ARCADE: A RAN Diagnosis Methodology in a Hybrid AI Environment for 6G Networks

Daniel Ricardo Cunha Oliveira[1], **Rodrigo Moreira**[1,2]
Flávio de Oliveira Silva[1,3]

[1] Faculty of Computing (FACOM) – Federal University of Uberlândia (UFU)
Uberlândia – MG – Brazil

[2]Federal University of Viçosa (UFV)
Rio Paranaíba – MG – Brazil

[3]Department of Informatics – School of Engineering
University of Minho – Braga, Portugal

drcoliveira@ufu.br, rodrigo@ufv.br, flavio@di.uminho.pt

Abstract. *Artificial Intelligence (AI) plays a key role in developing 6G networks. While current specifications already include Network Data Analytics Function (NWDAF) as a network element responsible for providing information about the core, a more comprehensive approach will be needed to enable automation of network segments that are not yet fully explored in the context of 5G. In this paper, we present Automated Radio Coverage Anomalies Detection and Evaluation (ARCADE), a methodology for identifying and diagnosing anomalies in the cellular access network. Furthermore, we demonstrate how a hybrid architecture of network analytics functions in the evolution toward 6G can enhance the application of AI in a broader network context, using ARCADE as a practical example of this approach.*

1. Introduction

In recent years, cellular networks have become an indispensable and essential component of society, particularly in a globalized digital community that relies on ubiquitous and reliable services. The exponential increase in connected devices has made Radio Access Networks (RANs) highly complex, especially considering the coexistence of multiple technology generations and the growing demand for Internet of Things (IoT)-based solutions. In this context, there is an increasing need for solutions that simplify RAN optimization, particularly since, in systems such as 4G and 5G—and potentially in 6G, should the next generation adopt an evolutionary approach to the access network—network self-interference is inversely correlated with efficiency and traffic capacity. Traditional network optimization methods, which rely on manual processes and highly specialized technical expertise, are becoming increasingly ineffective as system capacity and complexity expand.

This study presents ARCADE, a methodological approach for evaluating and identifying coverage anomalies in a standard 3GPP cellular system. More importantly, we address a key issue: how ARCADE acquires data from the RAN to analyze coverage and interference among cells within a specific cluster under investigation. This is

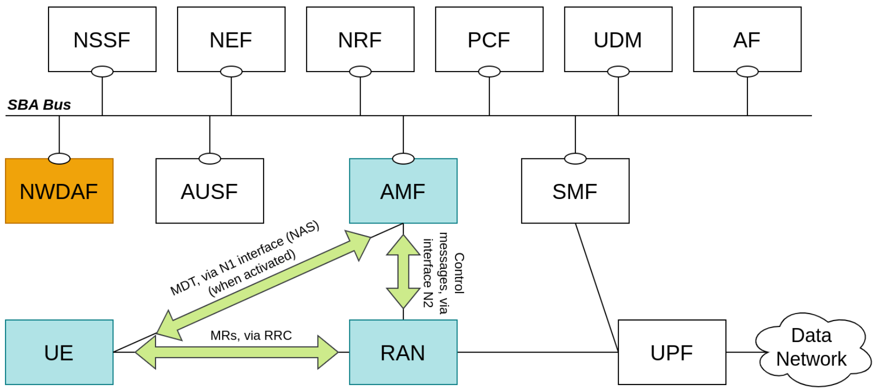

Figure 1. Simplified signaling diagram of MRs and Minimization of Drive Tests (MDT) between network elements. Figure created by the authors.

achieved through a hybrid architecture proposal for implementing AI in the future 6G core network.

This work is organized as follows: Section 2 presents an overview of related work in this area. Section 3 describes ARCADE. Section 4 contextualizes an NWDAF evolution. Section 5 correlates ARCADE and the NWDAF evolution. Section 6 concludes this study.

2. Related Work

To provide analytics services in mobile networks, 3GPP has defined NWDAF [3GPP 2023] as the central AI element in the 5G core topology.

NWDAF is integrated into the 5G Core (5GC) architecture as an independent network function, interacting with other network functions through standardized 3GPP interfaces. Although some researchers claim that NWDAF is responsible for AI and machine learning tasks, it was not initially designed for such purposes and is instead restricted to analytics services [Wu et al. 2021]. The current standardized architecture does not allow NWDAF to access raw radiofrequency (RF) data, particularly Measurement Reports (MRs), which contains rich information about coverage, neighboring cells, and network quality. This limitation arises from the centralized deployment of this network element, as illustrated in Figure 1, where NWDAF connects to other Network Functions (NFs) solely through the Service-Based Architecture (SBA) message bus.

Thus, a key challenge lies in developing a solution that enables RAN data to be stored and potentially processed by the network architecture's analytics elements, particularly NWDAF.

A hybrid approach to evolving an architecture that enables the use of AI in next-generation networks, as described in this study, was proposed in [Neto et al. 2024], contrasting with implementation perspectives based on centralized ([Abbas et al. 2022]) or distributed architectures ([Jeon et al. 2022]). Due to space constraints, a detailed overview of the proposed hybrid approach for the NWDAF cannot be included here. However, a comprehensive discussion is available in [Neto et al. 2024].

Furthermore, several studies have proposed approaches similar to ARCADE. Some efforts focus on modeling digital twins for mobile networks and RF environments

Table 1. Short State-of-the-Art Survey.

Work	ML/Neural network-based	Independent of project data	Independent of analytical prediction	Supports coverage anomalies	Independent of system KPIs	Applies to 3GPP systems (RAN)
Juan Deng *et* al. 2021 [Deng et al. 2021]	○	○	○	○	○	●
MingXue Wang *et* al. 2018 [Wang 2018]	●	●	○	●	●	●
Anita Gehlot *et* al. 2022 [Gehlot et al. 2022]	●	·	·	·	○	●
Kirkwood *et* al. 2022 [Kirkwood et al. 2022]	●	·	●	·	·	○
Skocaj *et* al. 2022 [Skocaj et al. 2022]	●	○	○	●	○	●
Nguyen, 2021 [Nguyen et al. 2021]	·	·	·	·	·	·
Zhou, 2023 [Zhou and Peng 2023]	●	·	●	·	·	○
Tang, 2023 [Tang et al. 2023]	●	○	●	●	●	○
Carmack, 2021 [Carmack et al. 2021]	●	·	·	·	·	●
Cheerla, 2018 [Cheerla et al. 2018]	●	○	○	·	●	●
Dreifuerst, 2021 [Dreifuerst et al. 2021]	●	○	●	●	·	●
Ojo, 2020 [Ojo et al. 2021]	●	○	●	·	·	●
This work	●	●	●	●	●	●

in general, as presented in [Deng et al. 2021] and [Nguyen et al. 2021]. However, these approaches are not exclusively focused on RAN or do not aim to model cellular coverage. None of the reviewed studies exhibits all the ideal characteristics within the proposed context. Therefore, we contrast our work with the existing literature in Table 1, highlighting essential features and their differences. Small dots indicate that a given criterion does not apply to the respective study.

3. The ARCADE Methodology

In this section, we briefly describe the ARCADE methodology, an approach for detecting and evaluating anomalies in the coverage of a cellular system. Due to the limited space available for publication, we emphasize that this work does not aim to provide a detailed description of ARCADE – an endeavor that will be addressed in future articles. Instead, our objective is to leverage ARCADE's framework to demonstrate how a hybrid architecture integrating AI elements within a 6G core could enhance applications that currently lack access to critical data due to the constraints of the existing 5G network topology.

3.1. Proposal

ARCADE introduces a methodology for the automated identification of cellular coverage anomalies, relying solely on georeferenced cell coverage level data, or Reference Signal Received Power (RSRP), without requiring information from Radio Base Station (RBS) design plans, network management system Key Performance Indicators (KPIs), or geographic and morphological databases—resources commonly used in most existing approaches [Wang 2018] [Skocaj et al. 2022] [Dreifuerst et al. 2021] [Ojo et al. 2021]. Moreover, the coverage data may be sparse and not fully encompass the analyzed area, necessitating coverage extrapolation based on the available data. In addition to the absence of design data, ARCADE does not employ mathematical prediction models, excluding, for instance, classical approaches such as Okumura, Hata, or COST-231 [Singh 2012]. Eliminating these dependencies simplifies the solution, given the complexity of obtaining reliable and up-to-date field data.

3.2. Data Acquisition

ARCADE takes as input RSRP samples from various cell Physical Cell IDs (PCIs) within the analyzed cluster. Additionally, the samples must be georeferenced, meaning they contain geographic coordinates indicating where each measurement was collected. These coordinates position the samples within a grid that maps the entire area of interest. Each grid element stores information on PCI and RSRP, as illustrated in Figure 2.

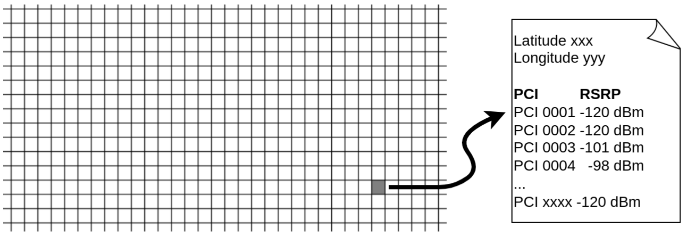

Figure 2. Georeferenced grid structure and information per grid element. Figure created by the authors.

Acquiring these samples is a challenge in itself. Several sources can be used to collect RF environment data and obtain information about the state of coverage and interference in a mobile system. One option is drive test (DT), which involves using probes on vehicles to capture and decode RBS signals, storing the data along with geolocation obtained via a Global Navigation Satellite System (GNSS). More recently, however, data acquisition has become possible through the 3GPP-defined MDT functionality [3GPP 2024]. This feature allows the cellular network to collect MRs, including georeferenced data, using modern smartphones' built-in GNSS capabilities. These devices provide critical RF measurement data, such as cell identification via PCI and RSRP, among other metrics, offering a valuable resource for mapping the radio environment of a cellular system. However, the industry adoption of MDT remains limited, and additional constraints on its usage persist, making the optimal solution for acquiring RF environment data in a cellular system an open problem.

3.3. Coverage Extrapolation

Conceptually, ARCADE requires a complete RF data table for each PCI within the grid covering the area of interest. Since not all cells have measurements in every grid element, coverage extrapolation for each cell is necessary. This generalization must model coverage as accurately as possible, considering both anomalous coverage samples (which may indicate suboptimal coverage configurations, such as improper antenna positioning) and outliers caused by device measurement errors (which should be discarded). The potential confusion between coverage anomalies and measurement errors necessitates a sample classification methodology. Once the samples are classified, the total grid area is divided into *normal* and *abnormal* regions, referring to areas where coverage is expected versus areas where coverage results from anomalies such as cell overshooting.

Once each cell's normal and abnormal regions are defined, sample augmentation is applied to the coverage anomaly areas. This augmentation serves two purposes: first, to establish boundary conditions that prevent anomalies from causing the effect of overfitting in the subsequent modeling with Artificial Neural Network (ANN); second, to emphasize the impact of anomalies so that the modeling process effectively captures potential overshooting effects. Our approach relies on data extrapolation using Gaussian processes with a spatial kernel, specifically the Radial Basis Function (RBF) kernel [Schölkopf 2002].

Once a sufficient number of samples allows for characterizing elements outside the normal coverage area—highlighting anomaly regions and establishing boundary conditions in areas distant from the cell's primary coverage—it is proposed that an ANN be trained with this data to generically model coverage across all grid elements in the total area. The structure of the ANN should be determined through experiments evaluating its efficiency and accuracy in generalizing the model. After this sequence of processes, the entire RF environment is ultimately described, enabling the classification of cells based on the normality of their coverage.

3.4. Coverage Anomaly Identification

Finally, once the final dataset to be evaluated is obtained, containing the information shown in Figure 2, each cell's coverage and interference assessment phase begins. This phase determines the cell interaction and how the system could be optimized by increasing or reducing their coverage. It is impossible to analyze a single cell's coverage without considering the context of other cells within the cluster. Even in a highly efficient and optimized system, interference between neighboring cells is inevitable. There is a minimum acceptable level of overlap between adjacent cells, below which coverage adjustments to reduce interference would degrade service quality. Therefore, it is necessary to define measurable parameters for evaluating both the coverage of a cell within the context of a cluster and the interference among them.

To this end, indicators are defined and associated with the PCIs, characterizing their coverage-related attributes, such as *Coverage Index*, *Interfering Index*, *Interfered Index*, *Overlap Index*, *Coverage Quality Index*, and *Coverage Matrix*. Analyzing these indicators and their correlations makes it possible to identify anomalous cells and diagnose the RAN in a cellular system.

4. Evolved NWDAF and the Hybrid Approach

The hybrid architecture for NWDAF and the definition of Enhanced Network Data Analytics Function (eNWDAF) were proposed in [Neto et al. 2024]. The architecture introduces *AI-Agents*, capable of collecting data from the RAN and other network points that are currently inaccessible to NWDAF as specified today. This enhancement expands the role of NWDAF, integrating broader AI functions in terms of network topology reach. The eNWDAF provides *AI for Networking*, *Networking for AI*, and *analytics* support within a single component. Its ability to access network elements beyond the core and its positioning within the SBA bus characterize it as a hybrid architecture. Figure 3 presents a conceptual illustration of this approach.

5. ARCADE in the Context of the Hybrid AI Approach

An example of ARCADE application within the new architecture illustrates the concepts proposed above. In this specific case, two key types of data may be collected: MRs and MDT data.

As previously mentioned, MRs data are transmitted from the User Equipment (UE) to the RAN and remain inaccessible to the NWDAF in the current 5G architecture. In the proposed architecture, the RBSs elements, which constitute the RAN, incorporate an *AI-Agent* in their baseband processing units. These agents would collect, anonymize, classify, and consolidate the measurement data received. Once the data are organized adequately into "packets", they can be transmitted to the eNWDAF for centralized analysis by ARCADE. This organization may include, for instance, the consolidation of measurements per UE over a short period, enabling the characterization of a measurement as a specific sample from a geographical point in the RF environment.

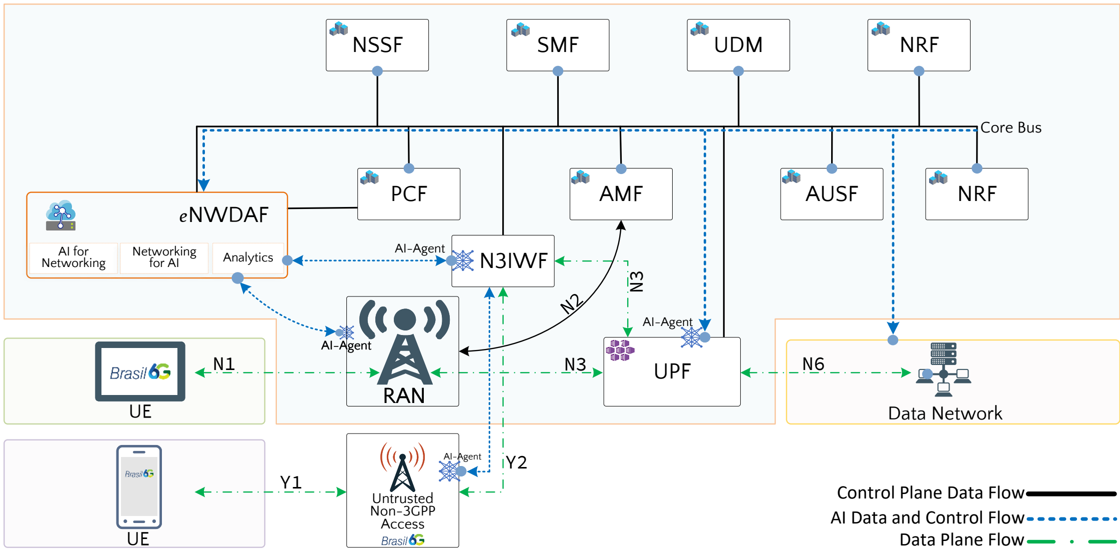

Figure 3. Advancements in AI-Related Core: The eNWDAF in 6G Architecture. Source: [Neto et al. 2024]

The transmission of MDT data is already standardized by 3GPP. However, the volume of available information may be reduced due to limited compatibility of UEs with the technology among the user base or due to 3GPP specification requirements, which mandate user consent for data collection. Although the available data volume may be lower, MDT data can serve as a highly valuable training base for AI, as they contain not only coverage information but also sample coordinates, unlike MR measurements, which lack georeferencing. In this case, the eNWDAF would use MDT data to train an ANN, enabling it to infer the coordinates of MRs samples sent to the primary element by the *AI-Agent* within the RBSs. This approach creates an optimal data provisioning scenario for ARCADE, ensuring a high volume of information even if user adoption of MDT functionality is low. Alternatively, in the absence of MDT data for training the ANN, information from alternative sources could be employed, such as proprietary solutions implemented by technology vendors or crowdsourced data made available by various app providers that collect such information.

This process clearly positions eNWDAF within the role of *Networking for AI* and Figure 4 illustrates how ARCADE can be implemented within this hybrid approach to 6G network elements. It highlights its contribution to information acquisition and AI application in the context of mobile network self-optimization, leveraging the concepts presented in this section.

6. Concluding Remarks

In this work, we introduced the concept of ARCADE. We demonstrated how a cellular coverage diagnostic system can benefit from a hybrid approach to positioning AI elements implemented centrally yet extending across all network segments, including the RAN. Such an approach can be highly valuable in 6G systems, where AI will be pervasive throughout the network, aiming to enhance automation, resource efficiency, and overall system and infrastructure management.

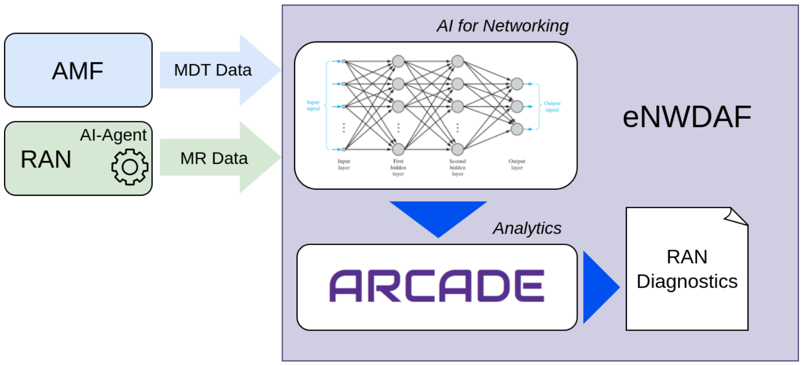

Figure 4. Implementation of ARCADE within the hybrid architecture and eNWDAF concepts. Figure created by the authors.

There are several ways to develop this proposal further. A practical implementation of a proof of concept for the eNWDAF, its interfaces, and the *AI-Agent* should be carried out as a first step. This should be followed by a validation phase using either simulated network data or a real-world dataset collected from a mobile operator. Additionally, the implementation of ARCADE must be completed and integrated into the proof of concept, thereby establishing a comprehensive environment for testing and validating the proposed approach.

Acknowledgement

The authors thank the support of FAPEMIG (Grant APQ00923-24) and FCT – Fundação para a Ciência e Tecnologia within the RD Unit Project Scope UID/00319/Centro ALGORITMI (ALGORITMI/UM) for supporting this work. The authors also thank Algar Telecom for their support in our research.

References

3GPP (2023). System architecture for the 5g system (5gs). Technical Specification TS 23.501, 3rd Generation Partnership Project (3GPP). Release 18.

3GPP (2024). 3rd generation partnership project; technical specification group radio access network; radio measurement collection for minimization of drive tests (mdt); overall description; stage 2. Technical Report TS 37.320, 3GPP.

Abbas, K., Khan, T. A., Afaq, M., and Song, W.-C. (2022). Ensemble Learning-based Network Data Analytics for Network Slice Orchestration and Management: An Intent-Based Networking Mechanism. In *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*, pages 1–5. ISSN: 2374-9709.

Carmack, J. et al. (2021). Neural Network Generative Models for Radio Frequency Data. In *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 0577–0582, New York, NY, USA. IEEE.

Cheerla, S., Ratnam, D. V., and Borra, H. S. (2018). Neural network-based path loss model for cellular mobile networks at 800 and 1800 MHz bands. *AEU - International Journal of Electronics and Communications*, 94:179–186.

Deng, J. et al. (2021). A Digital Twin Approach for Self-optimization of Mobile Networks. In *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 1–6, Nanjing, China. IEEE.

Dreifuerst, R. M. et al. (2021). Optimizing Coverage and Capacity in Cellular Networks using Machine Learning. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8138–8142, Toronto, ON, Canada. IEEE.

Gehlot, A. et al. (2022). Application of Neural Network in the Prediction Models of Machine Learning Based Design. In *2022 ICSES*, pages 1–6, Chennai, India. IEEE.

Jeon, Y., Jeong, H., Seo, S., Kim, T., Ko, H., and Pack, S. (2022). A distributed nwdaf architecture for federated learning in 5g. In *2022 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–2.

Kirkwood, C., Economou, T., Pugeault, N., and Odbert, H. (2022). Bayesian Deep Learning for Spatial Interpolation in the Presence of Auxiliary Information. *Mathematical Geosciences*, 54(3):507–531.

Neto, N. S., Gonçalves, M., Oliveira, D., Molinos, D., Moreira, R., and Silva, F. (2024). Evolved nwdaf towards a fully distributed artificial intelligence in the 6g network architecture. In *Anais do IV Workshop de Redes 6G*, pages 15–25, Porto Alegre, RS, Brasil. SBC.

Nguyen, H. X., Trestian, R., To, D., and Tatipamula, M. (2021). Digital Twin for 5G and Beyond. *IEEE Communications Magazine*, 59(2):10–15.

Ojo, S., Imoize, A., and Alienyi, D. (2021). Radial basis function neural network path loss prediction model for LTE networks in multitransmitter signal propagation environments. *International Journal of Communication Systems*, 34(3):e4680.

Schölkopf, B. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass.

Singh, Y. (2012). Comparison of Okumura, Hata and COST-231 Models on the Basis of Path Loss and Signal Strength. *International Journal of Computer Applications*, 59.

Skocaj, M. et al. (2022). Cellular Network Capacity and Coverage Enhancement with MDT Data and Deep Reinforcement Learning. *Computer Communications*, 195:403–415.

Tang, Z. et al. (2023). Multi-Output Gaussian Process-Based Data Augmentation for Multi-Building and Multi-Floor Indoor Localization. arXiv:2202.01980 [cs].

Wang, M. (2018). Anomaly Detection for Mobile Network Management. 9(2):19.

Wu, J. et al. (2021). Toward Native Artificial Intelligence in 6G Networks: System Design, Architectures, and Paradigms. *arXiv:2103.02823 [cs]*.

Zhou, T. and Peng, Y. (2023). Gaussian process regression based on deep neural network for reliability analysis in high dimensions. *Structural and Multidisciplinary Optimization*, 66(6):131.