Chaos
An Interdisciplinary Journal of Nonlinear Science

RESEARCH ARTICLE | JANUARY 25 2024

Predicting dynamic stability from static features in power grid models using machine learning

Special Collection: Data-Driven Models and Analysis of Complex Systems

Maurizio Titz ; Franz Kaiser ; Johannes Kruse ; Dirk Witthaut

Chaos 34, 013139 (2024)
https://doi.org/10.1063/5.0175372

Predicting dynamic stability from static features in power grid models using machine learning

Cite as: Chaos 34, 013139 (2024); doi: 10.1063/5.0175372
Submitted: 6 September 2023 · Accepted: 11 December 2023 ·
Published Online: 25 January 2024

Maurizio Titz,[1,2,3,a)] Franz Kaiser,[2,3] Johannes Kruse,[1,2,3] and Dirk Witthaut[1,2,3,b)]

AFFILIATIONS

[1]Forschungszentrum Jülich, Institute for Energy and Climate Research—Energy Systems Engineering (IEK-10), 52428 Jülich, Germany
[2]Forschungszentrum Jülich, Institute for Energy and Climate Research—Systems Analysis and Technology Evaluation (IEK-STE), 52428 Jülich, Germany
[3]Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany

Note: This paper is part of the Focus Issue: Data-Driven Models and Analysis of Complex Systems.
[a)]**Electronic mail:** m.titz@fz-juelich.de
[b)]**Author to whom correspondence should be addressed:** d.witthaut@fz-juelich.de

ABSTRACT

A reliable supply with electric power is vital for our society. Transmission line failures are among the biggest threats for power grid stability as they may lead to a splitting of the grid into mutual asynchronous fragments. New conceptual methods are needed to assess system stability that complement existing simulation models. In this article, we propose a combination of network science metrics and machine learning models to predict the risk of desynchronization events. Network science provides metrics for essential properties of transmission lines such as their redundancy or centrality. Machine learning models perform inherent feature selection and, thus, reveal key factors that determine network robustness and vulnerability. As a case study, we train and test such models on simulated data from several synthetic test grids. We find that the integrated models are capable of predicting desynchronization events after line failures with an average precision greater than 0.996 when averaging over all datasets. Learning transfer between different datasets is generally possible, at a slight loss of prediction performance. Our results suggest that power grid desynchronization is essentially governed by only a few network metrics that quantify the networks' ability to reroute the flow without creating exceedingly high static line loadings.

Modern society relies on a secure and stable supply of electric power, which makes a failure of the electric power system especially harmful.[1,2] It is, thus, of utmost importance to make and keep it as fail-proof as possible.[3] Large-scale power outages are typically triggered by the failure of a single transmission line in a period of a high grid load.[4] But not all failures are equally perilous. Most failures will not cause any major disturbance, while some are fatal. But what determines the importance of a single line? Can we find suitable indicators that identify critical lines and provide insights into network stability? Graph theory and network science provide various metrics for the importance of single nodes or lines, covering very different aspects of importance. In this article, we develop an interpretable machine learning model that integrates various metrics and identifies the ones with the highest predictive power. We apply the model to simulated data from synthetic test grids operating at high loads and observe a very good prediction performance.

I. INTRODUCTION

The ongoing transition to a sustainable energy system challenges the operation and stability of the electric power system. Renewable power sources are fluctuating and uncertain such that storage or backup infrastructures are required.[5,6] Furthermore, they are often built at places that offer favorable conditions for generation, far away from consumers, which increases grid loads.[7,8] Similarly, energy-intensive sectors, such as transport and heating, have to shift away from fossil fuels and toward electrification to reduce

carbon emissions,[9] putting further strain on the power grid.[10,11] Grid stability, in general, is, thus, not only an important but also a timely topic.

A central aspect of power system stability is synchronicity. All generators in a grid have to rotate in synchrony to guarantee a steady flow of electric power.[12–14] Stability is at risk if the grid is disturbed or damaged. In fact, most blackouts can be traced back to the outage of a single power system element such as a transmission line.[4] In transmission grids, which are generally strongly meshed, this initial failure can then lead to a cascade due to overloads of other transmission lines. At some point, either directly or after some steps of the cascade, the grid becomes dynamically unstable, and synchronicity is lost. This scenario occurred during the 2003 Italian power outage[15] and the 2006 Western European power outage.[16] Notably, a desynchronization does not necessarily induce a large-scale blackout if the asynchronous fragments can be stabilized by control actions or load shedding. In fact, two such events were observed in the Continental European grid in 2021.[17,18]

To divert these catastrophes, grid operators have to be able to judge risks *in time*. Warning systems that alert grid operators are in place, and the risk regarding any single contingency can, in principle, be assessed via simulations (see Ref. 19 for a review). However, the sheer number of potentially critical structural elements in power grid systems makes simulating all *possible* contingencies *in time* computationally impossible. Transmission grid operators, thus, have to, in part, rely on heuristics and experience. Modern machine learning methods may contribute to this assessment or the selection of relevant contingency cases to be simulated in detail.[20,21]

In this article, we explore the capability of supervised machine learning to identify line failures that lead to desynchronization. We employ computationally cheap, static inputs that are readily available to the transmission grid operators such that models can be evaluated rapidly. Furthermore, we focus on efficient models that are amenable to human interpretation and, thus, enable scientific insights.[22]

A key method is the combination of machine learning with methods from graph theory and network science to develop features that are both understandable and have a high predictive power. Ideally, these features will directly measure a relevant quantity of the respective line, such as its redundancy or centrality, which lend themselves to human interpretation. Machine learning models perform inherent feature selection and may, thus, reveal key factors that determine network robustness and vulnerability. Furthermore, they can incorporate different features to improve prediction.

The article is organized as follows. In Sec. II, we will first give an overview of the model used for power grid dynamics and the resulting datasets. We then introduce the machine learning models, performance metrics, and eXplainable Artificial Intelligence (XAI) methods used. After that, we introduce a multitude of different graph theory-based features and assess their ability to quantify desynchronization risk in Sec. III. We then go on to show that using machine learning produces improved predictions and quantify the contribution of different input features in Sec. IV. Next, we demonstrate that the machine learning models are capable of learning transfer between different datasets. Finally, we interpret the predictions of one model by showing under which circumstances it tends to fail and how it outperforms simpler models.

II. METHODS

Our approach is summarized in Fig. 1. Using a coarse-grained model, we simulate the dynamics of a power grid after the failure of a single transmission line. We then classify the line according to the final post-failure state: The grid can either relax to a new synchronous steady state or lose synchronicity permanently. This procedure is repeated for all lines in a set of synthetic test grids, providing the raw data for the development of machine learning models.

This work aims to predict the dynamic stability of a power system based solely on static features, i.e., features that can be derived from the pre-failure steady state, without making use of any simulation results. This way, our prediction method is computationally much cheaper than simulations, giving it a decisive advantage in an application case where real-time risk assessment with limited computational resources prohibits simulating all relevant contingencies. We make use of feature engineering based on concepts from graph theory and network science, which will be described in detail in Sec. III.

A. Simulating the impact of line failures

Models of varying complexity are used for power system stability analysis, depending on the scope of the analysis and the stability mechanism of interest.[23,24] Since this work investigates solely the synchronization behavior on large spatial scales, we focus on the voltage phase angles and frequencies, neglecting aspects of voltage stability or control. Furthermore, we focus on coarse spatial scales and, thus, consider aggregated models.[25–29] In this approach, each node of the network represents a small region or city, labeled by $i \in \{1, \ldots, N\}$. The dynamics of the local phase angle δ_i and frequency $\omega_i = \dot{\delta}_i$ is then determined by the aggregated swing equation

$$J_i \ddot{\delta}_i + D_i \dot{\delta}_i = P_i^{(in)} - P_i^{(el)}(t). \tag{1}$$

Here, J_i and D_i denote the inertia and damping constant, respectively; $P_i^{(in)}$ is the effective real power injection at node i; and $P_i^{(el)}$ is the real power exchanged with the grid. Throughout the paper, we assume that the power is balanced such that $\sum_i P_i^{(in)} = 0$. We neglect transmission losses as Ohmic resistance is typically small in high-voltage transmission grids. Hence, the real power exchanged with the grid can be written as

$$P_i^{(el)}(t) = \sum_{j=1}^{N} P_{ij}(t) = \sum_{j=1}^{N} K_{ij} \sin\left(\delta_i - \delta_j\right), \tag{2}$$

where P_{ij} is the real power flow on line (i, j). The "coupling strength" $K_{ij} = K_{ji}$ is determined by the susceptance of the respective transmission line (being zero if no line exists), and the voltage level of the grid.[24] In the context of network science, Eq. (1) is commonly referred to as the second-order Kuramoto model or Kuramoto model with inertia.[30]

In the simulations, we first find a steady state of the equations of motion (1) for the pre-failure grid. This is used as the initial state for the following simulations. Then, we select a transmission line (i, j) and remove it from the grid by setting K_{ij} to zero. We simulate the dynamics using *PowerDynamics.jl*[31] and assess stability in terms of

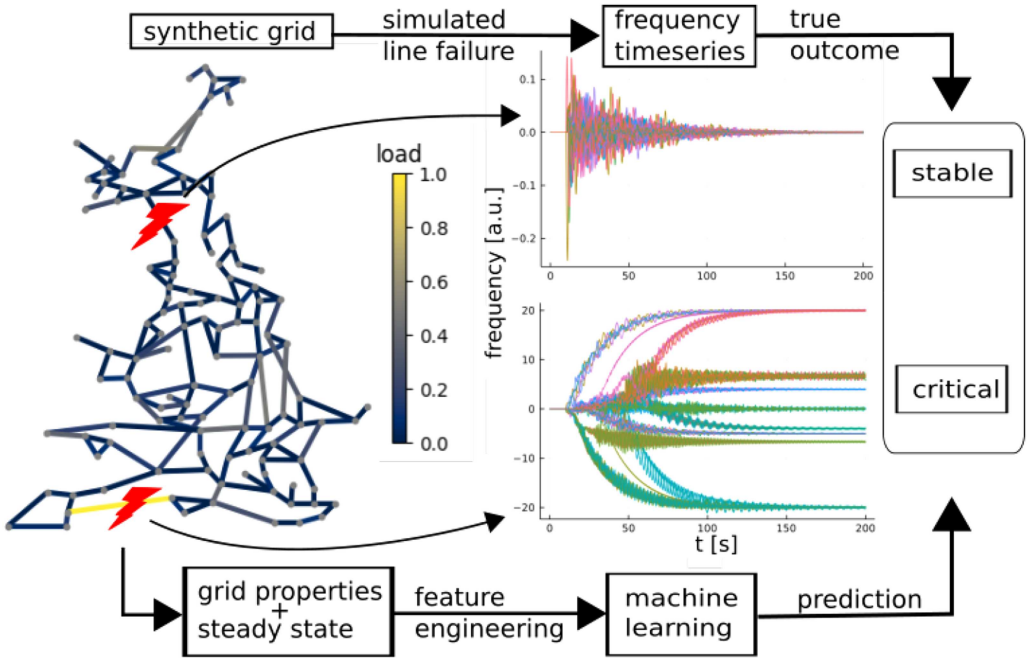

FIG. 1. Summary of the machine learning model developed in this paper. We simulated the dynamics of a power grid after the failure of a single transmission line and classified the outcome as either stable or critical depending on whether nodal frequencies relaxed to zero. Input features for stability prediction are engineered from the properties of the grid and the pre-failure steady state using concepts from graph theory and network science. We then trained an explainable machine learning model to predict the stability outcome.

the final state: A line is classified as stable if all nodal frequencies $\dot{\delta}_i(t)$ relax back to zero and critical otherwise. This procedure is repeated for all lines in the respective grid.

We remark that we assume that the power injections $P_i^{(in)}$ are fixed and perfectly known, while they may fluctuate in reality. Fast fluctuations of the power can deteriorate the stability of a power grid[32,33] and, thus, decrease the likeliness to relax back to a state with vanishing nodal frequencies $\dot{\delta}_i(t)$. Furthermore, the values of $P_i^{(in)}$ may slowly drift and, thus, differ from the originally planned dispatch. Hence, line loads change, which can affect the vulnerability of the grid.[34]

B. Power grid datasets

Training and evaluating machine learning models require a dataset of sufficient size. Real power grids are usually operated well within their margin of stability, such that desynchronization events are rare if potentially catastrophic. Hence, it is impossible to use real data or models of normal operation dispatch scenarios in our study. Instead, we resort to synthetic models that are close to actual grid topologies. In order to create datasets containing sufficient numbers of critical cases, the grids are designed as to be heavily loaded and are not necessarily *statically* $n-1$ stable, i.e., static stability is not guaranteed for all n possible single-line failures.

Eight different datasets were produced; an overview is given in Table I. All test cases in a dataset either have the British transmission grid topology (marked GB) or the topologies were generated synthetically using the random growth model developed in Ref. 35. In the latter case, parameters were chosen that produce topologies mimicking properties of the U.S. transmission grid (marked US). Next, we specify how many nodes in the grid act as effective generators $\left(P_i^{(in)} > 0\right)$ or effective consumers $\left(P_i^{(in)} < 0\right)$. Nodes are then randomly assigned to one of the two classes. The ratio of the number of generators and consumers N_G:N_C as well as the value of $\left|P_i^{(in)}\right|$ for the consumer nodes are given in Table I for all datasets. The transmission lines were either all modeled with the same effective coupling strength K_{ij}, or K_{ij} was set to be inversely proportional to the line length as derived from node positions. The latter case treats all transmission lines as conductors of the same electrical conductivity. Obviously, $K_{ij} = 0$ if two nodes i and j are not connected. Two datasets additionally contained grids of varying line loadings by rescaling node powers as shown in Table I. The dataset marked by $\mathrm{US}_\circ$ had all dead ends removed by connecting them to the next closest node. Finally, the dataset marked by GB_{pert} was created by adding randomly drawn perturbations of sum zero to every node power $P_i^{(in)}$ of a reference grid. In some cases, no steady state could be found for the pre-failure grid. These cases were discarded. The total numbers of simulated line failures vary by dataset; see Table I.

TABLE I. Synthetic grid datasets and their properties: K_{ij} is the line capacity and $|P_C|$ is the effective power demand of a consumer node. The generator power $|P_G|$ then follows from the generator–consumer ratio N_G:N_C since power is balanced. $N_G + N_C$ is the total number of nodes. Datasets US^P and US^P_B contain grids with $|P_C|$ = 1.5, 1.6, 1.7, 1.8, 1.9, and 2.

	US	$\mathrm{US}_\circ$	US^P	US^P_B	US_{het}	GB_{het}	GB	GB_{pert}
K_{ij}	L_{ij}^{-1}	L_{ij}^{-1}	L_{ij}^{-1}	4	L_{ij}^{-1}	L_{ij}^{-1}	L_{ij}^{-1}	L_{ij}^{-1}
$\|P_C\|$	2	2	[1.5, 2]	[1.5, 2]	0.75	1	1	1
N_G:N_C	1:1	1:1	1:1	1:1	1:4	1:4	1:1	1:1
$N_G + N_C$	50	50	50	50	50	120	120	120
Samples	43 825	26 734	126 508	65 568	21 936	118 400	29 600	73 556

C. Machine learning models

For our predictions, we use tree-based machine learning models because they provide state-of-the-art performance for many applications[36] as well as a high level of explainability.[37] The output of each model is a number $y \in [0, 1]$, which can be interpreted as the probability of a line failure leading to desynchronization. This number can be used to derive a classification by a simple thresholding procedure.

We compare the performance of state-of-the-art gradient boosted tree (GBT) models[36] to the gradient boosted stump model (stumps) and a simple decision tree (DT) model to find out how much complexity is needed to maximize performance. Here, a stump is a tree made up of only a root node and two leaf nodes. The architecture of the three model types is illustrated in Fig. 2. All models were subjected to hyperparameter optimization via random search.

Tree-based models can be made transparent by different XAI methods. We use SHapley Additive exPlanations (SHAP),[37] in particular, for the quantification of feature importance. The SHAP value quantifies how strongly a feature influences a given prediction made by a model. Averaging over all predictions then yields the global feature importance, i.e., how much a given feature contributes to a model overall. To reduce model complexity and increase interpretability, we try to find an "optimal" model of maximal performance and minimal dimensionality. Since evaluating all possible feature combinations is computationally prohibitive, we perform recursive feature elimination (RFE) in which the feature with the smallest feature importance is removed recursively from the model. Note that RFE does not necessarily lead to the optimal feature combination due to limitations in importance attribution.[38]

For each dataset, 80% of the samples are used for training and the remaining 20% for testing. We used fourfold cross-validation, i.e., the performance is calculated by averaging over four different train-test splits. Performance was assessed across multiple metrics, which generally agreed. For conciseness, in this paper, we, therefore, limit ourselves to the average precision (AP) as a numerical and the detection error trade-off curve (DET) plot[39] as a graphical performance metric. The AP is a well-suited metric for our purposes, as it quantifies the ability of an algorithm to rank samples by relevance, i.e., desynchronization risk in our case. In potential applications cases, machine learning would be combined with detailed numerical simulations. The output of the machine learning model would then be used to narrow down the vast amount of possible contingencies to those that should be investigated further via simulations.

III. METRICS OF LINK IMPORTANCE AND REDUNDANCY

A key idea of this study is to combine machine learning with concepts from graph theory and network science, which are used to engineer features that are interpretable and have a high predictive power, as for instance, measures of redundancy or centrality. Features are computed from the grid's topology (adjacency matrix),

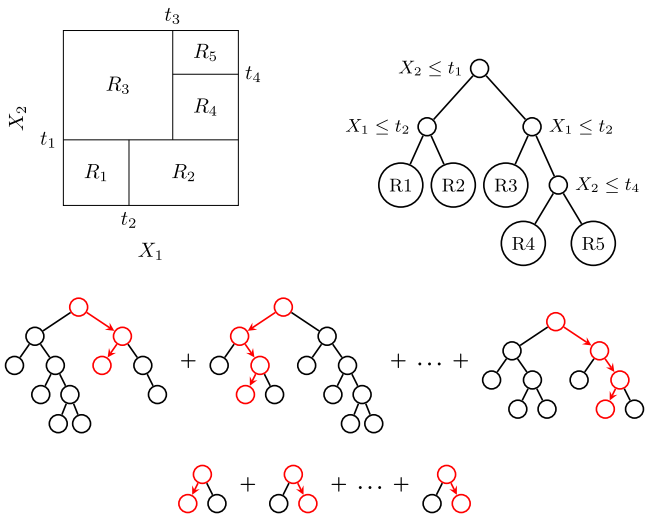

FIG. 2. Machine learning methods used in this work. Top row: Decision trees (DT) divide the feature space into hyperrectangles. For every sample, they predict the mean value of all training samples that lay in the hyper-rectangle the given sample is located in. This division of the feature space is done via the tree structure. Each leaf of the tree corresponds to a hyperrectangle in feature space. Middle row: The gradient boosted trees (GBT) model is initiated as a constant prediction and then iteratively adds trees to the model. Every new tree is trained so that its addition to the model minimizes the model's loss. Bottom row: We also use a gradient boosted trees model with a maximal tree depth of two, e.g., each tree can divide the samples only once, using one feature. A decision tree with only two leaf notes is also referred to as a stump.

TABLE II. List of the engineered features used to predict desynchronization. Features for which no symbol is listed showed bad performance and are omitted in Fig. 4 for conciseness. All features were used as features in the machine learning models.

Feature name/definition	Symbol
Redundant capacity ratio[40]	r_{ab}
Widest path redundant capacity ratio	r^{w}_{ab}
Shortest path redundant capacity ratio	r^{σ}_{ab}
Redundant capacity[40]	K^{red}_{ab}
Widest path redundant capacity	$K^{red,w}_{ab}$
Shortest path redundant capacity	$K^{red,\sigma}_{ab}$
Response theory pred. max. load[40]	l^{re}_{ab}
LODF pred. max. load[41]	l^{DC}_{ab}
$\sqrt{r^2_{ab} + (l^{DC}_{ab})^2}$[40]	c_{ab}
Max. load at operation point	l_{OP}
Load on failing line	l_{ab}
Flow on failing line	$\lvert P_{ab}\rvert$
Phase cohesion[13]	Not shown
Post-failure phase cohesion[13]	ρ_{ab}
$\max_i \lvert P_i\rvert - \sum_j K_{ij}$	b_{ab}
$\max_i \lvert P_i\rvert / \sum_j K_{ij}$	Not shown
Rerouting resistance distance	X^{re}_{ab}
Weighted rerouting distance	X^{σ}_{ab}
Normalized edge current betweenness[42]	ϵ^{CB}_{ab}
Pre-failure algebraic connectivity	Not shown
Post-failure algebraic connectivity	Not shown
Algebraic connectivity loss	$\Delta\lambda_{2,ab}$
Rerouting distance	d^{re}_{ab}
Edge betweenness	ϵ_{ab}
Edge connectivity	τ_{ab}
Edge k core	Not shown

the electric properties (P^{in}_i and K_{ij}), and the pre-outage state (phase angles δ_i and flows P_{ij}). In the following, we introduce the engineered features; a summary is given in Table II. The definition of the features l_{ab}, r_{ab}, and τ_{ab} is further illustrated in Fig. 3. We note that some of the features are widely used in network science. Hence, we introduce them very briefly. Furthermore, we will provide a first assessment of their predictive power using uni-variate models.

A. Definition of the features

The most simple features describing a failing line (a, b) are the flow $|P_{ab}|$ and the load $l_{ab} = |P_{ab}|/K_{ab}$ in the pre-outage state. Intuitively, we expect that the failure of a strongly loaded line will have a stronger impact than a weakly loaded line.

A variety of measures of connectivity, redundancy, and centrality were introduced in the context of network science. The *edge connectivity* τ_{ab} of a link (a, b) is defined as the number of edge-independent paths between the nodes a and b and provides an elementary measure of redundancy.[43] If $\tau_{ab} = 1$, then the removal of the edge (a, b) will disconnect the grid and almost surely lead to desynchronization. The *algebraic connectivity* or Fielder value λ_2 is of particular interest in flow networks.[44,45] It is defined as the smallest non-zero eigenvalue λ_2 of the graph Laplacian matrix $\mathbf{Y} \in \mathbb{R}^{N\times N}$,[44]

$$Y_{ij} = \begin{cases} -K_{ij} & \text{if } i \text{ is connected to } j, \\ \sum_{\ell} K_{i\ell} & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

We evaluate this eigenvalue before and after the removal of the line (a, b) (pre- and post-failure grid) and compute the difference $\Delta\lambda_{2,ab}$ as a measure for the loss of connectivity.

The *edge betweenness* measures the centrality of an edge.[43] The original version ϵ_{ab} is defined by computing the shortest path between all pairs of nodes in the network and counting how many of these paths cross the edge (a, b). Here, we also use the flow-based version ϵ^{CB}_{ab} where the shortest paths are replaced by current flows.[46] The *coreness* of a line has been shown to be related to its vulnerability.[47] The k-core of a graph is its maximal subgraph that contains only nodes of degree k or more. The coreness of a node a is defined as the highest k for which a is part of the k-core, and the coreness of a line is the smaller of the corenesses of its terminal nodes.

If an edge (a, b) fails, its flow must be rerouted via alternative paths. The properties of these paths, especially their length, may, thus, be relevant to determine the impact of the failure. The *rerouting distance* was introduced in[48] to predict flow changes in linear flow networks resulting from line failures. The rerouting distance of two lines is defined as the length of the shortest loop that contains both lines. Here, we introduce the rerouting distance of a single line (a, b) as the shortest loop that contains the line. Up to an offset of one, this is equivalent to the geodesic distance between a and b in the post-failure grid. Here, we extend this definition to incorporate line properties, interpreting K^{-1}_{ij} as the effective resistance of an edge (i, j). We then define the *weighted rerouting distance* X^{σ}_{ab}, as the effective series impedance of the weighted shortest path in the post-failure grid. We further define the *rerouting resistance distance* X^{re}_{ab} of a line (a, b) as the resistance distance[49] between a and b in the post-failure grid.

Line outage distribution factors (LODFs) are widely used in power system stability analysis.[50–52] Assuming small-phase angle differences between connected nodes, the equations describing the steady state are linearized, $\sin(\delta_i - \delta_j) \approx \delta_i - \delta_j$, which allows to compute the post-failure steady state analytically, implicitly assuming that such a steady state exists.[48,53] To the same end, in Ref. 54, *linear response theory* was applied to the second-order Kuramoto model by linearizing the sine around the pre-failure steady state. In our context, the maximal predicted line loading $\max_{ij} |P_{ij}|/K_{ij}$ in the post-failure grid is of particular importance. A value greater than one hints at an overload and, thus, a loss of stability. This quantity is denoted as l^{DC}_{ab} (computed using LODFs) and l^{re}_{ab} (linear response theory), respectively.

The maximal line loading at the pre-failure operation point, $\max_{(i,j)\in E} \sin\left(\delta_i - \delta_j\right)$, can be used as a proxy for the overall line loading. We denote it as l_{OP}.

A simple necessary condition for the existence of a steady state of Eq. (1) is given by $|P^{(in)}_i| \le \sum_j K_{ij}$. Hence, we consider the following quantities, evaluated after the removal of edge (a, b), as potential

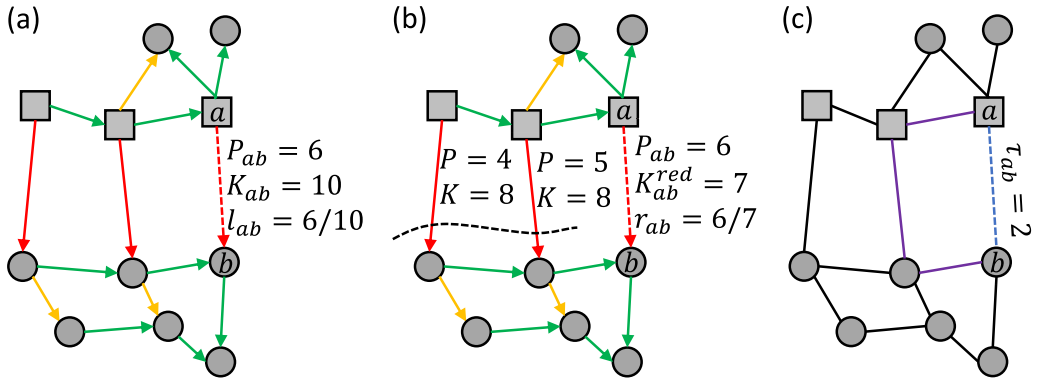

FIG. 3. Definition of the features l_{ab}, r_{ab}, and τ_{ab}. To illustrate the definition, we consider a small example network with three generator nodes (squares) and eight consumer nodes (circles). The color of the edges indicates the loading, where red corresponds to a high load. (a) The failing line (dashed) has a capacity $K_{ab} = 10$ and carries a flow $P_{ab} = 6$. The load is $l_{ab} = P_{ab}/K_{ab} = 6/10$. (b) If the line (a, b) fails, the power flow has to be rerouted via alternative paths. There is a bottleneck for the rerouting (dashed black line) with two lines having unused capacity $K_{ab}^{\text{red}} = 7$. The redundant capacity ratio is $r_{ab} = P_{ab}/K_{ab}^{\text{red}} = 6/7$. (c) The edge connectivity τ_{ab} counts the number of edge-independent paths, i.e., paths that do not share an edge, from a to b. In the example, we have two such paths highlighted in blue and violet.

features

$$\max_i \left|P_i^{(in)}\right| - \sum_j K_{ij}, \qquad \max_i \left|P_i^{(in)}\right| / \sum_j K_{ij}. \tag{4}$$

Dörfler *et al.*[13] introduced a synchronization condition for the existence of a stable steady state with maximal phase difference γ between any two connected oscillators. Adapted to our problem, the criteria reads

$$||\mathbf{Y}^{\dagger}\boldsymbol{P}||_{\epsilon,\infty} \leq \sin(\gamma), \tag{5}$$

where $\mathbf{Y}^{\dagger}$ is the pseudoinverse of the graph Laplacian matrix, $\boldsymbol{P} = \left(P_1^{(in)}, \ldots, P_N^{(in)}\right)^{\top}$, and $||x||_{\epsilon,\infty} = \max_{i,j\in E} |x_i - x_j|$. In the following, we use the left-hand side of the equation as an input feature and refer to it as the *cohesiveness* ρ. The respective value in the post-failure grid, i.e., after removing the edge (a, b) from the Laplacian, is denoted as ρ_{ab}.

The *redundant capacity* K_{ab}^{red} of a line (a, b) quantifies its redundancy in the pre-failure steady state of the grid.[40] This measure is inspired by graph theoretical flow problems, in particular, the min-cut max-flow theorem.[55] It is defined as the total additional flow that the grid can transport from a to b by using all possible paths that do not include line (a, b) itself. Since K_{ab}^{red} will be used to assess the impact of a line failure, we must take into account that lines typically carry a flow before the failure. Therefore, we do not ask how much flow a line (i, j) can carry in total (given by the coupling constant K_{ij}) but how much it can carry *in addition to the pre-failure state*. Accordingly, we introduce a residual network G^{res} in which lines are described by the residual capacity $K_{ij}^{res} = K_{ij} - P_{ij}$. Since $K_{ij}^{res} \neq K_{ji}^{res}$, we must view G^{res} as a directed graph, where every line is represented by two directed edges with different capacities. The redundant capacity K_{ab}^{red} of the line (a, b) is then obtained by removing the line (a, b) from G^{res} and then computing the maximum (a, b)-flow via the Edmonds–Karp algorithm.[55] By the virtue of the min-cut max-flow theorem, the redundant capacity K_{ab}^{red} equals the minimum capacity of all $(a - b)$-cuts in the G^{res} and, thus, identifies the bottlenecks between the nodes a and b. Furthermore, we relate the redundant capacity to the amount of flow that has to be rerouted, which is given by the pre-outage flow on the failing line $\left|P_{ab}^{\text{pre}}\right|$. The *redundant capacity ratio* $r_{ab} = \left|P_{ab}^{\text{pre}}\right| / K_{ab}^{red}$ has high predictive power for the impact of line failures as previously shown in Ref. 40. The definition of K_{ab}^{red} and r_{ab} is further illustrated in Fig. 3. Furthermore, we consider a combination of the redundant capacity ratio and the maximum load predicted from LODFs, $c_{ab} = \sqrt{r_{ab}^2 + \left(l_{ab}^{DC}\right)^2}$.[40]

The graph theoretical max flow provides an upper bound for real power flows but not the actual value. Hence, we introduce two variants of the redundant capacity $K_{ab}^{red,\sigma}$ and $K_{ab}^{red,w}$, which do not take into account all possible paths but only the shortest path (σ) path or the widest path (w) from a to b, respectively. Here, the widest path is defined as the single path with the widest bottleneck, i.e., the largest value of $\min_{ij\in\text{path}} K_{ij}^{res}$. As before, we additionally define the corresponding ratios r_{ab}^{σ} and r_{ab}^{w}.

B. Initial evaluation of the features

We now provide a first assessment of the predictive power of the engineered features in terms of simple uni-variate prediction models. For each feature, a model is setup as follows. Given a threshold h, an edge is predicted as critical if the metric exceeds h and predicted stable otherwise. By varying the value of h, one can derive the precision-recall curve and, thus, the AP. Figure 4 shows the AP scores for the different features and all datasets. Note that for conciseness, some of the features introduced before that performed badly are not shown. The features are ordered according to what information they take into account: from full information (topology, electric properties, and pre-failure state), over topological and electrical properties, to purely topological.

The features showing the best performance are c_{ab}, r_{ab}, r_{ab}^{w}, r_{ab}^{σ}, l_{re}, and l_{DC} , with c_{ab} clearly scoring the highest. Besides being strong predictors, they are also the most consistent, showing a comparably

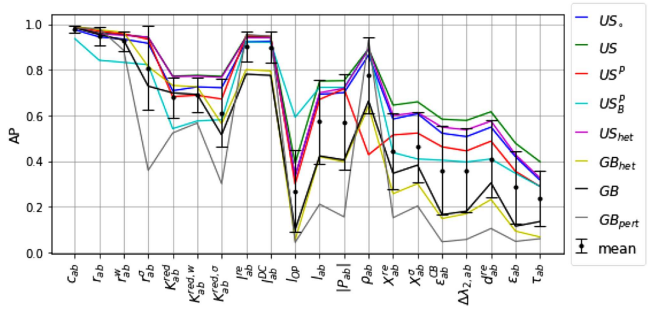

FIG. 4. Performance of uni-variate threshold models for different engineered features. The performance is quantified by the average precision (AP) score, evaluated for the eight datasets summarized in Table I. Black bars indicate the mean and standard deviation over the datasets. Note that the performance might differ across different datasets not only because of a feature's strengths and weaknesses but also because a data set can be intrinsically easier or harder to predict. Features take more information into account from right to left: first purely topological features, then including electrical properties, and finally also the pre-failure steady state. We find that features derived from the redundant capacity and the predicted maximum line load achieve the highest scores. They are also the most consistent over all datasets. Purely topological features show bad performance.

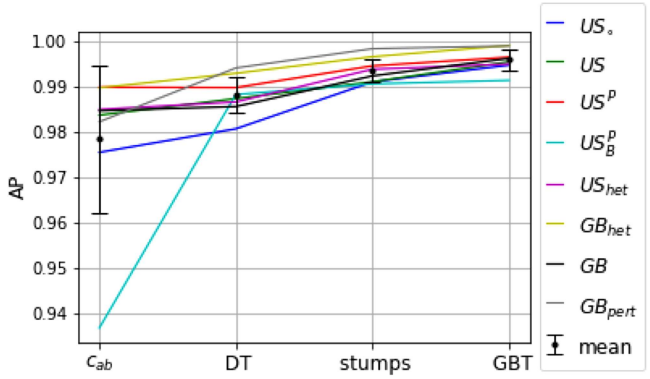

FIG. 5. The performance of machine learning models averaged over four train-test splits. We show the average precision (AP) score of the decision tree (DT), gradient boosted stumps (stumps), and gradient boosted trees (GBT) models with c_{ab} as a benchmark for all datasets. Black bars indicate the mean and standard deviation over the datasets. The GBT model consistently outperforms the other models.

low performance variance. The full redundant capacity ratio r_{ab} outperforms the related r^w_{ab} and r^σ_{ab}. The score of l^{re}_{ab} and l^{DC}_{ab} is almost identical on all metrics and datasets. All of the aforementioned features quantify the impact of the line failure on the flows. Notably, the next best feature, ρ_{ab}, also falls into this category as it also provides an estimate for the maximum line loading in the post-failure grid.

The ten best-performing features all take into account the full state of the grid. The next best features either take into account electrical and topological information or the pre-failure steady state. The best-performing purely topological feature is d^{re}_{ab}, which is outperformed by no less than 16 other features. Still, it outperforms more common and complex topological features, including the current flow centrality. This is likely due to the fact that rerouting features are more specific to the problem than centrality measures.

As described above, categorical predictions are derived by setting a threshold h. We find that the optimal threshold is mostly constant between different datasets for the best-performing features. This also hints at a potential for generalization as no further knowledge about a power grid is needed to gauge whether a failure might be critical or not. This is not given for other features. A very low impedance distance, for instance, does not prevent desynchronizations if the grid is already loaded to the point of failure.

We conclude that the best-performing features are all based on network flows—either employing a graph theoretical perspective or linearizing the steady-state equations.

IV. MACHINE LEARNING CRITICAL LINES

A. Predictability of network desynchronization

We now proceed to the main goal of our study, the combination of modern machine learning models and network science methods. The features introduced in Sec. III are used as inputs in models using gradient boosted trees (GBT), gradient boosted stumps (stumps), and a simple decision tree (DT). The feature c_{ab} is excluded here because it is already a combination of the two features r_{ab} and l^{DC}_{ab}. Instead, we use a univariate model based on c_{ab} as a benchmark.

Figure 5 shows the performance of the optimal models for all datasets. We find that all machine learning models outperform the univariate benchmark c_{ab}, i.e., combining features improves the predictability. The performance increases with model complexity, with GBT models reaching an average precision of more than 0.996 averaged over all datasets. We conclude that the proposed approach of integrating capable network features via machine learning is highly effective in assessing the network's robustness.

B. Reducing model complexity and identifying key features

Tree-based machine learning models perform inherent feature selection, which can be used to identify features with the highest predictive power. We apply recursive feature elimination to reduce model complexity and to improve the interpretability of the models. Figure 6 shows how the average performance over fourfold cross-validation evolves as features are gradually removed. In all cases, we find that many features can be discarded without loss of performance. Eliminating a feature does not harm performance if the feature is not a good predictor or when the feature is redundant with respect to the remaining features. For stumps and especially decision trees, we find the optimal number of features to be lower than for the GBT models. That is, the GBTs can better handle complex multi-dimensional inputs.

We now investigate the individual features and their role in machine learning prediction. Figure 7 shows the feature importances, i.e., how much each feature contributes to the predictions on average, for the optimal GBT models. Feature importances are

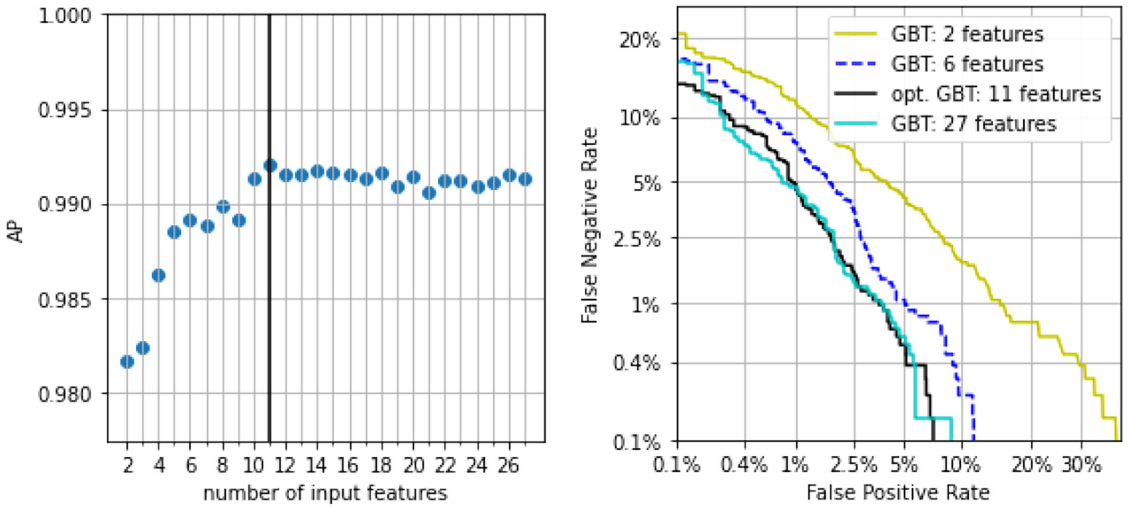

FIG. 6. Recursive feature elimination decreases model complexity and improves explainability. Left panel: Average precision score of the GBT model during recursive feature elimination. The vertical line marks the selected "optimal" model. Additional features do, in theory, not harm GBT performance; small differences occur by chance. Right panel: DET curves for GBT models with different numbers of input features. The optimal and the 27 feature model perform almost identically, confirming the choice of the optimal model. As evident from the 6-feature and 2-feature model, eliminating more features leads to significant performance loss.

normalized so that they sum to one. Results are mostly consistent between datasets with higher similarity also showing more similar feature importance. We find that, in general, feature importances are closely linked to the single feature performance shown in Fig. 4. That is, a feature with a high univariate predictive power will typically also show a high feature importance. The redundant capacity ratio and predicted line loading features have the highest feature importances. Purely topological features add very little value and are, thus, barely used by the models.

We note that two groups of features are based on similar concepts and are, thus, internally highly correlated. The two features l_{ab}^{DC} and l_{ab}^{re} are both based on a linearization of the power flow equations and, thus, partly redundant. Hence, models will use them interchangeably to a certain extent. Similarly, features based on the

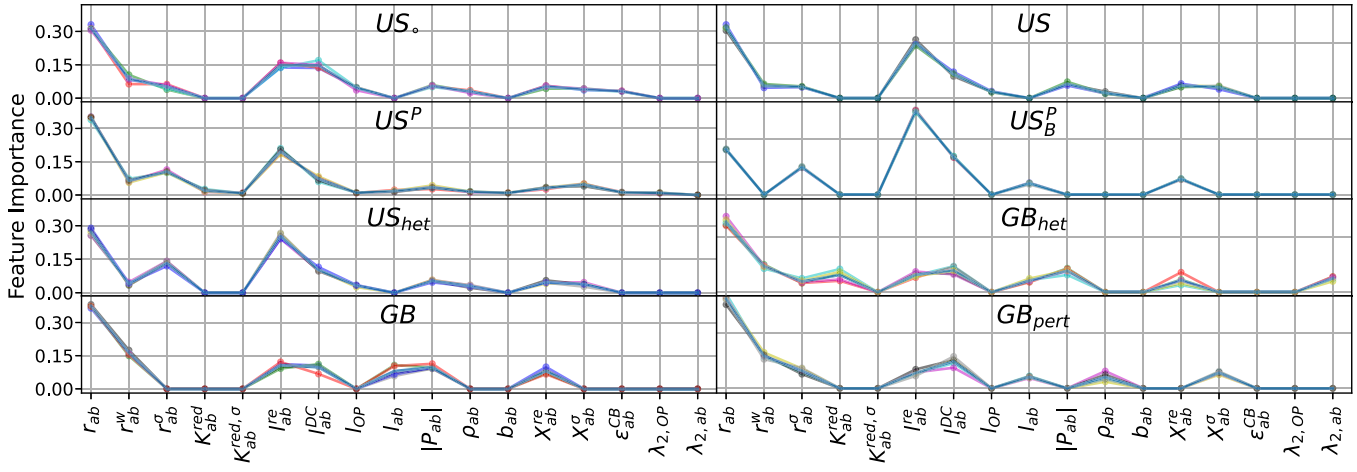

FIG. 7. GBT feature importance for all datasets. Results for five different train-test splits are shown, represented by different coloring. In most cases, they strongly overlap, so differences are barely visible. Due to recursive feature elimination, models have different input features, and no model uses all the features shown. Features not present in a model are represented by a zero feature importance. Redundant capacity ratio and l_{ab}^{DC} and l_{ab}^{re} dominate. Importances mostly fit the single-feature performance shown in Fig. 4. Correlated features compete for importance, so their importance turns out lower than expected from Fig. 4. Notably, ρ_{ab} has an overall lower importance than $|P_{ab}|$ even though it shows much better single-feature performance. Purely topological features barely contribute at all.

TABLE III. To assess the generalizability, models were trained and tested on different datasets. Every dataset was used as a test set once.

Train	US	GB	US_B^P	US	GB_{het}	US_{het}	US	$US_\circ$
Test	$US_\circ$	US	US^P	US_B^P	US_{het}	GB_{het}	GB	GB_{pert}

full redundant capacity and its shortest and widest path variants are, to some degree, redundant and interchangeable.

Another outlier in the correlation between single feature performances and feature importance is ρ_{ab}. Despite being a good univariate predictor, the GBT models barely rely on it. The post-failure cohesiveness ρ_{ab} is a proxy for the maximal line load in the post-failure steady state. With l_{ab}^{DC} and l_{ab}^{re}, though, two features exist that provide a similar proxy with substantially better performance. Hence, models will rather rely on l_{ab}^{DC} and l_{ab}^{re} than on ρ_{ab}.

C. Generalizability

Until now, all models were trained and tested on different subsets of the same data, i.e., the unknown test samples were drawn from the same underlying distribution as the training set. We now go a step further and analyze whether the learning translates between datasets. The rationale of this is the following: A machine learning model will only be deployed in a real-world application if it is expected to show good performance on real-world data, especially in a high-stakes situation. However, critical contingency events are rare, so there is not enough real-world data to test, let alone *train* a machine learning model on. Therefore, in an application case, training data will, by necessity, mostly be made up of synthetic test cases. We now use learning transfer between two different synthetic datasets as a proxy for learning transfer between synthetic and real word cases. Furthermore, a successful learning transfer would indicate that the machine learning model is not specific to one dataset or network topology. This would substantiate the hypothesis that the model and the most important features describe intrinsic physical aspects of network stability and not just statistical correlations.

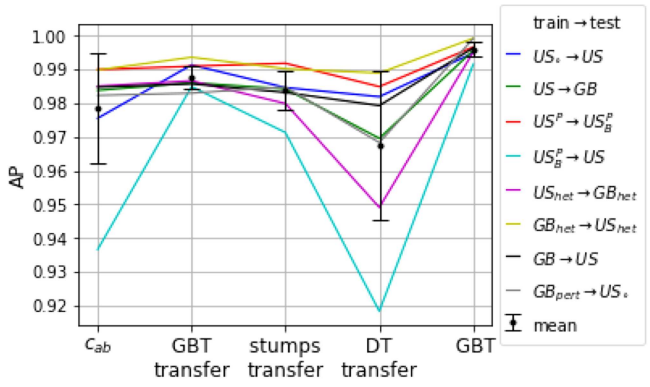

FIG. 8. AP score of the transfer performance for the decision tree (DT), gradient boosted stumps (stumps), and gradient boosted trees (GBT) model with c_{ab} as benchmark. While not as good as the original GBT model, the GBT transfer model clearly and consistently outperforms c_{ab}.

To test the generalizability, datasets were paired so that for every dataset, a transfer model was chosen, see Table III. Figure 8 shows the performance of the different transfer model types with c_{ab} and the original optimal non-transfer GBT model as benchmarks. We find that learning transfer is possible in principle, albeit at a reduced performance. The performance of the GBT and the stumps transfer model lie between the two non-transfer benchmark models: higher than the uni-variate model c_{ab} and lower than the full GBT model. In contrast, the decision tree transfer models do not outperform the c_{ab} model.

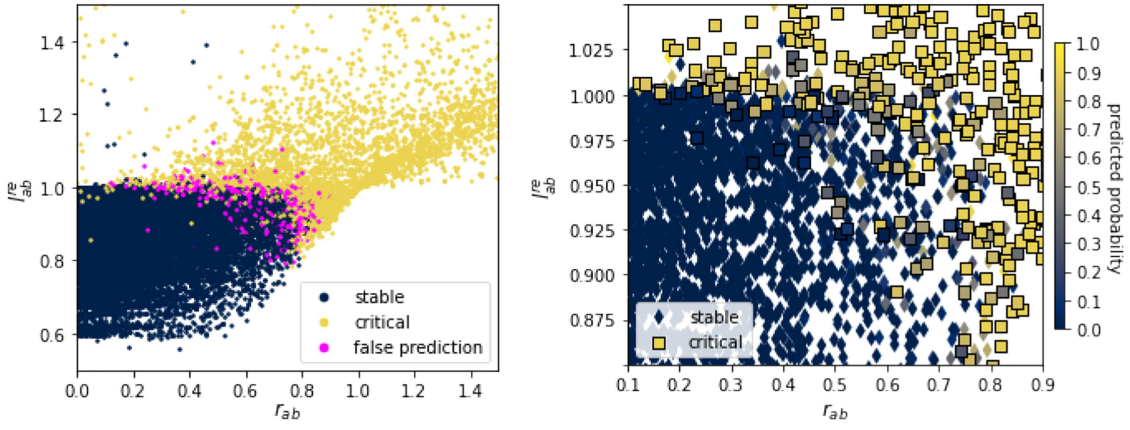

FIG. 9. Relation between key input features and model outcomes for the $US_\circ$ dataset. Left: All samples of the dataset projected into a two-dimensional feature space. Due to the high single-feature performance of r_{ab} and l_{ab}^{re}, stable and critical samples are mostly separated in two dimensions. Most misclassified samples are found in the overlap region. Right: Magnification of the data around the overlap region. The color code shows the probability output of the GBT model. In the overlapping region, the GBT model shows a significant improvement beyond the best prediction achievable by a two-dimensional model.

The transfer performance does, of course, depend on the two datasets. In this study, the choices of respective training sets were arbitrary and not necessarily optimized to enable easy learning transfer. Hence, test and training datasets may have differing topologies, electrical properties, line loads, and different class frequencies. It, thus, stands to reason that a machine learning model trained on data from high-quality simulation models would offer much-improved performance compared to uni-variate models based on single features such as c_{ab}, l^{re}_{ab}, or r_{ab}.

D. Interpreting the ML models

We complete our study with a more detailed look at the relation of the key input features to the GBT model output. The left panel in Fig. 9 is a two-dimensional projection of the $\mathtt{US}_\circ$ dataset, showing the two most important features r_{ab} and l^{re}_{ab}. Those samples that were correctly classified by the machine learning model are colored according to the outcome, while false predictions are colored in red. We find that while the two features already separate critical and stable samples with high accuracy, there is an area of overlap. By drawing on more features, the optimal GBT model is able to reduce errors significantly beyond what would be possible using only those two features (see Fig. 6). Nonetheless, most false predictions of the GBT model occur in this overlap region.

The right panel in Fig. 9 shows a magnification around the overlap region, with a color mapping according to the probability output y of the GBT model. As desired, critical samples generally have a higher probability output than stable samples. The figure elucidates the additional value of using additional features in the GBT model instead of relying on just the two best-performing ones. Both classes overlap in the two-dimensional feature space spanned by r_{ab} and l^{re}_{ab}. Still, the GBT model provides accurate predictions for many samples in this region. The additional features improve the model's predictions, even though especially the best performing among them are very highly correlated to either r_{ab} or l^{re}_{ab}.

V. CONCLUSION AND OUTLOOK

Power grid system stability is an important and timely topic, as large-scale outages can have catastrophic impacts. In this paper, we improved upon known ways of predicting the dynamic stability of a power grid to line failures by combining classical power system tools, graph theory-inspired feature engineering, and interpretable machine learning models. Since the prediction methods applied are computationally cheap, they should be well suited to real-world applications.

Using a second-order Kuramoto model, we simulated line failures for eight classes of synthetic grid models and classified whether they lead to a desynchronization event or not. We assessed the ability of 27 different features that quantify, among others, a line's load, connectivity, redundancy, or centrality to correctly classify dynamic stability outcomes. We then used these features as input for different tree-based machine learning models. After reducing model complexity by performing recursive feature elimination to improve interoperability, we compared the performance of the different models. We found that especially gradient tree boosting models produce very good predictions reaching an average precision (AP) that exceeds 0.996 when average over all datasets.

The engineered features contribute very differently to the models. In particular, we identified two classes of features that have a high predictive power and feature importance. First, a linearization in the spirit of line outage distribution factors allows us to predict the maximum flow in the post-failure grid from the pre-failure state. Second, graph theory provides an upper limit for the ability of a grid to reroute power flows and, thus, a measure of redundancy for each line in the grid.

Features that are not specific to flow networks generally performed very badly and had very low feature importance. This is particularly true for purely topological features including connectivity, centrality, or coreness. These findings are of particular relevance for studies of network robustness in network science, which frequently refer to power grids as potential applications cases. Many classical studies focus on purely topological network properties, see, e.g., Ref. 56–59. Based on our results, we conclude that purely topological metrics do not provide good predictors for the vulnerability of electric power grids, cf. Ref. 60 and 61.

Applications to real-world grids require sufficient training data that can only be obtained from simulations. Hence, it is essential to generalize and transfer learned results from one system to another. To assess generalizability, we tested the machine learning models on different datasets than they were trained on. The GBT model was still able to consistently outperform the best single-feature predictor. We conclude that good performance should be achievable, especially since training data would be engineered to mimic potential critical real-world contingencies, while our datasets were created to cover a wide range of grid properties.

In conclusion, we have demonstrated the potential of integrating machine learning models and network science metrics to assess the robustness of networked systems. A natural extension of our results would be to use larger and more realistic power system models. Our approach could also be applied to other stability risks, such as node failures or perturbations, and types, such as voltage stability, transient overloads, or overload cascades. By training and deploying different models, transmission grid operators should be able to greatly improve their ability to identify potential risks in time, thus allowing early interventions where necessary and ultimately further improving power grid stability.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the German Federal Ministry of Education and Research (BMBF Grant No. 03EK3055B) and the Helmholtz Association via the Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE), Germany.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Maurizio Titz: Conceptualization (equal); Data curation (equal); Investigation (equal); Methodology (equal); Software (equal); Visualization (equal); Writing – original draft (equal); Writing –

review & editing (equal). **Franz Kaiser:** Methodology (equal); Software (equal); Writing – review & editing (equal). **Johannes Kruse:** Methodology (equal); Software (equal); Writing – review & editing (equal). **Dirk Witthaut:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]K. H. LaCommare and J. H. Eto, "Cost of power interruptions to electricity consumers in the United States (US)," Energy **31**, 1845–1855 (2006).
[2]T. Petermann, H. Bradke, A. Lüllmann, M. Poetzsch, and U. Riehm, see https://www.tab-beim-bundestag.de/en/pdf/publications/books/petermann-etal-2011-141.pdf for "What happens during a blackout: Consequences of a prolonged and wide-ranging power outage" (2011), studie des Büros für Technikfolgenabschätzung beim Deutschen Bundestag.
[3]W. Kröger, "Critical infrastructures at risk: A need for a new conceptual approach and extended analytical tools," Reliab. Eng. Syst. Saf. **93**, 1781–1787 (2008).
[4]P. Pourbeik, P. S. Kundur, and C. W. Taylor, "The anatomy of a power grid blackout-root causes and dynamics of recent major blackouts," IEEE Power Energy Mag. **4**, 22–29 (2006).
[5]M. Anvari, G. Lohmann, M. Wächter, P. Milan, E. Lorenz, D. Heinemann, M. R. R. Tabar, and J. Peinke, "Short term fluctuations of wind and solar power systems," New J. Phys. **18**, 063027 (2016).
[6]I. Staffell and S. Pfenninger, "The increasing impact of weather on electricity supply and demand," Energy **145**, 65–78 (2018).
[7]T. Pesch, H.-J. Allelein, and J.-F. Hake, "Impacts of the transformation of the german energy system on the transmission grid," Eur. Phys. J. Spec. Top. **223**, 2561–2575 (2014).
[8]R. A. Rodriguez, S. Becker, G. B. Andresen, D. Heide, and M. Greiner, "Transmission needs across a fully renewable european power system," Renew. Energy **63**, 467–476 (2014).
[9]A. Orths, C. L. Anderson, T. Brown, J. Mulhern, D. Pudjianto, B. Ernst, O. Mark, J. McCalley, and G. Strbac, "Flexibility from energy systems integration: Supporting synergies among sectors," IEEE Power Energy Mag. **17**, 67–78 (2019).
[10]D. Steinberg, D. Bielen, J. Eichman, K. Eurek, J. Logan, T. Mai, C. McMillan, A. Parker, L. Vimmerstedt, and E. Wilson, "Electrification and decarbonization: Exploring us energy use and greenhouse gas emissions in scenarios with widespread electrification and power sector decarbonization," Tech. Rep. [National Renewable Energy Lab.(NREL), Golden, CO (United States), 2017].
[11]A. Guminski, F. Böing, A. Murmann, and S. von Roon, "System effects of high demand-side electrification rates: A scenario analysis for Germany in 2030," Wiley Interdiscip. Rev. Energy Environ. **8**, e327 (2019).
[12]A. E. Motter, S. A. Myers, M. Anghel, and T. Nishikawa, "Spontaneous synchrony in power-grid networks," Nat. Phys. **9**, 191–197 (2013).
[13]F. Dörfler, M. Chertkov, and F. Bullo, "Synchronization in complex oscillator networks and smart grids," Proc. Natl. Acad. Sci. U.S.A. **110**, 2005–2010 (2013).
[14]M. Rohden, A. Sorge, M. Timme, and D. Witthaut, "Self-organized synchronization in decentralized power grids," Phys. Rev. Lett. **109**, 064101 (2012).
[15]A. Berizzi, "The italian 2003 blackout," in *IEEE Power Engineering Society General Meeting, 2004* (IEEE, 2004), pp. 1673–1679.
[16]Union for the Coordination of Transmission of Electricity, see https://www.entsoe.eu/fileadmin/user_upload/_library/publications/ce/otherreports/Final-Report-20070130.pdf for "Final report on the system disturbance on 4 november 2006" (2007).
[17]European Network of Transmission System Operators for Electricity (ENTSO-E), see https://eepublicdownloads.azureedge.net/clean-documents/SOCdocuments/SOCReports/entso-e_CESysSep_Final_Report_210715.pdf for "Continental europe synchronous area separation on 08 January 2021. ics investigation expert panel—Final report" (2021).
[18]European Network of Transmission System Operators for Electricity (ENTSO-E), see https://www.entsoe.eu/news/2021/08/20/outage-of-french-spanish-interconnection-on-24-july-2021-update/ for "Outage of french-spanish interconnection on 24 July 2021—Update" (2021).
[19]N. Balu, T. Bertram, A. Bose, V. Brandwajn, G. Cauley, D. Curtice, A. Fouad, L. Fink, M. G. Lauby, B. F. Wollenberg, and J. N. Wrubel, "On-line power system security analysis," in *Proceedings of the IEEE* (IEEE, 1992), Vol. 80, pp. 262–282.
[20]L. Wehenkel, "Machine learning approaches to power-system security assessment," IEEE Expert **12**, 60–72 (1997).
[21]O. A. Alimi, K. Ouahada, and A. M. Abu-Mahfouz, "A review of machine learning approaches to power system security and stability," IEEE Access **8**, 113512–113531 (2020).
[22]R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke, "Explainable machine learning for scientific insights and discoveries," IEEE Access **8**, 42200–42216 (2020).
[23]J. Machowski, Z. Lubosny, J. W. Bialek, and J. R. Bumby, *Power System Dynamics: Stability and Control* (John Wiley & Sons, 2020).
[24]D. Witthaut, F. Hellmann, J. Kurths, S. Kettemann, H. Meyer-Ortmanns, and M. Timme, "Collective nonlinear dynamics and self-organization in decentralized power grids," Rev. Mod. Phys. **94**, 015005 (2022).
[25]Y. Zhang, L. Wehenkel, P. Rousseaux, and M. Pavella, "Sime: A hybrid approach to fast transient stability assessment and contingency selection," Int. J. Electr. Power Energy Syst. **19**, 195–208 (1997).
[26]H. You, V. Vittal, and X. Wang, "Slow coherency-based islanding," IEEE Trans. Power Syst. **19**, 483–491 (2004).
[27]G. Filatrella, A. H. Nielsen, and N. F. Pedersen, "Analysis of a power grid using a Kuramoto-like model," Eur. Phys. J. B **61**, 485 (2008).
[28]*Power System Coherency and Model Reduction*, edited by Joe H. Chow (Springer, New York, 2013).
[29]T. Nishikawa and A. Motter, "Comparative analysis of existing models for power-grid synchronization," New J. Phys. **17**, 015012 (2015).
[30]J. A. Acebrón, L. L. Bonilla, C. J. P. Vicente, F. Ritort, and R. Spigler, "The Kuramoto model: A simple paradigm for synchronization phenomena," Rev. Mod. Phys. **77**, 137 (2005).
[31]A. Plietzsch, R. Kogler, S. Auer, J. Merino, A. Gil-de Muro, J. Liße, C. Vogel, and F. Hellmann, "Powerdynamics.jl—An experimentally validated open-source package for the dynamical analysis of power grids," SoftwareX **17**, 100861 (2022).
[32]B. Schäfer, M. Matthiae, X. Zhang, M. Rohden, M. Timme, and D. Witthaut, "Escape routes, weak links, and desynchronization in fluctuation-driven networks," Phys. Rev. E **95**, 060203 (2017).
[33]K. Schmietendorf, J. Peinke, and O. Kamps, "The impact of turbulent renewable energy production on power grid stability and quality," Eur. Phys. J. B **90**, 1–6 (2017).
[34]T. Nesti, A. Zocca, and B. Zwart, "Emergent failures and cascades in power grids: A statistical physics perspective," Phys. Rev. Lett. **120**, 258301 (2018).
[35]P. Schultz, J. Heitzig, and J. Kurths, "A random growth model for power grids and other spatially embedded infrastructure networks," Eur. Phys. J. Spec. Top. **223**, 2593–2610 (2014).
[36]G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," Adv. Neural Inform. Process. Syst. **30**, 3146–3154 (2017).
[37]S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," Nat. Mach. Intell. **2**, 56–67 (2020).
[38]D. Fryer, I. Strümke, and H. Nguyen, "Shapley values for feature selection: The good, the bad, and the axioms," IEEE Access **9**, 144352–144360 (2021).
[39]A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The det curve in assessment of detection task performance," Tech. Rep. (National Inst of Standards and Technology Gaithersburg MD, 1997).
[40]D. Witthaut, M. Rohden, X. Zhang, S. Hallerberg, and M. Timme, "Critical links and nonlocal rerouting in complex supply networks," Phys. Rev. Lett. **116**, 138701 (2016).

[41] T. Guler, G. Gross, and M. Liu, "Generalized line outage distribution factors," IEEE Trans. Power Syst. **22**, 879–881 (2007).
[42] U. Brandes and D. Fleischer, "Centrality measures based on current flow," in *Annual Symposium on Theoretical Aspects of Computer Science* (Springer, 2005), pp. 533–544.
[43] M. E. J. Newman, *Networks—An Introduction* (Oxford University Press, Oxford, 2010).
[44] M. Fiedler, "Algebraic connectivity of graphs," Czechoslov. Math. J. **23**, 298 (1973).
[45] F. Dörfler and F. Bullo, "Exploring synchronization in complex oscillator networks," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)* (IEEE, 2012), pp. 7157–7170.
[46] M. E. Newman, "A measure of betweenness centrality based on random walks," Soc. Netw. **27**, 39–54 (2005).
[47] Y. Yang, T. Nishikawa, and A. E. Motter, "Small vulnerable sets determine large network cascades in power grids," Science **358**, eaan3184 (2017).
[48] J. Strake, F. Kaiser, F. Basiri, H. Ronellenfitsch, and D. Witthaut, "Nonlocal impact of link failures in linear flow networks," New J. Phys. **21**, 053009 (2019).
[49] D. J. Klein and M. Randić, "Resistance distance," J. Math. Chem. **12**, 81–95 (1993).
[50] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power Generation, Operation, and Control* (John Wiley & Sons, 2013).
[51] F. Kaiser, V. Latora, and D. Witthaut, "Network isolators inhibit failure spreading in complex networks," Nat. Commun. **12**, 3143 (2021).
[52] H. Ronellenfitsch, D. Manik, J. Hörsch, T. Brown, and D. Witthaut, "Dual theory of transmission line outages," IEEE Trans. Power Syst. **32**, 4060–4068 (2017).
[53] J. Guo, Y. Fu, Z. Li, and M. Shahidehpour, "Direct calculation of line outage distribution factors," IEEE Trans. Power Syst. **24**, 1633–1634 (2009).
[54] D. Manik, M. Rohden, H. Ronellenfitsch, X. Zhang, S. Hallerberg, D. Witthaut, and M. Timme, "Network susceptibilities: Theory and applications," Phys. Rev. E **95**, 012319 (2017).
[55] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications* (Prentice Hall, New Jersey, 1993).
[56] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the North American power grid," Phys. Rev. E **69**, 025103 (2004).
[57] R. Kinney, P. Crucitti, R. Albert, and V. Latora, "Modeling cascading failures in the North American power grid," Eur. Phys. J. B Condens. Matter Complex Syst. **46**, 101–107 (2005).
[58] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the US power grid," Saf. Sci. **47**, 1332–1336 (2009).
[59] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," Nature **464**, 1025–1028 (2010).
[60] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Do topological models provide good information about electricity infrastructure vulnerability?," Chaos **20**, 033122 (2010).
[61] M. Korkali, J. G. Veneman, B. F. Tivnan, J. P. Bagrow, and P. D. Hines, "Reducing cascading failure risk by increasing infrastructure network interdependence," Sci. Rep. **7**, 1–13 (2017).