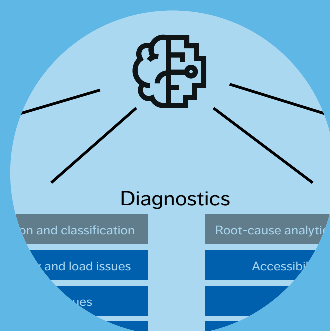
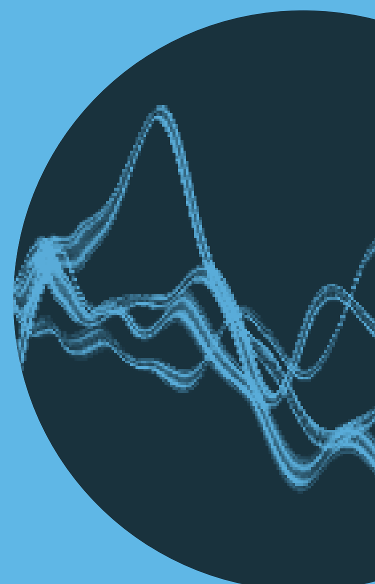
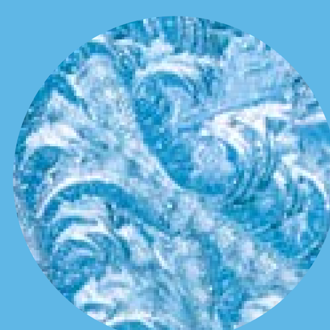
CHARTING THE FUTURE OF INNOVATION | #12·2019

ERICSSON TECHNOLOGY Review





ENHANCING RAN PERFORMANCE WITH AI



ERICSSON

Enhancing RAN performance with AI

Artificial intelligence (AI) has a key role to play in helping operators achieve a high degree of automation, increase network performance and shorten time to market for new features. Our research demonstrates that graph-based frameworks for both network design and network optimization can generate considerable benefits for operators. Even greater benefits can be achieved in the longer term through a comprehensive AI-based RAN redesign.

FRANCESCO DAVIDE CALABRESE, PHILIPP FRANK, EUHANNA GHADIMI, URSULA CHALLITA, PABLO SOLDATI

Advanced 5G use cases and services in areas such as ultra-reliable low latency communications, massive machine-type communications and enhanced mobile broadband place heavy demands on RANs in terms of performance, latency, reliability and efficiency.

The wide variety of network requirements, paired with a growing number of control parameters of modern RANs, has given rise to an overly complex system for which vendors are finding it increasingly difficult to write maintenance, operation and fast-control software. There is a clear need to both simplify the management and provisioning of the different services and improve the performance of the services offered.

The technical objectives of simplification and performance improvement can be roughly mapped to the business objectives of reducing operating and capital expenses respectively, which translate into reduced cost-per-byte for communication service providers and increased QoS for consumers. Embracing AI techniques for the design of cellular systems has the potential to address many challenges in the context of both simplification and

> EMBRACING AI TECHNIQUES FOR THE DESIGN OF CELLULAR SYSTEMS HAS THE POTENTIAL TO ADDRESS MANY CHALLENGES

performance improvement [1], making it possible to achieve new objectives that are beyond the reach of classical optimization and rule-based approaches.

In terms of simplification, AI has already shown the capability to significantly improve functionalities such as anomaly detection, predictive maintenance and the reduction of site interventions through automated site inspections with drones. Performance improvement in the RAN is a greater challenge, as it requires the replacement of classic rule-based network functionalities with their AI-based counterparts. Additional requirements include flexible and programmable data pipelines for data collection and storage; frameworks for the creation (training), execution (inference) and updating of the models; the adoption of graphical processing units for training; and the design of new chipsets for inference.

Three domains for RAN performance improvement

Improving RAN performance involves updating the RAN's control parameters across time, frequency and space to adapt the RAN operation to both static network characteristics, such as the 3D geometry of the surroundings and dynamic network changes in channel, users and traffic distributions. A key prerequisite to successfully apply AI in this context is a deep understanding of the nature and role of different classes of parameters affecting network performance, as well as the complexity of and potential to improve each class.

Artificial intelligence

Artificial intelligence (AI) has experienced an extraordinary renaissance in recent years. The abundance of data and computational capacity that are available today have finally made decades-old techniques like deep learning practically feasible. Substantial investments from both the public and private sectors have fueled the growth of an ecosystem comprised of libraries, platforms, publications and so on that has propelled the field forward and facilitated access to AI techniques for practitioners in various areas.

While the theoretical advances of the AI discipline often occur in domains such as image processing and games, the strengths exhibited by the resulting AI systems – such as the ability to optimize across multiple variables and identify patterns over complex time series – have attracted attention in many industries. In finance, manufacturing and logistics, for example, such capabilities show great potential to improve performance, reduce costs and speed up time to market.

Domain	Parameter type	Network entities	Update frequency
Network design	Deployment parameters	Basebands, cells, RAN configurations, and so on	Monthly/ weekly
Network optimization	Network hyperparameters	Cell clusters/ individual cells	Weekly/ daily/hourly
RAN algorithms	L3 to L1 transmission parameters	Cells and user equipment	Seconds/ milliseconds

Figure 1 Main performance improvement domains

Figure 1 illustrates the main domains for performance improvement that we have identified at Ericsson: network design, network optimization and RAN algorithms. The domains are characterized based on the type of parameters involved, the type and number of network entities and the frequency at which updates typically take place.

Network design domain

The network design domain focuses on improving the parameters that define network deployment – such as the number and location of new cells, the associations of cells to baseband (BB) units, the selection of BB units to form an elastic RAN (E-RAN) configuration, and so on. Network design traditionally relies on planning tools and the domain knowledge of engineers and is performed rather infrequently, such as when new cells are added to an existing network.

Network optimization domain

The network optimization domain focuses on tuning network hyperparameters. While the term hyperparameter has been strongly associated with machine learning in recent years, it generally refers to any parameter used to control the behavior of an underlying algorithm. The hyperparameters of the algorithm are tuned to produce, for the same measured input, a different output that is more appropriate for the given scenario.

While network hyperparameters encompass both the core network and the RAN, our focus here is on RAN hyperparameters such as static/semi-static configuration parameters for cells and user equipment as well as the hyperparameters of RAN algorithms.

Network hyperparameters are optimized to slowly adapt the RAN algorithms to different network scenarios and conditions and bring the performance of a certain area of the network (a particular cluster of cells, for example) into a steady state wherein specific key performance indicators (KPIs) are improved. Examples include hyperparameters for self-organizing networks algorithms and L3 algorithms (mobility, load balancing and so on) for coordination algorithms (such as coordinated multi-point (CoMP), multi-connectivity, carrier aggregation (CA) and supplementary uplink), as well as for L1/L2 algorithms (uplink power control, link adaptation, scheduling and the like).

RAN algorithms domain

The RAN algorithms domain focuses on optimizing the L3 to L1 control parameters that directly affect the signal transmitted to/from the user. Examples include handover and connectivity decisions and the allocation to users of resources such as modulation and coding scheme, resource blocks, power and beams. The L3 to L1 algorithms adapt these parameters on a fast timescale, for individual network entities (cells and UEs, for example), to the rapidly changing environment conditions in terms of channel, traffic, user distribution and so on.

Our path toward AI-based RAN optimization

A natural first step toward a wide integration of AI in RAN products for performance enhancement is the adoption of AI-based solutions in the network design and optimization domains. Optimizing the RAN by tuning the network hyperparameters is safer and easier than redesigning the RAN algorithms with AI-based solutions, as it consists of an outer control loop that does not modify the RAN algorithm design itself but only tunes its behavior.

Figure 2 demonstrates how different network hyperparameter values result in different behaviors for the underlying RAN algorithm, which are represented by different shapes. However, the performance improvement achievable by AI-based network optimization remains limited by the underlying design of the RAN algorithms and the frequency at which network hyperparameters c an be adapted, which affects the extent to which the system can be controlled.

At Ericsson, our long-term goal is to create an all-encompassing AI-based framework that spans the full hierarchy of control – that is, not only network design and optimization but also, importantly, AI-based RAN algorithms.

Examples of AI applications in today's networks

Based on our long-standing research in the area of how AI can be used to improve RAN performance, Ericsson has developed powerful AI-based frameworks for network design and network optimization, as well as several other AI-based solutions for specific use cases.

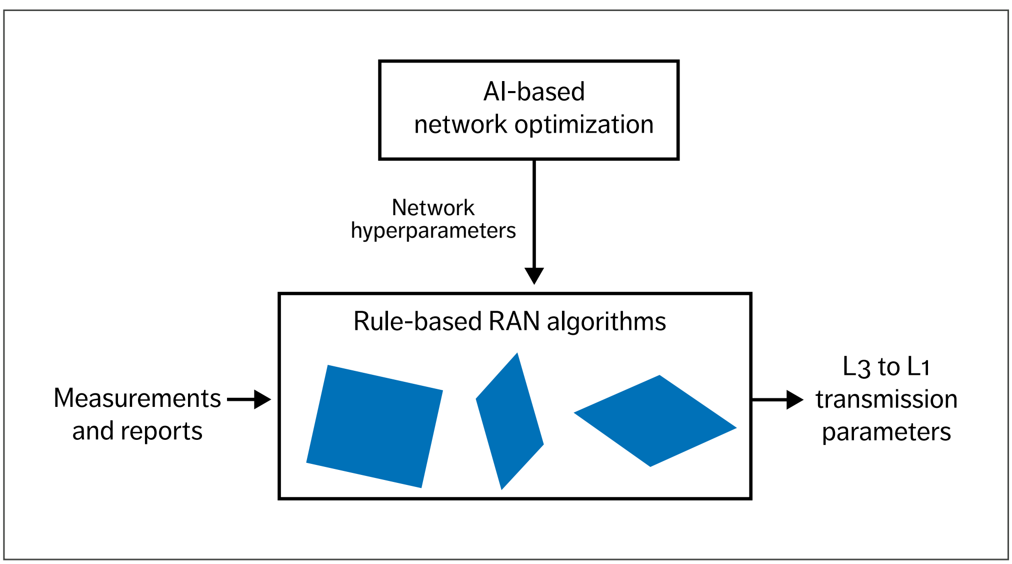


Figure 2 Impact of different hyperparameter values on the behavior of the underlying algorithms

Network design framework

In both 4G and 5G, our centralized RAN (C-RAN) and E-RAN interconnect BB units to allow optimal coordination across the entire network in a centralized, distributed or hybrid network architecture. To ensure that C-RAN and E-RAN performance is in line with customer expectations, a thorough network (re)design is required. In this regard, AI techniques based on advanced network graph methodologies are applied to understand and characterize the complex radio network and its underlying structures, such as the relations between cells and BB units. This approach leads to an optimal design that maximizes consumer throughput through optimized CoMP and CA techniques, and the design is also future-proof in terms of capacity and technology expansions. The design can be split into two main steps.

In the first step, with C-RAN, BB operation is shifted from site location to a centralized BB hub. The C-RAN design therefore focuses on the reconfiguration of the existing distributed RAN architecture to a centralized architecture, where cells are grouped in a centralized hub. This is done in such a way as to create the optimal coordination among cells belonging to the same BB unit, resulting in higher spectrum efficiency and improved consumer experience.

C-RAN configuration design is a highly complex task and difficult to solve using a traditional network design approach. This is because finding an optimal cell grouping that maximizes network performance among a large number of possible cell grouping combinations requires numerous aspects to be considered, such as:

- Intra and inter-frequency cell coverage overlap and neighbor signal strength
- Signal quality and diversity to improve coordination techniques
- Distance between cells
- Frequency band distribution per BB unit
- BB capacity design
- Future cells/sites deployment.

Using an AI-based network graph analysis, natural and hidden structures within cell relations (also known as communities) can be discovered. Based on the various network indicators listed above, the strength of each cell relationship can be measured by a weight factor. The higher the weight factor, the more likely it is that these cells should be grouped together into the same BB unit.

In the second step, E-RAN enables flexible coordination between BB units irrespective of the BB deployment. Similar to the C-RAN design, an AI-based network graph approach can also be applied here to obtain optimal BB clusters consisting of a set of interconnected BB units for borderless coordination across the entire network.

Terms and abbreviations

AI – Artificial Intelligence | **BB** – Baseband | **C-RAN** – Centralized RAN | **CA** – Carrier Aggregation | **CC** – Component Carrier | **CoMP** – Coordinated Multi-Point | **E-RAN** – Elastic RAN | **KPI** –Key Performance Indicator | **L1** – Layer 1 | **L2** – Layer 2 | **L3** – Layer 3 | **RL** – Reinforcement Learning

“E-RAN DESIGN IS ENTIRELY AUTOMATED”

Figure 3 shows the performance improvement in a 4G network operated by an Asian operator for three KPIs after an automated E-RAN redesign. The first bar graph indicates that the connections in CA mode using three component carriers (CCs) increased by 30 percent. The middle bar graph shows that the data volume carried by any secondary cell increased by 22 percent, while the bar graph on the right shows that downlink cell throughput increased by 4.3 percent. However, the most valuable benefit is that the E-RAN design is entirely automated and performed in minutes rather than t he months of work that would be required by human experts.

Network optimization framework

The monitoring and control of network performance is traditionally handled by a team of engineers supported by expert systems targeted at optimizing particular areas of the network (typically a cluster of cells). As such, network performance is often optimized by using a mix of manual and automated rule-based instructions combined with predetermined thresholds for each network performance metric. These rules and thresholds are completely based on human observations and expertise.

However, our solutions demonstrate that it is possible to create a fully scalable and automated closed-loop AI-based solution for network optimization consisting of automated network data processing, network issue identification and classification, detailed root-cause reasoning and automated parameter configuration

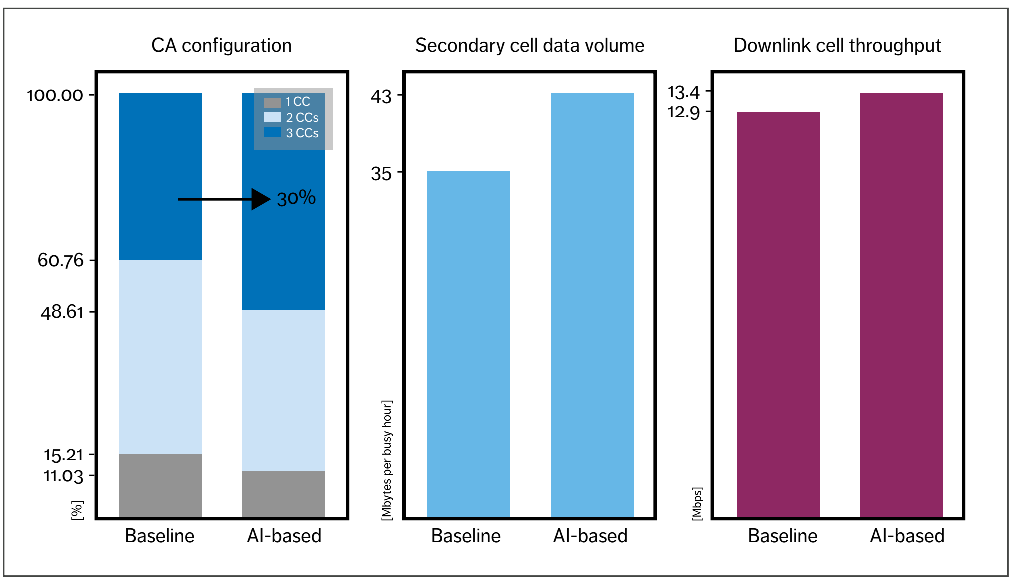


Figure 3 Performance improvement of three KPIs after an automated E-RAN design

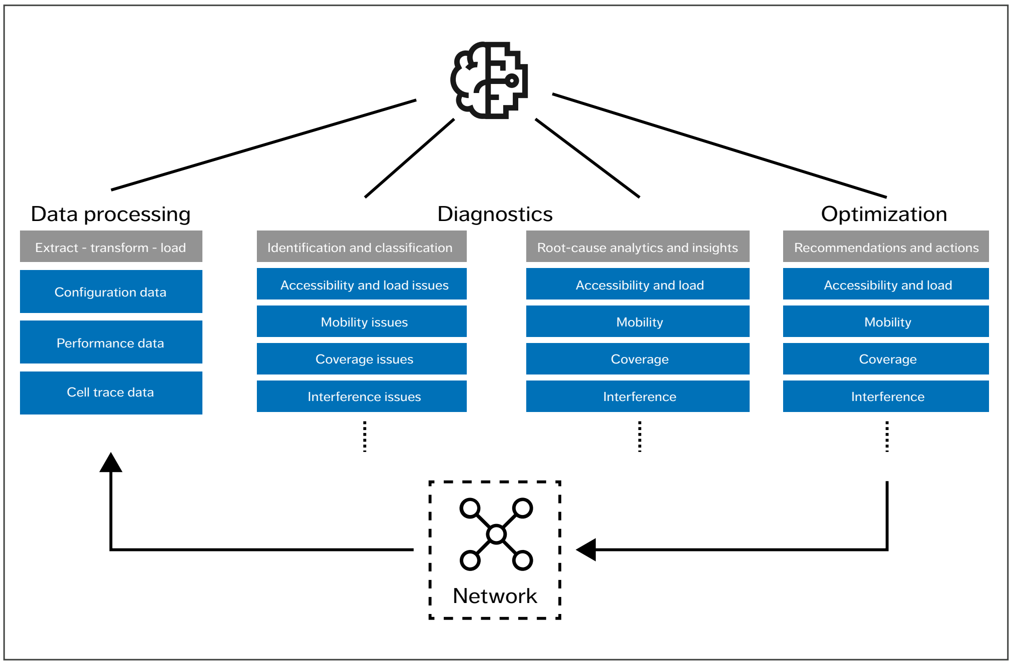


Figure 4 Flow of operations for Ericsson's network optimization framework

recommendations. *Figure 4* illustrates the operations flow for Ericsson's network optimization framework.

State-of-the-art unsupervised and semi-supervised learning techniques combined with expert domain knowledge lead to an efficient annotation of normal and abnormal performance patterns that can be utilized later for issue identification and classification using supervised learning techniques. By integrating network topologies and configurations with hundreds of performance metrics and their two-dimensional correlation in time and space, it is possible to generate a knowledge graph that reveals the specific root causes that lead to an identified network issue.

Closing the automated loop, network parameter changes are automatically suggested to resolve the specific root cause and further improve performance.

AI-based use cases

A non-exhaustive list of AI-based use cases that Ericsson has investigated includes handover [2], link adaptation [3], transmission optimization in C-RAN, interference management, rogue drone detection [4] and federated learning in RAN for privacy awareness [5]. Two of the use cases that are of particular interest in the context of RAN optimization are the prediction of performance on a secondary carrier using primary carrier data [6] and antenna tilting.

> ERICSSON CONTINUES TO INVEST SIGNIFICANT R&D RESOURCES IN THE USE OF AI

Secondary carrier prediction

The use of both high-frequency bands such as 28GHz and higher millimeter-wave bands will continue to increase in 5G radio networks and in future generations. A larger number of bands provides higher capacity but results in larger measurement overhead. For instance, initial deployments on the 28GHz frequency bands will provide spotty coverage. For users to be able to make use of potentially spotty coverage on higher frequencies, the UEs need to perform inter-frequency measurements, which could lead to high measurement overhead. We have used AI techniques to predict coverage on the 28GHz band based on measurements at the serving carrier (for example at 3.5GHz). This approach decreased the measurements on a secondary carrier, thus reducing the energy consumption and the delay for activating features like CA, inter-frequency handover and load balancing.

Antenna tilting

AI-based antenna tilting deserves particular attention among network optimization use cases, as it promises to enhance the coverage and capacity of mobile networks by adjusting base station antennas' electrical tilt based on the dynamics of the network environment. Unlike the conventional antenna tilt approach that follows a rule-based policy, AI techniques enable a self-evolving policy, learning from feedback through network KPIs. Using reinforcement learning (RL), an agent is trained to dynamically control the electrical tilt of multiple base stations jointly so as to improve the signal quality of a cell and reduce the interference on neighboring cells in response to changes in the environment, such as traffic and mobility patterns. This results in an overall improvement of network performance and QoE for the users while reducing operational costs.

Next steps

Ericsson continues to invest significant R&D resources in the use of AI in all three RAN performance improvement domains. We expect to see notable advancements in the network design and network optimization domains in the near term, while at the same time we are increasingly shifting our focus to the critically important RAN algorithms domain.

Network design

In the network design domain, we are currently working to make aspects such as cell-to-BB and BB-to-BB connections software defined. This development would enable the integration of automated AI-based network design in a closed loop, where the network continuously reshapes its graph depending on changing traffic patterns or the addition of new nodes to the network.

Network optimization

In the network optimization domain, our near-term goal is to extend the framework to optimize a larger number of hyperparameters at a higher update frequency. In the mid-term, we aim to integrate these new capabilities into our products and ultimately make them a native part of our product offering.

RAN algorithms

Addressing the optimization of the RAN algorithms domain is vital to our long-term vision of creating an all-encompassing single AI-based controller that spans the full hierarchy of control. The benefit of such a controller would be the inherent capability to optimize multiple transmission parameters across layers simultaneously. The creation of a controller with the ability to learn directly through exploration of the state space would remove the boundaries imposed by human-designed algorithms, making it possible to identify better combinations of transmission parameters within a layer and across layers. Moreover, a controller with the ability to learn from data would inherently be tuned to the environment and be free of network hyperparameters, which would lead to simplification of the software stack.

Nonetheless, replacing L3 to L1 RAN algorithms with a single AI-based controller presents more challenges than network design and optimization, on several levels. One challenge is that fast parameter changes introduce the problem of transients, and therefore require the AI controller to predict the short-term state evolution of the system due to channel and traffic changes, for example, as well as the actions the AI controller itself submits to the system.

Another challenge is the need to redefine the radio access problem in a way that enables learning through interaction with the RAN environment. Today's divide-and-conquer approach for providing radio access to UEs by breaking down the problem into many subproblems of manageable complexity, and designing specific solutions for each subproblem, is difficult to apply when using AI-based controllers. In other words, staying within the current fragmented RAN framework with different AI-based controllers, each trying to optimize a RAN feature while learning through interaction with the same RAN environment, would prevent the system from learning and jeopardize system performance.

One possible approach to address such challenges would be to adopt RL as the framework of choice for RAN control. RL has the necessary capabilities to deal with transients, but it remains challenging to deploy it in the context of the current fragmented RAN framework. To this end, one approach would be to redefine the problem and devise a solution with a single stage of state estimation and a single stage of downstream end-to-end control. This design choice would enable a state estimation as close as possible to the true system state and a controller capable of joint optimization over several transmission parameters.

Additionally, a true AI-based redesign of the system would require end-to-end integration of the different layers of the control hierarchy in such a way that slower (higher) layers of control (such as network design) can make decisions to improve overall system performance as a function of the models learned for the faster (lower) layers of control (such as RAN algorithms).

Conclusion

Interest in artificial intelligence (AI) is growing rapidly in the telecom industry as operators look for ways to automate RAN operations, boost network performance and shorten the time to market for new features. It is important to note, however, that the successful use of AI to optimize the performance of a radio communication network requires a deep understanding both of the nature and role of the different classes of parameters that affect network performance, as well as the complexity and optimization potential of each class.

At Ericsson, our long-term aim is to redefine the overall concept of radio access control with the intent to create a cellular network that constantly adapts itself to the static and dynamic characteristics of the scenarios as well as the requirements of the customers. To help get us there, we have identified three main RAN performance improvement domains based on the type of parameters involved, the type and number of network entities and the frequency at which updates typically take place.

Our work demonstrates that graph-based frameworks for both network design and network optimization can generate considerable benefits in terms of improved performance, simplified management and shorter time to market. Looking further ahead, we expect that the creation of a single AI controller that replaces RAN algorithms will play a key role in a comprehensive AI-based RAN redesign and ultimately make it possible to achieve performance targets that are unreachable in a traditional rule-based design.

References

1. **Ericsson, Employing AI techniques to enhance returns on 5G network investments, 2019, available at:** *https://www.ericsson.com/49b63f/assets/local/networks/offerings/machine-learning-and-ai-aw-screen.pdf*
2. **Cornell University, submitted to IEEE Globecom 2019, ArXiv preprint arXiv:1904.02572, 5G Handover using Reinforcement Learning, Yajnanarayana, V; Rydén, H; Hévizi, L; Jauhari, A; Cirkic, M, available at:** *https://arxiv.org/abs/1904.02572*
3. **In Proceedings of the 2019 Workshop on Network Meets AI & ML (NetAI'19) Pages 44-49, Contextual Multi-Armed Bandits for Link Adaptation in Cellular Networks, 2019, Saxena, V; Jaldén, J; Gonzalez, J E; Bengtsson, M; Tullberg, H; Stoica, I, available at:** *https://www.researchgate.net/publication/335183811_Contextual_Multi-Armed_Bandits_for_Link_Adaptation_in_Cellular_Networks*
4. **Cornell University, ArXiv preprint arXiv:1805.05138, Rogue Drone Detection: A Machine Learning Approach, 2018, Rydén, H; Redhwan, S B; Lin, X, available at:** *https://arxiv.org/abs/1805.05138*
5. **Ericsson Technology Review, Privacy-aware machine learning with low network footprint, October 21, 2019, Vandikas, K; Ickin, S; Dixit, G; Buisman, M; Åkeson, J, available at:** *https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/privacy-aware-machine-learning*
6. **IEEE, In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Predicting strongest cell on secondary carrier using primary carrier data, Ryden, H; Berglund, J; Isaksson, M; Cöster, R; Gunnarsson, F, available at:** *https://ieeexplore.ieee.org/document/8369000*

Further reading

- **Ericsson pioneers machine learning network design for SoftBank, available at:** *https://www.ericsson.com/en/press-releases/2018/5/ericsson-pioneers-machine-learning-network-design-for-softbank*
- **How will AI enable the switch to 5G?, available at:** *https://www.ericsson.com/en/networks/offerings/network-intelligence-and-automation/ai-report*
- **Towards zero-touch networks, available at:** *https://www.ericsson.com/en/ai-and-automation*
- **Ericsson launches unique AI functionality to boost radio access networks, available at:** *https://www.ericsson.com/en/news/2019/10/ericsson-ai-to-boost-ran*
- **AI in 5G networks: Highlights from our latest report, available at:** *https://www.ericsson.com/en/blog/2019/5/ai-in-5g-networks-report-key-highlights*
- **Automated network operations, available at:** *https://www.ericsson.com/en/digital-services/offerings/network-automation*
- **How to connect the dots of future network AI, available at:** *https://www.ericsson.com/en/blog/2019/7/connect-the-dots-of-future-network-AI*
- **An introduction to machine reasoning in networks, available at:** *https://www.ericsson.com/en/blog/2019/11/machine-reasoning-networks-introduction*
- **Supercharging customer experience through AI and automation: The inside view, available at:** *https://www.ericsson.com/assets/local/managed-services/ai-automation-report-screen-aw.pdf*

THE AUTHORS

Francesco Davide Calabrese
◆ joined Ericsson in 2017 as a concepts researcher. In his current role he works on concepts that redefine wireless communications through AI. Prior to joining Ericsson, he worked as a researcher at Nokia and Huawei. He holds a Ph.D. in wireless communication from Aalborg University in Denmark.

Philipp Frank
◆ joined Ericsson in 2014. In his current role, he heads AI development for network design and optimization within Ericsson's Managed Services business area. He holds a Ph.D. in electrical engineering and information technology from the University of Stuttgart, Germany, and a certificate from the executive program AI – Implications for Business Strategy from the Massachusetts Institute of Technology.

Euhanna Ghadimi
◆ joined Ericsson in 2018 where he works with AI concepts for future radio access products in Business Area Networks. Prior to joining Ericsson, he was employed at Huawei as a 5G networks researcher and at Scania, where his work focused on AI solutions for connected vehicles. Ghadimi received a Ph.D. in telecommunications from KTH Royal Institute of Technology in Stockholm, Sweden, in 2015. His research interests are in the areas of optimization theory, machine learning and wireless networks.

Ursula Challita
◆ joined Ericsson Research as a researcher in 2018, the same year she received a Ph.D. in machine learning for radio resource management at the University of Edinburgh, UK. She was a visiting research scholar at Virginia Tech in the US from 2016 to 2018. Her research interests include machine learning, optimization theory and wireless cellular networks.

Pablo Soldati
◆ joined Ericsson in 2018 as a standardization and concepts researcher for 5G New Radio and AI. He received a Ph.D. in wireless communications from KTH Royal Institute of Technology in Stockholm, Sweden, in 2010. He was a postdoctoral scholar at KTH and a visiting postdoctoral scholar at Stanford University before joining Huawei in 2011, where he served as a principal researcher.

ISSN 0014-0171
284 23-3336 | Uen

Ericsson
SE-164 83 Stockholm, Sweden
Phone: +46 10 719 0000