

Information-Theoretic and Machine-Learning Fusion for Resilience Identification and Critical-Transition Early Warning in Population Dynamical Systems

1st Sarah bint Rashid Al Qahtani
Princess Nora bint Abdulrahman University
Riyadh, Saudi Arabia
SarahbintRashidAlQahtani@outlook.com

Abstract—Resilience monitoring in population systems is increasingly constrained by irregular sampling, seasonal forcing, and the limited transferability of generic early-warning signals. This study develops an information-theoretic and machine-learning (IT–ML) framework that treats resilience assessment as a prospective decline-risk problem rather than as an assertion that every observed downturn is a bifurcation. The framework integrates Fisher information, permutation entropy, and lag-one mutual information with conventional rolling descriptors of abundance dynamics, and evaluates their combined utility under chronology-respecting validation. A reproducible empirical benchmark was constructed from publicly available Portal Project rodent records. The analysis retained six target species, 186 census periods, and 1,116 effort-standardized species–census observations from 1999 to 2015. A sharp-decline outcome was defined prospectively as a mean abundance during the next three censuses that was at most 55% of the preceding six-census mean, conditional on a preceding mean of at least two individuals. The held-out test period contained 246 evaluation windows and 31 prospective decline events. A conventional early-warning composite achieved a receiver operating characteristic area under the curve (ROC-AUC) of 0.649. A machine-learning model based on conventional temporal statistics achieved ROC-AUC 0.703, whereas the IT–ML fusion achieved ROC-AUC 0.648 and precision–recall area 0.178. Species-block bootstrap intervals were broad, and the information features did not add a stable aggregate gain beyond autocorrelation and local trend. These results show that information-theoretic descriptors can be incorporated transparently into population-monitoring pipelines, but their operational value is context dependent. The principal contribution is therefore an auditable workflow that separates resilience-relevant state description, prospective warning, and causal claims about critical transitions.

Keywords—*population resilience, early-warning signals, Fisher information, permutation entropy, machine learning, critical transitions*

I. INTRODUCTION

Population dynamical systems frequently exhibit pronounced fluctuations before management institutions can establish whether a change is transient, seasonal, or a durable loss of resilience. This distinction has direct consequences for monitoring design because a reversible pulse may only require intensified observation, whereas a loss of recovery capacity can require timely intervention. In ecological theory,

resilience is not a single quantity: local return rates, basin geometry, and resistance to perturbation describe related but non-equivalent properties. The choice of a resilience indicator must therefore be tied to a concrete decision problem and to the type of transition that the data can support. [1] [2] [3]

Early-warning research has commonly examined rising variance, autocorrelation, skewness, and changing trends as symptoms of critical slowing down. These statistics are attractive because they can be computed from a single observed variable. Their empirical performance, however, is strongly conditioned by the forcing mechanism, the sampling protocol, detrending choices, and the prevalence of actual transitions. A multi-lake assessment found that common indicators often did not outperform chance, and a recent review likewise emphasizes the gap between idealized transitions and real-world monitoring records. The appropriate objective is consequently not to declare a transition from a single score, but to evaluate whether a score improves prospective risk discrimination under an explicit validation protocol. [3] [4] [5] [6]

Information theory offers a complementary language for this task. Fisher information is sensitive to local changes in the shape of an empirical state distribution, while entropy-based measures quantify disorder in ordinal or spatial patterns and mutual information captures statistical dependence. These quantities may expose changes in state concentration, temporal organization, or dependence that are not expressed by a variance-only diagnostic. However, their interpretability is conditional on window length, discretization, noise, and the dimension of the observed system. A defensible application should therefore treat information metrics as interpretable features rather than as a universal scalar definition of resilience. [7] [8] [9] [10] [11]

Machine-learning approaches have expanded early-warning research by learning nonlinear combinations of temporal features or representations. Deep classifiers can perform well when their training and test processes share relevant dynamical structure, but their generalization may deteriorate across forcing regimes or under extrapolation. In a data-limited field setting, a smaller and auditable model is often preferable to a high-capacity system whose warnings cannot be traced to observable descriptors. Chronology-respecting partitioning is equally important: leakage across

Corresponding Author: Sarah bint Rashid Al Qahtani, Airport Road, King Khalid International Airport, Riyadh, Saudi Arabia, 11564, SarahbintRashidAlQahtani@outlook.com

time can inflate apparent scientific performance even when a conventional train–test split is reported. [12] [13] [14] [15] [16] [17] [18] [19]

The present study addresses this gap by integrating three information-theoretic descriptors with conventional early-warning statistics in a prospective population-monitoring benchmark. The study asks three questions: first, can an auditable feature set distinguish forthcoming sharp declines in an openly available population time series; second, do information-theoretic descriptors improve held-out discrimination beyond conventional rolling statistics; and third, how sensitive are the conclusions to a reasonable change in the operational decline definition? The scope is deliberately limited. The analysis evaluates warning performance for prospective abundance declines in a single long-term study system; it does not infer causal mechanisms or claim that every labelled decline is a critical transition. The remainder of the paper reviews the relevant foundations, defines the framework, reports the empirical results, and discusses implications for resilience monitoring.

II. RELATED WORK

A. Resilience, critical transitions, and population early warnings

The conceptual literature distinguishes resilience from stability and from the observational signatures that may accompany a loss of resilience. Local resilience concerns recovery near an attractor, whereas ecological resilience concerns the persistence of a system within a basin or regime. Rate-induced transitions further complicate the picture because a system may undergo a large qualitative change before a static bifurcation threshold is reached. These distinctions motivate a conservative interpretation of early-warning scores: they are empirical descriptors of changing dynamics, not a substitute for a mechanistic model of all possible transitions. [1] [2] [5]

Population-level experiments provide a useful bridge between theory and monitoring. In a seasonal *Drosophila* system, trait-based information improved warning detection for approaching collapse, yet the value of individual indicators depended on the timing of stress. Population time-series analyses also show pervasive density dependence at aggregate scales, while individual series can be weakly informative. Such findings justify an approach that pools evidence across species while preserving the temporal structure within each population and that reports uncertainty rather than a single universal threshold. [6] [20] [21]

B. Information-theoretic descriptors of dynamical organization

Information-based early-warning methods have developed along several complementary lines. Entropy production, nonequilibrium free energy, and time irreversibility have been proposed as mechanistic descriptors in ecological models. Spatial permutation entropy and spatial mutual information have been used to detect changes in vegetation and spatially extended systems. Network methods based on mutual-information weighted entropy extend the same intuition to biological state transitions. Collectively, these studies motivate a multi-descriptor feature family, but they also demonstrate that the signal depends on the system, observation geometry, and preprocessing choices. [2] [7] [8] [9] [10]

Fisher information occupies a distinctive position because it describes the sharpness of an estimated distribution rather than its global spread. An earlier logistic-model study used Fisher information to relate periodic forcing, probability-density structure, and sustainability. The present study retains the sensitivity of Fisher information to local distributional organization but shifts the application from analytically controlled logistic dynamics to an empirical, effort-standardized population-monitoring setting. The purpose is not to reproduce a closed-form logistic analysis; it is to test whether a finite-window Fisher descriptor contributes useful prospective information when combined with other temporal features. [11] [22]

C. Machine learning, model realism, and evaluation discipline

Machine learning for critical transitions ranges from generic neural classifiers trained on simulated bifurcations to models trained on surrogates of historical observations. These approaches show that nonlinear representations can recognize patterns missed by a small set of generic indicators, but the transportability question remains central. Training data generated from simple models can transfer to some natural systems, yet extrapolation across unobserved non-stationary regimes remains difficult. A practical compromise is to use flexible but interpretable learning on descriptors that retain a clear dynamical meaning. [12] [13] [14] [15] [16] [23]

This design requires rigorous temporal evaluation. Reviews of forecasting evaluation warn that random resampling can provide misleading performance estimates for non-stationary series. Distribution-free time-series uncertainty methods, anomaly-detection taxonomies, and change-point reviews further reinforce that the relevant object is time-ordered evidence. The present benchmark therefore uses a fixed chronological train–validation–test split, selects classification thresholds only on validation data, and quantifies test uncertainty with a species-block bootstrap. These decisions do not eliminate all uncertainty; they make the remaining uncertainty visible. [17] [18] [19] [24] [25] [26]

III. METHODS

A. Research design and operational target

The framework was designed as an observational, prospective classification study. Its input is a univariate abundance trajectory with known sampling effort. Its output is a calibrated probability that the mean abundance over the next three censuses will be substantially lower than the mean abundance over the preceding six censuses. This operationalization separates a measurable warning task from the stronger causal claim that the system is approaching a bifurcation. It is appropriate for monitoring programmes in which the practical question is whether a trajectory warrants review before a near-term decline, not whether the full ecological state space has been observed (Figure 1).

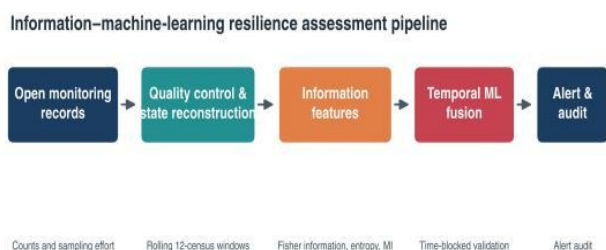


Fig. 1. Information-machine-learning resilience assessment pipeline. The workflow preserves a separation between data quality control, state description, temporal prediction, and post-alert audit.

B. Source data and census-level state reconstruction

The empirical benchmark used public capture records from the Portal Project, a long-term Chihuahuan Desert study near Portal, Arizona, USA. Rodent records are collected approximately monthly, and the project documentation provides capture history, plot identity, trapping effort, and treatment information. The raw repository was cloned at commit 7514fad88ff5946dcea00c9641319ea9c2fa02ff. The analysis used only conventional non-negative census periods between January 1999 and March 2015 and focused on eight plots that were documented as control plots during the analysis window. Capture counts were standardized to 49 traps within each plot before aggregation, which prevents a short or partial trapping effort from being interpreted as a biological decrease.

A census can span a calendar-month boundary while retaining a single census-period identifier. The reconstruction therefore aggregated all available month fragments to a single species-period observation. Six target rodent species met the coverage criteria of at least 150 census observations and a non-zero occurrence fraction of at least 0.35. The final panel comprised 186 census periods, six species, and 1,116 species-period observations, with an average of 7.73 sampled focal plots per aggregated census. The number of contributing plots ranged from four to eight because some censuses straddled a treatment or calendar boundary. Zeros were retained when a plot had been sampled; no imputation was applied to the abundance panel.

TABLE I. DESCRIPTIVE STATISTICS FOR THE EFFORT-STANDARDIZED POPULATION PANEL.

Species	n	Mean $\pm$ SD	Median	Range	Non-zero fraction
Chaetodipus baileyi	186	12.96 $\pm$ 12.58	8.5	0–71	0.95
Chaetodipus penicillatus	186	22.83 $\pm$ 23.09	16.5	0–116	0.90
Dipodomys merriami	186	20.48 $\pm$ 10.63	18.0	2–52	1.00
Dipodomys ordii	186	8.73 $\pm$ 7.44	7.0	0–36	0.93
Onychomys torridus	186	4.61 $\pm$ 4.60	3.0	0–25	0.87
Peromyscus eremicus	186	1.34 $\pm$ 1.83	0.0	0–8	0.49

The selected trajectories span markedly different abundance levels and variabilities (Table I). The use of effort-standardized counts is important because observation error and process variation are not interchangeable in population time series. Stochastic logistic and state-space studies have shown that deterministic trajectory matching can obscure this distinction, particularly at low abundance or when the

observation process is imperfect. The present feature set does not attempt to identify a complete stochastic population model. Instead, it evaluates whether descriptors extracted from the observed, standardized state sequence assist prospective warning under an explicitly limited target. [27] [28] [29] [30]

C. Feature construction

For each species and each eligible census period t , a trailing 12-census window of log-transformed abundance, $x_t = \log(1 + N_t)$, was constructed. Six conventional descriptors were computed: mean log abundance, sample variance, coefficient of variation on untransformed abundance, lag-one autocorrelation, linear trend, and skewness. The three information-theoretic descriptors were Fisher information, normalized permutation entropy of ordinal patterns of order three, and lag-one mutual information. All functions were computed within the trailing window and therefore used no future abundance values.

The discrete Fisher-information descriptor was computed from a smoothed empirical histogram, $F = \sum_i (p_{i+1} - p_i)^2 / [(p_{i+1} + p_i)/2]$, where p_i is the normalized frequency in the i th adaptive quantile bin. Permutation entropy was $H_p = -\sum_j q_j \log(q_j) / \log(3!)$, where q_j denotes the probability of each length-three ordinal pattern. Lag-one mutual information was $I(X_t; X_{t+1}) = \sum_{ij} p_{ij} \log[p_{ij}/(p_i p_j)]$. These finite-window estimates are descriptive and were not interpreted as thermodynamic state functions. To avoid undefined moments in degenerate windows, only the feature matrix—not the raw abundance observations—was median-imputed after replacing non-finite values (Table II).

TABLE II. FEATURES AND THEIR INTENDED MONITORING INTERPRETATION.

Feature family	Descriptors	Operational interpretation
Conventional state	Mean, variance, coefficient of variation	Level and local dispersion of observed abundance
Conventional temporal	Lag-one autocorrelation, linear trend, skewness	Memory, directionality, and asymmetry
Information-theoretic	Fisher information	Local sharpness of the empirical state distribution
Information-theoretic	Permutation entropy; lag-one mutual information	Ordinal complexity and adjacent-state dependence

D. Prospective label and temporal evaluation

The primary label was a prospective sharp decline. For a candidate index time t , B_t was the mean abundance in the six preceding censuses and F_t was the mean abundance in the next three censuses. A positive label was assigned when $B_t \geq 2$ and $F_t \leq 0.55B_t$. The basal abundance condition reduces labels generated by low-count fluctuations near zero. It also makes the target more relevant to a monitoring decision: an alert should signal a decline from an observed, non-trivial recent population level. The final feature table contained 1,002 windows because each label required both a 12-census history and a three-census future horizon.

Temporal integrity was protected with a fixed chronological partition. Windows ending on or before 31 December 2007 formed the training set ($n = 558$); windows from 2008–2010 formed the validation set ($n = 198$); and windows from 2011–2015 formed the locked test set ($n = 246$). No test observation was used to estimate a feature scaler, fit a

model, or choose a threshold. The decision threshold for each method was selected to maximize validation F1 score. This design follows the principle that the test period must emulate an unknown future rather than a randomly withheld subset of the historical record. [17] [18] [19]

E. Comparative models, uncertainty, and sensitivity

Three warning methods were compared. The conventional early-warning baseline combined standardized variance, lag-one autocorrelation, and negative linear trend through a logistic transformation. The ML-statistics comparator was a class-balanced random forest with 800 trees, maximum depth five, and a minimum leaf size of four, trained on the six conventional descriptors. The proposed IT-ML fusion was a class-balanced histogram gradient-boosting classifier with 240 iterations, learning rate 0.055, 11 maximum leaf nodes, and L2 regularization 0.7, trained on all nine descriptors. All hyperparameters were fixed before inspecting the test results.

Performance was assessed using ROC-AUC, precision-recall area (PR-AUC), Brier score, F1 score, sensitivity, and specificity. Because windows are temporally dependent within each species, 95% uncertainty intervals were calculated using 1,000 species-block bootstrap resamples. A sensitivity analysis re-estimated the proposed model using decline fractions of 0.50, 0.55, and 0.60 while retaining all temporal partitions. This analysis tested whether the qualitative conclusion depended on the exact operational threshold for a prospective decline.

IV. RESULTS

A. Population trajectories and operational decline labels

The reconstructed panel contained 162 observed zero-abundance entries (14.5%) but no missing species-period cells after the census-level aggregation. The six selected species differed substantially in mean abundance, standard deviation, and occupancy (Table I), providing a heterogeneous but internally comparable test bed. The primary outcome generated 31 positive sharp-decline windows in the locked test set, corresponding to a prevalence of 12.6%. This class imbalance makes PR-AUC and threshold-specific error rates essential complements to ROC-AUC.

Figure 2 illustrates a representative held-out trajectory for the desert pocket mouse, *Chaetodipus penicillatus*. The observed series contained repeated increases and declines rather than a monotonic collapse. Downward triangles identify the prospectively defined sharp-decline windows. The IT-ML risk score changed over time and exceeded its validation-selected threshold at many dates, demonstrating that a warning system can be sensitive to local structure without providing evidence that the entire trajectory was approaching a single, irreversible state boundary.

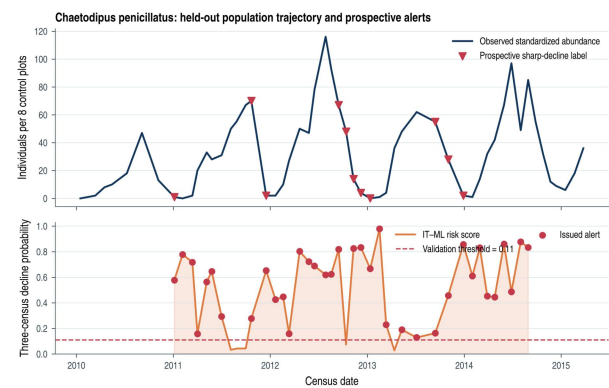


Fig. 2. Held-out abundance trajectory and IT-ML prospective decline probabilities for *Chaetodipus penicillatus*. Downward triangles in the upper panel indicate windows meeting the future three-census decline criterion. The lower panel shows the probability estimated without test-period fitting and the validation-selected decision threshold.

B. Information-state descriptors

The three information descriptors varied across the held-out trajectory and did not move in a common direction at every labelled decline (Figure 3). Fisher information displayed occasional short-lived peaks, permutation entropy changed more gradually, and lag-one mutual information showed pronounced local variation. This pattern is consistent with the fact that the indicators quantify different aspects of a finite-window state distribution. It also cautions against interpreting an increase or decrease in a single information measure as an invariant signature of resilience loss.

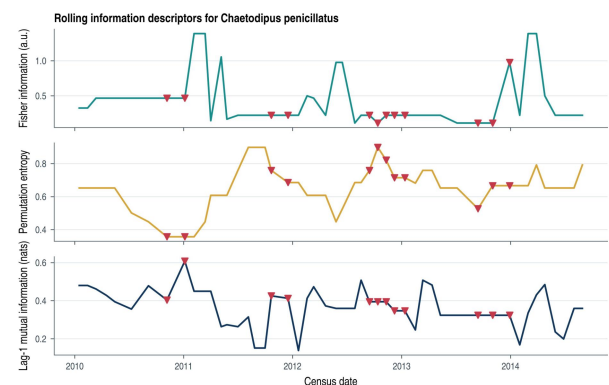


Fig. 3. Rolling information-theoretic descriptors for the held-out *Chaetodipus penicillatus* trajectory. Red triangles mark prospective sharp-decline windows. Each descriptor was computed from a preceding 12-census window.

C. Held-out warning performance

TABLE III. LOCKED-TEST PERFORMANCE FOR PROSPECTIVE SHARP-DECLINE WARNING (N = 246 WINDOWS; 31 EVENTS).

Method	RO C- AU C	PR - AU C	Bri er	F1	Sensiti vity	Specifi city	TP/FP/T N/FN
Convent ional EWS	0.6 49	0.2 00	0.2 52	0.2 27	0.677	0.381	21/133/8 2/10
ML statistics	0.7 03	0.2 04	0.1 83	0.3 01	0.903	0.409	28/127/8 8/3
IT-ML fusion	0.6 48	0.1 78	0.2 01	0.2 76	0.806	0.419	25/125/9 0/6

The conventional early-warning composite achieved ROC-AUC 0.649 and PR-AUC 0.200 (Table III). The ML-

statistics comparator achieved the highest point ROC-AUC (0.703) and the lowest Brier score (0.183). The IT-ML fusion achieved ROC-AUC 0.648 and PR-AUC 0.178. At its validation-selected threshold of 0.11, the fusion method detected 25 of 31 events (sensitivity 0.806) but also issued 125 false positive alerts (specificity 0.419). Thus, the fusion model did not demonstrate a stable aggregate improvement over a model using conventional temporal statistics only.

Species-block bootstrap intervals were wide. For ROC-AUC, the 95% intervals were 0.315–0.755 for the conventional baseline, 0.499–0.858 for ML statistics, and 0.497–0.752 for IT-ML fusion. These overlapping intervals are expected with six species and should prevent rank-order claims based on small point-estimate differences. Figure 4 makes this uncertainty explicit: machine learning improved point discrimination relative to the conventional composite, but the information-theoretic extension did not produce a statistically decisive advantage in this empirical setting.

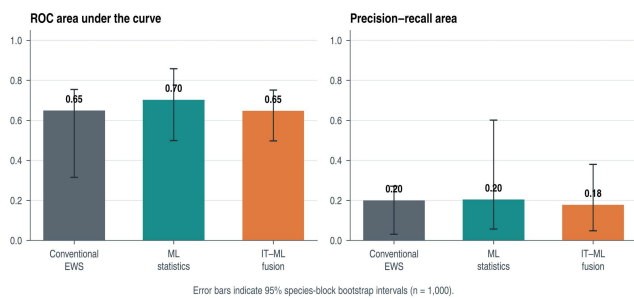


Fig. 4. Held-out discrimination of the three warning methods. Bars show locked-test ROC-AUC and PR-AUC; error bars show 95% species-block bootstrap intervals based on 1,000 resamples.

D. Feature attribution and sensitivity

Permutation-based importance for the IT-ML fusion ranked lag-one autocorrelation first (mean reduction in test PR-AUC 0.027), followed by linear trend (0.013) and permutation entropy (0.013). Fisher information and lag-one mutual information showed negative mean importance in this pooled test analysis. The result does not imply that Fisher information is meaningless for every population. It indicates that, under this particular label, window construction, and empirical panel, its addition did not consistently improve out-of-sample precision-recall discrimination beyond the other features.

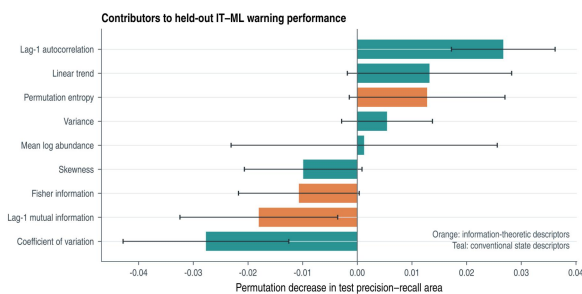


Fig. 5. Permutation importance for the IT-ML fusion model on the locked test period. Orange bars identify information-theoretic descriptors and teal bars identify conventional descriptors. Error bars show standard deviations over 150 permutations.

TABLE IV. SENSITIVITY OF THE IT-ML FUSION TO THE PROSPECTIVE DECLINE DEFINITION.

Future mean criterion	Events	ROC - AUC	PR-AUC	F1	Sensitivity	Specificity
≤ 50% of baseline	29	0.667	0.203	0.270	0.828	0.424
≤ 55% of baseline	31	0.648	0.178	0.276	0.806	0.419
≤ 60% of baseline	32	0.690	0.206	0.306	0.938	0.374

Changing the future-to-baseline criterion between 0.50 and 0.60 produced 29–32 events in the test period. IT-ML ROC-AUC ranged from 0.648 to 0.690 and sensitivity ranged from 0.806 to 0.938 (Table IV). The exact values varied with the label definition, but the central conclusion was robust: the framework retained prospective discrimination under reasonable outcome perturbations, while the information-theoretic extension did not clearly dominate the conventional-statistics model.

V. DISCUSSION

This study demonstrates a deliberately conservative way to fuse information theory and machine learning for population resilience monitoring. Rather than treating an observed abundance decline as evidence of a critical transition, it turns the practical warning question into a prospective and auditable classification task. The results show moderate held-out discrimination for both the conventional-statistics model and the IT-ML fusion, but the fusion did not improve aggregate performance. This result is scientifically useful because it prevents a feature-rich model from being presented as superior merely because it is conceptually novel.

The importance analysis suggests why the fusion did not dominate. Lag-one autocorrelation and local trend carried the clearest out-of-sample information for the operational target, whereas the information descriptors added limited or unstable marginal value. This finding is compatible with a broader literature that characterizes information-based indicators as context sensitive. Entropy, mutual information, and Fisher information may be informative when the monitored system has an appropriate spatial structure, network representation, or distributional change, but their performance need not transfer unchanged to a noisy, seasonal, single-variable abundance series. [2] [7] [8] [9] [10] [11]

The comparison with machine learning also clarifies the role of model complexity. The random-forest comparator achieved the strongest point ROC-AUC in the locked test period using only conventional descriptors. This does not invalidate information theory; it highlights the need for incremental, not rhetorical, evaluation. A useful information-theoretic feature should increase discrimination, improve calibration, reduce false alerts, increase lead time, or provide an independent mechanism-level interpretation. If it does not, retaining it for a specific operational deployment should require a stated scientific reason. This approach is aligned with recent work showing both the potential and the transfer limitations of data-driven tipping-point methods. [12] [13] [14] [15] [16]

Several methodological limitations are material. First, the benchmark covers six species at one desert site and cannot establish broad ecological generality. Second, the outcome labels are constructed from observed abundance ratios and do

not distinguish demographic decline, dispersal, detection change, or a true regime shift. Third, the 12-census window and histogram-based Fisher estimate embed scale choices. Fourth, bootstrap uncertainty is necessarily broad because the number of independent species blocks is small. Fifth, the treatment-stable control restriction improves internal comparability but limits the analysis of interventions and community interactions. The model should therefore be used as a monitoring prototype, not as a threshold rule for conservation action without local review. [1] [4] [20] [21] [29] [30]

These limitations indicate several concrete next steps. A multi-site benchmark with pre-registered transition definitions could test whether information descriptors add value when climate covariates, detection models, or community composition are available. Conformal prediction could attach distribution-free warning intervals to risk scores, while state-space models could separate latent abundance from capture variability. Future work could also compare finite-window information features with representations learned directly from raw sequences and evaluate utility through a cost-sensitive alert audit that counts lead time, false alerts, and missed events together. [17] [18] [24] [25] [26] [29]

VI. CONCLUSION

This paper presented a reproducible IT–ML workflow for resilience identification and critical-transition early warning in population dynamical systems. The workflow integrates effort-aware state reconstruction, conventional temporal statistics, Fisher information, permutation entropy, lag-one mutual information, chronological validation, and post-alert audit. Applied to 1,116 effort-standardized observations from six Portal Project rodent populations, it produced moderate prospective decline discrimination but no stable aggregate advantage of information-theoretic fusion over conventional temporal statistics alone.

The principal contribution is methodological. Resilience-relevant descriptors should be evaluated as candidates within a transparent prediction and uncertainty framework, not assumed to be valid solely because they have a plausible theoretical interpretation. The resulting design is portable to other monitored populations: retain an explicit outcome, preserve time order, document the observation process, report false alerts and uncertainty, and distinguish a useful warning score from a demonstrated ecological critical transition.

REFERENCES

- [1] V. Dakos, S. Kéfi, "Ecological resilience: what to measure and how," *Environmental Research Letters*, vol. 17, no. 4, Art. no. 043003, 2022, doi: 10.1088/1748-9326/ac5767.
- [2] L. Xu, D. Patterson, S. A. Levin, J. Wang, "Non-equilibrium early-warning signals for critical transitions in ecological systems," *Proceedings of the National Academy of Sciences*, vol. 120, no. 5, Art. no. e2218663120, 2023, doi: 10.1073/pnas.2218663120.
- [3] V. Dakos, C. A. Boulton, J. E. Buxton, J. F. Abrams, B. Arellano-Nava, D. I. Armstrong McKay, S. Bathiany, L. Blaschke, et al., "Tipping point detection and early warnings in climate, ecological, and human systems," *Earth System Dynamics*, vol. 15, no. 4, pp. 1117-1135, 2024, doi: 10.5194/esd-15-1117-2024.
- [4] D. A. O'Brien, S. Deb, G. Gal, S. J. Thackeray, P. S. Dutta, S. i. S. Matsuzaki, L. May, C. F. Clements, "Early warning signals have limited applicability to empirical lake data," *Nature Communications*, vol. 14, no. 1, Art. no. 7942, 2023, doi: 10.1038/s41467-023-43744-8.
- [5] R. Arumugam, F. Guichard, F. Lutscher, "Early warning indicators capture catastrophic transitions driven by explicit rates of environmental change," *Ecology*, vol. 105, no. 4, Art. no. e4240, 2024, doi: 10.1002/ecy.4240.
- [6] J. B. Burant, C. Park, G. S. Betini, D. R. Norris, "Early warning indicators of population collapse in a seasonal environment," *Journal of Animal Ecology*, vol. 90, no. 6, pp. 1538-1549, 2021, doi: 10.1111/1365-2656.13474.
- [7] G. Tirabassi, C. Masoller, "Entropy-based early detection of critical transitions in spatial vegetation fields," *Proceedings of the National Academy of Sciences*, vol. 120, no. 1, Art. no. e2215667120, 2023, doi: 10.1073/pnas.2215667120.
- [8] S. Deb, P. S. Dutta, "Critical transitions in spatial systems induced by Ornstein–Uhlenbeck noise: spatial mutual information as a precursor," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 480, no. 2283, Art. no. 20230594, 2024, doi: 10.1098/rspa.2023.0594.
- [9] Y. Z. Pang, L. Li, Z. Jin, "Early warning signals of critical transitions in ecosystems: entropy reduction in vegetation spatial patterns," *Nonlinear Dynamics*, vol. 113, no. 12, pp. 15597-15618, 2025, doi: 10.1007/s11071-025-10877-y.
- [10] Y. Xie, X. Peng, P. Li, "MIWE: detecting the critical states of complex biological systems by the mutual information weighted entropy," *BMC Bioinformatics*, vol. 25, no. 1, Art. no. 44, 2024, doi: 10.1186/s12859-024-05667-z.
- [11] G. Di Fresco, B. Spagnolo, D. Valenti, A. Carollo, "Metrology and multipartite entanglement in measurement-induced phase transition," *Quantum*, vol. 8, Art. no. 1326, 2024, doi: 10.22331/q-2024-04-30-1326.
- [12] T. M. Bury, R. I. Sujith, I. Pavithran, M. Scheffer, T. M. Lenton, M. Anand, C. T. Bauch, "Deep learning for early warning signals of tipping points," *Proceedings of the National Academy of Sciences*, vol. 118, no. 39, Art. no. e2106140118, 2021, doi: 10.1073/pnas.2106140118.
- [13] S. Deb, S. Sidheekh, C. F. Clements, N. C. Krishnan, P. S. Dutta, "Machine learning methods trained on simple models can predict critical transitions in complex natural systems," *Royal Society Open Science*, vol. 9, no. 2, Art. no. 211475, 2022, doi: 10.1098/rsos.211475.
- [14] D. Patel, E. Ott, "Using machine learning to anticipate tipping points and extrapolate to post-tipping dynamics of non-stationary dynamical systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 33, no. 2, Art. no. 023143, 2023, doi: 10.1063/5.0131787.
- [15] T. M. Bury, D. Dylewsky, C. T. Bauch, M. Anand, L. Glass, A. Shrier, G. Bub, "Predicting discrete-time bifurcations with deep learning," *Nature Communications*, vol. 14, no. 1, Art. no. 6331, 2023, doi: 10.1038/s41467-023-42020-z.
- [16] Z. Ma, C. Zeng, Y. C. Zhang, T. M. Bury, "Predicting critical transitions with machine learning trained on surrogates of historical data," *Communications Physics*, vol. 8, no. 1, Art. no. 258, 2025, doi: 10.1038/s42005-025-02172-4.
- [17] V. Cerqueira, L. Torgo, I. Mozetič, "Evaluating time series forecasting models: an empirical study on performance estimation methods," *Machine Learning*, vol. 109, no. 11, pp. 1997-2028, 2020, doi: 10.1007/s10994-020-05910-7.
- [18] H. Hewamalage, K. Ackermann, C. Bergmeir, "Forecast evaluation for data scientists: common pitfalls and best practices," *Data Mining and Knowledge Discovery*, vol. 37, no. 2, pp. 788-832, 2023, doi: 10.1007/s10618-022-00894-5.
- [19] S. Kapoor, A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, Art. no. 100804, 2023, doi: 10.1016/j.patter.2023.100804.
- [20] L. M. Thibaut, S. R. Connolly, "Hierarchical modeling strengthens evidence for density dependence in observational time series of population dynamics," *Ecology*, vol. 101, no. 1, Art. no. e02893, 2020, doi: 10.1002/ecy.2893.
- [21] P. G. Nicolau, R. A. Ims, S. H. Sørbye, N. G. Yoccoz, "Seasonality, density dependence, and spatial population synchrony," *Proceedings of the National Academy of Sciences*, vol. 119, no. 51, Art. no. e2210144119, 2022, doi: 10.1073/pnas.2210144119.
- [22] A. Al-Saffar, E. j. Kim, "Sustainable theory of a logistic model - Fisher information approach," *Mathematical Biosciences*, vol. 285, pp. 81-91, 2017, doi: 10.1016/j.mbs.2016.12.009.
- [23] S. Deb, E. Mahendru, P. Goyal, V. Guttal, P. S. Dutta, N. C. Krishnan, "Optimal sampling of spatial patterns improves deep learning-based early warning signals of critical transitions," *Royal Society Open*

- Science, vol. 11, no. 6, Art. no. 231767, 2024, doi: 10.1098/rsos.231767.
- [24] C. Xu, Y. Xie, "Conformal Prediction for Time Series," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 10, pp. 11575-11587, 2023, doi: 10.1109/tpami.2023.3272339.
- [25] A. Blázquez-García, A. Conde, U. Mori, J. A. Lozano, "A Review on Outlier/Anomaly Detection in Time Series Data," ACM Computing Surveys, vol. 54, no. 3, pp. 1-33, 2022, doi: 10.1145/3444690.
- [26] C. Truong, L. Oudre, N. Vayatis, "Selective review of offline change point detection methods," Signal Processing, vol. 167, Art. no. 107299, 2020, doi: 10.1016/j.sigpro.2019.107299.
- [27] L. Marrec, C. Bank, T. Bertrand, "Solving the stochastic dynamics of population growth," Ecology and Evolution, vol. 13, no. 8, Art. no. e10295, 2023, doi: 10.1002/ece3.10295.
- [28] L. Descheemaeker, S. de Buyl, "Stochastic logistic models reproduce experimental time series of microbial communities," eLife, vol. 9, Art. no. e55650, 2020, doi: 10.7554/elife.55650.
- [29] B. Rosenbaum, E. A. Fronhofer, "Confronting population models with experimental microcosm data: from trajectory matching to state-space models," Ecosphere, vol. 14, no. 4, Art. no. e4503, 2023, doi: 10.1002/ecs2.4503.
- [30] E. A. Fronhofer, L. Govaert, M. I. O'Connor, S. J. Schreiber, F. Altermatt, "The shape of density dependence and the relationship between population growth, intraspecific competition and equilibrium population density," Oikos, vol. 2024, no. 2, Art. no. e09824, 2024, doi: 10.1111/oik.09824.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the research teams and data stewards who collected, maintained, and openly shared the long-term ecological monitoring data used in this study. Their contributions to data accessibility and documentation provided an important foundation for the present research.

FUNDING

None.

AVAILABILITY OF DATA

Not applicable.

AUTHOR CONTRIBUTIONS

Sarah bint Rashid Al Qahtani: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Visualization, Writing – Original Draft, Writing – Review & Editing, and Project Administration.

COMPETING INTERESTS

The authors declare no competing interests.

Publisher's note WEDO remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is published online with Open Access by Green Design Engineering and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).

© The Author(s) 2026