Anomaly Detection and Root Cause Analysis of Retainability Contextual Anomalies in Radio Access Networks

David Santos[1(✉)], Paula Amaral[1,2], João Martins[3], and Rui Ramos[3]

[1] NOVA School of Science and Technology, University Nova de Lisboa, Lisbon, Portugal
dsm.santos@campus.fct.unl.pt
[2] NOVA Math CMA, University Nova de Lisboa, Lisbon, Portugal
[3] Multivision, Lisbon, Portugal

Abstract. The increasing complexity of mobile communication networks, driven by the rapid growth of connected devices and data traffic, has significantly challenged network monitoring and troubleshooting for Mobile Network Operators. Traditional fault diagnosis approaches rely heavily on manual analysis, which is time-consuming, error-prone, and difficult to scale. This work proposes an automated framework for anomaly detection and root cause analysis in Radio Access Networks (RAN) using Machine Learning techniques. The approach integrates unsupervised anomaly detection methods with explainability tools to identify abnormal patterns in Key Performance Indicators (KPIs) and uncover their underlying causes. Dimensionality reduction and clustering techniques are employed to reveal intrinsic data structures, while model interpretability methods, such as SHAP, provide insights into feature importance and the decision-making process. Initial experimental results are promising. The proposed methodology shows potential in detecting anomalous behavior and provides preliminary support for root cause identification, although this aspect is still under development. Ongoing work focuses on refining the identification of underlying causes, with current results indicating a positive trajectory. This progress is expected to contribute to reducing troubleshooting time, improving operational efficiency, and enhancing overall network reliability.

Keywords: Root Cause Analysis · Machine Learning · Radio Access Networks · SHAP

1 Introduction

The widespread adoption of mobile devices has driven a significant expansion of mobile communication infrastructures to support an increasingly large and demanding user base. To ensure reliable operation, Mobile Network Operators (MNOs) must continuously monitor network performance and promptly diagnose potential faults.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16766, pp. 259–269, 2027.
https://doi.org/10.1007/978-3-032-30539-8_19

Despite this need, a substantial portion of the diagnostic process relies on manual investigation, with only limited automation available. As a result, troubleshooting remains a slow and resource-intensive task. The time required to resolve network issues can vary considerably depending on their complexity, often requiring engineers to spend hours analyzing performance data before identifying the root cause.

In this context, the development of automated tools to assist in network analysis becomes essential. Such tools can streamline the troubleshooting workflow, enabling engineers to handle a larger number of incidents in less time, while improving the overall efficiency and performance of the network.

Recent advances in Machine Learning (ML) have created new opportunities for automating network monitoring tasks, particularly in the areas of Anomaly Detection (AD) and Root Cause Analysis (RCA). These techniques allow for the detection of anomalous patterns within large volumes of performance data and provide insights into the factors contributing to network degradation.

In this paper, we present the progress achieved so far towards a data-driven approach for anomaly detection and root cause analysis in Key Performance Indicators (KPIs). The proposed methodology leverages unsupervised learning techniques to detect anomalies in Key Performance Indicators (KPIs), combined with explainability methods to interpret the results and highlight the most relevant contributing factors. As this is ongoing work, particular emphasis is placed on the anomaly detection component, while the root cause analysis capabilities are still being refined.

The main contributions of this work are:

- A scalable framework for anomaly detection in RAN performance data using unsupervised machine learning techniques;
- The integration of explainable AI methods to support root cause analysis, currently under development;
- An experimental evaluation demonstrating the potential of the proposed approach in real-world scenarios.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 introduces the main concepts. Section 4 presents the proposed methodology. Section 5 discusses the results, and Sect. 6 concludes the paper.

2 Related Work

Some work has been done regarding automated processes for root cause analysis in radio access networks.

Cílino et al. [1] proposed a method to identify the causes of low downlink throughput by clustering Shapley Additive Explanations (SHAP) explanations obtained from a boosting classifier, and then labelling the resulting clusters with the corresponding root causes. This allowed for classifications of the causes on the whole dataset, however not in new observations.

In [4], Gómez-Andrades et al. developed an automatic diagnosis system for Long Time Evolution (LTE) networks using a Self-Organizing Map (SOM). After training the SOM, neurons were grouped through hierarchical clustering and new observations were classified according to the cluster of their best matching unit. The method was able to capture meaningful structures in Key Performance Indicators (KPIs) and reduce manual analysis, although its performance depends heavily on feature selection and assumptions about the statistical behavior of the metrics.

Ferreira et al. [3] analyzed the root causes of low accessibility using Principal Component Analysis (PCA) combined with feature-importance metrics. Cells with Evolved Radio Access Bearer (E-RAB) setup failures above the 90th percentile were labeled as faulty. Although feature importance could be traced back to the original variables, the percentile-based fault definition and the linearity of PCA limited the effectiveness of the approach, particularly at the individual cell level.

Recent work has explored Graph Neural Network (GNN) for root cause analysis in RAN environments. The Simba framework proposed in [5] combines a graph learning module, a Graph Convolutional Network (GCN), and a transformer to capture both spatial and temporal dependencies in the KPI time series. Trained on simulated data, the model achieved high F1-scores while identifying different fault types such as interference and excessive power reduction. Similarly, Yen et al. [7] introduced a framework that integrates Graph Structure Learning with GNN models, including GCNs and Graph Attention Networks (GANs), to infer dependencies between cells and classify faults from multivariate KPI time-series data. Although these GNN-based approaches achieve strong performance, their high computational cost limits scalability to large-scale networks.

In this paper, a different approach is proposed. SHAP values are used due to their capacity to generate interpretable insights from models trained on unsupervised data. In addition to providing clear explanations of model predictions, they enable the systematic assessment of feature importance, thereby allowing for root cause analysis.

3 Preliminary Concepts

In the context of mobile communication networks, particularly in Radio Access Networks (RAN), KPIs represent measurements that reflect the performance, quality, and reliability of the network. These metrics are continuously collected from network elements and provide engineers with insights into overall network behavior. A few of common examples of RAN KPIs include:

- **Throughput** – the rate at which data is successfully transmitted;
- **Latency** – the delay experienced during data transmission;
- **Packet loss rate** – the percentage of data packets that fail to reach their destination;
- **Call drop rate** – the proportion of calls that are unexpectedly terminated.

This paper focuses on the KPI: **E-RAB Retainability−Percentage Lost**, which will be referred to as "retainability" for brevity. It measures the percentage of established E-RABs that are abnormally released, therefore lower values indicate better performance.

3.1 Anomalies

Traditionally, MNOs use simple heuristics for detecting anomalies in KPIs where data points are flagged as anomalous whenever they exceed some predefined threshold. In the case of retainability, a common threshold is 2%, however, this criterion is very far from encompassing the type of faults that are actually relevant for maintaining a well-performing network. In reality, cells have distinct temporal patterns that can't be accounted for by simple heuristics. Thus, any anomaly detection method for this application should have into consideration the seasonal patterns typical to each cell. In general, cells tend to have much lower retainability at night, therefore a value that would be considered normal at midday may be anomalous if it occurs at midnight.

Another crucial aspect to consider is that cells may degrade momentarily, producing isolated anomalies in time before returning to normal operation, and such cases do not require intervention. The focus should therefore be on degradations that persist over time, since these are the ones that require corrective action. An anomaly score function S is a function from the set of data points $\mathcal{D}$ to the real line $\mathbb{R}$ that quantifies how anomalous each data point is.

Definition 1 (Sporadic Anomaly). *For a time series* $\{x_i\}_{i=1}^n$*, let* $S : \mathcal{D} = \{x_1, \ldots, x_n\} \to \mathbb{R}$ *be an anomaly score function and let* $T \in \mathbb{R}$ *be a given threshold. A data point* $x_i \in \mathcal{D}$ *is said to be a sporadic anomaly if* $S(x_i) \geq T$*.*

Definition 2 (Persistent Anomaly). *Given a time series* $\{x_i\}_{i=1}^n$ *and a window size* $k \in \mathbb{N}$*, any data point* x_i *such that* $x_{i-k+1}, x_{i-k+2}, \ldots, x_i$ *are sporadic anomalies is said to be a persistent anomaly.*

The only cells that need intervention are those with persistent anomalies, for the reasons specified above.

3.2 Shapley Additive Explanations

Shapley Additive Explanations (SHAP) is a framework for interpreting ML models that unifies several methods from the family of additive feature attribution models. It is particularly useful for explaining predictions of complex or black-box models. SHAP provides local explanations, meaning it attributes a contribution value to each feature for a specific prediction rather than only giving a global measure of feature importance. This makes it possible to understand how individual features influence the model's decision for a given observation. The explanation model in SHAP is defined as:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i, \tag{1}$$

where $z' \in \{0,1\}^M$ is a simplified binary representation of the input features, indicating the presence ($z' = 1$) or absence ($z' = 0$) of each feature. The term ϕ_0 represents the expected model output when no features are present, serving as a baseline. Each $\phi_i \in \mathbb{R}$ quantifies the contribution of feature i to the specific prediction under consideration, capturing how much that feature increases or decreases the model's output relative to the baseline. By summing these contributions, $g(z')$ approximates the original model's prediction for the given observation, allowing a clear interpretation of individual feature effects.

In the calculations of SHAP values, TreeExplainer [6] was used, which is optimized for tree-based models and offers significantly faster computation compared to the model-agnostic KernelSHAP.

4 Methodology

The dataset used includes real world data provided by Multivision, and comprises 11407 cells from a 4G network. It spans over a period of 21 d, and has granularity of 15 min. Each row in the dataset corresponds to a measurement from a specific cell at a given timestamp. Besides the cell identifier and timestamp columns, the dataset has one column for the retainability and 95 others for performance management counters, which were selected due to their potential to indicate root causes for anomalous retainability values.

The anomaly detection component described in Subsect. 4.1 operates on the time series of a single cell at a time. Prior to analysis, the time series is preprocessed by applying linear imputation to handle missing values. In contrast, the RCA model presented in Subsect. 4.2 takes as input both the explanatory performance counters and the retainability values. Unlike the AD model, no imputation is performed; instead, any rows containing missing values are removed.

4.1 Anomaly Detection

Seasonal-Trend Decomposition using LOESS (STL) [2] is a widely used statistical technique for decomposing a univariate time series into three components: trend, seasonality, and residuals (2). The trend component captures long-term changes in the series, the seasonal component captures periodic fluctuations, and the residual component contains the remaining variability not explained by the trend or seasonality. In the context of anomaly detection, the residual component is of particular interest, since it represents deviations from expected behavior after removing trend and seasonal components. STL was applied to decompose the retainability time series $\{y_i\}_{i=1}^n$ into trend t_i, seasonality s_i, and residuals r_i, such that

$$y_i = t_i + s_i + r_i. \tag{2}$$

Iterative robustness is incorporated by introducing weights that depend on the residuals and are updated across iterations. At each step, the decomposition is obtained by solving a weighted smoothing problem, where observations with large residuals are assigned smaller weights. This progressively reduces the

influence of anomalous points on the estimated trend and seasonal components, concentrating their effect in the residual term. This produces a decomposition with a clearer trend and seasonality and larger residuals for anomalous data points, facilitating their detection. Two anomaly scores were considered. The first score is the absolute value of the residual,

$$S(y_i) = |r_i|, \tag{3}$$

and the second score is the absolute value of the normalized residual,

$$\widetilde{S}(y_i) = \left| \frac{r_i - \bar{r}}{\widehat{\mathrm{Var}}(r)} \right|. \tag{4}$$

The score function $\widetilde{S}$ measures how anomalous an observation is relative to the variability of its own time series, allowing anomalies to be detected within each individual cell. However, relying solely on this score can lead to many non-relevant points being flagged as anomalies. For example, a cell with a very consistent daily cycle may be flagged as anomalous even if it deviates by a very small amount. Although this technically constitutes an anomaly relative to the cell's own behavior, such a small deviation is typically not meaningful for MNOs. For this reason, it was also necessary to consider the score S.

The data points of interest are those that are persistent anomalies according to both S and $\widetilde{S}$ relative to some thresholds $T_a \geq 0$ and $T_r \geq 0$, respectively. These thresholds are parameters of the procedure and should be chosen according to the desired sensitivity.

4.2 Root Cause Analysis

Cilínio et al. [1] categorize as anomalous any data point with throughput greater than 7 Mb/s, framing the task as a binary classification problem. However, this type of labeling is overly coarse, as it discards valuable information contained in the original KPI values. Instead of training a binary model to predict whether a data point is anomalous or not, we opt for a regression model. This approach preserves more information and provides greater flexibility in the choice of anomaly detection techniques. Therefore, a regression was performed on the raw KPI value and SHAP was applied only to those points that could be reliably predicted by the model.

First, the dataset was split into training, validation, and test sets. The split was performed temporally: the first 70% of the dataset was used for training, the next 15% for validation, and the final 15% for testing. The validations set was used for early stopping. Three gradient boosting models were trained to predict retainability: LightGBM, CatBoost, and XGBoost. Then the model achieving the lowest Root Mean Squared Error (RMSE) (5) was selected as the final predictor, however optimal predictive performance is not a strict requirement, as only the observations that are predicted with sufficient accuracy are subsequently

explained.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{5}$$

The reliability of SHAP explanations depends on the predictive accuracy of the underlying model: the smaller the prediction error for a data point, the more accurately the model captures its behavior and, consequently, the more informative the corresponding SHAP explanations become. For this reason, the explanation phase was restricted to observations that the model predicted with sufficient accuracy. To enforce this constraint, an error threshold T_e was defined, and a data point was considered well explained whenever

$$|y_i - \widehat{y}_i| \leq T_e, \tag{6}$$

where $\widehat{y}_i$ denotes the model's prediction for the i-th observation. The value of T_e was chosen based on practical considerations, reflecting both network operational constraints and the relevant business objectives.

The SHAP values were then processed using Uniform Manifold Approximation and Projection (UMAP), a dimensionality reduction technique that preserves the global and local structure of the data while projecting it into a lower-dimensional space. After dimensionality reduction, the data was clustered using Density-Based Spatial Clustering of Applications with Noise (DBSCAN), enabling the identification of groups of observations with similar SHAP value patterns. Each resulting cluster was subsequently labeled according to the most likely underlying cause of the anomaly.

Only cases with persistent anomalies are of interest to Mobile Network Operators (MNOs), therefore in order to determine the root cause of each cluster, an analysis was conducted on only the persistent anomalies within each cluster using the most broad criteria which consists in having three consecutive sporadic anomalies. This allows for any criteria to be used, because any cell with persistent anomalies needs to have at least three consecutive sporadic anomalies.

5 Results

All experiments were conducted on a standard computing environment equipped with an Intel Core i5-2520M processor and 16 GB of RAM, without the use of specialized hardware such as GPUs.

STL was used to decompose the time series, with the period parameter set to 96, corresponding to the number of data points in a day given that the granularity of the time series is 15 min. The threshold T_r was set to 2 in order to flag as anomalous any data points that deviate from the seasonal pattern more than 2 standard deviations, and T_a was set to 0.5 so that only absolute deviations greater than that are considered. This resulted in effective anomaly detection, as illustrated by the examples in Fig. 1. Specifically, the cell with ID "UNIBNCL" exhibits only sporadic anomalies, whereas cell "SHOPNC2L" shows multiple

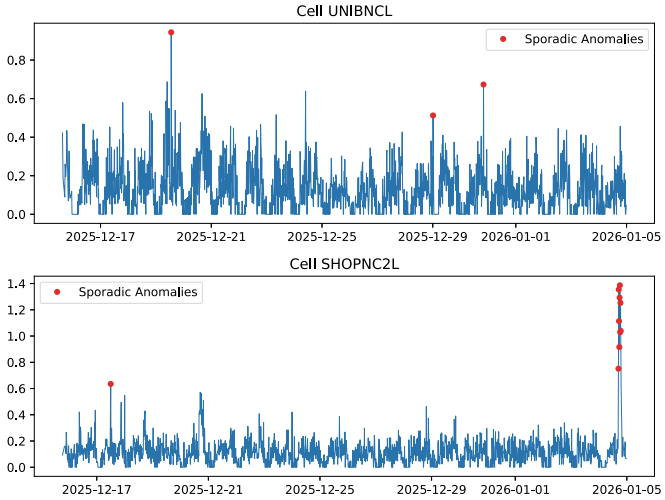

Fig. 1. Retainability time series for cells UNIBNCL (top) and SHOPNC2L (bottom) with flagged sporadic anomalies highlighted in red. (Color figure online)

sporadic anomalies toward the end of its time series, which could be classified as persistent depending on the chosen window size.

In the RCA phase, among the trained models, LightGBM achieved the lowest RMSE of 0.318, while XGBoost obtained 0.342 and CatBoost 0.357. Hence, LightGBM was selected as the final predictor.

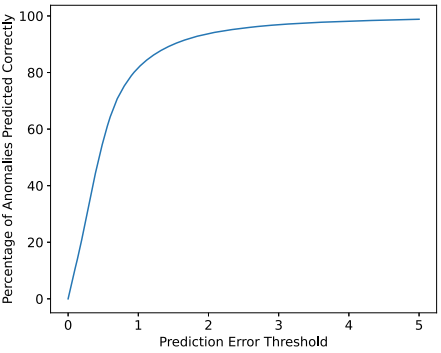

Fig. 2. Percentage of well-predicted anomalies plotted against prediction error threshold T_e.

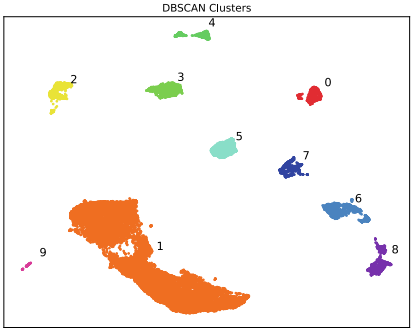

Fig. 3. Clusters of the embedded SHAP values.

An important aspect in evaluating the effectiveness of the model is the proportion of anomalies that can be explained. To illustrate how the explanation

coverage depends on the chosen threshold, Fig. 2 shows the percentage of anomalies that are well-predicted by the model for different values of the prediction error threshold. As T_e increases the percentage of well-predicted observations rises extremely quickly, meaning this threshold can be quite conservative and still capture a broad range of anomalous cases. The error threshold T_e was set to 1, under which 81% of the sporadic anomalies were well-predicted by LighGBM and could therefore be explained. After computing the SHAP values for all well-predicted anomalies, UMAP was applied, and the resulting two-dimensional representations were then clustered using DBSCAN with the parameters present in Table 1. This resulted in the clustering illustrated in Fig. 3, where 10 clearly distinct clusters are visible.

Table 1. Parameters used for UMAP and DBSCAN

Parameter	Description	Value
UMAP		
`n_neighbors`	Number of neighboring points used in local approximations of the manifold structure.	15
`min_dist`	Minimum distance between embedded points in the low-dimensional space.	0
`n_components`	Number of dimensions of the embedding space.	2
`metric`	Distance metric.	euclidean
DBSCAN		
`eps`	Maximum distance between two samples for them to be considered as neighbors.	0.8
`min_samples`	Minimum number of samples required to form a dense region (core point).	15
`metric`	Distance metric.	euclidean

Cases exhibiting three consecutive anomalies were analyzed to identify their root causes. Clusters containing persistent anomalies were clusters 1, 3, 6, 7, and 9, therefore the analysis focused exclusively on these clusters. Three distinct groups were found. One of these groups, Group A, maps directly to cluster 7, meanwhile the other two groups, Groups B and C, are spread across clusters 1, 3, 6, and 9. The root cause associated with Group A is an automated cell downtime event, consistent with a hardware or software fault activating the eNodeB supervision mechanisms within the 15-minute window. This is shown by a strong elevation in automated cell downtime, a simultaneous drop in coverage sampling, and a sharp increase in abnormal User Equipment (UE) context releases triggered by the downtime condition. Since this type of fault is indicative of a hardware issue, the radio engineer should redirect it to the field operations team for on-site investigation and corrective actions. Group B is characterised by Physical Random Access Channel (PRACH) and Contention-Based Random

Access (CBRA) congestion. The affected cells exhibit chronically low CBRA success rates well below the population median, while Contention-Free Random Access (CFRA) success rates remain near 100%. This contrast rules out a radio unit fault (where contention-free random access would also degrade under hardware or coverage failure) and points instead to saturation of the shared random access channel, most likely caused by high active UE density or a suboptimal Random Access Channel (RACH) configuration. In this context, the radio engineer should first verify the RACH configuration parameters, and if these are correctly set, the next step is to investigate the presence of temporary high UE concentration events (e.g., concerts, sports events) and apply appropriate optimization actions. Group C presents isolated retainability spikes against otherwise healthy KPI profiles, with no structural failure signature in adjacent counters. These anomalies are attributable to short-duration interference bursts or load spikes confined to the minimum detection window of 45 min and are not expected to persist under a more stringent threshold, making them candidates for exclusion from further investigation. In such cases, the radio engineer should keep the affected cells under observation to ensure that the degradation does not recur or evolve into a persistent issue.

The methodology successfully distinguished Group A, corresponding to cluster 7, from all other possible root causes. The SHAP values clustering isolated these events unambiguously because the model consistently attributed the elevated retainability prediction to downtime-related features, confirmed by the underlying counter evidence. While Groups B and C required indirect characterisation through cell-level counter analysis independently of the clustering result, the SHAP-based clusters having established only that these events were explained differently from Group A, the contrast between chronically depressed CBRA success rates and otherwise healthy structural indicators provided a principled basis for separating structural channel congestion from transient load events. The obtained execution times demonstrate the practical feasibility of the proposed methodology for real-world applications. Specifically, the anomaly detection stage required approximately 40 min, while the training phase of the root cause analysis model took around 19 min. Once trained, the RCA model exhibited a very low inference latency, requiring only 0.08 s to analyze a single data instance. These results indicate that, despite moderate offline training costs, the proposed framework is well-suited for near real-time root cause analysis in operational Radio Access Network environments, where rapid decision-making is critical.

6 Conclusion

In this paper, we presented a Machine Learning-based framework for anomaly detection and root cause analysis in Radio Access Networks. By combining unsupervised anomaly detection techniques with explainability methods, the proposed approach enables the identification of abnormal network behavior and provides meaningful insights into its underlying causes.

The results demonstrate that the framework is capable of effectively detecting anomalies in network performance data and supporting engineers in diagnosing issues more efficiently. The use of interpretability tools, such as SHAP, proved particularly valuable in understanding model decisions and highlighting the most relevant Key Performance Indicators contributing to anomalies.

Overall, this work shows that integrating anomaly detection with explainable AI can significantly enhance troubleshooting processes in modern communication networks, reducing manual effort and improving operational performance.

Future work may include the integration of real-time monitoring capabilities, the exploration of advanced deep learning models, and the extension of the approach to other domains within telecommunications networks.

Acknowledgments. This work is funded by national funds through the FCT – Fundação para a Ciência e a Tecnologia, I.P., under the scope of the projects UID/00297/2025 (https://doi.org/10.54499/UID/00297/2025) and UID/PRR/00297/2025 (https://doi.org/10.54499/UID/PRR/00297/2025) (Center for Mathematics and Applications – NOVA Math). This work was supported by the European Union via the European Regional Development Fund (ERDF) under the PORTUGAL 2030 and COMPETE 2030 Programme [COMPETE2030-FEDER-01176700].

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cilínio, M., Duarte, D., Vieira, P., Pereira, M., Mata, L.: Explainable fault analysis in mobile networks: a SHAP-based supervised clustering approach (2023). https://doi.org/10.1109/ICSPCS58109.2023.10261152
2. Cleveland, R.B., McRae, J.E., Terpenning, I.: STL: a seasonal-trend decomposition procedure based on loess. J. Off. Stat. **6**(1), 3–73 (1990)
3. Ferreira, D., et al.: Root Cause Analysis of Reduced Accessibility in 4g Networks. IEEE Press (2020)
4. Gómez-Andrades, A., Muñoz, P., Serrano, I., Barco, R.: Automatic root cause analysis for LTE networks based on unsupervised techniques. IEEE Trans. Veh. Technol. **65**(4), 2369–2386 (2016). https://doi.org/10.1109/TVT.2015.2431742
5. Hasan, A., Boeira, C., Papry, K., Ju, Y., Zhu, Z., Haque, I.: Root cause analysis of anomalies in 5g ran using graph neural network and transformer (2024)
6. Lundberg, S.M., et al.: From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. **2**(1), 2522–5839 (2020)
7. Yen, C., et al.: Graph neural network based root cause analysis using multivariate time-series KPIs for wireless networks. In: NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium. IEEE Press (2022). https://doi.org/10.1109/NOMS54207.2022.9789858