RESEARCH PAPER
*Future Research in AI & IoT, 2026, 2(2)*
*Homepage: https://frai.journals.umz.ac.ir/*
*DOI: 10.22080/frai.2026.30812.1042*

# Intelligent cell analysis in 5G/B5G using TOPSIS and ML

**Reza Moammer Yami[1]*, Hadi Soltanizadeh[1], Ali Shahzadi[1]**

[1] *Department of Electrical Engineering, Semnan University, Semnan, Iran*

**Article Info**
Received 19 Dec 2025
Accepted 14 June 2026
Available online 14 June 2026

**Keywords:**
MCDMA;
TOPSIS;
Cell Fault Detection;
SON;
5G &B5G.

**Abstract:**

The aim of this research is to accurately rank and increase the accuracy of abnormal cell detection in 5G and B5G networks by combining machine learning (ML) algorithms with the TOPSIS multi-criteria decision-making (MCDM) technique. This study was conducted in two main stages: the ranking stage and the failure estimation and detection stage. In the first stage, the TOPSIS technique was used to rank, score, and identify the best and worst cells. In the second stage, supervised ML algorithms were used to predict and estimate defective cells. The results show that this hybrid approach effectively addresses the challenges of data quality, scalability, complexity, and real-time processing; error interpretation is also able to classify the types of failures of each cell. Specifically, AdaBoost achieved 98.886% accuracy, 0.978 precision, and 0.989 F-Measure, while logistic regression achieved 97.7881% accuracy, 0.968 precision, and 0.979 F-Measure, both excellent failures. TOPSIS technique provides network operators with transparency and interpretability of each cell's performance by selecting and assigning weights to all the Indicators that affect the performance of each cell. This study represents a critical breakthrough in increasing the reliability, efficiency, and scalability of next-generation networks, thereby providing greater intelligence and agility in network operations.

***Corresponding Author:** r.moammer@semnan.ac.ir

**Supplementary information:** Supplementary information for this article is available at https://frai.journals.umz.ac.ir/

**Please cite this paper as:**

## 1 Introduction

The rapid evolution of wireless communication and mobile networks has underscored the critical need for efficient, scalable, and robust infrastructures to support next-generation telecommunications. Key drivers of this evolution include the growing demand for massive machine communication (MTC), ultra-reliable low-latency communication (URLLC), and enhanced mobile broadband (eMBB). To address these requirements, 5G networks and beyond must undergo significant advancements in both hardware and software, enabling faster data transfer, enhanced user experiences, and seamless integration of emerging technologies. Among these advancements, self-organizing networks (SONs), empowered by artificial intelligence (AI), have emerged as a cornerstone for achieving adaptability, autonomy and operational efficiency in modern telecommunications systems. Initially introduced in 3GPP Release 8 [1], SONs are designed to be scalable, stable, and agile, making them indispensable for meeting the dynamic demands of contemporary and future networks.

Recent research has increasingly focused on the integration of cutting-edge technologies such as AI, machine learning (ML), and data analytics into mobile communication infrastructures, particularly in the context of 5G and beyond 5G (B5G) networks. The transition to 6G is expected to bring even more groundbreaking innovations, such as the utilization of new spectrum bands up to terahertz (THz), optical wireless communications, and integrated sensing and communication (ISAC) for enhanced wireless capabilities and novel services. Furthermore, 6G envisions AI not only as a service but also as an intrinsic component of the communication system, enabling intelligent connectivity among devices and systems. Additional pillars of 6G include native trustworthiness based on multi-lateral trust models and advanced cryptographic technologies, integrated terrestrial and non-terrestrial networks for ubiquitous global coverage, and green networking solutions aimed at reducing the total cost of ownership (TCO) and promoting sustainable development worldwide [2-3]. These advancements collectively highlight the pivotal role of AI and ML in shaping the future of mobile networks, ensuring they are not only faster and more reliable but also smarter and more sustainable.

*ISSN 3092-7552*

Table 1: List of abbreviations

| | |
|---|---|
| SON | Self-Organizing Networks |
| COD | Cell outage detection |
| COC | Cell outage compensation |
| OPEX | Operating Expenditures |
| CAPEX | Capital Expenditures |
| ICIC | Inter-Cell Interference Coordination |
| UL | Uplink Traffic |
| DL | Downlink Traffic |
| RAN | Radio Access Network |
| DT | Decision Trees |
| AD | Anomaly Detector |
| FQL | Fuzzy Q-Learning |
| KNN | K-Nearest Neighbors |
| KPI | Key Performance Indicator |
| LTE | Long Term Evolution |
| MCDM | Multi-criteria decision-making |

To address the increasing complexities of network management, artificial intelligence algorithms and flexible machine learning (ML) methods, particularly those based on deep neural networks (NNs), have gained significant traction. These technologies enhance wireless communications by enabling advanced capabilities, including nonlinear mapping, distributed processing, and adaptive decision-making. ML algorithms, including unsupervised, supervised, and reinforcement learning, have been widely employed to enhance critical applications, including spectrum management, signal processing, and network optimization. The exponential growth in the application of ML in wireless communications underscores its importance as a dominant trend in academic, research, and industrial communities. In addition, multi-criteria decision-making (MCDM) methods have proven highly effective in managing and mitigating cell failures, as they enable decision-makers to evaluate multiple criteria simultaneously and select optimal solutions.

Despite these advances, traditional statistical algorithms often fall short in terms of accuracy, scalability, and flexibility, particularly when dealing with the dynamic and heterogeneous nature of modern networks. This limitation has prompted the adoption of ML-based approaches that offer superior adaptability and efficiency in network management. ML models, unsupervised, supervised, and reinforcement learning, offer a robust framework for analyzing network cells and predicting anomalies. In this research, we integrate rule-based approaches, such as multi-criteria decision systems for cell ranking, with ML techniques for predicting and estimating abnormal cells in the network. This hybrid approach seeks to bridge the gap between traditional methods and modern AI-based solutions.

Machine learning applications in wireless communications can be broadly categorized into three groups: (i) ML-based spectrum intelligence and adaptive radio resource management, (ii) ML-based transmission intelligence and adaptive baseband signal processing, and (iii) ML-based network intelligence and adaptive system-level optimization [4]. Self-organizing networks (SONs) fall into the third category with a particular emphasis on self-healing capabilities. Self-healing is a critical component of SONs, focusing on cell management through two main tasks: (1) detecting out-of-service cells and (2) compensating for their failure. Existing research on self-healing networks has explored three primary approaches: rule-based, algorithm-driven, and parametric-based approaches. While rule-based approaches rely on empirical rules and are suitable for small-scale networks, they often suffer from limited accuracy and scalability. Algorithm-based approaches, rooted in statistical theories, provide moderate accuracy but struggle with the complexity of modern network structures. In contrast, parametric-based approaches, leveraging ML, offer high flexibility, operational simplicity, and transparency, making them ideal for large and complex networks. Intelligent fault detection is a key infrastructure for next-generation networks. These faults are shown in Table 2.

Table 2: types of cell failure And causes

| Failure type | Description | Reasons for the incident |
|---|---|---|
| Complete cell failure | The cell completely fails and provides no service. | - Power Failure<br>- Hardware Failure<br>- Software Crash<br>- Physical damage to equipment (e.g., due to natural disasters or vandalism) |
| Partial cell failure | The cell does not fail completely, but some services or functions are disrupted | - Antenna Issues<br>- Signal Interference<br>- Congestion<br>-Configuration Errors |
| Temporary cell failure | Failure occurs temporarily and usually resolves automatically after a while. | - Temporary software glitches<br>- Temporary traffic overload<br>- Adverse weather conditions (such as storms or heavy rain) |
| Cell cover failure | The cell does not completely fail, but some areas covered by it experience signal loss or no coverage. | - Physical obstacles (such as tall buildings or hills)<br>- Network Planning Issues<br>- Antenna failure or reduced signal strength |
| Cell software failure | Failure occurs due to problems in the cell's management or operational software. | - Software Bugs<br>- Failed Software Updates<br>- Problems in Radio Resource Management (RRM) algorithms |

| | | |
|---|---|---|
| Cell hardware failure | Failure occurs due to problems in the cell's hardware equipment. | - Antenna Failure<br>- Radio Unit Failure<br>- Power Supply Failure |

However, several research gaps remain unaddressed. Integration of MCDM techniques with ML algorithms for cell failure management has not yet been fully explored. Furthermore, the scalability and generalizability of ML-based self-healing solutions in diverse network environments, overcoming the complex structures of next-generation networks, require further investigation. This study proposes an innovative hybrid framework that integrates the Order Preference Selection Technique based on Similarity to Ideal Solution (TOPSIS) and machine learning (ML) to enhance the efficiency of cell quality ranking and failure prediction accuracy, thereby addressing critical gaps in existing methods. Our objectives include (1) developing a hybrid model that integrates rule-based and ML-based approaches to improve accuracy and scalability, (2) evaluating the performance of this model in diverse network scenarios, and (3) providing practical insights for optimizing self-healing capabilities in next-generation networks. By addressing these challenges, this study seeks to contribute to the advancement of intelligent and autonomous network management systems.

## 2 Theoretical foundations and research history

The concept of self-healing in cellular networks embodies an intelligent and automated approach to detecting, diagnosing, and resolving network anomalies and failures. By leveraging machine learning (ML) techniques, self-healing systems facilitate precise and dynamic mapping of network parameters, thereby enhancing the robustness and efficiency of network operations. Such systems are indispensable in contemporary cellular networks, including 5G and beyond, where the increasing complexity and scale of infrastructure demand advanced automation to ensure sustained service quality and reliability. The self-healing process primarily focuses on cell management, encompassing three critical stages: (1) detecting cell outages or performance degradation, (2) diagnosing the root cause and type of failure, and (3) implementing corrective measures to restore normal functionality. This proactive approach not only minimizes downtime but also optimizes network performance, making it a cornerstone of modern autonomous network management systems.

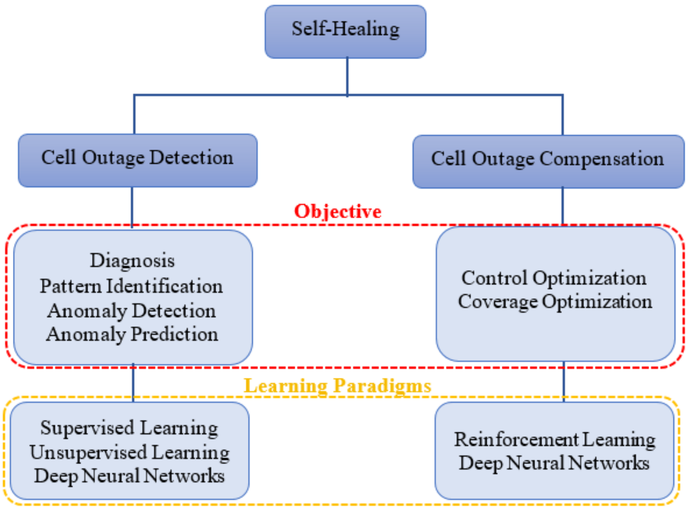

**Figure 1: Self-healing structure in next-generation wireless networks [5]**

As illustrated in Figure 1, the compensation process involves addressing challenges related to control, recovery, and functional optimization of Abnormal cells. In contrast, the detection process focuses on identifying these cells within the network. Supervised learning is predominantly used for this purpose. One commonly applied algorithm in this context is the decision tree [6], which employs an iterative process to traverse tree levels and determine sample classes by analyzing available features. While the decision tree algorithm is advantageous due to its noise resistance and low computational complexity with large datasets, its main drawback is the lengthy processing time. Notable applications of this algorithm include the work of Siva Kumar et al. [7], who investigated radio link failures, and Hangli and Zhuqiao [8], who utilized it for network selection based on available features. Additionally, research in [9] applied decision trees for classifying and predicting defective cells at the network level.

Logistic regression [10] is another widely used algorithm, particularly for binary classification problems. It employs a linear model based on gradient descent to optimize and adjust parameters, with class probabilities expressed through the sigmoid function. This method is particularly effective for imbalanced and binary datasets. Applications of logistic regression in network diagnostics include studies such as [11-12]. For instance, Kulkarni et al. [11] used logistic regression to identify failures and errors in fifth-generation network antennas by simulating various network Indicators.

The random forest method [13], which combines multiple base classifiers (decision trees) to create a robust and accurate model, is another popular algorithm in network diagnostics. Its primary goal is to address data variance issues. Research utilizing random forest for network diagnostics includes studies such as [14-16], where it served as the central algorithm. For example, [14] applied random forests to identify defective cells at the network level. Other ML methods, such as k-nearest neighbors (KNN) and support vector machines (SVM), have also been employed for anomaly detection in network components [17-19]. For instance, studies [17] and [18] combined SVM and KNN to detect anomalies in cell structures, focusing on identifying outlier-type defective cells and analyzing changes in adjacent cell indices, which often indicate defects in neighboring cells.

Further research, such as [20-22], has focused on identifying structural patterns within network cells. In the context of defective cell detection, supervised methods are typically preferred due to the nature of the data. However, reinforcement learning combined with deep learning approaches has shown superior performance in scenarios involving large, high-dimensional datasets. These methods are particularly effective in the compensation process, as they automate feature extraction and selection, reducing the likelihood of errors and computational inefficiencies. Examples of such applications include studies [23-24], which integrated reinforcement learning and deep learning for network optimization and defective cell management. In the research of Yami, R. M., and Khazaei [26], the VOTING method was used to identify nerve cells on three algorithms: Adaboost, Logistic Regression, and New Bayes, which achieved an accuracy of 98.07[26].

In summary, the approach to identifying and compensating for defective cells in network structures is heavily influenced by the nature and complexity of the available data. The methods employed in self-healing are aligned with the framework depicted in Figure 2, emphasizing the importance of data-driven solutions in addressing network challenges. Research activities in the field of self-healing based on the approaches used are shown in Figure 2.

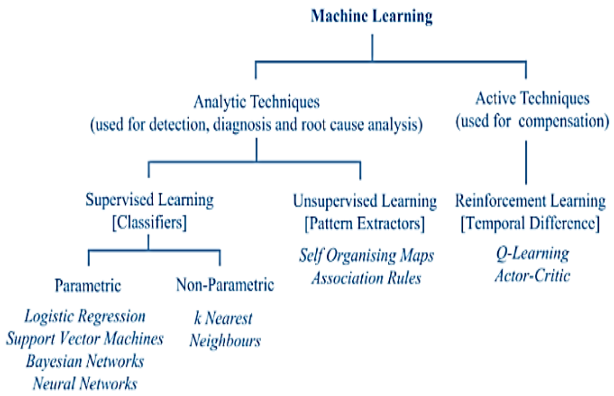

**Figure 2: Learning patterns for cell management in network self-healing [25]**

According to research, the challenges in this field are shown in Table 3.

**Table 3: challenges in Research background**

| Challenge | Description | Related examples | Needs improvement |
|---|---|---|---|
| Accuracy and efficiency of algorithms | Traditional methods do not have sufficient accuracy and efficiency when dealing with large and complex networks. | A decision tree has a long processing time. | Using hybrid approaches such as integrating machine learning and MCDM. |
| Scalability | Previous methods are designed for small networks and are not effective in large networks. | Rule-based algorithms are not suitable for large networks. | Using Deep Learning and Reinforcement Learning. |
| Flexibility | Traditional methods lack the flexibility to adapt to changing network conditions. | Logistic Regression is suitable for unbalanced data but is inflexible. | Using machine learning methods that adapt to dynamic network conditions. |
| Inefficient use of data | Network data is not being used optimally. | Limited use of signal quality data and lack of multi-level data integration. | Multi-Source Data Integration And Use of Deep Learning. |
| Computational complexity | Some methods have high computational complexity, making real-time implementation difficult. | Random Forest is time-consuming for large networks. | Optimization of algorithms to reduce complexity and increase execution speed. |
| Lack of transparency | Some methods, especially deep learning, act as "black boxes". | Deep Neural Networks have low interpretability. | Developing Explainable AI methods for greater transparency. |
| Lack of integration of methods | Previous research has focused on only one specific method. | Using only one algorithm (such as SVM or KNN) and not combining methods. | Developing hybrid methods that take advantage of the benefits of multiple algorithms. |

Considering the challenges and gaps identified in existing research, multi-criteria decision-making systems offer a promising solution to enhance self-healing mechanisms in cellular networks. These systems can effectively address important issues such as data quality, scalability, real-time processing, and interpretability, which are essential for the efficient operation of self-healing systems. Below, we review the role of the TOPSIS technique in overcoming these challenges, focusing on specific methodologies.

### *2.1* Key Challenges and Research Gaps:

1. **Data Quality and Availability**:
    - Network data is often incomplete, noisy, or heterogeneous.
    - There is a need for methods to prioritize high-quality data and filter out irrelevant or redundant information.
2. **Scalability**:
    - The advent of 5G and beyond introduces massive volumes of data and parameters.
    - Scalable methods are required to manage and prioritize key parameters effectively.
3. **Real-Time Processing**:
    - Real-time decision-making demands rapid and accurate evaluation of multiple criteria.
    - Techniques are needed to integrate diverse criteria and select optimal actions within minimal timeframes.
4. **Interpretability**:
    - Machine learning models often operate as "black boxes," lacking transparency.
    - There is a need for methods that provide clear, interpretable decisions to network operators.
5. **Unbalanced data:**
    - Due to the low number of defective cells in a network, there is an unbalanced data challenge.
6. **Inability to prioritize the type of failure**

### *2.2* Role of TOPSIS Technique:

To address these challenges, specific MCDM techniques can be employed as follows:

1. **Data Quality**: By assigning scores and weights to each indicator based on prioritization based on similarity to the ideal solution, the TOPSIS technique can prioritize the data in each cell with high quality and eliminate irrelevant information.
2. **Scalability**: Due to its very simple mathematical relationship, the TOPSIS technique can handle large data sets and identify and weigh critical parameters based on their impact on the performance of each cell.
3. **Real-Time Processing**: The TOPSIS technique enables the integration of multiple criteria and facilitates optimal decision-making based on the weight of each indicator in real time for each cell.
4. **Interpretability**: TOPSIS provides transparent and understandable decision-making frameworks based on the positive and negative impact of each indicator on cell performance.

### *2.3* Research Objectives:

The main objective of this study is to develop a hybrid framework that integrates the TOPSIS technique for cell ranking with machine learning (ML) algorithms for predicting and estimating faulty cells in 5G networks and beyond. The framework is designed to achieve two main objectives:

1. Phase 1: Anomaly Detection - Ranking cells using the TOPSIS technique to identify potential anomalies.

2. Phase 2: Prediction and Estimation - Using ML algorithms to predict and estimate abnormal cells, thereby enhancing the self-healing capabilities of the network.

The aim of this research is to address the aforementioned challenges and leverage the strengths of TOPSIS and ML to help develop efficient, scalable, and real-time interpretable self-healing systems for next-generation cellular networks.

## 3 The proposed algorithm

In this section, in the first phase, the method of identifying defective cells in the network is discussed, and in the second phase, the prediction of defective cells is done. First, using multi-criteria decision-making techniques, the cells are ranked, and the best and worst cells are determined. This algorithm is implemented in four stages, as shown in Figure 3.

**STEP1**

**Data generation (DG)**

**5G NR Network Design with ATOLL and genetrate KPI**

**STEP2**

**Data preprocessing**

**STEP3**

**Phase 1: Anomaly Detection**

**Scoring & Ranking with TOPSIS**

**STEP4**

**Phase 2: Prediction and Estimation**

**Machine learning algorithms for the prediction phase**

**Figure 3: Proposed Method**

- **STEP1**

The dataset utilized in this study was collected from the operational Radio Access Network (RAN) of a nationwide mobile network operator. The data encompasses 67,121 distinct cellular base stations (eNodeBs/cells) over a continuous 24-hour period, capturing the inherent diurnal patterns of network traffic and user behavior. This large-scale, real-world dataset provides a robust foundation for modeling and anomaly detection, reflecting the actual performance and challenges of a live 5G-ready network.

Key characteristics of the dataset are as follows:

- Temporal Granularity: Key Performance Indicators (KPIs) were logged at 15-minute intervals, resulting in 96 time samples per cell. This granularity is standard in network management systems (e.g., NMSS) as it provides a balance between capturing short-term fluctuations and maintaining computational tractability for large-scale analysis.
- Feature Set: For each cell and each time interval, a vector of 12 critical KPIs was extracted. These KPIs were meticulously selected based on 3GPP standards to holistically represent the four fundamental performance domains of a cell: Accessibility, Retainability, Integrity, and Mobility. The specific KPIs include ERAB_Success_Rate,S1Signal_ERAB_Setup_SR,RRC_Setup_Success_Rate, Call_Drop_Rate, IntraF_HOOut_SR, InterF_HOOut_SR,CSSR,Average_CQI,Average_UL_Packet_Loss,CSFB_SR, and RAN_Availability.
- Data Integrity and Cleansing: Raw network data often contains missing or erroneous values due to transient logging errors or cell restarts. A rigorous preprocessing pipeline was employed:

1. Handling Missing Values: Cells with more than 20% missing KPI records over the 24-hour period were entirely removed from the dataset to avoid biased imputation. For the remaining sporadic missing points, a linear interpolation method was applied along the time series for each cell.

2. Outlier Capping: To mitigate the effect of extreme, non-physical values, capping extreme values at the 1st and 99th percentiles for each KPI.

3. Normalization: Given the varying scales of different KPIs, the entire dataset was normalized using Standard Scaling. This process transforms each KPI to have a mean of zero and a standard deviation of one, which is a crucial step for the stability and convergence of many machine learning algorithms.

The final preprocessed dataset represents a high-fidelity snapshot of network performance, comprising over (67,121 cells × 96 time intervals × 12 KPIs).

The output data, including network performance Indicators, are shown in Table4.

**Table 4: Features of data**

| Features(KPI)s | Description |
|---|---|
| ERAB success Rate | Signaling indicator between the user and the network before a call is made. |
| S1Signal_ERAB_Setup_SR | Signaling index to determine user movement in the network |
| RAN_avail_Rate | Radio access index |
| HandOver | An indicator that determines the amount of movement of a user in the network. |
| InterF_HOOut_SR | An indicator that determines a user's frequency shift. |
| IntraF_HOOut_SR | Intercellular transmission determinant at a fixed frequency |
| Call_Drop_Rate | Indicator determining the number of unsuccessful calls |
| CSFB_Rate | Indicator that determines the success of a call |
| Call Setup Success Rate (CSSR) | An indicator expressing the percentage of successful calls in the network. |
| Average_CQI | An indicator that determines the quality of the channel in the network |
| Radio Resource Control (RRC) | Layer 3 signaling index in the network |
| Average_UL_Packet_Loss | An indicator that determines lost packets in the network |

The indices used in this study are listed in Table 4 and categorized in Table 5 based on their diagnostic applications for various cellular failures.

**Table 5: Classification of Indicators based on the type of cellular damage**

| Failure type | Related Indicators | Technical description |
|---|---|---|
| Complete cell failure | -ERAB Successs Rate<br>-S1Signal_ERAB<br>- RAN_avail_Rate | A decrease in these Indicators indicates complete cell failure and users' lack of access to the network. |
| Partial cell failure | - Call_Drop_Rate<br>- CSFB_Rate<br>- Call Setup Successs Rate | An increase in Call_Drop_Rate and a decrease in CSFB_Rate and CSSR indicate a partial cell service disruption. |
| Temporary cell failure | - InterF_HOOut_SR<br>- IntraF_HOOut_SR | A decrease in these Indicators indicates a temporary problem in user handover between cells or frequencies. |
| Cell cover failure | - Average_CQI<br>-Average_ Packet_Loss | A decrease in Average_CQI and an increase in Average_UL_Packet_Loss indicate a problem with cell coverage. |
| Cell software failure | -Radio Resource Control | Failure in RRC performance indicates a software problem in managing radio communications. |
| Cell hardware failure | -RAN_avail_Rate<br>- ERAB Successs Rate | A decrease in these Indicators indicates hardware failure in the cell equipment. |

The selection of Key Performance Indicators (KPIs) for evaluating cell performance is systematically aligned with the four fundamental network performance domains defined in standards such as those by 3GPP: Accessibility, Retainability, Integrity, and Mobility. The twelve KPIs chosen in this study collectively offer broad coverage across these domains with minimal overlap, enabling a holistic assessment of cell operational status.

The rationale for selecting these specific indicators is as follows:

1. Accessibility (Ability to successfully initiate a session):

- Relevant KPIs: ERAB_Success_Rate, S1Signal_ERAB_Setup_SR, Call Setup Success Rate (CSSR), Radio Resource Control (RRC) Setup Success Rate.

These indicators reflect the effectiveness of the initial signaling and bearer establishment procedures. Degradation in these metrics may indicate issues in the Random-Access Channel (RACH), radio resource congestion, or faults in core network connectivity, ultimately preventing users from initiating services.

2. Retainability (Ability to sustain an active service until normal termination):

- Relevant KPI: Call_Drop_Rate.

This metric serves as a direct indicator of service continuity. An elevated drop rate often results from radio link failure (RLF) due to poor coverage, significant interference, or handover failures, directly impacting user satisfaction.

3. Integrity (Perceived service quality):

- Relevant KPIs: Average_CQI, Average_UL_Packet_Loss.

These KPIs quantify user-experienced quality. The Channel Quality Indicator (CQI) reflects downlink signal conditions (SNR) and influences modulation and coding schemes (MCS), thereby affecting throughput. Uplink packet loss may indicate interference, power control issues, or hardware malfunctions. Degradation in these metrics results in reduced data rates, increased latency, and impaired multimedia services.

4. Mobility (Ability to maintain service during inter-cell movement):

- Relevant KPIs: IntraF_HOOut_SR, InterF_HOOut_SR, CSFB_Rate.

Handover success rates are critical for service continuity during user mobility. Failures in intra-frequency or inter-frequency handovers can lead to drops or service degradation. The Circuit-Switched Fallback (CSFB) success rate reflects the performance of legacy voice service interworking between 4G and 3G/2G networks. Joint monitoring of these KPIs allows accurate diagnosis of mobility-related anomalies.

5. Overall Network Health:

- Relevant KPI: RAN_avail_Rate.

This metric indicates the operational status of the cell. While total unavailability (0%) signifies a complete outage, subtle fluctuations may indicate partial failures, software reboots, or hardware instabilities that could affect other performance areas.

The KPI selection process adhered to the following methodology to ensure rigor and avoid arbitrariness:

- Domain-Driven Selection: KPIs were chosen based on comprehensive coverage of 3GPP-defined performance domains.

- Redundancy Analysis: Statistical correlation analysis was applied to a historical dataset to identify and minimize redundancy among KPIs. For instance, although CSSR and ERAB_SR are correlated, both were retained to isolate signaling-level from bearer-level failures. Model-based feature importance analysis further validated the contribution of each KPI.

- Causal Consistency: The selected KPIs exhibit well-established cause-effect relationships. For example, deterioration in CQI can lead to handover failures (IntraF_HOOut_SR) and eventually call drops (Call_Drop_Rate). Such causal links enhance the machine learning model's ability to identify root causes rather than merely detecting symptoms.

In summary, the chosen KPI set provides a non-redundant yet comprehensive feature space that captures the full lifecycle of user sessions—from access and retention to quality and mobility—while accounting for overall cell health. This makes the set particularly suitable for machine learning-based anomaly detection, enabling accurate identification of a wide range of cell impairments. The deficiencies identified by the selected key performance indicators for monitoring cell performance are shown in Table 6.

**Table6: Mapping Between Key Performance Indicators and Corresponding Network Impairments**

| KPI Category | Key Performance Indicator (KPI) | Technical Deficiencies Detectable |
|---|---|---|
| **Accessibility** | **RRC Setup Success Rate** | - Random Access Channel (RACH) problems (e.g., high collision).<br>- Control plane congestion / Signaling overload.<br>- eNodeB software/hardware faults in control processing.<br>- Misconfigured system information parameters. |
| | **S1Signal_ERAB_Setup_SR** | - S1 link failures (transport network issues).<br>- Core Network Element (MME) congestion or failure.<br>- MME-eNodeB version or configuration |

| | | |
|---|---|---|
| | | mismatch.<br>- Security procedure failures (Authentication/Authorization). |
| | **ERAB_Success_Rate** | - User plane path failures (eNodeB to SGW).<br>- Lack of available radio resources (PRB shortage).<br>- QoS parameter misconfiguration or policy failure.<br>- Serving Gateway (SGW/PGW) faults. |
| | **Call Setup Success Rate (CSSR)** | - A composite indicator reflecting all issues from RRC, S1, and ERAB setup.<br>- Specific IMS core network issues for VoLTE calls. |
| **Retainability** | **Call Drop Rate** | - Radio Link Failure (RLF) due to poor coverage or deep fading.<br>- Persistent high interference (leading to low SINR).<br>- Undetected handover failures (too late, too early).<br>- Unstable hardware causing intermittent failures. |
| **Integrity** | **Average CQI** | - Poor downlink coverage (cell edge, obstacles).<br>- Downlink interference (inter-cell, intra-cell).<br>- Faulty or misconfigured antenna (tilt, azimuth, gain).<br>- Unsuccessful load balancing concentrating users at the cell edge. |
| | **Average UL Packet Loss** | - Uplink interference (noise rise).<br>- Incorrect power control settings.<br>- Faulty hardware in the eNodeB uplink receiver chain.<br>- Uplink backhaul packet loss. |
| **Mobility** | **IntraFreq HO Out Success Rate** | - Missing or incorrect neighbor cell relations (NCL).<br>- Poor handover parameter tuning (e.g., A3 offset, hysteresis).<br>- Antenna overshooting causing pilot pollution.<br>- Coverage holes in the handover zone. |
| | **InterFreq HO Out Success Rate** | - All deficiencies listed for IntraFreq HO.<br>- Incorrect measurement gap configuration for UEs.<br>- Unoptimized inter-frequency load balancing. |
| | **CSFB Success Rate** | - SGs interface failure between MME and MSC.<br>- Misconfiguration or congestion in the target 2G/3G cell.<br>- Failure in the redirection or handback procedure to LTE. |
| **Availability** | **RAN Availability Rate** | - Complete cell outage (power, backhaul, hardware failure).<br>- Partial carrier failure in a multi-carrier sector.<br>- Software bugs causing frequent eNodeB reboots.<br>- Scheduled maintenance activities. |

- **STEP2**

TOPSIS method is a way to rank options in decision-making. It uses two main ideas: "ideal solution" and "similarity to ideal solution. " An ideal solution is the best option, which usually doesn't exist, and we aim to get as close as possible to it. We measure how similar a design is to both ideal and anti-ideal solutions by calculating distances. Options are ranked based on the ratio of distance from the anti-ideal to the total distance from both solutions. TOPSIS stands for Technique for Order of Preference by Similarity to Ideal Solution and treats decision-making problems as a geometric system with multiple points in space.

The method focuses on choosing the alternative that is closest to the best possible solution and farthest from the worst solution. TOPSIS uses an index to measure how similar an option is to the best solution and how distant it is from the worst one, selecting the option with the highest similarity. An ideal solution represents the best in all aspects and does not exist in reality, and we aim to get as close to it as possible by measuring distances to both ideal and non-ideal options.

- **STEP3**

To rank each cell based on the evaluated Indicators to determine the type of failure, a score and rank are determined for each cell using the following relationships:

The dataset is designed as a matrix of 67121 rows by 12 columns.

The rows are the grid cells. And columns are the 12 performance Indicators for each cell.

A: TOPSIS method and how to determine the score and rank of each cell:

1: Decision Matrix Normalization:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\Sigma_{i=1}^{n} x_{ij}^{2}}} \quad (1)$$

2: Determining the normalized matrix:

$$v_{ij} = w_j . r_{ij} \quad (2)$$

3: Calculating positive and negative ideals:

$$A^{+} = \{\max(v_{ij} | j \in J), \min(v_{ij} | j \in J')\}$$
$$A^{-} = \{\max(v_{ij} | j \in J), \min(v_{ij} | j \in J')\} \quad (3)$$

4: Calculating distance from the positive and negative ideal solution:

$$D_i^{+} = \sqrt{\sum_{j=1}^{m} (v_{ij} - A_j^{+})^2}$$
$$D_i^{-} = \sqrt{\sum_{j=1}^{m} (v_{ij} - A_j^{-})^2}$$
$$(4)$$

5: Final score calculation:

$$C_i = \frac{D_i^{-}}{D_i^{+} + D_i^{-}}$$
$$(5)$$

In the TOPSIS method, a coefficient of one is considered for 10 Indicators. And two Indicators, Average_Packet_Loss and Call_Drop_Rate, have coefficients of negative 1.

- **STEP4**

In this study, six machine learning algorithms were used to predict defective cells. Each of these algorithms was combined separately with the aforementioned multi-criteria decision-making methods.

**1. Naive Bayes algorithm:** The naive Bayes classification method is a probabilistic approach that assumes data samples and features are functionally independent. This assumption helps calculate the probability of a sample belonging to a certain class. In the naive Bayes method, since we assume independence of samples and features in the initial hypothetical model, we may encounter an increase in classification error. Advantage of the Naive Bayes algorithm: Predicting the class of the test data set is easy and fast. If the data has a normal distribution.

**2. Decision Tree:** A decision tree is a supervised learning method that uses a tree structure for classification or regression. The algorithm makes decisions by partitioning data based on features, with each branch representing a condition and each leaf representing an outcome. Decision trees are simple, interpretable, and fast, but they can suffer from overfitting.

**3. Random Forest:** A supervised learning method based on a set of decision trees. This algorithm increases the accuracy and robustness of the model by creating multiple decision trees and combining their results (usually by majority vote). Each tree is trained on a bootstrap sample of the data and uses a random subset of features for segmentation. This method avoids overfitting and is suitable for classification and regression problems.

**4. Multilayer Perceptron:** Multilayer Perceptron is an Artificial Neural Network with several layers: input, hidden, and output. It uses supervised learning for classification and regression tasks.

**5. Logistic Regression algorithm:** Logistic Regression models the relationship between two variables using a linear equation. One variable is explanatory, and the other is dependent. It is essential to check for a relationship, which is not the same as causation. A scatter plot can assess this, and a correlation coefficient shows the relationship's strength. In the equation Y = a + bX, Y is the dependent variable, X is the explanatory variable, b is the slope, and a is the intercept. Logistic regression performs well when the data set has an uneven class distribution.

**6. AdaBoost algorithm:** AdaBoost combines weak learners to improve performance.

# 4 Results

## 4.1 Results from Phase One of the Research

- **Phase 1: Anomaly Detection**

One of the most widely used methods in multi-criteria decision-making (MCDM) is the TOPSIS technique. In this part of the research, these techniques were used to determine the best and worst cells. The results of this study show that the best cell is cell 724, and the worst cell is cell 15562. In Table 7, the top 10 normal cells and 10 worst cells are ranked based on the methods used in this study.

**Table 7: The best and worst cells by applied techniques**

| Type of technique | 10 normal cells | SCORE normal cells | 10 abnormal cells | SCORE abnormal cells |
|---|---|---|---|---|
| **TOPSIS** | cell724 | 0.999716 | cell11494 | 0.733869 |
| | cell53459 | 0.998974 | cell63608 | 0.733644 |
| | cell43998 | 0.998533 | cell54280 | 0.732613 |
| | cell38399 | 0.997228 | cell10079 | 0.726769 |
| | cell14906 | 0.996328 | cell56673 | 0.725263 |
| | cell65102 | 0.994893 | cell47001 | 0.722343 |
| | cell36971 | 0.993321 | cell19613 | 0.690839 |
| | cell7308 | 0.992813 | cell53637 | 0.68492 |
| | cell43478 | 0.991826 | cell52784 | 0.683055 |
| | cell64067 | 0.991705 | cell15562 | 0.6697 |

## 4.2 Results from Phase Two of the Research

- **Phase 2: Prediction and Estimation**

The results of applying machine learning algorithms to predict defective cells are shown in Table 8.

**Table 8: Comprehensive evaluation of algorithms based on evaluation criteria**

| Algorithm | Accuracy | Kappa | MAE | Run time(s) |
|---|---|---|---|---|
| Naive Bayes | 91.143% | 0.1581 | 0.0485 | 0.18 |
| Decision Tree | 92.6878% | 0.3023 | 0.1696 | 0.04 |
| Random Forest | 93.658% | 0 | 0.1683 | 0.08 |
| Multilayer Perceptron | 91.4027% | 0.4458 | 0.0861 | 46.4 |
| Logistic Regression | 97.7881% | 0.9787 | 0.00023 | 9.84 |
| AdaBoost | 98.886% | 0.9886 | 0.00018 | 0.1298 |

**Table 9: Evaluation based on Precision, Recall, F-Measure, Normal cell=B, AB Normal cell=A**

| Algorithm | Class | Precision | Recall | F-Measure |
|---|---|---|---|---|
| Naive Bayes | A=0 | 0.913 | 0.964 | 0.938 |
| | B=1 | 0.866 | 0.887 | 0.872 |
| Decision Tree | A=0 | 0.904 | 0.914 | 0.900 |
| | B=1 | 0.923 | 0.985 | 0.953 |
| Random Forest | A=0 | 0.936 | 0.949 | 0.944 |
| | B=1 | 0.897 | 0.900 | 0.899 |
| Multilayer Perceptron | A=0 | 0.750 | 0.360 | 0.486 |
| | B=1 | 0.913 | 0.964 | 0.938 |
| Logistic Regression | A=0 | 0.968 | 0.979 | 0.979 |
| | B=1 | 0.987 | 0.988 | 0.988 |
| AdaBoost | A=0 | 0.978 | 0.989 | 0.989 |
| | B=1 | 0.988 | 0.988 | 0.988 |

The results of the first phase of the research show that the combination of machine learning and multi-criteria decision-making (MCDM) effectively overcomes the challenges of data quality, scalability, real-time processing, interpretability, and complexity of network structures. This approach has provided significant improvements in network reliability and efficiency. In the second phase of the research, based on the results obtained, it can be concluded that the machine learning algorithms have shown acceptable results (as shown in Table 8).

## 5 Summary

The aim of the research was to investigate the challenges of anomaly detection and self-healing in 5G networks and beyond by combining machine learning (ML) algorithms with the TOPSIS technique. The study was conducted in two stages of fault detection and prediction. In the first stage, the TOPSIS (Topic Preference Ranking Technique Based on Similarity to the Ideal Solution) method was used to systematically evaluate, rank, and classify network cells based on multi-criteria performance criteria, allowing for optimal identification of high-performing and low-performing cells for targeted corrective actions. In the second stage, supervised algorithms were used to predict and estimate faulty cells. The results showed significant improvements in network reliability, efficiency, real-time decision-making, and scalability. However, the following are suggested for future research in this area:

1. Use of advanced algorithms and techniques to overcome the challenge of unstructured data.
2. Application of multi-objective decision-making approaches to classify cell failures more effectively.
3. Utilization of generative artificial intelligence to build a mobile network-specific model for the Self-Organizing Network (SON) domain.
4. The combination of multi-user blind detection with asynchronous variable processing—a key advantage of wide-bandwidth IoT systems [27].Enables accurate assessment of cell operating status even in the absence of active users. This approach facilitates the detection of cells with no human user traffic while identifying potential issues such as hidden device activity or abnormal noise floor behavior. Consequently, network operators can proactively detect anomalies and perform maintenance on idle cells before failures occur.

# 6 References

[1] Moysen, J., & Giupponi, L. (2018). From 4G to 5G: Self-organized network management meets machine learning. *Computer Communications*, 129, 248-268. doi.org/10.1016/j.comcom.2018.07.015

[2] Klaine, P. V., Imran, M. A., Onireti, O., & Souza, R. D. (2017). A survey of machine learning techniques applied to self-organizing cellular networks. *IEEE Communications Surveys & Tutorials*, 19(4), 2392-2431. doi.org/10.1109/comst.2017.2727878.

[3] Zhang, T., Zhu, K., & Hossain, E. (2022). Data-driven machine learning techniques for self-healing in cellular wireless networks: Challenges and solutions. *Intelligent Computing.* doi.org/10.34133/2022/9758169

[4] Luo, F. L. (Ed.). (2020). Machine learning for future wireless communications.

[5] Farmani, J., & Zadeh, A. K. (2023). AI-based self-healing solutions applied to cellular networks: An overview. *arXiv preprint arXiv*:2311.02390. doi.org/10.48550/arXiv.2311.02390

[6] Charbuty, B., & Abdulazeez, A. (2021). Classification based on the decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28. doi.org/10.38094/jastt20165.

[7] Li, W. H., & Qi, Z. (2018). Network selection algorithm based on decision tree in heterogeneous wireless networks. *In MATEC Web of Conferences* (Vol. 189, p. 04010). EDP Sciences. doi.org/10.1051/matecconf/201818904010.

[8] Sivakumar, N., Srinivasan, P., Viswanath, N., & Venkateswar. (2022). Decision tree-based radio link failure prediction for 5G communication reliability. *ITU Journal on Future and Evolving Technologies*, 3(2), 142–156. doi.org/10.52953/lzlj8762.

[9] Saeed, U., Jan, S. U., Lee, Y.-D., & Koo, I. (2021). Fault diagnosis based on extremely randomized trees in wireless sensor networks. *Reliability Engineering & System Safety*, 205, Article 107284. doi.org/10.1016/j.ress.2020.107284

[10] Fernandes, A. A. T., Figueiredo Filho, D. B., da Rocha, E. C., & Nascimento, W. da S. (2020). Read this paper if you want to learn logistic regression. *Revista de Sociologia e Política*, 28(74). doi.org/10.1590/1678-987320287406en

[11] Kulkarni, A. M., Saini, G., & Pattnaik, S. S. (2023). Antenna array fault detection using logistic regression technique. In *Communications in Computer and Information Science* (pp. 13–29). Springer Nature Switzerland. doi.org/10.1007/978-3-031-48781-1_2

[12] Sharma, V., Hong, T., Cecchi, V., Hofmann, A., & Lee, J. Y. (2023). Forecasting weather-related power outages using weighted logistic regression. *IET Smart Grid*, 6(5), 470–479. doi.org/10.1049/stg2.12109

[13] Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: Random forest. In *Information Computing and Applications* (pp. 246–252). Springer Berlin Heidelberg. doi.org/10.1007/978-3-642-34062-8_32

[14] El-Aziz, L. A., et al. (2020). Cell outage detection and degradation classification based on alarms and KPI's correlation. In *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*. doi.org/10.1109/niles50944.2020.9257920

[15] Balyan, A. K., et al. (2022). A hybrid intrusion detection model using EGA-PSO and improved random forest method. *Sensors*, 22(16), Article 5986. doi.org/10.3390/s22165986

[16] Samidi, F. S., Mohamed Radzi, N. A., Mohd Azmi, K. H., Mohd Aripin, N., & Azhar, N. A. (2022). 5G technology: ML hyperparameter tuning analysis for subcarrier spacing prediction model. *Applied Sciences*, 12(16), Article 8271. doi.org/10.3390/app12168271

[17] Ali-Tolppa, J., Kajo, M., Gajic, B., Malanchini, I., Schultz, B., & Liao, Q. (2020). Cognitive autonomy for network self-healing. *In Toward Cognitive Autonomous Networks* (pp. 345–384). doi.org/10.1002/9781119586449.ch9

[18] Onireti, O., et al. (2016). A cell outage management framework for dense heterogeneous networks. *IEEE Transactions on Vehicular Technology*, 65(4), 2097–2113. doi.org/10.1109/tvt.2015.2431371

[19] Zoha, A., Saeed, A., Imran, A., Imran, M. A., & Abu-Dayya, A. (2015). A learning-based approach for autonomous outage detection and coverage optimization. *Transactions on Emerging Telecommunications Technologies*, 27(3), 439–450. doi.org/10.1002/ett.2971

[20] Asghar, M. Z., Ahmed, F., & Hamalainen, J. (2021). Artificial intelligence enabled self-healing for mobile network automation. In *2021, IEEE GlobeCom Workshops (GC Wkshps).* doi.org/10.1109/gcwkshps52748.2021.9681937

[21] Ruan, Y., Wang, Y., & Tang, Y. (2021). An intelligent cell outage detection method in cellular networks. In *2021, the 16th International Conference on Computer Science & Education (ICCSE).* doi.org/10.1109/iccse51940.2021.9569511

[22] Yu, P., Wang, H., & Chen, Z. (2022). Active cell outage detection algorithm for broadband service in 5G cloud radio access networks. In *2022, the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB).* doi.org/10.1109/bmsb55706.2022.9828778

[23] Yu, P., et al. (2020). Deep reinforcement learning aided cell outage compensation framework in 5G cloud radio access networks. *Mobile Networks and Applications*, 25(5), 1644–1654. doi.org/10.1007/s11036-020-01574-8

[24] Iwamoto, M., Suzuki, A., & Kobayashi, M. (2023). Deep reinforcement learning based antenna selection for cell outage compensation. In *ICC 2023 – IEEE International Conference on Communications*. doi.org/10.1109/icc45041.2023.10279017

[25] Klaine, P. V., Imran, M. A., Onireti, O., & Souza, R. D. (2017). A survey of machine learning techniques applied to self-organizing cellular networks. *IEEE Communications Surveys & Tutorials*, *19*(4), 2392–2431. doi.org/10.1109/COMST.2017.2727878

[26] Yami, R. M., Khazaei, A. A., & Quchani, S. R. (2024). Machine learning algorithms with multi-criteria-decision-making for cell outage detection in 5G/B5G. In *2024 11th International Symposium on Telecommunications (IST)* (pp. 44–52). IEEE. doi.org/10.1109/IST64061.2024.10843542

[27] Samsami Khodadad, F. (2025). Blind multiuser detection in asynchronous variable processing gain in spread spectrum IoT systems. *Future Research on AI and IoT*, 19–23. Doi.org/ 10.22080/frai.2025.5663