Towards Explainable Multimodal Analytics for 5G+ Network Monitoring and Management: Methods, Architecture, and Preliminary Results

Michał Błaszczak[1,2,*], Aleksandra Knapińska[2], Ireneusz Jabłoński[3,4], Krzysztof Walkowiak[2]
[1]Nokia Solutions and Networks, Wrocław, Poland
[2]Wrocław University of Science and Technology, Wrocław, Poland
[3]Fraunhofer Institute for Photonic Microsystems, Dresden, Germany
[4]Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany
[*]*michal.blaszczak@pwr.edu.pl*

***Abstract*—5G+ (5G and beyond) mobile networks are increasingly complex, making their monitoring and management challenging. This paper presents a PhD research focused on automating 5G+ network performance assessment through a family of Multimodal Data Analysis Methods (MDAM) in 5G+ cellular networks. We identify limitations of unimodal approaches, describe the design-science methodology adopted, and present the Module for Automated and Context-Preserving Telecommunication Network State Analysis (MACNSA) – a modular analytics platform hosting MDAM. The first MDAM implemented within MACNSA is an agentic RAG-based system (Agentic-RAG-MDAM) that fuses numerical KPI data with textual feature documentation. The evaluation protocol, based on both traditional information retrieval and LLM-as-a-judge metrics, is described. Preliminary experiments on real operator data demonstrate that Agentic-RAG-MDAM significantly reduces per-feature analysis time compared to traditional human expert approaches. Current limitations and planned future work are discussed.**

***Index Terms*—5G+ Mobile Networks, Network Performance Modelling, Multimodal Data Analysis Methods, Multimodal Data Fusion, Decision Support, Explainable AI, Agentic Systems**

I. Introduction

Traditional approaches to performance modeling of 5G+ network infrastructure face limitations due to data sparsity and heterogeneity. Network performance data includes numerical measurements such as Key Performance Indicators (KPIs), as well as textual technical documentation regarding feature activation, configuration, and interdependencies. Unimodal methods that rely on a single data type cannot capture cross-modal dependencies critical for accurate network assessment (see Section II), motivating the exploration of multimodal approaches. This research is conducted in collaboration between Wroclaw University of Science and Technology and Nokia, ensuring that the developed solutions are grounded in real industrial needs and validated on operational network data.

This paper presents a PhD research aimed at automating the monitoring and management of 5G+ cellular network conditions through a family of Multimodal Data Analysis Methods (MDAM) methods – including multimodal machine learning (MML), Multimodal Graph Learning (MGL), and Foundation Models adapted for Time Series analysis (TSFMs). The research is guided by the following Research Questions (RQs), with their current progress indicated:

- **RQ 1** *(in progress):* What is the accuracy of quantitative inference in multimodal data fusion models applied to monitoring the condition of 5G+ cellular networks?
- **RQ 2** *(in progress):* What is the impact of different alignment strategies on the performance of predictive models in multimodal data fusion frameworks?
- **RQ 3** *(in progress):* How does the integration of descriptive textual features with time series data influence the accuracy and robustness of predictive modeling?
- **RQ 4** *(planned):* Is it possible to develop a function for evaluating the quality of methods considering the specific properties of the data?
- **RQ 5** *(planned):* How does accuracy of explainable data analytics methods compare to "black-box" counterparts?
- **RQ 6** *(planned):* How does the incorporation of explainability into multimodal data analytics methods affect the trade-off between possible accuracy gains and computational cost?

The contributions are: (1) identification of research gaps in multimodal methods for telecommunications, (2) the MDAM family of methods and the Module for Automated and Context-Preserving Telecommunication Network State Analysis (MACNSA) developed to host them, and (3) evaluation protocol and preliminary findings from the first MDAM – the Agentic-RAG-MDAM.

II. Background and Related Work

A. Challenges in 5G+ network performance modeling

Mobile networks are experiencing growing complexity due to service diversification, network virtualization, and the expanding number of features, making human troubleshooting and manual monitoring infeasible [1]. User profiles and mobility patterns are evolving, and service distribution takes place over increasingly vast geographical areas. The growing scale and dynamism of the network infrastructure require

ISBN 978-3-903176-82-9

TABLE I
CROSS-DOMAIN EVIDENCE OF MULTIMODAL FUSION IMPROVEMENTS OVER UNIMODAL BASELINES.

Domain	Improvement	Ref.
Autonomous vehicles	3.7%	[14]
Social media video	11.6%	[15]
Medical imaging & clinical data	1.2–27.7%	[16]
Predictive maintenance	2.6–16.9%	[17]
Network traffic	11.6%	[18]

automation of monitoring and management processes, not only for efficiency, but also for meeting stringent Service-Level Agreements (SLAs). Self-Organizing Network (SON) [2] and Autonomous Network (AN) concepts aim to address these challenges, but face obstacles including training issues, lack of explainability, uncertainty in generalization, and lack of interoperability [3], [4]. Moreover, the Zero-touch network and Service Management (ZSM) framework envisions fully automated network operations, yet achieving this vision requires intelligent models capable of processing diverse data sources and making trustworthy decisions [5].

Unimodal methods, relying on a single data type (e.g., time series), can miss contextual information critical for accurate predictions [6]. Studies have shown that observing complexity using a single data modality is subject to inherent limitations [7]–[9]. For example, a shift in KPI data caused by feature activation remains unexplained when the model relies solely on numerical data, since the contextual information about the feature deployment resides in textual documentation. Moreover, deep learning models, although accurate, lack the explainability required for critical telecom infrastructure [10], limiting their trustworthiness and adoption in operational environments [11].

B. Multimodal machine learning

Multimodal Machine Learning (MML) leverages data of different modalities to improve modeling of a phenomenon of interest, enabling broader applicability by exploiting complementary information and facilitating knowledge transfer when one data type is scarce [12]. Key challenges include representation (how to jointly encode heterogeneous modalities), alignment (finding correspondences between modality elements), fusion (combining modality-specific representations for joint inference), and co-learning (transferring knowledge between modalities) [13]. Multimodal fusion has demonstrated consistent improvements over unimodal approaches across diverse domains, as summarized in Table I.

Despite these successes, the application of MML in telecommunications remains highly unexplored. Related work on multimodal time series analysis (e.g., spiking neural networks achieving 98.75% accuracy on biological data [19]), multimodal graph learning (e.g., TimeGNN with 4–80$\times$ faster inference [20]), and text-guided time series forecasting (e.g., 80% MSE improvement [21], modality-aware transformers [22]) demonstrate the potential of these approaches. Foundation models adapted for time series data analysis (TSFMs) represent yet another promising direction, with most current TSFMs being developed for a single modality. Developing multimodal TSFMs for the telecommunication domain offers an opportunity to bridge this gap. None of the aforementioned approaches have been applied to the telecom domain, constituting a significant research gap that this PhD aims to address.

III. METHODOLOGY

This work follows a design-science approach focused on iterative development and evaluation of artifacts. Two categories of artifacts are developed: (1) the MDAM family – concrete analytical approaches (MML, MGL, TSFM-based) fusing heterogeneous network data for predictive and prescriptive inference, and (2) MACNSA – a module hosting MDAM methods against real operator data. The development follows prototype-and-evaluate cycles using real Nokia network data, where each cycle produces an incremental MDAM integrated into MACNSA, evaluated and refined based on established metrics.

A. Multimodal data sources

The 5G+ network environment generates diverse multimodal data. Numerical data includes counters and KPIs organized into traffic condition indicators (e.g., number of user equipment, data volume, physical resource block utilization) and environmental condition indicators (e.g., channel quality indicator, block error rate). Non-numerical data includes feature documentation consisting of natural language texts, tables, diagrams, and images. Each new feature introduction is accompanied by documentation discussing: (a) the functionality being introduced, (b) the expected impact on network performance, (c) interdependencies with other features, and (d) configuration settings and activation procedures. The classical approach to assessing the impact of feature deployment relies on a domain expert whose knowledge is supported by this multimodal documentation – a process that is time-consuming and does not scale.

B. System architecture

Based on the literature review and conceptual studies, MACNSA has been developed as a modular platform for hosting MDAM methods. Its architecture, presented in Figure 1, supports multiple MDAM analytical paths: MGL, MML, and TSFMs, utilizing different fusion techniques for both predictive and prescriptive inference. The modular design enables incremental development, where each MDAM can be developed, evaluated, and integrated independently. The input layer accommodates both structured data (KPI time series) and unstructured data (documentation, logs), while the output layer provides both predictive assessments (e.g., expected KPI impact of a feature deployment) and prescriptive recommendations (e.g., suggested parameter configurations).

C. Current MDAM implementation: Agentic-RAG-MDAM

A data collection and preprocessing pipeline has been established, enabling the integration of multimodal data sources.

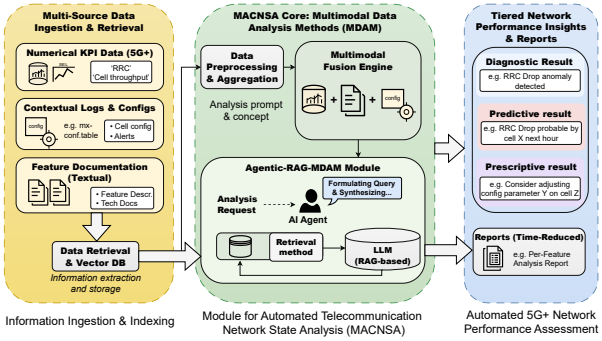

Fig. 1. Architecture of the MACNSA module.

The first MDAM developed within MACNSA – hereafter referred to as Agentic-RAG-MDAM – is an agentic system built using Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) that combines textual and numerical data for feature deployment assessment – a task classically performed by human domain experts requiring approximately 4 to 20 hours of manual analysis per feature. The numerical input consists of feature activation gain (or loss) results collected from the live network and aggregated at multiple granularity levels, enabling both global and fine-grained analysis. The second input is technical documentation related to the analyzed feature (functionality, expected impact, and configuration constraints); non-textual artifacts are converted into textual form during preprocessing to support unified retrieval. These two modalities are fused within the RAG pipeline, where documentation is retrieved using a vector-based mechanism and the LLM generates assessments grounded in both retrieved evidence and observed KPI trends. Viewed through the MML taxonomy of Section II, the Agentic-RAG-MDAM performs late fusion: each modality is first encoded independently (numerical KPIs as structured summaries, documentation as vector embeddings), and the LLM serves as the fusion function that jointly reasons over both representations to produce a unified assessment.

The system is "agentic" in the sense that it employs an autonomous reasoning loop: it decomposes the assessment task into sub-queries, decides which retrieval actions to execute at each step, and iteratively refines the analysis through multiple retrieval-generation cycles until a convergence criterion is met. This multi-step orchestration, implemented via a tool-augmented LLM agent, allows the system to progressively build a more complete picture of the feature deployment impact rather than relying on a single retrieval-and-generate pass. The system is designed to handle the inherent noise and incompleteness of real-world network data, where KPI measurements may contain gaps and documentation may be partially outdated or inconsistent. Agentic-RAG-MDAM primarily addresses RQ 1 and RQ 3, with ongoing work on alignment strategies (RQ 2). Future MDAM iterations within MACNSA will explore adapter-based approaches for tighter LLM-based fusion, as well as MGL and TSFM-based analytical paths shown in Figure 1.

IV. Evaluation

A. Protocol and metrics

The evaluation is conducted on data from a real operating network, providing a realistic context for testing the proposed MDAM. The evaluation protocol compares the multimodal approach against: (a) traditional human expert analysis, and (b) unimodal models relying solely on numerical data. This dual comparison enables assessment of both the practical utility (time savings and quality compared to human experts) and the methodological contribution (improvement over unimodal baselines) of the proposed MDAM.

For the current MDAM research two categories of metrics are employed. Traditional information retrieval metrics include hit rate and normalized discounted cumulative gain (nDCG), which measure the system's ability to retrieve relevant documentation and produce accurate assessments. LLM-as-a-judge metrics, implemented using the deepeval framework [23], include Faithfulness (whether the generated assessment is grounded in the retrieved context), Answer Relevance (whether the output addresses the query), and Context Relevance (whether the retrieved documents are pertinent to the question). These metrics are particularly suited for evaluating the quality of the RAG-based system outputs, as they capture dimensions beyond simple accuracy.

B. Preliminary findings

Experiments performed on real operator data show that Agentic-RAG-MDAM reduces per-feature analysis time from 4 to 20 hours (human expert) to under 3 minutes, while producing assessments that incorporate both quantitative KPI analysis and contextual documentation. With respect to the research questions, results provide initial evidence for RQ 1 (multimodal inference achieves operationally useful accuracy), RQ 3 (textual context improves assessment quality over numerical-only baselines), and partial evidence for RQ 2 (the late-fusion alignment via RAG is effective for this task class, though alternative alignment strategies remain to be compared). The system also maintains consistency across repeated analyses, reducing the variability inherent in human expert assessments. A detailed quantitative results are reserved for a forthcoming journal publication.

C. Reproducibility

Due to the sensitive nature of operator data, there are constraints on sharing the datasets. Detailed descriptions of the developed models and evaluation pipelines will be shared publicly to facilitate reproducibility. The evaluation framework is modular, allowing other researchers to substitute their own datasets while reusing the evaluation protocol. Future iterations will include assessment by human domain experts to validate practical applicability.

V. Conclusions and Future Work

This paper presented a PhD research on automating 5G+ network monitoring and management through a family of MDAM methods. The accomplished work includes: (a) a comprehensive literature review identifying research gaps in the application of MML, MGL, and TSFMs to telecommunications, (b) a data collection and preprocessing pipeline for multimodal network data, (c) the MACNSA module hosting the first MDAM – the Agentic-RAG-MDAM fusing numerical KPI data with textual feature documentation, and (d) evaluation on real operator data using both traditional and LLM-based metrics, showing a significant reduction in per-feature analysis time. The research demonstrates that multimodal data fusion is both feasible and beneficial for automating network performance assessment tasks that were previously dependent on scarce human domain expertise.

These outcomes push forward the identified limits by demonstrating the feasibility of integrating multimodal data sources for network performance assessment, addressing the over-reliance on numerically encoded data and the lack of automated monitoring solutions.

Current limitations include: (1) explainability has not yet been integrated into the MDAM (RQ 5, RQ 6), (2) only one MDAM (Agentic-RAG-MDAM) has been evaluated while MGL and TSFM-based MDAM remain to be explored, and (3) human domain expert assessment is pending. Future work will focus on developing additional MDAM including adapter-based approaches for LLM-based fusion (RQ 2), developing quality evaluation functions for domain-specific data properties (RQ 4), incorporating explainable AI techniques to enhance transparency and trustworthiness (RQ 5, RQ 6), and extending MACNSA with additional data modalities such as images and diagrams. Ablation studies utilizing explainability are also planned to better understand the contribution of individual modalities and fusion components. The planned timeline includes completing the explainability integration and comprehensive evaluation within the next phase of the PhD, conducting human domain expert validation, and preparing a journal publication with detailed quantitative results. The final phase will focus on thesis writing and implementation of the production version of MACNSA within Nokia's analytical infrastructure.

VI. Acknowledgments

This work is conducted within the Polish "Implementation Doctorate" program, in collaboration between Wroclaw University of Science and Technology and Nokia. It was financed by the Polish Ministry of Education and Science (Agreement no. DWD/8/0032/2024).

References

[1] N. P. Tran, O. Delgado, B. Jaumard, and F. Bishay, "ML KPI Prediction in 5G and B5G Networks," in *2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE, 2023.

[2] H. Fourati, R. Maaloul, L. Chaari, and M. Jmaiel, "Comprehensive survey on self-organizing cellular network approaches applied to 5G networks," *Computer Networks*, vol. 199, p. 108435, 2021.

[3] R. Shafin, L. Liu, V. Chandrasekhar, H. Chen, J. Reed, and J. C. Zhang, "Artificial intelligence-enabled cellular networks: A critical path to beyond-5G and 6G," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 212–217, 2020.

[4] J. Sifakis, D. Li, H. Huang, Y. Zhang, W. Dang, R. Huang, and Y. Yu, "A reference architecture for autonomous networks: An agent-based approach," *arXiv:2503.12871*, 2025.

[5] M. Liyanage, Q.-V. Pham, K. Dev, S. Bhattacharya, P. K. R. Maddikunta, T. R. Gadekallu, and G. Yenduri, "A survey on Zero touch network and Service Management (ZSM) for 5G and beyond networks," *Journal of Network and Computer Applications*, vol. 203, p. 103362, 2022.

[6] J. Summaira, X. Li, A. M. Shoib, and J. Abdul, "A review on methods and applications in multimodal deep learning," *arXiv:2202.09195*, 2022.

[7] A. Karchmer, "On stronger computational separations between multimodal and unimodal machine learning," *arXiv:2404.02254*, 2024.

[8] N. Zubić, F. Soldá, A. Sulser, and D. Scaramuzza, "Limits of deep learning: Sequence modeling through the lens of complexity theory," *arXiv:2405.16674*, 2025.

[9] M. Z. Esteghamati, B. Bean, H. V. Burton, and M. Z. Naser, "Beyond development: Challenges in deploying machine learning models for structural engineering applications," *arXiv:2404.12544*, 2024.

[10] M. Usama, R. N. Mitra, I. Ilahi, J. Qadir, and M. K. Marina, "Examining machine learning for 5G and beyond through an adversarial lens," *arXiv:2009.02473*, 2020.

[11] T. Senevirathna, V. H. La, S. Marcha, B. Siniarski, M. Liyanage, and S. Wang, "A survey on XAI for 5G and beyond security: Technical aspects, challenges and research directions," *IEEE Communications Surveys & Tutorials*, vol. 27, no. 2, p. 941–973, Apr. 2025.

[12] S. Li and H. Tang, "Multimodal alignment and fusion: A survey," *arXiv:2411.17040*, 2024.

[13] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

[14] M. Person, M. Jensen, A. O. Smith, and H. Gutierrez, "Multimodal fusion object detection system for autonomous vehicles," *Journal of Dynamic Systems, Measurement, and Control*, vol. 141, no. 7, p. 071017, 05 2019.

[15] T. Trzcinski, "Multimodal social media video classification with deep neural networks," in *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2018*, vol. 10808, International Society for Optics and Photonics. SPIE, 2018, p. 108082U.

[16] S.-C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, and M. P. Lungren, "Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines," *npj Digital Medicine*, vol. 3, no. 1, p. 136, 10 2020.

[17] Z. Liu and J. Hui, "Advancing predictive maintenance: a deep learning approach to sensor and event-log data fusion," *Sensor Review*, vol. 44, no. 5, pp. 563–574, 01 2024.

[18] R. Masukawa, S. Yun, S. Jeong, W. Huang, Y. Ni, I. Bryant, N. D. Bastian, and M. Imani, "Packetclip: Multi-modal embedding of network traffic and language for cybersecurity reasoning," *arXiv:2503.03747*, 2025.

[19] C. Liu, Z. Tao, Z. Luo, and C. Liu, "Mtsa-snn: A multi-modal time series analysis model based on spiking neural network," *arXiv:2402.05423*, 2024.

[20] N. Xu, C. Kosma, and M. Vazirgiannis, "Timegnn: Temporal dynamic graph learning for time series forecasting," *arXiv:2307.14680*, 2023.

[21] Z. Xu, Y. Bian, J. Zhong, X. Wen, and Q. Xu, "Beyond trend and periodicity: Guiding time series forecasting with textual cues," *arXiv:2405.13522*, 2024.

[22] H. Emami, X.-H. Dang, Y. Shah, and P. Zerfos, "Modality-aware transformer for financial time series forecasting," *arXiv:2310.01232*, 2024.

[23] J. Ip and K. Vongthongsri, "deepeval," 1 2026. [Online]. Available: https://confident-ai.com