DELFT UNIVERSITY OF TECHNOLOGY · EEMCS · CSE3000 BACHELOR RESEARCH PROJECT

Outlier and Anomaly-Handling for 6G Wireless Measurement Data

A systematic, downstream-centric comparison of statistical filters and unsupervised outlier detectors for tabular and time-series 6G network measurements

AUTHOR
Matei-Alexandru Stanescu

RESPONSIBLE PROFESSOR
Rihan Hai

AFFILIATION
EEMCS, Delft University of Technology

01 MAIN QUESTION

How do different denoising and outlier-handling methods compare in preserving data utility for machine-learning tasks when applied to noisy 6G network measurements across tabular and time-series modalities, and, based on this comparison, which strategy is most suitable for the dominant noise and anomaly types observed in each modality?

02 Background

METHODS

STATISTICAL	ML-BASED
Robust Z-score filtering	Isolation Forest
IQR-based clipping	Local Outlier Factor
Savitzky-Golay smoothing	PCA reconstruction

EVALUATION CRITERIA

1. Downstream performance

TABULAR Macro-F1 · ROC-AUC

TIME-SERIES RMSE · MAE

2. Detection diagnostic

Tabular dataset carries a ground-truth anomaly label — does the statistical-outlier notion these detectors encode align with the real anomalies present in the data?

3. Runtime

Results interpreted against the 10 ms near-RT RIC threshold — superior downstream performance is not always worthy.

GAP **No comparison of statistical vs ML in network data.**

03 Methodology

No clean reference → judge by downstream effect, not fidelity. No synthetic noise; all vs NoClean.

Two probes: RF (noise-tolerant) vs k-NN, k = 15 (noise-sensitive) → isolates the cleaning effect.

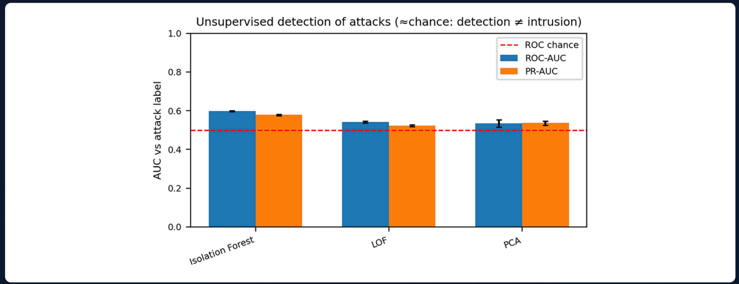
Tabular — 5G-NIDD: attack vs benign · 5 seeds × 10k, 80/20 → repair: drop+impute (detectors) / in-place (filters) → macro-F1, ROC-AUC.

Time-series — KPI/lat99: 1-step forecast, 15-ch KPI, causal, forward-fill → Blocked CV ×5 (≈stationary): 25 reps → RMSE, MAE.

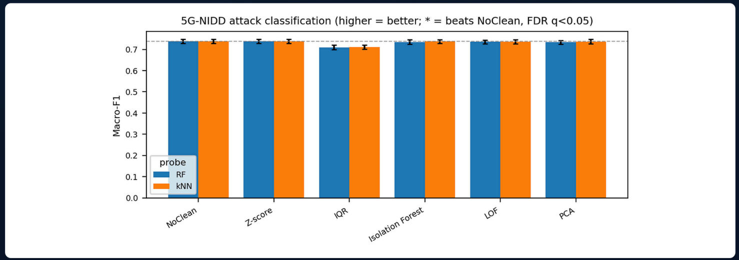
Detection diagnostic: anomaly score as soft classifier → ROC-AUC / PR-AUC — non-oracle Optuna tuning · paired t-test vs NoClean, BH-corrected.

04 Results

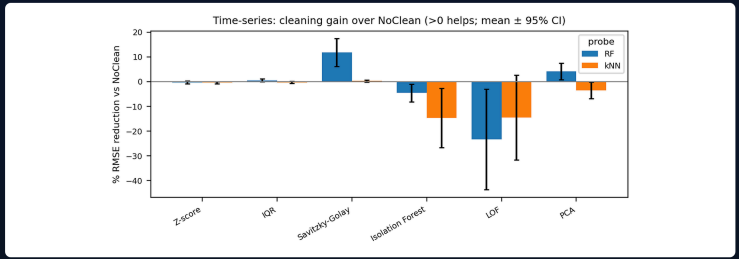
Detection diagnostic



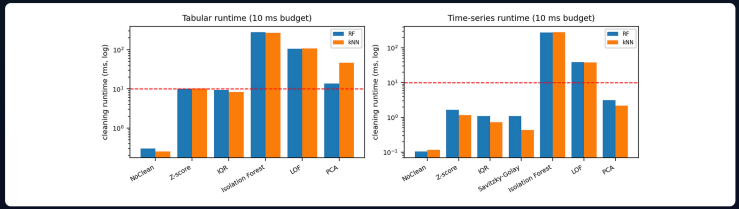
Tabular — attack classification



Time-series — gain over NoClean



Runtime — 10 ms budget



05 Discussions

Detection (Fig. 1): detectors barely above chance (ROC-AUC 0.54–0.60) vs supervised RF ≈0.86 → outlier detection ≠ intrusion detection.

Tabular (Fig. 2): 0/10 cleaners beat NoClean; IQR significantly worse (q = 0.009) → clean null, clipping harmful.

Time-series (Fig. 3): 0/12 beat NoClean after FDR; only Sav–Golay suggestive (≈17% lower RF RMSE) but fails correction.

Runtime (Fig. 4): filters / Sav–Golay / PCA within 10 ms budget; LOF & IForest too slow — costliest methods are also the most harmful.

06 Conclusions

Bottom line: generic cleaning gives no reliable benefit → must be demonstrated, not assumed.

Not the model: same null under RF & k-NN → property of data + task.

Why: real anomalies are structured signal, not rare points → detectors miss them, repair loses info.

Recommendation: skip cleaning by default, judge by downstream performance — only exception: a light causal smoother; costliest methods also most harmful → drop.

Contribution + next: leakage-free real-anomaly framework; future → more datasets, stronger repairs, stealthier attacks, joint accuracy–latency.

07 Limitations

External validity: one dataset per modality, a 5G/edge proxy for 6G; time-series folds are windows of a single series → generalises across temporal regimes, not independent series.

Balanced subsample (not the natural ~61% attack rate) → macro-F1 at a balanced operating point; attacks are DoS/scan-dominated, so the mismatch may shrink for stealthier, rarer intrusions.

Limited power: 5 tabular seeds, 25 overlapping-difficulty time-series reps → read as “no reliable effect detected,” not “no effect.”

One-step-ahead forecasting only: longer horizons or richer lag structure could change cleaning’s value.

Future work: more datasets & independent series, stronger non-destructive repairs, stealthier rarer attacks, joint accuracy–latency optimization.