well-discussed in the recent literature [12], [15]-[17].

In [12] and [13], parametric distributions model the uncertainties associated with climatology variables. In [14], Taylor series expansion is employed to find the mean and variance of DTLR in the coming hours based on the mean and variance of the forecasted values, which correspond to the meteorological variables. To consider the interdependency among meteorological variables in the DTLR prediction, multivariate Gaussian distributions, resulting from different meteorological variables, are used in a Monte Carlo simulation process to extract the distribution of DTLR [26].

In [12]-[24], [25], a parametric representation is assumed for the uncertainties associated with meteorological variables. However, this assumption may be erroneous due to the high nonstationarity of the meteorological time series (TS). To this end, [15] proposes a non-parametric autoregression framework for probabilistic prediction of DTLR based on quantile regression (QR). References [5] and [16] put forward a non-parametric prediction of DTLR based on QR forest (QRF), in which meteorological measurements and numerical weather predictions (NWPs) compose the input features of the prediction model. Overall, the studies on non-parametric DTLR prediction are still in their infancy. Thus, this paper is devoted to further contributing to the current literature.

Weather-based DTLR prediction models mainly limit their inputs to those features obtained directly from meteorological measurements. However, from the DTLR formulations proposed in [2] and [3], the relations of meteorological variables with DTLR value are evidently both complex and nonlinear. In addition, many latent variables, e. g., convection cooling, radiated heat loss rate, etc., are acquired in the process of DTLR calculation and may provide the information about the complex relationship between the DTLR value and meteorological variables. However, these important predictors have thus far not been considered in DTLR prediction. Hence, this paper scrutinizes the impact of latent variables in the DTLR prediction.

The remarkable advancements in deep learning and its successful implementation in prediction have resulted in the enhancement of prediction accuracy for a range of power system applications [27]-[30]. In deep neural network (DNN) architectures, highly efficient unsupervised dimension reduction blocks, i.e., autoencoder (AE) variants, can be employed to tackle the high-dimension feature space issues, posed due to numerous meteorological and latent variables [30], [31]. Using DNNs, more complex patterns, which cannot be identified in shallow networks, can be perceived. Compared with various building blocks of DNNs, long short-term memory (LSTM) [32], which is a recurrent neural network (RNN), has demonstrated superior performance in meteorological variable prediction problems [28]. The benefit of LSTM compared with conventional RNNs is its ability to capture long- and short-term dependencies in a sequence while addressing the vanishing and exploding gradient problems of prevalent RNNs [33]. Despite successful applications of RNNs, specifically LSTM, and the benefits of DNNs, these methods have not been adapted to the DTLR prediction. In this paper, LSTM is employed to develop the DTLR prediction, while the stacked denoising AE (SDAE) [31] is developed for unsupervised feature learning and extraction. Most of the DNN models developed so far have been trained deterministically in power system applications [27]-[30]. A major impediment that makes past methods incapable of offering a probabilistic model in DNNs may originate from the lack of a proficient probabilistic cost function. Such function should consider the reliability level and sharpness together so that the DNN parameters are tuned according to the system operator's preference. To address this need, inspired by the optimization function in QR problem, a cost function is presented. The proposed cost function can be used to train the model for the preferred POEs.

This paper proposes a DTLR for predicting the various POEs using accessible latent variables in addition to meteorological measurements. A DTLR model is developed using the SDAE and LSTM unit in the DNN architecture. The prediction engine is trained by considering a novel cost function to meet the JTF recommendation, while the sharpness is maximized. The performance of the proposed models is compared with the state-of-the-art DTLR prediction models using publicly available data. Briefly, the main contributions of this work are three-fold.

1) For the first time, latent variables are introduced in DTLR prediction as valuable predictors.

2) A cost function is proposed to train the deep-learning model for probabilistic prediction.

3) A deep learning model is trained for the DTLR prediction.

The remainder of the paper is organized as follows. Section II presents the proposed framework for DTLR prediction. The proposed training framework for probabilistic deep-learning-based prediction of DTLR is explained in Section III. Data, evaluation metrics, and simulations results are presented in Section IV, followed by discussion in Section V and remarkable conclusions in Section VI.

II. Proposed Framework for DTLR Prediction

Define $\{WS_t\}_{t=1}^{N}$ and $\{WD_t\}_{t=1}^{N}$ as TSs associated with wind speed and wind direction, respectively, where N is the length of the TSs, and t is the t^{th} sample of the TS. Also, the TSs of the wind speed components decomposed by a Cartesian coordinate system are denoted by $\{WSX_t\}_{t=1}^{N}$ and $\{WSY_t\}_{t=1}^{N}$, respectively. The TSs of ambient temperature and solar irradiance are denoted by $\{AT_t\}_{t=1}^{N}$ and $\{SI_t\}_{t=1}^{N}$, respectively.

Figure 1 demonstrates the overall framework of the proposed DTLR prediction model. The suitable number of lags l to form the input vector is identified using the embedding dimension method [34]. From the measurements, l lags of meteorological TS are imported to the calculation model of DTLR described in Section II-A. Using the model, $[DTLR_{t-l+1}, DTLR_{t-l+2}, \ldots, DTLR_t]$, $[qc_{t-l+1}, qc_{t-l+2}, \ldots, qc_t]$, and $[qr_{t-l+1}, qr_{t-l+2}, \ldots, qr_t]$, which represent the DTLR, convection cooling, and radiated heat loss rate for lags of meteorological inputs, respectively, are calculated. The features obtained from this step along with the meteorological input vector are utilized in a feature reduction and extraction stage as discussed in Section II-B. In Fig. 1, $[f_{t-l+1}^{m}, f_{t-l+2}^{m}, \ldots, f_t^{m}]$ $(m = 1, 2, \ldots, M)$ represents a vector associated with the m^{th} feature resulting from the dimension reduction and feature extraction stage, and M and l are the total number features and number of lags, respectively. Thereafter, various trained models are employed to acquire the final DTLR prediction $\underline{DTLR}_{t+1}$ as elaborated in Section III.

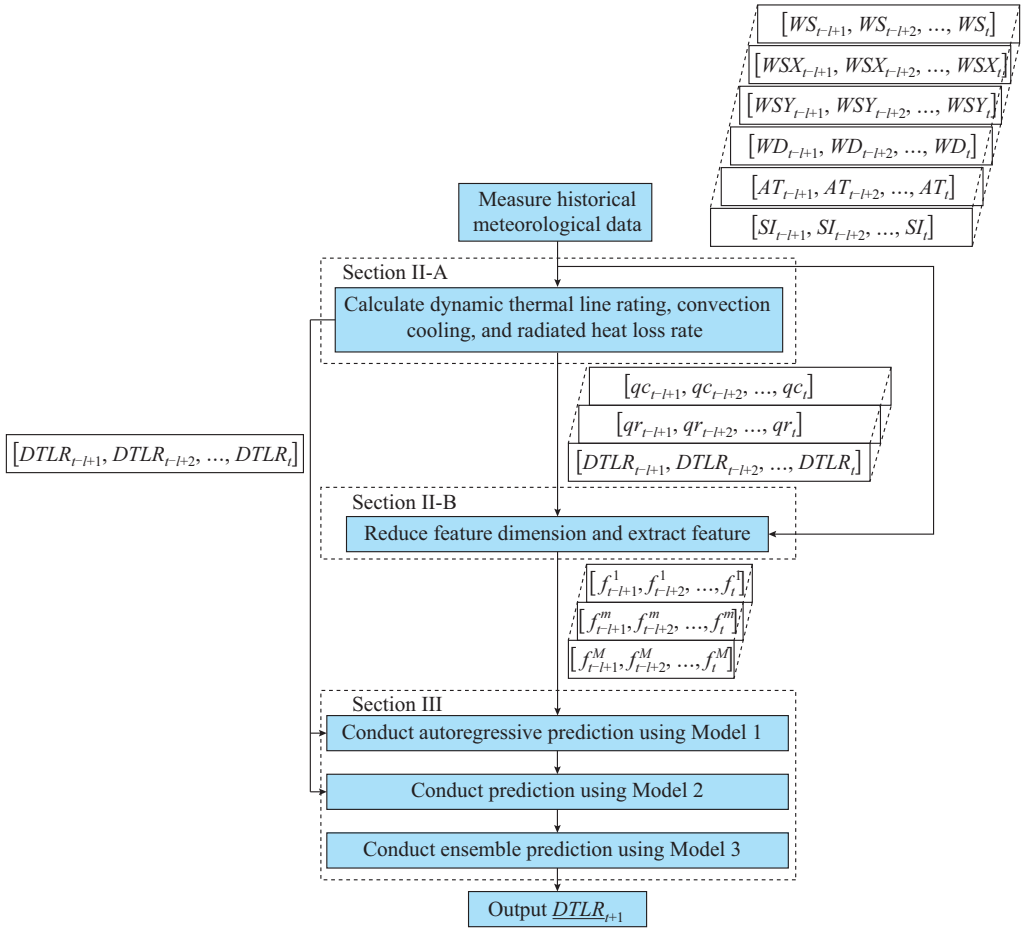

Fig. 1. Overall framework of proposed DTLR prediction model.

A. DTLR Calculation

In this study, the power system is assumed to be in normal operation. Therefore, the current fluctuations and the resulting variations in the OHL temperature can be considered negligible, provided that the system does not require any abrupt or temporary switching [21]. Studies reveal that the maximum time required to reach the steady state due to a step change in current is approximately 30 minutes [35]. Based on these points, the transients in DTLR can be ignored, and DTLR can be estimated for every hour.

Figure 2 schematically represents the factors influencing the DTLR of OHLs.

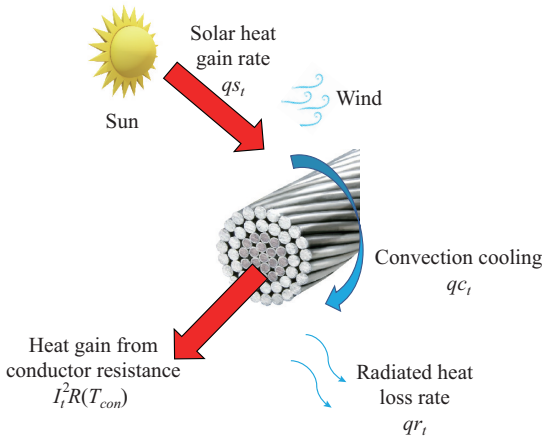

Fig. 2. A schematic diagram of DTLR.

As per IEEE Std 738-2012, in the steady state, the heat balance equation for an OHL at the t^{th} sample can be written as follows [2]:

$$qc_t + qr_t = qs_t + I_t^2 R(T_{con}) \tag{1}$$

where qc_t and qr_t are the convection cooling and radiated heat loss rates per unit length, respectively; qs_t is the heat gain rate from the sun; $R(T_{con})$ is the alternating current (AC) resistance associated with the conductor temperature T_{con}; and I_t is the allowable conductor current at T_{con}, which can be simply obtained as:

$$I_t = \sqrt{\frac{qc_t + qr_t - qs_t}{R(T_{con})}} \tag{2}$$

From (1) and (2), one can observe that qc_t and qr_t are the cooling elements in the heat-balance equation, and that their increase helps obtain more OHL ampacity. qs_t is a heating component and is a culprit of ampacity reduction. qc_t is a function of WS_t, WD_t, and AT_t. qc_t is calculated as [2]:

$$qc_t = \max(qc_t^{f1}, qc_t^{f2}, qc_t^{n}) \tag{3}$$

$$qc_t^{f1} = K_a\left[1.01 + 1.35 \times (N_{Re}^t)^{0.52}\right] k_f (T_{con} - AT_t) \tag{4}$$

$$qc_t^{f2} = K_a \times 0.754 \times (N_{Re}^t)^{0.6} k_f (T_{con} - AT_t) \tag{5}$$

$$qc_t^{n} = 3.645 \times \rho_f^{0.5} D_o^{0.75} (T_{con} - AT_t) \tag{6}$$

$$N_{Re}^t = \frac{D_o \rho_f WS_t}{\mu_f} \tag{7}$$

where qc_t^{f1}, qc_t^{f2}, qc_t^{n}, and N_{Re}^t are intermediate variables; K_a is the wind direction factor; ρ_f, μ_f, k_f are the air density, absolute air viscosity, and coefficient of thermal conductivity of air, respectively; and D_o is the outside diameter of the conductor. As can be seen from (3)-(7), qc_t is a nonlinear and complicated function of meteorological variables while the relationship between qc_t and I_t is simple as shown in (2). Therefore, it can be beneficial to consider $\left[qc_{t-l+1}, qc_{t-l+2}, \ldots, qc_t\right]$ as the elements of the predictor set for predicting $DTLR_{t+1}$. Moreover, qr_t in (1) and (2) can be calculated as [2]:

$$qr_t = 17.8 \times D_o \epsilon \left[\left(\frac{T_{con}+273}{100}\right)^4 - \left(\frac{AT_t+273}{100}\right)^4\right] \tag{8}$$

where ϵ is the emissivity and has a value between 0.23 and 0.91, which increases with the age of conductor. As can be seen from (8), qr_t is related to the fourth power of AT_t. Therefore, it may not be adequate to consider the simple vector of $\left[AT_{t-l+1}, AT_{t-l+2}, \ldots, AT_t\right]$ as the elements of the feature set to reflect the importance of ambient temperature and radiative heat loss in the DTLR prediction. It is worthful to analyze the influence of considering historical data of radiated heat loss rate. For a solar irradiance at time t SI_t, the rate of solar heat gain qs_t can be estimated by a linear function of SI_t [2]. Therefore, its consideration as a feature cannot be informative.

The calculated DTLR, $\left[DTLR_{t-l+1}, DTLR_{t-l+2}, \ldots, DTLR_t\right]$, as well as the latent variables obtained during the DTLR calculation, is used in a feature reduction and feature learning stage as elucidated below.

B. Feature Reduction and Extraction in DTLR Prediction

As detailed in Section II-A, several climatic variables strongly influence the DTLR value including wind speed, wind direction, wind speed Cartesian components, ambient temperature, and solar irradiance. A tensor, formed by a series of lags associated with these variables, contains the potential informative predictors for DTLR. Moreover, the latent variables, i.e., convection cooling and radiated heat loss rate, may also contain valuable information about the complex relationship between climatic variables and DTLR. Historical DTLR values could also contain useful information.

1) Feature Reduction

It is a principal stage in the DTLR prediction to properly optimize the input features of the prediction engine by eliminating the non-informative and redundant features and identifying the features, which can demonstrate the DTLR pattern more efficiently. To this end, we first employ a feature reduction stage based on minimal-redundancy and maximal-relevance (mRMR) [36]. mRMR is a mutual information (MI)-based method employed in various power system problems to identify the subset of features providing the most information of the observation (target variable) [37]. MI is widely used in the feature selection literature to evaluate the degree of uncertainty that a predictor can alleviate from an observation by measuring the mutual relevancy of predictor and target variable. For two random variables with domains A and B, MI is defined as follows [33]:

$$MI(A,B) = \sum_{a \in A}\sum_{b \in B} P(a,b) \ln \frac{P(a,b)}{P(a)P(b)} \tag{9}$$

where $P(a,b)$ is the joint probability density function; and $P(a)$ and $P(b)$ are the individual probability density functions of a and b random variables, respectively, which are the discretized format of continuous predictor and target variables. Based on (9), the iterative mRMR algorithm is carried out by the following optimization problem [33]:

$$\max_{\substack{a_j^{t-p}\in A-\Omega_{n-1}\\ j=\{1,2,\ldots,9\}\\ p=\{0,1,\ldots,l-1\}}}\left\{MI(a_j^{t-p};DTLR_{t+1})-\frac{1}{n-1}\sum_{\substack{a_i^{t-q}\in\Omega_{n-1}\\ i=\{1,2,\ldots,9\}\\ q=\{0,1,\ldots,l-1\}}}MI(a_j^{t-p};a_i^{t-q})\right\} \tag{10}$$

$$\begin{aligned}A=\{&a_1^{t-l+1}=WS_{t-l+1},a_1^{t-l+2}=WS_{t-l+2}\ldots,a_1^t=WS_t,\\ &a_2^{t-l+1}=WD_{t-l+1},a_2^{t-l+2}=WD_{t-l+2}\ldots,a_2^t=WD_t,\\ &a_3^{t-l+1}=AT_{t-l+1},a_3^{t-l+2}=AT_{t-l+2}\ldots,a_3^t=AT_t,\\ &a_4^{t-l+1}=SI_{t-l+1},a_4^{t-l+2}=SI_{t-l+2}\ldots,a_4^t=SI_t,\\ &a_5^{t-l+1}=WSX_{t-l+1},a_5^{t-l+2}=WSX_{t-l+2}\ldots,a_5^t=WSX_t,\\ &a_6^{t-l+1}=WSX_{t-l+1},a_6^{t-l+2}=WSX_{t-l+2}\ldots,a_6^t=WSX_t,\\ &a_7^{t-l+1}=qc_{t-l+1},a_7^{t-l+2}=qc_{t-l+2}\ldots,a_7^t=qc_t,\\ &a_8^{t-l+1}=qr_{t-l+1},a_8^{t-l+2}=qr_{t-l+2}\ldots,a_8^t=qr_t,\\ &a_9^{t-l+1}=DTLR_{t-l+1},a_9^{t-l+2}=DTLR_{t-l+2}\ldots,a_9^t=DTLR_t\}\end{aligned} \tag{11}$$

where Ω_{n-1} is the subset containing selected features at the $(n-1)^{\text{th}}$ iteration; a_j^{t-p} is the random variable describing the p^{th} lag of the j^{th} feature; and A is a set consisting of the random variables associated with feature candidates. In the first iteration of solving (10), $n=1$, $\Omega_0=\varnothing$. In this paper, the optimization problem in (10) is iteratively solved until the MI of the last component selected from solving (10) with $DTLR_{t+1}$ is negligible.

After conducting mRMR, the feature types that are not among the selected features in Ω_n or constitute a negligible portion of Ω_n are removed from the feature pool. The selected features, along with their corresponding l lags, are then used to build the tensor input of the deep-learning-based feature extraction described in Section II-B-3).

2) LSTM

LSTM is a unit of RNNs in which the temporal dependency among the elements of the TS can be captured. An LSTM block consists of a memory cell, an input gate, an output gate, and a forgetting gate. The memory cell stores the values for arbitrary time intervals. In LSTM, the three gates are neurons with activation functions. Figure 3 represents an LSTM unit, where $\boldsymbol{c}_{t-1}$ and $\boldsymbol{h}_{t-1}$ denote the cell memory state and hidden state at the previous time, respectively. The input vector $\boldsymbol{x}_t$, $\boldsymbol{c}_{t-1}$, and $\boldsymbol{h}_{t-1}$ are used to update the memory state $\boldsymbol{c}_t$, and attain the output $\boldsymbol{h}_t$ corresponding to $\boldsymbol{x}_t$. In Fig. 3, the blocks labelled by σ refer to sigmoid layers. The LSTM has been used in the proposed DTLR prediction as described in the next sections.

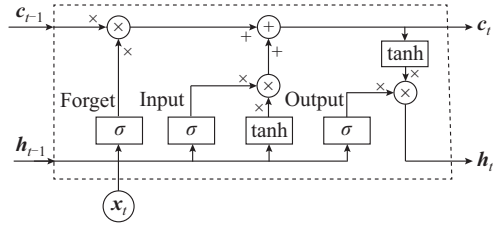

Fig. 3. Schematic of LSTM unit.

3) RNN-based SDAE

AE is a type of neural network mainly employed for unsupervised feature learning in a wide range of applications [30], [31]. An AE is composed of two fragments: the encoder and decoder. The input of the encoder is the original feature tensor, which is mapped to a different space in the output of the encoder, while the decoder uses the mapped features as the input and reconstructs the original feature space. The output of the encoder part of the AE can be used as the input features of the following hidden layers. In the conventional AE, in the case that the number of hidden layer nodes is higher than the input feature, the AE can potentially find the identity map. As a result, the AE can become inefficient. To circumvent this issue, denoising AE (DAE) is proposed [30], [31], where the original input features are reconstructed from the corrupted ones. In this way, the number of nodes in the DAE hidden layers can be larger than that of input features without identity mapping risk. Therefore, this overcomplete DAE may disclose some informative features from the interactions among the input features. In the following text, the DAE is briefly described, while a detailed description of DAE can be found in [31].

At time t, all features that remain after the feature reduction stage, which are described in Section II-B-1), are used to construct the input vector of the DAE $\boldsymbol{x}_t$, a three-dimensional tensor $(1\times l\times m)$, where m is the number of feature variants. To capture the sequential correlation of the TS, LSTM is used as the building blocks of the DAEs in this paper. An RNN-based DAE can be formulated as:

$$\min_{\theta,\theta'} L_{DAE}(\boldsymbol{x},\boldsymbol{z}) \tag{12}$$

$$\tilde{\boldsymbol{x}}\sim q_D(\boldsymbol{x}_t) \tag{13}$$

$$\boldsymbol{y}_t=\boldsymbol{f}_\theta(\tilde{\boldsymbol{x}}_t,\boldsymbol{m}_{t-1}) \tag{14}$$

$$\boldsymbol{z}_t=\boldsymbol{g}_{\theta'}(\boldsymbol{y}_t,\boldsymbol{m}'_{t-1}) \tag{15}$$

where $L_{DAE}(\cdot)$ is the loss function that employs the mean-square error (MSE); $\boldsymbol{x}$ $(n\times l\times m)$ is the tensor that contains all tensors $\boldsymbol{x}_t, t=1,2,\ldots,n$, n is the number of available points in the validation set; $\boldsymbol{z}$ is the tensor that consists of all tensors $\boldsymbol{z}_t$, $t=1,2,\ldots,n$, which are the outputs of a DAE corresponding to input $\boldsymbol{x}$; θ and θ' encapsulate the unknown parameters of encoder and decoder blocks, respectively; $\boldsymbol{y}_t$ and $\boldsymbol{z}_t$ are the outputs of encoder and decoder blocks, respectively; $\boldsymbol{m}_{t-1}$ and $\boldsymbol{m}'_{t-1}$ are the information passed from the calculation of $\boldsymbol{y}_{t-1}$ and $\boldsymbol{z}_{t-1}$, respectively, as the result of recurrent units; and $\boldsymbol{f}_\theta(\cdot)$ and $\boldsymbol{g}_{\theta'}(\cdot)$ are the functions associated with encoder and decoder blocks, respectively.

In (13), some elements of input vector $\boldsymbol{x}_t$ are destroyed in a stochastic process, i.e., $\boldsymbol{q}_D(\cdot)$, in order to form the corrupted input $\tilde{\boldsymbol{x}}$.

Stacking several DAE forms an SDAE, in which more informative features can be extracted. Using the ADAM stochastic optimization [38], the SDAE is trained layer-wise to find the initial parameters of all DAEs. Thereafter, a fine-tuning process is conducted to further tune the SDAE, as elaborated in Section III. The output of the encoder portion of the SDAE is used as the input of Model 2 in the proposed DTLR prediction, as discussed in Section III.

III. Proposed Training Framework for Probabilistic Deep-learning-based Prediction of DTLR

The training of the proposed DTLR is conducted in several steps according to the scheme presented in Fig. 4. First, the data are divided into three datasets: training, validation, and test datasets. The training dataset is used to tune the model parameters, while the validation dataset is used to assess the performance of the model during training. The test dataset is used to evaluate the trained model. The redundant or non-informative features are removed as described in Section II-B-1). Next, the SDAE is formed and trained as discussed in Section II-B-3). The autoregressive prediction model, named Model 1, is then trained using a series of DTLR TS lags, and simultaneously, a many-to-one prediction model, named Model 2, is trained using the feature tensor obtained from the encoder portion of the SDAE. Models 1 and 2 are both based on RNNs and trained using MSE as the cost function, while the weights of the SDAE are frozen. Then, the final model (Model 3) that employs the predictions of Models 1 and 2 as the input to provide a final prediction of the DTLR is trained deterministically using MSE as the cost function.

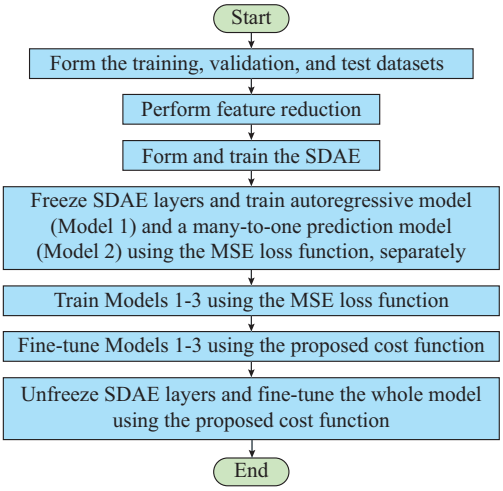

Fig. 4. General training scheme of proposed probabilistic prediction of DTLR.

Afterwards, Models 1-3 are further trained using the proposed cost function (16) - (20). Once Models 1-3 are tuned considering the proposed cost function, SDAE layers are unfrozen to tune the model further using the proposed cost function.

Inspired by QR formulation, the proposed cost function for training DNN for a preferred POE is defined as:

$$C(DTLR_{t+1}, \underline{DTLR}_{t+1}) = \sum_{i=1}^{2} a_i (sign(\epsilon_i) + 1)\epsilon_i \tag{16}$$

$$\epsilon_1 = \underline{DTLR}_{t+1} - DTLR_{t+1} \tag{17}$$

$$\epsilon_2 = DTLR_{t+1} - \underline{DTLR}_{t+1} \tag{18}$$

$$a_1 = POE^* \tag{19}$$

$$a_2 = 1 - POE^* \tag{20}$$

where $sign(\cdot)$ is the sign function; and POE^* is the preferred POE. In (16), the term associated with ϵ_1 penalizes the prediction model when the prediction model results in $\underline{DTLR}_{t+1}$ values above $DTLR_{t+1}$. The more $\underline{DTLR}_{t+1}$ value is higher than $DTLR_{t+1}$, the more the cost function penalizes the model. On the other hand, when $\underline{DTLR}_{t+1} < DTLR_{t+1}$, the term related to ϵ_1 in the cost function becomes zero. However, the term corresponding to ϵ_2 penalizes the prediction model if $\underline{DTLR}_{t+1}$ is deviated from the $DTLR_{t+1}$.

Equation (16) is non-differentiable at $\epsilon_i = 0$ due to $sign(\cdot)$. This can cause an issue in the backpropagation process during the training phase. In this regard, $sign(\cdot)$ is approximated by $\tanh(\cdot)$, and (16) can be rewritten as follows:

$$C(DTLR_{t+1}, \underline{DTLR}_{t+1}) = \sum_{i=1}^{2} a_i (\tanh(\lambda\epsilon_i) + 1)\epsilon_i \tag{21}$$

where λ is a large constant number so that $\tanh(\lambda\epsilon_i) \to 1$ for $\epsilon_i > 0$, $\tanh(\lambda\epsilon_i) \to -1$ for $\epsilon_i < 0$, and $\tanh(\lambda\epsilon_i) = 0$ for $\epsilon_i = 0$. Using $\tanh(\cdot)$ results in a differentiable function that can use the off-the-shelf DNN optimizers for training the model.

The proposed cost function enables to fine-tune the DTLR model so that the DTLR prediction model provides the lower bound of DTLR values with the minimum deviations from the actual DTLR values.

IV. Case Studies and Comparisons

A. Descriptions of Data, Simulation Tools, and Proposed Model Settings

We have performed the analysis based on a 5-year dataset (from January 1, 2010 to January 1, 2015), recorded from the M2 met tower at the National Wind Energy Center (NWEC) located in Denver, USA [39]. Linear interpolation is employed to fill the missing points in the dataset. It is assumed that the measurements correspond to an OHL constructed from 400 mm^2 Drake 26/7 ACSR conductor at the elevation of 1861 m from the sea level.

Based on [2], the specification of this conductor is summarized in Table I, where the STLR is calculated based on the low-speed perpendicular wind (0.6 m/s), high ambient temperature (40 °C), and full solar heating (1000 W/m^2) [40].

TABLE I
Specifications of Conductor

Specification	Value
STLR (A)	685
Outside diameter of conductor (mm)	28.12
Minimum conductor temperature (°C)	25
Maximum conductor temperature (°C)	75
Conductor resistance at the minimum temperature (Ω/km)	0.07284
Conductor resistance at the maximum temperature (Ω/km)	0.08689
Solar absorptivity and emissivity	0.5

The proposed model has been implemented in Python 3.7 on a Windows 10 PC with a 1.6 GHz Intel Core i5 CPU and 8 GB of memory. Based on the grid search, three DAE layers are selected as the suitable architecture for SDAE, re-

spectively. The numbers of hidden layers for Models 1 and 2 are one and three, respectively. Model 3 is a single-layer model which provides the final prediction by using the outputs from Models 1 and 2. The training is conducted using the process described in Section III.

B. Analysis of Feature Candidates

To evaluate the dependency of DTLR on different feature candidates, a 5-year meteorological dataset is used to generate historical $DTLR_t$, qc_t, and qr_t TS, using the procedure described in Section II-A.

Figure 5 represents the dispersion of DTLR with respect to some of the feature types in the upper triangle.

The Kendall τ rank coefficient is widely employed as a non-parametric statistical test in hypothesis testing to identify the statistical dependency between two random variables [41].

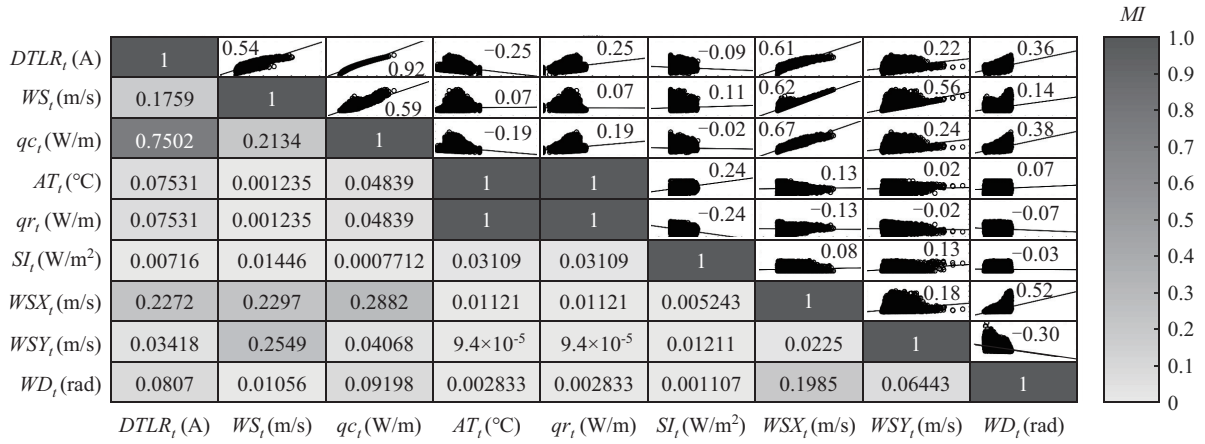

Fig. 5. Dispersion of DTLR with respect to different feature types and Kendall τ rank coefficient and mutual information of different features and DTLR.

The detailed explanations of the Kendall τ rank coefficient can be found in [38]. The presented quantities in the subplots, forming the upper triangle, represent the Kendall τ rank coefficient of different features vis-à-vis each other and $DTLR_t$. Based on the calculated Kendall τ rank coefficients, it can be concluded that there is a strong relation between $DTLR_t$ and latent variables as well as meteorological variables.

Furthermore, Fig. 5 shows that qc_t is the most influential factor in $DTLR_t$ compared with all other features. This strong relation can also be found from the *MI* values presented in the diagonal subplots and lower triangle of Fig. 5. The *MI* values testify to the high dependency of $DTLR_t$ on qc_t. As shown in Fig. 5, among the meteorological variables, *WS* demonstrates the strongest correlation with *DTLR*, while $MI(WS_t, DTLR_t)$ is significantly lower than $MI(qc_t, DTLR_t)$. On the other hand, qr_t and AT_t are fully correlated based on Kendall τ rank coefficient as well as *MI* values. Therefore, considering only one of them could be adequate in the DTLR prediction. Moreover, from $MI(DTLR_t, SI_t)$, it can be recognized that SI_t can provide the minimum information in the DTLR prediction. To this end, this feature species are eliminated in the feature reduction stage described in Section II-B. It is worth noting that the data are discretized with respect to the median values of different features to calculate the *MI* values. This type of discretization is widely used for calculating *MI* values among continuous variables [42].

C. Description of Benchmark Models

Three benchmark models are utilized in this paper: persistence [12], QR [15], and QRF [16] prediction models. The persistence model (PM) is a conventional prediction model that is widely used for short-term prediction of meteorological-related variables including DTLR [12]. The simplicity of PM can facilitate the comparison of the proposed model with other prediction models. QR has recently been proposed for the non-parametric probabilistic prediction of DTLR, which is the focus of this paper. Therefore, comparing the efficacy of the proposed model with QR provides some insights about the superiority of the proposed model with respect to other non-parametric probabilistic prediction models of DTLR. As the last benchmark model, the state-of-the-art QRF-based probabilistic prediction model of DTLR is employed [5], [16]. Optimal hyperparameters of the QRF are found using the Bayesian optimization. To carry out a fair comparison, the proposed model and benchmark models utilize similar input variables, as elaborated in Section II-A, but a notable exception is PM, for which the mean and variance of the latest DTLR values form a Gaussian distribution representing the uncertainty of upcoming samples.

D. Evaluation Metrics

Three evaluation metrics are used to appraise the performance of the different DTLR models. The POE of the prediction model is the most imperative evaluation criterion for a secure probabilistic DTLR prediction and is defined as:

$$POE = \frac{1}{N}\sum_{n=1}^{N}\delta_n \times 100\% \tag{22}$$

$$\delta_n = \begin{cases} 1 & \underline{DTLR}_{t+1} \le DTLR_{t+1} \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

where N is the number of points in the training, validation, or test datasets. Any deviation of the POE from the preferred POE^* can lead to unprecedented issues. As a measure of

sharpness of the predicted POEs, normalized mean absolute error (NMAE) is used, which is defined as:

$$NMAE = \frac{\sum_{t=1}^{N} \left| DTLR_{t+1} - \underline{DTLR}_{t+1} \right|}{N \cdot Range} \times 100\% \tag{24}$$

where *Range* is the range of DTLR values. In a perfect prediction of DTLR, $POE = POE^*$ while $NMAE \to 0$. Root-mean-squared-error (RMSE), which is a valuable measure and signifies large deviations of DTLR prediction from its actual value, is also employed as another evaluation metric, and can be calculated as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (DTLR_{t+1} - \underline{DTLR}_{t+1})^2} \tag{25}$$

E. Numerical Studies

As an instance, the performance of the prediction model on a two-year dataset (from January 1, 2010 to January 1, 2012) is used for numerical comparisons. 85% of the data are used for training and validation, while the remainder are employed for testing. In this section, the effectiveness of considering latent predictors, i.e., convection cooling and radiated heat loss rate, is first investigated. Thereafter, the proposed DTLR is compared with the benchmark models for different *CL* preferences using various evaluation metrics.

1) Impact of Latent Predictors

Section IV-B shows that in comparison to directly observed meteorological variables, the radiated heat loss rate has a more pretentious relation with DTLR. To empirically investigate the efficacy of considering a series of lags associated with the mentioned latent variables as predictors, a case study is conducted using the proposed DTLR prediction with and without the latent predictor. Table II summarizes the case study conducted for $POE^* = 99\%$. The table shows that the experiential results are in line with the theoretically expected outcome, and that considering latent variables can reduce NMAE and RMSE while satisfying the POE criterion. Therefore, using the proposed prediction method unlocks the OHL ampacity further, while the POE is satisfied. As a result, more power can be transferred through the transmission systems. Therefore, it can be concluded that the latent variables can provide further information about the DTLR pattern, and the prediction can be performed more precisely.

TABLE II
PERFORMANCE COMPARISON OF PROPOSED MODEL WITH OR WITHOUT LATENT VARIABLES AS INPUT

With or without latent variables	*POE* (%)	*NMAE* (%)	*RMSE*
Without	98.95	19.21	300.20
With	99.01	18.51	291.30

2) Numerical Comparisons of Different Prediction Models

The performance evaluations of different prediction models for $POE^* = 90\%$, 95%, and 99% are presented in Table III. The table shows that the PM is incompetent in providing secure prediction of DTLR for any of the POE^* values. Other benchmark models along with the proposed model can fulfill the POE^* requirements. From the perspectives of NMAE and RMSE, for $POE^* = 90\%$, the performance of QR, QRF, and the proposed model are closely comparable. For $POE^* = 95\%$, based on NMAE and RMSE, the proposed model slightly outperforms the benchmark models while POE^* is satisfied, i.e., $POE \geq POE^*$. For $POE^* = 99\%$, the case study demonstrates that the proposed model can facilitate the secure employment of DTLR, while NMAE and RMSE are reduced compared with the benchmark models. It is noted from Table III that for $POE^* = 99\%$, the proposed model results in reductions in NMAE and RMSE by at least 6.33% ((18.51 – 19.76)/19.76 × 100%) and 5.78% ((309.16 – 291.30)/309.16 × 100%), respectively.

TABLE III
PERFORMANCE EVALUATION OF DIFFERENT PREDICTION MODELS

POE^* (%)	Prediction model	*POE* (%)	*NMAE* (%)	*RMSE*
90	PM	83.42	13.80	251.10
	QR	92.20	11.90	202.55
	QRF	92.64	12.00	203.38
	Proposed model	91.90	11.91	203.60
95	PM	89.57	16.40	290.41
	QR	96.60	15.01	242.21
	QRF	95.56	14.73	241.27
	Proposed model	95.14	14.50	236.61
99	PM	95.75	22.12	372.52
	QR	99.80	19.96	311.80
	QRF	99.31	19.76	309.16
	Proposed model	99.01	18.51	291.30

To further investigate the performance of the proposed model compared with the benchmark models, the histograms of various prediction models of DTLR for $POE^* = 99\%$ is presented in Fig. 6. It is worth noting that the histograms are driven by considering only the predicted values that are equal to or less than the mixmum actual ampacity. Thus, the performance of the different prediction models in terms of unlocking the OHL capacity can be compared.

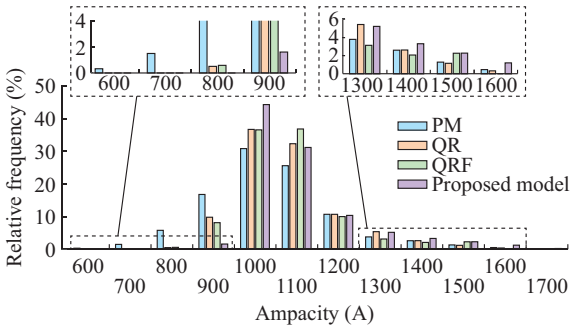

Fig. 6. Distribution of predicted DTLR for $POE^* = 99\%$ with predicted values equal or less than the maximum actual ampacity.

As can be observed from Fig. 6, for several points in the test dataset, the benchmark models result in ampacity predictions, which are equal to or less than 900 A, while the proposed model leads to the limited number of prediction values of DTLR which are smaller than 900 A. On the other hand, for more than about 10% of the times, the proposed model results in the DTLR prediction greater than 1200 A,

while the POE^* criterion is met. The frequency of predicting DTLR higher than 1200 A by using the benchmark models is reduced.

Figure 7 depicts DTLR prediction using proposed model and QRF for $POE^*=99\%$, which is the best benchmark model for $POE^*=99\%$. The figure covers the prediction for three successive days, i. e., 72 one-hour samples, to provide the view of high volatility in the DTLR values and the efficacy of the proposed model to track this volatile pattern. As can be observed, the DTLR can vary substantially within an hour, which mainly stems from the chaotic behaviour of its influential factors, e.g., wind speed, wind direction, etc. It is discerning from Fig. 7 that the proposed model can result in prediction values of DTLR that are 100 A higher than those obtained from QRF. Besides, comparing the STLR in Table I with the predictions values of DTLR shown in Fig. 7, it can be concluded that with minimal risk, a significant amount of OHL ampacity headroom can be unlocked.

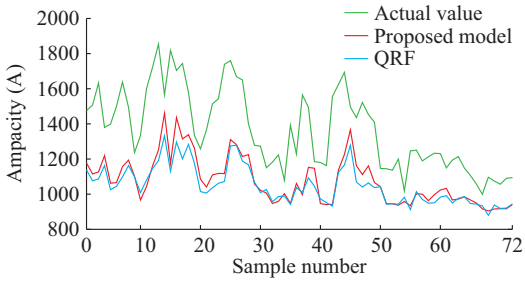

Fig. 7. DTLR prediction using proposed model and QRF for $POE^*=99\%$.

V. Discussion

As the case studies elucidate, the proposed prediction framework of DTLR results in a more accurate and reliable prediction compared with various well-established benchmark models. In Sections IV-E-1) and 2), the detailed analyses demonstrate that the superior performance of the proposed model results from both considerations of latent variables and the deep learning architecture. The proposed model can be used as a decision-supporting tool for system operators for unlocking the additional ampacity of OHL considering a pre-defined risk exposure acceptance, i.e., POE^*.

The hour-ahead secure prediction of DTLR can facilitate the accommodation of renewable generation, and alleviate the issues of transmission congestion. Consequently, the wind power curtailment is reduced, while the demand for transmission expansion can be postponed or even eliminated. Such a highly accurate hour-ahead prediction can be implemented in the energy management system for real-time security constraint economic dispatching [43].

In this study, the efficacy of the proposed model is validated for a range of POE^* values. In practice, the risk-based assessment of the transmission lines and the cost implications of the OHL annealing can be used to identify the optimal POE^*[44], as a consequence of operating the OHL higher than the conductor temperature limit.

The feature selection and extraction play crucial roles in the prediction framework. In this paper, mRMR and SDAE, which are well-known feature selection and extraction methods, are adapted, respectively. Despite the extensive research on the development of the advanced feature selection and extraction techniques in various power-system-related prediction applications [45], the development of these machine learning building blocks in the literature of DTLR is scarce. To this end, extensive research on feature selection and extraction in DTLR forecasting can be considered as an important future research direction.

The proposed prediction framework of DTLR is validated for a direct prediction application of DTLR, in which only meteorological variables and the DTLR formulations are employed. However, considering the generality of the framework, other features such as OHL sag/tension as well as NWP can be incorporated in the framework to improve the performance of DTLR forecasting further. Furthermore, using the NWP can also facilitate extending the prediction horizon, which is beneficial for day-ahead unit commitment and economic dispatching.

VI. Conclusion

This paper has proposed a deep-learning-based probabilistic prediction model of DTLR for hour-ahead power system operation problems. The latent variables, obtained in the process of calculating the DTLR values, are considered as new predictors of the proposed model, while SDAE is employed for feature learning and extraction. A training strategy is devised to train the DNN, and a cost function is put forward to train the prediction model in a probabilistic manner. The proposed prediction framework considers no hypothesis about the uncertainty of the DTLR. Simulation results confirm the efficacy of the proposed model and its superiority compared with the benchmark models.

References

[1] D. M. Greenwood, J. P. Gentle, K. S. Myers *et al*., "A comparison of real-time thermal rating systems in the U. S. and the U. K.," *IEEE Transactions on Power Delivery*, vol. 29, no. 4, pp. 1849-1858, Aug. 2014.

[2] *IEEE Standard for Calculating the Current-Temperature Relationship of Bare Overhead Conductors*, IEEE Std 738-2012, pp. 1-72, 2013.

[3] R. Stephen, "The thermal behaviour of overhead conductors, Section 3: mathematical model for evaluation of conductor temperature in the unsteady state," *Electra*, vol. 174, pp. 58-69, Oct. 1997.

[4] T. O. Seppa and A. Salehian, "Guide for selection of weather parameters for bare overhead conductor ratings," *CIGRE Working Group* B2.12, Paris, France, Aug. 2006.

[5] F. Teng, R. Dupin, A. Michiorri *et al*., "Understanding the benefits of dynamic line rating under multiple sources of uncertainty," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3306-3314, May 2018.

[6] S. Madadi, B. Mohammadi-Ivatloo, and S. Tohidi, "Integrated transmission expansion and PMU planning considering dynamic thermal rating in uncertain environment," *IET Generation, Transmission & Distribution*, vol. 14, no. 10, pp. 1973-1984, Feb. 2020.

[7] G. Murphy, "Implementation of real-time thermal ratings," SP Energy Networks, Merseyside, UK, Final Report SPT1001, Oct. 2013.

[8] B. J. Walker. (2019, Jun.). Dynamic line rating. [Online]. Available: https://www.energy.gov/sites/prod/files/2019/08/f66/Congressional_DLR_Report_June2019_final_508_0.pdf

[9] B. P. Bhattarai, J. P. Gentle, P. Hill *et al*., "Improvement of transmission line ampacity utilization by weather-based dynamic line rating," *IEEE Transactions on Power Delivery*, vol. 33, no. 4, pp. 1853-1863, Aug. 2018.

[10] J. Teh, C.-M. Lai, N. A. Muhamad *et al*., "Prospects of using the dynamic thermal rating system for reliable electrical networks: a review," *IEEE Access*, vol. 6, no. 1, pp. 26765-26778, Apr. 2018.

[11] S. Karimi, P. Musilek, and A. M. Knight, "Dynamic thermal rating of transmission lines: a review," *Renewable & Sustainable Energy Reviews*, vol. 91, no. 1, pp. 600-615, Aug. 2018.
[12] F. Fan, K. Bell, and D. Infield, "Probabilistic real-time thermal rating forecasting for overhead lines by conditionally heteroscedastic auto-regressive models," *IEEE Transactions on Power Delivery*, vol. 32, no. 4, pp. 1881-1890, Aug. 2017.
[13] T. Ringelband, P. Schäfer, and A. Moser, "Probabilistic ampacity forecasting for overhead lines using weather forecast ensembles," *Electrical Engineering*, vol. 95, no. 2, pp. 99-107, Jun. 2013.
[14] X. Sun, P. B. Luh, K.W. Cheung *et al*., "Probabilistic forecasting of dynamic line rating for over-head transmission lines," in *Proceedings of 2015 IEEE PES General Meeting*, Denver, USA, Jul. 2015, pp. 1-5.
[15] Z. Wei, M. Wang, X. Han *et al*.,"Probabilistic forecasting for the ampacity of overhead transmission lines using quantile regression method," in *Proceedings of 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, Xi'an, China, Oct. 2016, pp. 1632-1635.
[16] J. L. Aznarte and N. Siebert, "Dynamic line rating using numerical weather predictions and machine learning: a case study," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 335-343, Feb. 2017.
[17] S. Madadi, B. Mohammadi-Ivatloo, and S. Tohidi, "Probabilistic real-time dynamic line rating forecasting based on dynamic stochastic general equilibrium with stochastic volatility," *IEEE Transactions on Power Delivery.* doi: 10.1109/TPWRD.2020.3012205
[18] S. Madadi, M. Nazari-Heris, B. Mohammadi-Ivatloo *et al*., "Application of big data analysis to operation of smart power systems," *Big Data in Engineering Applications*, vol. 44, pp. 347-362, May 2018.
[19] A. Salehian, "ARIMA time series modeling for forecasting thermal rating of transmission lines," in *Proceedings of 2003 IEEE PES Transmission and Distribution Conference and Exposition*, Dallas, USA, Sept. 2003, pp. 875-879.
[20] S. Madadi, B. Mohammadi-Ivatloo, and S. Tohidi, "Dynamic line rating forecasting based on integrated factorized Ornstein-Uhlenbeck processes," *IEEE Transactions on Power Delivery*, vol. 35, no. 2, pp. 851-860, Apr. 2020.
[21] J. Jiang, Y. Liang, C. Chen *et al*., "On dispatching line ampacities of power grids using weather-based conductor temperature forecasts," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 406-415, Jan. 2018.
[22] Y. Yang, R. Harley, D. Divan *et al*., "Adaptive echo state network to maximize overhead power line dynamic thermal rating," in *Proceedings of 2009 IEEE Energy Conversion Congress and Exposition*, San Jose, USA, Sept. 2009, pp. 2247-2254.
[23] J. Jiang, J. Wan, X. Zheng *et al.*, "A novel weather information-based optimization algorithm for thermal sensor placement in smart grid," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 911-922, Mar. 2018.
[24] M. Matus, D. Saez, M. Favley *et al.*, "Identification of critical spans for monitoring systems in dynamic thermal rating," *IEEE Transactions on Power Delivery*, vol. 27, no. 2, pp. 1002-1009, Apr. 2012.
[25] J. Teh and I. Cotton, "Critical span identification model for dynamic thermal rating system placement," *IET Generation, Transmission & Distribution*, vol. 9, no. 16, pp. 2644-2652, Dec. 2015.
[26] S. Karimi, A. M. Knight, P. Musilek *et al*., "A probabilistic estimation for dynamic thermal rating of transmission lines," in *Proceedings of IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*, Florence, Italy, Jun. 2016, pp. 1-6.
[27] M. Khodayar and J. Wang, "Spatio-temporal graph deep neural network for short-term wind speed forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 670-681, Apr. 2019.
[28] Z. Zhang, L. Ye, H. Qin *et al*.,"Wind speed prediction method using shared weight long short-term memory network and Gaussian process regression," *Applied Energy*, vol. 247, no. 1, pp. 270-284, Aug. 2019.
[29] M. Khodayar, J. Wang, and Z. Wang, "Energy disaggregation via deep temporal dictionary learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1696-1709, Jul. 2020.
[30] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 2770-2779, Jul. 2017.
[31] P. Vincent, H. Larocheele, Y. Bengio *et al*.,"Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, Helsinki, Finland, Jul. 2008, pp. 1096-1103.
[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735-1780, Nov. 1997.
[33] K. Cho, B. van Merrienboer, C. Gulcehre *et al.* (2014, Sept.). Learning phrase representations using RNN encoder-decoder for statistical machine translation. [Online]. Available: https://arxiv.org/abs/1406.1078
[34] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D*, vol. 110, pp. 43-50, Dec. 1997.
[35] A. Pavlinić, V. Komen, and M. Uzelac, "Application of direct collocation method in short-term line ampacity calculation," *Electric Power Systems Research*, vol. 155, pp. 216-224, Feb. 2018.
[36] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226-1238, Aug. 2005.
[37] S. M. Mazhari, N. Safari, C. Y. Chung *et al*., "A hybrid fault cluster and Thévenin equivalent based framework for rotor angle stability prediction," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5594-5603, Sept. 2018.
[38] D. P. Kingma and J. L. Ba. (2017, Jan.). ADAM: a method for stochastic optimization. [Online]. Available: https://arxiv.org/abs/1412.6980
[39] NREL. (2020, Jul.). National Wind Energy Center. [Online]. Available: https://www.nrel.gov/nwtc/index.html
[40] D. Douglass, W. Chisholm, G. Davidson *et al.*, "Real-time overhead transmission-line monitoring for dynamic rating," *IEEE Transactions on Power Delivery*, vol. 31, no. 3, pp. 921-927, Dec. 2016.
[41] M.G. Kendall, *Rank Correltion Methods*, London: Griffin, 1970.
[42] N. Amjady and F. Keynia, "Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 306-318, Feb. 2009.
[43] K. W. Cheung and H. X. Wu, "Implementation of dynamic thermal ratings in the operational environment," Washington DC, USA, FERC Technical Conference Report, Jun. 2014.
[44] J. Teh and C. Lai, "Risk-based management of transmission lines enhanced with the dynamic thermal rating system," *IEEE Access*, vol. 7, pp. 76562-76572, Jun. 2019.
[45] J. Wang, T. Niu, H. Lu *et al*., "A novel framework of reservoir computing for deterministic and probabilistic wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 337-349, Jan. 2020.

N. Safari received the B.Sc. (Hon.) degree from the K. N. Toosi University of Technology, Tehran, Iran, in 2012, the M. Sc. degree from the Tehran Polytechnic University, Tehran, Iran, in 2014, and the Ph.D. degree in electrical engineering from the University of Saskatchewan, Saskatoon, Canada, in 2018. He is currently a Grid Operations Support Engineer-in-Training with SaskPower, Regina, Canada. His research interests include applications of machine learning algorithms in integration of large-scale renewable energy generation into power systems.

S. M. Mazhari received the M. Sc. (Hon.) degree from the University of Tehran, Tehran, Iran, in 2012, and the Ph.D. (Hon.) degree from the Amirkabir University of Technology, Tehran, Iran, in 2016, all in electrical engineering. He is currently a Postdoctoral Research Fellow at the University of Saskatchewan, Saskatoon, Canada. His research interests include power system transient stability, renewable energy interconnection, and artificial intelligence applications in power system automation.

C. Y. Chung received the B.Eng. (with First Class Honors) and Ph.D. degrees in electrical engineering from The Hong Kong Polytechnic University, Hong Kong, China, in 1995 and 1999, respectively. He is currently a Professor, the NSERC/SaskPower (Senior) Industrial Research Chair in Smart Grid Technologies, and the SaskPower Chair in Power Systems Engineering in the Department of Electrical and Computer Engineering, the University of Saskatchewan, Saskatoon, Canada. His research interests include smart grid technologies, renewable energy, power system stability/control, planning and operation, computational intelligence applications, power markets, and electric vehicle charging.

S. B. Ko received his Ph.D. in electrical and computer engineering at the University of Rhode Island, Kingston, USA, in 2002. He is currently a Professor in the Department of Electrical and Computer Engineering, the University of Saskatchewan, Saskatoon, Canada. He worked as a Member of technical staff for Korea Telecom Research and Development Group, Korea, from 1993 to 1998. His research interests include computer arithmetic, computer architecture, deep learning processor architecture, efficient hardware implementation of compute-intensive applications, cryptography, and biomedical engineering.