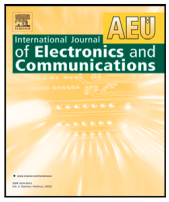
Int. J. Electron. Commun. (AEÜ) 181 (2024) 155339

Contents lists available at ScienceDirect

Int. J. Electron. Commun. (AEÜ)

journal homepage: www.elsevier.com/locate/aeue



Unraveling the root causes of faults in mobile communications: A comparative analysis of different model explainability techniques

M. Cilínio [a,b,*], M. Pereira [a,b], D. Duarte [a,b], L. Mata [a,b,c,d], P. Vieira [b,d]

[a] Celfinet- A Cyient Company, R. João Chagas 53, Lisbon 1495-072, Portugal
[b] Instituto de Telecomunicações, Av. Rovisco Pais 1, Lisbon 1049-001, Portugal
[c] Instituto Superior Técnico, Av. Rovisco Pais 1, Lisbon 1049-001, Portugal
[d] Instituto Superior de Engenharia de Lisboa, R. Conselheiro Emídio Navarro 1, Lisbon 1959-007, Portugal

ARTICLE INFO

Keywords:
Mobile networks
Root cause analysis
Machine learning
Explainable AI
SHAP

ABSTRACT

The escalating demand and complexity of monitoring services handled by Network Operations Centers (NOCs) have led Mobile Network Operators (MNOs) to prioritize automated solutions for network fault detection and diagnosis. Consequently, various Machine Learning (ML)-based Root Cause Analysis (RCA) systems have been developed, however their lack of explainability poses a challenge due to the predominantly black-box nature of ML models. This paper addresses this issue by presenting a supervised clustering methodology capable of integrating both glass-box and black-box models, the latter complemented by post-hoc explainability techniques. While black-box models excel in predictive capabilities, necessitating post-hoc techniques for explainability, glass-box models prioritize transparent decision-making, fostering a clearer understanding of the model's behavior. This work delineates a methodology for performing RCA of faults in the User Downlink (DL) Average Throughput Key Performance Indicator (KPI), simultaneously comparing the performance of black-box models (Light Gradient-Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost)) with glass-box models (Logistic Regression (LR) and Explainable Boosting Machine (EBM)). Results revealed that the LightGBM black-box algorithm coupled with the SHapley Additive exPlanations (SHAP) method demonstrated superior performance in fault detection and diagnosis, without compromising the overall explainability. Consequently, it was possible to identify faults related to radio conditions, low network usage in specific user groups, low network capacity, and mobility issues. The paper concludes with practical mitigation strategies for each identified fault cluster.

1. Introduction

The Coronavirus Disease 2019 (COVID-19) pandemic accelerated society's digitalization, leading to increased traffic demand and stringent Quality of Service (QoS) requirements for MNOs. Despite significant CAPital EXpenditures (CAPEX) investments in network improvement over the past decade, major European MNOs have shown minimal growth in customer revenues. This challenges the long-term sustainability of the mobile telecommunications business, prompting a shift in operational models for NOCs from traditional approaches to a new paradigm called Smart Operations. This concept comprises new applications and methods aimed at replacing the MNO's traditional operative model and reducing OPerational EXpenditures (OPEX) through a more efficient usage of all network resources and infrastructures. Hence, it comes as no surprise that many recent research works have focused on several problems that are posed by the creation of intelligent NOCs which, relying on Artificial Intelligence (AI) algorithms, are key enablers to automated techniques that optimize resource utilization and improve QoS. These techniques include automated fault analysis, preventive fault detection, self-regenerative actions, and reduction of overall energy consumption.

A challenge for MNOs, considering the need to reduce OPEX, is the development of models capable of learning patterns from network-generated data, which may allow them to automate manual repetitive tasks while maintaining the high quality of service delivered to the end users. However, despite the recent success of AI techniques in creating models to address many different applications, a common criticism is the lack of explainability in the generated outputs [1]. For the specific purpose of wireless communications applications, this can hamper the adoption of such models by MNOs, unless additional modules that provide interpretable and explainable results are available.

* Corresponding author at: Celfinet- A Cyient Company, R. João Chagas 53, Lisbon 1495-072, Portugal.
E-mail addresses: madalena.cilinio@cyient.com (M. Cilínio), marcio.pereira@cyient.com (M. Pereira), david.duarte@cyient.com (D. Duarte), luis.mata@tecnico.ulisboa.pt (L. Mata), pedro.vieira@isel.pt (P. Vieira).

https://doi.org/10.1016/j.aeue.2024.155339
Received 28 December 2023; Accepted 15 May 2024
Available online 18 May 2024

While explainability unravels the model's decision-making process, interpretability guarantees that humans can easily comprehend the model's outputs.

Thus, the main objective of this work is to develop an explainable and interpretable ML approach for identifying faults in mobile networks. The objective is to speed up fault diagnosis, allowing engineers to move from detailed root cause searches to a validation role, trusting the model's transparent and easy-to-understand results.

In this context, the main contributions of this work are summarized as follows: (i) Implements a modular architecture capable of accommodating different fault detection models with different levels of explainability to support the RCA module, (ii) Offers interpretable visualizations of faults, facilitating the differentiation of various groups with distinct root causes and (iii) Introduces an optimization framework based on the categorization of symptoms within each fault cluster. Additionally, the use of real network data enhances the result's reliability, mitigating typical data access limitations in this research area.

This paper is organized as follows. After the introduction, Section 2 overviews the related work. Section 3 presents the proposed methodology. Section 4 discusses and analyzes the main results. Finally, Section 5 presents the main conclusions and research guidelines for future work.

2. Related work

This section delves into the related literature, briefly introducing the fundamental concepts of ML and eXplainable Artificial Intelligence (XAI), examining their associated challenges, and conducting a review of their application in the RCA of faults in mobile networks.

ML techniques encompass a wide range of models, making them a versatile tool in the field of AI. These models vary from sophisticated black-box models, renowned for their predictive abilities but requiring post-hoc techniques to yield explainable outcomes, to glass-box models that offer transparent decision-making, thereby facilitating human understanding of the model's behavior.

As the endeavor to extract trust from ML models increases, so does the focus on model explainability. However, recent studies [2,3] have highlighted that the pursuit of explainability may imply a trade-off resulting in reduced model performance. This finding highlights the importance of balancing model transparency with predictive performance in real-world applications.

Therefore, understanding the fundamental principles of XAI holds significant importance in the context of RCA for mobile networks, as AI-based processes must exhibit strong predictive capabilities in fault detection while also providing comprehensible diagnostics for humans. This process is vital for swiftly addressing network faults, enabling MNOs to easily identify issues and implement efficient intervention strategies. The relevance of this topic is underscored by the increasing body of literature dedicated to automating RCA tasks, notably through the use of ML techniques. One widely adopted method involves using Supervised Learning (SL) models for predicting failures and focuses on identifying key predictive features to diagnose the underlying root causes. A practical implementation of this approach can be found in [4], where the authors identify potential causes of reduced accessibility in Fourth Generation (4G) networks by testing both black-box and glass-box fault classification algorithms and subsequently evaluating the model's feature importance. Additionally, in [5], the authors propose a framework for identifying the primary causes of low DL throughput in mobile networks, leveraging the SHAP method to assess feature importance. In another approach discussed in [6], an architecture based on Deep Neural Network (DNN) is introduced to analyze the causes of poor throughput in Third Generation (3G) networks, with the Local Interpretable Model-Agnostic Explanations (LIME) technique being employed to evaluate the contribution of each input variable. An alternative approach involves the implementation of Unsupervised Learning (UL) models to identify clusters of failures with common causes. In a related study [7], the authors present a diagnostic system for 4G networks based on Self-Organizing Maps (SOM) and Ward's hierarchical method. Another solution proposed in [8] relies on clustering both traffic and signaling channel KPIs using the Expectation Maximization (EM) algorithm.

The current work applies a recent methodology, firstly introduced in [9], and further developed in [10], to analyze the root causes of poor DL throughput in mobile networks. Leveraging previously outlined strategies, this methodology employs supervised clustering to forecast real-time fault events, aiming to identify clusters with shared symptoms. The strength of this methodology lies in the incorporation of a clustering diagnostic mechanism stemming directly from the supervised fault detection model. Specifically, it uses the feature importance values obtained from the fault detection model to conduct clustering. This fusion of supervised learning and clustering techniques ensures a robust and insightful analysis of fault patterns. Thus, both black-box models, interpreted through the SHAP method, and glass-box models were employed to compare the outcomes of both approaches.

3. Research methodology

This section outlines the research methodology for conducting RCA of mobile network faults through a supervised clustering approach. The methodology comprises three main phases: data characterization, fault detection, and fault diagnosis, as illustrated in Fig. 1. In the data characterization phase, the collected data underwent processing and cleansing. Moving to the fault detection phase, a comparison was made between two sets of models: black-box models and glass-box models. The best-performing model for each set was then selected. Finally, in the fault diagnostic phase, clustering was applied to the feature importance values extracted from both the black-box model (using the post-hoc SHAP method) and the glass-box model (extracted directly from the model). These values measure the relevance of each feature in prediction-making, aiding in the identification of fault clusters with similar root causes. This consideration takes into account the potential for common problems resulting from similar combinations of these values. Each of these three phases is detailed below.

3.1. Data characterization

For this work, a total of 223 Performance Management (PM) indicators were used to detect and diagnose faults in the User DL Average Throughput KPI. The data was collected from 19 4G Base Stations (BSs), consisting of 51 cells with the characteristics described in Table 1. The data extraction was performed with an hourly granularity, over 28 days. Since the data was collected from a live network, being subject to extraction errors, it was necessary to perform several data preprocessing steps that included removing rows with missing values and data artifacts.

3.2. Fault detection

The fault detection process was automated using binary classification models. These models used the 223 PM indicators as input variables, X, to estimate a mapping function, f, which predicted a binary value of fault/non-fault in the KPI under study, Y, by inferring $f(X)$.

3.2.1. Dataset labeling for fault definition

The binary variable was established using a fault threshold value determined by the MNO to meet its network's requirements for the User DL Average Throughput KPI. Consequently, Class 1 was assigned to DL Average Throughput values below 7 Mbps (indicating a fault), while Class 0 was assigned to KPI values above 7 Mbps (indicating a non-fault).

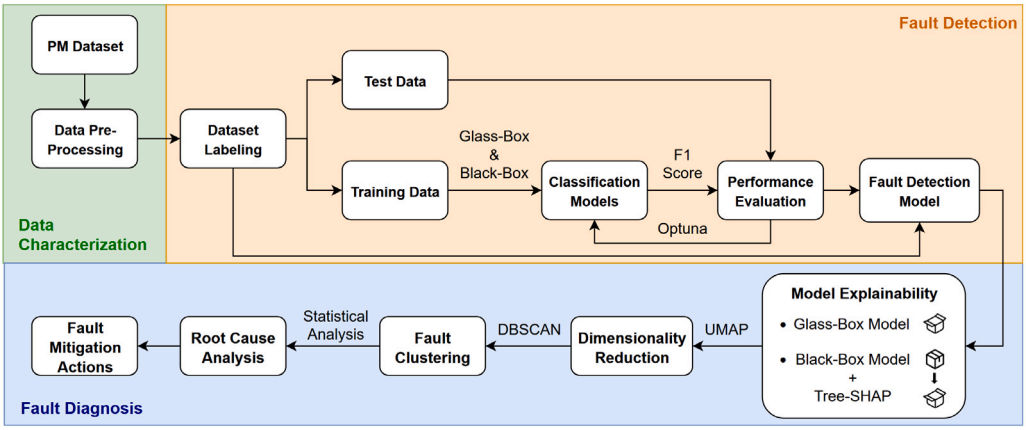


Fig. 1. Proposed methodology.

Table 1
Frequency band characteristics from the used dataset (51 cells).

N° of Cells	Frequency band	Channel bandwidth	Physical Resource Blocks (PRBs)
38	800 MHz	10 MHz	50
7	900 MHz	5 MHz	25
3	1800 MHz	20 MHz	100
3	2100 MHz	10 MHz	50

3.2.2. Fault detection models: Explainable approaches

Throughout the implementation of binary classification models, two distinct model sets were evaluated. The first set included the black-box models XGBoost [11] and LightGBM [12], while the second set consisted of the glass-box models LR [13] and EBM [14]. The XGBoost and LightGBM are both gradient boosting models that use a set of multiple decision trees that are sequentially trained to improve the system's overall predictive ability. While XGBoost uses a level-wise tree growth strategy, considering all possible splits at each level, LightGBM employs a leaf-wise approach, adding new leaves with the maximum improvement in loss, resulting in faster training and potentially deeper trees. Both algorithms were chosen for their intrinsic capabilities to address data imbalance, which is an essential consideration in scenarios involving the detection of uncommon faults, as well as for their capacity to integrate the SHAP method. Furthermore, they often outperform other models in terms of predictive accuracy and execution speed.

In turn, LR is a linear model well-known for its simplicity, explainability and robust application in various domains. LR's explainability lies in its linear equations with feature-specific coefficients, enabling a clear feature impact understanding, simple output-to-probability conversion and an absence of complex interactions, making it ideal for transparent decision-making. Conversely, EBM operates as a tree-based, cyclic gradient boosting model, essentially embodying a Generalized Additive Model (GAM) with automatic interaction detection. While GAMs offer more flexibility than linear models, they are less flexible than most ML algorithms. However, their restricted additive structure offers the advantage of perfect explainability. During prediction, each feature contributes with a single score, allowing for exact local feature importance measures that can be easily compared and interpreted [15]. The EBM model was thus designed to maintain a level of predictive performance on par with the Boosted Trees algorithms while being highly interpretable [14]. Therefore, the choice of these two glass-box models was based on LR's widespread use and computational efficiency, along with EBM's effective modeling of non-linear relationships and high-accuracy predictions.

3.2.3. Model performance evaluation

Once the classification models to be tested were determined, it was necessary to define an evaluation metric that should consider the imbalance between both classes, since the occurrence of faults in mobile networks is significantly less frequent when compared to their normal operation. The evaluation metric F1-score was thus selected, as in [16], where a comparison of various evaluation metrics applied to imbalanced datasets is conducted. The F1-score metric assumes a maximum value of 1 in the best evaluation scenario and a minimum value of 0 in the worst scenario, being calculated from the harmonic mean between the precision and recall values. While precision quantifies the number of correct positive predictions, recall measures the number of correct positive predictions made from all positive predictions. The optimization of the F1-score value for each algorithm was achieved using the Optuna software framework [17].

3.3. Fault diagnosis

After the fault detection phase, the methodology advanced to diagnose the identified faults. In this stage, a supervised clustering approach was implemented, using the feature importance weights derived from the best-performing black-box and glass-box fault detection models. The goal was to pinpoint clusters showcasing common fault patterns, characterized by mutual sets of relevant features that might indicate shared symptoms.

3.3.1. Black-box model explainability: Post-Hoc method

Given the unintelligible nature of black-box models, a post-hoc explainability technique was selected to extract the feature weights from the best-performing algorithm. Therefore, the SHAP method was chosen for its advantages over other techniques, such as LIME. Its strengths include providing consistent global explanations, solid theoretical grounding in Shapley value estimation, versatile applicability across ML frameworks, and a focus on fairness and equity in feature importance attributions. Moreover, this technique automatically rescales raw data and assigns weights based on importance, reducing the impact of irrelevant features on the model's predictions [9]. Thus, the SHAP method [18], based on the theory of cooperative games, explains the model outputs by calculating the contribution of each input variable in the produced predictions. To this end, the method calculates the Shapley values, given by the average of all variable's contributions, considering all possible coalitions. The Shapley values

are thus represented as an additive variable attribution method, whose explanatory model, $g(z')$, is given by:

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j \quad (1)$$

where $z' \in \{0,1\}^M$ represents the vector of combinations, M defines the maximum dimension of the coalition, and $\phi_j \in \mathbb{R}$ corresponds to the feature attribution of a given variable j [19]. In this work, it was used a faster and more accurate variation of the SHAP method for ML models based on decision trees called Tree-SHAP.

3.3.2. Dimensionality reduction

After computing the SHAP values for the best-performing black-box model and obtaining the internal feature importance weights for the leading glass-box model, the scores underwent processing using a non-linear data dimensionality reduction technique called Uniform Manifold Approximation and Projection (UMAP) [20]. The implementation of this technique involves two stages. Firstly, the method learns the structure of the data in a high-dimensional space. Secondly, it finds a representation of the data in a reduced-dimensional space by minimizing the cross entropy. The main advantages of this dimensionality reduction technique include its low computational cost and the preservation of the local and global structure of the data. Thus, the feature score data was reduced to a two-dimensional (2D) space to simplify its visualization and subsequent analysis.

3.3.3. 2D data clustering

Thereafter, the obtained 2D feature importance values were clustered using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) model [21]. This model's results rely on the initial set of two constant parameters: the maximum neighborhood distance between two samples (ϵ), and the minimum number of samples required in a neighborhood for a point to be considered a central point ($MinPts$) [22]. The set of elements of D contained in a region of radius (ϵ) and central point p is thus described by [23]:

$$Eps(p) = \{q \in D | distance(p,q) \leq \epsilon\} \quad (2)$$

where $distance(p,q)$ denotes the Euclidean distance between points p and q belonging to the dataset D. The obtained clusters were evaluated using the Silhouette Score, which measures the average similarity within clusters and their proximity to neighboring ones. While a score of 1 denotes highly distinct clusters, 0 suggests overlap or a weaker structure, and -1 implies incorrect cluster assignments [24]. Lastly, an analysis of the statistical characteristics of each fault cluster was conducted to determine the underlying causes and establish appropriate mitigation measures.

4. Results and discussion

This section not only highlights the outcomes from the implementation of the methodology presented earlier, but also offers an in-depth discussion and analysis of these results.

4.1. Fault detection results

This subsection presents the results covering the data characterization and fault detection stages. Thus, once the data pre-processing and labeling steps were completed, the black-box and glass-box models were implemented, using 75% of the dataset to train the classification algorithms, and the remaining 25% to test their performance. Table 2 summarizes the distribution of samples by class used in the training and testing phases, indicating the presence of an imbalanced dataset.

The hyperparameter values used to train the binary classification models were then obtained using the Optuna software tool, which optimized the algorithm's input parameters to maximize the F1-Score metric. Fig. 2 illustrates the confusion matrices obtained for each model. Hence, comparing the confusion matrices in Fig. 2, which led to the precision, recall, and F1-score values in Table 3, it turns out that the LightGBM model stood out in the prediction of User DL Average Throughput faults among the 'black-box' models, while the EBM excelled among the 'glass-box' models. Thus, these two models were chosen for the succeeding fault clustering phase.

Table 2
Data split between training and test sets with the respective class distribution.

	Class 0 (Non-Faults)	Class 1 (Faults)	Total
Train Set	18505	5634	24139 (75%)
Test Set	6226	1821	8047 (25%)
Total	24731 (77%)	7455 (23%)	32186

Table 3
Precision, Recall and F1-score values obtained for each model.

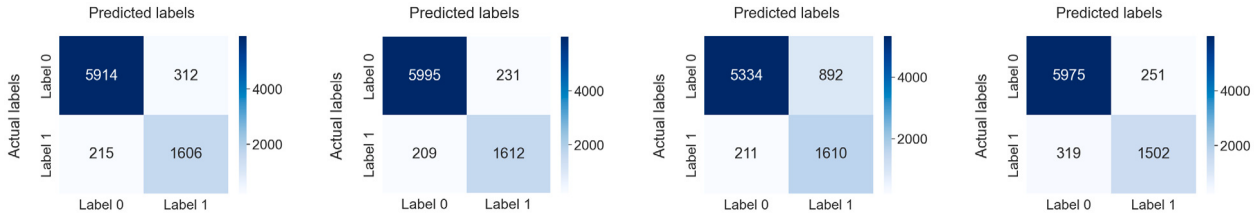
Model Explainability	Model	Precision	Recall	F1-Score
Black-Box	XGBoost	0.84	0.88	0.86
	LightGBM	0.87	0.89	**0.88**
Glass-Box	LR	0.64	0.88	0.75
	EBM	0.86	0.83	**0.84**

4.2. Fault supervised clustering results

This subsection presents the results covering the fault diagnosis stage excluding the RCA and fault mitigation steps. Thus, after concluding the fault detection phase, where a 'black-box' and a 'glass-box' model were trained to successfully predict faults in the User DL Average Throughput KPI, the methodology progressed to the diagnostic stage. In this analysis, two distinct strategies were applied to enhance model explainability. For the EBM model, the approach involved directly extracting the internal feature weights from the model's predictions, while for the LightGBM algorithm, the post-hoc Tree-SHAP method was employed. In both scenarios, the models were implemented with the optimal parameterization determined by Optuna, using the entire dataset spanning one month. The feature scores were then transformed into a 2D representation using the UMAP dimensionality reduction technique. Following this transformation, the DBSCAN clustering algorithm was implemented to group the feature weight values for each model. Fig. 3 shows the 2D representations of the raw indicator values, the SHAP values derived from the LightGBM model, and the internal feature weights obtained from the EBM model. Additionally, it also highlights the clusters identified by the DBSCAN algorithm for each model. While the plot of the raw variables depicted a scattered and poorly structured representation of the two sample classes, the 2D representations obtained for the LightGBM and EBM models revealed clear separation of clusters for each class, highlighting the advantages of this methodology.

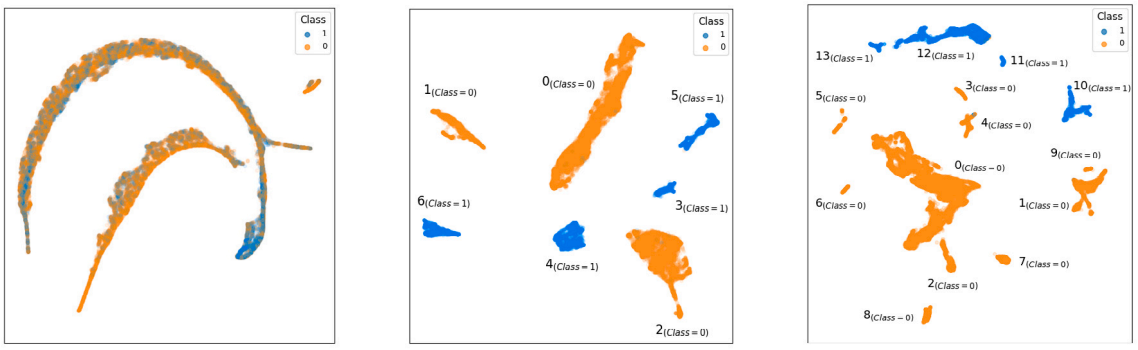
The quality of the obtained clusters was evaluated using the Silhouette Score. Consequently, the LightGBM + Tree-SHAP approach achieved a score of 0.55, surpassing the EBM model's score of 0.31. This indicates the superior performance of the LightGBM + Tree-SHAP approach, not only in fault detection but also in cluster formation. Additionally, in contrast to the LightGBM + Tree-SHAP, the EBM model exhibited clusters exclusively populated by samples with erroneous predictions. Specifically, the model's predictions obtained in Fig. 3(c) for clusters 3 and 9 indicated the absence of fault instances, whereas in reality, all samples proved to be faulty. Conversely, the opposite scenario unfolded for cluster 11. Consequently, the EBM model yields unreliable outcomes, as explainability cannot be inferred from poorly made predictions.

Therefore, for this particular study, the comparison between both approaches has highlighted the advantages of utilizing a black-box model with a post-hoc explainability method. This approach not only



(a) XGBoost (b) LightGBM (c) LR (d) EBM

Fig. 2. Confusion matrices for both 'black-box' (XGBoost and LightGBM) and 'glass-box' (LR and EBM) models.



(a) Raw Values (b) LightGBM + TreeSHAP (c) EBM

Fig. 3. 2D representation of each observation raw values color-coded by their actual classes, SHAP values color-coded by LightGBM predicted classes, and EBM internal feature weights color-coded by the predicted classes.

Table 4
Clusters generated by the DBSCAN model with $\epsilon = 1.2$ and $MinPts = 20$.

Dominant class	Cluster	N° of samples	DBSCAN cluster coordinates in 2D
0 (Non-Fault)	0	13747	[7.22 , 10.63]
	1	2988	[−4.28 , 9.60]
	2	8008	[−12.60 , −1.23]
1 (Fault)	3	986	[12.89 , 4.49]
	4	2794	[5.01 , 0.53]
	5	1966	[15.86 , 9.38]
	6	1697	[−5.71 , 1.13]

improves predictive performance but also enhances the separation of clusters, all while still ensuring predictions remain explainable. The reliability of LightGBM explainability is ensured by its strong predictive capabilities along with the robustness of the SHAP method. Consequently, the RCA was exclusively based on the examination of the clusters obtained in Fig. 3(b), described in greater detail in Table 4. The identified fault clusters encompass 7443 of the 7455 Class 1 samples, with the remaining fault samples having been appended to the non-fault clusters.

4.3. Root cause analysis

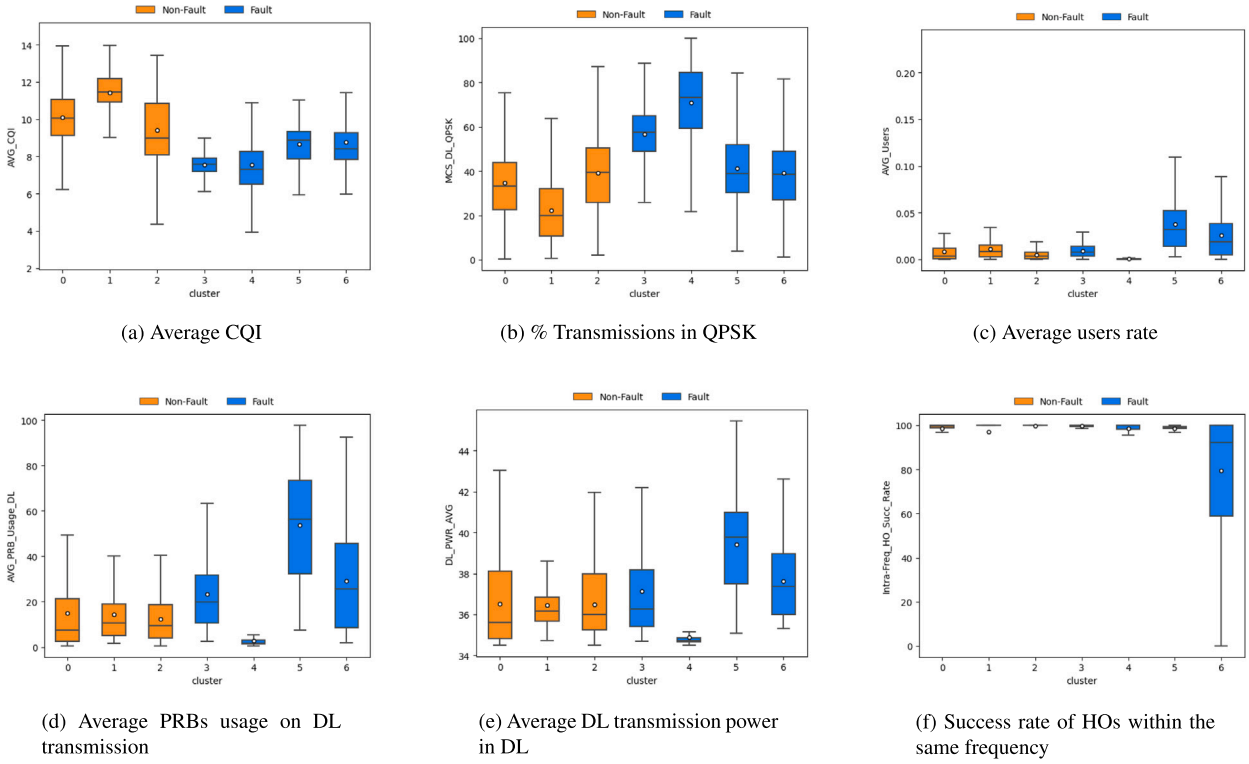
Finally, this subsection presents the results encompassing the RCA and fault mitigation phases within the fault diagnosis stage. The RCA phase was based on the statistical analysis of each fault cluster, focusing on the most critical PM indicators for diagnosing faults in the User DL Average Throughput KPI. Based on the expertise of NOC engineers, the following PM indicators were selected for thorough analysis:

- **Channel Quality Indicator (CQI):** The average CQI value stores information regarding the radio channel conditions, spanning from 0 (poorest conditions) to 15 (optimal conditions), as determined by the Signal-to-Noise and Interference Ratio (SINR) measurements;
- **Average Modulation and Coding Scheme (MCS):** The MCS index determines the modulation scheme for DL transmissions based on the reported CQI value;
- **Average Users Rate:** This indicator measures the average number of users in a cell, being used to evaluate its demand and occupancy rate. The average number of active users was divided by the total number of available PRBs to ensure fair comparability across cells with different channel bandwidths;
- **DL Average Transmission Power:** This parameter measures the average DL transmit power (dBm) assigned by the eNodeB in each Transmission Time Interval (TTI);
- **Average PRBs Usage on DL:** This indicator evaluates the cell's used capacity;
- **Handover Success Rate:** This metric offers mobility insights by assessing both Intra-frequency and Intra-Radio Access Technology (RAT) successful handovers.

Therefore, Fig. 4 presents a compilation of the key PM indicators analyzed during the diagnostic phase. The diagnostic analysis primarily focused on the four fault clusters. However, for comparison purposes, the three non-fault clusters were also included in the boxplots. Thus, it was possible to identify the main differentiating characteristics for each fault cluster. This information, along with the incidence of each cluster in the network and the corresponding diagnosis are summarized in Table 5.

Consequently, as diagnostic outcomes indicate, two primary causes were identified for the User DL Average Throughput KPI faults: poor Radio Frequency (RF) conditions (Clusters 3 and 4, with fault incidences of 13.23% and 37.48%, respectively) and limited network capacity (Clusters 5 and 6, with fault incidences of 26.37% and 22.76%, respectively). Despite some clusters sharing a common diagnosis, their distinct symptoms potentially suggest different underlying issues.

Fig. 5 illustrates a spatio-temporal analysis of the occurrences of failures, with Fig. 5(a) specifically showing the distribution of faults by cluster for each day of the month, represented as a time-series. Therefore, Cluster 4 faults were predominant throughout the entire



(a) Average CQI

(b) % Transmissions in QPSK

(c) Average users rate

(d) Average PRBs usage on DL transmission

(e) Average DL transmission power in DL

(f) Success rate of HOs within the same frequency

Fig. 4. Statistical comparison of the most critical PM indicators by clusters.

Table 5
Diagnosis of each fault cluster.

Cluster	Fault incidence	Diagnosis	Main characteristics
3	13.23%	RF Conditions	Low CQI and MCS index with high QPSK (low order) transmission.
4	37.48%		Low CQI and MCS index with extremely high QPSK transmission. Extremely low number of active users and PRB usage.
5	26.37%	Capacity Problems	High number of active users and PRB usage. High transmission power in DL.
6	22.76%		High PRBs usage and numerous handover faults.

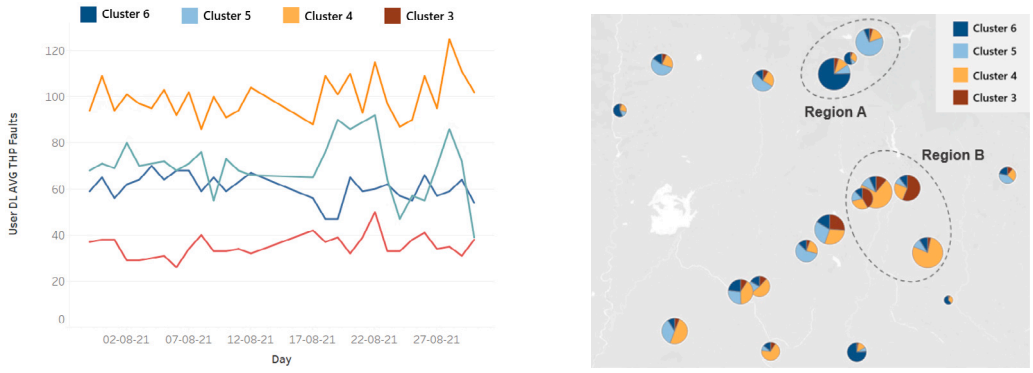
month, consistently exhibiting the highest incidence. Conversely, faults belonging to Cluster 3 consistently had the lowest incidence. In turn, Fig. 5(b) illustrates the spatial distribution of fault clusters across the 19 BSs. The pie charts were sized proportionally to the total number of faults per site. Thus, out of the total 19 sites analyzed, 11 sites were primarily associated with low network capacity issues, aligning with the pattern observed in the BSs of Region A. Meanwhile, the remaining 8 sites were primarily diagnosed with RF channel faults, as observed with the BSs in Region B.

Addressing the challenges posed by non-deterministic factors in mobile networks, including simultaneous network usage, multipath conditions, interference, and signal variations influenced by user location, some fault mitigation techniques were developed in collaboration with NOC engineers, who manually validated some of the results obtained. Their implementation is expected to improve network performance and speed up service restoration for users. These mitigation techniques, outlined in Table 6, aim to effectively tackle the identified faults by considering factors such as network usage patterns, interference management and signal optimization. Despite collaborative efforts with NOC engineers to assess the methodology's real-time performance on a large scale, obtaining these results has been hindered by network access limitations.

5. Conclusions

The automation of mobile networks monitoring through the development of self-detection and self-diagnosis algorithms is becoming increasingly relevant in the mobile communications domain. However, a typical criticism of the development of exclusively ML-based algorithms designed for mobile network performance regeneration is their lack of explainability. In this context, this work aims to address the transparency issues in fault detection and diagnosis systems in mobile networks. To achieve this, a supervised clustering framework comprising three phases was employed. The first phase involved data characterization, entailing its processing and cleansing. Subsequently, the User DL Average Throughput fault detection phase was initiated, where the performance of two sets of models was assessed: the black-box models LightGBM and XGBoost (explained using the post-hoc SHAP technique), and the glass-box models LR and EBM. The best model for each set was then selected. Finally, in the fault diagnostic phase, clustering was applied to the feature importance values obtained from the best-performing black-box and glass-box models to identify sets of faults with similar root causes.

Therefore, the proposed methodology was used to accelerate the RCA of faults in the User DL Average Throughput KPI, with the most favorable results being achieved by employing the LightGBM model alongside the SHAP method. By utilizing this approach, it was possible to diagnose 13.23% of faults related to radio conditions problems. Additionally, 37.48% of faults were observed in areas characterized by low network usage, where the degradation of performance within a specific user group impacted the overall site performance. Another 26.37% of faults were attributed to network capacity issues, while the remaining 22.76% experienced mobility issues in conjunction with



(a) Temporal analysis (b) Spatial analysis

Fig. 5. Spatio-Temporal analysis of faults by cluster.

Table 6
Mitigation Strategies for each fault cluster.

Cluster	Mitigation strategies
3	Since these faults show typical signs of RF issues, it is recommended to adjust antenna tilts and azimuths, reduce overshooting, or increase the inter-site distance.
4	As these issues mainly occur in sparsely populated areas, a single user's degradation can impact the site's overall performance. Analyzing base station traces and conducting drive tests to assess parameters like SINR, Received Signal Strength Indicator (RSSI) and Reference Signals Received Power (RSRP) is recommended.
5	Immediate action is not warranted as radio conditions remain unaffected. If they become compromised, it may be necessary to add new cells (e.g. small cells) or more carriers.
6	Immediate action is not warranted as radio conditions remain unaffected. If they deteriorate, a potential solution involves adjusting mobility parameters linked to Automatic Neighbor Relation (ANR), such as Hysteresis and Time-To-Trigger (TTT).

capacity problems. Additionally, drawing on the expertise of NOC engineers, it was possible to outline some mitigation strategies according to the characteristics of the diagnosed fault clusters.

In future research, this methodology will undergo real-time testing, expanding its scope to include other KPIs and incorporating data from Fifth Generation (5G) networks.

CRediT authorship contribution statement

M. Cilínio: Writing – original draft, Validation, Software, Methodology, Investigation. **M. Pereira:** Writing – review & editing, Validation, Software, Investigation. **D. Duarte:** Writing – review & editing, Resources, Investigation, Funding acquisition. **L. Mata:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition. **P. Vieira:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

The authors would like to acknowledge the importance of EUREKA Clusters project AI Call 2021 (AI2021-061 CANOPY), supported by PORTUGAL 2020 Partnership Agreement (SI I&DT Empresarial – Copromoção N° 181417). This work has also received funding from UIDB/00760/2020. Moreover, a heartfelt acknowledgment is due to the Instituto de Telecomunicações (IT) and CELFINET (a Cyient Company), who made this work possible.

References

[1] Gossen F, Margaria T, Steffen B. Towards explainability in machine learning: The formal methods way. IT Prof 2020;22(4):8–12. http://dx.doi.org/10.1109/MITP.2020.3005640.
[2] Hulsen T. Explainable artificial intelligence (XAI): Concepts and challenges in healthcare. AI 2023;4(3):652–66. http://dx.doi.org/10.3390/ai4030034.
[3] Sanii J, Chan WY. Explainable machine learning models for pneumonia mortality risk prediction using MIMIC-III data. In: 2022 9th international conference on soft computing & machine intelligence. ISCMI, 2022, p. 68–73. http://dx.doi.org/10.1109/ISCMI56532.2022.10068438.
[4] Ferreira D, Senna C, Salvador P, Cortesão L, Pires C, Pedro R, et al. Root cause analysis of reduced accessibility in 4G networks. In: Boumerdassi S, Renault E, Mühlethaler P, editors. 2nd international conference on machine learning for networking. Machine learning for networking, vol. LNCS-12081, Paris, France: Springer International Publishing; 2019, p. 117–33. http://dx.doi.org/10.1007/978-3-030-45778-5_9.
[5] Cilínio M, Duarte D, Vieira P, Queluz MP, Rodrigues A. Root cause analysis of low throughput situations using boosting algorithms and the TreeShap analysis. In: 2022 IEEE 95th vehicular technology conference. 2022, p. 1–5. http://dx.doi.org/10.1109/VTC2022-Spring54318.2022.9860734.
[6] Mampaka MM, Sumbwanyambe M. Poor data throughput root cause analysis in mobile networks using deep neural network. In: 2019 IEEE 2nd wireless Africa conference. 2019, p. 1–6. http://dx.doi.org/10.1109/AFRICA.2019.8843409.
[7] Gómez-Andrades A, Muñoz P, Serrano I, Barco R. Automatic root cause analysis for LTE networks based on unsupervised techniques. IEEE Trans Veh Technol 2016;65(4):2369–86. http://dx.doi.org/10.1109/TVT.2015.2431742.
[8] Rezaei S, Radmanesh H, Alavizadeh P, Nikoofar H, Lahouti F. Automatic fault detection and diagnosis in cellular networks using operations support systems data. In: NOMS 2016 - 2016 IEEE/iFIP network operations and management symposium. 2016, p. 468–73. http://dx.doi.org/10.1109/NOMS.2016.7502845.
[9] Cooper A, Doyle O, Bourke A. Supervised clustering for subgroup discovery: An application to COVID-19 symptomatology. In: Machine learning and principles and practice of knowledge discovery in databases. Springer International Publishing; 2021, p. 408–22. http://dx.doi.org/10.1007/978-3-030-93733-1_29.
[10] Cilínio M, Pereira M, Duarte D, Mata L, Vieira P. Explainable fault analysis in mobile networks: A SHAP-based supervised clustering approach. In: 2023 16th international conference on signal processing and communication system. 2023, p. 1–9. http://dx.doi.org/10.1109/ICSPCS58109.2023.10261152.
[11] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA: Association for Computing Machinery; 2016, p. 785–94. http://dx.doi.org/10.1145/2939672.2939785.

[12] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A highly efficient gradient boosting decision tree. In: Proceedings of the 31st international conference on neural information processing systems. Red Hook, NY, USA: Curran Associates Inc.; 2017, p. 3149–57.

[13] Molnar C. Interpretable machine learning. 2nd ed. Independently Published; 2022, Available online: https://christophm.github.io/interpretable-ml-book. [Last Accessed on 26 December 2023].

[14] Lou Y, Caruana R, Gehrke J, Hooker G. Accurate intelligible models with pairwise interactions. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA: Association for Computing Machinery; 2013, p. 623–31. http://dx.doi.org/10.1145/2487575.2487579.

[15] Nori H, Caruana R, Bu Z, Shen JH, Kulkarni J. Accuracy, interpretability, and differential privacy via explainable boosting. In: Meila M, Zhang T, editors. Proceedings of the 38th international conference on machine learning. Proceedings of machine learning research, vol. 139, PMLR; 2021, p. 8227–37, Available online: https://proceedings.mlr.press/v139/nori21a/nori21a.pdf. [Last Accessed on 26 December 2023].

[16] Jeni L, Cohn J, De la Torre F. Facing imbalanced data - Recommendations for the use of performance metrics. In: Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, vol. 2013, 2013, http://dx.doi.org/10.1109/ACII.2013.47.

[17] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining. 2019, http://dx.doi.org/10.1145/3292500.3330701.

[18] Lundberg S, Erion G, Chen H, DeGrave A, Prutkin J, Nair B, et al. Explainable AI for trees: From local explanations to global understanding. Nat Mach Intell 2020;2(1):56–67. http://dx.doi.org/10.1038/s42256-019-0138-9.

[19] 9.6 SHAP (SHapley Additive exPlanations) | Interpretable Machine Learning. 2022, Available online: https://christophm.github.io/interpretable-ml-book/shap.html. [Last Accessed on 09 November 2023].

[20] McInnes L, Healy J, Saul N, Grossberger L. UMAP: Uniform manifold approximation and projection. J Open Source Softw 2018;3:861. http://dx.doi.org/10.21105/joss.00861.

[21] Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the second international conference on knowledge discovery and data mining. AAAI Press; 1996, p. 226–31. http://dx.doi.org/10.5120/739-1038.

[22] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.

[23] Li Y, Yang Z, Jiao S, Li Y. Partition KMNN-DBSCAN algorithm and its application in extraction of rail damage data. Math Probl Eng 2022;2022:1–10. http://dx.doi.org/10.1155/2022/4699573.

[24] Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65. http://dx.doi.org/10.1016/0377-0427(87)90125-7.