Active Learning for Power Grid Security Assessment: Reducing Simulation Cost with Informative Sampling

Gašper Leskovec
Jožef Stefan Institute
Slovenia
leskovecg@gmail.com

Costas Mylonas
UBITECH
Greece
kmylonas@ubitech.eu

Klemen Kenda
Jožef Stefan Institute
klemen.kenda@ijs.si

Abstract

Power grid security assessment under the N-1 criterion requires extensive contingency simulations, which are computationally intensive and costly to label. In this work, we explore the use of active learning (AL) to train binary classifiers that can accurately predict the outcome of contingency scenarios using fewer labeled samples. We evaluate several AL strategies, such as entropy, margin, and uncertainty sampling against a random baseline. Our results show that AL methods achieve the same predictive performance with significantly fewer labels, reducing labeling effort and simulator runtime. These findings demonstrate the effectiveness of integrating AL with power system simulators to enable scalable and efficient N-1 security assessment without sacrificing model accuracy.

Keywords

active learning, smart grids, security assessment, simulation cost reduction

ACM Reference Format:
Gašper Leskovec, Costas Mylonas, and Klemen Kenda. 2025. Active Learning for Power Grid Security Assessment: Reducing Simulation Cost with Informative Sampling. In *Proceedings of the Slovenian KDD Conference (SiKDD 2025), October 6th, 2025, Ljubljana, Slovenia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Ensuring secure operation of power systems under the N-1 criterion is a cornerstone of grid reliability. The criterion requires that the system remains within operational limits following the loss of any single component (e.g., line, transformer, or generator). In practice, this involves simulating a large number of contingencies and checking for violations of thermal or voltage constraints. While essential, such simulations are computationally intensive, particularly when performed on high-fidelity grid models, and their interpretation often requires expert judgment. This creates a bottleneck for both real-time applications and large-scale scenario analyses, where scalability and efficiency are important.

Classical approaches to N-1 assessment rely on exhaustive AC power flow simulations combined with contingency ranking heuristics such as performance indices (PIs). While useful for screening, these heuristics may mis-rank contingencies or overlook borderline cases due to masking effects [3]. Moreover, exhaustive analysis does not scale well with system size, making it unsuitable for fast or repeated assessments.

To overcome these challenges, researchers have proposed machine learning (ML) and deep learning (DL) approaches that approximate N-1 contingency outcomes directly from operating point features. One of the earliest contributions in this direction applied convolutional neural networks (CNNs) to contingency datasets, showing that deep models could achieve over 99% accuracy in detecting insecure cases while being more than 200 times faster than traditional power flow calculations [1]. Building on this, more recent work explored pooling-ensemble multi-graph learning to design scalable contingency screening schemes based on steady-state information, demonstrating improved adaptability for large-scale systems [2]. These approaches enable fast security screening without solving power flows for every contingency. However, their reliability hinges on the availability of large labeled datasets covering all relevant operating points and contingencies. Such datasets are typically generated by running exhaustive offline N-1 simulations, which is computationally expensive, or require significant expert effort to label secure versus insecure cases. This dependence on costly and large-scale data generation remains a major limitation of existing ML-based frameworks for steady-state security assessment.

To reduce labeling costs, active learning (AL) has recently been explored in other areas of power systems. For example, authors of [5] used AL to enhance stability assessment and dominant instability mode identification, showing that models could be trained with far fewer labeled samples while maintaining accuracy. Similarly, authors of [4] demonstrated an AL-enhanced digital twin for day-ahead load forecasting, where the model iteratively refined predictions by querying only the most uncertain cases. These studies confirm the potential of AL to reduce expert effort and simulation cost by strategically selecting informative samples. However, AL has not yet been applied to N-1 steady-state security assessment, where the need to cut down on contingency simulations is especially critical.

In this work, we propose a novel framework for active learning-driven N-1 security assessment. Our contributions are threefold:

(1) We design a binary classification model that predicts whether a given contingency is secure or insecure based on steady-state features.

SiKDD 2025, Ljubljana, Slovenia

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

(2) We integrate active learning strategies (entropy, margin, and uncertainty sampling) with the classifier to selectively query the most informative contingencies for simulation, reducing the number of labels required.
(3) We demonstrate through a case study that our approach achieves the same predictive accuracy as fully supervised baselines while reducing simulation cost and labeling effort by up to 40–50%.

This work provides the first evidence that active learning can be directly leveraged for N-1 security assessment, offering a scalable and label-efficient alternative to exhaustive simulation or purely supervised ML approaches.

Although machine-learning and deep-learning approaches are increasingly being used in smart-grid applications, most of them rely on fully supervised training with large labelled datasets. Active learning (AL), which iteratively queries only the most informative samples, has been used to reduce labelling costs in related tasks such as stability assessment and energy-theft detection, but to date it has not been systematically applied to N-1 steady-state security assessment. As a result, exhaustive simulations remain the norm for contingency screening, even though they are expensive and often unnecessary. By addressing this gap, our work demonstrates that informative sampling can achieve comparable predictive performance while substantially lowering simulation overhead.

2 Methodology

The dataset for training and evaluating the proposed framework was generated using a digital twin of the transmission network, where time-series power flow simulations were performed for both base case and N-1 contingency conditions. At each timestamp, load and generation profiles (including renewables) were assigned, and a base-case power flow provided steady-state indicators such as maximum line loading, minimum and maximum bus voltages, and active power injections from loads and generators. The N-1 criterion was then applied by sequentially removing each line, transformer, or generator and re-running the power flow, with the worst-case line loading and bus voltages recorded. An operating point was labeled secure if neither the base case nor any contingency violated standard limits (line loading $>$ 100%, bus voltages outside [0.90, 1.10] p.u.), and insecure otherwise. Non-convergent cases were also marked insecure. The resulting dataset consists of timestamped operating states with base-case and worst-case contingency features, each paired with a binary security label, providing the foundation for classifier training and active learning evaluation.

Table 1: Dataset description

Attribute	Value
Total contingency cases	8 769
Secure / Insecure	51.28% / 48.72%
Features	Line loadings, Bus voltages, Generator and load injections

We trained a Random Forest classifier with an initial labeled set of 100 samples. Active learning proceeded in batches: at each iteration a batch of 50 additional samples was selected using one of the strategies described below. We ran 20 such iterations, yielding roughly 1 100 labeled samples per run (100 initial labels plus 20×50 additional samples).

- **Random**: randomly selected samples (baseline).
- **Entropy**: selects samples with the highest predictive entropy.
- **Margin**: selects samples with the smallest difference between the top two predicted class probabilities.
- **Uncertainty**: selects samples on which the classifier is most uncertain, measured by one minus the maximum class probability.

2.1 Evaluation

We evaluate after each AL iteration on a held-out validation set. The initial labeled pool has n_0=100 samples; each iteration adds a batch of b=50 labels for T=20 iterations (total n_0+Tb). Results are reported as average over $R = 3$ random seeds.

2.2 Metrics

Besides Accuracy and ROC AUC, we use two label-efficiency metrics.

Time-to-Target (TTT).. For a threshold τ,

$$\mathrm{TTT}_\tau \;=\; \min\{\, n \;:\; \mathrm{Acc}(n) \ge \tau \,\}.$$

We also track cumulative simulator runtime required to obtain labels.

3 Results

We first analyze learning curves. As shown in Fig. 1 and Fig. 2, all active learning strategies reach target performance with substantially fewer labels than the random baseline. Figure 1 reports accuracy as a function of labeled samples, while Fig. 2 shows the corresponding ROC AUC.

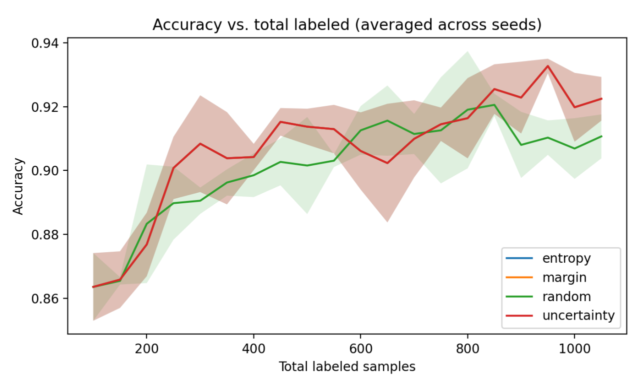

Figure 1: Accuracy vs. total labeled samples (mean across runs).

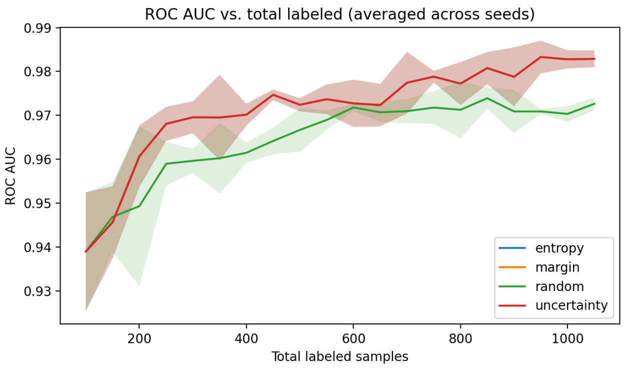

Figure 2: ROC AUC vs. total labeled samples (mean across runs).

Table 2 summarizes key performance indicators (KPIs) at convergence and for predefined accuracy thresholds.

Table 2: Summary of KPIs across strategies

Strategy	Final Accuracy	Final AUC	TTT@0.90	Sim. Time (s)
Random	0.91	0.973	450	2400
Entropy	0.92	0.982	250	4300
Margin	0.92	0.981	250	3600
Uncertainty	0.92	0.983	250	2150

Next, we compare label efficiency and simulation cost. Figure 3 and Fig. 4 plot time-to-target (TTT) for accuracy thresholds 0.90 and 0.92, respectively, while Fig. 5 reports the cumulative simulator runtime. Across thresholds, entropy/margin reach the target after roughly 250 labels, whereas random sampling requires about 450; uncertainty sampling matches their label efficiency and achieves the lowest runtime.

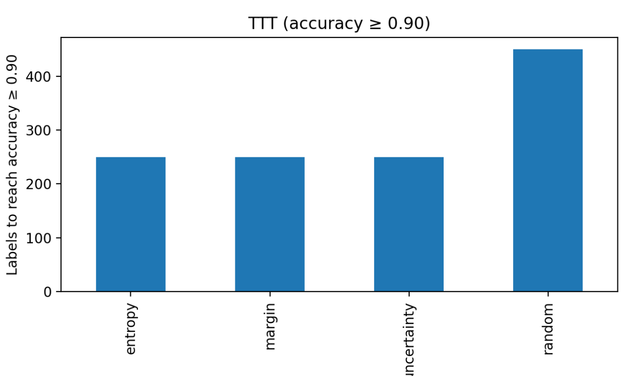

Figure 3: TTT (accuracy ≥ 0.90): labeled samples required to reach the target.

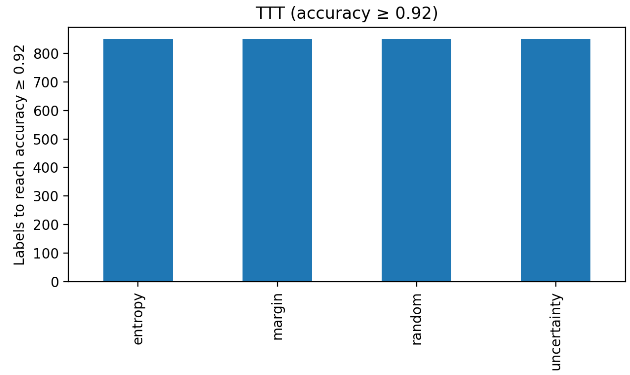

Figure 4: TTT (accuracy ≥ 0.92): labeled samples required to reach the target.

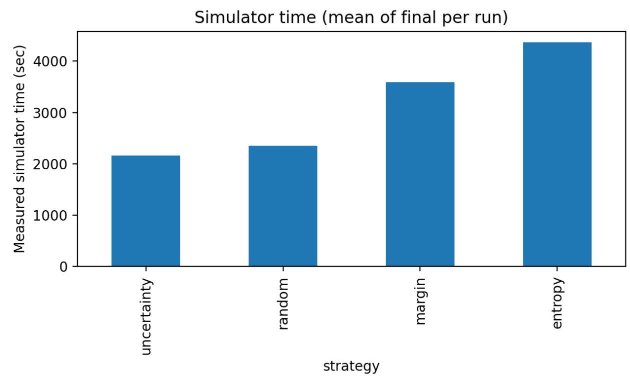

Figure 5: Cumulative simulator runtime across strategies (final mean).

The experiments confirm that all active learning strategies substantially outperform random sampling in terms of label efficiency and simulator cost. In particular, entropy and margin sampling reached an accuracy of 0.90 after roughly 250 labeled samples, whereas random sampling required about 450 samples—a reduction of almost 45% in labeling effort. Uncertainty sampling achieved similar label efficiency but offered the lowest simulator runtime (about 2 150 s compared to about 4 300 s for entropy and about 3 600 s for margin), making it the most cost-effective strategy. Random sampling was consistently less efficient, needing nearly twice as many labels to meet the same accuracy thresholds.

Despite the reduced number of simulator calls, all active learning methods achieved a final validation accuracy of approximately 0.92 and an AUC around 0.98, close to the offline baselines obtained by training on the fully labeled dataset. These findings indicate that integrating active learning with a digital twin simulator can preserve predictive performance while significantly reducing simulation cost.

4 Conclusion

This paper demonstrates that active learning is a viable strategy for reducing simulation costs in power-grid security assessment. By selectively querying informative contingencies, the number of simulator calls can be reduced by roughly 40–50 % without sacrificing predictive accuracy. Fewer simulator calls translate into shorter training times and lower computational and memory requirements, which are particularly important for real-time or resource-constrained applications. Moreover, integrating AL within a digital-twin pipeline enables a feedback loop in which the classifier continuously refines itself using only the most informative contingencies. These findings suggest that exhaustive N-1 simulations are not always necessary for reliable security assessment, paving the way for more scalable and efficient grid-analysis tools.

The present study focuses on a single test system and a Random Forest classifier. In future work we plan to evaluate the proposed framework on larger and more diverse grid topologies (e.g., IEEE 39-bus, 118-bus or national transmission networks) and under varying operating conditions. Another direction is to explore more advanced models such as gradient-boosting machines, deep neural networks or graph neural networks, which may capture complex relationships among grid variables. We also intend to investigate alternative sampling strategies—including diversity-based selection, query-by-committee and Bayesian active learning—to further improve label efficiency. Finally, extending the methodology to multi-contingency (N-k) and dynamic security assessments (e.g., transient stability) will broaden its applicability in future smart-grid deployments.

Acknowledgements

This work was supported by European Union's funded Project HUMAINE [grant number 101120218]. The authors acknowledge the use of LLMs for language optimization. While the LLMs contributed to enhancing efficiency and refining the presentation of this work, all conceptual frameworks, analyses, and interpretations remain the sole responsibility of the authors.

References

[1] José-María Hidalgo Arteaga, Fiodar Hancharou, Florian Thams, and Spyros Chatzivasileiadis. 2019. Deep learning for power system security assessment. In *2019 IEEE Milan PowerTech*. IEEE, 1–6.

[2] Jiyu Huang, Lin Guan, Yinsheng Su, Haicheng Yao, Mengxuan Guo, and Zhi Zhong. 2021. System-scale-free transient contingency screening scheme based on steady-state information: A pooling-ensemble multi-graph learning approach. *IEEE Transactions on Power Systems* 37, 1 (2021), 294–305.

[3] Kip Morison, Lei Wang, and Prabha Kundur. 2004. Power system security assessment. *IEEE power and energy magazine* 2, 5 (2004), 30–39.

[4] Costas Mylonas, Titos Georgoulakis, and Magda Foti. 2024. Facilitating AI and System Operator Synergy: Active Learning-Enhanced Digital Twin Architecture for Day-Ahead Load Forecasting. In *2024 International Conference on Smart Energy Systems and Technologies (SEST)*. IEEE, 1–6.

[5] Zhongtuo Shi, Wei Yao, Yong Tang, Xiaomeng Ai, Jinyu Wen, and Shijie Cheng. 2023. Intelligent power system stability assessment and dominant instability mode identification using integrated active deep learning. *IEEE Transactions on Neural Networks and Learning Systems* 35, 7 (2023), 9970–9984.