

Explaining Temporal Resilience-State Inference of Wireless Industrial Communication Systems

^{1st} Shashank Jhansale Anil Kumar 
Technische Universität Braunschweig
Germany
jhansale@ibr.cs.tu-bs.de

^{2nd} Ahmad Ashraf Zargar 
Technische Universität Braunschweig
Germany
zargar@ibr.cs.tu-bs.de

^{3rd} Lars Wolf 
Technische Universität Braunschweig
Germany
wolf@ibr.cs.tu-bs.de

Abstract—Industrial wireless networks require monitoring of communication degradation to be resilient against disruptions, despite limited access to measurements in the application-layer caused by encrypted or proprietary protocols. Temporal learning models can link infrastructure telemetry to application-relevant resilience-states through inference, but how they use this evidence remains unclear. This work investigates explainable resilience-state inference through conditional feature masking, feature-time occlusion, and bootstrap resampling. The analysis characterizes how the telemetry contributions vary across inferred resilience-states, where the influential observations occur within temporal input windows, and how much application-visible Round-Trip Time (RTT) closes the visibility gap.

The results show that feature reliance is strongly state-dependent. Longer temporal contexts improve balanced multi-class inference, while CNN’s feature-time occlusion identifies the strongest evidence near the decision instant. The WinIT-inspired analysis yields consistent feature rankings and temporal localization. Application-visible RTT substantially improves instantaneous classifiers but adds little to temporal models, suggesting that temporal infrastructure telemetry already captures part of the missing RTT timing information. Bootstrap analysis confirms that the main performance and explanation trends are stable across test-set resampling.

Index Terms—Explainable artificial intelligence, resilience-state inference, industrial wireless networks, network telemetry, temporal inference, limited visibility.

I. INTRODUCTION

Wireless communication increases industrial flexibility but remains vulnerable to interference, congestion, and propagation effects that can compromise the reliable, low-latency communication required by cyber-physical applications. Therefore, Industry 5.0 identifies resilient communication as an important requirement to maintain industrial operation under changing conditions [1].

To enable timely resilience actions before application operation is compromised, our previous work [2] formulated three application-relevant communication states: i) *Operational* (S_o), where application requirements are satisfied; ii) *Partially Disrupted* (S_p), where communication degradation is observed but application failure is not guaranteed; and iii) *Completely Disrupted* (S_c), where communication requirements are guaranteed to be violated.

We evaluated different instantaneous threshold-based methods and temporal learning-based models to recognize their suitability for inferring resilience states and concluded that

temporal models provide a balanced three-class classification. Explaining these inferences is important for understanding how otherwise black-box ML models use network telemetry. Building on our previous work, we therefore investigate state-wise feature contribution and feature-time evidence.

Since industrial application traffic may be encrypted, proprietary, or unavailable to the network controller, resilience-state inference uses infrastructure telemetry. At every telemetry-sample index n , the infrastructure provides the feature vector

$$\mathbf{x}_n = [\text{RSSI}_n, \text{CBR}_n, \text{RET}_n], \quad (1)$$

corresponding to Received Signal Strength Indicator (RSSI), Channel Busy Ratio (CBR), and MAC-layer Retransmissions (RET), respectively. These measurements characterize the radio and medium-access conditions, but do not directly indicate whether the application requirements are satisfied.

For periodic industrial application traffic, particularly traffic patterns standardized by 3GPP [3], a telemetry sampling interval T_{sample} for an application inter-packet gap T_{cycle} contains

$$N = \left\lceil \frac{T_{\text{sample}}}{T_{\text{cycle}}} \right\rceil \quad (2)$$

application-transmission opportunities. Let L_n denote the number of lost application packets within the interval n , and let B_n denote the longest burst of consecutive losses. Based on the application’s survival time S_{time} , the minimum burst length that violates the survival requirement and the corresponding critical accumulated-loss threshold are

$$K = \left\lceil \frac{S_{\text{time}}}{T_{\text{cycle}}} \right\rceil + 1, \quad L_{\text{crit}} = N - \left\lfloor \frac{N}{K} \right\rfloor + 1. \quad (3)$$

The threshold L_{crit} is the minimum number of packet losses that implies a burst of at least K consecutive losses, irrespective of their packet order. The derivation of this deterministic bound is provided in [2].

The resilience state associated with the telemetry sample n is defined as

$$S_n = \begin{cases} S_o, & L_n = 0, \\ S_p, & 0 < L_n < L_{\text{crit}} \wedge B_n < K, \\ S_c, & L_n \geq L_{\text{crit}} \vee B_n \geq K. \end{cases} \quad (4)$$

Thus, S_o represents packet-loss-free operational communication, S_p represents degraded communication without survival-time violations, and S_c denotes an observed or guaranteed violation of the application survival requirement.

To link these application-defined communication states to infrastructure-observable telemetry, this paper investigates explainable resilience-state inference under limited application visibility. The main contributions of this work are:

- Identification of state-wise telemetry contributions using conditional feature masking.
- Localization of influential feature–time evidence within the temporal observation window, with a WinIT-inspired temporal-XAI consistency analysis.
- Quantification of the visibility gap between infrastructure-only and application-visible inference.
- Assessment of performance and explanation stability through bootstrap resampling.

The remainder of this paper is structured as follows. Section II introduces the methodology and experimental setting, while Section III presents the explainability and visibility-gap evaluation. Section IV discusses related work before concluding the paper in Section V.

II. METHODOLOGY

This section summarizes the dataset, inference, perturbation, and robustness analysis. The dataset partitions, model architectures and hyperparameters follow our previous study [2] and were retained without retuning for the present evaluation.

A. Dataset and Label Calibration

The dataset was collected using a testbed comprising an industrial-grade access point, a wireless client, and a software-defined radio introducing controlled interference over a fixed 2.4 GHz channel. The application messages were inspired by the 3GPP process-automation traffic pattern [3] and were transmitted every $T_{\text{cycle}} = 10$ ms, while the infrastructure telemetry was sampled every $T_{\text{sample}} = 100$ ms. The dataset contains approximately 220,000 telemetry samples and 2.2 million application messages collected under operational and disrupted conditions.

Ground-truth labels were generated exclusively from the application packet outcomes using the formulation in Eqs. (2)–(4). For an application survival time of $S_{\text{time}} = 20$ ms, the calibration gives $N = 10$, $K = 3$, and $L_{\text{crit}} = 8$. Consequently, S_o denotes no packet loss, S_p denotes $0 < L_n < 8$ with $B_n < 3$, and S_c denotes $L_n \geq 8$ or $B_n \geq 3$.

The dataset contains approximately 51%, 16%, and 33% samples of S_o , S_p , and S_c , respectively. Figure 1 shows the data distribution of RSSI and CBR and illustrates substantial overlap between states, particularly between S_p and S_c , and motivates the use of temporal inference. The measurement sequences were divided into fixed training, validation, and test partitions using a 70%–15%–15% split before temporal-input construction, preventing samples from the same measurement sequence from appearing in different subsets.

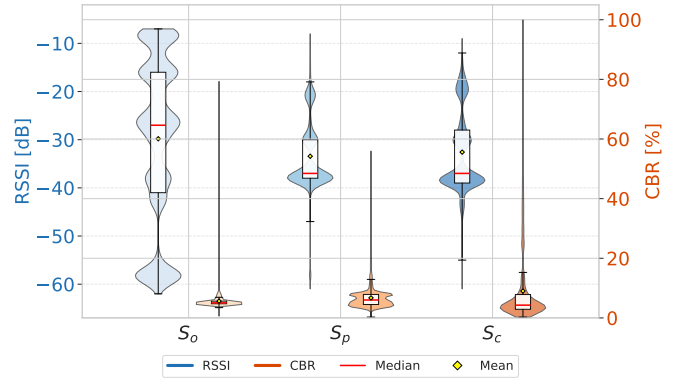


Fig. 1: Distribution of RSSI and CBR across the three resilience states.

B. Inference and Evaluation Protocol

Two feature-visibility settings have been evaluated. The infrastructure-only inference setting uses RSSI, CBR, and RET. The application-visible reference setting additionally includes Round-Trip Time (RTT), which is aggregated over the same 100 ms telemetry sampling interval. RTT is therefore used only to quantify the benefit of application information unavailable to an infrastructure-based monitor.

Infrastructure-only inference evaluates Kernel Density Estimator (KDE)-based thresholding, instantaneous Random Forest (RF), a Convolutional Neural Network (CNN) with $W = 10$ telemetry samples, a Long Short-Term Memory (LSTM) with $W = 100$, and a CNN using $W = 150$ telemetry samples. The temporal inputs correspond to the contexts of 1 s, 10 s, and 15 s, respectively. Under application-visible conditions, the same methods are evaluated except CNN-10, which provides no substantial additional insight.

The binary evaluation combines S_p and S_c into a disruption class, while S_o represents operational communication. Binary evaluation reports accuracy, F1 score and False Negative Rate (FNR), while three-state evaluation reports accuracy, macro F1, and state-wise recall.

C. Reference-Based Perturbation Analysis

Feature reliance was evaluated by masking one feature at a time and comparing the perturbed predictions with the original predictions. Each masked feature was conditionally estimated from the remaining telemetry features using an ExtraTrees regressor fitted only on the training split. For temporal CNN and LSTM inputs, conditional replacement was applied in raw feature space over the complete input sequence before applying the existing model scaler. Changes in macro F1 and state-wise recall quantify global and state-specific reliance, respectively. Negative recall changes indicate evidence supporting the corresponding state, while positive changes indicate prediction redistribution after masking. The procedure has been applied under both feature-visibility settings.

Feature–time reliance was evaluated for CNN-150, over the final 10 s of its 15 s input, divided into ten non-overlapping 1 s

intervals for localization. State-wise recall changes localized the largest effects; the final 10 s were selected because earlier effects were negligible.

As a temporal-XAI consistency check, a WinIT-inspired conditional analysis was additionally applied to CNN-150. For each feature-time block, the analysis conditionally replaced the selected feature and measured the change in the model probability assigned to the true state. The resulting feature rankings and temporal localization were compared with the recall-based occlusion results.

D. Bootstrap Analysis

Variability was assessed using 1,000 bootstrap resamples of test instances. For each resample, the ground-truth label, the original prediction, and the corresponding perturbed predictions were sampled jointly to preserve pairing. The percentile-based 95% confidence intervals were then computed for the performance and perturbation quantities. Since the models were not retrained, the intervals represent variability in test-set composition rather than training variability.

III. EVALUATION

The evaluation addresses three research questions:

- RQ1* How do infrastructure-observable features contribute to the recognition of each resilience state?
- RQ2* How does temporal context support resilience-state inference, and where within the context is the relevant evidence concentrated?
- RQ3* How does application-visible RTT change predictive performance and feature reliance compared with infrastructure-only telemetry?

A. Predictive Reference

The predictive performance under infrastructure-only and application-visible conditions is summarized in Table I. The infrastructure-only results in Table I(a) provide the reference for the feature and temporal attribution analyses, while the RTT-visible results in Table I(b) are discussed in Section III-D.

Under infrastructure-only visibility, the temporal models achieve the strongest balanced three-state performance. CNN-150 achieves the highest macro F1 of 0.6788 and the highest S_p recall of 0.6331, while the instantaneous RF and KDE methods provide weaker separation of the intermediate state. In contrast, the lowest binary FNR of 0.0322 is obtained by CNN-10. These results suggest that a short temporal context supports sensitive disruption detection, whereas a longer context is associated with better distinction among operational, partially disrupted, and completely disrupted communication states.

B. *RQ1: State-Wise Reliance on Infrastructure Features*

Figure 2 reports the signed percentage-point change in state-wise recall after masking each infrastructure feature. Negative values indicate that the feature supports recognition of the corresponding state, whereas positive values indicate a redistribution of predictions after masking.

The resulting reliance pattern is strongly state-dependent. For the operational state S_o , RET reduces recall by 14–32 percentage points across the evaluated models, making RET the dominant feature for recognizing operational communication. CBR has only a small effect on S_o recall.

For the completely disrupted state S_c , CBR provides the most consistent evidence across models. Its removal reduces recall by approximately 21–51 percentage points across the evaluated models. The effect on S_p is indirect: masking CBR increases S_p recall because samples previously assigned to S_c are redistributed towards this intermediate state. Thus, CBR mainly provides evidence of severe disruption rather than a unique indication of S_p . No individual feature produces a similar dominant reduction in S_p recall. Therefore, the recognition of S_p appears to depend on less separable feature combinations or interactions, which is consistent with its lower baseline recall and overlapping feature distribution presented in Fig. 1. RSSI shows more model-dependent effects, including substantial S_c reliance for CNN-150. This is consistent with the multi-modal RSSI distribution and overlap between resilience states.

In general, RET provides the strongest evidence for S_o , while CBR provides the most consistent evidence for S_c .

C. *RQ2: Contribution and Localization of Temporal Context*

Table I(a) suggests that longer temporal context supports balanced three-state inference, with the largest gain observed for S_p recall. Figure 3 reports the state-wise recall change produced by feature-time occlusion within the analyzed final 10 s of the CNN-150 input. Masking most individual intervals earlier than the final second produces only small marginal recall changes. The largest effects occur in the final second before prediction.

Within the final second, conditional masking of CBR reduces S_c recall by approximately 24 percentage points, while masking RET reduces S_o and S_c recall by approximately 13 and 11 percentage points, respectively. The corresponding increase in S_p recall is consistent with predictions shifting to the intermediate state when evidence supporting the neighboring states is removed. These results show that the strongest individually localized evidence within the analyzed 10 s occurs close to the decision instant.

This does not render earlier context irrelevant, as information distributed across preceding intervals may have only a limited effect when intervals are removed individually. However, CNN-150 outperforming CNN-10 is consistent with a longer temporal context that improves the distinction between partial and critical disruptions. This pattern suggests that recent observations provide stronger evidence, while earlier observations may contribute through distributed temporal dependencies within the window.

The WinIT-inspired analysis produced the same qualitative temporal pattern. It identified the same top-ranked feature for each state, with Spearman rank correlation $\rho = 1.0$ relative to the conditional occlusion ranking. The final 1 s block also

TABLE I: Predictive performance of threshold-based, non-temporal, and temporal resilience-state inference under infrastructure-only and application-visible conditions. Arrows indicate the preferred metric direction.

(a) Infrastructure-only telemetry											
Paradigm	Method	Context Window		Binary			Multi-class		Recall		
		Samples	Time	Acc. $\uparrow$	F1 $\uparrow$	FNR $\downarrow$	Acc. $\uparrow$	Macro F1 $\uparrow$	S_o $\uparrow$	S_p $\uparrow$	S_c $\uparrow$
Threshold	KDE	1	100 ms	0.7660	0.7510	0.2930	0.6514	0.4892	0.8249	0.0240	0.7457
Non-temporal	RF	1	100 ms	0.7548	0.7921	0.0644	0.6163	0.5723	0.5746	0.3762	0.8250
Temporal	CNN-10	10	1 s	0.7817	0.8157	0.0322	0.6550	0.6303	0.5963	0.5615	0.8041
Temporal	LSTM	100	10 s	0.8120	0.8307	0.0764	0.7027	0.6702	0.7009	0.5698	0.7843
Temporal	CNN-150	150	15 s	0.8244	0.8428	0.0564	0.7057	0.6788	0.7056	0.6331	0.7489

(b) Infrastructure telemetry with application-visible RTT											
Paradigm	Method	Context Window		Binary			Multi-class		Recall		
		Samples	Time	Acc. $\uparrow$	F1 $\uparrow$	FNR $\downarrow$	Acc. $\uparrow$	Macro F1 $\uparrow$	S_o $\uparrow$	S_p $\uparrow$	S_c $\uparrow$
Threshold	KDE	1	100 ms	0.9103	0.9082	0.0230	0.8110	0.7062	0.9381	0.1486	0.9580
Non-temporal	RF	1	100 ms	0.9233	0.9264	0.0323	0.8693	0.8371	0.8791	0.7352	0.9329
Temporal	LSTM	100	10 s	0.8151	0.8313	0.0870	0.7096	0.6730	0.7175	0.5435	0.7953
Temporal	CNN-150	150	15 s	0.8174	0.8404	0.0366	0.7147	0.6893	0.6720	0.6412	0.8264

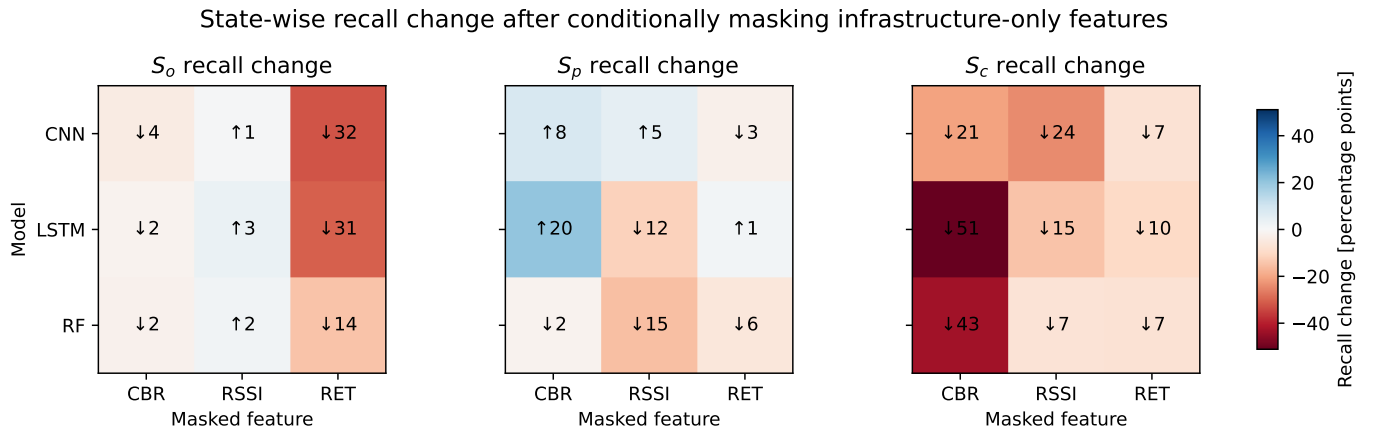


Fig. 2: State-wise recall change after masking individual infrastructure features. Each cell reports the signed percentage-point change relative to the unmasked predictions. Negative values indicate reduced recognition of the corresponding state, while positive values indicate prediction redistribution after masking.

contained the strongest localized importance, supporting the concentration of evidence near the prediction instant.

D. RQ3: Information Added by Application-Visible RTT

Table I(b) shows that the inclusion of application-visible RTT substantially improves instantaneous inference. RTT reflects the latency behavior experienced by the application and is therefore closely related to the timing constraints used to define the resilience states. Consequently, the RTT-visible setting serves only as a reference for quantifying information unavailable under infrastructure-only visibility.

For KDE, the inclusion of RTT increases macro F1 from 0.4892 to 0.7062 and enables strong recognition of S_o and S_c . However, S_p recall remains limited to 0.1486, indicating that RTT alone does not resolve the intermediate state. RF combines RTT with nonlinear feature interactions and achieves

the highest RTT-visible macro F1 of 0.8371, including an S_p recall of 0.7352.

Smaller gains are observed for the temporal models. CNN-150 increases from 0.6788 to 0.6893 in macro F1, while LSTM increases from 0.6702 to 0.6730. This suggests that temporal infrastructure telemetry already contains information correlated with the application-level latency behavior exposed by RTT.

Figure 4 further shows that RF relies strongly on RTT. Its masking reduces S_o and S_p recall by approximately 82 and 64 percentage points, respectively. In contrast, the temporal models retain stronger reliance on infrastructure telemetry. For CNN-150, RTT contributes to S_o recognition, but RET remains more influential. For LSTM, RTT has only limited influence, while CBR and RET remain the dominant features for S_c and S_o , respectively. In general, RTT narrows the visibility gap mainly for instantaneous methods. Its limited

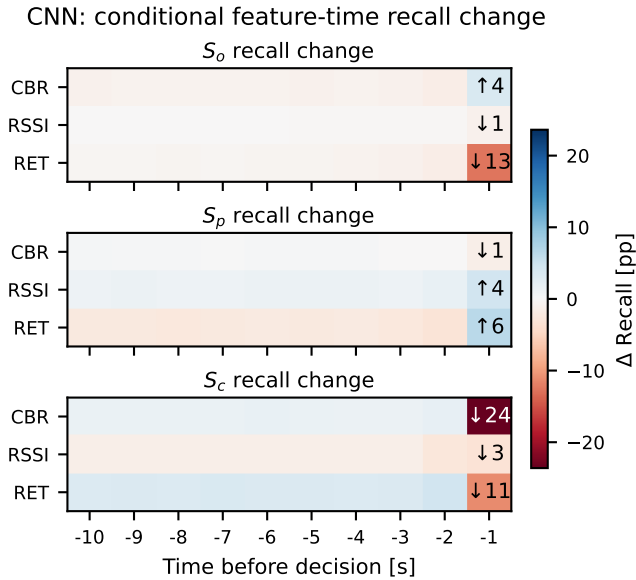


Fig. 3: Feature-time occlusion for CNN-150 under infrastructure-only visibility. Each cell reports the signed percentage-point change in state-wise recall when one feature is masked within a 1s interval of the final 10s before the prediction instant.

benefit to temporal models suggests that temporal infrastructure telemetry already captures part of the application-relevant timing information.

Robustness:

The bootstrap analysis produced narrow 95% confidence intervals for the main performance and perturbation trends. The dominant RET effect for S_o and the consistent CBR effect for S_c remained stable across test-set resamples. By contrast, several weaker effects were not consistently different from no recall change, including the impact of masking RET on LSTM-based S_p recall. Thus, the main trends are robust to test-set composition, but not necessarily to model retraining or alternative perturbation references. The findings are limited to the evaluated controlled testbed; feature reliance may shift under different radio environments, interference patterns, or deployment conditions.

IV. RELATED WORK

Literature relevant to this study spans temporal attribution, explainable network management, and cross-layer inference under restricted application visibility. For sequential models, TimeSHAP extends Shapley-value explanations to recurrent architectures and decomposes predictions into feature, event, and feature-time contributions through sequence perturbations [4]. Dynamask learns sparse feature-time masks using temporally informed perturbations [5], whereas WinIT evaluates feature removal over prediction windows through

conditional replacements that preserve temporal dependencies [6]. These methods establish temporal masking as a model-agnostic explanation mechanism.

Temporal explanations have also been investigated in mobile-network settings. AIChronoLens relates attribution scores to patterns in mobile-traffic histories and shows that similar aggregate relevance may conceal different temporal model behavior [7]. RAN anomaly prediction has been addressed by forecasting user-equipment indicators with an LSTM and applying SHAP-based attribution to the predicted features [8]. These studies motivate temporal analysis in wireless systems.

Related efforts in explainable wireless-network management primarily target control or diagnosis. EXPLORA associates deep-reinforcement-learning actions with attributed network-state graphs for O-RAN control [9]. SpotLight combines distributed anomaly detection with causal filtering to identify responsible RAN and platform indicators [10]. XAI anomaly explains radio measurements that contribute to O-RAN anomaly scores through SHAP-based attribution [11], while XMLAD exposes feature thresholds using interpretable decision trees [12]. These studies therefore motivate feature-level localization explanations.

Limited application visibility has separately motivated the inference from network-observable measurements. DeepQoE estimates web and mobile-application quality from network-visible traffic features [13], while WebRTC quality indicators have been inferred from IP/UDP features and compared with methods using application headers [14]. Similarly, standardized telemetry schemas have been proposed to combine network statistics with application-specific quality indicators [15]. These works show that infrastructure observations can approximate application behavior.

Against this background, the present work combines state-wise conditional feature masking and feature-time occlusion, supported by a WinIT-inspired temporal-XAI consistency analysis, to explain operational, partially disrupted, and completely disrupted communication states and quantify the application-visibility gap.

V. CONCLUSION

We investigated explainable resilience-state inference under limited infrastructure visibility using conditional feature masking, feature-time occlusion, and bootstrap resampling. The results show strongly state-dependent feature reliance, with different telemetry variables contributing differently across inferred conditions and no single feature dominating the intermediate condition. Longer temporal context improves multi-state inference, while feature-time analysis localizes the strongest evidence near the prediction instant. The WinIT-inspired analysis yields consistent feature rankings and temporal localization.

Application-visible RTT substantially improves instantaneous inference but provides limited gains for temporal models, indicating that temporal infrastructure telemetry already captures part of the application-relevant timing information.

State-wise recall change after conditionally masking RTT-aware features

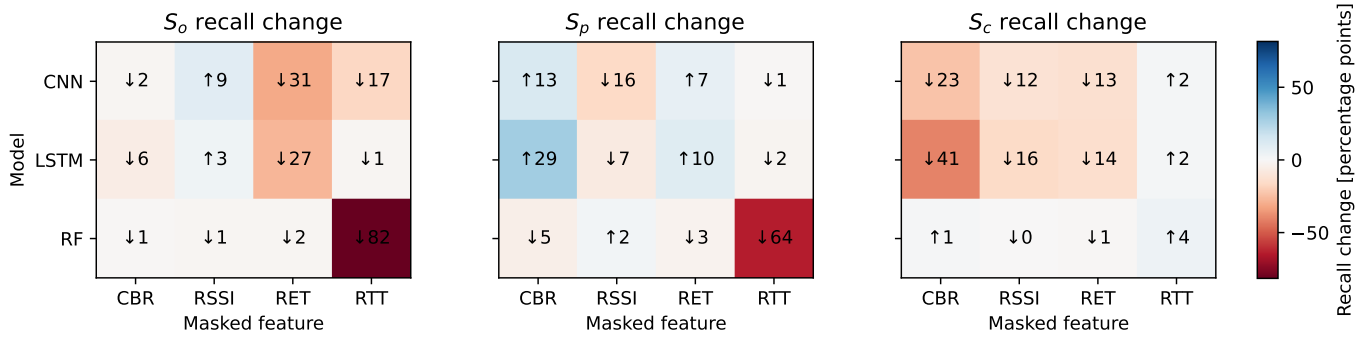


Fig. 4: State-wise recall change after conditional feature masking in the application-visible reference setting.

Bootstrap resampling shows that the main performance and explanation trends are stable with respect to test-set composition. Although the explanations quantify model reliance rather than causal relationships, they support transparent validation of telemetry-based monitors before operational deployment.

ACKNOWLEDGMENT

This work was in part funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 503131139 (REINDEER). Priority Programme SPP 2378 ”Resilience in Connected Worlds – Mastering Failures, Overload, Attacks, and the Unexpected (Resilient Worlds)” and by the Federal Ministry of Education and Research (BMBF), grant number (16KISR015K), ”RePro” project.

REFERENCES

- [1] European Commission, ”Industry 5.0,” Online, 2021, luxembourg: Publications Office of the European Union. PDF ISBN: 978-92-76-25308-2, DOI: 10.2777/308407. [Online]. Available: <http://europa.eu>
- [2] S. Jhansale Anil Kumar, A. A. Zargar, D. Chen, R. Raghunath, E. Jorswieck, and L. Wolf, ”Resilience-state inference from wireless telemetry in industrial communication systems,” in *2026 International Conference on Future and Intelligent Networking (FINE)*. IEEE, 2026, pp. 262–269. [Online]. Available: <https://doi.org/10.1109/FINE71157.2026.00050>
- [3] 3rd Generation Partnership Project (3GPP), ”5G; Service requirements for cyber-physical control applications in vertical domains,” 3rd Generation Partnership Project (3GPP), Tech. Rep. TR 22.104, 2022. [Online]. Available: <https://www.3gpp.org/specifications-technologies/releases/release-17>
- [4] J. Bento, P. Saleiro, A. F. Cruz, M. A. T. Figueiredo, and P. Bizarro, ”TimeSHAP: Explaining recurrent models through sequence perturbations,” in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD ’21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 2565–2573. [Online]. Available: <https://doi.org/10.1145/3447548.3467166>
- [5] J. Crabbé and M. van der Schaar, ”Explaining time series predictions with dynamic masks,” in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 139. PMLR, 2021, pp. 2166–2177. [Online]. Available: <https://proceedings.mlr.press/v139/crabbé21a.html>
- [6] K. K. Leung, C. Rooke, J. Smith, S. Zuberi, and M. Volkovs, ”Temporal dependencies in feature importance for time series prediction,” in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: <https://openreview.net/forum?id=C0q9oBc3n4>
- [7] C. Fiandrino, E. Pérez Gómez, P. Fernández Pérez, H. Mohammadizadeh, M. Fiore, and J. Widmer, ”AIChroneLens: Advancing explainability for time series AI forecasting in mobile networks,” in *IEEE INFOCOM 2024 – IEEE Conference on Computer Communications*. IEEE, 2024, pp. 1521–1530. [Online]. Available: <https://doi.org/10.1109/INFOCOM52122.2024.10621134>
- [8] P. Marantis, K. Ramantas, and C. Verikoukis, ”RAN explainable anomaly prediction for 6G networks,” in *2024 IEEE 29th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. IEEE, 2024, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/CAMAD62243.2024.10942895>
- [9] C. Fiandrino, L. Bonati, S. D’Oro, M. Polese, T. Melodia, and J. Widmer, ”EXPLORA: AI/ML explainability for the open RAN,” *Proceedings of the ACM on Networking*, vol. 1, no. CoNEXT3, pp. 19:1–19:26, Dec. 2023. [Online]. Available: <https://doi.org/10.1145/3629141>
- [10] C. Sun, U. Pawar, M. Khoja, X. Foukas, M. K. Marina, and B. Radunovic, ”SpotLight: Accurate, explainable and efficient anomaly detection for open RAN,” in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, ser. ACM MobiCom ’24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 923–937. [Online]. Available: <https://doi.org/10.1145/3636534.3649380>
- [11] O. T. Basaran and F. Dressler, ”XAIomaly: Explainable and interpretable deep contractive autoencoder for O-RAN traffic anomaly detection,” *Computer Networks*, vol. 261, p. 111145, Apr. 2025. [Online]. Available: <https://doi.org/10.1016/j.comnet.2025.111145>
- [12] J. M. Ramírez, F. Díez, P. Rojo, V. Mancuso, and A. Fernández-Anta, ”Explainable machine learning for performance anomaly detection and classification in mobile networks,” *Computer Communications*, vol. 200, pp. 113–131, Feb. 2023. [Online]. Available: <https://doi.org/10.1016/j.comcom.2023.01.003>
- [13] P. Casas, S. Wassermann, M. Seufert, N. Wehner, O. Dinica, and T. Hoßfeld, ”X-ray goggles for the ISP: Improving in-network web and app QoE monitoring with deep learning,” in *6th IFIP Network Traffic Measurement and Analysis Conference (TMA)*, Enschede, The Netherlands, Jun. 2022. [Online]. Available: <https://dl.ifip.org/db/conf/tma/tma2022/tma2022-paper30.pdf>
- [14] T. Sharma, T. Mangla, A. Gupta, J. Jiang, and N. Feamster, ”Estimating WebRTC video QoE metrics without using application headers,” in *Proceedings of the 2023 ACM on Internet Measurement Conference*, ser. IMC ’23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 485–500. [Online]. Available: <https://doi.org/10.1145/3618257.3624828>
- [15] Y. Wang, M. Lyu, and V. Sivaraman, ”Standardizing multimedia QoE telemetry from telecommunications networks for open analytics,” in *Proceedings of the 2024 SIGCOMM Workshop on Emerging Multimedia Systems*, ser. EMS ’24. New York, NY, USA: Association for Computing Machinery, 2024, pp. 14–20. [Online]. Available: <https://doi.org/10.1145/3672196.3673400>