Enabling Selective Classification in NN-based Small-Signal Stability Analysis

Galadrielle Humblot-Renaux gegeh@create.aau.dk
Visual Analysis and Perception Lab, Aalborg University & Pioneer Centre for AI, DK
Yang Wu yawu@energy.aau.dk
Department of Energy, Aalborg University, DK
Sergio Escalera sescalera@ub.edu
University of Barcelona and the Computer Vision Center, ES
Thomas B. Moeslund tbm@create.aau.dk
Visual Analysis and Perception Lab, Aalborg University & Pioneer Centre for AI, DK
Xiongfei Wang xiongfei@kth.se
Div. of Electric Power and Energy Systems, KTH Royal Institute of Technology, SE
Heng Wu heng_wu@ieee.org
School of Electrical Engineering, Southeast University, China

Abstract

Neural network (NN)-based analysis methods have the potential to accelerate stability screening of modern power systems, but cannot guarantee accurate and reliable stability predictions for unseen operating scenarios (OSs), posing safety risks. To address this limitation, we propose a selective classification framework leveraging deep ensembles for uncertainty and asymmetric thresholding of predicted probabilities to identify safety-critical misclassifications. These uncertain OSs are then flagged for further analysis using physical-based methods, ensuring safety and robustness. We validate the proposed method both in simulation and on a physical system. This paper is an aggressively abridged version of the interdisciplinary work by [Humblot-Renaux et al., 2025], published in IEEE Transactions on Power Electronics. Code is available at `https://github.com/glhr/ibr-stability-ensemble`.

Keywords: Power electronics, stability, neural networks, uncertainty, selective classification.

1 Introduction

The decarbonization of global energy system accelerates the deployment of renewable energy resources, which are mostly connected to the power grid via power electronic inverters. Those inverter-based resources (IBRs) may interact with one another and with grid dynamics, leading to power system oscillations or even blackout incidents that are increasingly reported in recent years [Wang and Blaabjerg, 2019]. Hence, stability studies are of vital importance for transmission system operators (TSOs) to guarantee the secure and reliable operation.

Thanks to their scalability and computational efficiency, NNs can enable the assessment of all OSs within a reasonable timeframe, and have shown promising results [Chatzivasileiadis et al., 2022, Liao et al., 2024, Li et al., 2024, Zhang

and Xu, 2024]. However, NN-based stability analysis cannot guarantee 100% stability prediction accuracy [Li et al., 2023]. Yet, incorrect stability estimation in certain OSs can be safety-critical in practice and even lead to a blackout event. **In practice, it is crucial to identify OSs where NN predictions cannot be trusted.** Prior work using a single NN for stability assessment of OSs achieves imperfect performance, yet does not consider uncertainty and does not provide a mechanism for identifying NN errors [Zhang and Xu, 2024].

We propose an uncertainty-aware framework to systematically identify OSs with unreliable predictions and flag them for further analysis using physics-based methods. Different from standard selective classification approaches [Pugnana et al., 2024] and recognizing that the false negative (FN) errors (misclassifying a stable OS as unstable) are less critical (as all unstable cases will be re-investigated in subsequent analysis), the proposed dual-thresholding approach prioritizes the identification of safety-critical false positive (FP) errors (misclassifying an unstable OS as stable, which will be ignored in subsequent analysis, but might ultimately lead to blackout events in practice). The combination of the NN ensemble (which reliably assigns high uncertainty to incorrect predictions) and the dual-thresholding approach (which rejects uncertain predictions) allows safety-critical errors to be avoided.

2 System and Dataset Description

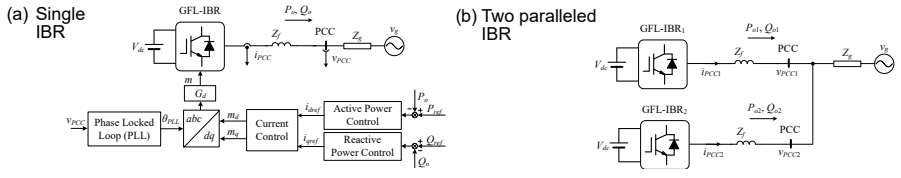

Figure 1: Grid-Following IBR connected to the weak ac system.

Fig. 1 illustrates the single-line diagram of the investigated IBR systems, where the single IBR (Fig. 1a) and two-paralleled IBRs (Fig. 1b) connecting to the weak ac grid are considered. In both cases, the IBRs are operated with the standard grid-following (GFL) control [Wang and Blaabjerg, 2019]. The focus of this work is the small-signal stability analysis of IBR-dominated systems, which is affected by different power flows of each IBR, i.e., different combinations of active/reactive power and PCC voltage [Xie]. Table 1 gives an overview of the generated datasets. We refer to Humblot-Renaux et al. [2025] for details.

Dataset	input dimension d	num. OSs	num. stable / unstable	set
Single (sparse)	3 (V, P, Q)	9,261	3,044 / **6,217**	train/val
Single (dense)	3 (V, P, Q)	3,232,080	1,141,933 / **2,090,147**	test
Parallel	5 (V, P_1, Q_1, P_2, Q_2)	14,406	**12,471** / 1,935	train/val/test

Table 1: Dataset overview

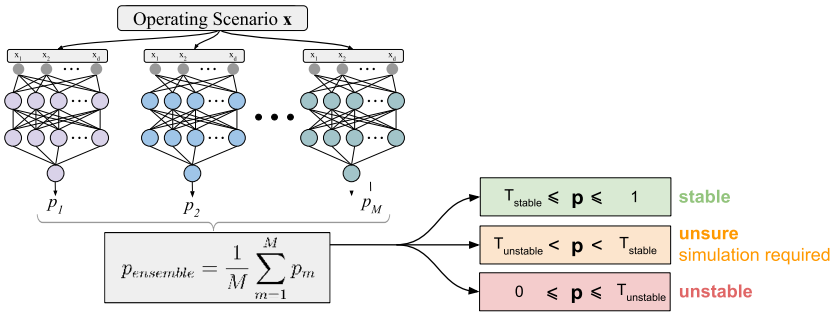

Figure 2: Flowchart showing how an OS is classified based on the predicted p. The proposed NN ensemble and dual-threshold framework utilize the NN ensemble itself to process the stability analysis of **most** OSs with high prediction confidence, while the **remaining small subset** of OSs with low stability prediction confidence from NN are identified for further verification using physics-based stability analysis (like time-domain simulation). The proposed method leverages the strengths of both machine learning and physical-based approaches.

3 Approach

The approach is summarized in Figure 2. The non-linear relationship between OSs and stability is modelled by fully connected layers (2 hidden layers with 64 neurons each) and logistic sigmoid activations. The output is the predicted probability $p \in [0, 1]$ that the OS is stable. Instead of training a single NN [Zhang and Xu, 2024], we leverage deep ensembles [Lakshminarayanan et al., 2017], as they have shown not only to improve predictive performance compared to a single NN, but also to provide reliable uncertainty estimates. Each NN is initialized with a different random seed and trained independently on the full training set $\mathcal{D}_{train}$, thus providing a different plausible solution to the stability learning problem. The intuition is that combining different viewpoints from a group of experts offers more balanced, nuanced predictions than any single expert could provide. During testing, a stability estimate p is obtained by taking the average over the ensemble outputs: $p = \frac{1}{M}\sum_{m=1}^{M} f_{\theta_m}(\mathbf{x})$ where $f_{\theta_m}(\mathbf{x})$ is the output of a single member of the ensemble. The estimate p approaches 0.5 when the disagreement between individual NN increases, or when all NNs' estimates individually approach 0.5. This indicates a high prediction uncertainty that requires further analysis with physical-based methods.

To enable selective classification, we apply two thresholds $T_{unstable}$ and T_{stable} which define the range for which the estimated p is not trusted and should be rejected, as shown in Fig. 2. The rejection rate r is the proportion of OSs in $\mathcal{D}_{test}$ for which $T_{stable} < p < T_{unstable}$. Ideally, r should be as small as possible, but without compromising classification performance. Rejected OSs are excluded from evaluation. For the remaining (non-rejected) OSs, classification performance is evaluated by comparing predicted stability vs. known

small-signal stability in terms of Precision and Recall. Classifying an unstable OS as stable (False Positive) is a safety-critical issue, and must be strictly avoided. The thresholds $T_{unstable} < T_{stable}$ are tuned on a validation set $\mathcal{D}_{val}$ based on a desired rejection rate r_{target} as follows:

1. Find the *highest* possible threshold for which the model achieves 100% Recall on $\mathcal{D}_{val}$. Set this to be $T_{unstable}$.
2. Find the *lowest* possible threshold for which the model achieves 100% Precision on $\mathcal{D}_{val}$. Set this to be T_{stable}.
3. Check the resulting rejection rate r_{val} - that is, the proportion of $\mathcal{D}_{val}$ classified with $T_{unstable} < p < T_{stable}$. If $r_{val} < r_{target}$, increase T_{stable} until $r_{val} = r_{target}\%$. This ensures that at minimum, r_{target} of validation OSs fall between the two thresholds.

4 Results

Figure 3 summarizes the quantitative and qualitative results, considering both a Single NN and Ensemble. With the proposed dual-thresholding approach, the Ensemble can successfully reject **all unstable OSs whose stability predictions are unreliable** (with a rejection rate around 20-21%.), thereby guaranteeing 100% precision on the remaining OSs. On the other hand, while the precision of Single NN is also high under the dual-thresholding approach, there are still safety-critical errors, especially for the Parallel dataset.

	Single NN	Ensemble x100
Single dataset		
Precision	99.962% (±0.143%)	**100% (±0%)**
Recall	97.945% (±4.957%)	**99.216% (±1.418%)**
Rejection	27.327% (±12.590%)	**21.209% (±0.799%)**
Parallel dataset		
Precision	96.889% (±16.575%)	**100% (±0%)**
Recall	96.516% (±16.231%)	**99.862% (±0.018%)**
Rejection	28.732% (±20.875%)	**20.107% (±1.385%)**

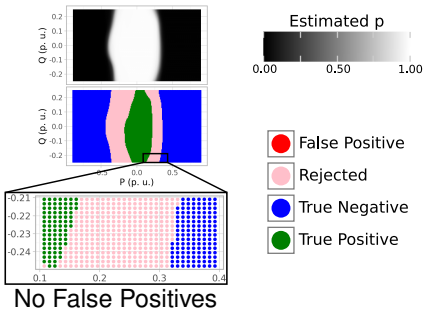

Figure 3: Left: Mean (± std dev) test set classification performance across 30 runs. Right: Ensemble's estimated p (top) and stability analysis after dual-thresholding (bottom) on unseen OSs from a 2D slice of the dense Single dataset ($V = 99.22$). Note that False Positives (red) are safety-critical.

We also validate the approach on a real IBR system, assessing two OSs in the single and parallel set-up [Wu et al., 2025]. These cases are confidently misjudged as stable by the standard single NN (safety-critical), but rejected due to uncertainty by the proposed approach. These results highlight the importance of both reliable uncertainty estimation and effective thresholding to avoid safety-critical prediction errors. We refer to Humblot-Renaux et al. [2025] for experimental details and further analysis.

Acknowledgements

This work was supported by the Danish Data Science Academy, which is funded by the Novo Nordisk Foundation (NNF21SA0069429) and VILLUM FONDEN (40516). It was also supported by the European Union's Horizon 2020 Research and Innovation Programme through the Marie Skłodowska-Curie under Grant 101107634 (PhyDAWN). It was partially supported by the Spanish project PID2022-136436NB-I00 and by ICREA under the ICREA Academia programme. Lastly, thanks to the Pioneer Centre for AI (DNRF grant P1).

References

S. Chatzivasileiadis, A. Venzke, J. Stiasny, and G. Misyris. Machine learning in power systems: Is it time to trust it? *IEEE Power Energy Mag.*, 20(3): 32–41, 2022. ISSN 1540-7977. doi: 10.1109/MPE.2022.3150810.

G. Humblot-Renaux, Y. Wu, S. Escalera, T. B. Moeslund, X. Wang, and H. Wu. Uncertainty-aware stability analysis of ibr-dominated power system with neural networks. *IEEE Transactions on Power Electronics*, pages 1–6, 2025. doi: 10.1109/TPEL.2025.3560236.

B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

B. Li, P. Qi, B. Liu, S. Di, J. Liu, J. Pei, J. Yi, and B. Zhou. Trustworthy ai: From principles to practices. *ACM Comput. Surv.*, 55(9), 2023. ISSN 0360-0300. doi: 10.1145/3555803.

Y. Li, Y. Liao, L. Zhao, M. Chen, X. Wang, L. Nordström, P. Mittal, and H. Vincent Poor. Machine learning at the grid edge: Data-driven impedance models for model-free inverters. *IEEE Trans. Power Electron.*, 39(8):10465–10481, 2024. doi: 10.1109/TPEL.2024.3399776.

Y. Liao, Y. Li, M. Chen, L. Nordström, X. Wang, P. Mittal, and H. V. Poor. Neural network design for impedance modeling of power electronic systems based on latent features. *IEEE Trans. Neural Netw. Learn. Syst.*, 35(5):5968–5980, 2024. doi: 10.1109/TNNLS.2023.3235806.

A. Pugnana, L. Perini, J. Davis, and S. Ruggieri. Deep neural network benchmarks for selective classification. *Journal of Data-centric Machine Learning Research*, 2024. URL `https://openreview.net/forum?id=xDPzHbtAEs`.

X. Wang and F. Blaabjerg. Harmonic stability in power electronic-based power systems: Concept, modeling, and analysis. *IEEE Trans. Smart Grid*, 10(3): 2858–2870, 2019. doi: 10.1109/TSG.2018.2812712.

Y. Wu, H. Wu, L. Cheng, J. Zhou, Z. Zhou, M. Chen, and X. Wang. Impedance profile prediction for grid-connected vscs with data-driven feature extraction. *IEEE Transactions on Power Electronics*, 40(2):3043–3061, 2025. doi: 10.1109/TPEL.2024.3495214.

X. Xie. Analysis and control of oscillatory stability in renewable power systems: China's experience. URL `https://resourcecenter.ieee-pels.org/education/webinars/pelsweb041524v`.

M. Zhang and Q. Xu. Deep neural network-based stability region estimation for grid-converter interaction systems. *IEEE Trans. Ind. Electron.*, 71(10):12233–12243, 2024. doi: 10.1109/TIE.2024.3355525.