Aalborg Universitet

AALBORG UNIVERSITY

Uncertainty-Aware Stability Analysis of IBR-dominated Power System with Neural Networks

Humblot-Renaux, Galadrielle; Wu, Yang; Escalera Guerrero, Sergio; Moeslund, Thomas B.; Wang, Xiongfei; Wu, Heng

Published in:
I E E E Transactions on Power Electronics

DOI (link to publication from Publisher):
10.1109/TPEL.2025.3560236

Publication date:
2025

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Humblot-Renaux, G., Wu, Y., Escalera Guerrero, S., Moeslund, T. B., Wang, X., & Wu, H. (2025). Uncertainty-Aware Stability Analysis of IBR-dominated Power System with Neural Networks. *I E E E Transactions on Power Electronics*, *40*(11), 16055-16060. https://doi.org/10.1109/TPEL.2025.3560236

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: 11 07, 2026

Uncertainty-Aware Stability Analysis of IBR-dominated Power System with Neural Networks

Galadrielle Humblot-Renaux, Yang Wu, *Member, IEEE*, Sergio Escalera, *Senior Member, IEEE*, Thomas B. Moeslund, Xiongfei Wang, *Fellow, IEEE*, and Heng Wu, *Senior Member, IEEE*

***Abstract*—Machine learning (ML) technologies have significant potential in accelerating stability screening of modern power systems that are dominated by inverter-based resources (IBRs). Nonetheless, neural network (NN)-based analysis methods cannot guarantee accurate and reliable stability predictions for unseen operating scenarios (OSs), posing safety risks. To address this limitation, this letter proposes an approach combining neural network ensembles with a dual-thresholding framework, which enables the reliable identification of OSs where ML predictions may fail. These uncertain OSs are then flagged for further analysis using physical-based methods, ensuring safety and robustness. The effectiveness of the proposed method is verified by simulation and experimental test.**

***Index Terms*—Stability, inverter-based resources, machine learning, uncertainty estimation.**

I. 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 [1]. Hence, stability studies are of vital importance for transmission system operators (TSOs) to guarantee the secure and reliable operation of IBR-dominated power systems.

The methodology for assessing the stability of IBR-dominated power system under a single operating scenario (OS) is well developed [1]. However, a large power system with thousands of IBRs can have more than 1 billion OSs [2]. Such a high number of OSs makes it impossible for TSOs to carry out in-depth stability studies for each specific OS. Therefore, stability analysis of IBR-dominated power system with extremely high number of OSs has become one of the most challenging tasks for TSOs, as identified in a recent survey from IEEE task force [3].

In recent years, many research efforts have been devoted to leveraging data-driven machine learning (ML) approaches to analyze the stability of the IBR-dominated power system. In particular, neural networks (NNs) have shown promising results for stability analysis [4]–[7]. Thanks to their scalability and computational efficiency, NNs can enable the assessment of all OSs within a reasonable timeframe. However, the critical disadvantage of ML-based stability analysis is that it cannot guarantee 100% stability prediction accuracy [8]. Yet, incorrect stability estimation in certain OSs can be safety-critical in practice and even lead to a blackout event.

A potential solution for such a problem is adopting physical-based approaches (e.g. the impedance-based stability analysis, EMT simulations [1]) to analyze the stability of OSs where ML-based predictions are likely to fail. However, this solution relies on the reliable identification of OSs where ML-based predictions cannot be trusted. This is challenging in the conventional NN-based framework, because a single NN cannot capture uncertainty in its parameters (i.e. *epistemic* uncertainty) and thus cannot provide reliable uncertainty estimation beyond the training data [9]. Prior work using a single NN for stability assessment of OSs achieves imperfect classification performance, yet does not consider uncertainty and does not provide a mechanism for identifying NN errors [7].

To tackle this challenge, this letter is the first to apply deep ensembles [10] in the stability analysis of IBR-dominated power system. In contrast to a single NN whose predictions are sensitive to initialization and prone to being confidently incorrect, a deep ensemble aggregates multiple NNs, each with a different set of parameters. This approach not only provides more robust and stable predictions by averaging the outputs of multiple NNs, but is also a well-established approach for uncertainty estimation in the machine learning literature [11], [12]. Compared to Bayesian NNs, deep ensembles are simple, scalable, architecture-agnostic and do not modify the training procedure. Compared to MC-Dropout [13], they offer state-of-the-art performance but at a greater computational cost, due to the need to train multiple NNs.

While previous studies also adopt NN ensembles in other power system applications, e.g., load forecasting [14], [15], their primary aim is to improve predictive performance. They do not incorporate mechanisms to identify OSs where NN-based predictions may fail and take further actions. In contrast, this work proposes an uncertainty-aware framework to

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). Lastly, it was partially supported by the Spanish project PID2022-136436NB-I00 and by ICREA under the ICREA Academia programme. *(Corresponding author: Heng Wu)*

Galadrielle Humblot-Renaux and Thomas B. Moeslund are with the Visual Analysis and Perception Lab, Aalborg University, 9200 Aalborg, Denmark. (e-mail: gegeh@create.aau.dk; tbm@create.aau.dk)

Yang Wu and Heng Wu are with the Department of Energy, Aalborg University, 9220 Aalborg Denmark (e-mail: yawu@energy.aau.dk; hew@energy.aau.dk).

Sergio Escalera is with the University of Barcelona and the Computer Vision Center, Spain, and also with the Visual Analysis and Perception Lab, Aalborg University, 9200 Aalborg, Denmark. (email: sescalera@ub.edu).

Xiongfei Wang is with the Division of Electric Power and Energy Systems, KTH Royal Institute of Technology, 10044 Stockholm, Sweden, and also with the Department of Energy, Aalborg University, 9220 Aalborg, Denmark (e-mail: xiongfei@kth.se).

systematically identify OSs with unreliable predictions and flag them for further analysis using physics-based methods. Furthermore, different from standard selective classification approaches [16] 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). This strategy strictly minimizes FP errors without rejecting an impractical high number of OSs. 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. Finally, simulation and experimental test results verify the effectiveness of the proposed method.

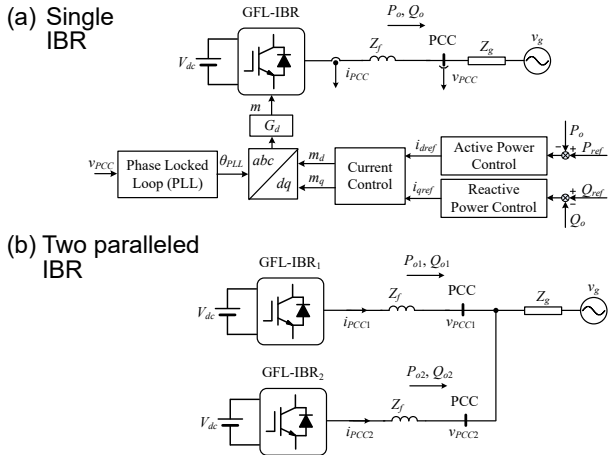

Fig. 1. Grid-Following IBR connected to the weak ac system.

II. System Description

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 [1], where the current control loop is used to regulate the output current of IBR to track its current references generated by the active/reactive power control. The phase-locked loop (PLL) is used for grid synchronization. Z_f and Z_g represent the filter and grid impedance, respectively. V_{PCC} and V_g represent the voltage at the point of common coupling (PCC) and grid voltage, respectively. 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 [2]. Therefore, the OS for a single IBR can be prescribed by a 3-dimensional vector (P_1, Q_1, V_{PCC}), while the OS for two paralleled IBRs can be prescribed by 5-dimensional vector (P_1, Q_1, P_2, Q_2, V_{PCC}).

III. Approach

A. Basic neural network

First, the basic NN used for stability analysis in previous literature is reviewed [7]. The input of the NN is the OS that is characterized by a d-dimensional vector $\mathbf{x} \in \mathbb{R}^d$ ($d = 3$ for single IBR, $d = 5$ for two parallel IBRs). The nonlinear relationship between OSs and stability is modelled by fully connected layers (2 hidden layers with 64 neurons each) and logistic sigmoid activations. The output of the NN is the predicted probability $p \in [0, 1]$ that the OS is stable, which is compared to a binary stability label $y \in \{0, 1\}$ (1 for stable and 0 for unstable). In practice, NN parameters θ are first initialized randomly. They are then optimized using Stochastic Gradient Descent by minimizing the negative Bernoulli log-likelihood of the training data $\mathcal{D}_{train}$: $\hat{\theta} = \arg\min_\theta - \sum_{n=1}^{N} [y_n \log(p_n) + (1 - y_n) \log(1 - p_n)]$. It is important to note that while the basic NN's output p is probabilistic, **it is not a reliable measure of uncertainty** [4].

B. Ensemble of neural networks

Instead of training a single NN [7], we propose to leverage deep ensembles [10], as they have shown not only to improve predictive performance compared to a single NN, but also to provide reliable uncertainty estimates. The intuition behind deep ensembles is that combining different viewpoints from a group of experts offers more balanced, nuanced predictions than any single expert could provide.

As illustrated in Fig. 2b, an ensemble consists of M **distinct** NNs, each with their own set of learned parameters θ_m for $m = 1, 2, \ldots, M$. To obtain different parameters θ_m across the ensemble, each NN is initialized with a different random seed before training. The NNs are then trained independently on the full training set $\mathcal{D}_{train}$. Each NN in the ensemble provides a different plausible solution to the stability learning problem.

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. In contrast, p approaching 1 (0) indicates a stable

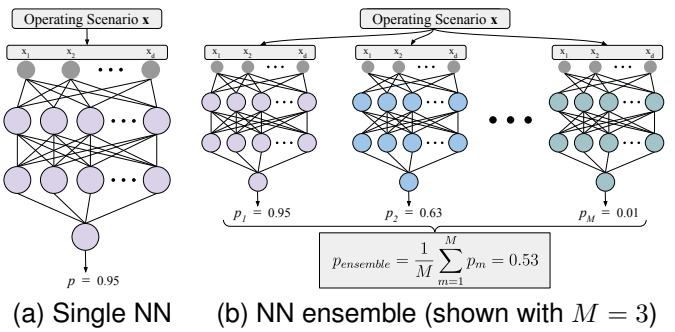

Fig. 2. Default (a) vs. proposed (b) NN architecture for stability estimation, illustrated with example outputs (arbitrary values). Each color represents a different set of trained NN parameters.

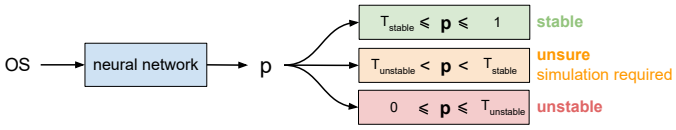

Fig. 3. Flowchart showing how an OS is classified based on the predicted p. In case of low confidence, the model abstains.

(unstable) prediction with a low uncertainty. In practice, ensemble members are likely to agree for OSs in $\mathcal{D}_{train}$, but may not agree for OSs which were not seen during training.

C. Classification and evaluation

A held-out test set $\mathcal{D}_{test}$ is used to evaluate the performance of the final trained model. For this, a discrete class (*stable* or *unstable*) must be extracted from the estimate p. Typically, a single fixed threshold of 0.5 is used such that $p \geq 0.5$ indicates stable, and $p < 0.5$ indicates unstable [7]. However, this method cannot identify OSs whose stability prediction cannot be trusted. To tackle this challenge, the dual-thresholding approach is proposed in this work, where $T_{unstable}$ and T_{stable} are two thresholds that define a range for which the estimated p is not trusted and should be rejected, as shown in Fig. 3. 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 $\text{Precision} = \frac{TP}{TP+\mathbf{FP}}$ and $\text{Recall} = \frac{TP}{TP+\mathbf{FN}}$, where TP, FP, and FN are True Positives, False Positives and False Negatives, respectively. Classifying an unstable OS as stable (FP) is a safety-critical issue, and must be strictly avoided. Therefore, the two thresholds will be carefully tuned to maximize the Precision, which will be detailed in the following.

D. Choice of classification thresholds

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.

Note that performance on $\mathcal{D}_{val}$ does not necessarily reflect the performance on $\mathcal{D}_{test}$. Intuitively, the value for T_{stable} is thus chosen more conservatively than $T_{unstable}$, since it is safety-critical to achieve high Precision.

E. Overall stability analysis methodology

As shown in Fig. 3, 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, and thus, improving the computational efficiency without jeopardizing the overall stability prediction accuracy.

F. Scalability

The proposed framework can be generalized to large-scale IBR-dominated power systems, due to the following reasons:

1) Data generation: the power flow calculation in bulk IBR-dominated power systems is very mature and is widely supported by commercial software such as DIgSILENT PowerFactory and PSS/E.
2) NN ensemble: The feasibility of training the single NN for stability analysis in bulk IBR-dominated power systems has been extensively investigated [17]. Since an NN ensemble is simply a collection of multiple independent NNs (and allows for bigger or alternative NN architectures to be used if needed), training the ensemble does not introduce additional complexity or constraints beyond what has already been demonstrated for the single NN in bulk power systems. The training cost scales linearly with M - thus, there is a tradeoff between computational efficiency and reliable uncertainty estimation. However, since the NNs are fully independent, they can be trained in parallel, thus greatly reducing the training time. Similarly, after training, the predictions of the individual NNs can be obtained in parallel.

IV. Case Study

A. Datasets and splits

The proposed approach is validated on two datasets corresponding to the single and paralleled IBRs, as summarized in Table I. The datasets consist of OS values and their associated stability label, obtained via small-signal stability analysis in simulation models of grid-connected converters [18]. A grid condition with short-circuit-ratio (SCR) of 2.4 is applied to testify the method under weak grid conditions.

For single IBR, a *sparse* and a *dense* version of datasets are generated for fine-grained evaluation along the stability boundaries (see Fig. 4). For *Single (sparse)*, V_{PCC} is set as 0.9: 0.01: 1.1. P is set as -1: 0.1: 1, while Q is set as -0.3: 0.03: 0.3, all in per unit (p.u.). These ranges are set to balance the proportion of stable and unstable cases. For *Single (dense)*, the sampling intervals of V_{PCC}, P, and Q are shrunk to 0.1

TABLE I
Dataset Overview

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

times the original values. The OSs present in *Single (sparse)* are removed from *Single (dense)* to avoid any overlap.

For paralleled IBRs, V_{PCC} is set as above. P ranges in [-0.2, -0.06, -0.03, 0, 0.03, 0.06, 0.2] and Q ranges in [-0.1, -0.03, -0.01, 0, 0.01, 0.03, 0.1] for each inverter.

Several proportions (20%, 50% and 80%) of the *Single (sparse)* and the *Parallel* datasets are selected as training sets via random splitting. For the Parallel case, the remaining portion is further randomly split into 20% validation, 80% test. For the Single case, *Single (dense)* is used as a test set. This allows for a very fine-grained evaluation in unseen regions along the stability boundary where prediction errors are most likely to occur.

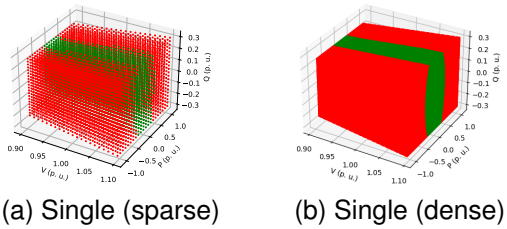

Fig. 4. OSs in the two versions of the Single dataset, color-coded by stability (red: unstable, green: stable).

B. Model comparison

The **single NN** (as described in Sec. III-A) is considered as a baseline and is compared to the **Ensemble x10** and **Ensemble x100** (NN ensemble with $M = 10$ and $M = 100$, as described in Sec. III-B). The same architecture and hyper-parameters are used for all models. NN parameters are optimized with Adam [19] for a maximum of 1000 epochs with an initial learning rate of 10^{-1} and a batch size of 200. Following standard practice, the input data is standardized based on $\mathcal{D}_{train}$ statistics. The proposed thresholding approach is applied with $r_{target} = 20\%$ (a conservative choice) and is compared to the standard approach [7] of having a single threshold at 0.5 and no rejected samples. To ensure that the performance of the approach is consistent, for every dataset and training set size, training/threshold tuning/ testing are repeated with 10 different random seeds, to account for the stochasticity in parameter initialization, dataset shuffling and random data splits. This results in 30 evaluation runs per dataset. Note that all reported performance is on *unseen* OSs, as there is no overlap between train, val and test sets.

V. Results

A. Quantitative evaluation

Tab. II summarizes the quantitative results of the stability prediction based on the proposed dual-thresholding approach (Tab. IIb) and the standard single-thresholding approach adopted in prior work [7] (Tab. IIa), considering both Single NN and Ensembles. In the single threshold scenario, although the Ensemble improves the prediction accuracy compared to Single NN, all of the models make safety-critical errors on both datasets (the precision is less than 100%). However, it can be observed from Tab. IIb that with the proposed

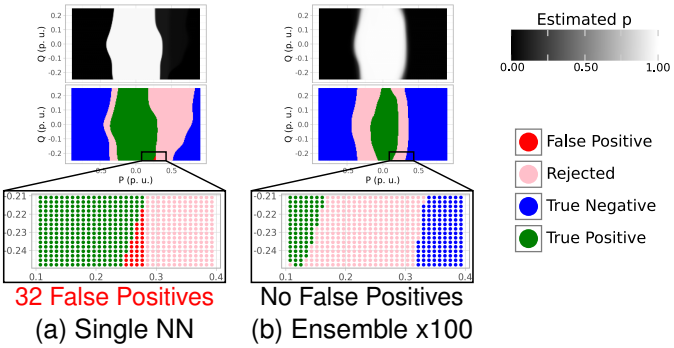

Fig. 5. Estimated p (top) and stability analysis after dual-thresholding (bottom) on unseen OSs from a 2D slice of the dense Single dataset ($V = 99.22$). Note that False Positives (red) are safety-critical.

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, and thus, the safety-critical prediction errors are avoided. On the other hand, while the precision of Single NN is also increased under the dual-thresholding approach, there are still safety-critical errors, especially for the Parallel dataset, as shown in Tab. IIb. These results highlight the importance of adopting both Ensemble and dual-thresholding approaches to avoid safety-critical prediction errors. Comparing Ensemble x10 and Ensemble x100, a higher M ensures 100% precision across all runs and a slightly lower rejection rate for both the Single and Parallel case.

Fig. 5 further provides a visualized example of the stability prediction with the proposed dual-thresholding approach. Compared to the Single NN, the Ensemble x100 exhibits higher and more consistent uncertainty along the decision boundary due to disagreement across ensemble members, with smoother transitions between low and high uncertainty regions. It does not make any FP predictions thanks to the reliable uncertainty estimation. In contrast, the Single NN makes 32 safety-critical FP predictions which cannot be intercepted. This shows that while the dual-thresholding approach can be applied to a Single NN, its effectiveness requires a reliable estimation of p, which the Single NN does not provide.

B. Computational cost

Using a consumer laptop, the Ensemble x100 takes 1 minute to train on the *Single (sparse)* dataset and 1.5 minutes on the *Parallel* dataset, with the NNs in the Ensemble being trained sequentially. Parallelization reduces the training time to 16 seconds and 23 seconds, respectively. After training, the Ensemble x100 is compared to EMT simulation in PSCAD in terms of computation speed on the same hardware. Results are shown in Table III. It can be seen that Ensemble x100 is significantly faster for stability analysis ($10^8\times$ faster in the Parallel case). Assuming a rejection rate of 20% by the Ensemble, the stability analysis can be sped up by a factor of 5. As a concrete example, given the paralleled IBR system and a set of 10,000 OSs to assess, stability assessment using EMT

TABLE II
MEAN (± STANDARD DEVIATION) CLASSIFICATION PERFORMANCE ON THE TEST SET ACROSS 30 RUNS. BEST RESULTS ARE HIGHLIGHTED IN BOLD.

	Single NN	Ensemble x10	Ensemble x100
Single dataset			
Precision	95.119% (±17.978%)	99.111% (±0.385%)	**99.229% (±0.340%)**
Recall	92.590% (±17.602%)	97.091% (±0.565%)	**97.157% (±0.444%)**
Rejection Rate	0%	0%	0%
Parallel dataset			
Precision	98.096% (±3.109%)	99.159% (±0.349%)	**99.195% (±0.317%)**
Recall	99.172% (±0.445%)	99.588% (±0.212%)	**99.707% (±0.203%)**
Rejection Rate	0%	0%	0%

(a) Single threshold of 0.5 (no rejected OSs).

	Single NN	Ensemble x10	Ensemble x100
Single dataset			
Precision	99.962% (±0.143%)	**100% (±0%)**	**100% (±0%)**
Recall	97.945% (±4.957%)	99.209% (±1.211%)	**99.216% (±1.418%)**
Rejection Rate	27.327% (±12.590%)	21.354% (±0.987%)	**21.209% (±0.799%)**
Parallel dataset			
Precision	96.889% (±16.575%)	100% (±0.003%)	**100% (±0%)**
Recall	96.516% (±16.231%)	99.856% (±0.020%)	**99.862% (±0.018%)**
Rejection Rate	28.732% (±20.875%)	20.463% (±4.789%)	**20.107% (±1.385%)**

(b) Thresholds $T_{unstable}$ and T_{stable} tuned on $\mathcal{D}_{val}$.

TABLE III
THROUGHPUT (NUMBER OF ESTIMATES PER SECOND).

	Single IBR ($d = 3$)	Paralleled IBR ($d = 5$)
Simulation (ground truth)	1.553 OSs / sec	0.123 OSs / sec
Ensemble x100	1.083 $\times 10^7$ OSs / sec	1.113 $\times 10^7$ OSs / sec
Combination (20% Simulation)	7.765 OSs / sec	0.615 OSs / sec

TABLE IV
CASES FOR THE EXPERIMENTS

	Single IBR ($d = 3$)	Paralleled IBR ($d = 5$)
OS (V, P1, Q1, P2, Q2)	99.22, 830.47, -744.15	121, 1089, 544.5, 0, 544.5
Ensemble x100	Unsure ($p = 0.324$)	Unsure ($p = 0.783$)
Single NN	Stable ($p = 0.978$)	Stable ($p = 0.999$)

simulation takes 81,301 seconds (approximately 22.6 hours). Using our approach (with only around 20% of OSs processed by simulation), the computation time (including the training) reduces to 16,283 seconds (approximately 4.5 hours).

C. Experiment validation

Experiments have been conducted to prove the superiority of the proposed method. Two Danfoss converters are controlled in GFL mode, connected to the grid simulator Chroma 61845. Detailed description of the setup can be found in [20].

Two cases from the single and paralleled converters have been selected, with parameters in Table IV. These cases are confidently misjudged as stable by traditional single NN (safety-critical), but rejected due to uncertainty by the proposed ensemble NN. Experiment waveforms are shown in Fig. 6 (the OS is changed by varying the reactive power to switch the system from stable case to unstable case), where unstable resonances can be observed. It should be noted that the studied cases are critically unstable, very close to the stability boundary. Although the oscillations are not very large due to inherent damping (parasitic resistance) in the hardware platform, a prominent increase of system resonance can still be observed in the experiments, supporting the effectiveness of the proposed method.

VI. CONCLUSION

This letter represents the first application of NN ensembles in the stability analysis of IBR-dominated power system.

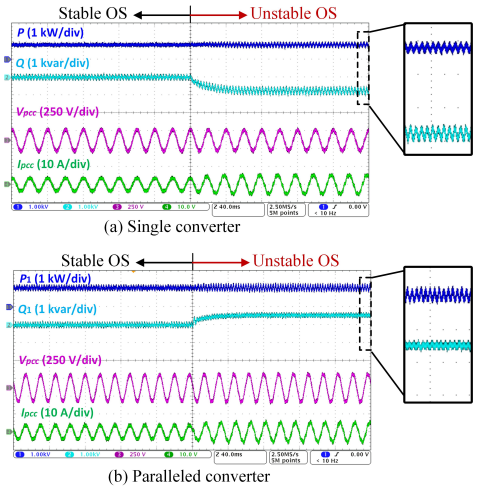

Fig. 6. Experiment waveforms for single and paralleled converters.

Complemented by the proposed dual-thresholding approach, the framework offers robust and reliable uncertainty estimates for stability predictions. Simulation and experimental tests verify the effectiveness of the proposed method.

REFERENCES

[1] X. Wang and F. Blaabjerg, "Harmonic stability in power electronic-based power systems: Concept, modeling, and analysis," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2858–2870, 2019.
[2] X. Xie, "Analysis and control of oscillatory stability in renewable power systems: China's experience." [Online]. Available: https://resourcecenter.ieee-pels.org/education/webinars/pelsweb041524v
[3] H. Wu, F. Zhao, and X. Wang, "A survey on impedance-based dynamics analysis method for inverter-based resources," *IEEE Power Electron. Mag.*, vol. 10, no. 3, pp. 43–51, 2023.
[4] S. Chatzivasileiadis, A. Venzke, J. Stiasny, and G. Misyris, "Machine learning in power systems: Is it time to trust it?" *IEEE Power Energy Mag.*, vol. 20, no. 3, pp. 32–41, 2022.
[5] 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.*, vol. 35, no. 5, pp. 5968–5980, 2024.
[6] 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.*, vol. 39, no. 8, pp. 10 465–10 481, 2024.
[7] M. Zhang and Q. Xu, "Deep neural network-based stability region estimation for grid-converter interaction systems," *IEEE Trans. Ind. Electron.*, vol. 71, no. 10, pp. 12 233–12 243, 2024.

[8] 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.*, vol. 55, no. 9, 2023.

[9] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods," *Machine Learning*, vol. 110, pp. 457 – 506, 2019.

[10] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[11] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, M. Shahzad, W. Yang, R. Bamler, and X. X. Zhu, "A survey of uncertainty in deep neural networks," *Artif. Intell. Rev.*, vol. 56, no. Suppl 1, p. 1513–1589, Jul. 2023. [Online]. Available: https://doi.org/10.1007/s10462-023-10562-9

[12] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. Dillon, B. Lakshminarayanan, and J. Snoek, "Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/8558cb408c1d76621371888657d2eb1d-Paper.pdf

[13] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *Proceedings of The 33rd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 1050–1059. [Online]. Available: https://proceedings.mlr.press/v48/gal16.html

[14] Z. Cao, C. Wan, Z. Zhang, F. Li, and Y. Song, "Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1881–1897, 2019.

[15] Y. Yang, W. Hong, and S. Li, "Deep ensemble learning based probabilistic load forecasting in smart grids," *Energy*, vol. 189, p. 116324, 2019.

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

[17] Y. Du, F. Li, J. Li, and T. Zheng, "Achieving 100x acceleration for n-1 contingency screening with uncertain scenarios using deep convolutional neural network," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3303–3305, 2019.

[18] B. Wen, D. Boroyevich, R. Burgos, P. Mattavelli, and Z. Shen, "Analysis of d-q small-signal impedance of grid-tied inverters," *IEEE Trans. Power Electron.*, vol. 31, no. 1, pp. 675–687, 2016.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

[20] 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 Trans. Power Electron.*, Early Access.