Ensemble adversarial training-based robust model for multi-horizon dynamic line rating forecasting against adversarial attacks

Najmul Alam[a], M. A. Rahman[a], Md. Rashidul Islam[a,*] and M. J. Hossain[b]

[a]Department of Electrical & Electronic Engineering, Rajshahi University of Engineering & Technology, Kazla, Rajshahi 6204, Bangladesh
[b]School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia

ARTICLE INFO

Keywords:
Dynamic line rating (DLR)
Forecasting
Ensemble learning
Adversarial attacks
Ensemble adversarial training (EAT)

ABSTRACT

Dynamic line rating (DLR) forecasting is critical in the effective and economical utilization of overhead lines (OHLs) in smart grids, which facilitates the integration of renewable energy sources and reduces infrastructure upgrade costs. The forecasting techniques used for DLR rely on weather data collected from sensors as well as data communication, which can introduce a potential vulnerability to adversarial attacks. Hence, this work utilizes extreme gradient boosting (XgBoost), categorical boosting (CatBoost), and random forest as ensemble learning techniques for multi-horizon forecasting, while investigating their vulnerability by introducing adversarial attacks using two different attack models with variable data contamination and perturbations. Additionally, ensemble adversarial training (EAT)-based countermeasure is proposed for robust and accurate DLR forecasting. Experimental results indicate the outperformance of the CatBoost method compared to XgBoost and random forest models under normal conditions, while highlighting the vulnerability of all models to adversarial attacks in terms of root mean square error (RMSE) and mean absolute percentage error (MAPE). The proposed CatBoost with EAT significantly mitigates the impacts of adversarial attacks and retains accuracy under normal conditions. This research contributes to developing an accurate, cyber-resilience, and reliable forecasting methodology for line rating technology, leading towards academic and industrial developments in smart grids.

1. Introduction

Technological innovations, industrial growth, and increasing population have escalated the demand for electrical energy, causing an increase in load on existing transmission systems in smart grids. Also, as conventional fossil fuels are declining and not environmentally friendly as well, renewable energy sources are now supplying the power grid in a large proportion [1]. The integration of these various renewable energy sources is another critical factor that causes energy transmission congestion. Consequently, the existing transmission system of smart grids needs to be upgraded to meet the increasing power demand and to allow the quick integration of renewable energy sources [2]. Changing the physical infrastructure of transmission systems, such as the transmission line, entails a significant financial investment, an extended time frame, and substantial spatial requirements. The alternative is finding a way to use the existing resources optimally and effectively without any damage [3]. Consequently, the concept of dynamic line rating (DLR) or dynamic thermal rating (DTR) is introduced to adaptively control electric energy transmission through overhead lines (OHLs), while offering transmission congestion reduction economically with improved efficiency [4–6]. Although direct and indirect methods based on real-time monitoring of OHLs exist [7], their dependence on real-time measurements sometimes makes them less appealing, especially for planning as well as economical and effective utilization of existing resources. As addressed in [8] while presenting a comprehensive analysis of DLR forecasting, forecasting methods can be an alternative solution for predicting the ampacity from an hour to days ahead to support the effective implementation of DLR, which is focused on this work.

This work analyzes different ensemble learning-based algorithms for multi-horizon dynamic line rating forecasting and investigates their vulnerabilities to cyber-attacks by introducing adversarial attacks. As a countermeasure, ensemble adversarial training-based proactive defense is presented for robust forecasting against FGSM and BIM-based adversarial attacks.

*Corresponding author
najmulalam2872@gmail.com (N. Alam); mdabdur.rahman.1995@ieee.org (M. A. Rahman); rashidul@eee.ruet.ac.bd (Md.R. Islam); jahangir.hossain@uts.edu.au (M. J. Hossain)
ORCID(s): 0009-0000-5490-8178 (N. Alam); 0000-0003-0864-4882 (M. A. Rahman); 0000-0001-8415-0206 (Md.R. Islam); 0000-0001-7602-3581 (M. J. Hossain)

Based on the time horizon, forecasting methods can be categorized into short-term (0-6h), medium-term (6-48h), and long-term (more than 48 hours) [9]. Although long-term forecasting is crucial for grid management and the strategic planning of future grid development, most existing works have limited their focus on short- and medium-term DLR forecasting. In [10], a time series forecasting scheme has been discussed to estimate line rating. Adopted generalized linear models, multivariate adaptive regression splines, random forests, and quantile random forests have been used in [11] to predict ampacity up to 27 hours in advance for two conductor lines in Northern Ireland. For short-term forecasting, probabilistic, quantile regression (QR), and super-quantile regression (SQR) methods have been discussed in existing literature [9, 12]. With the intention of developing accurate generalized models, ensemble learning techniques have been discussed in [13] for short- and long-term DLR forecasting, where decision tree ensembles include random forest, adaptive boosting (AdaBoost), gradient boosting (GBoosting), extreme gradient boosting (XgBoost), light gradient boosting machine (LGBM), and categorical boosting (CatBoost). These notable DLR forecasting schemes are mostly data-driven, resulting in their dependency on data quality for accurate forecasting, as signified in [14]. Hence, they can be vulnerable to cyber-attacks, which can corrupt data or disrupt data flow.

Although DTR plays a critical role in maintaining the reliability of smart grids and supports effective integrations of renewable energy resources, very few research works have addressed its vulnerability to cyber-attacks due to its dependence on information and communications technology (ICT), which is very alarming, as addressed in [15]. As highlighted in [16–18], sensitive data in generation and transmission areas of a power grid can be compromised due to cyber-attacks, especially data contamination or corruption through cyberinfrastructure. Similarly, DLR systems can be affected by data contamination attacks, as discussed in [19]. Several studies [20–22] have addressed the effects of cyber-attacks on the cyber layer of DTR. The study in [20] has discussed the impacts of data integrity, communication network vulnerability, device spoofing and tampering, and coordinated cyber-physical attacks on cyber-physical power networks by focusing on applications of phasor measurement units (PMUs). Also, the impacts of discontinuous service of ICT in terms of PMUs' failure/malfunction on network reliability in a power system integrating DTR and system integrity protection schemes (SIPS) have been discussed in [21]. In [22], a framework for synchrophasor-based DTR and SIPS has been presented, revealing that communication network outages can reduce reliability by up to 99.15% without SIPS, highlighting the adverse effects of single-path failures on network performance. Furthermore, the importance of DLR forecasting in ensuring the secure and reliable operation of DTR systems has been signified in [23]. Although the studies in [20–23] are notable for securing DTR systems, they have not addressed the vulnerability of DLR forecasting schemes.

In the case of DLR forecasting schemes, the threats of cyber-attacks can be prominent, as they significantly depend on meteorological information collected from different weather monitoring stations using ICT. While the vulnerability of forecasting algorithms could be investigated in general, existing works [24] have investigated the vulnerability of the forecasting algorithms, focusing on application areas in smart grids to highlight the severity in that particular sector. For example, among different data contamination attacks, false data injection (FDI) attacks have been considered in different application areas of forecasting algorithms in smart grids, such as load forecasting [25], generation forecasting [26], transformer's lifetime forecasting [27], and DLR forecasting [13]. In [13], FDI attacks have been introduced to address vulnerabilities of DLR forecasting without any countermeasure, where ensemble learning algorithms have been used for forecasting across various time horizons. The research studies discussed in [28, 29] are notable for presenting countermeasures against FDI attacks to improve the resiliency of DLR forecasting models, but the models employed in them are not generalized. Additionally, the countermeasures discussed in [28, 29] involve data pre-processing schemes whose failure can significantly affect forecasting models.

As the investigation of cyber-attacks on DLR forecasting is still in its infancy, the literature [24] is mostly limited to scaling model-based FDI attacks consideration despite the possibility of different cyber-attacks. Forecasting models in smart grids can be vulnerable to adversarial examples, which has been signified in [30, 31] for solar power forecasting and load forecasting. Similarly, adversarial attacks on DLR forecasting can be threatening to the economical, stable, and robust operations of transmission systems, which is yet to be addressed, as indicated in [24]. Hence, this work investigates the severity of different adversarial attacks on DLR forecasting models. As the failure of any pre-processing countermeasure can decrease forecasting accuracy, this work focuses on improving the robustness of forecasting models. The contributions of this work are presented as follows.

- Presenting a comparative analysis of ensemble learning algorithms for multi-time horizon forecasting,
- Investigating adversarial attacks using two different attack models on DLR forecasting methodologies, especially on ensemble learning-based multi-time horizon DLR forecasting methodologies,

- Presenting a comparative analysis of the severity of adversarial attacks on DLR forecasting algorithms during diverse attack cases,
- Proposing an ensemble adversarial training-based robust DLR forecasting scheme.

The organization of the paper is outlined as follows. In Section 2, line rating technology and existing works on DLR forecasting are discussed briefly, including a foundational understanding of the DLR calculation of an OHL. Following that, Section 3 presents the methodologies for DLR forecasting along with the proposed scheme for countermeasures, covering machine learning (ML) models, adversarial training, and ensemble adversarial training. The adversarial attack methodologies used in this work are discussed in Section 4. Section 5 presents preparations for the investigations and experimental observations, encompassing the outcomes for clean, attacked, and mitigated scenarios. A thorough analysis of the dataset and evaluation metrics are discussed as experimental preparation in Section 5. The paper is concluded in Section 6 with a concise summary.

2. Background

In the conventional transmission system, an ampacity value is assigned beforehand for an OHL to allow a fixed power flow transmission, where ampacity is defined as the current carrying capacity of the electrical power transmission line at a given voltage. Assigning a fixed ampacity rating based on predetermined environmental conditions and physical circumstances for safe energy transmission is known as static line rating (SLR). For ampacity rating selection, physical circumstances include maximum operating line temperature, sag, electrical resistance, and mechanical strength of the line conductor, while environmental conditions include ambient temperature, wind speed, wind direction, and GHI [10].

In SLR scheme, the rating is calculated taking into account the worst case of the metrological conditions, which includes 40^0C as summer air temperature, 1000 W/m^2 as full sun solar irradiance, $0.6m/s^2$ wind speed, and perpendicular wind direction. However, environmental conditions are variable. Consequently, employing SLR hinders the effective and optimum utilization of transmission lines, as SLR is confined to a constant value to avoid conductor overheating and premature aging, considering the worst ambient conditions. As a flexible scheme, the concept of seasonal line rating is considered to determine the line rating by taking into account different sets of ambient condition assumptions for distinct seasons. Although it better utilizes existing systems than SLR, season-wise line rating selection is still insufficient and ineffective.

Due to adaptively adjusting the ampacity rating according to ambient conditions, DLR or DTR is a viable solution, which can prevail over the shortcomings of the SLR as well as seasonal line rating schemes and exaggerate the transmission capacity [32, 33]. Here, the DTR of a transmission line indicates the highest electric current that the line can support without exceeding temperature limits considering the ambient conditions, which is essential to ensure reliable and safe operation with maximized utilization. As addressed in research studies [34–36], utilization of DLR can potentially enhance line capacity by 10%–30%, which can be up to 50% in regions with strong winds. For instance, a case study conducted in Texas on the 138 kV lines has demonstrated an average real-time capacity increase from 8% to 12%, while an average real-time capacity increase ranging from 6% to 14% has been reported for the 345 kV lines, as discussed in [37]. Hence, DLR is focused on this work. This section briefly discusses the process of determining the DLR of OHLs and existing notable works on DLR forecasting.

2.1. Evaluation of DLR of an OHL

Dynamic line rating involves calculating the rating by considering variations in meteorological parameters. The influence of these parameters on the capacity of transmission lines can be examined based on IEEE-738 standard [38] and CIGRE Technical Brochure TB 601 [39]. IEEE-728 2012 standard provides a comprehensive framework for determining the current-temperature relationship of uncovered OHLs. The thermal balance process of an OHL is illustrated in Fig. 1, whose characteristics can be expressed as in Eqn. (1).

$$Q_r + Q_c = Q_j + Q_s \tag{1}$$

where Q_r, Q_c, Q_j, and Q_s represent radiated heat loss, convective loss, Joule heating, and heat gain from solar irradiation, respectively.

The radiated heat loss (Q_r) transpires when the surrounding temperature is higher than the conductor, whose rate relies on the temperature difference between the conductor and surroundings as well as on the emissivity of the

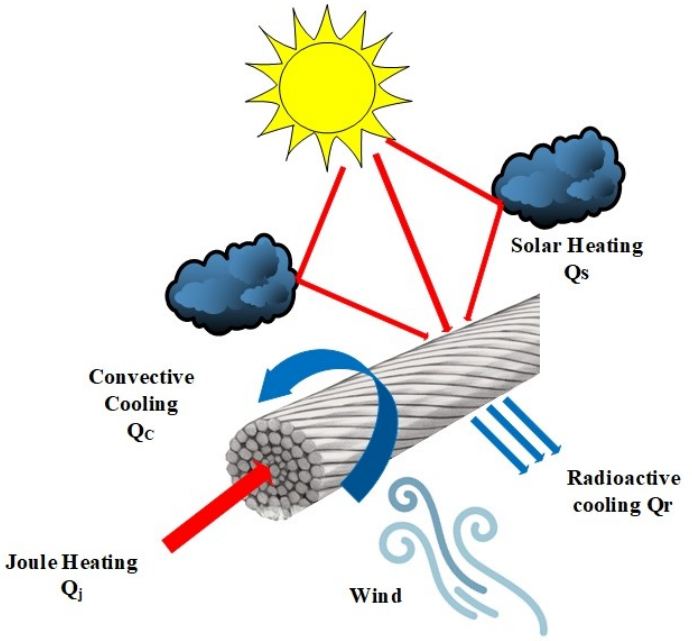

Figure 1: Thermal balance process of OHLs.

conductor. The convective loss (Q_c) can be classified into natural convection (Q_{cn}) and forced convection Q_{cf}. Natural convection (Q_{cn}) occurs in zero-wind conditions where warm air is replaced by cool air and can be mathematically expressed as follows.

$$Q_{cn} = 3.3645 \cdot \rho_f^{0.5} \cdot D^{0.75} \cdot (T_{con} - T_{air})^{1.25} \tag{2}$$

where ρ_f, D, T_{con}, and T_{air} denote air density, conductor diameter, conductor surface temperature, and ambient air temperature, respectively. Forced convection (Q_{cf}) is initiated when the wind causes the air to circulate around the conductor, compelling cooler air to make contact with the conductor. Force convection at low wind speed (Q_{cfl}) and force convection at high wind speed (Q_{cfh})can be expressed as follows.

$$Q_{cfl} = K_{angle} \cdot K_f \cdot [1.01 + 1.13 \cdot N_{Re}^{0.52}] \cdot (T_{con} - T_{air}) \tag{3}$$

$$Q_{cfh} = K_{angle} \cdot K_f \cdot 0.754 \cdot N_{Re}^{0.6} \cdot (T_{con} - T_{air}) \tag{4}$$

where N_{Re} is the Reynolds number, K_f is the thermal conductivity of the conductor, and K_{angle} is the wind direction factor. K_f is a function of film temperature (T_f). K_{angle} is a function of the angle of wind direction and is perpendicular to the conductor axis ((β)). K_f and K_{angle} can be mathematically expressed as follows.

$$K_f = 2.424 \times 10^{-2} + 7.477 \times 10^{-5} T_f - 4.407 \times 10^{-9} T_f^2 \tag{5}$$

$$K_{angle} = 1.194 - \sin\beta - 0.194 \cdot \cos 2\beta + 0.368 \sin 2\beta \tag{6}$$

On the other hand, the mathematical model of Joule heating (Q_j) is evaluated as a function line current (I) and AC resistance of the conductor ($R(T_{avg})$) at temperature T_{avg}, which can be mathematically expressed as follows.

$$Q_j = I^2 \cdot R(T_{avg}) \tag{7}$$

The heat gain from solar irradiation (Q_s) calculated from the absorptivity of the conductor ((α)), radiant heat intensity from the sun and sky, adjusted for elevation ((Q_{se})), effective angle of incidence of the sun's rays ((θ)), and projected area of conductor per-unit length ((A')). The mathematical model of Q_s is expressed as follows.

$$Q_s = \alpha \cdot Q_{se} \cdot A' \sin(\theta) \tag{8}$$

After evaluating Q_r, Q_c, and Q_s, the line current (I) is determined based on the mathematical model of the thermal balance of an OHL expressed in Eqn. (1) and the mathematical model of Q_r expressed in Eqn. (7), which is expressed as follows.

$$I = \sqrt{\frac{Q_r + Q_c - Q_s}{R(T_{avg})}} \tag{9}$$

Due to consideration of ambient conditions, Eqn. (9) can be used to evaluate line rating dynamically, which maximizes the utilization of existing resources.

2.2. Forecasting methods

There are several notable research works on DLR forecasting, including synthetic data preparation. In [10], an artificial dataset has been prepared from statistical analysis of weather data, which has been used to estimate line rating by time series forecasting. Numerical weather predictions have been utilized to predict ampacity up to 27 hours in advance for two conductor lines in Northern Ireland in [11] by employing generalized linear models, multivariate adaptive regression splines, random forests, and quantile random forests. Quantile regression (QR) and super-quantile regression (SQR) techniques have been discussed for a very short-term DLR prediction in [12]. The study in [9] has focused on reliable as well as accurate DLR forecasting and safe rating value of DLR based on probabilistic forecasting for 24h ahead. In another research work [40], an integrated artificial neural network (ANN) has been used to predict DTR as well as conductor temperature.

In addition to focusing on DLR forecasting, existing works have explored system behavior as well as data dependency while developing models for multi-time horizon forecasting. Most forecasting models use meteorological data, resulting in their dependency on data quality to ensure accuracy. This issue has been highlighted in [14] by discussing experimental observations of tree-based algorithms, where the data quality has affected the forecasting accuracy. In another work [33], the system's behavior under varying conditions is explored based on sequential Monte Carlo (SMC) simulation, while capturing the time-dependent patterns in line ratings and wind power using auto-regressive and moving-average (ARMA) model to analyze their correlation. As most works have focused on single-term forecasting, ensemble learning techniques have been discussed in [13] for short- and long-term DLR forecasting using metrological data, while offering an increase in capacity up to 30% of a 400 kV line. The research work presented in [13] is notable for developing generalized forecasting models by employing decision tree ensembles like random forest, AdaBoost, GBoosting, XgBoost, LGBM, and CatBoost due to their flexible learning ability. Although these works are notable, they are usually limited to forecasting over a single or two time horizon. Hence, this work aims to present a robust multi-time horizon forecasting model.

3. Methodologies for forecasting & countermeasures

This section presents a summarized discussion of forecasting models used in this work, along with their hyper-parameter configuration. Additionally, a countermeasure is proposed in this section as a defense against adversarial attacks. The methodologies are presented as follows.

Table 1
Hyperparameter configurations for XgBoost, CatBoost, and random forest algorithms

Algorithm	Hyperparameters
XgBoost	{n_estimators=2500, learning_rate=0.08, max_depth=5, subsample=0.7, reg_alpha=0.1, reg_lambda=0.5}
CatBoost	{iterations=1000, depth: 7, learning_rate: 0.1, l2_leaf_reg: 1.0 }
Random forest	{n_estimators: 100, random_state: 42, max_depth: None, min_samples_split: 2, min_samples_leaf: 1 }

3.1. Forecasting models

ML has become a trending issue in every engineering field [41], while already establishing its vast popularity in computer vision, data science, natural language processing, image processing, and so on. Among different methodologies, ensemble learning is an ML technique that can enhance prediction accuracy by combining the output of multiple models. As addressed in [13], ensemble learning can be employed for two-time horizon DLR forecasting, indicating its generalizability. Hence, ensemble techniques are utilized in this work as it aims to develop a multi-horizon forecasting model. The commonly employed ensemble techniques are bagging, random forest, stacking, and boosting. The ensemble techniques used in this work are summarized as follows, while presenting their hyperparameter configurations in Table 1.

- **XgBoost:** XgBoost refers to the extreme gradient boosting method, which is a scalable end-to-end tree boosting algorithm [42]. The gradient-boosting architecture on which XgBoost is built combines the predictions of several weak learners to produce a powerful predictive model. XgBoost includes a feature with built-in regularization techniques that help prevent overfitting and improve model generalizability. It also uses pruning, which entails chopping out tree branches that do not make an important impact on the model's overall performance and regulates the decision trees' complexity.

- **CatBoost:** CatBoost is a gradient-boosted decision tree ensemble technique introduced in 2018 [43]. It is a useful tool for dealing with large amounts of data. CatBoost is endorsed for its capability to handle categorical and heterogeneous data effectively. The approach employs ordered target statistics to prevent target leakage in categorical feature encoding, leveraging statistics from prior trees. Additionally, exclusive feature combinations efficiently manage high-cardinality categorical features by grouping sparse combinations, reducing dimensionality. Catboost supports order boosting to reduce the chance of overfitting and leverages graphics processing unit (GPU) acceleration for convenience in fast training and predictions.

- **Random forest:** Random forest is an ensemble learning method that combines the predictions of multiple decision trees during training and determines the final output by either taking the mode (for classification tasks) or the mean (for regression tasks) of the predictions from each tree [44]. Each tree in the random forest is constructed using a random subset of the training data and features, which introduces diversity and lowers the risk of overfitting. Random forest is suitable for the application of classification and regression tasks.

3.2. Countermeasure methods

This section presents countermeasure methodologies against adversarial attacks. As addressed in [45], the countermeasures against adversarial attacks can be categorized into two strategies, which are the reactive approach and the proactive approach. A reactive approach is a pre-processing approach, which involves identifying and dealing with adversarial examples before feeding data to an existing model. This type of defensive measure causes no change in an existing. Consequently, it is deployed after constructing the model. However, the failure of a reactive approach can severely affect performance. On the other hand, a proactive approach enhances the robustness of the model preemptively. This approach is implemented before deploying the model for applications. As it ensures a model's

performance without involving an external process during applications, this work utilizes a proactive approach as a countermeasure, which is presented below.

Adversarial training is a proactive defense strategy introduced in [46]. The key concept of adversarial training is to inject adversarial examples into the training dataset to improve the robustness of the model. During training, adversarial examples are generated by applying carefully crafted small perturbations to the input data. These adversarial examples are combined with the original training data to create a new training dataset, known as the augmented dataset, which includes clean and adversarial samples. The model is then trained on that augmented dataset, where it learns to handle both clean and adversarial inputs. The objective is to make the model more robust by preventing it from being overly sensitive to adversarial attacks, which is achieved by minimizing the total loss. In adversarial training, the total loss is calculated as the summation of the original loss and the adversarial loss, which can be mathematically expressed as follows.

$$\overline{J}_\theta(X, Y) = \gamma(J_\theta(X, Y)) + (1 - \gamma)J_\theta(X^{adv}, Y) \tag{10}$$

where X is clean input data, Y is the output data, and X^{adv} is the adversarially corrupted input data. J_θ is the Jacobian matrix dependent on both input and output data. γ is a weight parameter, which is a constant between 0 and 1, to determine the balance between the original loss and the adversarial loss.

Adversarial training involves training a single model for developing a robust system. On the other hand, instead of using a single model, ensemble adversarial training (EAT) uses multiple models to develop a robust system [47]. An ensemble refers to the utilization of a diverse collection of individual models, combining their outputs to make predictions. This approach aims to enhance overall performance by leveraging the strengths of each constituent model and mitigating individual weaknesses, where these models may differ in architecture, initialization, or training data. As the corrupted data portion and the strength of an attack can vary in real-life scenarios, different combinations of training data are used in this work for EAT to defend against diverse adversarial attack cases, which is investigated later in Section 5.2.

In EAT, the individual predictions are combined for a more robust and accurate outcome. This can be achieved through methods like averaging, weighted averaging, or majority voting, where each model contributes a vote for the final decision [48]. A stacking approach to combine individual predictions goes a step further by training a meta-model to make the ultimate prediction. Dynamic weight averaging (DWA) is another technique that dynamically adjusts the weights of individual model predictions based on their current performance. This adaptive approach allows the model to assign more influence to models with higher accuracy, enhancing overall ensemble robustness and predictive accuracy, as addressed in [49]. Considering the adaptivity and advantages offered by the DWA, this work utilizes it to combine the individual outcomes in EAT. The algorithm for an EAT-based defensive scheme is presented below.

4. Modeling of cyber-attacks

The massive integration of ICT for measurement, monitoring, and control of the physical infrastructure in smart grid systems has created a large attack surface. With the employment of ML and deep learning (DL) models, adversarial attacks have become a very concerning issue nowadays for smart grids. Adversarial examples are corrupted inputs to machine learning or DL models, where these corrupted inputs are generated by intentionally manipulating input data or adding perturbation to input data in a subtle way to cause the model to make an incorrect prediction or classification. These perturbations are often so slight that they are imperceptible to humans. However, as DL models are trained with normal data, small perturbations due to adversarial attacks can significantly alter the output of a model. The adversarial attack was first introduced in [50], which has been discussed widely for image processing applications, especially for image classification. However, as addressed in [51], adversarial attacks can be used on time series data. Similarly, the effects of adversarial attacks can be severe for applications of forecasting models in smart grids. Although adversarial attacks on generation and load forecasting have been discussed in existing works [30, 31], to the best of the authors' knowledge, no study has investigated adversarial attacks on DLR forecasting, which is the objective of this work.

Adversarial attacks can be categorized based on different factors. For instance, based on the information available to an attacker, an adversarial attack can be classified into a white-box attack and a black-box attack [45]. In case of a white-box attack, the attacker has complete knowledge about the model, including the training model architecture, training data, hyper-parameter, and model weights. On the contrary, the black-box attack adversary does not have any knowledge about the trained model. The attacker will act as a standard user and only can manipulate the input data. As

Algorithm 1: Defensive approach using ensemble adversarial training

Input: Clean test dataset $D_{clean} = [X_0, X_1 \dots X_N]$
Attacked test dataset $D_{attacked}$=$[X_0, X_1, X_2^{adv}, X_3, \dots, X_N^{adv}]$
Output: Forecasted ampacity
Step 1: Generate augmented datasets.
Initialize: Percentage of attack data p , magnitude of perturbation ϵ
Input: Clean sample $X_0, X_1 \dots X_N$
for $i <= N$ **do**
  Generate FGSM adversarial example: X_N^{adv}
  `// attack on p% random data`
  $X^{clean} \leftarrow X^{adv}$
end
Return: $D_{mix} = D_{clean} + D_{attacked}$
Step 2: Train the models
Initialize: Number of models, M
for $j <= M$ **do**
  $m_j \leftarrow train\_model(\mathrm{D}_{mix})$
end
Return: M numbers adversarial trained model
Step 3: Forecast ampacity from individual model
for $k <= M$ **do**
  f_k= Prediction(Dataset)
  `// dataset can be` D_{clean} `or` $D_{attacked}$
end
Step 4: Combine the predictions using dynamic weighted average
for $k <= M$ **do**
  $P[k] \leftarrow \mathrm{f}_k$
  `// Predictions store is an empty array P[]`
end
$P[k] \leftarrow$P[k] '
$dw \leftarrow \frac{f}{Sum(f)}$
`// each element of row of array` $P[k]$, f
$E_f \leftarrow$ (dw * P[k])
`// Ensemble forecast,` E_f
Return: Forecasted ampacity, E_f

a black-box attack has a close resemblance to real scenarios, considering the fact that the attacker will have difficulty bypassing security to access the model for a white-box attack, this work has focused on black-box attacks to explore the vulnerability of DLR forecasting models.

Another criterion for categorizing adversarial attacks focuses on the attack implementation process. Depending upon the number of steps required for successful attack implementation, adversarial attacks can be single-step or iterative. The fast gradient signed method (FGSM) is an example of a single-step attack, where the perturbation for each example is generated by calculating the loss gradient only once. An iterative attack can be implemented using the basic iterative method (BIM), where the perturbation is determined iteratively to achieve the maximized effect. Compared to an FGSM attack, a BIM attack involves more computational complexities but offers a higher attack success rate. As both FGSM and BIM attacks can be possible in real scenarios, this work considers both to investigate the impacts of adversarial attacks on DLR forecasting. In this section, the attacks are described briefly, along with their implementation process for this work.

4.1. FGSM attack

An FGSM attack [46] is a one-step adversarial attack, which is one of the most widely used methods for adversarial attacks in different sectors. The attack implementation involves a single iteration to decide the perturbations to be added to an image solely in the direction indicated by the sign of the gradient of the loss for that image data. The mathematical expression of the perturbation (η) for an input data (X) and the corresponding desired output (Y) is presented as follows.

$$\eta = \epsilon \cdot \text{sign}(\nabla_X J_\theta(X, Y)) \tag{11}$$

where ϵ is a hyper-parameter that controls the magnitude of the perturbation. J_θ is the Jacobian matrix, dependent on both X and Y. θ represents the model parameters. The $sign(\cdot)$ function ensures that the perturbation causes the maximum deviation from the desired output. The generated adversarial example for the clean sample (X) is computed as follows.

$$X^{adv} = X + \eta \tag{12}$$

An FGSM attack is a white-box attack and is computationally inexpensive. In the high-dimension problem, the drastic change in the output can be seen due to the small perturbation at each dimension. As this work aims to investigate the impacts of FGSM attacks as a black-box attack, a surrogate model is used to determine the perturbations. The algorithm for a black-box FGSM attack is presented as follows.

Algorithm 2: Black-box FGSM attack

Input: Clean sample X, magnitude of perturbation ϵ
Output: FGSM attacked sample X^{adv}
Step 1: Forecast using surrogate model
Train Linear regression model
Forecast using test data, F_s
Step 2: Attack implementation
Compute gradient $\nabla_x J_\theta(X, F_s)$
$\eta \leftarrow \epsilon \cdot \text{sign}(\nabla_x J_\theta(X, F_s))$
$X^{adv} \leftarrow X + \eta$
Return: X^{adv}

4.2. BIM attack

A BIM attack [52] is an iterative extension of the FGSM attack. For determining perturbations, BIM takes an iterative approach by applying FGSM attack multiple times with a defined step size α. A clipping operation is integrated so that the perturbations remain imperceptible and restricted within a defined numerical range in proximity to the original value. By assigning the input data (X) as the initial value of the adversarially corrupted input (X_o^{adv}), the mathematical model to determine the adversarially corrupted input through the iterative approach can be expressed as follows.

$$X_{N+1}^{adv} = Clip_{X,\epsilon}\{X_N^{adv} + \alpha \cdot \text{sign}\nabla_X J(X_N^{adv}, Y)\} \tag{13}$$

where X_{N+1}^{adv} indicates the adversarially corrupted input after $N+1$ iteration, X_N^{adv} indicates the adversarially corrupted input after N iteration, J is the Jacobian matrix, Y is the desired output, and ϵ is the perturbation's magnitude controlling hyper-parameter.

BIM requires more computational resources because it involves a predefined iteration number. BIM attack is also a white box attack like FGSM. By tuning the ϵ, the attacker has the opportunity to control how far the adversarial sample is pushed beyond the decision boundary. The algorithm for implementing a black-box BIM attack is as follows.

Algorithm 3: Black-box BIM attack

Input: Clean sample X, magnitude of perturbation ϵ, step size α, number of iteration I
Output: BIM attacked sample X^{adv}_{N+1}
Step 1: Forecast using surrogate model
Train Linear regression model
Forecast using test data, F_s
Step 2: Attack implementation
$X^{adv}_{o} \leftarrow X$
for $n < I$ **do**
  $X^{adv}_{N+1} = Clip_{X,\epsilon}\{X_N + \alpha \cdot \text{sign}\nabla_X J(X^{adv}_N, F_s)\}$
end
Return: X^{adv}_{N+1}

4.3. Attack implementation strategy

DLR forecasting methods rely on meteorological data, where the meteorological data shows fluctuations over time. Hence, adversarial attacks are implemented to corrupt the meteorological data in this study. This study considers FGSM and BIM attacks for implementing adversarial attacks based on the abovementioned algorithms. Although these are white-box attacks, they are implemented as black-box attacks, as mentioned earlier. As a surrogate model, a linear regression model is used to replicate the target model [53]. During each attack's implementation, 20% and 50% of the test dataset for temperature data are corrupted with varying values of ϵ, where $\epsilon = [0.5, 1, 10]$. Furthermore, for the BIM attack, 0.05 is used as the step size for change per iteration (α), while 200 is assigned as the number of iterations. The impacts of these attacks are investigated in detail in Section 5.2.

5. Experimental analysis

This section presents a summarized discussion of the experimental setup, including an overview of the dataset and performance evaluation metrics. Also, a detailed analysis of experimental observations is presented in this section to address the severity of adversarial attacks on multi-horizon forecasting and to verify the effectiveness of the proposed robust forecasting scheme.

5.1. Experimental setup

This section presents an overview of the data used in this work. Additionally, mathematical models of the performance evaluation metrics used in this work are discussed briefly.

5.1.1. Data description

For DLR forecasting, historical and real-time data of the meteorological parameters are required, which are usually obtainable on the website of the nearby weather stations. To reflect the practical scenarios, the data used in this study is from Trang-Thap Cham 220 kV OHL with a length of 117.794 km, and the line is situated in the South Central Coast region in Vietnam [54]. The site selection was based on its significant renewable energy potential in Vietnam, encompassing onshore and offshore wind power as well as solar power. This choice is motivated by the necessity to augment transmission capacity to address the escalating energy demand and facilitate the integration of new generation sources. The characteristics of OHL for the considered case are summarized in Table II.

The dataset includes four meteorological parameters: ambient temperature, wind speed, wind direction, and global horizontal irradiance (GHI). The actual ampacity is calculated using IEEE-738 2012 at one-hour intervals over a period of eighteen months. One year of data has been designated for training, with the remaining six months allocated for testing across different time horizons. Data has taken one-hour intervals because the study in [55] shows that attaining a steady state in response to a step change in current requires approximately half an hour. Consequently, the transients in DLR can be disregarded, allowing for the estimation of DLR on an hourly basis.

While the considered four meteorological parameters are critical [56], each of them correlates differently with the ampacity, as illustrated in Fig. 2. As illustrated in Fig. 2, wind speed and wind direction have a propitious effect on the line rating. The wind direction exhibits a sinusoidal pattern, which indicates that the cooling influence of the wind is

Table II
Properties of the OHL.

Parameter	Value
Line length	117.794 km
Nominal Voltage	220 kV
Conductor type	ACSR
Conductor sections	394/51.1 mm^2
Conductor diameter	394 mm
Emissivity	0.7
Absorbability	0.9
Elevation	30-35 m
Static Line Rating	850A

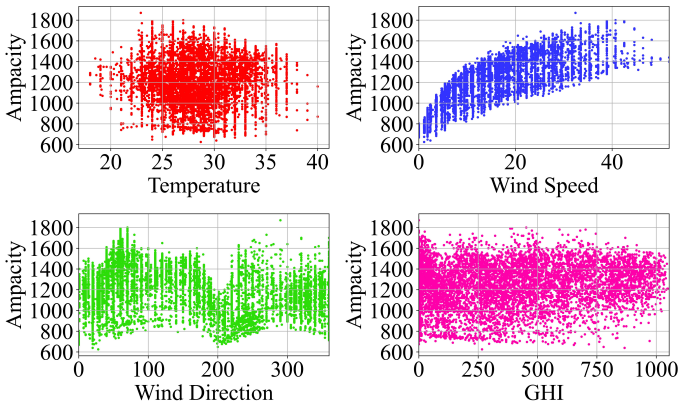

Figure 2: Correlation of DLR with meteorological parameters.

consistent for directions perpendicular to the line. As presented in Fig. 2, the effect of wind speed allows an increase in the ampacity in DLR proportionally with an increase in wind speed. On the other hand, the temperature and GHI have an adverse impact on DLR, which is illustrated in Fig. 2. As the existing works investigating the impacts of cyber-attacks on forecasting algorithms in smart grids focus on corrupting single input data to the best knowledge of the authors, attacks are implemented on single input data utilizing the attack models discussed in Section 4. As temperature is a critical parameter in determining the DLR of an OHL based on Eqn. (9) and has an adverse relationship with DLR, attacks are implemented on temperature in this work.

5.1.2. Performance evaluation metrics

Accuracy is a commonly employed criterion for assessing the efficacy of forecasting models. It is often expressed as a percentage and can be calculated using various metrics depending on the nature of the problem. In the context of DLR forecasting models, accuracy is crucial to ensure that the predicted ampacity values are close to the actual values. In this work, root mean square error (RMSE) and mean absolute percentage error (MAPE) are used as performance evaluation metrics. The mathematical models of the evaluation metrics used in this work are presented as follows, where C_i represents the calculated or actual ampacity, F_i represents the forecasted ampacity, and N is the number of samples.

- RMSE: RMSE is a well-known error measurement metric, which indicates accuracy as the average difference or distance between actual and estimated values. It can be mathematically expressed as in Eqn. (14).

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(C_i - F_i)^2} \tag{14}$$

Table III
Comparison of the performance of forecasting models under normal conditions.

Algorithm	Horizon	RMSE	MAPE
CatBoost	24h	**9.78**	**0.48**
	1m	**7.08**	**0.33**
	3m	**6.02**	**0.31**
	6m	**5.14**	**0.29**
Random forest	24h	35.98	1.81
	1m	27.58	1.16
	3m	22.77	0.99
	6m	18.20	0.82
XgBoost	24h	12.85	0.59
	1m	9.14	0.45
	3m	7.92	0.40
	6m	6.71	0.36

Bold font – Best performance

- MAPE: MAPE represents the accuracy of forecasts by calculating the average percentage difference between predicted and actual values to highlight larger errors, while offering a straightforward percentage measure for interpretation. The mathematical model of MAPE can be expressed as in Eqn. (15).

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{C_i - F_i}{C_i}\right| \times 100\% \tag{15}$$

5.2. Results & discussions

The experimental results are organized in three parts. At first, the performance of the selected DLR models is evaluated at normal conditions for multi-horizon DLR forecasting to identify an optimum model for normal cases. Next, two adversarial attacks discussed in Section 4 are implemented for input data contamination to analyze the impacts of adversarial attacks on the selected models. Different attack cases are considered for a detailed analysis of attack severity and for identifying a comparatively less affected model. Finally, the performance of the optimum model with EAT is evaluated to verify the effectiveness of the proposed countermeasure scheme.

5.2.1. Normal cases

The performance of forecasting models, discussed in Section 3.1, is evaluated in terms of RMSE and MAPE under normal conditions, which is presented in Table III in an organized manner. The forecasting models are used for multi-horizon forecasting, which includes forecasting for 24 hours, 1 month, 3 months, and 6 months. As presented in Table III, CatBoost stands out with superior performance compared to random forest and XgBoost, which is evident by the low values of RMSE and MAPE. XgBoost is found to be better than random forest during the experimental analysis. Irrespective of the model, the forecasting accuracy is improved with the increase in time horizon in forecasting, resulting in the best performance for 6 months. This phenomenon can be attributed to the superior ability of ensemble methods to capture the inherent seasonal patterns and trends inherent in meteorological data, particularly over longer time horizons such as daily, weekly, and yearly cycles. During forecasting over short-term horizons, it may be challenging to capture these trends effectively, resulting in less accurate predictions. However, longer-term forecasts benefit from the capabilities of ensemble learning, allowing for more accurate predictions by adequately capturing the intricate long-term patterns and trends in meteorological data. This behavior of each model during multi-horizon forecasting indicates a critical consideration for effective grid planning and further development strategies.

At the same time, the accuracy, varying from 0.29 to 0.48 in terms of MAPE and from 5.14 to 9.78 in terms of RMSE, indicates the CatBoost as the optimum model, which can be applied for multi-horizon forecasting with reasonable accuracy, as indicated in Table III. The optimum performance of CatBoost can also be observed visually in Fig. 3, which depicts the forecast results for DLR. As illustrated in Fig. 3, the difference between actual data and forecasts of CatBoost is barely noticeable, while slightly noticeable and noticeable differences are observed for XgBoost and random forest, respectively. As a consequence of using CatBoost for DLR selection, a 41.4% increase in the

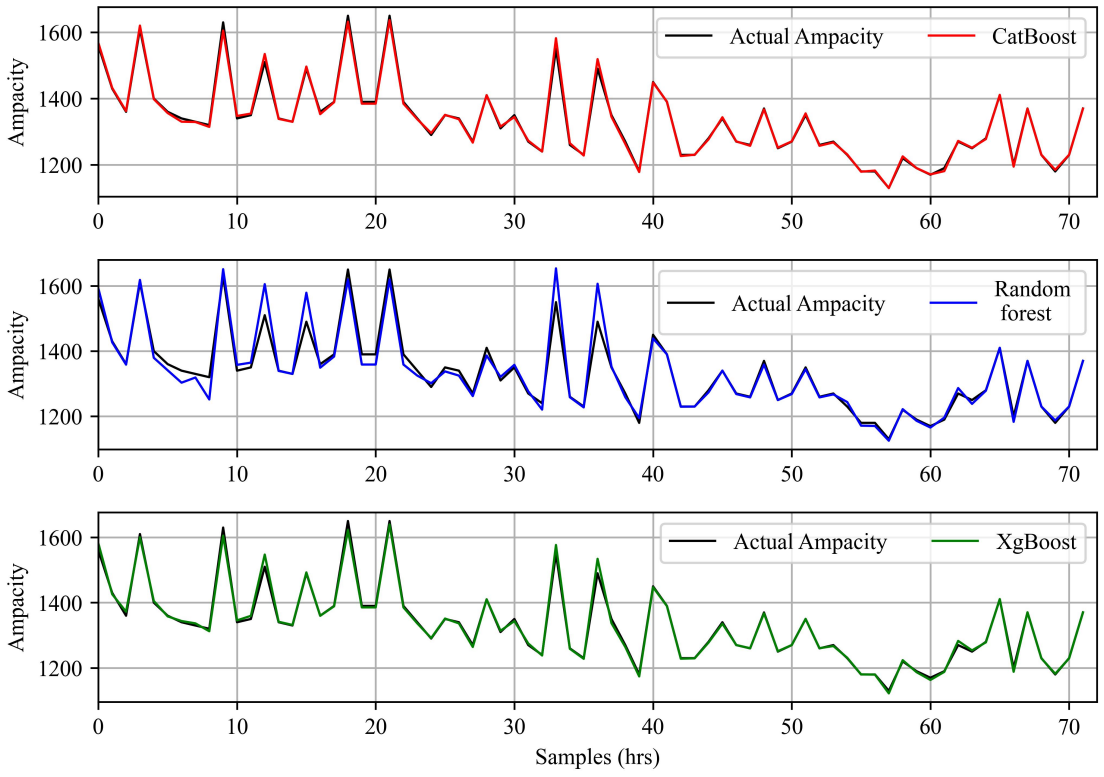

Figure 3: Comparison of the actual ampacity and the forecasted data using CatBoost, random forest, and XgBoost for 72 hours.

ampacity can be achieved compared to the ampacity selected by SLR, emphasizing the effectiveness of the forecasting models.

5.2.2. Cyber-attack cases

The vulnerabilities of DLR forecasting models to adversarial attacks are investigated using FGSM and BIM attacks with different perturbations and two different amounts of data contamination, as discussed in Section 4.3. From the attack perspective, both types of attacks are capable of causing significant degradation in forecasting accuracy for any algorithm, as presented in Table IV. The decrease in forecasting accuracy increases with increases in data contamination percentage and the value of ϵ, irrespective of attack type. Consequently, 20% data contamination with $\epsilon = 10$ is found to be severe in most cases for any forecasting models during FGSM and BIM attacks, emphasizing the necessity of developing robust forecasting models. For the selected forecasting algorithms, the impacts of FGSM and BIM attacks are found to be different for any percentage of data contamination as well as for any value of ϵ. For instance, the FGMS attack is found to be comparatively severe for CatBoost and XgBoost, whereas the BIM attack is found to be comparatively severe for the random forest. Hence, for a black-box attack, the selection of attack type will depend on the availability of computational resources for attack implementation.

On the other hand, from the algorithm perspective, there are slight decreases in forecasting accuracy for any forecasting model during adversarial attacks with any percentage of data contamination and $\epsilon = [0.5, 1]$, as presented in Table IV. For these attack scenarios, CatBoost outperforms others in most cases during adversarial attacks due to having the best performance under normal circumstances. However, only for adversarial attacks with $\epsilon = 10$, random forest is found to be better than others, as presented in Table IV. Consequently, during this extreme case of ϵ, random forest and CatBoost are the least and most affected models, respectively. For example, during forecasting for a 24-hour time horizon, the accuracy of random forest can be decreased up to 29.35% in terms of RMSE and 46.96% in terms of MAPE for its worst-attack case scenario with $\epsilon = 10$, whereas the accuracy of CatBoost can be decreased up to 601.64% in terms of RMSE and 656.25% in terms of MAPE for its worst-attack case scenario with $\epsilon = 10$. In the

Table IV
Comparison of forecasting accuracies of different models under diverse adversarial attack scenarios.

Algorithm	Attack information: Attack type	Attack information: Contamination percentage	ϵ	24h RMSE	24h MAPE	1m RMSE	1m MAPE	3m RMSE	3m MAPE	6m RMSE	6m MAPE
CatBoost	FGSM	20%	0.5	**10.18**	**0.50**	**7.32**	**0.35**	**6.32**	**0.33**	**5.49**	**0.31**
			1	**10.58**	**0.51**	**8.14**	**0.41**	**7.32**	**0.39**	**6.48**	**0.37**
			10	34.53	1.28	39.08	1.57	38.74	1.54	35.55	1.48
		50%	0.5	**11.77**	**0.64**	**7.59**	**0.38**	**6.84**	**0.37**	**6.17**	**0.37**
			1	**11.31**	**0.58**	**9.27**	**0.53**	**8.74**	**0.52**	**8.04**	**0.52**
			10	68.62	3.63	63.58	3.68	61.71	3.52	55.94	3.26
	BIM	20%	0.5	**10.05**	**0.51**	**7.24**	**0.35**	**6.30**	**0.33**	**5.52**	**0.31**
			1	**10.68**	**0.53**	**7.87**	**0.39**	**7.18**	**0.38**	**6.48**	**0.38**
			10	**39.70**	**1.42**	39.37	**1.57**	38.77	1.56	35.31	1.48
		50%	0.5	**10.76**	**0.56**	**7.58**	**0.38**	**6.82**	**0.37**	**6.13**	**0.37**
			1	13.20	0.77	**9.13**	**0.51**	**8.54**	0.51	**7.92**	**0.51**
			10	62.88	3.01	62.07	3.54	60.98	3.48	55.48	3.25
Random forest	FGSM	20%	0.5	36.37	1.83	27.73	1.18	22.87	1.00	18.28	0.82
			1	36.03	1.81	27.85	1.21	23.07	1.03	18.50	0.86
			10	40.62	2.02	**34.44**	1.68	31.49	1.52	**25.81**	**1.26**
		50%	0.5	36.24	1.83	27.80	1.20	22.97	1.02	18.36	0.84
			1	36.30	1.83	28.06	1.26	23.34	1.10	18.86	0.94
			10	**40.12**	**2.22**	**44.29**	**2.54**	**41.69**	**2.36**	**34.35**	**1.94**
	BIM	20%	0.5	36.36	1.81	27.67	1.17	22.84	0.99	18.25	0.82
			1	36.05	1.81	27.83	1.19	23.02	1.02	18.45	0.86
			10	41.16	2.22	**35.11**	1.69	**31.50**	**1.52**	**25.72**	**1.26**
		50%	0.5	36.35	1.82	27.89	1.21	23.00	1.02	18.38	0.85
			1	36.62	1.92	28.13	1.27	23.34	1.09	18.84	0.93
			10	**46.54**	2.66	**42.70**	**2.44**	**41.14**	**2.33**	**33.90**	**1.93**
XgBoost	FGSM	20%	0.5	12.86	0.62	9.33	0.47	8.17	0.43	6.98	0.39
			1	12.60	0.60	10.01	0.51	8.93	0.48	7.74	0.44
			10	**22.97**	**1.00**	34.57	**1.46**	**34.62**	**1.43**	**31.18**	1.35
		50%	0.5	15.61	0.82	10.13	0.53	8.87	0.49	7.68	0.44
			1	14.33	0.65	11.06	0.62	9.95	0.58	8.98	0.56
			10	57.17	2.98	55.20	3.18	54.52	3.09	48.78	2.86
	BIM	20%	0.5	14.12	0.66	9.52	0.48	8.31	0.43	7.11	0.39
			1	12.91	0.57	9.87	0.51	8.80	0.47	7.73	0.44
			10	57.17	2.98	55.20	3.18	54.52	3.09	30.96	1.36
		50%	0.5	14.12	0.66	9.52	0.48	8.31	0.43	7.66	0.44
			1	**12.91**	**0.57**	9.87	0.51	8.80	**0.47**	8.74	0.55
			10	54.17	**2.63**	53.80	3.05	53.68	3.04	48.16	2.83

Bold font – Best performance

case of XgBoost, the forecasting accuracy decreases to 344.90% in terms of RMSE and 405.08% in terms of MAPE for its worst-attack case scenario with $\epsilon = 10$, during forecasting for a 24-hour time horizon. Nonetheless, as the forecasting accuracy of each algorithm during adversarial attacks is degraded significantly, none of the algorithms can be considered robust in overall. Hence, the experimental observations presented in Table IV signifies the necessity of developing a robust DLR forecasting scheme, which is capable of defending against adversarial attacks with any attack type, data contamination percentage, and the value of ϵ.

5.2.3. Cyber-attack cases with EAT

To defend against adversarial attacks, an ensemble adversarial training approach is employed, wherein multiple models trained under adversarial conditions are utilized for predicting DLR under attack scenarios and normal conditions. The primary objective is to mitigate the severe impacts of adversarial attacks. For this purpose, this

work employs four distinct models, all trained under adversarial attacked data. These four models share the same architecture as the CatBoost model used for predicting DLR using clean samples. The choice of the CatBoost model is motivated by its superior accuracy in forecasting, both in normal and adversarial attack scenarios, as presented in Table III and Table IV. Each model undergoes training using a specific percentage of clean and attacked data, contributing to a comprehensive defense mechanism against adversarial manipulations. Fig. 4 presents the percentage of data distribution used for training those models. One model is designed to prioritize the predominance of clean data, utilizing 95% clean data and only 5% attacked data. Meanwhile, the remaining three models employ an equal amount of clean and malicious data. A dynamic weighted technique is employed to combine the predictions of the four models.

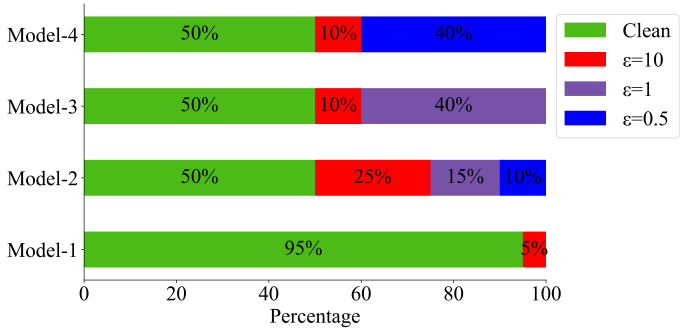

Figure 4: Percentage of clean and attacked data in the training set for EAT.

The performance of the proposed CatBoost model with EAT is presented in Table V and Table VI for attack scenarios and normal conditions, respectively. As presented in Table V, the proposed model forecasts DLR with very low error for any attack type with any percentage of data contamination and $\epsilon = [0.5, 1]$, which is very close to normal performance presented in Table III. Also, the proposed model can significantly reduce errors for DLR forecasting during adversarial attacks with $\epsilon = 10$, which is found to be severe for a normal CatBoost model. For instance, during forecasting for a 24-hour time horizon under FGSM attack with 50% data contamination and $\epsilon = 10$, the forecasting accuracy of 68.62 in terms of RMSE and 3.63 in terms of MAPE can be improved to 24.26 in terms of RMSE and 1.14 in terms of MAPE using the proposed defensive mechanism. A visual presentation of the effectiveness of the proposed model is provided in Fig. 5. As illustrated in Fig. 5, there are barely visible differences between actual ampacity and predicted DLR values using the proposed model under diverse attack scenarios in most cases. For attack cases with $\epsilon = 10$, a small difference with the actual ampacity is noticeable. Hence, as presented in Table V and Fig. 5, the improvements in forecasting accuracy in terms of RMSE and MAPE and the closeness with the actual ampacity indicate the efficacy of the proposed in providing robust DLR forecasting under diverse attack scenarios.

During improving robustness against any cyber-attacks, there is always a trade-off between accuracy at normal conditions and the model's robustness against cyber-attacks. Hence, the performance of the proposed CatBoost with EAT is also evaluated for clean data under normal conditions, which is presented in Table VI. As indicated in Table VI, the forecasting accuracy of the proposed model is almost similar to the performance of a CatBoost model under normal conditions presented in Table III. For the 24-hour time horizon, the proposed model maintains low forecasting errors, which are 10.94 in terms of RMSE and 0.49 in terms of MAPE. These metrics collectively indicate a satisfactory performance in clean samples of the technique over a 24-hour period. As the time horizon extends to 1 month, 3 months, and 6 months, the RMSE and MAPE show a decreasing trend. Notably, the MAPE continues to decline, reaching to 0.30 for the 6-month horizon, indicating a high level of accuracy in predicting clean samples over an extended period. As the proposed scheme trains one of the models in ensembles prioritizing the clean data, it helps to achieve low accuracy in normal conditions. Hence, based on the observations in Table V and Table VI, the proposed CatBoost model with EAT is capable of forecasting DLR with high accuracy under normal conditions, while significantly mitigating the impacts of diverse black-box based adversarial attacks to ensure robust forecasting.

Fig. 6 illustrates the RMSE error metric comparison over a 6-month forecasting horizon, presenting predictions for clean and attacked data using the CatBoost model. Fig. 6 also includes results for the attacked data mitigated by the EAT model, highlighting the impact of attacks and the effectiveness of mitigation. Under a severe attack scenario, where single-model performance drops notably, EAT maintains lower error rates with a minimal RMSE increase, showing its capacity to effectively forecast under manipulated conditions. This resilience is further demonstrated by the balanced

Table V
Performance of the CatBoost model with EAT on adversarially corrupted data.

Time horizon	Contamination percentage	ϵ	FGSM RMSE	FGSM MAPE	BIM RMSE	BIM MAPE
24h	20%	0.5	11.20	0.49	11.47	0.54
		1	11.70	0.54	11.41	0.52
		10	11.23	0.52	17.81	0.78
	50%	0.5	12.84	0.62	12.48	0.58
		1	13.16	0.62	13.76	0.74
		10	24.16	1.14	20.22	1.01
1m	20%	0.5	7.81	0.36	7.82	0.35
		1	8.56	0.41	8.30	0.40
		10	18.21	0.72	16.68	0.67
	50%	0.5	7.95	0.36	7.94	0.37
		1	9.34	0.51	9.18	0.51
		10	26.26	1.20	24.46	1.09
3m	20%	0.5	6.39	0.32	6.40	0.31
		1	7.29	0.38	7.16	0.37
		10	15.69	0.61	15.14	0.59
	50%	0.5	6.61	0.33	6.58	0.33
		1	8.34	0.48	8.14	0.47
		10	23.41	1.03	22.84	1.00
6m	20%	0.5	5.61	0.31	5.63	0.31
		1	6.44	0.37	6.40	0.36
		10	12.04	0.49	11.70	0.48
	50%	0.5	5.85	0.32	5.78	0.32
		1	7.46	0.47	7.34	0.46
		10	17.79	0.76	17.27	0.73

Table VI
Performance of the CatBoost model with EAT under normal conditions.

Time Horizon	RMSE	MAPE
24h	10.94	0.49
1m	7.69	0.34
3m	6.29	0.30
6m	5.55	0.30

performance of EAT across different attack intensities. The dynamic weighted technique combines the strengths of each ensemble member, allowing EAT to adjust its response based on each model's specialized robustness to specific attacks. This combination leads to a more accurate, attack-resistant forecast in DLR predictions, underscoring EAT's practical advantage for real-world applications.

5.2.4. Comparison with related works

The recent surveys [24, 57] reveal a lack of sufficient research studies investigating adversarial attacks and their impact on DLR forecasting, resulting in few research studies in this area. A comparison between this work and the existing literature on the cyber security of DLR forecasting is summarized in Table VII, highlighting the main contributions, forecasting horizons, implemented attacks, and defense approaches. As presented in Table VII, the existing literature is limited to scaling-based FDI attack consideration for DLR forecasting, which is emphasized in [24]. As DL-based approaches can be susceptible to adversarial attacks, the novelty of this work is signified in Table VII because of considering two different adversarial attacks. From the defense perspective, the existing literature either relies on pre-processing schemes or addresses the inherent resiliency of their considered learning schemes, which may

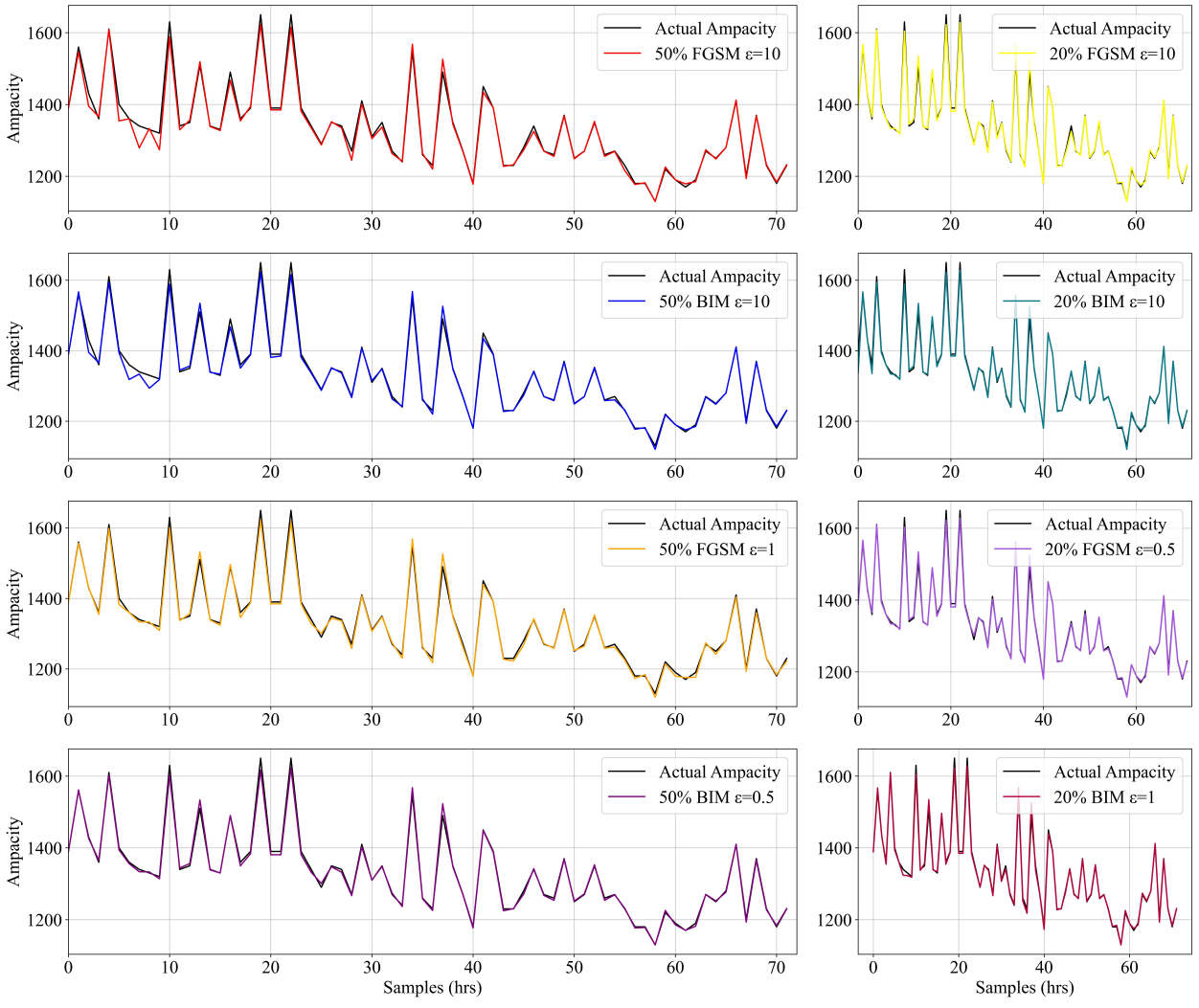

Figure 5: Comparison of the actual ampacity and the forecasted data using CatBoost with EAT for 72 hours under diverse attack scenarios.

keep the forecasting model still vulnerable during well-designed attack schemes like adversarial attacks. However, as this work addresses an EAT-based defensive scheme, the forecasting model becomes more robust in defending against such attacks. Furthermore, as only this work discusses multi-horizon forecasting while presenting a defensive measure against cyber-attacks, this work can be a stepping stone for the further development of robust and accurate forecasting schemes.

6. Conclusions

This paper delves into the DLR forecasting of OHL, while investigating their vulnerability to cyber-attacks. Ensemble learning techniques are employed to forecast over multi-horizon, where a comparative analysis of multiple models under normal conditions and diverse attack scenarios is performed to identify an optimum model. Additionally, ensemble adversarial training (EAT) is proposed to improve the robustness of the optimum model. The experimental observations can be summarized in the following points.

- The employment of ensemble techniques results in low forecasting error, while allowing forecasting for 24 hours, 1 month, 3 months, and 6 months. The forecasting accuracy improves with the increase in time horizon, where the CatBoost outperforms XgBoost and random forest models under normal conditions. The CatBoost-based DLR

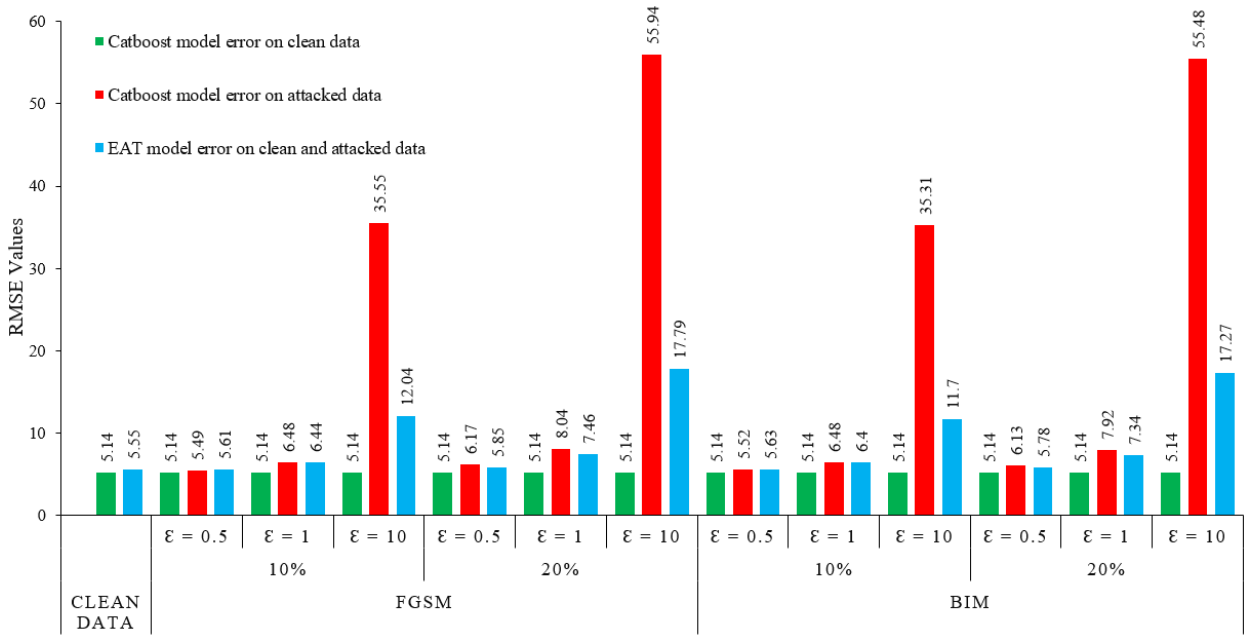

Figure 6: Comparison of RMSE errors between clean and attacked data on the CatBoost model as well as the EAT model for 6-month horizon forecasting.

Table VII
Comparison with related research works on cyber security of DLR forecasting schemes

Reference	Key contribution	Horizon	Attack	Defence
[58]	Presenting CNN-based federated learning (FL) approach for cyber-resilient DLR forecasting	Single	FDI: scaling	FL's inherent data privacy preserving
[59]	Presenting LSTM-based DLR forecasting and an image processing-based data contamination detector	Single	FDI: scaling	Image processing for attack detection
[29]	Presenting a hybrid deep learning approach by combining autoencoder and bidirectional LSTM models for robust and accurate DLR forecasting	Single	FDI: scaling	AE for data reconstruction
[13]	Investigating FDI attacks on ensemble learning-based DLR forecasting models	Multiple	FDI: scaling	×
This work	Investigating adversarial attacks on ensemble learning-based DLR forecasting models and presenting EAT (ensemble adversarial training)-based robust forecasting schemes	Multiple	Adversarial attacks: FGSM, BIM	EAT for model robustification

forecasting achieves a 41.4% increase in the ampacity compared to the ampacity selected by SLR, highlighting its effectiveness in DLR selection.

- The introduction of adversarial attacks in DLR forecasting signifies the vulnerability of DLR forecasting models to FGSM and BIM-based black box attacks, where increases in data contamination percentage and perturbations increase the severity of attacks, resulting in notable degradations in forecasting accuracy for all models. Except for the worst attack cases with $\epsilon = 10$, CatBoost performs comparatively better than others in overall. However, this optimum model's accuracy decreases up to 601.64% in terms of RMSE and 656.25% in terms of MAPE

for the worst attack case during the smallest time horizon forecasting, which may worsen with the time horizon, signifying the attacks' severeness.

- The utilization of a variable combination of clean and corrupted data for the proposed EAT scheme allows for retaining accuracy under normal conditions, resulting in a forecasting accuracy of 0.49 in terms of MAPE for a 24-hour time horizon, which can be reduced to 0.30 for 6 months. Simultaneously, the proposed scheme achieves almost similar accuracy for adversarial attacks with $\epsilon = [0.5, 1]$, resulting in a barely visible difference between the actual ampacity and the ampacity selected by the proposed CatBoost with EAT. Even for attack cases with $\epsilon = 10$, the forecasting accuracy is significantly improved, indicating the effectiveness of the proposed scheme for robust and accurate multi-horizon DLR forecasting.

In the future, this work can be extended to simultaneously consider physical and cyber-oriented data anomalies to develop a robust DLR forecasting method. Additionally, other types of cyber-attacks can be considered to evaluate their impacts on both point and probabilistic method-based DLR forecasting while developing an attack-resistant scheme against them.

CRediT authorship contribution statement

Najmul Alam: Conceptualization; Data curation; Investigation; Methodology; Resources; Writing - Original draft. **M. A. Rahman:** Conceptualization; Methodology; Resources; Writing - Original draft. **Md. Rashidul Islam:** Conceptualization; Supervision; Writing - Review & Editing. **M. J. Hossain:** Supervision, Writing - review & editing.

References

[1] M. R. Islam, J. Hasan, M. R. Islam, A. Z. Kouzani, M. A. P. Mahmud, Transient performance augmentation of dfig based wind farms by nonlinear control of flux-coupling-type superconducting fault current limiter, IEEE Transactions on Applied Superconductivity 31 (8) (2021) 1–5. doi:10.1109/TASC.2021.3091061.

[2] F. Teng, R. Dupin, A. Michiorri, G. Kariniotakis, Y. Chen, G. Strbac, Understanding the benefits of dynamic line rating under multiple sources of uncertainty, IEEE Transactions on Power Systems 33 (3) (2018) 3306–3314. doi:10.1109/TPWRS.2017.2786470.

[3] A. K. Rao, P. Kundu, System integrity protection scheme for minimizing wind curtailment considering transmission line thermal limits, Sustainable Energy, Grids and Networks 33 (2023) 100970. doi:10.1016/j.segan.2022.100970.

[4] X. He, J. Teh, Impacts of dynamic thermal rating systems on wind generations: A review, Electric Power Systems Research 233 (2024) 110492. doi:10.1016/j.epsr.2024.110492.

[5] O. A. Lawal, J. Teh, A framework for modelling the reliability of dynamic line rating operations in a cyber–physical power system network, Sustainable Energy, Grids and Networks 35 (2023) 101140. doi:10.1016/j.segan.2023.101140.

[6] Z. Gao, S. Hu, H. Sun, Z. Wang, S. Liu, F. Yang, Day-ahead dynamic thermal line rating forecasting and power transmission capacity calculation based on forecastnet, Electric Power Systems Research 220 (2023) 109350. doi:10.1016/j.epsr.2023.109350.

[7] A. Pepiciello, G. Coletta, A. Vaccaro, D. Villacci, The role of learning techniques in synchrophasor-based dynamic thermal rating, International Journal of Electrical Power & Energy Systems 115 (2020) 105435. doi:10.1016/j.ijepes.2019.105435.

[8] O. A. Lawal, J. Teh, Assessment of dynamic line rating forecasting methods, Electric Power Systems Research 214 (2023) 108807. doi:10.1016/j.epsr.2022.108807.

[9] R. Dupin, G. Kariniotakis, A. Michiorri, Overhead lines dynamic line rating based on probabilistic day-ahead forecasting and risk assessment, International Journal of Electrical Power Energy Systems 110 (2019) 565–578. doi:10.1016/j.ijepes.2019.03.043.

[10] J. Heckenbergerová, P. Musilek, K. Filimonenkov, Quantification of gains and risks of static thermal rating based on typical meteorological year, International Journal of Electrical Power & Energy Systems 44 (1) (2013) 227–235. doi:10.1016/j.ijepes.2012.07.005.

[11] J. L. Aznarte, N. Siebert, Dynamic line rating using numerical weather predictions and machine learning: A case study, IEEE Transactions on Power Delivery 32 (1) (2017) 335–343. doi:10.1109/TPWRD.2016.2543818.

[12] A. Kirilenko, M. Esmaili, C. Y. Chung, Risk-averse stochastic dynamic line rating models, IEEE Transactions on Power Systems 36 (4) (2021) 3070–3079. doi:10.1109/TPWRS.2020.3045589.

[13] A. Ahmadi, M. Nabipour, B. Mohammadi-Ivatloo, V. Vahidinasab, Ensemble learning-based dynamic line rating forecasting under cyberattacks, IEEE Transactions on Power Delivery 37 (1) (2022) 230–238. doi:10.1109/TPWRD.2021.3056055.

[14] A. Ahmadi, M. Nabipour, B. Mohammadi-Ivatloo, A. M. Amani, S. Rho, M. J. Piran, Long-term wind power forecasting using tree-based learning algorithms, IEEE Access 8 (2020) 151511–151522. doi:10.1109/ACCESS.2020.3017442.

[15] B. Jimada-Ojuolape, J. Teh, Surveys on the reliability impacts of power system cyber–physical layers, Sustainable Cities and Society 62 (2020) 102384. doi:10.1016/j.scs.2020.102384.

[16] P. Eder-Neuhauser, T. Zseby, J. Fabini, G. Vormayr, Cyber attack models for smart grid environments, Sustainable Energy, Grids and Networks 12 (2017) 10–29. doi:10.1016/j.segan.2017.08.002.

[17] M. A. Rahman, M. S. Rana, H. R. Pota, Mitigation of frequency and voltage disruptions in smart grid during cyber-attack, Journal of Control, Automation and Electrical Systems 31 (2020) 412–421. doi:10.1007/s40313-020-00574-z.

[18] A. Devnath, M. A. Rahman, M. S. Rana, Impact analysis of cyber–attack on mmc–hvdc control system with countermeasures, International Journal of Dynamics and Control (2023) 1–11doi:10.1007/s40435-023-01313-3.

[19] O. A. Lawal, J. Teh, B. Alharbi, C.-M. Lai, Data-driven learning-based classification model for mitigating false data injection attacks on dynamic line rating systems, Sustainable Energy, Grids and Networks (2024) 101347doi:10.1016/j.segan.2024.101347.

[20] B. Jimada-Ojuolape, J. Teh, C.-M. Lai, Securing the grid: A comprehensive analysis of cybersecurity challenges in pmu-based cyber-physical power networks, Electric Power Systems Research 233 (2024) 110509. doi:10.1016/j.epsr.2024.110509.

[21] B. Jimada-Ojuolape, J. Teh, Composite reliability impacts of synchrophasor-based dtr and sips cyber–physical systems, IEEE Systems Journal 16 (3) (2022) 3927–3938. doi:10.1109/JSYST.2021.3132657.

[22] B. Jimada-Ojuolape, J. Teh, Impacts of communication network availability on synchrophasor-based dtr and sips reliability, IEEE Systems Journal 16 (4) (2022) 6231–6242. doi:10.1109/JSYST.2021.3122022.

[23] O. A. Lawal, J. Teh, Dynamic line rating forecasting algorithm for a secure power system network, Expert Systems with Applications 219 (2023) 119635. doi:10.1016/j.eswa.2023.119635.

[24] M. A. Rahman, M. R. Islam, M. A. Hossain, M. S. Rana, M. J. Hossain, E. M. Gray, Resiliency of forecasting methods in different application areas of smart grids: A review and future prospects, Engineering Applications of Artificial Intelligence 135 (2024) 108785. doi:10.1016/j.engappai.2024.108785.

[25] J. Luo, T. Hong, S.-C. Fang, Benchmarking robustness of load forecasting models under data integrity attacks, International Journal of Forecasting 34 (1) (2018) 89–104. doi:10.1016/j.ijforecast.2017.08.004.

[26] A. Ahmadi, M. Nabipour, S. Taheri, B. Mohammadi-Ivatloo, V. Vahidinasab, A new false data injection attack detection model for cyberattack resilient energy forecasting, IEEE Transactions on Industrial Informatics 19 (1) (2023) 371–381. doi:10.1109/TII.2022.3151748.

[27] A. S. Tusher, M. A. Rahman, M. R. Islam, M. J. Hossain, Adversarial training-based robust lifetime prediction system for power transformers, Electric Power Systems Research 231 (2024) 110351. doi:10.1016/j.epsr.2024.110351.

[28] A. Moradzadeh, H. Moayyed, B. Mohammadi-Ivatloo, A. Anvari-Moghaddam, Z. Vale, R. Ghorbani, Image Processing-based Data Integrity Attack Detection in Dynamic Line Rating Forecasting Applications, in: 2022 10th International Conference on Smart Grid (icSmartGrid), 2022, pp. 249–254. doi:10.1109/icSmartGrid55722.2022.9848657.

[29] A. Moradzadeh, M. Mohammadpourfard, I. Genc, Ş. S. Şeker, B. Mohammadi-Ivatloo, Deep learning-based cyber resilient dynamic line rating forecasting, International Journal of Electrical Power & Energy Systems 142 (2022) 108257. doi:10.1016/j.ijepes.2022.108257.

[30] N. Tang, S. Mao, R. M. Nelms, Adversarial attacks to solar power forecast, in: 2021 IEEE Global Communications Conference (GLOBECOM), 2021, pp. 1–6. doi:10.1109/GLOBECOM46510.2021.9685910.

[31] Y. Zhou, Z. Ding, Q. Wen, Y. Wang, Robust load forecasting towards adversarial attacks via bayesian learning, IEEE Transactions on Power Systems 38 (2) (2023) 1445–1459. doi:10.1109/TPWRS.2022.3175252.

[32] M. Davis, A new thermal rating approach: The real time thermal rating system for strategic overhead conductor transmission lines – part i: General description and justification of the real time thermal rating system, IEEE Transactions on Power Apparatus and Systems 96 (3) (1977) 803–809. doi:10.1109/T-PAS.1977.32393.

[33] J. Teh, I. Cotton, Reliability impact of dynamic thermal rating system in wind power integrated network, IEEE Transactions on Reliability 65 (2) (2016) 1081–1089. doi:10.1109/TR.2015.2495173.

[34] D. Douglass, A.-A. Edris, Real-time monitoring and dynamic thermal rating of power transmission circuits, IEEE Transactions on Power Delivery 11 (3) (1996) 1407–1418. doi:10.1109/61.517499.

[35] T. Barton, P. Musilek, Day-ahead dynamic thermal line rating using numerical weather prediction, in: 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), 2019, pp. 1–7. doi:10.1109/CCECE.2019.8861883.

[36] J. Teh, C.-M. Lai, N. A. Muhamad, C. A. Ooi, Y.-H. Cheng, M. A. A. Mohd Zainuri, M. K. Ishak, Prospects of using the dynamic thermal rating system for reliable electrical networks: A review, IEEE Access 6 (2018) 26765–26778. doi:10.1109/ACCESS.2018.2824238.

[37] A. W. Abboud, J. P. Gentle, E. E. Bukowski, M. J. Culler, J. P. Meng, S. Morash, A guide to case studies of grid enhancing technologies, Tech. rep., Idaho National Lab. (INL), Idaho Falls, ID, United States (2022). doi:10.2172/1957788.

[38] Ieee standard for calculating the current-temperature relationship of bare overhead conductors, IEEE Std 738-2012 (Revision of IEEE Std 738-2006 - Incorporates IEEE Std 738-2012 Cor 1-2013) (2013) 1–72doi:10.1109/IEEESTD.2013.6692858.

[39] J. Iglesias, G. Watt, D. Douglass, V. Morgan, R. Stephen, M. Bertinat, D. Muftic, R. Puffer, D. Guery, S. Ueda, et al., Guide for Thermal Rating Calculations of Overhead Lines, CIGRE, 2014.

[40] R. Alberdi, E. Fernandez, I. Albizu, M. T. Bedialauneta, R. Fernandez, Overhead line ampacity forecasting and a methodology for assessing risk and line capacity utilization, International Journal of Electrical Power & Energy Systems 133 (2021) 107305. doi:10.1016/j.ijepes.2021.107305.

[41] M. O. Faruque, M. A. J. Rabby, M. A. Hossain, M. R. Islam, M. M. U. Rashid, S. M. Muyeen, A comparative analysis to forecast carbon dioxide emissions, Energy Reports 8 (2022) 8046–8060. doi:10.1016/j.egyr.2022.06.025.

[42] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.

[43] J. T. Hancock, T. M. Khoshgoftaar, Catboost for big data: an interdisciplinary review, Journal of big data 7 (1) (2020) 1–45. doi:10.1186/s40537-020-00369-8.

[44] L. Breiman, Random forests, Machine learning 45 (2001) 5–32. doi:10.1023/A:1010933404324.

[45] X. Yuan, P. He, Q. Zhu, X. Li, Adversarial examples: Attacks and defenses for deep learning, IEEE Transactions on Neural Networks and Learning Systems 30 (9) (2019) 2805–2824. doi:10.1109/TNNLS.2018.2886017.

[46] I. J. Goodfellow, J. Shlens, C. Szegedy, Explaining and harnessing adversarial examples, arXiv preprint arXiv:1412.6572 (2014). doi:10.48550/arXiv.1412.6572.

[47] F. Tramèr, A. Kurakin, N. Papernot, I. Goodfellow, D. Boneh, P. McDaniel, Ensemble adversarial training: Attacks and defenses (2020). arXiv:1705.07204, doi:10.48550/arXiv.1705.07204.

[48] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, P. N. Suganthan, Ensemble deep learning: A review, Engineering Applications of Artificial Intelligence 115 (2022) 105151. doi:10.1016/j.engappai.2022.105151.

[49] Q. Yan, Z. Lu, H. Liu, X. He, X. Zhang, J. Guo, Short-term prediction of integrated energy load aggregation using a bi-directional simple recurrent unit network with feature-temporal attention mechanism ensemble learning model, Applied Energy 355 (2024) 122159. doi:10.1016/j.apenergy.2023.122159.

[50] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, R. Fergus, Intriguing properties of neural networks, arXiv e-prints (2013) arXiv:1312.6199arXiv:1312.6199, doi:10.48550/arXiv.1312.6199.

[51] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Adversarial attacks on deep neural networks for time series classification, in: 2019 International Joint Conference on Neural Networks (IJCNN), 2019, pp. 1–8. doi:10.1109/IJCNN.2019.8851936.

[52] A. Kurakin, I. J. Goodfellow, S. Bengio, Adversarial examples in the physical world, in: Artificial intelligence safety and security, Chapman and Hall/CRC, 2018, pp. 99–112.

[53] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, A. Swami, Practical black-box attacks against machine learning, in: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, 2017, pp. 506–519. doi:10.1145/3052973.3053009.

[54] X. T. Nguyen, T. D. Nguyen, Dynamic line rating solution: Deployment opportunities for the power transmission grid of vietnam, International Journal of Energy and Power Engineering 11 (2) (2022) 56–67. doi:10.11648/j.ijepe.20221102.15.

[55] A. Pavlinić, V. Komen, M. Uzelac, Application of direct collocation method in short-term line ampacity calculation, Electric Power Systems Research 155 (2018) 216–224. doi:10.1016/j.epsr.2017.10.018.

[56] F. Muñoz, F. Torres, S. Martínez, C. Roa, L. García, Case study of the increase in capacity of transmission lines in the chilean system through probabilistic calculation model based on dynamic thermal rating, Electric Power Systems Research 170 (2019) 35–47. doi:10.1016/j.epsr.2019.01.008.

[57] O. A. Lawal, J. Teh, Dynamic thermal rating forecasting methods: A systematic survey, IEEE Access 10 (2022) 65193–65205. doi:10.1109/ACCESS.2022.3183606.

[58] H. Moayyed, A. Moradzadeh, A. Mansour-Saatloo, B. Mohammadi-Ivatloo, M. Abapour, Z. Vale, A global cyber-resilient model for dynamic line rating forecasting based on deep federated learning, IEEE Systems Journal 17 (4) (2023) 6390–6400. doi:10.1109/JSYST.2023.3287413.

[59] A. Moradzadeh, H. Moayyed, B. Mohammadi-Ivatloo, A. Anvari-Moghaddam, Z. Vale, R. Ghorbani, Image processing-based data integrity attack detection in dynamic line rating forecasting applications, in: 2022 10th International Conference on Smart Grid (icSmartGrid), 2022, pp. 249–254. doi:10.1109/icSmartGrid55722.2022.9848657.