IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 December 2023, accepted 24 January 2024, date of publication 1 February 2024, date of current version 12 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3361284

APPLIED RESEARCH

On the Use of Spatial Graphs for Performance Degradation Root-Cause Analysis Toward Self-Healing Mobile Networks

LUÍS MATA[1,2,3,4], MARCO SOUSA[3,4], PEDRO VIEIRA[1,3], MARIA PAULA QUELUZ[1,2], AND ANTÓNIO RODRIGUES[1,2], (Member, IEEE)

[1] Instituto de Telecomunicações, 1049-001 Lisbon, Portugal
[2] Instituto Superior Técnico, University of Lisbon, 1049-001 Lisbon, Portugal
[3] Instituto Superior de Engenharia de Lisboa, 1959-007 Lisbon, Portugal
[4] CELFINET—a CYIENT Group Company, 1495-764 Lisbon, Portugal

Corresponding author: Luís Mata (luis.mata@tecnico.ulisboa.pt)

This work was supported in part by the EUREKA Clusters Project AI Call 2021 (https://eureka-clusters-ai.eu/) through CANOPY under Grant AI2021-061, in part by the Cognitive and Automated Network Operations for Present and Beyond (Sistema de Incentivos á Investigação e Desenvolvimento Tecnológico (SI I&DT) Empresarial—Copromoção) under Grant 181417, and in part by the Norte Portugal Regional Operational Program (NORTE 2020) under the PORTUGAL 2020 Partnership Agreement through the European Regional Development Fund (ERDF).

ABSTRACT On the road to the sixth generation of cellular networks (6G), the need to ensure a sustainable usage of natural resources, amid increased competition and cost pressures, has driven the adoption of Self-Healing Mobile Networks to enhance operational efficiency of current and future wireless networks. This paradigm shift relies on Artificial Intelligence (AI) to increase automation of network functions, notably by applying predictive fault detection and automatic root-cause analysis. In this context, this paper proposes a Deep Learning (DL) model for self-healing operations based on a Spatial Graph Convolutional Neural Network (SGCN), which is applied to evaluate the performance degradation of Base Stations (BSs) and uncover the underlying root-causes. The advantages of the proposed DL model are threefold. Firstly, it is especially suited for wireless network applications, leveraging the SGCN to account for spatial dependencies among BSs and their physical characteristics. Secondly, the proposed model offers the flexibility to process diverse types of predictive features, including Performance Management (PM), Fault Management (FM), or other data types. Thirdly, it incorporates an explainability module that pinpoints the input features, such as PM counters, with the most significant influence on BS performance, thereby shedding light on its root-cause factors. The proposed model was evaluated on a live 4G network dataset and the results confirmed its effectiveness in identifying BS performance degradation. An F1-score of 89.6% was achieved in the classification of performance failures, which includes a 27% reduction in false negatives compared to prior research outcomes. In a live network environment, this reduction translates into substantial improvements in Quality of Experience (QoE) for the end users and cost savings for the Mobile Network Operators (MNOs).

INDEX TERMS Artificial intelligence, deep learning, self-healing operations, mobile network performance, root-cause analysis.

I. INTRODUCTION

Self-healing operations in telecommunications networks encompass a diverse range of applications where network data and AI are used to enhance operational efficiency. These applications are focused on identifying performance degradation (*e.g.,* using Key Performance Indicators (KPIs)), root-cause analysis, predictive fault detection, and self-healing actions [1], [2], [3], [4], [5].

Currently, the operational paradigm adopted by most MNOs involves initiating problem analysis and troubleshooting only after a service fault has been reported [6]. This

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa Zaman Chowdhury.

approach is inefficient and time-consuming, as it entails manually analysing a multitude of information sources, including network alarms, PM counters, configuration data, and other data types. Consequently, this leads to prolonged Mean Time to Repair (MTTR) values, resulting in reduced service availability and increased customer complaints. Self-healing operations applications use automated data analysis leveraging AI to predict network failures and their root-cause factors, enabling the design of preemptive actions to avoid the end users' QoE being impacted. Moreover, such applications yield positive environmental outcomes, potentially reducing carbon emissions and energy consumption, contributing to mobile network sustainability.

A challenge in wireless networks' self-healing operations is the development of models able to predict the BSs' performance, notably due to mutual coverage/interference levels considering the influence of neighbouring base stations and user mobility patterns. A possible solution to represent the spatial relationship between neighbouring BSs is using SGCNs, which combine Graph Neural Networks (GNNs) with spatial information associated with each BS. However, contrary to traditional methodologies oriented to the physical factors conditioning the wireless network performance, AI-based approaches, such as SGCNs, are often seen as very complex and hard to interpret, which may hamper their adoption by MNOs.

This paper proposes a DL model named as 4D-SGCN-Gx, which aims at classifying the BSs' performance and determining its root-cause performance factors. The proposed model is composed of two complementary modules: the first one relies on a SGCN to implement a binary classification model aimed at assessing BSs' downlink average throughput degradation; the second one, uses the saliency algorithm [7] to identify the root-cause factors of such performance degradation. The proposed model was evaluated using a dataset from a 4G live network, and the results were compared with previous related research provided in [8]. However, the methodology can be used with other wireless access technologies. The advantages of the proposed model are three-fold. Firstly, it is especially suited for self-healing operations in wireless networks, taking advantage of a four-dimensional (4D) spatial representation of BSs, encompassing the latitude, longitude, antenna height and antenna orientation, which is incorporated in the proposed model by using a SGCN. Secondly, it is technology agnostic and can be applied to a variety of self-healing operations applications, such as in network alarms prediction, or KPI classification using PM or FM data. Thirdly, the proposed model provides interpretable results, identifying the input features that contributed the most to classify the BSs' performance, thus providing immediate insights to is performance root cause factors. Although SGCNs have been successfully applied to different problems where the data presents a geometric structure, to the best of our knowledge, SGCN-based models have never been applied to performance degradation classification in live wireless networks, notably considering an explainability module and a 4D spatial representation of BSs, as in the proposed model. The main contributions of the paper are:

- Proposal of the 4D-SGCN-Gx model using graph-based representations of BSs to assess their performance by considering PM data.
- Inclusion of a 4D spatial representation of BSs, considering the BSs' height, antenna orientation, and geographical coordinates, which are exploited by the SGCN to enhance the 4D-SGCN-Gx's predictive capabilities.
- Development of an explainability module to assess the root-cause factors of BSs' performance degradation.

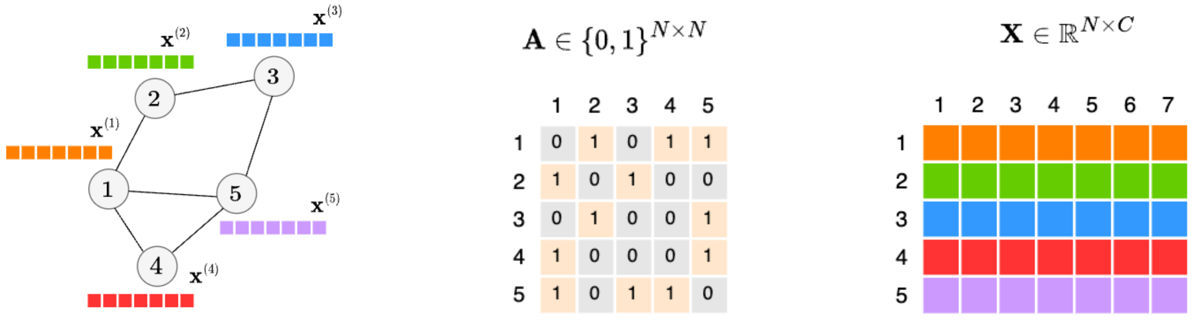
This paper is organized as follows. After the introduction provided in Section I, the theoretical background and related work are reviewed in Section II. In Section III, the 4D-SGCN-Gx model, designed for self-healing operations applications, is described. In SectionIV, the proposed model is evaluated and compared with previous research in a BS performance degradation classification problem, relying on downlink throughput analysis. Section V presents a detailed analysis of the model's results explainability, providing a comprehensive exploration of their interpretation and consistency. Section VI, investigates the versatility of applying the proposed model across diverse self-healing operation scenarios, by examining an additional performance degradation problem. In Section VII, the main conclusions are drawn, and future work is outlined.

II. THEORETICAL BACKGROUND AND RELATED WORK

This section provides an overview of GNNs and introduces the key concepts used in this paper, such as graph theory, convolution, spatial graphs and the range of problems that can be tackled with this approach. Notably, it emphasises the significance of using graphs to address performance optimization issues in mobile networks. Furthermore, a summary of related work presents the most relevant studies that have previously applied GNNs to similar problems, focusing on the distinctions with the approach presented in this paper.

A. GRAPH NEURAL NETWORKS

In recent times, a proliferation of research initiatives has focused on the development of GNNs extending DL approaches for the analysis of graph-based data. Graphs are versatile data structures, comprising nodes that are connected by edges. In reality, numerous systems can be effectively represented by graphs. For instance, chemical molecules, where nodes correspond to atoms, and edges symbolise chemical bonds. Similarly, in transportation networks like subway systems, grah nodes represent stations and edges depict the connecting routes. This concept extends to describing wireless networks, where BSs are represented by the graph nodes and edges can signify established neighbour relations among them. Beyond the graph structure, each node typically holds characteristic information, representing node features stored as node embeddings. For example, in a mobile network, each BS can be characterized by a vector of performance indicators at a specific time stamp. Furthermore,



(a) (b) (c)

FIGURE 1. Example of an undirected graph with $N = 5$ nodes and its embeddings with $C = 7$ characteristic features: a) Graph representation; b) Adjacency matrix; c) Node embeddings.

characteristic information can also be associated with edges, representing edge features, and stored as edge embeddings. For instance, edge information between BSs may encompass details related to daily handovers between neighbouring nodes or radio link indicators, such as path loss.

Formally, a graph comprises N nodes connected through a set of E edges. Graphs can be mathematically characterized using three matrices: $\mathbf{A}$, $\mathbf{X}$, and $\mathbf{E}$, which respectively represent the adjacency matrix, node embeddings, and edge embeddings. The adjacency matrix describes the graph's structural connections, where a_{ij} is 1 if there is an edge between node i and node j, and 0 otherwise. Moreover, if the edges have a direction the matrix $\mathbf{A}$ is asymmetric, indicating that the graph is directed (digraph). In this context, the presence of an edge from node 1 to node 2 does not imply the existence of an edge from node 2 to node 1. Conversely, when the edges lack direction, the graph is considered undirected, and in such cases, the matrix $\mathbf{A}$ is symmetric.

In Fig. 1-a), an undirected graph consisting of five nodes is depicted and its structure is defined by $\mathbf{A}$ in Fig. 1-b). For simplicity, in this example only the node embeddings, $\mathbf{X}$, are described, where each row represents the vector of characteristic features of the n^{th} node, $\mathbf{x}^{(n)}$, with size C. Upon defining an object as a graph, three types of problems can be tackled using GNNs, relying both on the graph's structure and its node embeddings [9]:

- **Graph-Level Tasks** - the objective is assigning labels or estimating values for the entire graph by aggregating the node embeddings. An example is the detection of anomalies in a mobile network by analysing the traffic flow between BSs (classification), or predicting the average Received Signal Strength Indicator (RSSI) in a wireless network (regression).
- **Node-Level Tasks** - the objective is assigning a label or predicting a value to each node of the graph. An example is the prediction of alarm occurrence intervals in BSs (classification), or predicting the energy consumption of BSs (regression).
- **Edge-Level Tasks** - involve making predictions or computations related to individual edges within a graph. An example could be predicting the path loss between two BSs (regression) or estimating whether two adjacent BS should be defined as neighbours (classification).

GNNs can be categorised into four groups considering their different approaches [10]:

- **Graph Recurrent Neural Networks (GRNs)** - learn node representations relying on recurrent DL architectures.
- **Graph Convolutional Neural Networks (GCNs)** - learn node representations by leveraging a convolutional-based DL architecture.
- **Graph Auto Encoders (GAEs)** - comprise unsupervised learning frameworks that encode nodes or graphs into a latent vector space and subsequently reconstruct graph data using the encoded information.
- **Spatial-temporal Graph Neural Networks (STGNs)** - include spatial and temporal dependencies to uncover complex data patterns.

A taxonomy of GNNs is depicted in Fig. 2, denoting the four main categories and specific branches, such as the spatial methods within the Graph Convolutional Neural Networks (GCNs) approach, also called SGCNs.

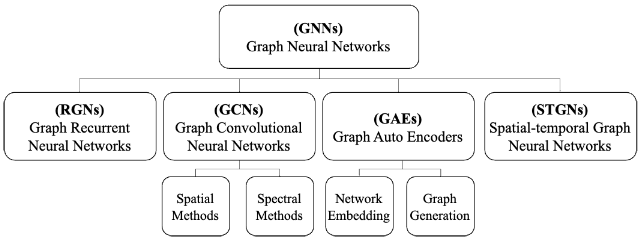
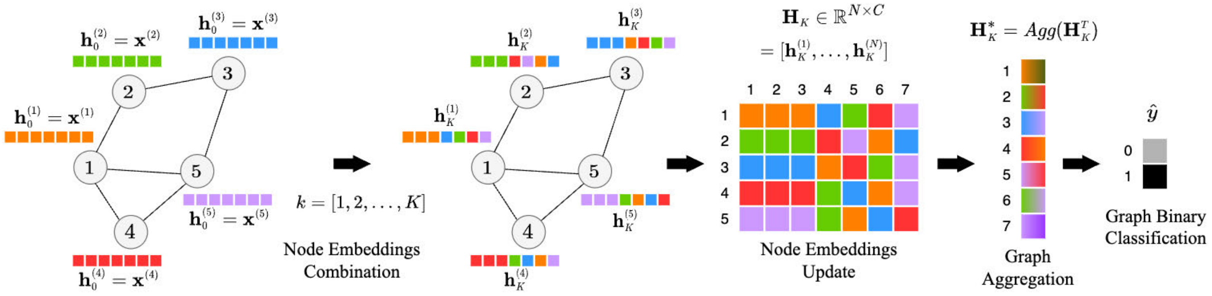


FIGURE 2. Taxonomy of GNNs (adapted from [10]).

Among the various types of GNNs, this paper focuses on SGCNs. Unlike other GCNs approaches, in SGCNs each node is associated with spatial information defined in a multi-dimensional space. Moreover, these models are composed of multiple convolutional layers, where the initial features



$\mathbf{h}_0^{(2)} = \mathbf{x}^{(2)}$
$\mathbf{h}_0^{(3)} = \mathbf{x}^{(3)}$
$\mathbf{h}_0^{(1)} = \mathbf{x}^{(1)}$
$\mathbf{h}_0^{(5)} = \mathbf{x}^{(5)}$
$\mathbf{h}_0^{(4)} = \mathbf{x}^{(4)}$
1
2
3
4
5
$k = [1, 2, \ldots, K]$
Node Embeddings Combination
$\mathbf{h}_K^{(2)}$
$\mathbf{h}_K^{(3)}$
$\mathbf{h}_K^{(1)}$
$\mathbf{h}_K^{(5)}$
$\mathbf{h}_K^{(4)}$
$\mathbf{H}_K \in \mathbb{R}^{N\times C}$
$= [\mathbf{h}_K^{(1)}, \ldots, \mathbf{h}_K^{(N)}]$
1 2 3 4 5 6 7
1 2 3 4 5
Node Embeddings Update
$\mathbf{H}_K^* = Agg(\mathbf{H}_K^T)$
1 2 3 4 5 6 7
Graph Aggregation
$\hat{y}$
0
1
Graph Binary Classification

FIGURE 3. Example of a graph-level classification using a SGCN.

of each node are sequentially updated across K iterations through convolution operations.

In SGCNs, the node embeddings, $\mathbf{X}$, are transformed by considering information from neighbouring nodes, defined through the matrix $\mathbf{A} \in \{0, 1\}^{N\times N}$, where N is the number of nodes in the graph. This process is conditioned on spatial information associated to the nodes, $\mathbf{P} \in \mathbb{R}^{N\times S}$, where S denotes the number of spatial dimensions considered. The spatial information matrix, $\mathbf{P}$ is given by:

$$\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N] \tag{1}$$

where the spatial information of the n^{th} node is denoted as $\mathbf{p}_n \in \mathbb{R}^S$. Each convolutional layer of the SGCN, $\mathbf{H}_k$, is defined by a function $f(\bullet)$ with parameters $\mathbf{U}$, where $\mathbf{H_0} = \mathbf{X}$ such that the intermediate node embeddings (hidden features) at each convolution layer, can be written as [9]:

$$\begin{aligned} \mathbf{H}_1 &= f(\mathbf{H}_0, \mathbf{U}_1, \mathbf{A}, \mathbf{P}) \\ \mathbf{H}_2 &= f(\mathbf{H}_1, \mathbf{U}_2, \mathbf{A}, \mathbf{P}) \\ \mathbf{H}_K &= f(\mathbf{H}_{K-1}, \mathbf{U}_K, \mathbf{A}, \mathbf{P}) \end{aligned} \tag{2}$$

In this process, unlike $\mathbf{H}$ and $\mathbf{U}$, which undergo changes at each convolutional layer, $\mathbf{A}$ and $\mathbf{P}$ remain unchanged and are employed for updating node embeddings. Fig. 3 depicts an example of an SGCN with K convolutional layers, where the objective is to perform graph-level classification. The process transforms the initial node embeddings, $\mathbf{H_0}$, into the final hidden representation $\mathbf{H}_K$. As illustrated, in each k_{th} convolutional layer, at each node i, information is collected from its neighbouring nodes j by combining their respective node embeddings $\mathbf{h}_{k-1}^{(j)}$, through a process called message passing. This results in a new representation that mixes information from adjacent nodes while considering their spatial positions, $\mathbf{p}_i$ and $\mathbf{p}_j$. For each k^{th} convolutional layer, the i^{th} node is updated according to:

$$\mathbf{h}_k^{(i)} = \sum_{j\in\Omega_i} u_k^{(ij)}(\mathbf{p}_j - \mathbf{p}_i)\mathbf{h}_{k-1}^{(j)} \tag{3}$$

where, $\Omega_i = \{j : a_{ij} = 1\}$ identifies the set of indices of the neighbours of node i, with $\mathbf{p}_i$ and $\mathbf{p}_j$ representing the spatial information vectors for nodes i and j, respectively. Additionally, $u_k^{(ij)}$ is a learnable parameter enabling the SGCN to assign distinct weights to individual neighbours during each convolution operation; to illustrate in a mobile network, certain neighbouring BSs may introduce more interference than others. The ultimate hidden representation of the node embeddings, denoted as $\mathbf{H}_K \in \mathbb{R}^{N\times C}$, is defined as:

$$\mathbf{H}_K = [\mathbf{h}_K^{(1)}, \mathbf{h}_K^{(2)}, \ldots, \mathbf{h}_K^{(N)}] \tag{4}$$

In the example of Fig. 3, where the goal is to perform a graph-level classification, the output node embeddings are aggregated using an aggregation function, such as computing the mean of the node embeddings. The resulting vector is then mapped via a linear transformation or neural network to a fixed-size vector denoted as $\mathbf{H}_k^*$. This process effectively reduces the graph's spatial resolution, providing an aggregated perspective of the entire graph:

$$\mathbf{H}_K^* = Agg(\mathbf{H}_K^T) \tag{5}$$

Finally, for the graph classification, the probability of each class is calculated by passing the vector $\mathbf{H}_k^*$ to the generalized linear model, also called softmax [9], which returns the normalised probabilities for each class. In binary classification, the softmax function simplifies to the logistic sigmoid function, yielding the graph's prediction $\hat{y}$.

B. RELATED WORK

The recent success of AI, namely Machine Learning (ML) and DL, has fostered an increasing adoption of learning-based solutions to plan, manage, and optimize wireless networks using path-loss predictions, such as in [11], [12], and [13]. AI can contribute to solving cornerstone problems in self-healing operations, such as performance management, predictive fault detection, automatic root-cause analysis and self-healing regenerative mechanisms. Some early works addressed dynamic network performance optimisation using Deep Reinforcement Learning (DRL) architectures [14], [15], [16], [17]. However, this approach requires large volumes of *a priori* labelled data, which is often unavailable. Other works have applied ML to supervised root-cause analysis in fixed networks [18], [19], and some recent examples have expanded this application to mobile networks [20], [21]. An application relying on real network data was proposed in [8] to analyse the root-cause of low throughput scenarios, using the TreeShap [22] analysis to provide an explainable interpretation of the results.

Recently, GCNs have become increasingly popular in a wide range of applications [10]. Some previous works attempted to use graph structures for network optimization, where the nodes may represent users, BSs, or antennas, and edges can represent network coverage quality using Received Signal Strength Indication (RSSI), interference or other metrics [23]. A few approaches have attempted to apply GCNs to model wireless network planning and performance optimization problems, relying on graphs to represent individual BSs' features [24], [25]. A very active research branch within GCNs is how to aggregate the node features at the graph level, aiming at predicting the label for the aggregated graph while providing interpretable explanations for the prediction. An example of this approach can be found in [26], where a learnable aggregator for GCNs uses a specific mask for each neighbour of a given node, allowing the model to learn different importance weights to each node and feature. Other examples within the same line of work can be found in [27], where a graph pooling (gPool) mechanism is proposed, or in [28], where a graph pooling method considers the topology of the graph rather than barely independent node features. Another strategy, aimed at capturing the interdependent influences among nodes in non-telecommunications scenarios, utilises SGCNs, considering the spatial associations among them [29], [30], [31]. An example application can be found in [29], where the authors confirmed its effectiveness in generic classification tasks. Another work proposes a positional encoder GNN-based learning framework for regression problems using geographic data [30]. In [31], a geometric graph aggregation scheme is proposed for classification tasks in graphs. As for the case of wireless networks' applications, some previous works attempted to perform cellular traffic prediction with spatio-temporal graph networks [32], [33], [34]. However, contrary to the model proposed in this paper, all these previous examples lack an explainability module that provides detailed insights about the model outputs.

III. BS PERFORMANCE CLASSIFICATION MODEL

This section begins by outlining the key definitions employed in this paper. Subsequently, it introduces the proposed model framework, which focuses on classifying the performance state of BSs. This classification is based on the PM counters and utilises a spatial graph representation of these BSs. Finally, an explainability model is presented, which quantifies the contributions of each PM counter and each base station (BS) within the graph to the estimated classification.

A. DEFINITIONS

Some definitions are worth clarifying beforehand due to their prevalence in the paper:

- **BS** - a base station which, in this paper, denotes a unitary cell in a specific Radio Access Technology (RAT) and deployed on a given network location that may host multiple RATs (*e.g.*, 4G and 5G). Moreover, active and passive equipment is required for the BSs' operation, notably the Baseband Units (BBUs), the Remote Radio Units (RRUs), and the antennas.
- **PM counters** - the performance indicators generated by the Operational Support System (OSS), which are used to quantify the performance of each BS.
- **Target BS** - BS where the model classification is applied to estimate the BS's performance state, considering its PM counters and those from neighbouring BSs.
- **Source BS** - BS located in the vicinity of the target BS whose PM counters are also considered by the model to estimate the performance state of the target BS.
- **Graph** - a set of nodes linked by edges. In the proposed model, the nodes are the target and the source BSs.
- **Adjacency Matrix** - a matrix representing the edges defined between nodes in a graph.
- **BS Spatial Features** - a set of characteristic features that describe the spatial positioning of a BS (e.g., latitude, longitude, azimuth, height).
- **Graph Convolution** - a process where, in the proposed model, the PM counters of each BS are mixed with the ones from other BS considering the defined adjacency matrix and the spatial features. As a result, at each BS, a new representation of its PM counters is generated.

A summary of the notations used throughout this paper is presented in Table 1.

TABLE 1. Main notations used in this paper.

Notations	Descriptions
N	The number of BSs
C	The number of PM counters
T	The time-series length
$\mathbf{X} \in \mathbb{R}^{N \times C \times T}$	The PM counters of a set of N BSs
$\mathbf{x}^{(n)} \in \mathbb{R}^{C \times T}$	The PM counters of the n^{th} BS
$\mathbf{x}^{(n)}(t) \in \mathbb{R}^{C}$	$\mathbf{x}^{(n)}$ at time-index t
$\mathbf{Y} \in \mathbb{R}^{N \times T}$	The performance states of the N BSs
$\mathbf{y}^{(n)} \in \mathbb{R}^{T}$	The performance states of the n^{th} BS
$\mathbf{y}^{(n)}(t) \in [0, 1]$	$\mathbf{y}^{(n)}$ at time-index t (0 non-failure |1 failure)
$\hat{\mathbf{y}}^{(n)}(t)$	The predicted $\mathbf{y}^{(n)}$ at time-index t
$\mathbf{G} = (\mathbf{V}, \mathbf{A})$	The graph of the target BS
$\mathbf{V}$	The nodes (BSs) of $\mathbf{G}$
D	The number of nodes (BSs) in $\mathbf{V}$
$\mathbf{A}$	Adjacency matrix containing the edges of $\mathbf{G}$
a_{ij}	The edge between nodes i and j
$\Omega_i = \{j : a_{ij} = 1\}$	The set of indices of the neighbours of node i
S	The number of the spatial hidden features
B	The number of the BS hidden features
K	The number of graph convolutions
k	The k^{th} graph convolution
$\mathbf{h}_k^{(i)}(t) \in \mathbb{R}^{C \times S}$	The k^{th} hidden features of node i at time-index t
$\mathbf{p}_i \in \mathbb{R}^4$	The spatial features of node i
$u_k^{(ij)}$	The k^{th} learned weights for edge index a_{ij}

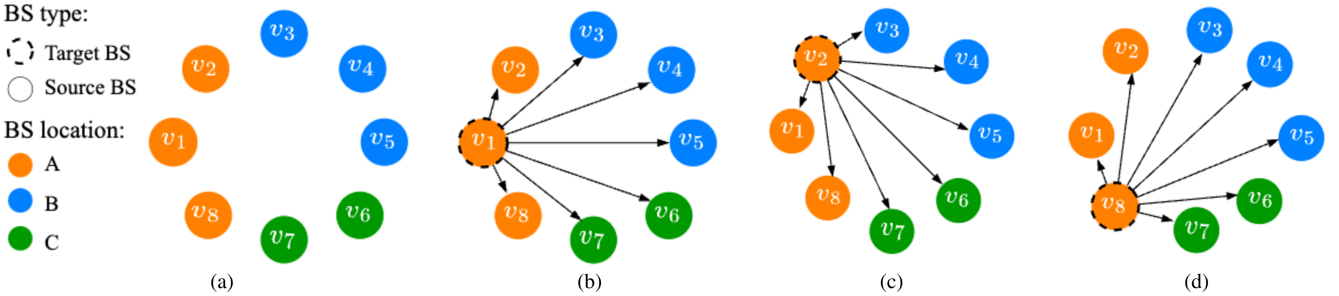


FIGURE 4. Example showing the used methodology in the definition of individual graphs for each considered target BS, where directed edges are created between the target and the source BSs: a) An example of a network with 8 BSs; b) The resulting graph for the target BS v_1; c) The resulting graph for the target BS v_2; d) The resulting graph for the target BS v_8.

The model considers a set of N BSs and the corresponding PM counters, $\mathbf{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}\}$, where $\mathbf{x}^{(n)} \in \mathbb{R}^{C\times T}$ are the counters associated to the n^{th} BS, with C indicating the total number of counters, and T the respective time-series length. Moreover, the BSs are characterized by their performance states, $\mathbf{Y} = \{\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(N)}\}$, where $\mathbf{y}^{(n)} \in \mathbb{R}^T$ and $\mathbf{y}^{(n)}(t) \in [0, 1]$ is the t^{th} element of $\mathbf{y}^{(n)}$ and denotes the classification of the n^{th} BS performance state, at each time index t, as failure ($\mathbf{y}^{(n)}(t) = 0$) or non-failure ($\mathbf{y}^{(n)}(t) = 1$). When considering an arbitrary BS as the target BS, the associated graph comprising both the target and the source BSs is denoted by $\mathbf{G} = \{\mathbf{V}, \mathbf{A}\}$. Here, $\mathbf{V} = \{v_1, v_2, \dots, v_D\}$ represents the graph nodes. Specifically, v_1 represents the target BS, while the remaining nodes correspond to the source BSs. Furthermore, for the D BSs comprising the graph $\mathbf{V}$, an adjacency matrix is established, with edges denoted as:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1D} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{iD} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{D1} & \cdots & a_{Dj} & \cdots & a_{DD} \end{pmatrix}$$

In this matrix, $a_{ij} = 1$ if there is a directed edge from v_i to v_j, and $a_{ij} = 0$ otherwise.

B. SGCN PROPOSED MODEL ARCHITECTURE

The primary objective of the proposed SGCN model is to classify the performance state of a target BS using a supervised approach, where PM indicators serve as characteristic features to analyse performance degradation. This classification considers source BSs' feature contributions, also accounting for their relative spatial positions. This involves the generation of an individual graph for each target BS, at each timestamp. Furthermore, the model uses an explainability module to identify root-cause factors explaining BS performance degradation.

1) GRAPHS DEFINITION

The dataset provided consists of target base BSs that are geographically distributed, as illustrated in Fig. 4-a) for a group of eight BSs located in three distinct physical locations. Subsequently, for each arbitrary target BS, a corresponding graph is constructed based on predefined criteria, namely considering the number of nearby locations to be included. This recurrent process is depicted in Figs. 4 b), c), and d) for a group of eight BSs, where a criterion of three locations was adopted to establish directed edges linking each target BS with the selected source BSs. Although not depicted in the figure, self-loop edges have been added to all nodes (represented by diagonal entries in the adjacency matrix, $\mathbf{A}$). The inclusion of these self-loops is crucial as they ensure that during each graph convolution, an updated representation of each BS' features takes into account both its previous values and the information from neighbouring BSs. Upon defining the required set of N directed graphs to evaluate the BSs, the proposed model conducts a graph-level classification task by aggregating the entire graph to estimate the performance state, $\hat{\mathbf{y}}^{(n)}(t)$, of an arbitrary target BS, n, at time index t. This aggregation considers the target BS's features and those of its neighbours, sequentially applied to encompass the entire set of target BSs. Additionally, the explainability module enhances the model's functionality by providing interpretable insights regarding the model output. This is particularly valuable when the model predicts a performance degradation in a specific BS, indicated by a binary output of $\hat{\mathbf{y}}^{(n)}(t) = 1$ (failure). The interpretation of the results produced by the SGCNs model can contribute to explaining the root causes of BS performance degradation at two levels: feature-wise, by examining the PM counters of the source and target BS; and node-wise, by analysing the most impacting nodes to the classification. This approach enables the capture of complex inter-dependencies between BSs.

2) GRAPHS CREATION CRITERION

To assess the impact of spatially distributed BSs on the model's output, it is essential to establish selection criteria for determining which source BSs should be included in the graphs supporting the SGCN model. A trade-off arises between selecting a limited number of source BSs, resulting in graphs that only contain co-located or nearby BSs, and choosing a larger number of source BSs, which increases the

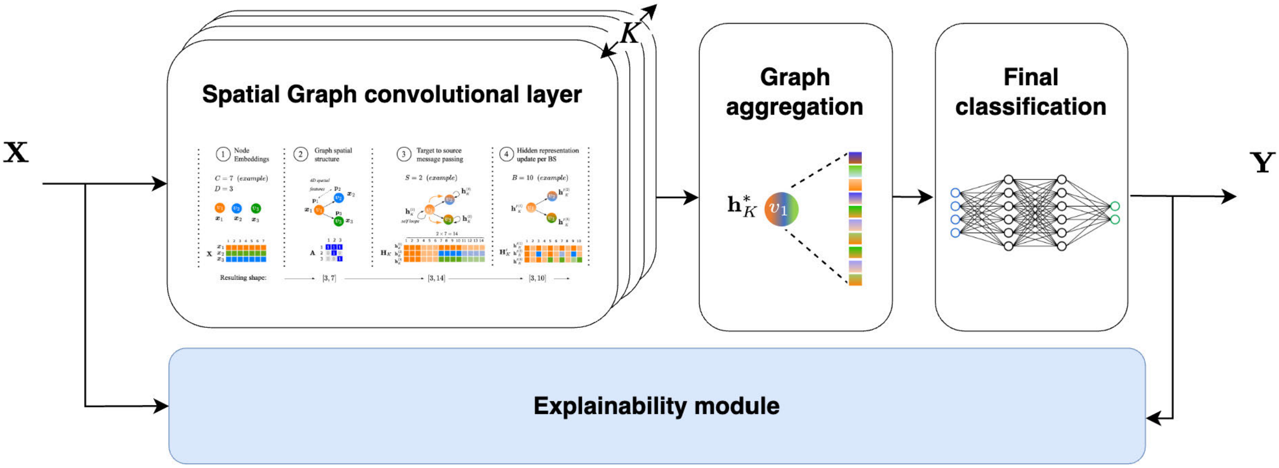


FIGURE 5. Proposed model architecture, combining an explainability module and a SGCN module using a 4D spatial representation for the BS, a graph aggregation module and a final Artificial Neural Network (ANN) to predict the label performance state.

computational complexity of the model. The approach in this paper defines a fixed number of distinct physical BS locations in the graph. In the simplified example provided in Fig. 4 considering a distance criterion of three locations, the number of BSs included in any given graph equals $D = 8$. However, it is worth emphasising that under more complex and irregular distributions found in live networks, the number of BSs varies based on the network topology surrounding each target BS.

3) SPATIAL INFORMATION

In the proposed SGCN model, each BS, v_i, within an arbitrary graph, $\mathbf{G}$, is also represented by its spatial features described by the vector $p_i \in R^4 =$ [latitude, longitude, elevation, azimuth] comprising the following components:

- **latitude**, **longitude** are the geographical coordinates;
- **elevation** is the total antenna installation height, including the terrain elevation plus the height of the installation infrastructure;
- **azimuth** is the angle between the axis of the BS's antenna and the geographical north.

The initial two features capture the Euclidean distance between the nodes, a critical variable in any wireless network modelling problem because of its direct influence on the Received Signal Strength Indicator (RSSI), as shown in [11]. The third component allows us to calculate the difference in antenna height between the source and target BSs by combining the antenna installation height with the terrain elevation. This effective height introduces a new source of learnable information that can also impact the mutual interactions between neighbouring BSs in mobile networks, such as interference. Additionally, the azimuth of the antenna influences the coverage zone of a particular BS. Collectively, these spatial features enable the model to discern the coverage overlap between two adjacent BSs, a crucial factor in predicting their expected impact on performance states.

4) PROPOSED ARCHITECTURE

The architecture of the SGCN model proposed in this paper is presented in Fig. 5. The initial spatial graph convolutional layer is responsible for updating the aggregated graph information after each of the K convolutions; subsequently, the final classification is carried out through the output layer of an Artificial Neural Network (ANN). Additionally, there is an explainability module whose objective is to provide insights into the model output. Fig. 6 describes in four steps the various transformations applied to the characteristic features of every node (BS), during each convolution iteration. In the initial stage, node embeddings are generated, pertaining to the PM indicators that characterise the BS's performance at a specific time index; these indicators are represented in the example by a set of $C = 7$ PM features and $D = 3$ nodes (BSs). The second step illustrates the graph structure, as specified in the adjacency matrix, $\mathbf{A}$. For each node, a vector of spatial variables is defined comprising a 4D spatial encoding. In the third step, a new k^{th} hidden representation, denoted as $\mathbf{h}_k^{(i)}$, is learned for each node i, based on the PM indicators of the BS, as well as the relative values of the spatial features between neighbour BSs. Additionally, a SGCN layer is employed to adjust the dimensionality of the vector $\mathbf{h}_k^{(i)}$ by a factor S, referred to as "the number of the spatial hidden features", which is one of the hyper-parameters of the proposed model. The S factor allows the model to increase the number of parameters used to weigh the influence of the spatial features on the PM indicators. Hence, a new feature representation of the PM indicators for the target BS is transmitted to its neighbours with a dimensionality $S \times C$, which in the example corresponds to $2 \times 7 = 14$. In the fourth step, a new representation of the initial features for each BS is generated, incorporating contributions from both the target and source BSs. During this stage, each BS in the graph updates its initial features to obtain a new representation, denoted as $\mathbf{h}'^{(i)}_k$. This update is achieved using

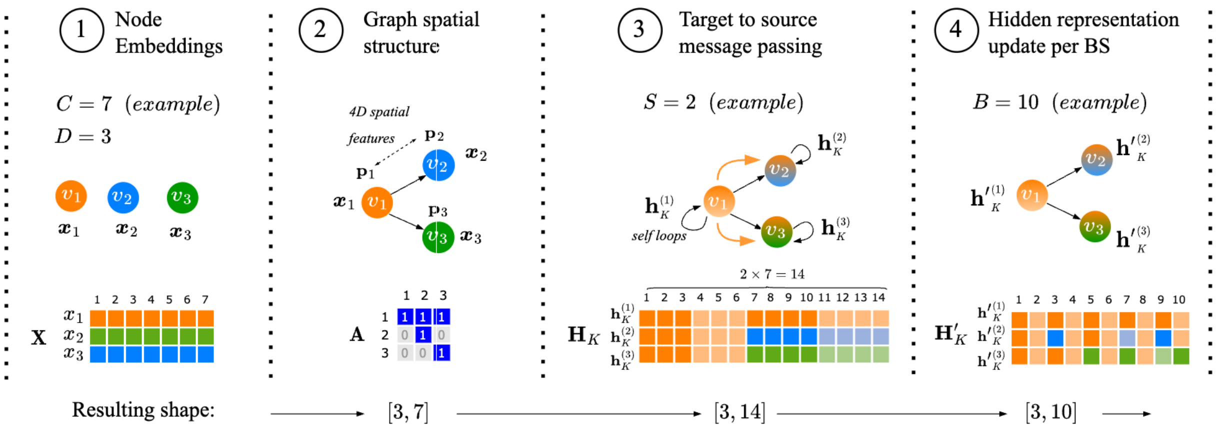


FIGURE 6. Illustration of the graph convolution process implemented in the proposed SGCN model.

a linear transformation to learn a new representation of the BS's features with a dimensionality of B. Here, B is defined as "the number of the BS hidden features" and serves as another hyperparameter of the model. The output of the linear transformation is subsequently passed through a nonlinear activation function, which will ensure that the learned weights are non-negative. Although the proposed model can use any activation function, the ReLU is a common choice because it has the merit of being easily interpretable [9]. The sequence of steps 1 to 4 is iterated K times, here K denotes "the number of graph convolutions", which is yet another hyperparameter of the proposed model.

Finally, after each convolution operation, a graph-level aggregation step is executed, resulting in the graph being transformed into a single virtual node. This virtual node's features are computed by aggregating weighted contributions from all the nodes in the original graph. The primary goal is to learn a function z that, through this graph-wide aggregation process, generates a new representation for the target BS's features, denoted as $\mathbf{h}^*_K$, which serves as the final output. Subsequently, this graph-level aggregation output is fed into an ANN responsible for predicting the performance state of the target BS. In this context,"1" signifies a failure, while "0" indicates a non-failure. It's worth noting that, within the framework of the proposed model, a "failure" is defined as a scenario where the average downlink throughput falls below 7 Mbps, a reference threshold commonly used by MNOs to monitor 4G and 5G performance.

The foundation of the proposed SGCN implementation draws from [29], incorporating significant adaptations tailored for self-healing applications in mobile networks. The architectural modifications unfolded across several dimensions. First, an explainability module was introduced, facilitating performance root-cause analysis. Second, the new 4D positional approach was implemented to describe the spatial aspects of BS spatial positioning. Third, a flexible graph creation mechanism was implemented to simulate graphs of variable size (*e.g.* single network location or multiple network locations). Fourth, a multi-aggregation layer was implemented to produce the final aggregation at the graph level. Finally, an ANN layer was implemented to predict the BS performance state using a final softmax layer that assigns probabilities to each binary class, thus supporting the model's classification prediction $\hat{\mathbf{y}}$. The softmax function was selected in this implementation as it offers the flexibility to extend the problem into a multi-class classification task using the negative-loss likelihood loss function. Moreover, it is a continuously differentiable function, so it allows calculating the derivative of the loss function concerning every weight in the SGCN architecture.

C. EXPLAINABILITY MODULE

Considering the objective of providing interpretable results, an explainability module was added to the proposed model, which calculates the contribution of each PM counter and of each node, to the model prediction, as shown in Fig. 7.

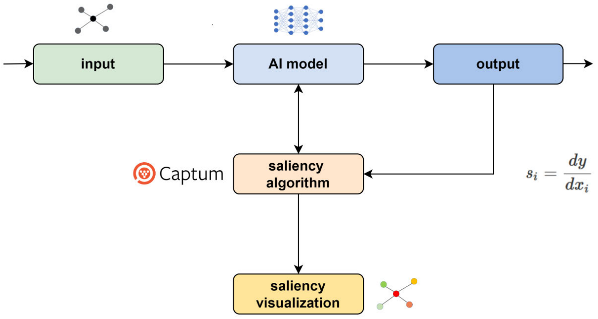


FIGURE 7. Explainability module that aimed to evaluate the input features' importance to the model's output.

The explainability module's implementation relies on the Captum platform [35], an open-source, extensible library for model interpretability built on PyTorch. Captum contains general-purpose implementations of integrated gradients, saliency maps, smoothgrad, vargrad and others. The integrated gradients' output is an attribution score for each input element and a convergence delta. The saliency maps are based on computing the gradient of the class score with respect to the input. The smoothgrad can sharpen gradient-based sensitivity maps, and vargrad is an unbiased gradient estimator for variational inference. In this paper, the choice was the saliency maps algorithm [7], given its straightforward process of correlating the gradient of the output regarding the input, where a first-order Taylor network expansion is used at the input, and the gradients are the coefficients of each feature in the model's linear representation. The absolute value of these coefficients indicates the relevance of each graph node and the input features, producing insightful visualizations of the model's output.

IV. MODEL EVALUATION

The evaluation of the proposed SGCN model was done with a dataset of real BS, which included physical installation details (such as geographical coordinates, antenna heights, and azimuths) and historical PM data. The model performance assessment focused on the "User Downlink Average Throughput" KPI using as benchmark the results of [8]. The primary goal was to predict the performance state of the target BS at each timestamp, employing binary classification with a 7 Mbps failure threshold. The model classification relied on the observed PM counters per timestamp, excluding those already considered in the KPI formula.

A. DATASETS

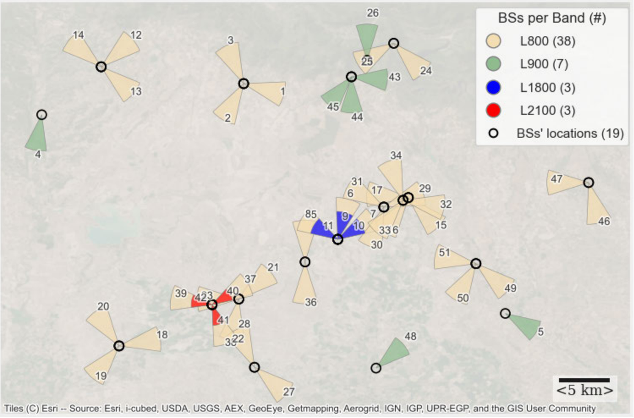
The BS dataset encompasses 19 different network locations, hosting a total of 51 BSs operating across various 4G frequency bands, as described in Table 2. These locations can be categorised into two groups: single-band installations and dual-band installations, the latter serving to enhance capacity.

TABLE 2. BSs dataset and radio access frequency bands.

Number of BSs *(by location type)*	Frequency Bands [MHz] 800	900	1800	2100	**Total BSs**	**Distinct Locations**
Single Band	32				32	12
		7			7	5
Dual Band	3		3		6	1
	3			3	6	1
Total	38	7	3	3	**51**	**19**

Fig. 8 shows the spatial distribution of the 19 mobile network locations and 51 BSs over an area of approximately 2 500 km^2, denoting irregular proximity patterns.

The PM dataset comprises 232 distinct hourly counters collected over 28 days for each of the 51 BSs, resulting in 51 BSs × 672 time stamps = 34 272 classification examples. For training and testing purposes, the same dataset split used by [8] was considered to ensure comparable results, using 75% of the dataset for training (25 704 examples) and reserving 25% of the dataset (8 568 examples) for the final test. Both the train and test sets contain samples for all individual BSs in the whole dataset, which means that the train and test datasets contain different time stamp samples for the same BSs. In Table 3, a more comprehensive description is provided, notably showing the train and test dataset splits.



Tiles (C) Esri -- Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS User Community

FIGURE 8. Spatial distribution of the 51 considered BSs dataset spread over an area of approximately 2 500 km^2.

TABLE 3. PM Dataset downlink throughput distribution.

Dataset	Mean [Mbps]	Median [Mbps]	*P*75% [Mbps]	Failures $\times 10^3$ (#)	Total $\times 10^3$ (#)	Failure Ratio %
Total	**13.0**	**10.5**	**17.0**	**7.4**	**34.3**	**21.8%**
Train	13.0	10.5	17.0	5.5	25.7	21.6%
Test	12.9	10.4	17.0	1.9	8.6	22.5%

As observed, the dataset exhibits an even failure ratio between the training and testing partitions. Furthermore, both partitions demonstrate similar values in the statistical distribution of the downlink throughput KPI, which serves as the reference metric for assessing performance degradation.

B. FEATURE SELECTION

Feature selection is a popular topic in DL to remove redundant features, noisy and irrelevant data, to improve learning feature accuracy and to reduce the training time [36]. Considering the high number of distinct PM counters in the dataset, the contribution of each of the 232 PM counters for the model classification was evaluated by applying the saliency algorithm to a pre-trained model and using the graphs with one BS location. An indicative threshold of 95% was considered for the cumulative feature importance, which resulted in the selection of 25 features required for the 4D-SGCN model to reach this level. In Table 4, the obtained reduced set of PM counters is described, grouping into four categories: volume, users, channel quality and availability.

Another analysis was performed to assess the contribution of the neighbour BSs to the model classification, when

TABLE 4. List of a reduced set of 25 PM counters included in the graph representation for a single network location.

#	Simplified designation	Category	Vendor designation	Brief description
1	Traf_CarrierAggr	Volume	L.Thrp.bits.DL.CAUser	The amount of data transmitted using carrier aggregation
2	Traf_Stream (QCI6)	Volume	L.Thrp.bits.UL.QCI.6	The amount of TCP Buffered streaming traffic (*ex:* www)
3	N_Packets_Schedul	Volume	L.Traffic.DL.PktUuLoss.Tot	The number of packets scheduled for transmission
4	N_Act_Users	Users	L.Traffic.ActiveUser.DL.Avg	The number of active users in a BS
5	N_Act_Users_QCI	Users	L.Traffic.ActiveUser.DL.QCI.Total.Max	The number of active users in a BS with measured QCI
6	Time_CarrierAggr	Time	L.Thrp.Time.DL.CAUser	The duration of Carrier Aggregation time utilization
7	Time_Session_Setup	Time	L.Traffic.DL.EmptyBuf.PDCPLat.Time	The time taken for a data session setup
8	Av_ChQual_Index	Channel	AVG_CQI	The average channel quality indicator (CQI)
9	N_MIMO1_Streams	Channel	L.ChMeas.MIMO.PRB.CL.Rank1	The number of DL blocks using MIMO with 1 spatial stream
10	N_MIMO2_Streams	Channel	L.ChMeas.MIMO.PRB.CL.Rank2	The number of DL blocks using MIMO with 2 spatial streams
11	N_16QAM (MCS10)	Channel	L.ChMeas.PDSCH.MCS.10	The number of times 16QAM (MCS10) indexes are scheduled
12	N_16QAM (MCS11)	Channel	L.ChMeas.PDSCH.MCS.11	The number of times 16QAM (MCS11) indexes are scheduled
13	N_64QAM (MCS17)	Channel	L.ChMeas.PDSCH.MCS.17	The number of times 66QAM (MCS17) indexes are scheduled
14	N_QPSK (MCS2)	Channel	L.ChMeas.PDSCH.MCS.2	The number of times QPSK (MCS2) indexes are scheduled
15	N_QPSK (MCS27)	Channel	L.ChMeas.PDSCH.MCS.27	The number of times QPSK (MCS27) indexes are scheduled
16	N_64QAM (MCS29)	Channel	L.ChMeas.PDSCH.MCS.29	The number of times 64QAM (MCS29) indexes are scheduled
17	N_Ret_16QAM (MCS30)	Channel	L.ChMeas.PDSCH.MCS.30	The number of retransmitted packets 16QAM and MCS30
18	N_QPSK (MCS5)	Channel	L.ChMeas.PDSCH.MCS.5	The number of times QPSK (MCS5) indexes are scheduled
19	N_QPSK (MCS7)	Channel	L.ChMeas.PDSCH.MCS.7	The number of times QPSK (MCS7) indexes are scheduled
20	N_Rank1_Meas	Channel	L.ChMeas.RI.Rank1	The number of BS measures reported in rank 1
21	N_Rank2_Meas	Channel	L.ChMeas.RI.Rank2	The number of BS measures reported in rank 2
22	N_DRX_Activation	Channel	L.Signal.Num.DRX.Reconfig	The number of tines that data reception was disabled
23	N_ARQ_PackLoss	Channel	L.Traffic.DL.AM.TxDropPackets	The number of packets lost due to automatic error correction
24	Time_BS_Available	Availability	L.Cell.Avail.Dur	The duration of BS up time availability
25	Time_BS_Faults	Availability	L.Cell.Unavail.Dur.Sys	The duration of BS unavailability due to faults

the reduced feature set of PM indicators is applied with different graph sizes. The hypothesis is that the spatial data representation, enriched with the information conveyed by the adjacent BSs, will reduce the required number of PM counters to achieve the objective of 95% of the model explainability. The assessment of the graph size impact on the model performance relied on the empirical definition of three scenarios comprising one, three and five network locations. The first scenario denotes a graph encompassing only the adjacent BSs within a physical network location (graph 1). The second scenario considers all the BSs in the first group plus those in the nearest two network locations (graph 3). Finally, the third scenario considers the first group plus the BSs in the nearest four network locations (graph 5). The underlying empirical assumptions for the definition of the scenarios relied on the limited geographical span of the dataset, where the distance between adjacent BS's locations ranges from 0 km to 25 km across the 19 BS locations. Hence, it would not make sense to consider larger sized graphs.

The attained results provided in Fig. 9 confirm the hypothesis, *i.e.*, the objective of 95% of the model explainability can be met with a lesser number of PM counters by using larger sized graphs.

When using graphs of a single BS location, 25 out of the 232 available PM counters are required to meet the defined criteria of 95% explainability. Conversely, with graphs of three and five BS locations, the required number of PM counters is reduced to 20 and 15, respectively. This result reinforces the hypothesis that the 4D spatial graphs feature representation can capture complex relations between the graph nodes (*i.e.*, BSs), thus reducing the need for larger and more complex datasets.

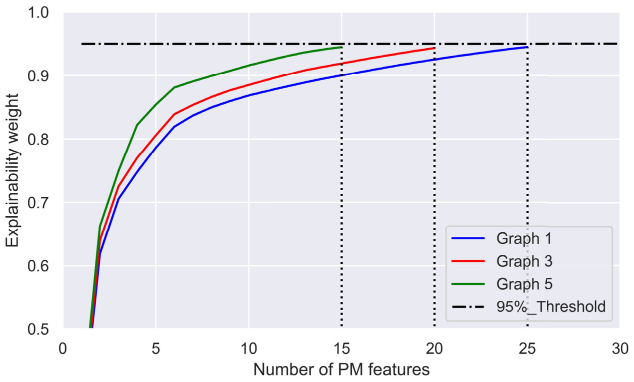


FIGURE 9. Minimum number of required PM counters to reach 95% of model explainability, by graph size.

Considering the importance of transparency and interpretability in AI algorithms, a theoretical comparative analysis was conducted on the selected PM counters across various graph sizes. Fig. 10 provides a summary of the most influential PM counters, considering graphs with one, three, and five BS locations. The comparative analysis reveals that the nine primary PM counters, which have the greatest impact on model classifications, are consistent across all graph configurations. Therefore, an in-depth examination of these counters is provided:

1) **N_Act_Users** - the number of active users connected to the BS, which allows to predict throughput based on resource demand.
2) **N_Packets_Schedul** - a resource allocation control indicator that reflects the number of scheduled packets.

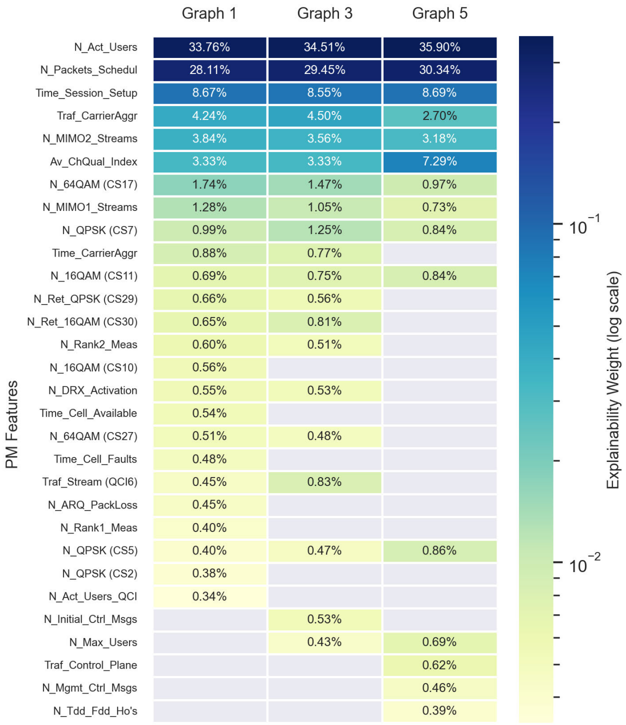


FIGURE 10. Comparative analysis using three graph sizes of the PM counters that account for 95% of the total explainability.

3) **Time_Session_Setup** - a metric for session setup time delay.
4) **Traf_CarrierAggr** - a metric for traffic volume utilising carrier aggregation between frequency bands.
5) **N_MIMO2_Streams** - a channel quality measure, associated with Multiple Input Multiple Output (MIMO) spatial streams and the radio propagation environment.
6) **Av_ChQual_Index** - the average Channel Quality Index (CQI), which signifies the quality of the wireless channel measured by the user equipment and reported to the BS.
7) **N_64QAM (CS17)** - the top modulation usage, such as 64 Quadrature Amplitude Modulation (QAM), denoting excellent channel quality.
8) **N_MIMO1_Streams** - the measure of MIMO single spatial streams, impacting maximum throughput.
9) **N_QPSK (CS7)** - the usage of Quadrature Phase Shift Keying (QPSK) modulation, which is linked to lower channel quality.

This analysis validates the choice of the reduced set of PM counters selected by the explainability module. As shown, the reduced feature set comprises a theoretically coherent summary of the most relevant PM indicators and denoting root-cause effects with the KPI addressed in this paper. Hence, the 25 features described in Table 4 were used for the final model development and evaluation.

C. PERFORMANCE METRICS

The model's classification output yields a 2×2 confusion matrix that captures the interplay between actual and predicted values for each instance in the dataset. From the confusion matrix, four basic measures can be derived - True Positives (TPs), True Negatives (TNs), False Positives (FPs), and False Negatives (FNs) - and from these, four performance metrics, typically used in binary classification problems, can be computed:

- **Accuracy** - the number of correctly classified examples over the total number of examples:

$$Accuracy = (TN + TP)/(TN + FP + TP + FN) \quad (6)$$

- **Precision** - represents the percentage of failures which are correctly classified:

$$Precision = (TP)/(TP + FP) \quad (7)$$

- **Recall** - denotes the percentage of failures which have been classified as such:

$$Recall = (TP)/(TP + FN) \quad (8)$$

- **F1-Score** - represents the harmonic mean between the Precision and Recall, such that:

$$F1 = 2 \times (\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \quad (9)$$

D. IMPLEMENTATION DETAILS

The model was trained on a NVIDIA Quadro P4000 GPU. Training and validation relied on the k-fold cross-validation methodology, with parameter $k = 5$ and using 75% of the total dataset. At each iteration, this technique considers $k - 1$ folds for model training and the remaining fold for model validation. The F1-Score served as the primary indicator for evaluating the model, with accuracy, precision, and recall serving as complementary metrics. The model estimates weight parameters, **U**, based on output probabilities using Maximum Likelihood Estimation (MLE). This involves minimising the Negative Log-Likelihood (nLL) loss function [37] through an optimization process, where the the Adam optimizer was employed [38]. Adam is a common choice for handling large datasets and parameters in similar works, as it converges faster than other alternatives using a different learning rate for each iteration. The choice for the activation function used in the model's implementation was the ReLU [39], considering its well-known benefits for DL networks, notably not suffering from vanishing gradients, as described in [40].

The hyper-parameter optimization relied on the Optuna framework [41], considering its modularity, easy adaptation to complex code and support of test parallelization and pruning of unsuccessful tests. In total, 100 independent trials were performed across the k-fold cross-validation process to compare the attained average results. To reduce computation

TABLE 5. Summary of hyper-parameters used.

Hyper-parameters	Values
Learning rate	7.63×10^{-5}
L2-decay	4.07×10^{-5}
Dropout rate	0.36
K - Number of graph convolutions	5
B - Number of the BS hidden features	113
S - Number of the spatial hidden features	12
Batch size	200
Number of epochs	2000

time, the hyper-parameters optimization used the reduced feature set of 25 PM and a graph size of one network location. The final model's hyper-parameters resulting from the optimization process are shown in Table 5.

The optimal hyper-parameter setting was used to compare model results across graph sizes. The setting which produced the best average F1-Score in the test set across all trials, was selected to train the model on the complete train set, and its performance was finally evaluated using the test set, comprising 25% of the whole dataset.

E. RESULTS COMPARISON WITH PREVIOUS RESEARCH

The results achieved with the application of the proposed 4D-SGCN-Gx model in the classification of low throughput situations were compared with the results obtained in [8], where various boosting algorithms were employed. These algorithms combine multiple decision trees through sequential learning, with each subsequent model aiming to correct the errors of its predecessor. These specific algorithms include:

- **AdaBoost** [42] - an adaptive boosting algorithm that pays more attention to under-fitted training instances by the previous model,focusing on the harder cases.
- **Grad Boosting** [43] - a gradient boosting model that tries to fit the new predictor to the residual errors made by the previous predictor.
- **XGBoost** [44] - an extreme gradient boosting model that minimizes a loss function based on the difference between the predicted and target outputs and adds a penalty term for model complexity.
- **Catboost** [45] - a categorical boosting model that supports categorical values, offering enhanced interpretability of the results and higher computational efficiency.
- **Light GBM** [46] - an increasingly popular light gradient boosting model because it focuses on the accuracy of results, can support GPU learning, and offers enhanced efficiency compared to other boosting frameworks, notably by reducing memory usage.

Table 6 presents the achieved results for the "Average Downlink Throughput" performance taken from [8], and obtained with the proposed 4D-SGCN-Gx model, where Gx denotes the number of considered BSs' locations in the graphs.

TABLE 6. F1-Score (%) on the test set for the downlink throughput KPI failure classification.

Models	F1-Score
AdaBoost [42]	0.849
Grad Boosting [43]	0.863
XGBoost [44]	0.867
Catboost [45]	0.852
LightGBM [46]	0.877
4D-SGCN-G1	**0.893**
4D-SGCN-G3	**0.896**
4D-SGCN-G5	**0.893**

As observed, the proposed 4D-SGCN-Gx model outperforms the previous results in all tested graph sizes. However, the best results were attained with the graph of three network locations, while the graphs with one and five BSs' locations present a similar performance. These results confirm that the combined representation of the spatial positioning of the target and source BSs and their PM counters encompasses a richer and more realistic view of the scenarios posed to live mobile networks. Hence, the graph's spatial structure seems adequate to represent the far-reach effects that create mutual dependence between the PM counters in each BS. However, a comprehensive analysis is essential to gain a deeper understanding of the performance fluctuations observed when different graph sizes are employed. This analysis aims to reveal the dataset's distinctive characteristics, with a specific focus on inter-location distances. The in-depth examination is thoroughly discussed in Section V, drawing insights from individual node contributions to the model's output, a feature made accessible through the use of the explainability module.

V. DETAILED RESULT ANALYSIS

This section presents an in-depth analysis of the proposed SGCN model's results, leveraging the explainability module outlined in Section III. Firstly, in conjunction with the evaluation of the F1-Score, the Accuracy, the Precision and the Recall of the proposed model are assessed. Secondly, a detailed evaluation of individual BS performance is conducted to assess the consistency of prediction results across the considered area, taking into account the coverage area of each BS. Following this, a spatial analysis is conducted to quantify the influence of the source BSs on the classification of the target BS, focusing on the impact of distance and the number of locations included in the graph definition. Lastly, a temporal analysis investigates the model's performance over the course of a day.

A. DETAILED MODEL PERFORMANCE METRICS

To complement the proposed model performance evaluation using the F1-Score, a more thorough evaluation entails the computation of additional metrics, specifically: Accuracy,

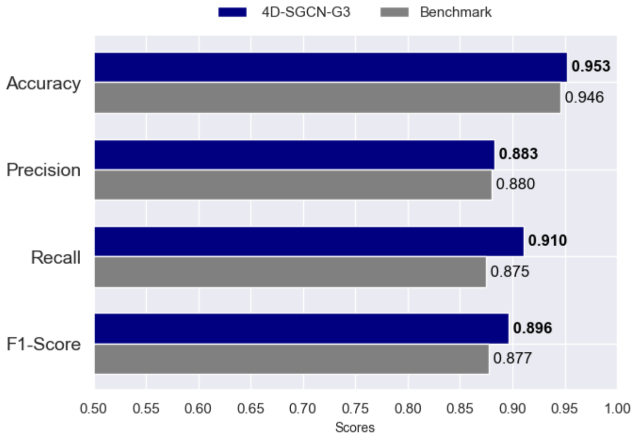


FIGURE 11. Model performance metrics for the 3 locations graphs.

Precision and Recall. These metrics were computed for the test set, using the graph size of three locations, where the proposed model yielded the best performance results. Fig. 11 presents the detailed results of the performance metrics, using the best results of [8] as benchmark.

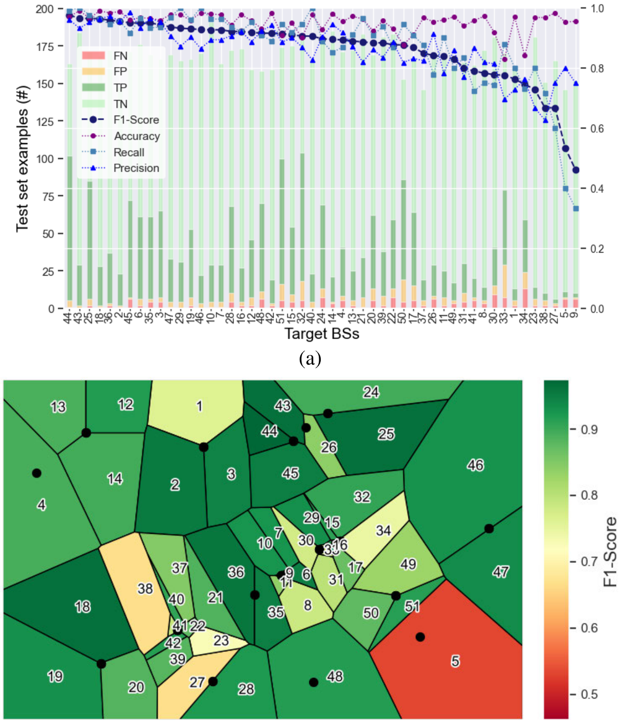
From Fig. 11, the following insights stand out:

- **Better performance in all metrics** - the classification performance gain accomplished in terms of F1-Score is approximately 2 *p.p.*, which comprises an improvement of circa 4 *p.p.* in the Recall. A deeper inspection of the confusion metrics reveals that this improvement is founded on less FNs (-27%), more TPs (+6%), a slight improvement of 0.3 *p.p.* in the Precision caused by an increase in TPs, and a small degradation of FPs (+3%).
- **Relevance for self-healing operations** - the improvement in the Recall aligns with the objective of early detection and prevention of network failures and performance degradation. Despite the slight increase in TPs, if the effort to treat such cases remains manageable, it can be considered a preventive maintenance approach, a fundamental aspect of self-healing applications.

B. MODEL PERFORMANCE PER BASE STATION

The analysis was extended by computing the model's performance metrics at the BS level, aiming at offering a more comprehensive evaluation of the model performance across different radio environments. The outcomes are presented in Fig. 12, offering two complementary views: overall metrics per BS and, specifically for the F1-Score, the spatial distribution within each BS's coverage area. The coverage areas are defined using Voronoi maps, which consist of proximal regions around the 51 BSs, considering their spatial orientation. Moreover, the x-axis of Fig. 12-a) depicts the number identifying each BS, which is also displayed in the spatial map in Fig. 12-b).

The primary insights that can be drawn from Fig. 12 are summarised in the following three points:



(a)

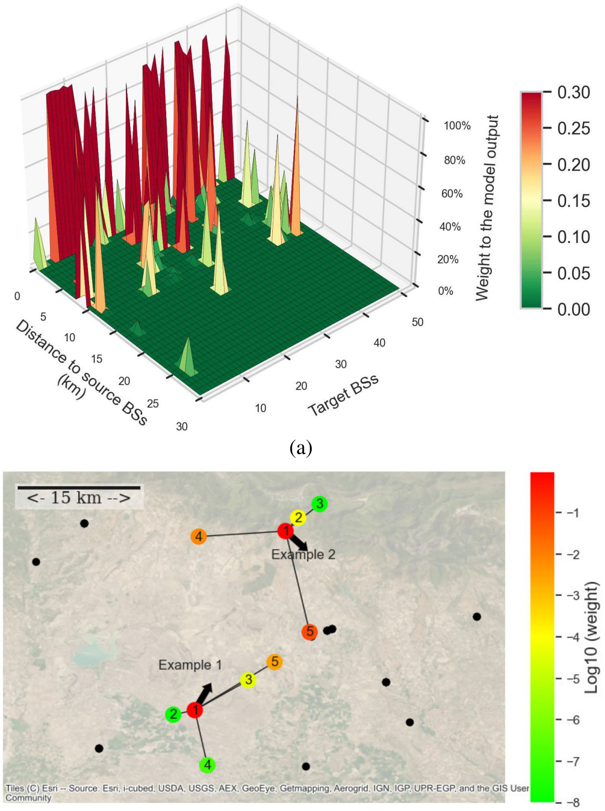
(b)

FIGURE 12. Detailed model performance for the 51 BSs: a) Overall analysis of the performance metrics per BS; b) Spatial distribution of the F1-Score on each BS's coverage area using Voronoi maps considering their spatial orientation (azimuth).

- **Performance conditioned by number of failures** - the model reveals lower F1-Scores in BSs with fewer failure examples, namely in number 5, 9, 27 and 38, indicating reduced statistical significance for a per-BS analysis.
- **Area generalisation** - the model generalises well with high F1-Scores in different radio environments, whereas some discrepancies are observed in specific peripheral BSs, notably in number 1, 5 and 27.
- **Peripheral BSs caveats** - the peripheral BSs are constrained by the lack of informative PM indicators from neighbours, which, in conjunction with the smaller number of failures, might help explain the lower F1-Scores achieved is these specific examples.

C. IMPACT OF DISTANCE AND RELATIVE ORIENTATION

The influence of distance to the target BSs was evaluated by computing the weight of the source BSs on the model results, as depicted in Fig. 13-a). The x and y axes represent, respectively, the 51 target BSs and the distances to all source BS. The z-axis indicates the weight of each source BS to the model's prediction. Fig. 13-b) presents two illustrative examples to investigate the effect of the BSs relative orientation. The node colours quantify the significance of each source BS in classifying the target BS, with the orientation of the target BS indicated by a black arrow.



(a)

(b)

FIGURE 13. Analysis of distance and azimuth: a) Impact of distance between the target BS and the source BSs; b) Impact of relative orientation (azimuth) for two illustrative examples.

In the previous analyses, the results were obtained with the 5 locations' graphs, where the primary contribution of the target BS was omitted due to its predominant effect. Consequently, the source BSs situated at a distance of 0 km represent co-located BSs with distinct spatial orientations. Furthermore, the reduced set of 25 PM indicators was employed for both analyses.

The primary insights that can be drawn from Fig. 13 are the following:

- **Impact of distance** - the results in Fig. 13-a) confirm that the source BSs with smaller distances to the target BS have a more pronounced effect on the final classification, particularly those co-located.
- **Overshooting effects** - it is evident that certain target BSs exhibit long-range impacts from source BSs, notably BS located more than 10 km away, comprising possible overshooting effects which are captured by the model.
- **Spatial orientation impact** - in Fig. 13-b) a preliminary examination was conducted using two illustrative examples. In these examples, while the target BS remains the primary contributor, the most influential source BSs are those aligned with the target's orientation, which validates the model's ability to capture dependencies taking into account both their distance and relative orientation.

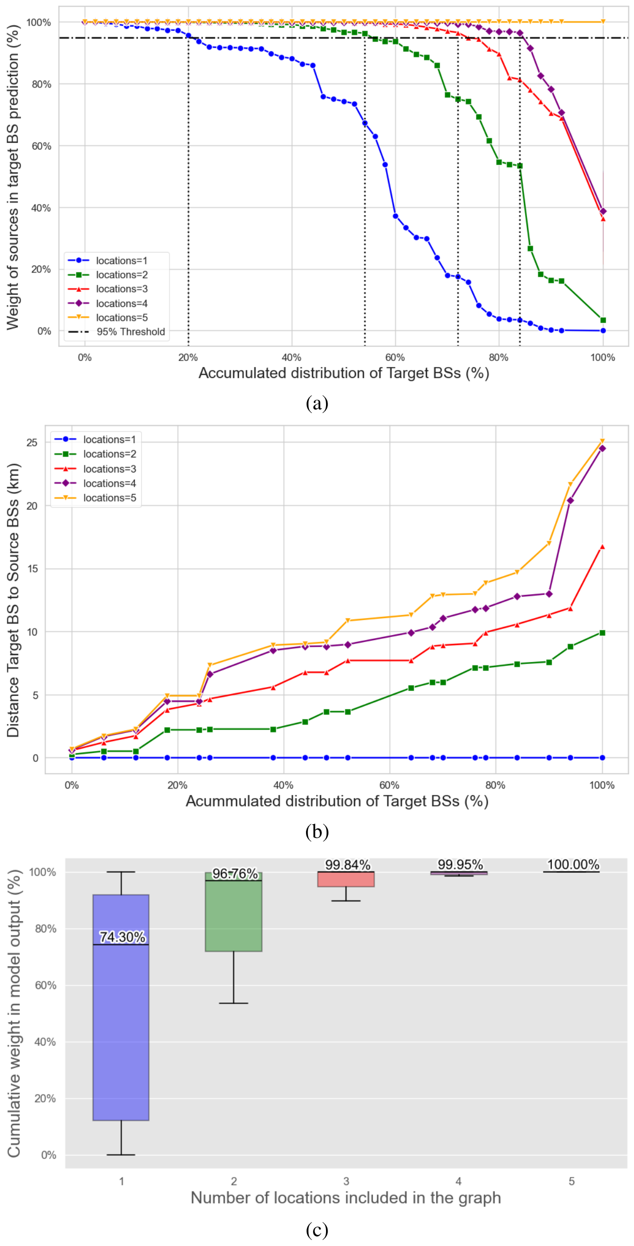
D. IMPACT OF GRAPH SIZE

Another significant question pertains to determining the appropriate number of locations to be included in the graphs. Including too few source BSs may result in insufficient information for the model, while incorporating an extensive range can lead to increased model complexity without relevant performance improvement. To address this, several analyses were conducted to assess the cumulative weight contribution of each additional location included in the graph to the model predictions, aiming to determine what is the optimal number of locations that should be considered upon creating the characteristic graph for each target BS.

The results are visualised in Fig. 14, where three complementary analyses are provided. The first introduces a criterion to evaluate the contribution of each location; the second shows the relation between graph size and distance between BSs; and the third concludes with the marginal contribution of each location to the model results. In Fig. 14-a) the x-axis represents the cumulative distribution of the 51 target BSs, while the y-axis shows the cumulative weight of their source BSs' contributions to the model predictions, with intermediate distributions displayed for each location, up to the fifth one, which represents the 100% contribution limit. A horizontal line at the 95% threshold serves as a criterion for establishing the minimum cumulative source BSs contribution, indicating when 95% of the total source BSs contribution is reached. In Fig. 14-b) the x-axis still represents the distribution of target BSs, while the y-axis illustrates the cumulative distribution of distances between each target BS and its corresponding source BSs. Finally, Fig. 14-c) presents a box plot of the cumulative weight contribution to the model output, accounting for the inclusion of each supplementary location within the graph definition. The average values of the cumulative contribution for different graph sizes are highlighted within each box plot.

The following primary insights can be extracted from Fig. 14:

- **Additional locations bring diminishing returns** - considering the 95% weight criterion, Fig. 14-a) shows that if the graphs contained just one location, approximately 20% of the target BSs meet the objective, while the remaining 80% receive contributions from the other four BS locations, of at least 5%. Extending the graph to include two locations, as indicated by the vertical dotted lines, shows that already 55% of the target BSs meet the criteria. However, beyond the inclusion of the first three locations, the source BSs within the fourth and fifth ones contribute diminishingly to the model predictions. This leads to an undesired increase in the model's complexity without a significant improvement in model performance.



(a)

(b)

(c)

FIGURE 14. Analysis of the optimal graph size: a) Number of target BSs meeting the 95% weight criteria per additional graph location; b) Impact on source to target distance when the number of considered graph locations increase; c) Cumulative weight on model results per additional graph location.

- **Dataset constrained up to 5 locations** - the results in Fig. 14-b) reveal that, as expected, starting from the first location (at a distance of 0 km from the target BS), the incorporation of additional locations in the graphs progressively extends the range of distances. These extensions result in median values of roughly 7.5 km when three locations are included and approximately 10 km when four locations are considered. The distribution associated with the fifth location closely mirrors that of the fourth, which denotes that in this dataset the fifth location does not significantly contribute new distance-related information, illustrating the principle of diminishing returns regarding distance data as additional locations are incorporated. Moreover, the maximum distance reached both in the four and five locations is 25 km, which is the limit span of the considered dataset. This confirms that it would not make sense in this case to employ graphs with size greater than five locations.
- **Optimal results with 3 locations' graphs** - the cumulative weight contribution to the model output for each supplementary location, as presented in Fig. 14-c), shows that the first location is, by large, the most important one for the model final classification contributing to roughly 74% of the weight, albeit displaying a high standard deviation. The second location extends this contribution up to around 97%, and the third location reaches more than 99% of the total contribution. This confirms that the appropriate graph size for this dataset comprises three BS locations, as the incremental increase in cumulative weight is minimal beyond this value. These findings emphasise the principle of diminishing returns discussed in the previous analyses. This outcome aligns with the results obtained for the proposed model, which yielded the best results while employing graphs with three locations.

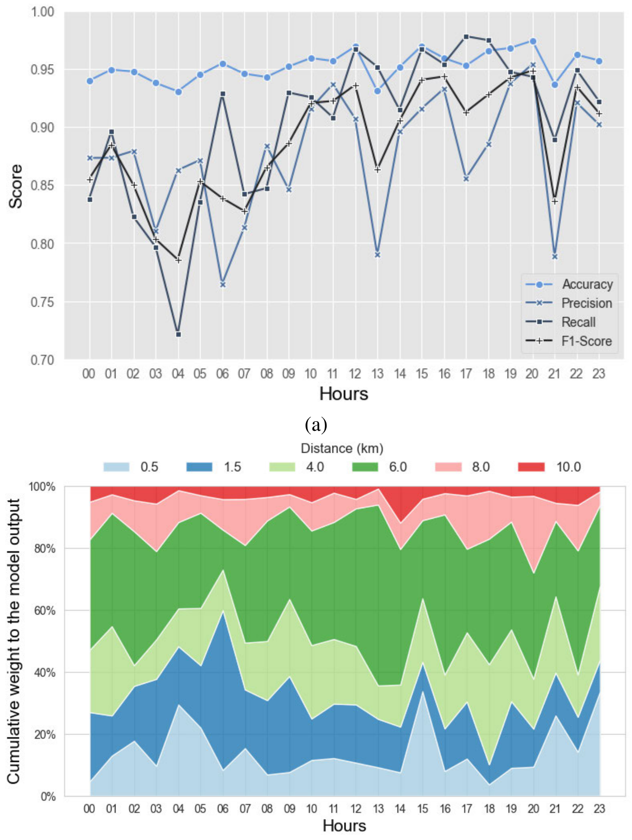
E. INTRA-DAY ANALYSIS OF MODEL PERFORMANCE

Contrary to other applications where the aggregated values are sufficient for the model performance evaluation, in self-healing operations, the near real-time classification of events can play a crucial role. Hence, an investigation was performed on the intra-day variation of the model results, encompassing two complementary analysis presented in Fig. 15. The first, as shown in Fig. 15-a), investigates the time dependence of the average model performance throughout the day. The second, as shown in Fig. 15-b), focuses on tracking the evolution of neighbourly influence over a 24-hour period in the model classification results. This involved computing the weight impact of individual source BSs on the classification of the target BS, for each hourly period.

In both figures the x-axis denotes the 24 hours of the day. As for the y-axis, Fig. 15-a) presents the evolution of the average values for the Accuracy, Precision, Recall and F1-Score computed for each hourly period, which accounts for the minimal time granularity available in this dataset. Conversely, the y-axis in Fig. 15-b) represents the weights of the source BSs contributing to the model predictions within cumulative average distance intervals between the source BS and the target BSs. Similarly to the previous analyses, only the source BSs weights are considered, given the dominant influence of the target BS.

The following primary conclusions can be extracted from Fig. 15:

- **Model performance varies with traffic load** - the results in Fig. 15-a) indicate that albeit the model Accuracy shows stability throughout the day, the



(a)

(b)

FIGURE 15. Analysis of model results throughout the day: a) Performance metrics evolution across the 24-hour period; b) Influence from neighbour BS in the model classification across the 24-hour period.

Precision, Recall, and, consequently, the F1-Score show degradation during the night period. A more detailed analysis reveals that this is caused by a higher number of FP and FN predictions during night-time. A possible explanation is the decrease in user traffic, thus reducing the variability of the PM indicators and hampering the model's performance, notably due to less failures examples. From a Network Operations Centre (NOC) perspective, having a higher performance in fault detection during the higher traffic periods is preferable, considering the end users' QoE.

- **Influence from far-away neighbours is constant** - the results in Fig. 15-b) show that the impact of source BSs positioned beyond an 8 km radius from the target BS remains fairly consistent throughout the day. This stability can be attributed to persistent interference originating from these more distant BSs, which may comprise permanent overshooting effects.
- **Mid-range neighbours have more impact during the day** - the results in Fig. 15-b) also indicate that during typical working hours, a larger portion of the weight contribution to the model predictions comes from source BSs located between 4 km and 6 km away from the target BS. This group collectively represents roughly 60% of the total contribution, particularly around 18:00 when many individuals are commuting from work to their homes.
- **Closest neighbours have more impact at night** - during the period from 00:00 to 6:00, when people are likely at home, and network activity reflects reduced user mobility, the weight assigned to source BSs within 1.5 km of the target BS increases to approximately 50%, indicating reduced reliance on information from source BSs in more distant locations.
- **Spatial features complement PM indicators** - the previous results reinforce the conclusion that the spatial graphs can complement the PM indicators with informative additional features extracted from the spatial relation between BSs, which improves the model's performance in the addressed classification objective.

VI. MODEL ADAPTABILITY TO DIFFERENT PM DATA

An additional performance evaluation of the proposed model was done using an uplink throughput KPI. The objective is to assess the model's flexibility in adapting to other use cases in self-healing operations. This evaluation involved comparing the results of the proposed model for average uplink throughput with those presented in [47], where the same set of BSs and PM dataset were employed, along with the same threshold for failure classification (0.5 Mbps). The results, as presented in Table. 7, confirm that, in the uplink throughput KPI classification, the proposed model outperforms the previous work by a significant margin. However, contrary to what was observed in the downlink analysis, the best performance is achieved with the smaller-sized graph of a single network location. Moreover, the absolute F1-Scores achieved are lower. Although the detailed analysis of the uplink throughput classification problem is out of the scope of this paper, some general comments regarding the specific nature of this problem should be underlined. Contrary to the downlink throughput classification problem, where the most relevant factor is the downlink interference that neighbour BSs exert on each other, in the uplink throughput classification problem, as explained in [48], the main root-cause effect is the inter-cell uplink interference that users in each BS suffer from co-channel interference

TABLE 7. F1-Score (%) on the test set for the uplink throughput KPI failure classification.

Models	F1-Score
AdaBoost [42]	0.588
Grad Boosting [43]	0.617
XGBoost [44]	0.716
Catboost [45]	0.687
LightGBM [46]	0.729
4D-SGCN-G1	**0.741**
4D-SGCN-G3	**0.732**
4D-SGCN-G5	**0.731**

imposed by users in other surrounding BSs. Hence, it stands to reason that the proposed SGCN model is well adapted to the analysis of the downlink throughput classification problem since the graphs have comprehensive information regarding the relative spatial distribution of the BSs. This was confirmed by the results in the downlink throughput classification using the PM indicators collected by each BS. However, for the classification problem using the uplink throughout KPI, the location information of users connected to nearby BSs is unknown to the model, which hampers its capability of capturing the uplink interference effects.

As expected, lower model performance was observed in the uplink classification problem compared to the downlink throughput. Considering the technical aspects of uplink interference, having the user spatial distribution would be essential to achieve a higher performance in the model classification. This information could be included in the graphs as if the mobile users were "individual" BSs with edge connections to other nearby users and to surrounding real BSs.

VII. CONCLUSION AND FUTURE WORK

In this work, a SGCN-based model relying on a 4D spatial representation is proposed to classify the performance of individual BSs in a mobile network, taking into account the current state of the PM indicators and using the data downlink and uplink average throughput KPI as the target classification metric. The model was evaluated using real network data, relying on fourth generation (4G) indicators due to the constrained accessibility of recent fifth generation (5G) network data. Nonetheless, it is essential to note that the spatial graphs methodology introduced in this paper is technology-agnostic. The methodology can be seamlessly applied to a wide range of cellular wireless network applications, transcending specific technology boundaries.

The application of spatial graphs to model a mobile communications network aimed at capturing the mutual influence effects that neighbour BSs exert on each other, thus improving the model prediction of each BSs performance states. The results were compared with those from previous research, where ML boosting models were applied to the same problem. This confirmed the benefits of the proposed model, achieving significant performance gains both in the downlink and uplink throughout KPI classification. The classification performance gain accomplished in terms of F1-Score is approximately 2 *p.p.* for the downlink throughput KPI, which comprises an improvement of circa 4 *p.p.* in the recall, founded on a smaller number of false negatives (-27%) along with more true positives (+6%) in the failure classification. This new balance of the F1-Score, as the harmonic mean between precision and recall, is more favourable to self-healing applications, such as predictive fault detection use cases. Considering the mission of NOC, aimed at detecting and preventing network failures and performance degradation, it is preferable to take preventive action towards a false alarm rather than not handling an alarm event. In live networks, such a reduction of circa 27% in false negatives will have a proportional impact on QoE improvement, reduction in customer complaints and field maintenance costs.

In addition to better performance, the proposed model contributes to improved model interpretability by introducing an explainability model developed using the saliency algorithm, returning node and feature level individual weights to the final classification. Leveraging the insights from the explainability model allows for a feature reduction process that specifically targets those features responsible for 95% of the total model explainability. This contribution holds significant relevance for live NOCs applications as it streamlines the real-time operation of the model by reducing data infrastructure requirements and alleviating the need for extensive processing of all available features. Additionally, the explainability module introduced here can provide root-cause explanations for network failures and KPI degradation.

In future work, two main continuation areas are identified. Firstly, the model should be able to process data in time-series format, comprising PM, FM or Configuration Management (CM) indicators, and provide time-shifted anticipatory predictions, as the basis for effective preventive maintenance and self-healing regenerative actions based on the root-cause analysis. Secondly, regarding the SGCN architecture, it is envisaged to test a single graph implementation to higher-order analysis in the node convolutions. One possibility to explore involves using satellite imagery for terrain features, similar to the approach in [11], where the graphs can be embedded within image pixels. This approach facilitates the simulation of intricate scenarios, with individual pixels representing elements such as BSs, users, sensors, or other pertinent network variables. Additionally, there is an aim to investigate alternative graph aggregation methods, integrating node and feature-level learning into the aggregation process.

ACKNOWLEDGMENT

The authors would like to thank Instituto de Telecomunicações and CELFINET, who made this work possible.

REFERENCES

[1] C. Zhang, "Intelligent Internet of Things service based on artificial intelligence technology," in *Proc. IEEE 2nd Int. Conf. Big Data, Artif. Intell. Internet Things Eng. (ICBAIE)*, 2021, pp. 731–734, doi: 10.1109/ICBAIE52039.2021.9390061.

[2] M. Z. Asghar, F. Ahmed, and J. Hämäläinen, "Artificial intelligence enabled self-healing for mobile network automation," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2021, pp. 1–6, doi: 10.1109/GCWkshps52748.2021.9681937.

[3] M. S. Riaz, H. N. Qureshi, U. Masood, A. Rizwan, A. Abu-Dayya, and A. Imran, "Deep learning-based framework for multi-fault diagnosis in self-healing cellular networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2022, pp. 746–751, doi: 10.1109/WCNC51071.2022.9771947.

[4] T. Zhang, K. Zhu, and E. Hossain, "Data-driven machine learning techniques for self-healing in cellular wireless networks: Challenges and solutions," *Intell. Comput.*, vol. 2022, pp. 104–111, Jan. 2022, doi: 10.34133/2022/9758169.

[5] G. Carrozzo, M. S. Siddiqui, A. Betzler, J. Bonnet, G. M. Perez, A. Ramos, and T. Subramanya, "AI-driven zero-touch operations, security and trust in multi-operator 5G networks: A conceptual architecture," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2020, pp. 254–258, doi: 10.1109/EuCNC48522.2020.9200928.

[6] S. A. Mohammed, A. R. Mohammed, D. Côté, and S. Shirmohammadi, "A machine-learning-based action recommender for network operation centers," *IEEE Trans. Netw. Service Manage.*, vol. 18, no. 3, pp. 2702–2713, Sep. 2021, doi: 10.1109/TNSM.2021.3095463.

[7] R. Rashid, S. Ubaid, M. Idrees, R. Rafi, and I. S. Bajwa, "Visualization of salient object with saliency maps using residual neural networks," *IEEE Access*, vol. 9, pp. 104626–104635, 2021, doi: 10.1109/ACCESS.2021.3100155.

[8] M. Cilínio, D. Duarte, P. Vieira, M. P. Queluz, and A. Rodrigues, "Root cause analysis of low throughput situations using boosting algorithms and the TreeShap analysis," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2022, pp. 1–5, doi: 10.1109/VTC2022-Spring54318.2022.9860734.

[9] S. J. Prince, *Understanding Deep Learning*. Cambridge, MA, USA: MIT Press, 2023.

[10] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021, doi: 10.1109/TNNLS.2020.2978386.

[11] M. Sousa, P. Vieira, M. P. Queluz, and A. Rodrigues, "An ubiquitous 2.6 GHz radio propagation model for wireless networks using self-supervised learning from satellite images," *IEEE Access*, vol. 10, pp. 78597–78615, 2022, doi: 10.1109/ACCESS.2022.3193486.

[12] K. J. Jang, S. Park, J. Kim, Y. Yoon, C.-S. Kim, Y.-J. Chong, and G. Hwang, "Path loss model based on machine learning using multi-dimensional Gaussian process regression," *IEEE Access*, vol. 10, pp. 115061–115073, 2022, doi: 10.1109/ACCESS.2022.3217912.

[13] N. Moraitis, L. Tsipi, and D. Vouyioukas, "Machine learning-based methods for path loss prediction in urban environment for LTE networks," in *Proc. 16th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2020, pp. 1–6, doi: 10.1109/WiMob50308.2020.9253369.

[14] Z. Xu, Y. Wang, J. Tang, J. Wang, and M. C. Gursoy, "A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2017, pp. 1–6, doi: 10.1109/ICC.2017.7997286.

[15] Z. Luo, C. Wu, Z. Li, and W. Zhou, "Scaling geo-distributed network function chains: A prediction and learning framework," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 8, pp. 1838–1850, Aug. 2019, doi: 10.1109/JSAC.2019.2927068.

[16] M.-L. Tham, A. Iqbal, and Y. C. Chang, "Deep reinforcement learning for resource allocation in 5G communications," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2019, pp. 1852–1855, doi: 10.1109/APSIPAASC47483.2019.9023112.

[17] A. Feriani and E. Hossain, "Single and multi-agent deep reinforcement learning for AI-enabled wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1226–1252, 2nd Quart., 2021, doi: 10.1109/COMST.2021.3063822.

[18] V. Tong, S. Souihi, H. A. Tran, and A. Mellouk, "Machine learning based root cause analysis for SDN network," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6, doi: 10.1109/GLOBECOM46510.2021.9685185.

[19] R. Pan, X. Li, and K. Chakrabarty, "Unsupervised root-cause analysis with transfer learning for integrated systems," in *Proc. IEEE 39th VLSI Test Symp. (VTS)*, Apr. 2021, pp. 1–6, doi: 10.1109/VTS50974.2021.9441030.

[20] A. Saleem, M. Raeiszadeh, A. Ebrahimzadeh, R. H. Glitho, J. Eker, and R. A. F. Mini, "A deep learning approach for root cause analysis in real-time IIoT edge networks," in *Proc. NOMS IEEE/IFIP Netw. Oper. Manage. Symp.*, 2023, pp. 1–5, doi: 10.1109/NOMS56928.2023.10154218.

[21] L. Liu, K. Zhang, L. Liu, L. Zhang, and J. Zhang, "Root cause analysis of network fault based on random forest," in *Proc. 7th Int. Conf. Commun., Image Signal Process. (CCISP)*, Nov. 2022, pp. 1–5, doi: 10.1109/CCISP55629.2022.9974518.

[22] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," 2018, *arXiv:1802.03888*.

[23] Y. Shen, J. Zhang, S. H. Song, and K. B. Letaief, "Graph neural networks for wireless communications: From theory to practice," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3554–3569, May 2023, doi: 10.1109/TWC.2022.3219840.

[24] M. Zhang and D. Vasal, "Mechanism design for large scale network utility maximization," in *Proc. 55th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2021, pp. 1–6, doi: 10.1109/CISS50987.2021.9400216.

[25] C. Rattaro, F. Larroca, and G. Capdehourat, "Predicting wireless RSSI using machine learning on graphs," in *Proc. IEEE URUCON*, Nov. 2021, pp. 372–376, doi: 10.1109/URUCON53396.2021.9647374.

[26] L. Zhang and H. Lu, "A feature-importance-aware and robust aggregator for GCN," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, p. 1813, doi: 10.1145/3340531.3411983.

[27] H. Gao and S. Ji, "Graph U-nets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4948–4960, Sep. 2022, doi: 10.1109/TPAMI.2021.3081010.

[28] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," 2019, *arXiv:1904.08082*.

[29] P. Spurek, T. Danel, J. Tabor, M. Smieja, L. Struski, A. Slowik, and L. Maziarka, "Spatial graph convolutional networks," in *Proc. 27th Int. Conf. ICONIP*, 2020, pp. 668–675, doi: 10.1007/978-3-030-63823-8_76.

[30] K. Klemmer, N. Safir, and D. B. Neill, "Positional encoder graph neural networks for geographic data," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2021, pp. 1–11.

[31] H. Pei, B. Wei, K. C.-C. Chang, Y. Lei, and B. Yang, "Geom-GCN: Geometric graph convolutional networks," in *Proc. 8th Int. Conf. Learn. Represent. (ICLR)*, 2020, pp. 1–12.

[32] Y. Yao, B. Gu, Z. Su, and M. Guizani, "MVSTGN: A multi-view spatial–temporal graph network for cellular traffic prediction," *IEEE Trans. Mobile Comput.*, vol. 22, no. 5, pp. 2837–2849, May 2023, doi: 10.1109/TMC.2021.3129796.

[33] Y. Chen, T. Shu, X. Zhou, X. Zheng, A. Kawai, K. Fueda, Z. Yan, W. Liang, and K. I. Wang, "Graph attention network with spatial–temporal clustering for traffic flow forecasting in intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 8727–8737, Aug. 2022, doi: 10.1109/TITS.2022.3208952.

[34] N. Zhao, A. Wu, Y. Pei, Y.-C. Liang, and D. Niyato, "Spatial–temporal aggregation graph convolution network for efficient mobile cellular traffic prediction," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 587–591, Mar. 2022, doi: 10.1109/LCOMM.2021.3138075.

[35] N. Kokhlikyan, V. Miglani, M. Martin, E. Wang, B. Alsallakh, J. Reynolds, A. Melnikov, N. Kliushkina, C. Araya, S. Yan, and O. Reblitz-Richardson, "Captum: A unified and generic model interpretability library for PyTorch," 2020, *arXiv:2009.07896*.

[36] S. Velliangiri, S. Alagumuthukrishnan, and S. I. T. Joseph, "A review of dimensionality reduction techniques for efficient computation," *Proc. Comput. Sci.*, vol. 165, pp. 104–111, 2019, doi: 10.1016/j.procs.2020.01.079.

[37] H. Yao, D.-L. Zhu, B. Jiang, and P. Yu, "Negative log likelihood ratio loss for deep neural network classification," in *Proc. Future Technol. Conf. (FTC)*, 2019, pp. 276–282, doi: 10.1007/978-3-030-32520-6_22.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–11.

[39] A. Fred Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.

[40] G. Shen and Y. Yuan, "On theoretical analysis of single hidden layer feedforward neural networks with relu activations," in *Proc. 34rd Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, Jun. 2019, pp. 706–709, doi: 10.1109/YAC.2019.8787645.

[41] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 2623–2631, doi: 10.1145/3292500.3330701.

[42] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference*. Berlin, Germany: Springer, Jan. 2013, pp. 37–52, doi: 10.1007/978-3-642-41136-6_5.

[43] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.

[44] T. Chen and C. Guestrin, "XGBoost," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[45] A. Veronika Dorogush, V. Ershov, and A. Gulin, "CatBoost: Gradient boosting with categorical features support," 2018, *arXiv:1810.11363*.

[46] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–9. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017, doi: 10.5555/3294996.3295074.

[47] M. R. Cilínio, "Network performance enhancement by implementing self-healing functions in mobile network operations," Master thesis, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal, 2022.

[48] D. X. Almeida, L. M. Correia, and M. Serrazina, "Antenna characteristics impact on LTE inter-cell interference performance in urban scenarios," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2014, pp. 1–5, doi: 10.1109/eucnc.2014.6882629.

LUÍS MATA received the B.S. degree in electrotechnical engineering and computers from Instituto Superior Técnico, Universidade de Lisboa (IST/UL), in 1996, the Executive M.B.A. degree from the AESE/IESE Business School, Lisbon, and the Specialist degree in telecommunications from the Polytechnic Institute of Lisbon (IPL), in 2022. He is currently pursuing the Ph.D. degree with IST/UL. He is also an Adjunct Invited Professor in computer and internet networks with Instituto Superior de Engenharia de Lisboa (ISEL). In 2021, he joined the Research and Development Department, CELFINET—a CYIENT Group Company, as a Lead Researcher. His Ph.D. thesis is focused on the subject of "Using AI to Enhance the Operational Efficiency of Present and Future Mobile Networks." He has over 25 years of professional experience working in mobile and fixed network engineering, marketing, product management, and business strategy. His research interests include data science, artificial intelligence, machine learning, 5G, 6G, and industry 4.0.

MARCO SOUSA received the B.Sc. degree in electronic and telecommunications engineering from Universidade da Madeira (UMa), Portugal, in 2014, the M.Sc. degree in electronic and telecommunications from Instituto Superior de Engenharia de Lisboa (ISEL), Portugal, in 2017, and the Ph.D. degree in electrical and computer engineering from Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal, in 2023. Since 2015, he has been an integral part of the Research Department, CELFINET—a CYIENT Group Company, where he has taken a leadership role in driving research activities focused on mobile network configuration management and optimization. Additionally, he holds the position of an invited Professor with the Department of Electronics, Telecommunications, and Computer Engineering, ISEL. His primary research interests include mobile network optimization, radio access networks, and the use of artificial intelligence in the realm of mobile networks.

PEDRO VIEIRA was born in Lisbon, Portugal, in 1974. He received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from Instituto Superior Técnico (IST), University of Lisbon, Portugal, in 1997, 2003, and 2008, respectively, and the Specialist degree in telecommunications from the Polytechnic Institute of Lisbon (IPL), in 2016. He is currently pursuing the Executive Master of Business Administration (M.B.A.) degree in ISCTE executive education. He is also an Adjunct Professor with ISEL, Lisbon Polytechnic Institute, where he lectures radio communications courses, including mobile communications. He is also a Researcher with Instituto de Telecomunicações, working on wireless communications, including radio and visible light propagation, wireless network planning and optimization, and SON systems. He is also a Senior Engineer registered in the Portuguese Engineering Order (OE) and has been an OE Telecommunications Specialist, since 2008. He is also leading the research activities in CELFINET—a CYIENT Group Company, where he is engaging in applied research to create new methodologies, tools, and algorithms for the mobile communications business.

MARIA PAULA QUELUZ received the B.S. and M.S. degrees in electrical and computer engineering from Instituto Superior Técnico (IST), University of Lisbon, Portugal, and the Ph.D. degree from the Catholic University of Louvain, Louvain-la-Neuve, Belgium. She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, IST, and a Senior Research Member of Instituto de Telecomunicações, Lisbon, Portugal. Her main scientific and research interests include image/video quality assessment, image/video processing, and wireless communications.

ANTÓNIO RODRIGUES (Member, IEEE) received the B.S. and M.S. degrees in electrical and computer engineering from Instituto Superior Técnico (IST), Technical University of Lisbon, Lisbon, Portugal, in 1985 and 1989, respectively, and the Ph.D. degree from the Catholic University of Louvain, Louvain-la-Neuve, Belgium, in 1997. Since 1985, he has been with the Department of Electrical and Computer Engineering, IST, where he is currently an Associate Professor. He is also a Senior Research Member of Instituto de Telecomunicações, Lisbon. His current research interests include mobile and satellite communications, wireless networks, modulation, coding, and multiple access techniques.