International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Data-Driven Intelligence for Real-Time Traffic Prediction and Resource Allocation in Heterogeneous 5G Networks

K. Anusha Reddy[1*], K. Vamshee Krishna[1*], Kadari Shashi Kumar[2], Vuppu Vinay Naidu[2], Mudavath Shivaraj[2], Jakkani Arun Kumar[2]
[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering
[1,2]Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.
*Correspondence: K. Anusha Reddy (anusha.kundanapally@gmail.com), K. Vamshee Krishna (vamshik825@gmail.com)

ABSTRACT

The emergence of fifth-generation (5G) cellular networks represents a transformative advancement in communication technology, offering ultra-low latency, high bandwidth, and the ability to support massive device connectivity. As mobile applications, Internet of Things (IoT) devices, and data-intensive services such as autonomous driving and augmented reality continue to expand, 5G is expected to serve as the foundation of next-generation digital ecosystems. While previous generations of mobile networks improved speed and efficiency, they still face challenges related to spectrum utilization, energy efficiency, and dynamic resource management. Traditional approaches, including static resource allocation and rule-based scheduling, are inadequate in highly dynamic environments, often resulting in network congestion, reduced Quality of Service (QoS), and inefficient spectrum usage. To address these limitations, this study proposes a data-driven framework that integrates machine learning techniques with 5G network analytics. A deep autoencoder is utilized to extract compact and meaningful latent representations from high-dimensional network performance data, effectively reducing noise and redundancy. These latent features are then evaluated using multiple machine learning classifiers, including K-Nearest Neighbours (KNN), Categorical Boosting (CatBoost), and Extreme Gradient Boosting (XGBoost), for baseline comparison. The proposed model, termed Deep Latent-Forest (DLF), combines the autoencoder with a Random Forest classifier to enhance prediction capability. Experimental results demonstrate that the DLF model significantly improves accuracy, robustness, and scalability in detecting dropped connections, making it highly suitable for real-time, intelligent 5G network performance management.

Keywords: machine learning, deep autoencoder, latent representation, Random Forest, Deep Latent-Forest (DLF), K-Nearest Neighbours (KNN), CatBoost, XGBoost, network analytics, Quality of Service (QoS), resource allocation.

1. INTRODUCTION

Fifth-generation (5G) and beyond networks are envisioned to support a diverse range of advanced applications such as Industry 4.0, connected vehicles, and smart cities, all of which demand significantly higher performance and stricter cost efficiency compared to conventional mobile broadband services. To accommodate these diverse and stringent requirements, future networks must adopt flexible and scalable architectures capable of addressing multiple constraints, including performance, security, reliability, availability, and cost optimization. To meet these challenges, network slicing has emerged as a critical solution, widely supported by both academia and industry, for delivering tailored 5G services over a shared physical infrastructure, as illustrated in Figure 1.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

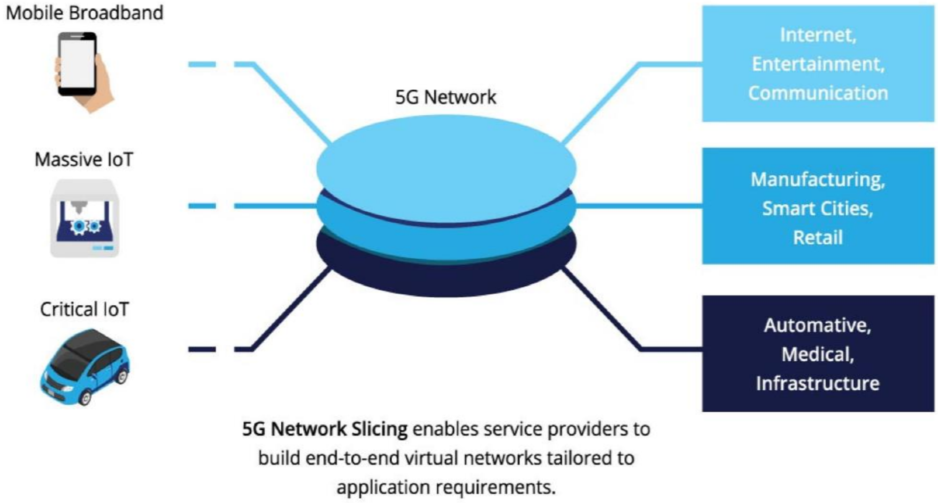

Figure 1: 5G network optimization

In 5G networks, large amounts of data should be analysed before decisions are made to select network slices to ensure that the network can adequately meet QoS requirements. Therefore, machine and deep learning models can be used to analyze large amounts of data and make the most accurate predictions of network slices in 5G networks. Additionally, these models should be optimized in terms of complexity to provide fast decisions for network slicing. The hybridization of different deep learning methods can leverage the complementary strengths of each model and improve generalization, therefore leading to better performance, robustness, or adaptability.

2. LITERATURE SURVEY

Several works have been carried out in the investigation and evaluation of the performance of the existing MBB. In [1], the authors highlight the importance of QoE in cellular networks with various radio access technologies (fourth-generation (4G), 5G and beyond), where they provide the literature on the most advanced measurement methods in QoE. In addition, the QoE is further investigated by different metrics and models for web QoE estimation. The work in [2] analyzed data measurement using several key performance indicators (KPIs) in 4G networks, such as signal quality and download throughput. The drive test considered the actual road traffic conditions at a vehicle speed of 30 km/h. The experimental results demonstrated that the achieved throughput leads to different profiles in terms of time evolution. In [3], the authors investigated the performance of nine MNOs in Europe during times when restrictions were in place due to the COVID-19 outbreak. This investigation included several KPIs such as web QoE, signal coverage, throughput and round-trip-time (RTT). The measurement results showed approximate 46% increases in page load time at different times during the COVID-19 period. The findings indicated that the MNOs had managed their network performance during the pandemic period, although some short-term performance degradations were observable. In [4], the authors presented time and space mapping of outdoor electromagnetic field exposure induced by base station antennas in 4G cellular networks using artificial neural networks. The data were obtained from electromagnetic field exposure sensor networks. In [5], an embedded vehicle-to-everything platform was used to perform drive tests on a public cellular network. The field measurements were conducted based on existing passive network quality indicators and application-level information to forecast uplink transmission power using a novel machine learning technique. Concering indoor environments, the authors of [6] conducted MBB measurements to analyze the network performance in indoor buildings of several areas in Malaysia. The measurement data were collected from three MNOs using mobile smartphones for two types of MBB services: video streaming and web browsing. The

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

measurement results demonstrated that 80% of 4G and 3G networks coverage have good received signal strengths, whereas only 20% reach the threshold level. In addition, 4G networks were the most accessed web and video cycles compared with 3G networks. The authors of [7] conducted multiple measurements (static and dynamic) in live LTE networks in terms of throughput to benchmark Austrian MNOs. The finding showed that the static measurements are not applicable for benchmarking of MNOs due to different small-scale fading patterns, whereas excellent benchmarking can be obtained by using dynamic measurements. According to Jahng and Park [8], it is crucial to obtain an accurate forecast of the potential size of the new mobile market that will capitalise on the potential benefits of 5G technology. In order to better comprehend the development of 5G services, the study presents a customer adoption model based on system dynamics paired with an agent-based model that takes into account 5G adoption estimates under categories with three possible scenarios. One of the most significant findings is that the initial rate of acceptance for 5G is higher than that of 4G. Considering customer preferences and acquisition delay behaviour, Maeng et al. [9] assessed the accuracy of prediction for the 5G service industry. The study found that customer preference and purchase delay behaviour are key to the demand for 5G services, and that consumers have a significant degree of heterogeneity regarding the characteristics of 5G services. As the study also establishes, understanding mobile communication service customers, who are anticipated to be crucial initial users for the creation and diffusion of 5G services, is essential at a time when 5G commercialization is in its infancy.
Maimó et al. [10] present a deep learning anomaly detection method for network flows. The study used a deep belief network (DBN) and a long short-term memory (LSTM)-based anomaly prediction method to analyse the network data in real time. The finding reveals that an alert was generated once a flaw was detected in the network's traffic.

3. PROPOSED SYSTEM

The research proposes an autoencoder-based machine learning framework for intelligent performance monitoring in 5G networks. It preprocesses network data and uses a deep autoencoder to extract compact features from high-dimensional parameters, reducing noise and redundancy. These features are classified using a Random Forest model to accurately predict dropped connections. The approach outperforms traditional methods and is implemented through a secure, user-friendly interface, enabling scalable and real-time 5G network performance management as shown in the figure 2.

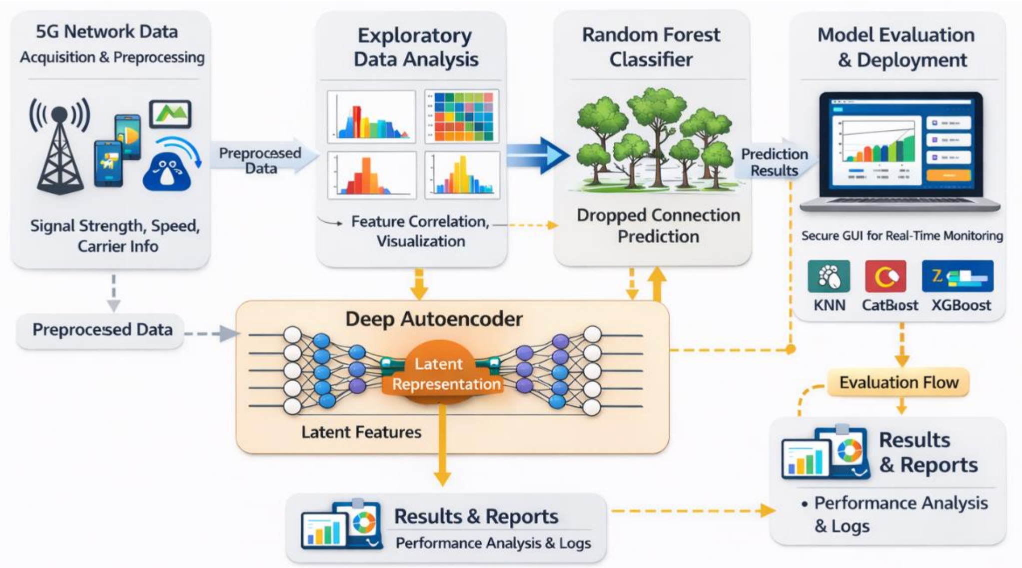

Figure 2: Proposed system architecture

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

1. Intelligent 5G Network Data Acquisition and Preprocessing

The proposed system begins by collecting 5G network performance data containing parameters such as signal strength, download speed, carrier information, and network type. The raw dataset is cleaned by removing irrelevant fields, handling missing values, and converting categorical attributes into numerical form using label encoding. Feature scaling is applied to ensure uniform input distribution for deep learning models. This preprocessing stage ensures consistency, data integrity, and model-ready input.

2. Exploratory Data Analysis for Performance Insight

After preprocessing, exploratory data analysis is performed to understand the relationship between network parameters and dropped connections. Visualizations such as class distribution plots, signal strength density plots, throughput comparisons, and correlation heatmaps are generated. This step provides insight into influential features affecting network performance and validates the suitability of the dataset for predictive modeling. It also helps identify patterns that are difficult to capture using traditional rules.

3. Deep Autoencoder-Based Latent Feature Learning

A deep autoencoder is trained on normalized 5G performance data to learn compact and informative latent representations. The encoder compresses high-dimensional input features into a lower-dimensional bottleneck layer that captures hidden network behavior patterns. This process reduces noise, eliminates redundancy, and preserves essential performance characteristics. The learned latent features form a robust foundation for accurate classification.

4. Hybrid Classification Using Random Forest

The latent features extracted from the autoencoder are used as input to a Random Forest classifier. The ensemble learning mechanism improves generalization and handles non-linear relationships between network conditions and dropped connections. By combining deep feature learning with ensemble decision-making, the model achieves superior prediction accuracy. This hybrid structure forms the core of the proposed DeepLatent-Forest model.

5. Model Evaluation and Comparative Analysis

The proposed system evaluates model performance using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrices. Results are compared against traditional machine learning classifiers including KNN, CatBoost, and XGBoost. This step validates the effectiveness and robustness of the hybrid model under diverse network conditions. All evaluation results are visualized and stored for analysis.

6. Secure Deployment and Real-Time Prediction

A GUI-based deployment framework with role-based authentication is implemented using Tkinter and Redis. Administrators can upload datasets, train models, and analyze performance, while users can perform real-time prediction on unseen 5G data. The system appends predicted dropped connection labels to input datasets for easy interpretation. This step ensures practical usability, security, and scalability of the proposed solution.

4. RESULTS AND DISCUSSIONS

The implementation of the proposed Hybrid Autoencoder Driven Machine Learning system begins with importing required Python libraries and initializing a Tkinter-based GUI with role-based authentication. The 5G network dataset is uploaded, preprocessed through cleaning, encoding, scaling, and exploratory analysis, and then split for model training and evaluation. Baseline machine learning models (KNN, CatBoost, and XGBoost) are trained and compared with the proposed DLF model. The autoencoder learns deep latent features, which are classified using Random Forest to predict dropped connection events. Finally, the system evaluates performance metrics, displays results through the GUI, and supports prediction on new 5G data with result export functionality.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

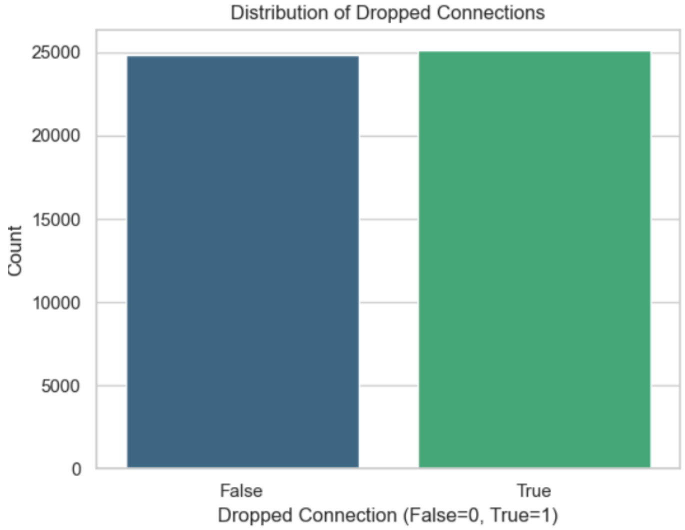

Figure 3: Countplot obtained for target column

The figure 3 shows count plot illustrates the distribution of the target variable Dropped Connection in the 5G dataset. It shows that the number of dropped (True) and non-dropped (False) connection instances is nearly balanced, indicating no significant class imbalance. This balanced distribution is beneficial for training machine learning and deep learning models, as it reduces bias toward any single class and supports reliable performance evaluation.

The figure 4 shows confusion matrix of the KNN classifier that shows the model correctly identifies a large number of non-dropped connections, but it struggles to accurately detect dropped connection cases. A significant number of dropped connections are misclassified as non-dropped, indicating high false negatives. This behavior highlights the limitation of KNN in capturing complex and non-linear patterns present in 5G network data, motivating the use of more advanced and hybrid learning models.

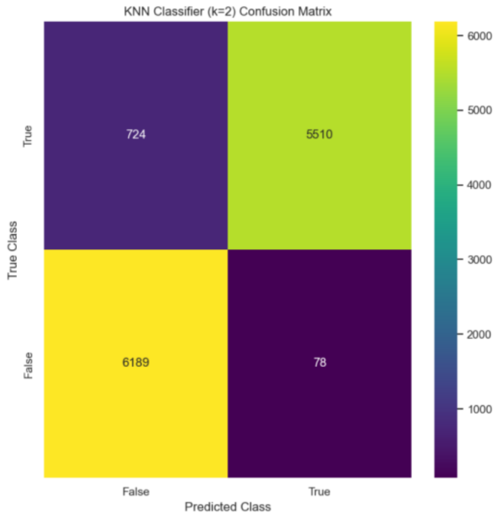

Figure 4. Illustration of confusion matrix using KNN classifier

The figure 5 shows confusion matrix of the proposed DeepLatent-Forest model shows a significant improvement in classification performance compared to all baseline models. The model accurately identifies the majority of dropped connection cases with very few false negatives, while also maintaining high correctness for non-dropped connections. This strong performance demonstrates the effectiveness of autoencoder-based latent feature extraction combined with Random Forest classification in capturing complex and hidden patterns in 5G network performance data, resulting in more reliable dropped connection prediction.

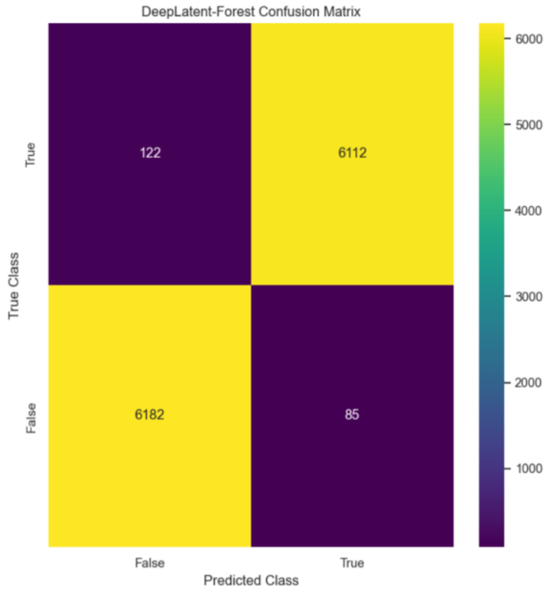

Figure 5: Illustration of confusion matrix using DLF model

Table 1 presents a comprehensive performance comparison of the KNN, CatBoost, XGBoost, and the proposed DLF models for predicting dropped connections in 5G cellular networks. The KNN classifier achieves a reasonably high accuracy of 93.58%, indicating good performance on the dataset; however, its recall and F-score values suggest limitations in consistently identifying all dropped connection cases. The CatBoost model shows comparatively lower performance, with accuracy, precision, recall, and F-score values around 69.6%, indicating difficulty in effectively capturing complex patterns in the 5G network data. XGBoost demonstrates improved performance over CatBoost, achieving an accuracy of 87.15%, reflecting its stronger ability to model non-linear feature interactions, though it still falls short of optimal prediction reliability. In contrast, the proposed DLF model significantly outperforms all baseline models, achieving the highest accuracy, precision, recall, and F-score of 98.34%. This superior performance highlights the effectiveness of combining autoencoder-based latent feature extraction with RFC, enabling the model to capture hidden and complex patterns in 5G network performance data and provide highly reliable dropped connection prediction.

Table 1: Performance comparison for the KNN, Catboost, XGBoost and proposed Deep latent forest

Algorithms Name	Accuracy	Precision	Recall	F-score
KNN model	93.58%	94.06%	93.57%	93.56%
Catboost model	69.61%	69.63%	69.62%	69.61%
XGboost model	87.15%	87.15%	87.15%	87.15%
DLF model	98.34%	98.34%	98.34%	98.34%

The figure 6 shows the row-wise prediction output generated by the proposed DLF model for unseen 5G network data. Each row represents an individual 5G session with complete network, device, and QoS parameters displayed horizontally along with the predicted dropped connection label. The model successfully analyses complex interactions among signal strength, throughput, latency, congestion level, and mobility factors to classify each session as dropped or non-dropped. This detailed row-level prediction demonstrates the effectiveness of the deep latent feature extraction combined with Random Forest classification for accurate and reliable 5G network performance monitoring.

ISSN: 3068-272X www.ijdim.com Original Research Paper

Row-wise Prediction Results:

Row 1 : Timestamp = 09:51.2 | Location = Berlin | Signal Strength (dBm) = 75.4 | Download Speed (Mbps) = 609.09 | Upload Speed (Mbps) = 126.18 | Latency (ms) = 17.8 | Jitter (ms) = 4.09 | Network Type = 5G SA | Device Model = Galaxy S23 | Carrier = Verizon | Band = n78 | Battery Level (%) = 87 | Temperature (°C) = 37.8 | Connected Duration (min) = 34 | Handover Count = 2 | Data Usage (MB) = 120.4 | Video Streaming Quality = 4 | VoNR Enabled = True | Network Congestion Level = Medium | Ping to Google (ms) = 21.9 | Predicted Label = False
Row 2 : Timestamp = 59:51.1 | Location = Mumbai | Signal Strength (dBm) = 79.6 | Download Speed (Mbps) = 149.52 | Upload Speed (Mbps) = 94.3 | Latency (ms) = 4.8 | Jitter (ms) = 4.86 | Network Type = 4G | Device Model = Pixel 7 | Carrier = Jio | Band = n28 | Battery Level (%) = 83 | Temperature (°C) = 23.0 | Connected Duration (min) = 30 | Handover Count = 4 | Data Usage (MB) = 227.42 | Video Streaming Quality = 5 | VoNR Enabled = False | Network Congestion Level = Medium | Ping to Google (ms) = 85.2 | Predicted Label = False
Row 3 : Timestamp = 49:51.2 | Location = Kolkata | Signal Strength (dBm) = 87.6 | Download Speed (Mbps) = 345.67 | Upload Speed (Mbps) = 20.97 | Latency (ms) = 9.7 | Jitter (ms) = 1.16 | Network Type = 4G | Device Model = Nord 4 | Carrier = Jio | Band = n258 | Battery Level (%) = 66 | Temperature (°C) = 28.9 | Connected Duration (min) = 17 | Handover Count = 3 | Data Usage (MB) = 319.69 | Video Streaming Quality = 3 | VoNR Enabled = True | Network Congestion Level = Low | Ping to Google (ms) = 28.4 | Predicted Label = True
Row 4 : Timestamp = 59:51.2 | Location = Berlin | Signal Strength (dBm) = 107.0 | Download Speed (Mbps) = 928.06 | Upload Speed (Mbps) = 107.67 | Latency (ms) = 17.7 | Jitter (ms) = 0.63 | Network Type = 5G SA | Device Model = iPhone 14 | Carrier = Verizon | Band = n260 | Battery Level (%) = 40 | Temperature (°C) = 42.9 | Connected Duration (min) = 21 | Handover Count = 1 | Data Usage (MB) = 258.6 | Video Streaming Quality = 3 | VoNR Enabled = False | Network Congestion Level = Low | Ping to Google (ms) = 77.4 | Predicted Label = False

Figure 6: Prediction on test data using DLF classifier

5. CONCLUSION

By incorporating traditional machine learning models such as KNN, CatBoost, and XGBoost alongside the proposed Deep Latent-Forest (DLF) approach, the system enables a comprehensive comparison of network performance prediction methods. Extensive data preprocessing and exploratory analysis reveal important relationships among factors such as signal strength, throughput, latency, congestion, and mobility that affect connection stability. Experimental findings show that while baseline models can capture basic trends, they struggle with complex and non-linear patterns present in large-scale 5G datasets. In contrast, the proposed DLF model delivers superior performance by utilizing an autoencoder to learn compact and meaningful latent features, combined with a Random Forest Classifier for robust prediction. The system is implemented through a GUI-based platform with role-based access control, improving usability and enabling efficient dataset handling, model training, evaluation, and real-time prediction. High accuracy, precision, recall, and F-score values validate the effectiveness and reliability of the approach. Overall, this research emphasizes the significance of deep latent feature learning in addressing modern 5G challenges and offers a scalable, intelligent solution for proactive network performance monitoring and decision support.

REFERENCES

[1]. Bouraqia, K.; Sabir, E.; Sadik, M.; Ladid, L. Quality of experience for streaming services: Measurements, challenges and insights. IEEE Access 2020, 8, 13341–13361.

[2]. Imoize, A.L.; Orolu, K.; Atayero, A.A.-A. Analysis of key performance indicators of a 4G LTE network based on experimental data obtained from a densely populated smart city. Data Brief 2020, 29, 105304.

[3]. Rajiullah, M.; Khatouni, A.S.; Midoglu, C.; Alay, Ö.; Brunstrom, A.; Griwodz, C. Mobile network performance during the COVID-19 outbreak from a testbed perspective. In Proceedings of the 14th International Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization, London, UK, 25 September 2020.

[4]. Wang, S.; Wiart, J. Sensor-Aided EMF Exposure Assessments in an Urban Environment Using Artificial Neural Networks. Int. J. Environ. Res. Public Health 2022, 17, 3052.

[5]. Falkenberg, R.; Sliwa, B.; Piatkowski, N.; Wietfeld, C. Machine learning based uplink transmission power prediction for LTE and upcoming 5G networks using passive downlink indicators. In Proceedings of the 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), Chicago, IL, USA, 27–30 August 2022.

[6]. Shayea, I.; Rahman, T.A.; Azmi, M.H.; Han, C.T.; Arsad, A. Indoor network signal coverage of mobile telecommunication networks in West Malaysia: Selangor and Johor Bahru. In Proceedings of the 2017 IEEE 13th Malaysia International Conference on Communications (MICC), Johor Bahru, Malaysia, 28–30 November 2022.

[7]. Raida, V.; Svoboda, P.; Rupp, M. On the Inappropriateness of Static Measurements for Benchmarking in Wireless Networks. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2021.

[8]. Maeng, K.; Kim, J.; Shin, J. Demand forecasting for the 5G service market considering consumer preference and purchase delay behavior. Telemat. Inform. 2020, 47, 101327.

[9]. Jahng, J.H.; Park, S.K. Simulation-based prediction for 5G mobile adoption. ICT Express 2020, 6, 109–112.

[10]. Maimó, L.F.; Clemente, F.J.G.; Pérez, M.G.; Pérez, G.M. On the performance of a deep learning-based anomaly detection system for 5G networks. In Proceedings of the 2017 IEEE Smart World, Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing & Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation, San Francisco, CA, USA, 4–8 August 2017; pp. 1–8.