Decomposition-based Stacked Bagging Boosting Ensemble for Dynamic Line Rating Forecasting

Amirhossein Ahmadi, *Student Member, IEEE,* Saman Taheri,
Reza Ghorbani, Vahid Vahidinasab, *Senior Member, IEEE*, and Behnam Mohammadi-ivatloo, *Senior Member, IEEE*.

***Abstract*—Effective exploitation of overhead transmission lines needs reliable and precise dynamic line rating forecasting. High-accuracy dynamic line rating forecasting, in particular, is an important short-term method for coping with grid congestion, enhancing grid stability, and accommodating high renewable energy penetration. Due to the non-stationarity and stochasticity of the meteorological variables, a single model is often not sufficient to accurately predict the dynamic line rating. Herein, a new stacked bagging boosting ensemble is developed based on multivariate empirical mode decomposition to overcome single models' restrictions and increase the dynamic line rating forecasting performance. The developed ensemble is utilized on the data gathered from a 400 kV aluminum conductor steel-reinforced overhead power line with a length of 32.85 Km between Ghadamgah and Binalood wind farms, located in the northeast of Iran. The simulation results substantiate that the proposed ensemble can capture meteorological variables' non-linear characteristics, yielding more accurate yet robust to noisy data forecasts than single forecasting models.**

***Index Terms*—Dynamic line rating, forecasting, ensemble, empirical mode decomposition, stacked bagging boosting.**

I. Introduction

ECONOMIC and environmental benefits have fueled the growth and integration of renewable energy sources into power systems [1], [2]. However, growing variability and uncertainty imposed by high penetration of renewable energies have adverse impacts on the reliability of modern power systems [3]. Enhancing flexibility on the supply, grid, and demand sides has been identified as a realistic approach for dealing with the fluctuations and uncertainty associated with renewable energies, balancing supply and demand, and accommodating the high penetration of renewable energy sources. [4]. Grid-side flexibility can be achieved by allowing better use of transmission infrastructure via line rating [5].

Dynamic line rating (DLR) is a platform that allows electric transmission lines to raise their existing carrying power dynamically. The overhead line's capacity is dictated by its ability to dissipate heat generated by the Joule effect into the atmosphere. The maximum current that will satisfy the construction, protection and safety requirements of a particular line on which the conductor is used is known as line ampacity [7]. This, in turn, is influenced by external factors, including air temperature, wind intensity, and solar radiation. Although dynamic, ampacity is currently computed statically, which is called Static Line Rating (SLR), based on a worst-case scenario like high air temperature ($40°C$), full solar radiation ($1000W/m^2$) and low wind speeds ($0.6m/s$). These conservative ambient conditions can result in a reduction in line capability during less demanding weather conditions.

In the DLR context, ampacity is treated as a dynamic parameter, providing an approximation of the values at which the lines are operated at each time step. This is particularly noticeable on overhead transmission lines, where DLR may deliver significant uprating. Conductor-based, weather-based, and sag/vibration/tension-based DLR approaches are the three most common types. Amongst them, the weather-based approach is simple as it gathers data solely from weather sensors mounted on towers or local weather stations. The main issues here are those of uncertainty and variability associated with meteorological data. A low-cost and easy-to-implement solution to tackle these challenges is the accurate forecasting of line ratings. By ensuring accurate DLR forecasting, the grid operation can be more reliable, resulting in higher penetration of renewable energy. As such, developing accurate DLR forecasts is seen as a critical phase in incorporating DLR into power system management and reaping the intended benefits. Among all DLR forecasting approaches, machine learning (ML) algorithms have gained a lot of attention due to their ability to learn transmission line behavior with a wide degree of variability [9].

A. Literature Survey

Over the last few years, multiple forecasting methods have been proposed, as thoroughly reviewed in [10]–[12], to enrich the DLR technology. Ahmadi et al. [8] investigated decision tree ensemble-based DLR forecasts to prepare for overhead transmission line capability changes across multiple forecasting horizons. For 400 kV overhead transmission lines, while XGBoost outperformed the other decision trees, the survey resulted in a near 30% capacity boost. Madadi et al. Talpur et al. [13] looked at the cost savings of using DLR for a 130 kV sub-transmission system. The findings of such a study affirm the DLR's ability to increase capacity and

A. Ahmadi is with the Department of Electrical and Software Engineering, University of Calgary, Calgary, Canada, e-mail: amirhossein.ahmadi@ucalgary.ca.

S. Taheri is with the Department of Mechanical Engineering, Indiana University-Purdue University Indianapolis, Indianapolis, USA e-mail: Staheri@iu.com.

R. Ghorbani is with the Department of Mechanical Engineering, University of Hawai'i at Manoa, Honolulu, USA, e-mail: Rezag@Hawaii.edu

V. Vahidinasab is with the School of Science and Technology, Nottingham Trent University, Nottingham, UK, e-mail: vahid.vahidinasab@ntu.ac.uk

B. Mohammadi-ivatloo is with the Department of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran, and also with the School of Energy Systems, LUT University, Lappeenranta, Finland, e-mail: mohammadi@tabrizu.ac.ir

promote large-scale wind energy integration. Talpur et al. [14] addressed the uncertainty and variability associated with DLR forecasting using integrated factorized Ornstein-Uhlenbeck processes. Bhattarai et al. [15] looked at increasing transmitting capability using weather-based DLR and several case studies. As opposed to the current SLR, the proposed method's scalability and usefulness were shown. Sugihara et al. [16] developed a real-time DLR model based on the line current variations for a highly renewable-penetrated network. Dupin et al. [17] introduced a method to use the real-time latent current-carrying capacity of overhead conductors–considering the operator's risk aversion in high-risk situations.

To cope with the uncertainty associated with DLR, a new time-series modeling approach based on autoregressive integrated moving average is introduced in [18]. Abboud et al. [19] used fluid mechanics to improve the accuracy of wind speed measurements along conductor paths. As such, steady-state ampacity is predicted using wind simulations coupled with meteorological records. Bosisio et al. [20] built a step-by-step method for evaluating all of the stochastic processes of atmospheric variables usable for DLR forecasting. Albizu et al. [21] suggested an adaptive SLR for static line rating situations that could reach a maximum temperature limit. Minguez [22] demonstrated a promising DLR method based on ML techniques for reducing wind farm outages caused by excess power generation. Saatloo et al. [23] proposed hierarchical extreme learning machine-enabled short-term DLR forecasting based on meteorological parameters. Albizu et al. [24] compared the results of many DLR forecasting methods, focusing on appropriate forecast ratios and safety indicators. Because of their lower maximum temperature exceedance and adequate overestimate ranges, the SLR and selective ambient-adjusted ratings were found to be the most stable.

Based on the association between temperature, conductor resistance, and voltage drop-through lines, Dawson and Knight [25] investigated the applicability of extending DLR technology to non-thermally confined lines. Dupin et al. [26] created a probabilistic method for day-ahead real-time ampacity forecasting based on artificial intelligence and using meteorological station measurements and numerical weather forecasts as input(s). Centered on stochastic general equilibrium with stochastic volatility, Madadi et al. [27] constructed a probabilistic real-time DLR forecasting method. Following their work, Cheng et al. [28] presented DLR forecasting by implementing a real-time simulation model via Tabu Search. Dong [29] introduced a data-driven long-term DLR forecasting for power transformers. Kirilenko et al. [30] proposed quantile regression and super-quantile regression approaches to present very short-term risk-averse stochastic DLR of overhead conductors.

B. Research gap and paper contribution

The research history shows that few studies have discussed DLR forecasting using ML algorithms. ML is a form of artificial intelligence that can recognize trends in data and forecast them. To maximize learning accuracy, precision, and robustness while minimizing bias/variance, two or more ML algorithms can be combined by the ensembling technique. Ensembles can be generally divided into homogeneous ensembles founded on the same base learners and heterogeneous ensembles founded on various base learners. Homogeneous ensembles can be further classified into parallel and sequential ensembles. In parallel ensembles, such as bagging and random forest (RF), homogeneous base learners are trained independently and in parallel (the distribution of the training set is changed stochastically) and then are combined by simple averaging (or majority voting in classification case). Sequential ensembles, such as Adaboost and gradient boosting, on the other hand, train homogeneous base learners sequentially in an adaptive way (the training set's distribution adjusts adaptively depending on the results of the previous model) and then combine them through weighted averaging. In heterogeneous ensembles, such as stacking, learners of various types work in parallel, and their predictions are then fed as inputs to a second layer, which creates a new collection of predictions. Different learning algorithms, hyper-parameter environments, function subsets, and training sets can all be used to provide variety to the dependent learners. Parallel ensembles strive to produce strong models that have less variance than their components, while sequential and stacking ensembles strive to produce strong models that are less biased than their components (although variance can also be reduced). On a noiseless dataset, sequential ensembles are more efficient than bagging, whereas on noisy data, parallel ensembles are more efficient than sequential ensembles [31]. Ensemble systems are effectively employed in various applications, including confidence estimation, feature selection, error correction, incremental learning, data fusion, class-imbalanced data, and concept drift.

It can be concluded that DLR forecasting must be as precise as possible to avoid reaching the line's maximum operating temperature during times of high current loading, which will result in transmission line lifetime loss and breakdown and safety risks. Developed models must be generalizable to be properly functional over the line without requiring another training procedure. Furthermore, robustness against inaccurate measurements and cyber intrusions is another issue that should be provided by such forecasting models. Due to the nonstationarity and stochasticity of the meteorological variables, a single model is often not sufficient to accurately predict the dynamic line rating. Accordingly, the major contributions of this study can be summarized as follows:

- Bagging-boosting ensemble is proposed to provide forecasts with less bias and variance and good robustness on noisy data.
- A decomposition technique is utilized to transform the non-stationary historic dataset into a series of relatively simple and stationary subsets. This enables stacking a bagging-boosting ensemble as a novel DLR forecasting model to overcome single models' restrictions and increase the forecasting performance.
- The proposed model is applied to a 400 kV aluminum conductor steel-reinforced overhead power line located in the northeast of Iran.
- The effects of forecasting horizon on generalization and

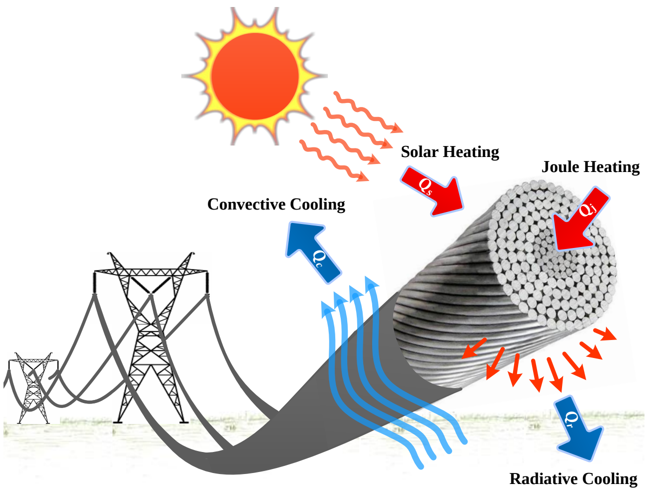

Fig. 1. Thermal balance of overhead transmission lines.

robustness of DLR forecasting models are analyzed. A systematic analysis is performed considering different forecasting horizons, and multiple performance indicators.
- Comprehensive comparisons are made on widely used machine learning models for different datasets.

Finally, Section V outlines the key takeaways from the simulation results and concludes the paper.

II. Preliminary

A. Overhead Line Thermal Modelling

As per IEEE standard 738, the line ampacity of overhead lines can be formulated as follows [8]:

$$I_m = \sqrt{\frac{Q_r + Q_c - Q_s}{R(T_c)}} \tag{1}$$

Equation (1) is based on the thermal heat balance of overhead lines, as depicted in Fig. 1, where $Q_j = Q_r + Q_c - Q_s = R(T_c)I^2$, I and $R(T_c)$ stand for Joule heating, line current, and conductor resistance, respectively. Moreover, Q_s, Q_r and Q_c denote solar heating, radiative cooling and convective cooling, respectively.

B. Multivariate Empirical Mode Decomposition

Empirical mode decomposition (EMD) is a data-driven technique for non-linear and non-stationary signal time-frequency analysis [32]. Via an approach known as sifting, it attempts to decompose a given signal ($x(t)$) into a linear combination of a finite set of localized intrinsic mode functions (IMFs) ($c(t)$) plus a non-zero mean low-degree polynomial residual ($r(t)$). The EMD process can be formulated as:

$$x(t) = \sum_{i=1}^{d} c_i(t) + r_d(t) \tag{2}$$

where $c_i(t)$ is amplitude/frequency modulated zero mean oscillatory components defined as:

$$c_i(t) = a_i(t)\cos(\varphi_i(t)) \tag{3}$$

where φ_i and $a_i(t)$ are instantaneous phase and amplitude, respectively. Different from wavelet decomposition, the EMD is free from predefined basis functions for signal decomposition. For all advantages the EMD provides, it considers only univariate signals and computes the local mean by averaging the lower and upper envelopes. The local minima and maxima cannot be directly defined when dealing with multivariate signals [33]. Multivariate EMD (MEMD) is used to solve these problems, where a vector-valued version of regular EMD is used to decompose a p-variate signal as:

$$\mathbf{x}(t) = \sum_{i=1}^{d} \mathbf{c}_i(t) + \mathbf{r}_d(t) \tag{4}$$

where $\mathbf{r}_d(t)$ denotes the p-variate residual.

III. Proposed DLR Forecasting Model

The idea behind ensembling is achieving an improved composite learner by aggregating a set of diverse learners to counteract deficiencies associated with constituents and provide more effective predictions. This section outlines representative ensemble models, including bagging, boosting, and stacking.

A. Bagging

Bagging (or bootstrap aggregation) is a parallel ensemble formed through majority voting for classification and averaging for regression by aggregating independent homogeneous base learners trained with bootstrap data sets. A bootstrap is generated independently based on sampling with replacement from the original training collection, which creates the diversity needed for the assembly. Dagging is a new Bagging method developed for large training sets that avoid bootstrapping by sampling a set of equal-sized segments. Another variant of Bagging is Wagging, which uses the entire original training set and assigns a stochastic weight to each record. In addition, RFs employ a subset array of features (feature bagging) as well as bootstrapping to minimize correlation between base learners, making them more sophisticated forms of Bagging. Another form of RF is rotation forest, which generates new training sets using principal component analysis.

Bagging combines weak learners with similar characteristics. The weak learners are all of the same model types, for instance, they may all be decision trees. With bagging, each of the weak learners is trained independently, allowing us to easily parallelize model training over multiple cores or computers. The predictions from each of the weak learners are then aggregated and some type of averaging is performed in order to obtain predictions on unknown data. After dividing our data into a training set and a test set, we produce multiple bootstrap sample sets and train a weak learner on each of these sample sets. To make predictions using the bagged model, an observation is passed through each of the weak learners, and then the predictions are averaged, as is common for regression models. The distinction between a bagging model and an RF is that, while an RF functions similarly to a bagging method, with a random forest, various subsets of observations are used to train each individual decision tree. In addition, we use a distinct subset of data to train each individual decision tree.

B. Boosting

Boosting is a sequential ensemble formed by aggregating independent homogeneous weak learners, each of which is sequentially trained with a new training set containing higher weights for the errors of the previous model. The training set for each subsequent learner is updated such that it minimizes mistakes of previously trained learners. Adaptive boosting or AdaBoost is a new version of boosting in which weak learners are sequentially trained with a new training set containing higher weights for errors of the previous model. Based on learners' training errors, weak learners are then aggregated using weighted majority voting for classification and weighted averaging for regression.

Gradient boosting or GBoost is a generalized version of AdaBoost that utilizes residuals, gradients of the loss function, to identify the shortcomings of weak learners and then adopts the gradient descent algorithm to correct them. Stochastic gradient boosting or SGBoost is another deviant of GBoost in which a random subsample of the training set is drawn without replacement to train weak learners at each iteration. Extreme GBoost or XGBoost is a scalable and generalized version of GBoost with higher model performance and computational speed that uses pre-sorted and histogram-based algorithms for finding the best split. It is founded on a new regularized objective function containing training loss measuring how the model fits training data and the regularization term measuring learners' complexity to avoid overfitting. Optimizing training loss leads to predictive models fitting well in the training set while optimizing regularization leads to simple yet stable models. Light GBoost or LGBoost is a leaf-wise version of GBoost providing higher accuracy and computational speed. In contrast to pre-sorted and histogram-based algorithms, LGBoost uses a gradient-based one-side sampling technique to approximate the information gain with a much smaller data size and faster speed. It also uses an exclusive feature bundling technique to bundle mutually exclusive features and reduce the number of features. Categorical boosting or CatBoost is a depth-wise version of GBoost with symmetric or oblivious trees that automatically handles categorical data and provides high accuracy and computational speed. Note that the main difference between XGBoost, LGBoost, and CatBoost is in learning the decision trees and finding the best split points.

Boosting is an assemblage technique that progressively or iteratively trains models. Therefore, weak students lack independence. The current training of the model is dependent on prior models. Adaboost is an example of a boosting algorithm that operates as follows. Each successive model is trained by giving greater weight to the observations that were most challenging to predict in the previous step. Specifically, the weights of previously inaccurate observations are enhanced. To create the ultimate strong student from the weak learners, a weighted portion of the weak learners is selected at the end. The weights for each of the weak learners are proportional to their performance. The weights for each of the weak learners are proportional to their performance. In other words, the weaker the learner's performance, the greater its weight. The gradient boosting algorithm is another instance of a boosting algorithm. The final model in gradient boosting is also a weighted sum of the weak learners. However, gradient descent is utilized to identify how to improve at each successive step. Gradient boosting is an extension of boosting in which optimization can be based on any arbitrary loss function that is differentiable. Frequently, gradient boosting is applied to decision trees. Typically, gradient-enhanced trees outperform RFs.

C. Stacking

Stacked generalization or stacking is a heterogeneous ensemble combining multiple diverse base learners (also called first-level learners) by training a meta-learner (also called second-level learner). The base learners are trained on the original training set, and then their output predictions are fed as a new training set to train a meta-learner making final predictions. The base learners are often from different learning algorithms; however, it is also possible to use diverse homogeneous base learners using different hyper-parameters, features, or datasets. Multi-level stacking is the extended version of stacking that comprises stacking with multiple layers.

Unlike bagging, which is a less variance ensemble with efficiency on noisy data, boosting appears to be less biased with efficiency on noise-free data. Thus, it is expected that using boosting as a base learner for bagging, as shown in Fig. 2, can yield forecasts with less bias and variance. Despite this, bagging boosting still struggles to recognize patterns with high accuracy and robustness because of the non-stationarity of meteorological variables. Alternatively, the non-stationary dataset can be decomposed into relatively simple and stationary chunks. As a result, stacked bagging boosting models can be applied where each model focuses on the frequency band components of a single subset, thus improving their overall performance. Fig. 3 illustrates the proposed decomposition-based stacked bagging boosting ensemble for DLR forecasting, where MEMD is adopted to decompose the dataset due to the multivariate nature of DLR.

Stacking is a mechanism for assembling diverse weak learners. For instance, you may mix neural networks with decision trees, GLM, etc. It is also important to remember those bagged and boosted models frequently serve as poor learners in stacked ensembles. Consequently, stacked models might be challenging to interpret meaningfully. However, they are typically high-performing models. Since interpretability is not our primary concern in this investigation, we seek the most accurate model possible. Ensemble approaches are therefore highly appealing.

The bagging and stacking models used in this paper are inspired by [34]. The boosting model utilized in this paper is constructed following the work of Prettenhofer et al. [38].

D. Performance Metrics

The performance of forecasting models should be assessed using various statistical measures. The quality of a forecasting model is defined by several elements including bias, reliability/calibration, uncertainty, sharpness/refinement, accuracy, association, resolution, and discrimination. Nonetheless, the majority of research in this area has been

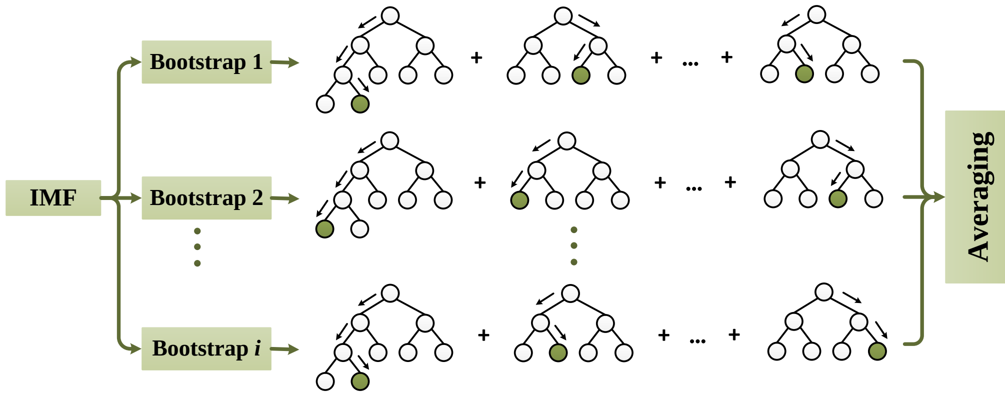

Fig. 2. Bagging boosting ensemble.

validated using accuracy measures such as mean absolute error (MAE) as defined by equation (9):

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \tag{5}$$

root mean square error (RMSE) by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{6}$$

mean absolute percentage error (MAPE) by:

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|.100\%, \tag{7}$$

and coefficient of determination (R^2) by:

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}. \tag{8}$$

where y_i is the observation, N is the number of samples, $\bar{y}$ is the mean of observations, and $\hat{y}_i$ is the prediction. The MAE measures the mean of the absolute errors or bias while the RMSE measures the standard deviation of the errors or variance [35]. The lower the MAE and RMSE values, the more accurate the forecasts. A principal limitation of the MAE and RMSE metrics is that they do not consider the size of actual values. The MAPE measures the mean of the absolute values of percentage errors taking into account the magnitude of actual values. The MAPE values less than 10% mean a highly accurate forecast, 11% to 20% imply a good forecast, 21% to 50% denote a reasonable forecast and more than 50% indicate an inaccurate forecast [35]. The R^2, with values normally ranging from 0 to 1, measures how well the predictions approximate the observations. In other words, it offers valuable knowledge about the forecasting model's ability to fit, with a value closer to 1 suggesting better prediction precision.

IV. Simulation Results

The presented forecasting model was tested on a 400 kV Aluminium conductor steel-reinforced cable (ACSR) overhead power line with a length of 32.85 kilometers connecting the Ghadamgah and Binalood wind farms in northeast Iran (shown in Fig. 9). The configuration files were created and executed using TensorFlow2 on top of Python. Table I lists the features of the selected ACSR line. To record weather data over the line, three additional weather stations have been installed between these wind turbines. The proposed algorithm is compared with XGBoost which outperformed the other decision tree ensembles including baggings and boosting [8].

A. Data Analysis

Data from over 15 months from 02/09/2017 11:30 to 30/11/2018 16:50 was collected, which includes 73,500 meteorological values and corresponding ampacities with a resolution of 10-min. The dataset includes wind speed, wind direction, solar radiation, and ambient temperature as features (input), and the line ampacity is considered as the label (output). The data is split to train and test sets (10% as test and the rest for training) and then fed to the algorithms. The correlations between the line ampacity and the corresponding features are illustrated in Fig. 5. Wind speed has a positive correlation of 0.33 on the rating, i.e., the more wind speed, the more convective heat loss, and therefore line ampacity. There is, however, a negative correlation between air temperature and solar radiation, -0.27 and -0.16, meaning that higher temperature or radiation causes lower radiative,

TABLE I
Data Characteristics of the Selected Overhead Power Lines

Property	Value	Unit
Aluminum mass per unit length	1.401	kh/m
Thermo-resistivity coefficient	19.4×10^6	-
Steel mass per unit length	0.522	kh/m
Elevation above sea level	950	m
Maximum temperature	75	$^\circ C$
Conductor diameter	31.5	mm
Conductor section	585.5	mm^2
Nominal voltage	400	kV
Absorptivity	0.8	-
Emissivity	0.85	-

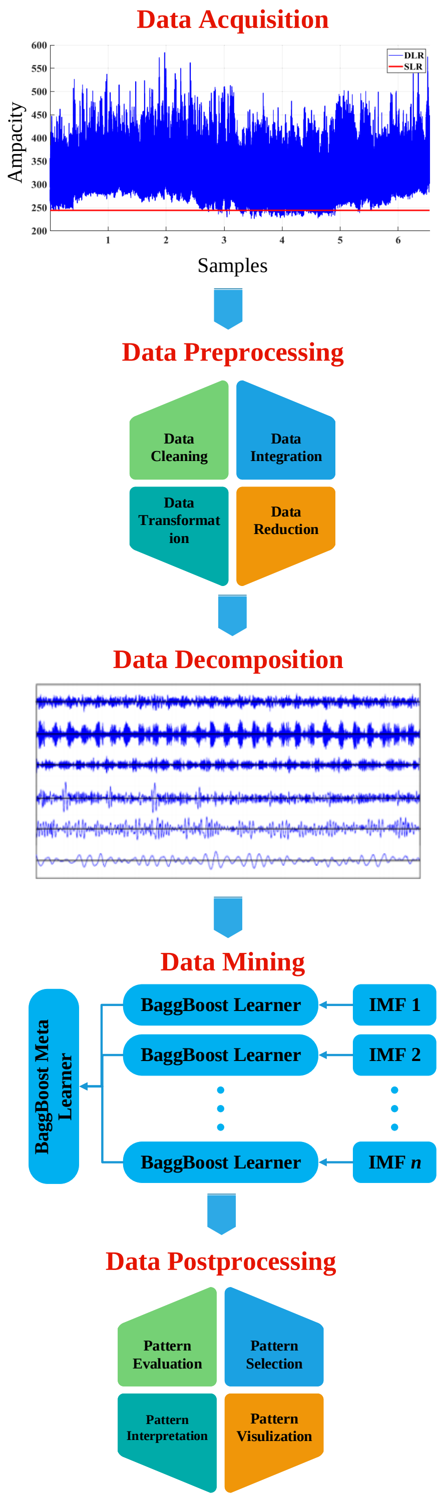

Fig. 3. Framework of the proposed decomposition-based stacked bagging boosting forecasting model.

and convective cooling, which causes more solar heating and less line ampacity. Convective cooling caused by wind is equivalent to orthogonal directions since wind direction shows a sinusoidal correlation. Weibull distributions of wind speed and temperature for all stations are presented in Fig. 6. It is evident from Station 3 that the wind and temperature profiles are diverse. Observations at Station 1 have a higher level of average speeds than those at other sites, while those at Station 3 have higher wind speeds. A broader temperature profile can be seen at Ghadamgah and Station 1, while the most common temperatures are seen at Station 3.

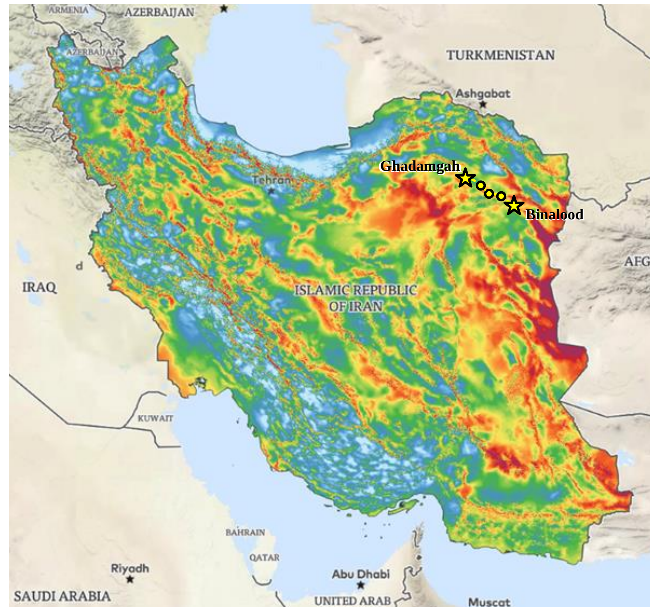

Fig. 4. Selected overhead power line with three weather stations.

B. Mode Decomposition

Fig. 7 represents the original and decomposed normalized ampacities for Ghadamgah using the MEMD technique. The original ampacity profile contains numerous variations and spikes that can be attributed to the randomness and uncertainty of meteorological conditions. It can also be seen that the first IMFs contain high frequencies while the last IMFs contain low frequencies.

C. Case 1

The first case looks at DLR forecasts for short- and mid-term forecasting horizons of 6 hours, 12 hours, 24 hours, and 48 hours. The related performance indexes for different forecasting horizons are summarized in Table II. In terms of MAE, RMSE, and MAPE, as well as R^2, the proposed multivariate analytical mode decomposition-based stacked bagging boosting models generated more reliable forecasts than XGBoost. Fig. 8 displays DLR predictions for Ghadamgah records with a 10-min sampling interval from 6 hours to 48 hours forward. The proposed ensemble successfully tackled the non-stationary problem and properly detected the DLR variation utilizing the mode decomposition technique. It is worth mentioning that XGBoost hyper-parameters, though not well-tuned to highlight the potential of the decomposition technique, are the same as those utilized in the proposed model. The DLR estimates for the entire forecasting duration are better than the SLR, with an expected capability increase of about 22.7%. The proposed model's precision was maintained with a minor variation as the forecasting horizon was extended from 6 hours to 48 hours.

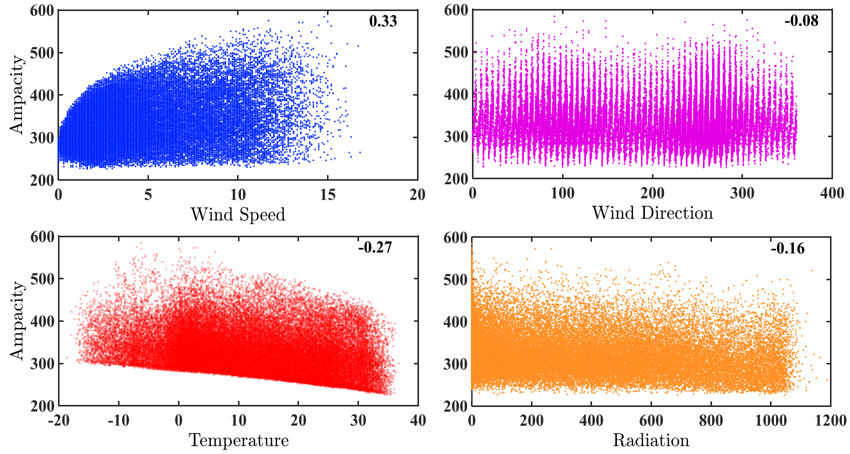

Fig. 5. Depiction of Correlation between ampacity and meteorological variables.

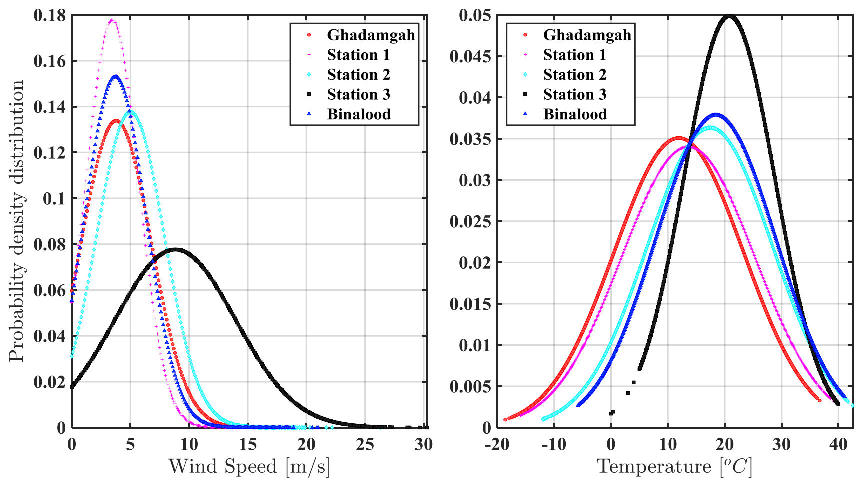

Fig. 6. Probability density distribution of most correlated features.

D. Case 2

Generalizability refers to the degree to which the model can perform well on new, previously unseen data without being neither underfit nor overfit. Overfitting refers to a model that performs too well on the training data but performs too poorly on new, previously unseen data. Underfitting is a concept used to describe simple models which are unable to understand the related patterns in the training datasets. Bias is a measure of underfitting, while variance is a measure of overfitting quantified by MAE and RMSE metrics, respectively. In order to provide accurate DLR forecasting, however, ampacity must be calculated at many points along the lines. It was thus hypothesized that stations one, two, three, and Binalood would serve as valid predictors for assessing the generalizability of the representative models.

TABLE II
PERFORMANCE INDICES OF THE PREDICTION MODELS IN CASE 1

Algorithm	Horizon	RMSE	MAPE	MAE	R^2
XGBoost	6h	7.8	10.6	12.3	0.87
	12h	7.6	11.6	16.1	0.81
	24h	8.7	13.6	15.4	0.79
	48h	9.7	15.1	15.3	0.75
Proposed	6h	6.1	6.5	9.2	0.91
	12h	7.3	7.6	9.4	0.89
	24h	9.4	8.3	10.4	0.85
	48h	10.1	9.9	11.6	0.81

Table III summarizes results for 6-hour-ahead DLR forecasting, from which their generalizability is substantiated for stations different from the model-trained station, i.e., Ghadamgah. The proposed model is compared to various forecasting models including support vector machines (SVM), multi-layer perceptron (MLP) and Long Short-term Memory (LSTM) recurrent neural networks. The findings show that the proposed stacked bagging boosting model outperforms other existing models in terms of MSE, MAPE, RMSE, and R^2. Finally, as seen in Fig. 9, the DLR can be forecasted as the minimum of the predicted ampacity along the line in five stations. The proposed approach, as seen in the diagram, could forecast the DLR curve with greater precision, robustness, and a wider range of stations. This reduces the need for operators to use additional measurement devices or communication networks in addition to calibration.

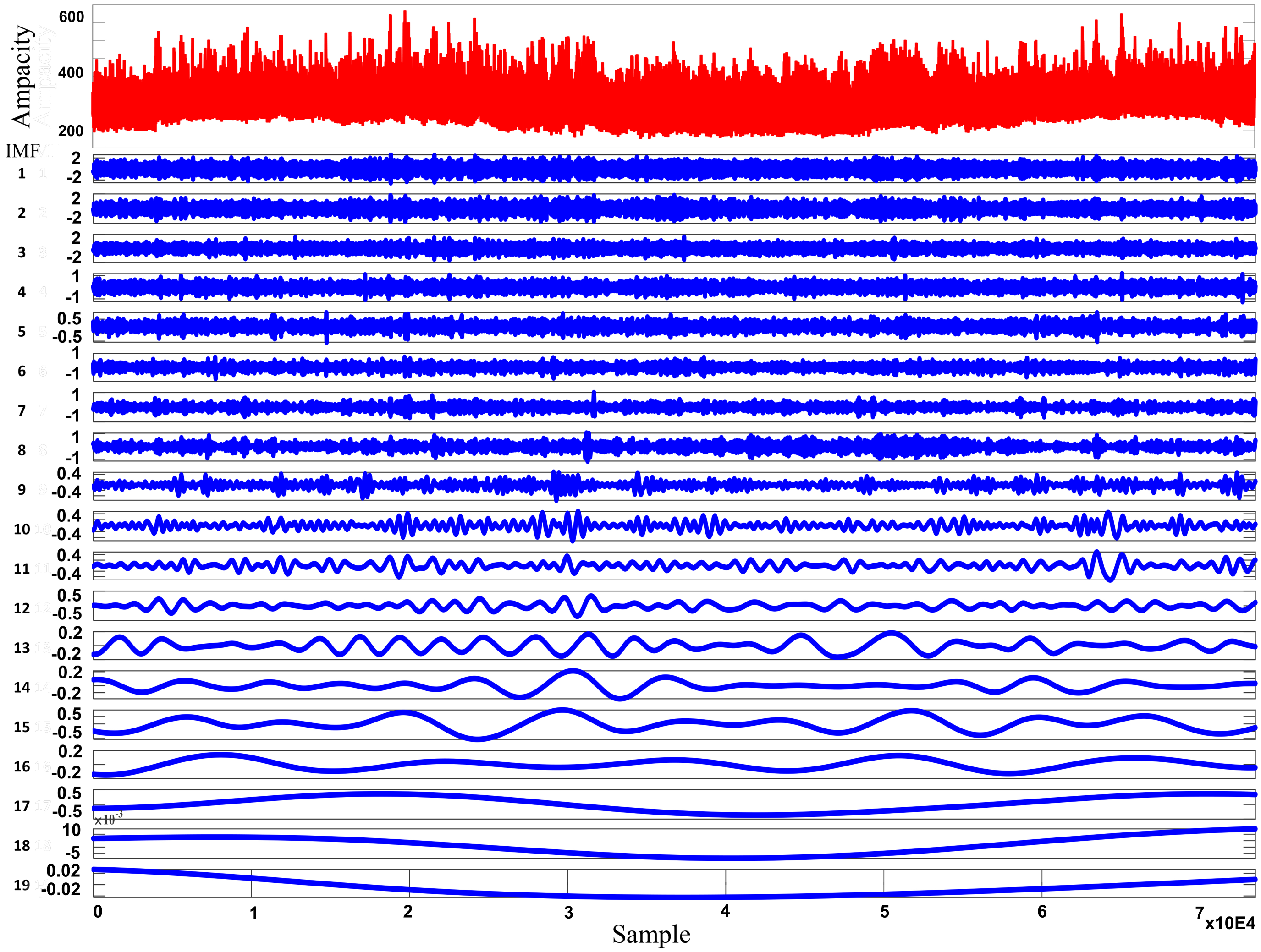

Fig. 7. Decomposition results of Ghadamgah ampacity.

E. Computational efficiency

The computational efficiency of models is also compared in Table IV. The computational time is considered per epoch (average time for 35 epochs) for each algorithm

While the proposed method is not the most efficient network in terms of computing time, it follows SVM and XGBosst, both of which are known as very fast ML networks. Furthermore, it should be mentioned that the amount of data used to train the ML model might have a significant impact on computing efficiency. The literature advises examining the scalability of ML algorithms to provide a fair comparison in terms of computational time. Scalable ML refers to learning that can analyze large amounts of data without exhausting available resources like memory. There are a number of reasons why machine learning requires scalability: 1) The process of training a model may take a long time. 2) The working memory of a training device may not be large enough to hold the model. Even if we acquire a huge computer with sufficient memory and processing capability, the price will be more than if we utilize a collection of smaller machines. In other words, vertical scaling is excessively expensive. As a result, machine learning's scalability presents several issues, including the management of huge datasets, model training, and model evaluation. As mentioned in [36], the development of algorithms executable on a distributed infrastructure is a crucial tactic for addressing these obstacles. Determining the highest pace at which data may be sent is the key to accelerating this sort of processing. Small batch sizes allow us to overcome this obstacle. Mini-batch is a technique for computing gradients with a minute sample size. Mini-batching facilitates stability modeling by updating gradients on fragments rather than in a single time step. Another benefit of mini-batches over regular SGD is the speed improvement GPUs give while doing matrix computations [37]. We divided the dataset into several chunk sizes, namely 10, 15, 20, 25, and 30, in order to increase prediction accuracy and processing performance. After several trials with 20 mini-batches containing a total of 3,675 samples, the optimal result was attained. By decreasing the size of the mini-batch, the accuracy of the model is drastically reduced. This large decline in precision may be attributable to the decreased variance of the smaller mini-batches, which leads to less stable gradients. In addition, increasing the chunk size

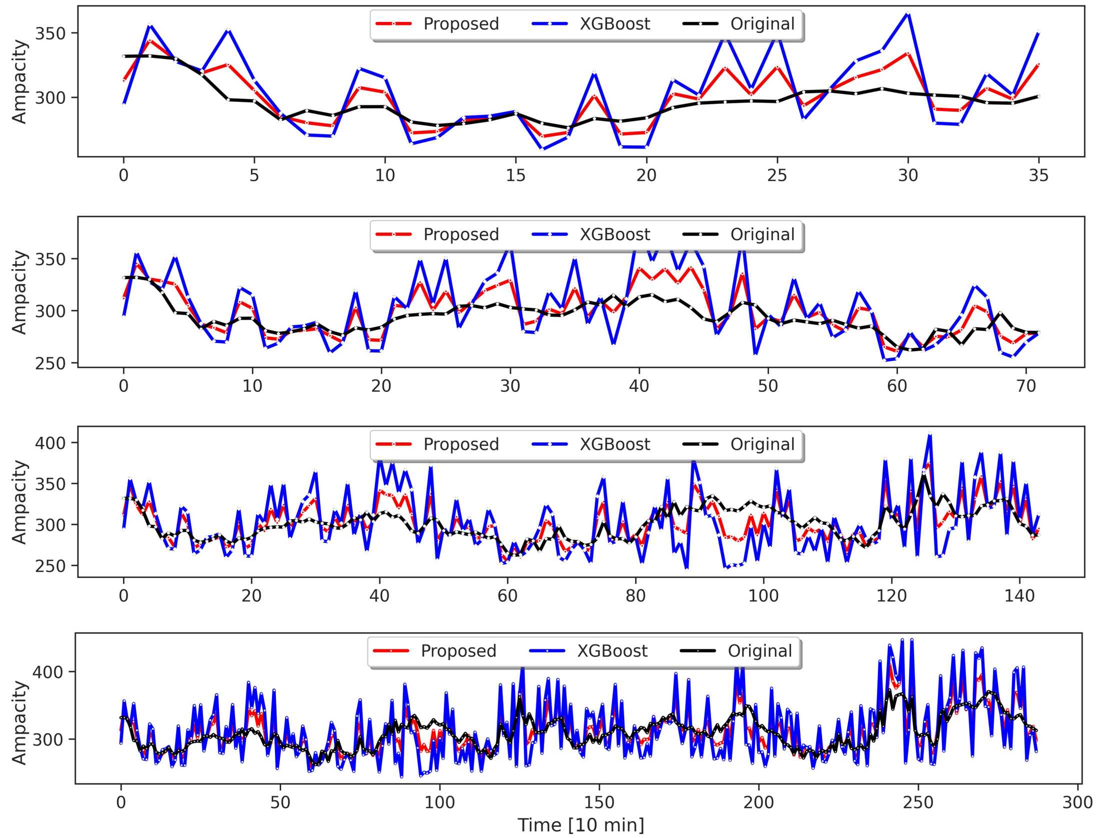

Fig. 8. Forecasting accuracy for different horizons (case 1).

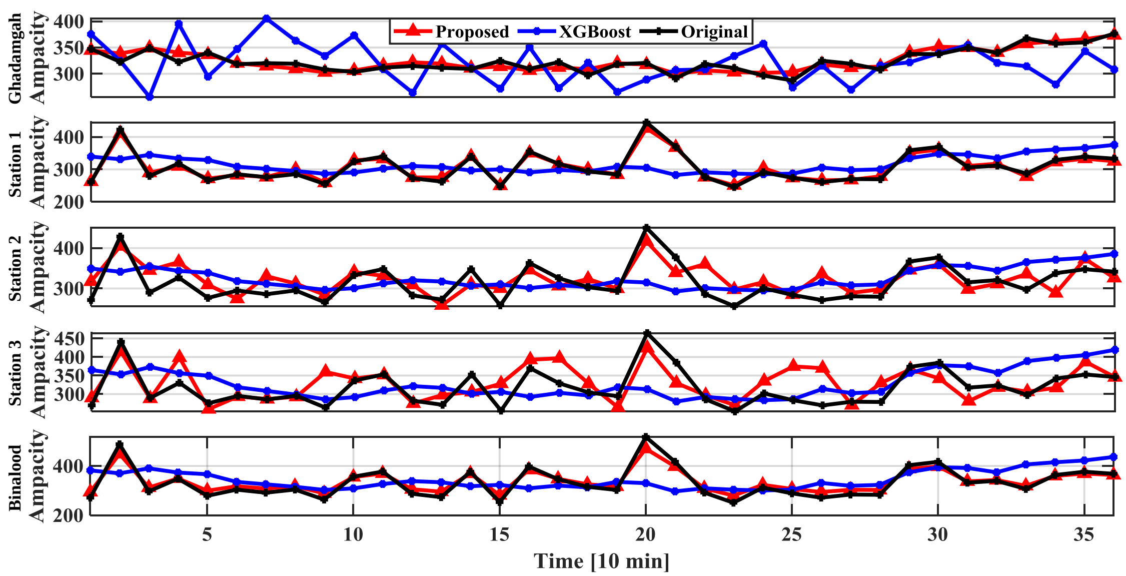

Fig. 9. Forecasting accuracy for different stations (case 2).

to 25 or 30 does not improve the performance of the model. Comparing epoch 25 utilizing best practice, the average calculation time was reduced by just 26 minutes. In addition, we used GeForce GTX 3050, 3060, and 3070 graphics cards that operate in parallel chunks to expedite the calculation of gradients. The GTX 3070 outperformed the GTX 3050 and 3060 in our tests by 1.28 and 1.86 times, respectively, with a low standard deviation. In conclusion, by distributing the suggested architecture over many GPUs, we were able to make it both commercially viable and highly efficient. In the case of the suggested EMD-based algorithms, all data storage is needed is the model parameters, which generally range

TABLE III
PERFORMANCE INDICES OF THE PREDICTION MODELS IN CASE 2

Location	Index	SVM	MLP	LSTM	XGBoost	Proposed
		Model				
Ghadamgah	RMSE	14.8	13.7	13.2	7.8	**6.1**
	MAPE	9.6	8.4	7.4	10.6	**6.5**
	MAE	18.1	28.2	21.3	12.3	**9.2**
	R^2	0.75	0.83	0.85	0.87	**0.91**
Station 1	RMSE	26.3	14.7	19.0	13.4	**12.7**
	MAPE	13.9	20.1	12.6	13.6	**13.2**
	MAE	12.1	13.6	14.4	15.1	**11.7**
	R^2	0.78	0.82	0.81	0.82	**0.89**
Station 2	RMSE	26.4	24.6	21.7	18.9	**13.3**
	MAPE	14.1	18.1	18.2	16.7	**16.2**
	MAE	22.6	19.4	17.5	17.6	**11.5**
	R^2	0.78	0.72	0.76	0.88	**0.91**
Station 3	RMSE	17.3	13.7	13.9	9.93	**8.3**
	MAPE	18.7	16.1	14.7	19.2	**14.3**
	MAE	23.2	17.6	23.5	19.9	**16.9**
	R^2	0.79	0.82	0.79	0.84	**0.85**
Binalood	RMSE	16.3	14.0	22.3	17.6	**12.3**
	MAPE	13.9	20.1	10.3	13.3	**11.8**
	MAE	26.1	24.2	22.8	21.8	**18.9**
	R^2	0.74	0.72	0.78	0.80	**0.86**

TABLE IV
COMPUTATIONAL EFFICIENCY OF THE PREDICTION MODELS

Method	Computational time (s)
SVM	23
MLP	138
LSTM	165
XGboost	103
Proposed	126

from a few tens to several hundred thousand double-precision floating point values, or a few gigabytes overall.

V. CONCLUSION

This paper introduces a novel model for DLR forecasting of overhead transmission lines over short and medium-term periods. To tackle the difficulty associated with the nonstationarity nature of meteorological variables, multivariate empirical mode decomposition is utilized. A new stacked bagging boosting ensemble is developed based on intrinsic mode functions to produce forecasts with less bias and variance along with improved generalizability and robustness. For the entire forecasting duration and various forecasting horizons, the predicted ampacity was greater than that given by the SLR. The findings revealed that the 400 kV overhead transmission line between the Ghadamgah and Binalood wind farms has a minimum capacity improvement of 22%, which is necessary to alleviate congestion and boost grid stability. Furthermore, our findings imply that the suggested method for DLR prediction can yield a reasonable prediction accuracy, which would inevitably facilitate the demand for intricate physical models as well as the usage of additional sensors and communication networks over the transmission lines.

REFERENCES

[1] A. Kirilenko, Y. Gong and C. Y. Chung, *A Framework for Power System Operational Planning under Uncertainty Using Coherent Risk Measures*, IEEE Transactions on Power Systems, Feb.2021.

[2] M. Numan, D. Feng, F. Abbas, S.n Habib, S. Hao, *Coordinated operation of reconfigurable networks with dynamic line rating for optimal utilization of renewable generation*, International Journal of Electrical Power & Energy Systems, vol. 125, p.106473, Feb.2021.

[3] Taheri S, Hosseini P, Razban A, *Model predictive control of heating, ventilation, and air conditioning (HVAC) systems: A state-of-the-art review*, Journal of Building Engineering, vol. 60, p. 105067, Nov, 2022.

[4] Zhang Y, Liao Y, *Kalman filter based method for tracking dynamic transmission line parameters*, SN Applied Sciences, Jan.2021.

[5] M. Manana, A.F. Zobaa, A. Vaccaro, A. Arroyo, R. Martinez, P. Castro, A. Laso, S. Bustamante, *Increase of capacity in electric arc-furnace steel mill factories by means of a demand-side management strategy and ampacity techniques*, International Journal of Electrical Power & Energy Systems, vol. 124, Jan.2020.

[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, 2020.

[7] F. Teng, R. Dupin, A. Michiorri, G. Kariniotakis, Y. Chen, and G. Strbac, *Understanding the benefits of dynamic line rating under multiple sources of uncertainty*, IEEE Transactions on Power Systems, vol. 33, no. 3, pp. 3306–3314, 2017.

[8] A. Ahmadi, M. Nabipour, B. Mohammadi-Ivatloo, V. Vahidinasab, *Ensemble Learning-based Dynamic Line Rating Forecasting under Cyberattacks*, IEEE Transactions on Power Delivery, 2021.

[9] A. Moradzadeh, S. Zakeri, M. Shoaran, B. Mohammadi-Ivatloo, F. Mohamamdi, *Short-term load forecasting of microgrid via hybrid support vector regression and long short-term memory algorithms*, Sustainability, vol. 12, no. 17, p.7076, 2020.

[10] S. Karimi, P. Musilek, and A. M. Knight, *Dynamic thermal rating of transmission lines: A review*, Renewable and Sustainable Energy Reviews, vol. 91, pp. 600–612, 2018.

[11] D. A. Douglass, J. Gentle, H. M. Nguyen, W. Chisholm, C. Xu, T. Goodwin, H. Chen et al, *A Review of Dynamic Thermal Line Rating Methods With Forecasting*, IEEE Transactions on Power Delivery, vol. 34, no. 6, pp. 2100–2109, 2019.

[12] N. M. Zainuddin, M. S. Abd Rahman, M. Z. A. Ab Kadir, NH Nik Ali, Z. Ali, M. Osman, M. Mansor et al, *Review of Thermal Stress and Condition Monitoring Technologies for Overhead Transmission Lines: Issues and Challenges*, IEEE Access, 2020.

[13] 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, 2019.

[14] S. Talpur, C. J. Wallnerstrom, P. Hilber, and S. N. Saqib, *Implementation of dynamic line rating technique in a 130 kv regional network*, In 17th IEEE Int. Multi Topic Conference 2014, pp. 477-482. IEEE, 2014.

[15] B. P. Bhattarai, J. P. Gentle, T. McJunkin, P. J. Hill, K. S. Myers, A. W. Abboud, R. Renwick, and D. Hengst, *Improvement of transmission line ampacity utilization by weather-based dynamic line rating*, IEEE Transactions on Power Delivery, vol. 33, no. 4, pp. 1853–1863, 2018.

[16] H. Sugihara, T. Funaki, and N. Yamaguchi, *Evaluation method for real-time dynamic line ratings based on line current variation model for representing forecast error of intermittent renewable generation*, Energies, vol. 10, no. 4, p. 503, 2017.

[17] R. Dupin, A. Michiorri, and G. Kariniotakis, *Optimal dynamic line rating forecasts selection based on ampacity probabilistic forecasting and network operators' risk aversion*, IEEE Transactions on Power Systems, vol. 34, no. 4, pp. 2836–2845, 2019.

[18] J. Zhan, C. Y. Chung, and Elemer Demeter, *Time series modeling for dynamic thermal rating of overhead lines*, IEEE Transactions on Power Systems, vol. 32, no. 3, pp. 2172–2182, 2016.

[19] A. W. Abboud, J. P. Gentle, T. R. McJunkin, and J. P. Lehmer, *Using Computational Fluid Dynamics of Wind Simulations Coupled With Weather Data to Calculate Dynamic Line Ratings*, IEEE Transactions on Power Delivery, vol. 35, no. 2, pp. 745–753, 2019.

[20] A. Bosisio, A. Berizzi, D. D. Le, F. Bassi, and G. Giannuzzi, *Improving DTR assessment by means of PCA applied to wind data*, Electric Power Systems Research, vol. 172, pp. 193–200, 2019.

[21] I. Albizu, E. Fernandez, R. Alberdi, M. T. Bedialauneta, and A. J. Mazon, *Adaptive Static Line Rating for Systems With HTLS Conductors*, IEEE Tran. on Power Delivery, vol. 33, no. 6, pp. 2849–2855, 2018.

[22] 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, 2016.

[23] A.M. Saatloo, A. Moradzadeh, H. Moayyed, M. Mohammadpourfard, and B. Mohammadi-Ivatloo, *Hierarchical Extreme Learning Machine Enabled Dynamic Line Rating Forecasting*, IEEE Systems Journal, 2021.

[24] I. Albizu, E. Fernandez, A. Javier Mazon, and R. Alberdi, *Forecast ratio and security analysis of rating forecasting methods in an overhead line* IET Generation, Transmission & Distribution, vol. 11, no. 6, pp. 1598–1604, 2017.

[25] L. Dawson, and A. M. Knight, *Applicability of dynamic thermal line rating for long lines*, IEEE Transactions on Power Delivery, vol. 33, no. 2, pp. 719–727, 2017.

[26] R. Dupin, G. Kariniotakis, and A. Michiorri, *Overhead lines Dynamic Line rating based on probabilistic day-ahead forecasting and risk assessment*, International Journal of Electrical Power & Energy Systems, vol. 110, pp. 565–578, 2019.

[27] 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, 2020.

[28] Y. Cheng, P. Liu, Z. Zhang, Y. Dai, *Real-time Dynamic Line Rating of Transmission Lines Using Live Simulation Model and Tabu Search*, IEEE Transactions on Power Delivery, 2020.

[29] M. Dong, *A data-driven long-term dynamic rating estimating method for power transformers*, IEEE Transactions on Power Delivery, 2020.

[30] A. Kirilenko, M. Esmaili, C. Y. Chung, *Risk-Averse Stochastic Dynamic Line Rating Models*, IEEE Transactions on Power Systems, 2020.

[31] Cao B, Gu Y, Lv Z, Yang S, Zhao J, Li Y, *RFID reader anticollision based on distributed parallel particle swarm optimization.*, In IEEE Internet of Things Journal, vol. 8(5), pp. 3099–3107, Oct 2020.

[32] Cao B, Yan Y, Wang Y, Liu X, Lin JC, Sangaiah AK, Lv Z, *A multiobjective intelligent decision-making method for multistage placement of PMU in power grid enterprises*, IEEE Transactions on Industrial Informatics, pp.1–9, Oct 2022.

[33] Lai X, Yang B, Ma B, Liu M, Yin Z, Yin L, Zheng W, *An Improved Stereo Matching Algorithm Based on Joint Similarity Measure and Adaptive Weights*, Applied Sciences, vol. 13, no. 1, p.514, Dec. 2022.

[34] Morteza, Azita and Sadipour, Masod and Fard, Reza Saadati and Taheri, Saman and Ahmadi, Amirhossein, *A dagging-based deep learning framework for transmission line flexibility assessment*, IET Renewable Power Generation, 2022.

[35] Guo C, Ye C, Ding Y, Wang P, *A Multi-State Model for Transmission System Resilience Enhancement Against Short-Circuit Faults Caused by Extreme Weather Events*, In IEEE Transactions on Power Delivery, vol. 36, no. 4, pp. 2374–2385, Dec. 2020.

[36] Guo L, Ye C, Ding Y, Wang P, *Allocation of Centrally Switched Fault Current Limiters Enabled by 5G in Transmission System*, In IEEE Transactions on Power Delivery, vol. 36, no. 5, pp. 3231–32415, Nov. 2020.

[37] K. Serizawa, O. Tatebe, *Accelerating machine learning i/o by overlapping data staging and mini-batch generations*, In Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies, 2019 Dec 2, pp. 31-34.

[38] Prettenhofer P, Louppe G, *Gradient boosted regression trees in scikit-learn.*, InPyData 2014, Feb 2014.