WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Multi-Scenario Validation of Voltage Stability Indices and Machine-Learning-Based Weak Line Detection on the IEEE 14-Bus System

Rowtu Venkatesh[1], Hemanth Kumar Chappa[2], Vempadapu Divya[3], Mudunuru Shanmukha Varma [4],And Dr.T.S.Kishore(Corresponding Author) [5]

Email:rowtuvenkatesh@gmail.com

Department of Electrical and Electronics Engineering, GMR Institute of Technology, Rajam, Andhra Pradesh, India

ABSTRACT—

Voltage stability indices (VSIs) are widely used to flag the transmission line most likely to trigger voltage collapse, yet published indices are usually validated at a handful of loading snapshots along a single trajectory. This paper audits twelve line VSIs on the IEEE 14bus system using continuation power flow (CPF) over 30 scenarios (a base case, 19 successful single-line outages and 10 randomized loading directions), each traced to its collapse point and yielding 31,660 line-step samples. Indices are scored on saturation, trend correlation with the loading factor, and single-line bias. We show analytically, and verify on 30,693 in-service samples, that three pairs of published formulas (BVSI/Lp, Lij/Lmn and VQILine/FVSI) are identical, and that the R-multiplied pair BVSI/Lp is structurally blind to the five zero-resistance transformer branches, including the heavily loaded branch 5–6. The Novel Voltage Stability Index (NVSI) is the most reliable index (trend correlation 0.99, 0.4% saturation) and supplies a severity label, while a consensus vote across all twelve indices supplies a critical-line label. A Gradient Boosting classifier trained on 14 raw electrical measurements, with no index as input, reaches F1 = 0.824, recall 0.901 and ROCAUC 0.987 on scenario-disjoint test data, and a Random Forest regressor reproduces NVSI with R^{2} = 0.999. We also observe that the weakest line migrates with loading (5–6, 7–8, 2–3, then 5–6 again in the base case) and that residual errors arise almost entirely among the lines that trade the critical label. A Streamlit application serves both models.

Index Terms—*voltage stability index, continuation power flow, weak line detection, N-1 contingency, gradient boosting, IEEE 14-bus system*

I. INTRODUCTION

Power systems are increasingly operated close to their stability limits because of rising demand, renewable integration and the cost of new transmission [1]. Voltage instability, the progressive decline of bus voltages under stress, is a recurring contributor to major blackouts [1], [2]. Unlike many other disturbances it can develop with little warning from voltage magnitude alone, since voltages often stay within limits until the system is close to its collapse point [3]. Operators therefore need indicators that translate local measurements into a proximity-to-collapse signal, and that point to the specific transmission line where reinforcement or control action is most urgent.

A line voltage stability index (VSI) is a closed-form expression of one line's sending- and receiving-end quantities that is intended to approach a threshold (conventionally 1) as that line nears its transfer limit. Dozens of such indices have been proposed over three decades [19], [20], including FVSI [7], Lij [8], Lmn [9], LQP [10], Lp [11], NLSI [12], NVSI [13], VQILine [14], VLSI [15], L [16], VSI2 [17] and BVSI [18]. Each is normally validated in its originating paper on a few loading snapshots of one or two test systems. Two practical questions are left open: which of these indices can be trusted on a given network, and whether apparently distinct formulas actually carry different information.

Machine learning (ML) offers a way to replace repeated formula evaluation with a learned surrogate that maps raw measurements to a weak-line decision [21], [22], [23]. Its value, however, depends entirely on the quality of the labels it is trained on, and the labels in this setting are themselves produced by VSIs. This paper therefore treats index validation and ML as one pipeline: the indices are audited first, and only the audited result is used as ground truth.

The contributions are as follows. (i) A multi-scenario CPF protocol (base case, N-1 outages, randomized loading directions) that scores each index on saturation, trend fidelity and single-line bias instead of a few snapshots. (ii) A proof, confirmed numerically on tens of thousands of operating points, that three pairs of published indices are identical, together with the consequence that BVSI and Lp vanish on zero-resistance branches. (iii) A labeled dataset of 31,660 line-step samples built from this audit, and a classifier and regressor that learn weak-line status and severity from raw measurements only. (iv) A characterization of how the weakest line migrates with loading and where the classifier fails, with an honest account of the limits of consensus-based labels.

II. BACKGROUND AND RELATED WORK

A. Voltage collapse and continuation power flow

For a given load pattern, the bus voltages follow a PV curve: voltage declines slowly with loading, then rapidly near the maximum loadability point, or

nose, beyond which no steady-state solution exists [2], [3]. Conventional NewtonRaphson power flow becomes ill-conditioned there because the Jacobian is singular at the nose. Continuation power flow (CPF) avoids this by introducing a loading parameter λ and tracing the solution branch with a predictor-corrector scheme under a parameterization (e.g. pseudo-arclength) that remains well-posed through the nose [4]. Here $\lambda = 0$ is the unloaded reference, $\lambda = 1$ the nominal case and λ_{max} the loadability limit.

B. Line voltage stability indices

All twelve indices in this study follow from a common twobus model in which a sending bus ($V_S\delta_S$) supplies a receiving bus ($V_R\delta_R$) through a line of impedance $R + jX = Z\theta$. Combining $S_R = V_R I^*$ with Kirchhoff's voltage law and writing $\delta = \delta_S - \delta_R$ gives a quadratic in the receiving-end voltage [18]:

$$4R\,V_R^2 - V_S V_R(R\cos\delta + X\sin\delta) + P_R(R^2 + X^2) = 0 \quad (1)$$

A real receiving-end voltage exists only if the discriminant is non-negative:

$$[V_S(R\cos\delta + X\sin\delta)]^2 - 4R\,P_R(R^2 + X^2) \geq 0 \quad (2)$$

Rearranging (2) gives BVSI ≤ 1 [18]. The other indices arise from the same condition after different simplifications: neglecting δ (FVSI), neglecting R (NVSI), neglecting reactive power (Lp), or expressing the line through Z and θ (Lmn). The resulting expressions, in the form used in this work, are listed in Table I.

TABLE I
LINE VOLTAGE STABILITY INDICES (VI, VJ: SENDING/RECEIVING VOLTAGE; PI: SENDING-END REAL POWER; PJ, QJ: RECEIVING-END POWER; Θ = ATAN2(X, R))

Index	Expression	Ref.
FVSI	$4Z^2Qj / (Vi^2X)$	[7]
Lij	$4Z^2QjX / [Vi^2(R\sin\delta - X\cos\delta)^2]$	[8]
Lmn	$4XQj / [Vi\sin(\theta-\delta)]^2$	[9]
LQP	$4(X/Vi^2)[(X/Vi^2)Pi^2 + Qj]$	[10]
Lp	$4R\,Pj / [Vi\cos(\theta-\delta)]^2$	[11]
NLSI	$(Pj\,R + Qj\,X) / (0.25\,Vi^2)$	[12]
NVSI	$2X\cdot\mathrm{sqrt}(Pj^2 + Qj^2) / (2Qj\,X - Vi^2)$	[13]
VQILine	$4Qj / (\lvert\mathrm{Im}(1/(R+jX))\rvert\,Vi^2)$	[14]
VLSI	$4(Vi\,Vj\cos\delta - Vj^2\cos^2\delta) / Vi^2$	[15]
L	$4(Vi\,Vj - Vj^2) / Vi^2$	[16]
VSI2	$4Qj\,(R+X)^2 / [X(Vi^2 + 8R\,Qj)]$	[17]
BVSI	$4R\,Pj\,Z^2 / [Vi(R\cos\delta + X\sin\delta)]^2 \leq 1$	[18]

C. Machine learning for voltage stability assessment

ML has been applied to estimate stability margins and indices directly from operating data. Feed-forward networks have been used to predict loadability and index values [21]; tree ensembles have reproduced FVSI on real distribution feeders with very high accuracy [22]; and clustering with classifiers has been used to identify weak areas online from phasor measurements [23]. These studies take the target index as given. This paper differs in auditing the candidate target indices first, because the audit shows that the choice of target changes what the model learns.

III. PROBLEM FORMULATION

Let a sample correspond to one line ℓ at one CPF step k of one scenario. Its feature vector **x** contains 14 raw quantities: R, X, Z, V_i, V_j, $\Delta V = V_i - V_j$, the real and reactive power at both line ends, the real and reactive losses, λ and the distance to collapse $\lambda_{max} - \lambda$. Two targets are defined. The *critical-line label* $y \in \{0, 1\}$ equals 1 for exactly one line per step and is produced by a consensus rule among the twelve indices (Section IV-E). The *severity target* $s \in \mathbb{R}$ is the NVSI value of the line. The learning tasks are binary classification of y and regression of s from **x**. No VSI is part of **x**, so a model cannot succeed merely by re-evaluating a known formula.

Because the positive class is rare (about 5%), accuracy is uninformative. Classification is evaluated by precision, recall, F1 and ROC-AUC on the positive class; regression by R^2, MAE and RMSE. Because adjacent CPF steps are nearly identical, train and test sets are split by scenario, never by row.

IV. METHODOLOGY

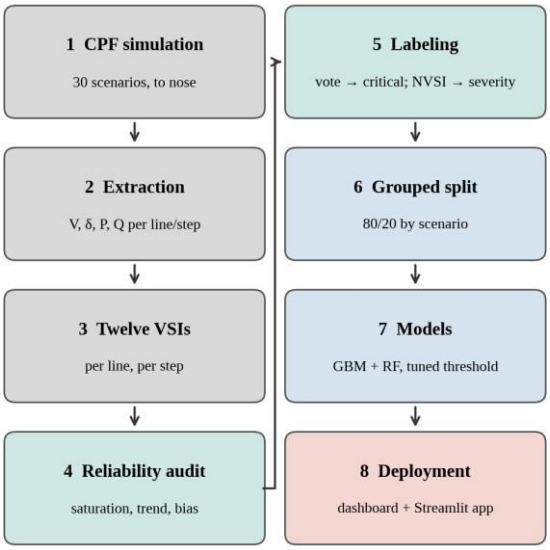


Fig. 1. End-to-end pipeline. Grey: simulation and index computation; green: audit and labeling; blue: model development; red: delivery.

A. Test system and tools

The IEEE 14-bus system [6] has 14 buses, 20 branches and 5 generators, three of which (buses 3, 6 and 8) are synchronous condensers with zero real output. Five branches (lines 8, 9, 10, 14 and 15) are transformers with R = 0 in the case data. Fig. 2 shows the topology. Simulations use MATPOWER 8 [5] in MATLAB; learning uses scikit-learn [26] in Python. Generator reactive-power, real-power and voltage limits are deliberately disabled so that each trace ends at the mathematical nose; the resulting λ_{max} values are therefore upper bounds on secure loadability, not operating limits.

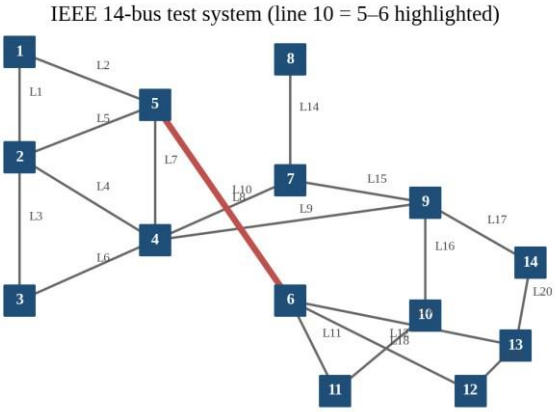


Fig. 2. IEEE 14-bus system with line numbers (line 10 = branch 5–6 highlighted).

B. Scenario design

Thirty-one scenarios were defined: one base case; 20 N-1 outages, each tripping one line under nominal loading; and 10 randomized loading directions in which every bus load is scaled by an independent factor drawn from U(0.5, 1.5). Tripping line 14 (7–8) islands bus 8 and the CPF cannot converge, so it was excluded. The remaining 30 scenarios contain 1583 CPF steps and 31,660 line-step samples (30,693 for in-service lines and 967 for tripped lines).

C. CPF procedure

Algorithm 1: scenario simulation and feature extraction

1. Build the target case: apply the scenario's load scaling or set the tripped branch out of service.
2. Build the base case with the same topology, loads set to 10^{-6} of the target loads and generator real power set to zero.
3. Run CPF from base to target and continue past $\lambda = 1$ until the nose; skip the scenario if the solver fails.
4. At every step compute the branch admittance matrices from the scenario's own branch status and obtain sending- and receiving-end power per unit;

convert to MW/MVAr only for storage.
5. Store V_i, V_j, δ, P, Q, losses, R, X, Z, λ and $\lambda_{max} - \lambda$ for all 20 lines.

Three implementation details proved essential. First, MATPOWER builds a load element only for buses with nonzero demand, so an exactly zero base load changes the network structure relative to the target and aborts the CPF; the 10^{-6} scaling avoids this. Second, VSI expressions mix voltages, impedances and powers and must be evaluated entirely in per unit; passing MW/MVAr inflates every index by the base MVA. Third, the standard external-to-internal renumbering routine deletes out-of-service branches, which shortens per-line arrays under N-1 outages; admittance matrices were therefore built directly from the unfiltered branch table so that line numbering is identical across scenarios.

D. Index computation

At every step all twelve indices of Table I are evaluated for every line, with a small constant added to denominators to avoid division by zero. A tripped line carries no current, and several formulas diverge on it, so its indices are forced to zero and it is excluded from the critical-line vote.

E. Reliability metrics and labeling

An index is scored by (a) its *saturation rate*, the fraction of in-service samples with value above 1 before any clamping; (b) its *trend correlation*, the Pearson correlation between index value and λ along a reference line, averaged over scenarios; and (c) its *single-line bias*, the share of per-step argmax selections taken by its most frequently chosen line. Two reference lines are used for (b). The *consensus line* is the line that wins the majority vote of the twelve per-index argmax selections at the final CPF step. As this definition is partly circular, a label-independent *voltage-weakest line* is also used: the in-service line carrying the largest power among those incident to the lowest-voltage bus at the nose.

Labels are then assigned per step. The critical-line label y marks the line chosen by the majority vote of the twelve indices, which dilutes the idiosyncrasies of any single formula. The severity target is the NVSI value, and a stability zone follows from thresholds of 0.4 and 0.8 chosen from the NVSI distribution (median 0.21, 90th percentile 0.71): Safe below 0.4, Warning from 0.4 to 0.8, Critical above 0.8.

F. Models and training

Data were split with grouped sampling on the scenario identifier (80/20), giving 24 training scenarios (24,980 samples) and 6 test scenarios (6,680 samples, 334 critical), listed as scenarios 9, 10, 17, 19, 25, 29. The classifier is a Gradient Boosting model [24] with 300 trees of depth 3 and learning rate 0.05; a class-weighted Random Forest [25] with 300 trees served as baseline. The regressor is a 300-tree Random Forest. The classifier outputs a probability, and the decision threshold is chosen on the test scenarios as the value maximizing precision subject to recall ≥ 0.90, reflecting that a missed weak line is costlier than a false alarm.

Algorithm 2: labeling and learning
1. For each step, evaluate the twelve indices for every in-service line and record each index's argmax line.
2. Set y = 1 for the line with the most votes, y = 0 for the others; set s = NVSI of each line (0 for a tripped line).
3. Split scenarios 80/20 with grouped sampling.
4. Fit the Gradient Boosting classifier on **x** → y and the Random Forest regressor on **x** → s.
5. Choose the threshold maximizing precision subject to recall ≥ 0.90; report metrics on held-out scenarios.

V. RESULTS

A. Loadability and contingency severity

The base case reaches $\lambda_{max} = 4.06$ with bus 5 the lowestvoltage bus at the nose (0.683 p.u.; Fig. 3). The N-1 outages differ strongly in severity (Fig. 4). Losing line 1 (1–2), the connection of the slack bus to the rest of the network, cuts loadability to 1.34, and losing line 3 (2–3) or line 10 (5–6) to 2.27 and 2.35. Line 15 (7–9) follows at 2.95, while the remaining 15 outages leave λ_{max} between 3.27 and 4.05. The random-direction runs span 3.30 to 4.83, showing how strongly loadability depends on where the load grows.

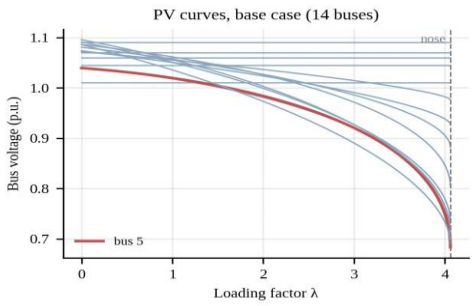


Fig. 3. PV curves of all 14 buses in the base case; bus 5 is lowest at the nose.

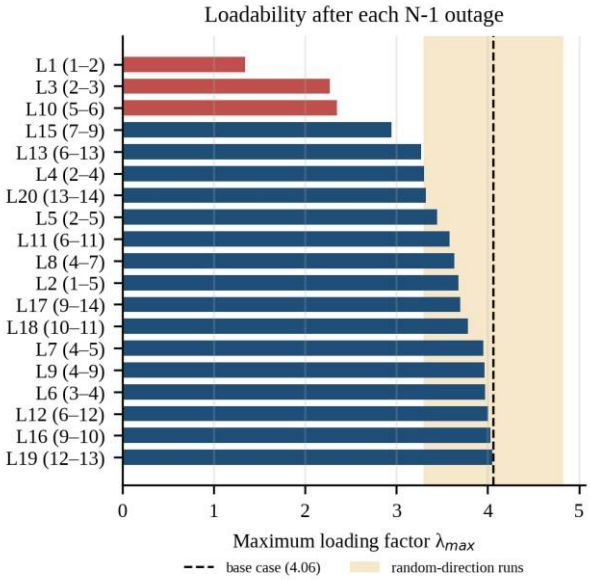


Fig. 4. Maximum loading factor after each N-1 outage; the base case and the range of the random-direction runs are shown for reference.

B. Reliability of the twelve indices

TABLE II INDEX RELIABILITY OVER 30 SCENARIOS (30,693 IN-SERVICE SAMPLES)

Index	Sat. (%)	Trend r (consensus)	Trend r (voltageweakest)	Top-line share (%)
NVSI	0.4	0.99	0.96	44
L	3.6	0.85	0.93	78
FVSI	6.6	0.80	0.91	63
VQILine	6.6	0.80	0.91	63
VSI2	13.0	0.79	0.89	49
NLSI	4.6	0.79	0.81	67
VLSI	1.7	0.76	0.95	44
Lij	11.0	0.73	0.82	41
Lmn	11.0	0.73	0.82	41
Lp	7.2	0.47	0.68	44
BVSI	7.2	0.47	0.68	44
LQP	2.1	0.45	0.89	66

Table II and Fig. 5 summarize the audit. NVSI is the best index under both reference definitions (r = 0.99 and 0.96) and saturates in only 0.4% of samples; its value rises steadily with loading (Fig. 7). L, FVSI/VQILine, VSI2, NLSI and VLSI also track loading ($r \geq 0.76$ under the consensus definition), but L, NLSI and FVSI select a single line for 78%, 67% and 63% of steps respectively, a sign of strong dependence on fixed line parameters. BVSI and Lp are the weakest under both definitions (r = 0.47 and 0.68). LQP ranks last under the consensus definition (0.45) but near the top under the voltage-weakest definition (0.89), so its ranking depends on how the reference line is defined and should not be overinterpreted. The agreement of the two definitions at the top and bottom of the ranking supports the choice of NVSI as ground truth.

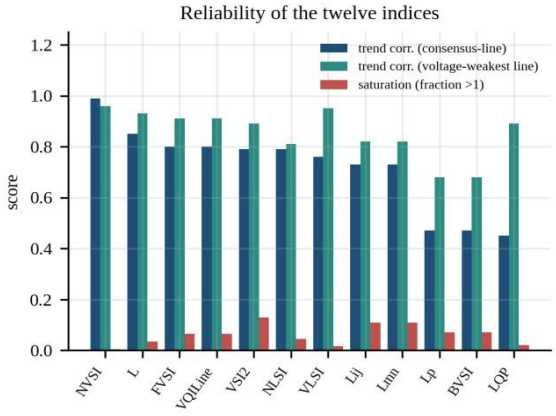


Fig. 5. Trend correlation under two reference-line definitions and saturation rate for each index.

C. Algebraic equivalences and a structural blind spot

Since $\theta = \text{atan2}(X, R)$, $\cos\theta = R/Z$ and $\sin\theta = X/Z$, hence

(3)

Substituting (3) into the Lp expression of Table I gives exactly the BVSI expression, and the same substitution makes Lij and Lmn identical, because $(R\sin\delta - X\cos\delta)^2 = Z^2\sin^2(\theta - \delta)$. Because $|\text{Im}(1/(R + jX))| = X/Z^2$, VQILine reduces to FVSI. Only nine of the twelve formulas are therefore distinct. Numerically, the largest absolute difference between BVSI and Lp (both clamped at 1) over all 30,693 in-service samples is below 10^{-8} (Fig. 6); before clamping the largest relative difference is 3.4×10^{-5}, at an extreme-valued sample. Lij and Lmn differ by at most 1.8×10^{-7} and FVSI and VQILine by at most 1.7×10^{-8}. This also accounts for the observation in [18] that BVSI and Lp move together in every test case.

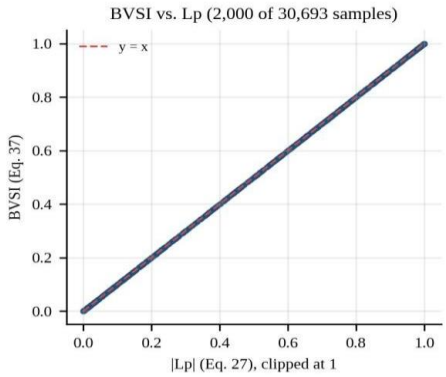


Fig. 6. BVSI against Lp for 2,000 of the in-service samples; all points lie on y = x.

The identity has a practical consequence. BVSI and Lp carry the factor R in the numerator, so both are identically zero on the five R = 0 transformer branches. Branch 10 (5–6) is such a branch, yet it is the line most often identified as critical by the consensus vote (955 of 1583 steps, 60.3%). Along its trajectory in the base case, NVSI, FVSI and L climb toward and beyond 1 while BVSI stays at zero (Fig. 7). An operator relying on BVSI or Lp alone would never see the stress on this branch.

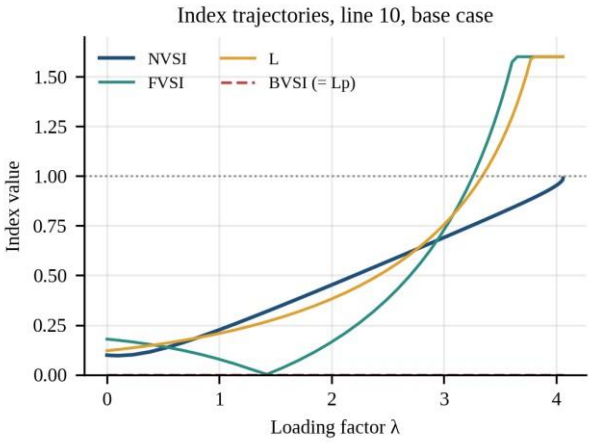


Fig. 7. Index trajectories along line 10 (branch 5–6) in the base case; values above 1.6 are clipped. BVSI (= Lp) is zero because R = 0.

D. Dataset and label statistics

Table III summarizes the labeled dataset. The critical-line label is highly concentrated: lines 10, 3 and 14 account for 89.8% of the 1,583 labeled steps (955, 244 and 222), 8 further lines are labeled between 1 and 46 times, and 9 lines are never critical (Fig. 8). The severity zones are imbalanced as well (Safe 71%, Warning 25%, Critical 4%).

TABLE III

LABELED DATASET BY SCENARIO TYPE

Type	Scen.	Samples	Safe	Warning	Critical
Base	1	1,140	771	311	58
N-1 outage	19	19,340	14,045	4,613	682
Random dir.	10	11,180	7,630	3,030	520
Total	30	31,660	22,446	7,954	1,260

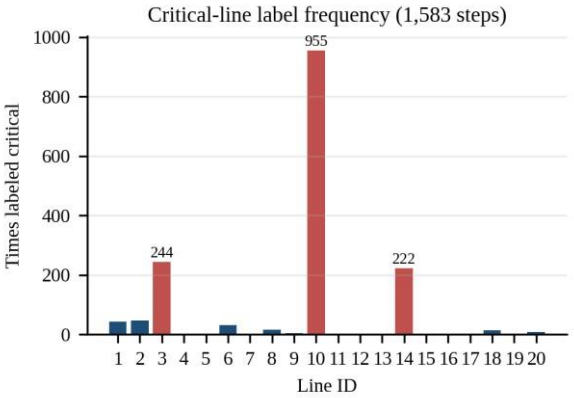


Fig. 8. Number of steps for which each line carries the consensus critical label.

The consensus-critical line also *migrates* as loading increases. In the base case it is line 10 (5–6) for λ = 0–0.84, then line 14 (7–8) for λ = 0.98–1.85, then line 3 (2–3) for λ = 1.99–2.97, then line 10 (5–6) for λ = 3.05–4.06. Line 3 is therefore not a contingency-only weak line: it is critical in 10 base-case steps, 153 outage steps and 81 steps of the randomdirection scenarios 22, 23, 24, 28, 30. Weakness is thus tied to the region of the loading path, and the same few lines trade the label as it is traversed (Fig. 9).

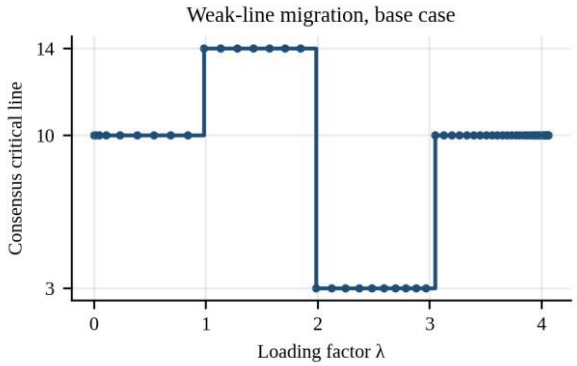


Fig. 9. Consensus critical line versus loading factor in the base case.

E. Classification performance TABLE IV CLASSIFICATION ON THE 6 HELD-OUT SCENARIOS (334 CRITICAL

SAMPLES)

Model, threshold	Prec.	Rec.	F1	FP	FN
Random Forest, 0.5	0.857	0.787	0.821	44	71
Random Forest, 0.213	0.749	0.901	0.818	101	33
Gradient Boosting, 0.5	0.862	0.784	0.821	42	72
Gradient Boosting, 0.218	0.758	0.901	0.824	96	33

Table IV compares the two models at the default threshold and at the recall-constrained threshold. At 0.5 both miss about one in five critical samples. Lowering the threshold to 0.218 raises Gradient Boosting recall from 0.784 to 0.901 (72 to 33 missed samples) at the cost of 54 additional false alarms, with precision 0.758 and F1 0.824. The ROC-AUC is 0.987 (Fig. 10), which shows that the ranking of lines is accurate independently of the threshold; the operating point is a policy choice (Fig. 11). The confusion matrix at the tuned threshold is 6250 / 96 / 33 / 301 (TN / FP / FN / TP; Fig. 12). Gradient Boosting and Random Forest reach the same recall, but Gradient Boosting does so with fewer false alarms (96 versus 101).

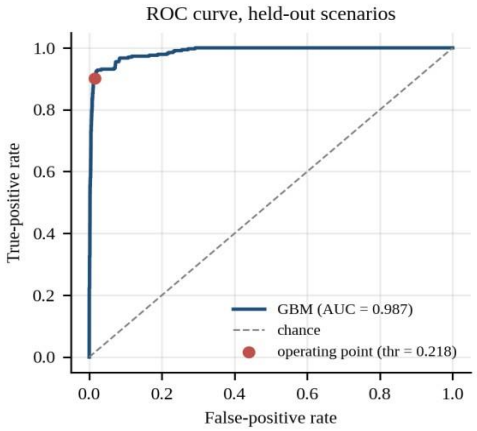


Fig. 10. ROC curve on the held-out scenarios; the marker is the tuned operating point.

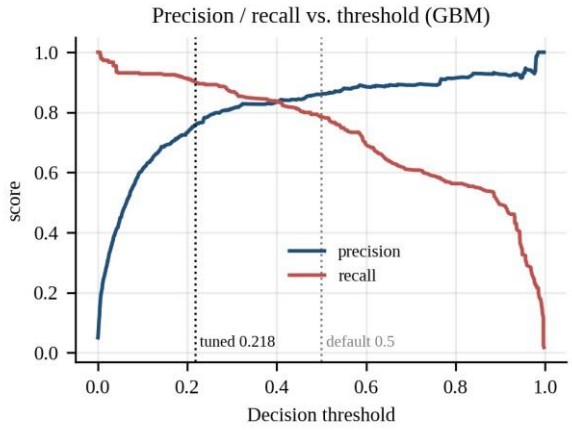


Fig. 11. Precision and recall as functions of the decision threshold.

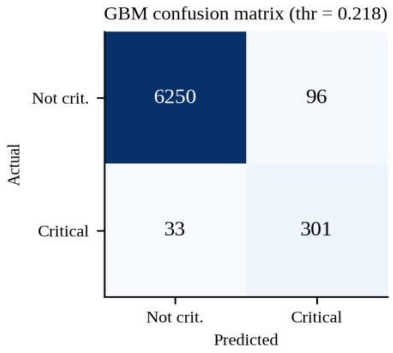


Fig. 12. Confusion matrix of the Gradient Boosting classifier at the tuned threshold.

F. Severity regression

The Random Forest regressor reproduces NVSI on the heldout scenarios with $R^2 = 0.999$, MAE = 0.0024 and RMSE = 0.0085 (Fig. 13). This is expected: NVSI is a deterministic function of X, V_i, P_{to} and Q_{to}, which are among the inputs, so the result shows that the surrogate captures the formula from measurements rather than that the task is hard. Deviations are largest above NVSI = 1, where training samples are scarce and tree ensembles cannot extrapolate. A smaller 50-tree, depth-10 regressor used for deployment attains $R^2 = 0.9987$ and MAE = 0.0046 at 5 MB instead of 627 MB.

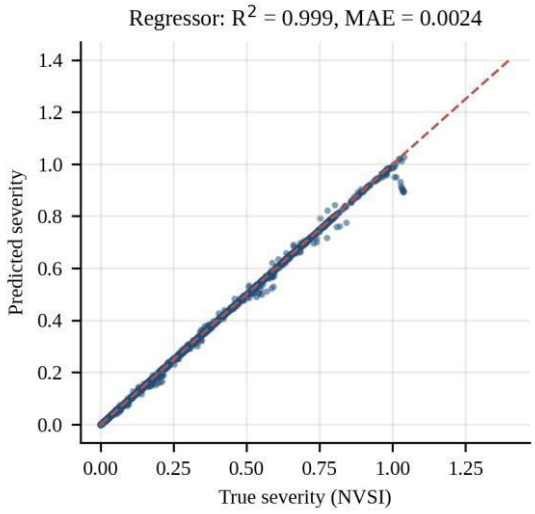


Fig. 13. Predicted against true NVSI severity for 2,500 test samples.

G. Error analysis

At the tuned threshold the 33 missed critical samples fall on lines 3 (12), 10 (6), 2 (5), 6 (4), 14 (3), 9 (2) and 8 (1), while all 96 false alarms fall on three lines: 14 (41), 3 (40) and 10 (15) (Fig. 14). The errors are therefore confined to the lines that trade the critical label and to rarely critical lines with few training examples. Missed samples lie farther from collapse (median distance to collapse 1.47) than critical samples overall (0.82), so the classifier is less likely to fail close to the nose. Because the label is a single winner per step, steps near a hand-over between two lines (Fig. 9) are inherently ambiguous: their physical state varies smoothly while the label flips.

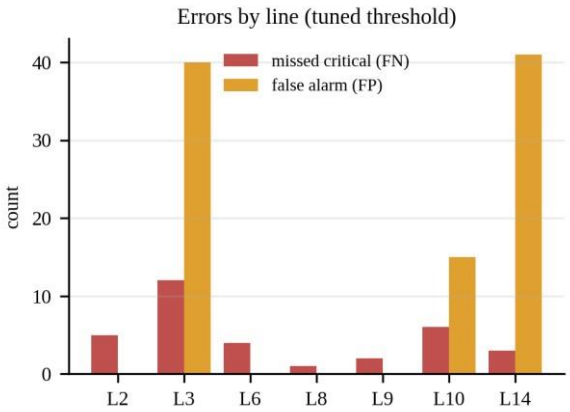


Fig. 14. False negatives and false positives by line at the tuned threshold.

TABLE V

FEATURE-ENGINEERING EXPERIMENTS (RANDOM FOREST, DEFAULT THRESHOLD)

Feature set	F1	Recall
14 raw features (baseline)	0.821	0.787
+ R/X ratio	0.816	0.781
+ line identifier	0.811	0.772

Two interventions aimed at the hard lines left performance unchanged (Table V): an explicit R/X ratio to expose the transformer branches, and a line identifier, which is legitimate here because the target is this fixed network. The tree ensemble already extracts this structure from R and X, so the residual error is a property of the label boundary rather than of missing inputs.

H. Feature importance

The Gradient Boosting importances (Fig. 15) are led by the voltage drop ΔV (26.7%) and receiving-end reactive power Q_{to} (22.3%), followed by P_{from}, P_{to}, λ, Q_{loss}, V_j, the distance to collapse and Q_{from} (4–9% each); R, X, Z, V_i and P_{loss} each contribute about 1% or less. Voltage drop and reactive power flow are the classical drivers of collapse, so the ranking is physically plausible. Importances from tree ensembles are shared among correlated features ($\Delta V = V_i - V_j$; P_{from} and P_{to}) and should be read as indicative only.

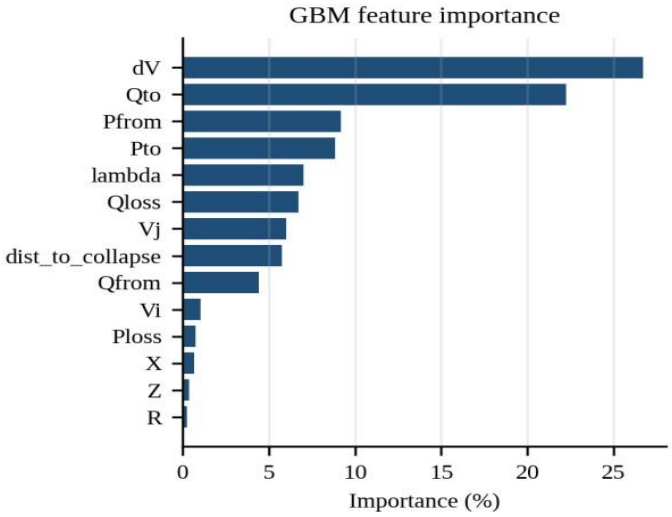


Fig. 15. Feature importance of the Gradient Boosting classifier.

VI. DISCUSSION AND LIMITATIONS

Deployment. All 14 inputs are quantities available from SCADA or phasor measurements, and inference is one pass through a fixed tree ensemble, so the classifier avoids evaluating twelve trigonometric formulas for every line at every interval. The decision threshold can be moved after training to match an operator's tolerance for false alarms. A Streamlit application serves both models, with a scenario replay on the one-line diagram, a what-if predictor, and the analytical views of this paper.

Limitations. (1) All results are from one 14-bus network; the equivalences are general, but the reliability ranking and error pattern need confirmation on larger systems. (2) The data are noise-free simulations; robustness to measurement error has not been tested. (3) Labels are consensus-based and not an independent physical ground truth; the voltage-weakest-line check preserves the top and bottom of the ranking but changes the rank of LQP. (4) At very light loading all indices are small, and the critical label identifies the relatively most stressed line, not a line near collapse; the distance-to-collapse feature mitigates this. (5) Generator limits are disabled, so λ_{max} overstates secure loadability. (6) Only N-1 outages are covered, and the test set contains only 6 scenarios, so per-line statistics are coarse; repeated grouped cross-validation would tighten them. (7) Tree ensembles do not extrapolate beyond the training range, as seen in the regression tail.

VII. CONCLUSION

A multi-scenario CPF audit of twelve line voltage stability indices on the IEEE 14-bus system showed that NVSI is the most reliable index, that three pairs of published formulas are identical, and that BVSI and Lp are blind to zero-resistance branches that the other indices identify as the most stressed. Using NVSI and a consensus vote as labels, a Gradient Boosting classifier trained on raw measurements reached F1 = 0.824, recall 0.901 and ROC-AUC 0.987 on unseen scenarios, with its errors confined to the few lines that trade the critical label as loading increases. Future work will repeat the protocol on larger systems and radial feeders, add N-2 outages and measurement noise, enforce generator limits, use sequence models that exploit the loading trajectory, and test the classifier on real phasor data.

ACKNOWLEDGMENT

The authors thank the Department of Electrical and Electronics Engineering, GMR Institute of Technology, for support. Code and data: repository link,App Link: Click here.

REFERENCES

[1] H. H. Alhelou, M. E. Hamedani-Golshan, T. C. Njenda, and P. Siano, "A survey on power system blackout and cascading events: Research motivations and challenges," *Energies*, vol. 12, no. 4, p. 682, 2019.
[2] P. Kundur, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill, 1994.
[3] T. Van Cutsem and C. Vournas, *Voltage Stability of Electric Power Systems*. Boston, MA, USA: Springer, 1998.

[4] V. Ajjarapu and C. Christy, "The continuation power flow: A tool for steady state voltage stability analysis," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 416–423, Feb. 1992.
[5] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.
[6] Univ. of Washington, "Power Systems Test Case Archive," [Online]. Available: https://labs.ece.uw.edu/pstca/
[7] I. Musirin and T. K. A. Rahman, "Novel fast voltage stability index (FVSI) for voltage stability analysis in power transmission system," in *Proc. Student Conf. Res. Develop.*, 2002, pp. 265–268.
[8] C. Subramani, S. S. Dash, M. A. Bhaskar, M. Jagadeeshkumar, K. Sureshkumar, and R. Parthipan, "Line outage contingency screening and ranking for voltage stability assessment," in *Proc. Int. Conf. Power Syst.*, 2009, pp. 1–5.
[9] M. Moghavvemi and F. M. Omar, "Technique for contingency monitoring and voltage collapse prediction," *IEE Proc. Gener. Transm. Distrib.*, vol. 145, no. 6, pp. 634–640, Nov. 1998.
[10] A. Mohamed, G. B. Jasmon, and S. Yusoff, "A static voltage collapse indicator using line stability factors," *J. Ind. Technol.*, vol. 7, no. 1, pp. 73–85, 1989.
[11] M. Moghavvemi and M. O. Faruque, "Technique for assessment of voltage stability in ill-conditioned radial distribution network," *IEEE Power Eng. Rev.*, vol. 21, no. 1, pp. 58–60, 2001.
[12] A. Yazdanpanah-Goharrizi and R. Asghari, "A novel line stability index (NLSI) for voltage stability assessment of power systems," in *Proc. Int. Conf. Power Syst.*, 2007, pp. 164–167.
[13] R. Kanimozhi and K. Selvi, "A novel line stability index for voltage stability analysis and contingency ranking in power system using fuzzy based load flow," *J. Electr. Eng. Technol.*, vol. 8, no. 4, pp. 694–703, 2013.
[14] F. A. Althowibi and M. W. Mustafa, "Line voltage stability calculations in power systems," in *Proc. IEEE Int. Conf. Power Energy*, Nov. 2010, pp. 396–401.
[15] T. K. A. Rahman and G. B. Jasmon, "A new technique for voltage stability analysis in a power system and improved loadflow algorithm for distribution network," in *Proc. Int. Conf. Energy Manage. Power Del.*, 1995, pp. 714– 719.
[16] S. Sahari, A. F. Abidin, and T. K. A. Rahman, "Development of artificial neural network for voltage stability monitoring," in *Proc. Nat. Power Eng. Conf.*, 2003, pp. 37–42.
[17] T. K. Chattopadhyay, S. Banerjee, and C. K. Chanda, "Impact of distributed generator on voltage stability analysis of distribution networks under critical loading conditions for composite loads," in *Proc. Int. Conf. Electron. Commun. Instrum.*, Jan. 2014, pp. 1–4.
[18] B. Ismail, N. I. A. Wahab, M. L. Othman, M. A. M. Radzi, K. N. Vijayakumar, M. K. Rahmat, and M. N. M. Naain, "New line voltage stability index (BVSI) for voltage stability assessment in power system: The comparative studies," *IEEE Access*, vol. 10, pp. 103906–103931, 2022.
[19] J. Modarresi, E. Gholipour, and A. Khodabakhshian, "A comprehensive review of the voltage stability indices," *Renew. Sustain. Energy Rev.*, vol. 63, pp. 1–12, Sep. 2016.
[20] M. S. S. Danish, T. Senjyu, S. M. S. Danish, N. R. Sabory, N. K, and P. Mandal, "A recap of voltage stability indices in the past three decades," *Energies*, vol. 12, no. 8, p. 1544, Apr. 2019.
[21] S. Mokred, Y. Wang, and T. Chen, "Modern voltage stability index for prediction of voltage collapse and estimation of maximum load-ability for weak buses and critical lines identification," *Int. J. Electr. Power Energy Syst.*, vol. 145, Art. no. 108596, 2023.
[22] M. A. Mossie, T. T. Yetayew, G. T. Bitew, M. G. Yenealem, and T. M. Beza, "Machine learning algorithms for voltage stability assessment in electrical distribution systems," *Sci. Rep.*, vol. 15, Art. no. 31980, 2025.
[23] S. Perez-Londono, G. Olivar-Tost, and J. J. Mora-Florez, "Online determination of voltage stability weak areas for situational awareness improvement," *Electr. Power Syst. Res.*, vol. 145, pp. 112–121, 2017.
[24] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.
[25] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[26] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.