

Physics-Informed Forecasting of Short Circuit Current Waveforms in LVDC Microgrids

Daniel-Cătălin Mitroi

Automation and Industrial Informatics
University Politehnica of Bucharest
Bucharest, Romania
daniel.mitroi1806@upb.ro

Iulia Stamatescu

Automation and Industrial Informatics
University Politehnica of Bucharest
Bucharest, Romania
iulia.stamatescu@upb.ro

Cristina Nichiforov

Automation and Industrial Informatics
University Politehnica of Bucharest
Bucharest, Romania
cristina.nichiforov@upb.ro

Grigore Stamatescu

Automation and Industrial Informatics
University Politehnica of Bucharest
Bucharest, Romania
grigore.stamatescu@upb.ro

Abstract—Fast short circuit transients in low-voltage DC (LVDC) microgrids challenge conventional threshold-based protection strategies. This paper presents a physics-guided, feature-based framework for multi-step forecasting of post-fault short circuit current trajectories from early transient measurements. A publicly available parametric dataset generated through Typhoon HIL and MATLAB/Simulink co-simulation was segmented into 10,906 instances, each comprising a 200-sample (400 μ s) observation window and a 500-sample (1 ms) prediction horizon. Seven physically meaningful electrical descriptors were used as inputs to Random Forest, XGBoost, and 1D-CNN models. Generalization was evaluated using five-fold GroupKFold cross-validation at the simulation-file level, preventing windows from the same simulation case from appearing in both training and evaluation folds. Random Forest achieved the best average performance, with an RMSE of 9.14 A, an MAE of 3.98 A, and an R^2 of 0.981. These results support the feasibility of waveform forecasting on simulation-derived fault data, while end-to-end latency and robustness to measurement noise require further experimental validation.

Index Terms—data-driven modelling, DC microgrids, machine learning, physics-guided features, short circuit current forecasting.

I. INTRODUCTION

The accelerated integration of renewable energy sources, battery storage systems, electric vehicle charging infrastructures, and modern power electronic loads has significantly increased the interest in low-voltage direct current (LVDC) microgrids. Compared to traditional AC distribution systems, LVDC architectures offer several advantages, including reduced conversion stages, improved efficiency, simplified integration of photovoltaic systems and batteries, lower distribution losses, and enhanced controllability of distributed energy [1].

This work has been partially supported by the European Union's Horizon Europe research and innovation programme under grant agreement No. 101192615 (NOVETROL)

Despite these advantages, protection remains one of the major technical challenges limiting the large-scale deployment of LVDC systems. In DC systems, fault currents rise extremely rapidly due to the low impedance of the network and the discharge of DC-link capacitors, often reaching dangerous levels within microseconds to milliseconds. Unlike AC grids, DC fault currents lack natural zero-crossings, making conventional protection schemes based on overcurrent thresholds or distance relays less effective, especially under high-impedance faults or converter-limited conditions [2].

Traditional overcurrent protection methods based on fixed thresholds or steady-state measurements often prove inadequate in modern LVDC microgrids, especially under high-impedance faults, converter-limited currents, or highly dynamic transients. In many practical situations, power electronic converters actively limit the fault current, reducing the effectiveness of conventional protection strategies that rely exclusively on current magnitude [3]. As a result, recent research has increasingly focused on data-driven fault detection approaches based on machine learning techniques [4].

Recent studies have demonstrated the potential of machine learning and deep learning techniques for fault detection, classification, and localization in DC microgrids. Data-driven approaches based on support vector machines, ensemble learning algorithms, convolutional neural networks, and transfer learning strategies have shown promising results for extracting transient fault patterns from electrical measurements [5]. In parallel, several recent works have investigated transient waveform analysis using CNN-based architectures and signal decomposition methods for high-speed DC fault discrimination [6].

More recently, increasing attention has been directed toward incorporating physical knowledge into data-driven models for intelligent power system applications. Physics-informed machine learning typically integrates governing equations, initial or boundary conditions, or other physical constraints into

the learning process [7]. The present study adopts a feature-based approach in which domain knowledge is introduced through electrically meaningful transient descriptors, including current slopes, voltage-collapse indicators, current-magnitude and current-integral measures, and statistical waveform characteristics. Accordingly, the proposed methodology is more precisely characterized as physics-guided feature engineering. In this paper, the term physics-informed forecasting is used in this broader, feature-based sense and does not refer to a physics-informed neural network or a physics-constrained learning formulation.

Most existing methods focus on binary fault detection or location, with limited attention to predicting the future evolution of the fault current waveform itself. However, the capability to forecast the future evolution of the fault current trajectory may support future research on predictive protection and adaptive converter control.

To address these challenges, the present paper proposes a physics-guided multi-step forecasting framework for transient short circuit current waveforms in low-voltage DC microgrids. The proposed methodology combines sliding-window transient segmentation with physics-guided feature extraction and supervised regression models in order to predict the post-fault evolution of the short circuit current waveform over a finite forecasting horizon. Instead of supplying raw sampled signals directly to the forecasting models, the proposed framework relies on a reduced set of transient descriptors specifically designed to capture both magnitude-related and dynamic properties associated with DC fault propagation.

This study builds on the publicly available parametric dataset introduced in [8] and the short circuit measurement and co-simulation framework presented in [9]. While these prior contributions focused on dataset generation and transient-current characterization, the present work formulates and evaluates post-fault current evolution as a direct multi-step forecasting problem. The main contribution lies in this forecasting formulation, the compact physics-guided feature representation, and the leakage-aware GroupKFold evaluation across held-out simulation files, rather than in the development of new regression algorithms.

The main contributions of this work can be summarized as follows:

- The design of a physics-guided forecasting pipeline combining sliding-window segmentation with transient feature extraction;
- The comparative evaluation of Random Forest (RF), XGBoost and 1D-CNN models for multi-step short circuit current waveform forecasting;
- The validation of model generalization using leakage-aware GroupKFold cross-validation, together with a Random Forest feature-importance analysis to support model interpretability.

The rest of the paper is structured as follows. Section II describes the dataset and the scenarios analyzed. Section III presents the proposed methodology, including sliding window

segmentation, physics-guided feature extraction, and the forecasting models used. Section IV presents the experimental results and comparative performance analysis, while Section V concludes the paper and outlines future research directions.

II. DATASET DESCRIPTION

The proposed approach leverages our recently published “Parametric Short Circuit Dataset for Fault Analysis and Protection of DC Microgrids” dataset, available through IEEE DataPort [8]. The dataset was specifically developed for the analysis of transient short circuit phenomena in low-voltage DC (LVDC) residential microgrids and contains high-resolution time-domain electrical measurements obtained from co-simulation environments implemented in Typhoon HIL and MATLAB/Simulink.

The database is organized around three representative residential high-level use cases: *Cottage in the Sun* – photovoltaic-based residential LVDC microgrids, *Safe Storage at Home* – residential battery energy storage systems, and *Safe DC at Home* – residential DC distribution networks supplying household loads.

For each use case, several DC busbar voltage levels were considered, namely 48 VDC, 220 VDC, and 350 VDC, depending on the specific operating architecture. Multiple parametric operating scenarios were generated by varying quantities such as photovoltaic nominal current, load power, battery model type, and short circuit resistance.

Each scenario contains synchronized transient recordings captured during controlled short circuit events. Depending on the operating configuration, the recorded quantities include the DC busbar voltage (V_{out}), short circuit current (I_{sc}), photovoltaic source voltage (V_{pv1}), source-side currents, battery currents, capacitor currents, and additional load-specific electrical variables.

The transient current responses presented in Figure 1 correspond to the short circuit current (I_{sc}) measured in the fault branch. Immediately after fault inception, the pronounced initial (I_{sc}) peak is dominated by the discharge current (I_c) of the DC-link capacitor. The peak magnitude increases with the DC bus voltage level, reaching approximately 16 A at 48 V and 28 A at 350 V. Following the voltage collapse, converter protection mechanisms disconnect the photovoltaic source, causing the residual current to rapidly decay toward zero.

The acquisition process was performed using the Typhoon HIL Control Center capture functionality with a sampling rate of 500 kS/s, corresponding to a temporal resolution of 2 μ s over a total acquisition interval of 0.2 s. This high temporal resolution enables accurate representation of the fast transient dynamics associated with DC faults, including steep current rise rates and rapid voltage collapses. A more detailed analysis of transient short circuit current measurement and co-simulation methodologies in LVDC systems is presented in [9].

Figures 1–3 illustrate representative I_{sc} trajectories across the three residential use cases, highlighting variations in peak

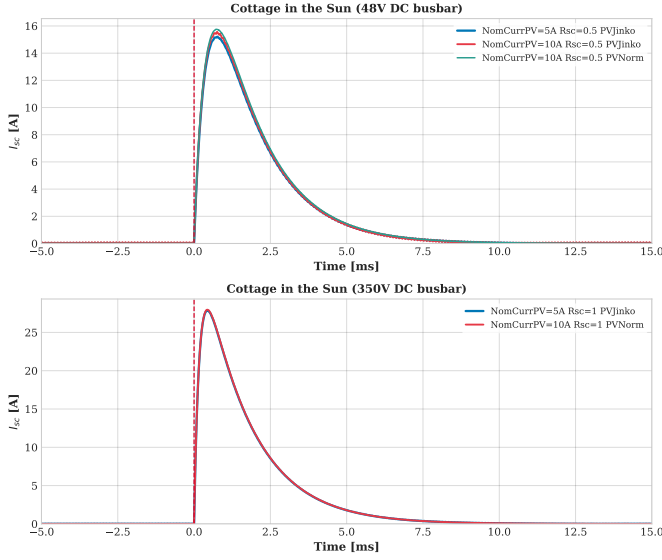


Fig. 1: Transient short circuit current overlays for the Cottage in the Sun residential LVDC use case under different photovoltaic operating scenarios and DC busbar voltage levels.

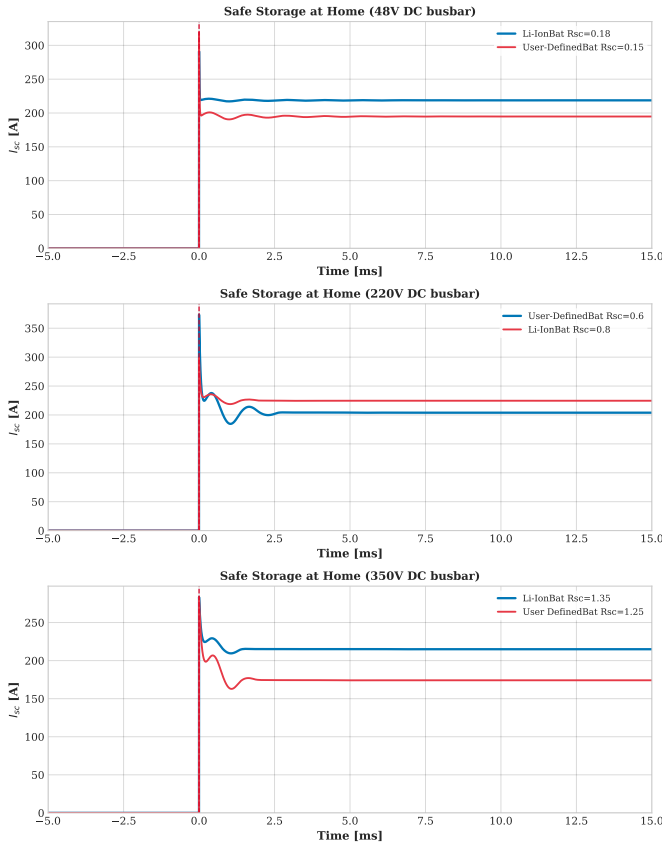


Fig. 2: Transient short circuit current overlays for the Safe Storage at Home residential battery storage use case for multiple battery models and DC busbar voltage levels.

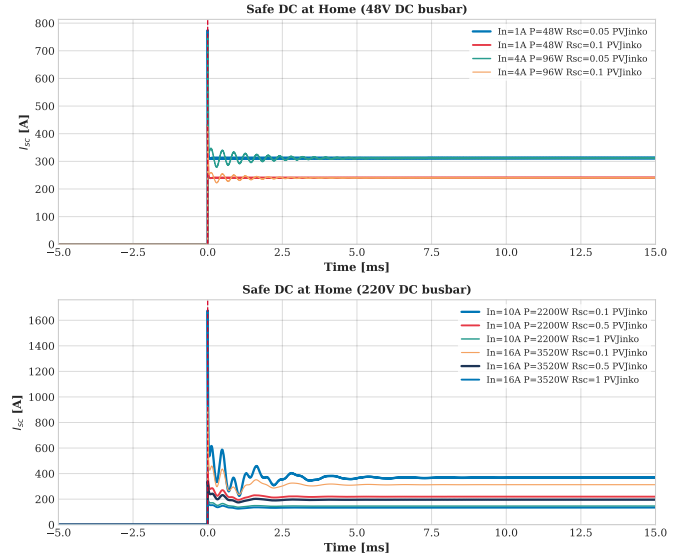


Fig. 3: Transient short circuit current overlays for the Safe DC at Home residential DC distribution use case under different load and fault resistance configurations.

magnitude and transient behavior with DC-bus voltage, source configuration, and fault resistance.

Only post-fault observation windows aligned with the known fault-inception instant ($t=0$) are used in this study; no separate non-fault or benign-transient events are included. Accordingly, the framework forecasts current evolution after fault inception rather than performing fault detection or fault-versus-non-fault discrimination. The results represent a simulation-based proof of concept, while robustness to measurement noise and hardware non-idealities remains to be experimentally validated.

III. METHODOLOGY

A. Overall Forecasting Framework

The proposed methodology aims to forecast the future evolution of short circuit current waveforms during the sub-transient stage immediately following fault inception in residential LVDC microgrids. Unlike conventional fault detection approaches that focus on identifying the occurrence of a fault, the proposed framework attempts to predict the subsequent fault-current trajectory using information extracted from the earliest transient response.

The overall processing pipeline is illustrated in Figure 4. The workflow consists of five sequential stages. Transient short circuit waveforms obtained from the public IEEE DataPort dataset are first processed using a sliding-window segmentation strategy in order to generate multiple forecasting samples from each simulation case. Each observation window is subsequently transformed into a compact physics-guided feature representation designed to capture the dominant electrical dynamics associated with DC fault propagation, including rapid current rise rates, capacitor discharge effects, and voltage collapse phenomena.

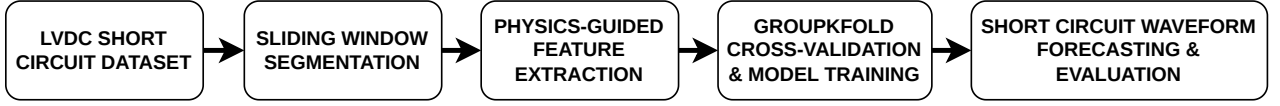


Fig. 4: Overall methodology pipeline

The extracted feature vectors are then supplied to three forecasting models, namely Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and a one-dimensional Convolutional Neural Network (1D-CNN). Model generalization capability is assessed through GroupKFold cross-validation, where all forecasting samples generated from a given simulation file are assigned to the same validation fold. This strategy prevents information leakage between training and testing subsets and provides a more realistic estimation of model performance under unseen operating conditions. The forecasting accuracy is quantified using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2).

B. Physics-Guided Feature Extraction

To transform the raw transient waveforms into a compact and interpretable representation, a fixed-length sliding-window segmentation strategy is applied to the fault recordings. Considering the dataset sampling frequency of 500 kS/s, corresponding to a temporal resolution of 2 μ s, each observation window contains 200 samples, representing approximately 400 μ s of transient behavior.

Each 200-sample observation window is transformed into a compact seven-dimensional physics-guided feature vector. The selected descriptors capture complementary aspects of the early fault transient, including current magnitude, current rate of change, accumulated current, voltage collapse, waveform impulsiveness, and statistical dispersion [7].

For every observation window, the subsequent 500 samples of the short circuit current waveform are selected as the forecasting target. Consequently, each training instance consists of an input feature vector extracted from the observation interval and a corresponding future waveform segment spanning the prediction horizon.

The extracted feature set includes seven representative transient indicators: peak short circuit current ($I_{sc,peak}$), maximum absolute current derivative (di/dt_{max}), DC bus voltage drop (V_{drop}), transient current integral (Q_{abs}), RMS transient current (I_{rms}), crest factor (C_f), and current standard deviation ($\sigma_{I_{sc}}$).

- **Peak Current ($I_{sc,peak}$):** captures the maximum instantaneous current amplitude observed during the transient fault interval.

$$I_{sc_peak} = \max_{k \in [1, N]} |I_{sc}(k)| \quad (1)$$

- **Maximum Absolute Current Derivative (di/dt_{max}):** quantifies the maximum transient current rising slope

generated during the initial capacitor discharge stage of the fault.

$$di/dt_{max} = \max_k \left| \frac{I_{sc}(k) - I_{sc}(k-1)}{\Delta T} \right| \quad (2)$$

- **Voltage Drop (V_{drop}):** characterizes the voltage excursion within the post-fault observation window relative to its maximum observed value.

$$V_{drop} = \frac{V_{ref} - \min(V_{out})}{V_{ref} + \epsilon}, \quad (3)$$

where V_{ref} denotes the maximum DC-bus voltage observed within the post-fault observation window, and $\epsilon = 10^{-8}$ is a small positive constant introduced to avoid numerical division by zero.

- **Transient Current Integral (Q_{abs}):** quantifies the accumulated absolute current over the observation window.

$$Q_{abs} = \int_0^{T_w} |I_{sc}(t)| dt \quad (4)$$

$$\approx \frac{\Delta T}{2} \sum_{k=1}^{N-1} (|I_{sc}(k)| + |I_{sc}(k+1)|)$$

- **Root-Mean-Square Current (I_{rms}):** characterizes the effective magnitude of the transient current waveform.

$$I_{rms} = \sqrt{\frac{1}{N} \sum_{k=1}^N I_{sc}(k)^2} \quad (5)$$

- **Crest Factor (C_f):** evaluates the impulsive behavior of the transient waveform through the ratio between peak and RMS current.

$$C_f = \frac{I_{sc_peak}}{I_{rms}} \quad (6)$$

- **Current Standard Deviation ($\sigma_{I_{sc}}$):** quantifies the statistical dispersion of the transient current samples within the analysis window.

$$\sigma_{I_{sc}} = \sqrt{\frac{1}{N} \sum_{k=1}^N (I_{sc}(k) - \bar{I}_{sc})^2}, \quad (7)$$

where

$$\bar{I}_{sc} = \frac{1}{N} \sum_{k=1}^N I_{sc}(k). \quad (8)$$

C. Forecasting Models and Validation Strategy

Three forecasting models were investigated in this study: Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and a one-dimensional Convolutional Neural Network (1D-CNN). These models were selected due to their complementary learning mechanisms and their widespread adoption in intelligent power-system monitoring applications.

The **Random Forest (RF)** model was implemented using the *MultiOutputRegressor* strategy, which trains an independent Random Forest regressor for each of the $H = 500$ steps of the forecasting horizon. Each output-specific regressor comprises $K = 100$ decision trees with a maximum depth of 12. For an input feature vector $\mathbf{x}_i \in \mathbb{R}^7$, the prediction corresponding to the h -th forecasting step is computed as [10]

$$\hat{y}_{i,h} = \frac{1}{K} \sum_{k=1}^K T_{h,k}(\mathbf{x}_i), \quad h = 1, \dots, H, \quad (9)$$

where $T_{h,k}$ denotes the prediction produced by the k -th decision tree of the Random Forest regressor associated with forecasting step h . The complete forecast is obtained by concatenating the predictions corresponding to all H forecasting steps:

$$\hat{\mathbf{y}}_i = [\hat{y}_{i,1}, \hat{y}_{i,2}, \dots, \hat{y}_{i,H}]^T \in \mathbb{R}^{500}. \quad (10)$$

The **eXtreme Gradient Boosting (XGBoost)** model was also implemented as a *MultiOutputRegressor* using 100 gradient-boosted trees, with a maximum depth of 6 and a learning rate of 0.08. Through sequential tree construction, each new tree corrects the residual errors of the previous ensemble, enabling effective modeling of intricate patterns in the physics-guided feature space.

The **1D-CNN** comprises two convolutional layers with 32 and 64 filters, respectively, and a kernel size of 3, followed by batch normalization, global average pooling, two fully connected layers with 128 and 64 ReLU neurons, dropout regularization, and a linear 500-output layer. Adam, mean squared error loss, early stopping, and learning-rate reduction were used during training. The model was included as an exploratory deep-learning baseline motivated by prior CNN-based transient fault-analysis studies in DC distribution systems [6]. However, because it operates on the seven-dimensional physics-guided feature vector rather than raw waveform samples, its comparison with tree-based models is interpreted only within this compact tabular representation.

All models were trained and evaluated using a strict 5-fold **GroupKFold cross-validation** strategy. Each original simulation file received a unique group identifier, ensuring that all forecasting instances originating from the same parametric scenario remained together in either the training or validation set. This grouping prevents data leakage across different voltage levels, source configurations, and fault conditions, providing a reliable measure of generalization performance.

Sliding windows were advanced by 50 samples (100 μs), corresponding to 75% overlap. No separate external test set

was used; in each iteration, the held-out GroupKFold fold served as the test set. The reported hyperparameters were held fixed across all folds, with no fold-specific tuning. The 400 μs observation interval contributes directly to the decision delay; since feature-extraction and inference times were not benchmarked on target hardware, the present results do not establish real-time feasibility.

IV. EXPERIMENTAL RESULTS

After applying the sliding-window approach with a window size of 200 samples (400 μs) and a prediction horizon of 500 samples (1 ms), a total of 10,906 valid forecasting instances were generated. Each instance consists of a seven-dimensional physics-guided feature vector as input and the corresponding future short circuit current waveform segment as the target.

For each validation iteration, the models were trained using four folds and evaluated on the remaining unseen fold. The final reported results represent the average performance across all five validation folds.

Forecasting accuracy was assessed using three standard regression metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2).

The *Root Mean Square Error (RMSE)* measures the standard deviation of the prediction errors, giving higher weight to larger deviations: [11]

$$\text{RMSE} = \sqrt{\frac{1}{N_e} \sum_{i=1}^{N_e} (y_i - \hat{y}_i)^2} \quad (11)$$

The *Mean Absolute Error (MAE)* represents the average magnitude of the errors in the predicted values: [11]

$$\text{MAE} = \frac{1}{N_e} \sum_{i=1}^{N_e} |y_i - \hat{y}_i| \quad (12)$$

The *coefficient of determination (R^2)* indicates the proportion of the variance in the actual short circuit current that is explained by the model: [12]

$$R^2 = 1 - \frac{\sum_{i=1}^{N_e} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_e} (y_i - \bar{y})^2} \quad (13)$$

where y_i and $\hat{y}_i$ denote the actual and predicted values, respectively, and $\bar{y}$ is the mean of the observed values. Higher R^2 values (closer to 1) and lower RMSE/MAE values indicate better model performance.

Table I summarizes the average forecasting performance obtained across all folds. The Random Forest model achieved the best overall results, with a mean RMSE of 9.14 A, MAE of 3.98 A, and an excellent coefficient of determination ($R^2 = 0.981$). This indicates that the model successfully explains more than 98% of the variance in the future short circuit current waveform using only the seven physics-guided features.

Under the considered compact feature representation, XGBoost achieved the second-best forecasting performance, while

TABLE I: Averaged Performance Metrics for Multi-Step Short Circuit Current Forecasting

Model	RMSE [A]	MAE [A]	R^2
Random Forest	9.14	3.98	0.981
XGBoost	10.68	5.05	0.967
1D-CNN	15.24	8.11	0.937

the 1D-CNN produced higher prediction errors, particularly during highly dynamic transient regions. These results should be interpreted in relation to the tabular seven-feature input representation used by all three models.

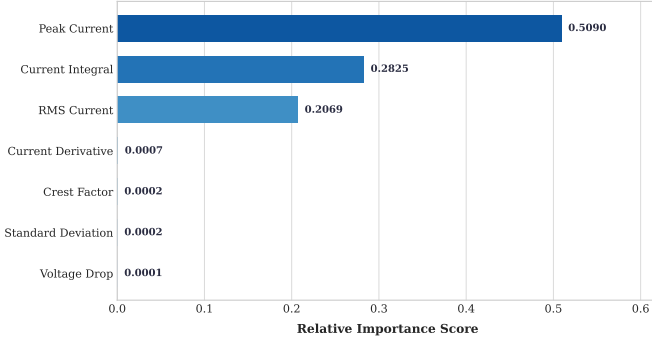


Fig. 5: Feature Importance Analysis of Physics-Guided Descriptors

Fig. 5 presents the impurity-based Random Forest feature-importance scores. Peak current, current integral, and RMS current received the highest scores, indicating that the model relies primarily on magnitude-related information. These scores describe model reliance rather than causal physical mechanisms and may be influenced by correlations among the input descriptors.

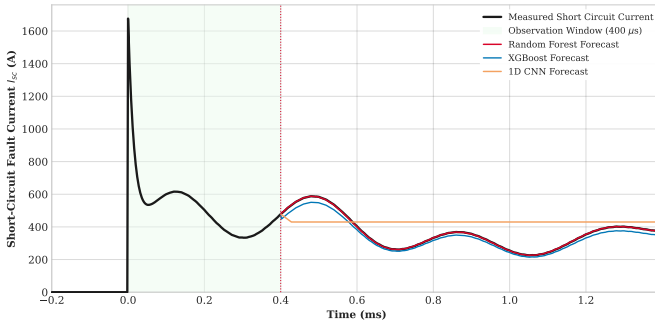


Fig. 6: Multi-model forecasting of LVDC short circuit current evolution following fault inception

Figure 6 presents a representative forecasting example for a short circuit scenario. The prediction is generated using only the information contained within the 400 μ s observation window following fault inception. Consistent with the results reported in Table I, the Random Forest model provides the closest approximation to the measured short circuit current

waveform, while XGBoost and 1D-CNN exhibit larger deviations during the transient evolution.

V. CONCLUSION

This paper presented a physics-guided, feature-based framework for direct multi-step forecasting of post-fault short circuit current waveforms. Under leakage-aware five-fold GroupK-Fold validation, Random Forest achieved the best mean performance, with an RMSE of 9.14 A, an MAE of 3.98 A, and an R^2 of 0.981. Under the compact seven-feature representation considered in this study, the tree-based ensemble models also outperformed the exploratory 1D-CNN baseline. Furthermore, the Random Forest feature-importance analysis identified peak current, current integral, and RMS current as the most influential descriptors, although these scores indicate model reliance rather than causal physical relevance. The results demonstrate the feasibility of forecasting current evolution from early post-fault observations, but remain limited to simulation-derived fault data and do not establish fault-detection or real-time protection performance. Future work will investigate raw-waveform and analytical baselines, robustness to measurement noise, and end-to-end latency on target hardware.

REFERENCES

- [1] M. S. Alam, F. S. Al-Ismael, S. M. Rahman, M. Shafiullah, and M. A. Hossain, "Planning and protection of dc microgrid: A critical review on recent developments," *Engineering Science and Technology, an International Journal*, vol. 41, p. 101404, 2023.
- [2] H. Binqadhi, W. M. Hamanah, M. Shafiullah, M. S. Alam, M. M. AlMuhaini, and M. A. Abido, "A comprehensive survey on advancement and challenges of DC microgrid protection," *Sustainability*, vol. 16, no. 14, p. 6008, 2024.
- [3] M. H. Ibrahim, E. A. Badran, and M. H. Abdel-Rahman, "Detect, classify, and locate faults in DC microgrids based on support vector machines and bagged trees in the machine learning approach," *IEEE Access*, vol. 9, pp. 75 718–75 732, 2021.
- [4] X. Miao, S. Zhuang, J. Li, and L. Tang, "Review of short circuit current prediction technology in electric power system," in *Proceedings of 2021 Chinese Intelligent Systems Conference: Volume I*. Springer, 2021, pp. 311–327.
- [5] T. Wang, C. Zhang, Z. Hao, A. Monti, and F. Ponci, "Data-driven fault detection and isolation in DC microgrids without prior fault data: A transfer learning approach," *Applied Energy*, vol. 336, p. 120708, 2023.
- [6] F. Zheng, Y. Peng, C. Jiang, Y. Lin, and N. Liang, "Research on the identification of high-resistance ground faults in the flexible DC distribution network based on VMD-inception-CNN," *Frontiers in Energy Research*, vol. 11, p. 1258619, 2023.
- [7] P. Mishra and P. Zhang, "Physics-informed transient stability assessment of microgrids," *iEnergy*, vol. 2, no. 3, pp. 231–239, 2023.
- [8] D.-C. Mitroi, G. Stamatescu, R. Plamanescu, and M. Albu, "Parametric short circuit dataset for fault analysis and protection of dc microgrids," 2026. [Online]. Available: <https://dx.doi.org/10.21227/yk0q-v117>
- [9] D.-C. Mitroi, G. Stamatescu, R. Plămănescu, and M. Albu, "Co-simulation of shortcircuit current measurement in lvdC systems," in *2025 9th International Symposium on Electrical and Electronics Engineering (ISEEE)*, vol. CFP2593K-ART, 2025, pp. 1–6.
- [10] B. Magalhães, P. Bento, J. Pombo, M. d. R. Calado, and S. Mariano, "Short-term load forecasting based on optimized random forest and optimal feature selection," *Energies*, vol. 17, no. 8, 2024.
- [11] D.-C. Mitroi, I. Stamatescu, N. Arghira, and I. Făgărășan, "Lithium-ion battery soc estimation: A neural network approach," in *2024 28th International Conference on System Theory, Control and Computing (ICSTCC)*, 2024, pp. 432–437.
- [12] D.-C. Mitroi, C. Nichiforov, I. Stamatescu, and G. Stamatescu, "Analysis of energy data forecasting performance in low-voltage dc microgrids," in *2025 29th International Conference on System Theory, Control and Computing (ICSTCC)*, 2025, pp. 31–36.