Fjord5G: A Comprehensive 5G Dataset for Coastal Maritime Connectivity

Birkan Denizer[1], Nils Dohse[2], Olaf Landsiedel[1,3]
[1]*Kiel University, Germany*
[2]*ADDIX GmbH, Germany*
[3]*Hamburg University of Technology, Germany*
birkan.denizer@cs.uni-kiel.de, nd@addix.net, olaf.landsiedel@tuhh.de

***Abstract*—In recent years, the use of machine learning (ML) in cellular networking has increased significantly, enabled by the availability of new ML algorithms and cellular datasets. However, existing 5G datasets focus mainly on land-based vehicular networks, which do not capture the unique challenges of the coastal maritime domain, such as large distances from base stations, dynamic sea states, such as waves and tides, and varying interference from water surface reflections and nearby vessels.**

This paper introduces the Fjord5G[1], a 5G dataset for coastal maritime connectivity research. We conduct an extensive measurement campaign aboard research and public ferries in the Kiel Fjord, Germany, collecting GPS-located cellular data along maritime routes. These measurements cover the network conditions encountered in coastal and near-shore regions and provide insights into metrics such as signal strength, modulation, and bandwidth. The resulting dataset includes cellular measurements at a sampling rate of 1 Hz from two mobile network operators, four 5G routers, and two ferries for up to 12 months per router. Initial data analysis reveals key challenges for ML, such as dealing with varying bandwidth and handover events, while highlighting potential features, such as signal strength metrics, that can be exploited to improve coastal maritime connectivity.

***Index Terms*—Dataset, coastal, maritime, LTE, 5G, machine learning, QoS prediction, remote control, autonomous ferry**

Fig. 1: The research vessel MS Wavelab in the Kiel Fjord: For remote and autonomous operation, the ferry features 5G and Starlink connections, cameras, LiDAR, and mmWave sensors.

I. Introduction

As wireless communication technologies evolve toward 6G and beyond, machine learning (ML) enables adaptive and intelligent behavior in network components [1], [2]. Especially in remotely controlled and autonomous systems, these capabilities are critical for maintaining a high quality of service (QoS) under dynamic network conditions. Unlike land-based environments, the coastal maritime domain poses unique communication challenges, such as long distances from base stations, varying waves and tidal movements that affect line-of-sight to base stations, and dynamic interference from water surface reflections and nearby vessels. Combined with varying signal propagation conditions, these challenges often result in sudden bandwidth drops, increased latency due to handovers, and fluctuations in other QoS parameters. Therefore, we need new datasets (1) to train existing ML algorithms to learn the specifics of cellular networking in the near-shore and coastal regions and improve QoS [3]–[7], (2) to provide the basis of new ML algorithms tailored to coastal maritime settings, for example, to manage handovers [8], [9].

In this paper, we introduce the *Fjord5G* dataset, collected during an extensive measurement campaign in the Kiel Fjord, Germany, over 12 months, with more than 240 hours of data collection time over 48 individual ferry rides using research and public vessels, see Figure 1. Our dataset provides GPS-located cellular measurements at a 1 Hz sampling interval along coastal maritime routes and harbor areas, capturing diverse network conditions encountered at sea. The dataset includes cellular measurements from two mobile network operators (MNOs), Vodafone and Deutsche Telekom. It also incorporates data from four 5G routers using two different antenna arrays, making it a comprehensive dataset for maritime wireless communications research. Our measurements allow researchers to investigate how such transitions impact network performance and how ML algorithms can optimize connectivity during handovers. Key features such as signal strength metrics (e.g., CQI, RSRP, and SINR), network deployment characteristics (e.g., LTE, 5G NSA, and 5G SA), a peak uplink bandwidth of more than 300 Mbps, a peak downlink bandwidth of more than 700 Mbps, and handover events provide valuable insights into maritime network performance.

[1]https://github.com/ds-kiel/Fjord5G

This is the author’s accepted manuscript of an article published by IEEE.
The final authenticated version is available online at: https://doi.org/10.1109/VTC2025-Spring65109.2025.11174898
DOI: 10.1109/VTC2025-Spring65109.2025.11174898

Our initial analysis of the dataset reveals key challenges in maritime wireless communications, such as handling intermittent connectivity and predicting QoS parameters in changing environments. The dataset highlights potential features, such as RSRP and SINR, that ML models can exploit to predict QoS, optimize network resources, and manage adaptive resources. Furthermore, the Fjord5G dataset supports transfer learning by providing real-world data, allowing ML models trained in one domain to adapt to other domains.

Overall, this paper makes the following contributions:

1) Introduces the *Fjord5G*, a 5G dataset with GPS-labeled cellular metrics for coastal maritime connectivity.
2) Offers detailed measurements from two mobile network operators (MNOs), four 5G routers, and two vessels, capturing a wide range of network conditions.
3) Demonstrates the utility of the dataset in addressing key challenges for ML tasks in maritime communication environments, such as handling intermittent connectivity.

The remainder of this paper is organized as follows. Section II details the measurement campaign conducted on ferries in the Kiel Fjord. Section III presents an initial analysis of the data set, highlighting the challenges and opportunities. Section IV discusses potential studies that use the Fjord5G dataset. Section V concludes the paper with a discussion of future work and potential uses of the dataset.

II. Related Work

4G Datasets Bokani et al. [10] provide a 4G bandwidth dataset for urban vehicular networks around Sydney, Australia. Raca et al. [11] present a 4G LTE dataset with channel and context metrics, including different mobility patterns, such as static, pedestrian, car, bus, and train around Cork, Ireland. NYU-METS [12] is an LTE mobile bandwidth dataset that covers several transportation scenarios like buses, subways, ferries, cars, and trains in New York City, United States.

5G Datasets 5Gophers [13] is a measurement study of the 5G performance in U.S. cities to analyze the handoff mechanism and predict network performance in urban environments. Raca et al. [14] present a 5G trace dataset with channel and context metrics from two mobility patterns (static and car) and two application patterns (video streaming and file download) in Ireland. Farthofer et al. [15] publish a 5G dataset with channel metrics on a highway in Austria to analyze mobile communications and their use for machine learning. Berlin V2X [16] is a vehicular ML dataset from multiple vehicles using cars in urban environments around Berlin, Germany.

Unlike prior vehicular datasets, Fjord5G provides a comprehensive 5G dataset with channel metrics recorded with research and public ferries for coastal maritime connectivity. We include GPS-labeled measurements at 1 Hz from multiple ferries, mobile virtual network operators, 5G routers, and antennas in the Kiel Fjord, Germany.

III. Data Measurement Campaign

We collect data from research and public ferries operating around the Kiel Fjord along the routes shown in Figure 2.

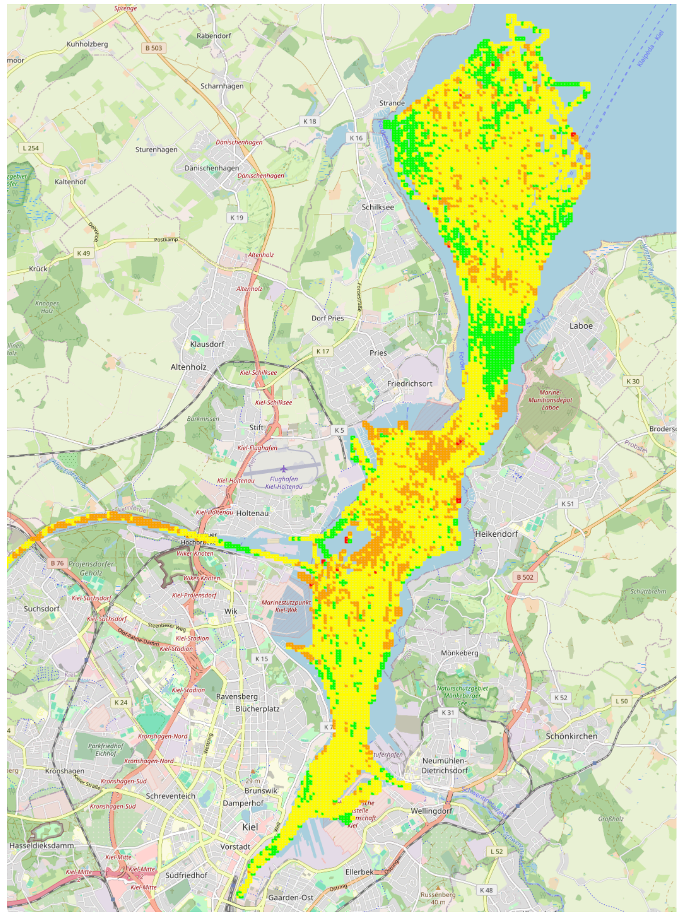


Fig. 2: Measurement region in the Kiel Fjord: We plot the CQI coverage for the 5G router #1 by averaging the measurements in square regions. Regions and values: green ($CQI \geq 11$), yellow ($CQI \geq 7$), orange ($CQI \geq 3$), and red ($CQI \leq 2$).

The Kiel Fjord, originating from the Hörn in the city center, stretches nearly 17 km. It then merges into the Bay of Kiel and connects to the Kiel Canal. Our measurements cover most of the Kiel Fjord, the Bay of Kiel, and parts of the Kiel Canal.

The dataset spans between 5–12 months, depending on the router and ferry, with a sample rate of 1 Hz. The research ferry operates on average once per week, while the public ferry runs practically daily. With the research ferry, the data collection time is over 240 hours over 48 individual rides, with an average of 5 hours per ride. The public ferry runs on fixed routes, on average, for 8 hours a day for 5–7 days per week unless there is maintenance work. We collect data from four 5G routers installed on two ferries with the following setup:

1) Research ferry MS Wavelab: 5G router #1 with external antennas (operator 1)
2) Research ferry MS Wavelab: 5G router #2 with external antennas (operator 2)
3) Research ferry MS Wavelab: 5G router #3 with default antennas (operator 1)
4) Public ferry MS Gaarden: 5G router #4 with default antennas (operator 1)

TABLE I: Overview of captured features

Data Category	Features
Position	Latitude, Longitude, Altitude, Speed, Heading
Operator information	Operator name (e.g., 1 or 2) Cell ID, Enb ID, Sector ID, PHY Cell ID
Network deployment	Data Class (e.g., LTE, 5G NSA, or 5G SA) Primary Band (e.g., n3 at 25Mhz) Carrier Aggregation (CA) Band (e.g., n78, B7) UL CA Band (e.g., B1, B7) (only for LTE)
PHY metrics	CQI (between 0–15) RI (e.g., 1 or 2) MCS (between 0-31) RSSI (dBm), RSRP (dBm) RSRQ (dB), SINR (dB) NR RSRP, NR RSRQ, NR SINR (for 5G NSA)
Modulation	DL Modulation (e.g., QPSK, 64-QAM) NR DL Modulation
Rx/Tx information	Rx/Tx packets per second Rx/Tx bits per second

For our measurement campaign, we use MikroTik Chateau 5G AX[2] routers, which rely on Quectel RG502Q-EA 5G sub-6GHz modules. These modules support 3rd Generation Partnership Project (3GPP) Release 15 specifications, with both standalone (SA) and non-standalone (NSA) modes. The 5G modules use 4×4 MIMO, supporting a theoretical maximum download rate of 5.0 Gbps in 5G NSA and 4.2 Gbps in 5G SA. The theoretical maximum upload rate in 5G SA is 900 Mbps (using 2×2 MIMO), compared to 650 Mbps in 5G NSA. The routers are backward-compatible with 4G and 3G networks.

The 5G routers utilize two types of antenna arrays: The routers #1 and #2 use external marine-grade POYNTING RIPPLE-16[3] MIMO antennas, while the routers #3 and #4 use the default antennas provided by the manufacturer.

We record the network status every second and include GPS position and baseband information, see Table I. We use the public Vodafone and Deutsche Telekom networks across frequency bands B1, B20, B28, B3, B7, B8, n28, n3, and n78, with channels ranging from 5 MHz to 80 MHz. Vodafone operates LTE, 5G NSA, and 5G SA networks, while Deutsche Telekom supports LTE and 5G NSA. We map them as operators 1 and 2, respectively.

Using position information, we can identify regions with lower network connectivity. Operator information allows for the analysis of handover behavior between cell towers. Network deployment details provide insights into accessible technologies, including data class, frequency bands, and channel bandwidth. Physical layer metrics reflect the signal quality received and transmitted by the router. Modulation features indicate the downlink modulation for both primary and secondary bands. We collect received (Rx) and transmitted (Tx) packets and bits per second using the modem interface.

IV. Data Analysis

This section presents a preliminary analysis of the dataset focusing on coverage and the correlation between operators and devices. In Figure 2, we analyze CQI coverage by averaging the measurements from the router #1 in square regions, revealing variations in coverage across different areas. The CQI value is between 0 and 15, with higher values showing better channel quality. We define regions and values as follows: green ($CQI \geq 11$), yellow ($CQI \geq 7$), orange ($CQI \geq 3$), and red ($CQI \leq 2$). We explore the CQI comparison between routers in Figure 3. The routers #1 and #3 exhibit similar CQI distributions, with primarily high CQI metrics. Both routers are on the same ferry and use the same MNO despite having different antenna systems. The router #4 is on a public ferry with a fixed route and shows a uniform CQI distribution.

Next, we analyze the correlation between features from operators 1 and 2 using the routers #1 and #2, as shown in Figure 4. We observe a correlation between physical layer (PHY) features for both operators. For instance, RSSI shows a 58% correlation, RSRP 36%, SINR 30%, and RSRQ 19% between operators 1 and 2. Due to resource constraints, we

[2]https://mikrotik.com/product/chateau_5g_ax

[3]https://poynting.tech/antennas/ripple-16/

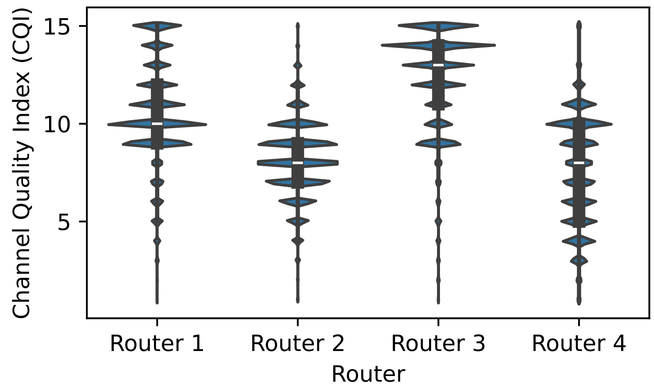


Fig. 3: Comparison of Channel Quality Index (CQI) across routers: The CQI value is between 0 and 15, with higher values showing better channel quality. Routers #1, #3, and #4 use operator 1, while router #2 uses operator 2.

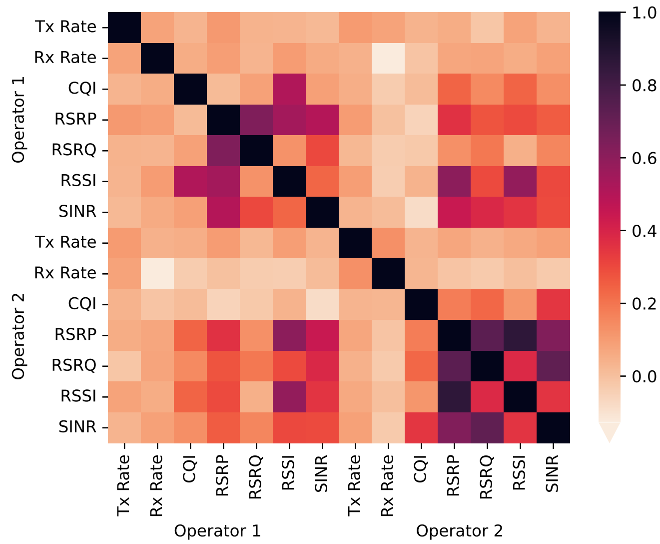


Fig. 4: Operator correlation matrix for 5G routers #1 and #2: We compare the correlation of features between operators 1 and 2 using the same routers and external MIMO antennas.

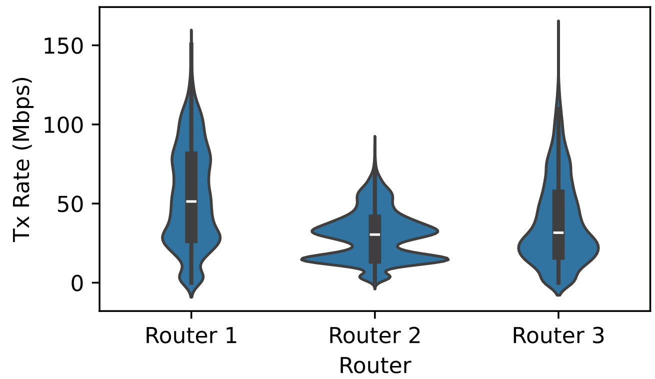


Fig. 5: We compare uplink data rates between routers #1, #2, and #3. Routers #1 and #3 offer higher peak data rates using operator 1 than router #2, which uses operator 2. This measurement takes place on the 28th of June for 270 minutes.

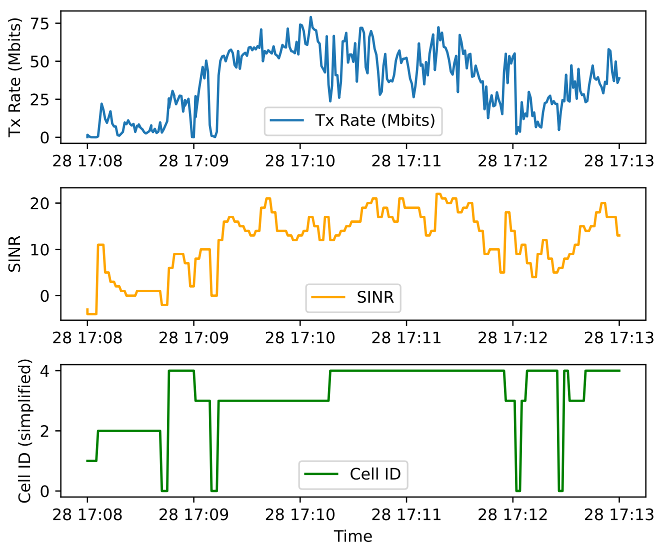


Fig. 6: Analysis of handover events with 5G SA on router #3: Handover events impact SINR, which in turn affects Tx rates.

only perform bandwidth measurements on select occasions. Figure 5 compares uplink bandwidth for the routers #1, #2, and #3 during a ride on June 28th. The routers #1 and #3 show a 33% correlation for NR RSRP and a 27% correlation for NR SINR, and 33% for NR RSRP and 39% for NR SINR to uplink bandwidth, respectively.

In Figure 6, we examine SINR, handover events, and uplink bandwidth. During handover periods, SINR drops, resulting in lower transmission rates. Handovers are inevitable due to mobility, thus significantly impacting achievable data rates.

The SINR exhibits higher variability in 5G NR than LTE, as seen in Figure 8. Higher frequency bands, such as n78, and 5G deployment approaches, such as NSA and SA, further complicate understanding SINR performance. We observe differences in signal quality between the routers #1 and #3, which are on the same MNO but have different antenna systems. The external antennas on the router #1 improve the average SINR,

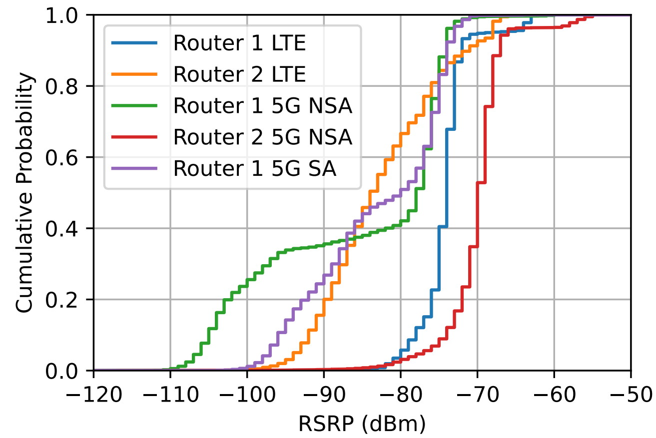


Fig. 7: Comparison of measured RSRP by routers #1 and #2 with the Empirical Cumulative Distribution Function (eCDF): We plot RSRP measurements for LTE, 5G NSA, and 5G SA.

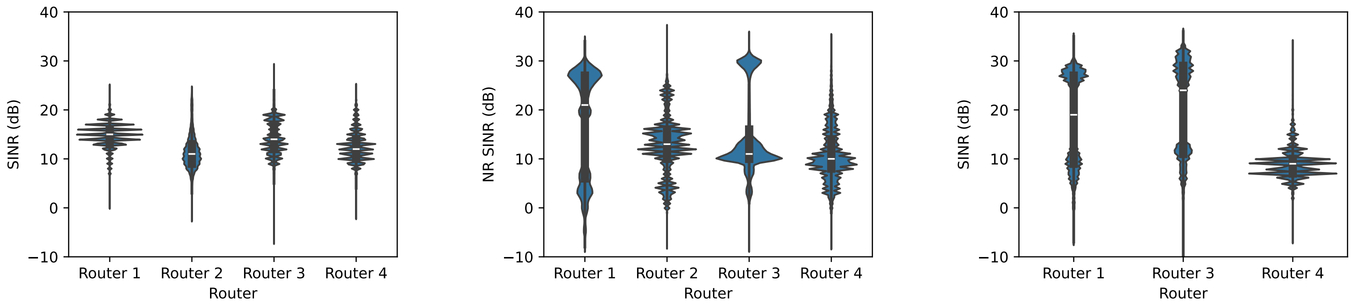
preventing low values in LTE networks compared to the router #3. In 5G NSA networks, the router #1 shows a higher median SINR than the router #3. Both routers exhibit similar SINR distributions in 5G SA networks, with the router #3 slightly outperforming the router #1 in median SINR.

We compare the RSRP measurements using the Empirical Cumulative Distribution Function (eCDF) in Figure 7. For LTE connectivity, routers #1 and #2 exhibit distinct RSRP profiles for operators 1 and 2. This discrepancy is likely due to the location of base stations and the number of users connected to those stations. Since operator 1 also supports 5G SA, there is less reliance on the LTE core for coordinating user allocations. In 5G NR, operator 1 shows similar RSRP profiles for both 5G NSA and 5G SA but with lower average quality than operator 2, which achieves higher average RSRP values.

V. Potential Studies

Firstly, the Fjord5G dataset enables the development of ML-based predictive models to improve QoS. Proactive radio resource management technologies, which use predictions to adapt to upcoming changes in the wireless environment, are crucial for remote-controlled or autonomous systems. For example, bandwidth prediction models provide insights into achievable data rates. Handover prediction models alert the system to potential handovers that may affect data rates. Dual connectivity-aware load-balancing models adjust the load of data transmission between links based on network coverage.

Secondly, the dataset compares cellular network rollouts from two mobile network operators in a maritime coastal area. Identifying similarities between network rollouts enables the development of models that can generalize across operators without the need for additional data collection. The correlations between operators and devices shown in Figure 4 highlight similarities that transfer learning approaches may use. The use of transfer learning enables models trained on one operator or device to apply to others with minimal adaptation.



(a) LTE SINR

(b) 5G NSA NR SINR

(c) 5G SA SINR

Fig. 8: Signal-to-Interface plus Noise Ratio (SINR) ratios for LTE, 5G NSA, and 5G SA networks. The SINR range recorded by the modem is -20 to 30 dB in LTE and -23 to 40 dB in 5G NR modes. Fig. a shows most SINR measurements are between 10 and 20 for LTE. Fig. b and Fig. c show a much higher variability in the SINR measurements for 5G NR than LTE.

Finally, simulation environments modeled and calibrated using real-world data generate more realistic simulation outputs, which allows ML models to be pre-trained before deployment in production systems. Thus, accurate and precise 5G data contribute to faster development cycles for ML-based models.

VI. Conclusion

In this paper, we present Fjord5G, a coastal maritime dataset containing GPS-located cellular measurements at 1 Hz sampling rate from multiple 5G routers, antennas, operators, and ferries. The dataset captures the unique challenges of the coastal maritime domain and allows the training of existing ML models in this domain. It also provides the training data for new ML models tailored to coastal maritime settings.

The Fjord5G dataset enables comparisons between network deployments such as 5G NSA and 5G SA by giving insights into PHY features and achievable data rates. We present a macro view of the correlation between selected features, followed by a detailed discussion of CQI, SINR, and RSRP. We support transfer learning studies with the dataset by providing real-world data to calibrate simulations, generating more realistic pre-training data for ML models.

Acknowledgment

This research has been partially funded by the German Ministry of Transport and Digital Infrastructure within the project CAPTN Förde Areal II (45DTWV08D). ADDIX

References

[1] H. Ye, L. Liang, G. Ye Li, J. Kim, L. Lu, and M. Wu, "Machine learning for vehicular networks: Recent advances and application examples," *IEEE Vehicular Technology Magazine*, vol. 13, no. 2, pp. 94–101, 2018.

[2] D. F. Külzer, M. Kasparick, A. Palaios, R. Sattiraju, O. D. Ramos-Cantor, D. Wieruch, H. Tchouankem, F. Göttsch, P. Geuer, J. Schwardmann, G. Fettweis, H. D. Schotten, and S. Stańczak, "Ai4mobile: Use cases and challenges of ai-based qos prediction for high-mobility scenarios," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, pp. 1–7, 2021.

[3] D. C. Moreira, I. M. Guerreiro, W. Sun, C. C. Cavalcante, and D. A. Sousa, "Qos predictability in v2x communication with machine learning," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pp. 1–5, 2020.

[4] M. Boban, M. Giordani, and M. Zorzi, "Predictive quality of service: The next frontier for fully autonomous systems," *IEEE Network*, vol. 35, no. 6, pp. 104–110, 2021.

[5] L. Torres-Figueroa, H. F. Schepker, and J. Jiru, "Qos evaluation and prediction for c-v2x communication in commercially-deployed lte and mobile edge networks," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pp. 1–7, 2020.

[6] M. Boban, C. Jiao, and M. Gharba, "Measurement-based evaluation of uplink throughput prediction," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, pp. 1–6, 2022.

[7] B. Denizer and O. Landsiedel, "Bandseer: Bandwidth prediction for cellular networks," in *2024 IEEE 49th Conference on Local Computer Networks (LCN)*, pp. 1–8, 2024.

[8] A. Langolf and S. Pachnicke, "Handover prediction for nsa 5g systems in maritime environments using machine learning," in *Mobile Communication - Technologies and Applications; 27th ITG-Symposium*, pp. 74–77, 2023.

[9] A. Langolf, S. Kühl, and S. Pachnicke, "Transfer learning for handover prediction in 5g maritime vehicular networks," in *2024 19th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–5, 2024.

[10] A. Bokani, M. Hassan, S. S. Kanhere, J. Yao, and G. Zhong, "Comprehensive mobile bandwidth traces from vehicular networks," in *Proceedings of the 7th International Conference on Multimedia Systems*, MMSys '16, (New York, NY, USA), Association for Computing Machinery, 2016.

[11] D. Raca, J. J. Quinlan, A. H. Zahran, and C. J. Sreenan, "Beyond throughput: a 4g lte dataset with channel and context metrics," in *Proceedings of the 9th ACM Multimedia Systems Conference*, MMSys '18, (New York, NY, USA), p. 460–465, Association for Computing Machinery, 2018.

[12] L. Mei, R. Hu, H. Cao, Y. Liu, Z. Han, F. Li, and J. Li, "Realtime mobile bandwidth prediction using lstm neural network," in *Passive and Active Measurement* (D. Choffnes and M. Barcellos, eds.), (Cham), pp. 34–47, Springer International Publishing, 2019.

[13] A. Narayanan, E. Ramadan, J. Carpenter, Q. Liu, Y. Liu, F. Qian, and Z.-L. Zhang, "A first look at commercial 5g performance on smartphones," in *Proceedings of The Web Conference 2020*, WWW '20, (New York, NY, USA), p. 894–905, Association for Computing Machinery, 2020.

[14] D. Raca, D. Leahy, C. J. Sreenan, and J. J. Quinlan, "Beyond throughput, the next generation: a 5g dataset with channel and context metrics," in *Proceedings of the 11th ACM Multimedia Systems Conference*, MMSys '20, (New York, NY, USA), p. 303–308, Association for Computing Machinery, 2020.

[15] S. Farthofer, M. Herlich, C. Maier, S. Pochaba, J. Lackner, and P. Dorfinger, "An open mobile communications drive test data set and its use for machine learning," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1688–1701, 2022.

[16] R. Hernangómez, P. Geuer, A. Palaios, D. Schäufele, C. Watermann, K. Taleb-Bouhemadi, M. Parvini, A. Krause, S. Partani, C. Vielhaus, *et al.*, "Berlin v2x: A machine learning dataset from multiple vehicles and radio access technologies," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, pp. 1–5, IEEE, 2023.