

3-4 マルチモーダル人工知能 (AI) を用いた遠隔制御ロボットにおける受信信号強度予測

3-4 *Multimodal Wireless Link Quality Prediction for Resilient Mobile Robot Operation*

Nguyen Nam Khanh 滝沢 賢一

NGUYEN Nam Khanh and TAKIZAWA Kenichi

人の立ち入りが困難な環境で作業を行う遠隔制御ロボットにとって、無線通信による制御回線は命綱となるため、途絶を防ぐことが重要である。しかし、複雑な構造を持つプラント内等では、ロボット移動に伴い電波強度は時々刻々と変動するため、安心・安全な遠隔制御の実現には、強度低下リスクを事前に把握し、途絶前にロボットを適応的に制御することが求められる。本研究では、作業用ロボットが備える複数センサ (カメラやLiDAR等) で取得したデータを入力として、マルチモーダル人工知能による電波強度予測手法を考案した。プラント模擬環境における2.4 GHz 帯Wi-Fiを対象とした実証実験では、ロボットに実装可能な規模のモデルで、約2.3 dB (二乗平均平方根誤差) の精度で1秒先の受信信号強度を予測できることを確認した。この結果から、提案手法が制御回線の途絶リスクを予測し、ロボットの適応的な制御へ活用できる可能性を示した。

Reliable wireless communication is essential for mobile robots operating in complex environments where signal blockage, multipath, and interference can cause sudden link degradation. This paper describes a multimodal artificial intelligence (AI)-based framework that predicts wireless link quality one second ahead by jointly using camera images, Light Detection and Ranging (LiDAR) point clouds, and 2.4 GHz Wi-Fi received signal strength indicator (RSSI). The experimental results show that multimodal prediction captures RSSI variations more accurately than single-modality approaches, enabling robots to make proactive decisions such as slowing down, stopping, changing routes, or switching access points before communication failure occurs. The framework also supports distributed learning for cooperative robots, where local models can share knowledge without exchanging raw sensor data. Building on this foundation, the research is also being extended toward quality-aware prediction, which considers sensor reliability, and toward multi-band prediction across Local 5G, mmWave, and other wireless bands.

1 Introduction

Japan and many other countries are facing labor shortages caused by declining birthrates and aging populations. Against this background, mobile robots are increasingly expected to support inspection, monitoring, logistics, disaster response, and other tasks in environments that are difficult or dangerous for humans. Many of these robots rely on wireless communication to receive control commands and transmit sensor data, video streams, and status information. Therefore, maintaining a reliable wireless link is essential for safe and continuous robot operation.

Within NICT's research on resilient information and communication technologies, this study aims to develop

robot communication systems that can adapt before wireless conditions become critical. Conventional robot systems often respond only after link quality has already degraded, which may be too late for remotely controlled or networked robots. In practical environments such as disaster sites, factories, tunnels, and infrastructure facilities, radio waves are strongly affected by obstacles, metallic structures, reflection, diffraction, shadowing, multipath propagation, and interference. Multipath propagation occurs when signals reach the receiver through multiple paths, causing rapid fluctuations in received signal strength. Because these effects vary with the robot's position and surrounding environment, conventional channel models may not adequately represent such dynamic and site-specific conditions [1].

The purpose of this research is to enable proactive wireless-aware robot operation by using both wireless and sensing information to predict future link degradation before communication failure occurs. Instead of waiting until the received signal becomes weak, the robot can use the predicted information to prepare an appropriate action in advance. Such actions may include slowing down, stopping safely, changing the route, switching to another access point, or preparing an alternative communication link. This approach is related to “make-before-break” control, in which a new or more reliable communication condition is prepared before the current link becomes unusable. A related proactive handover approach based on camera-derived environmental information has also been investigated in previous work [2].

The present study first focuses on multimodal prediction of 2.4 GHz Wi-Fi received signal strength using camera, LiDAR, and wireless measurements. The prediction model is designed to be sufficiently lightweight for execution on small robots, adaptable to environmental changes, and robust against unreliable or missing sensor data. In the future, this framework will be extended to multi-band wireless systems and integrated with real-time robot and network control. The long-term objective is to establish a common technological foundation that can help robots avoid communication outages in advance and operate more safely and autonomously in a wide range of practical environments.

2 Multimodal Wireless Link Quality Prediction Framework

In this study, we develop a multimodal AI-based wireless link quality prediction framework as shown in Fig. 1. “Multimodal” means that the AI model uses multiple types of information at the same time. The current framework

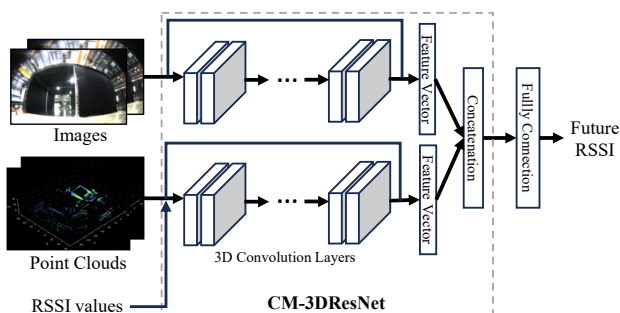


Fig. 1 Proposed prediction framework for wireless link quality prediction.

combines camera images, LiDAR point clouds, and 2.4 GHz Wi-Fi received signal information. Camera and LiDAR data provide spatial information about the surrounding environment, while wireless signal history provides temporal information about how the link quality changes over time. By combining these complementary data sources, the AI model can learn the relationship between the physical environment and future wireless conditions. We propose a customized multimodal three-dimensional residual network (CM-3DResNet) model to predict future wireless link quality by using both wireless measurements and environmental sensing data obtained from a mobile robot [3]. The model is trained in a supervised manner: current and recent multimodal data are used as inputs, and the RSSI measured one second later is used as the teaching signal. This prediction can be used as prior information for proactive robot or network control.

The CM-3DResNet prediction model is a residual neural network architecture, as in Fig. 2, that can learn deep features efficiently by using shortcut connections. In this study, three-dimensional convolution is used to capture both spatial and temporal features from consecutive frames. This is important because wireless link quality is not determined only by the current robot position. It also depends on recent movement, surrounding structures, and short-term signal changes.

Compared with simply applying large general-purpose neural networks, CM-3DResNet is designed specifically for robot wireless prediction. It integrates camera, LiDAR, and wireless-related features while keeping the model relatively compact. This is important for future implementation on edge devices installed on robots, where computation resources, power consumption, and inference time are limited. The original study also compared the proposed model with larger architectures such as 3DResNet-18,

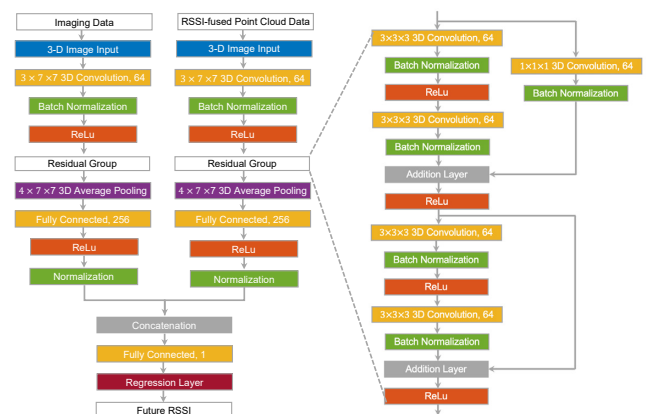


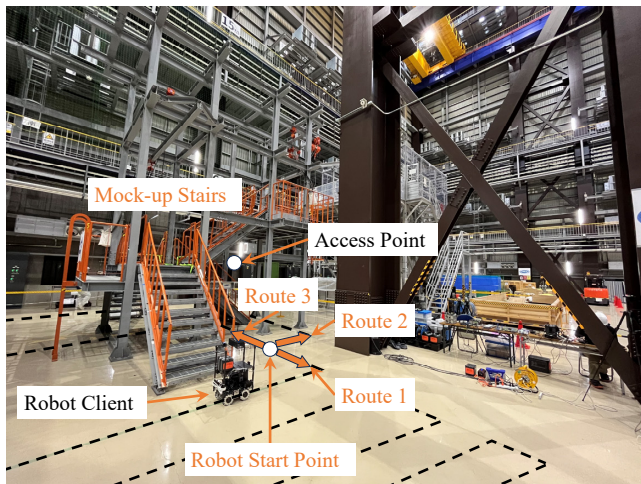
Fig. 2 Architecture of the CM-3DResNet model for multimodal RSSI prediction.

3DResNet-50, and Vision Transformer.

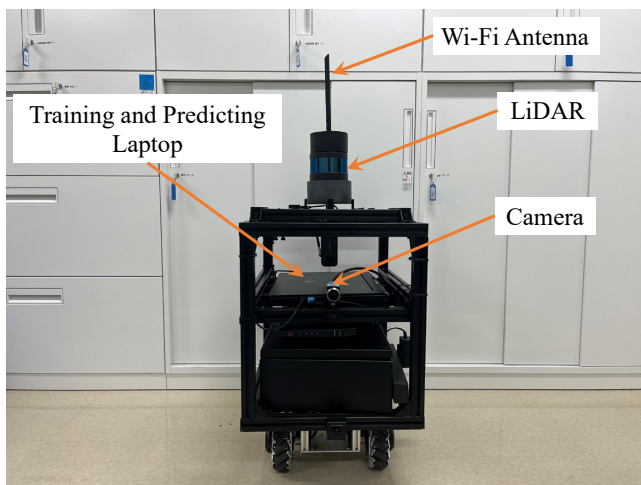
3 Experimental Setup and Data Preparation

To evaluate the proposed framework, an indoor robot experiment, as described in Fig. 3, was conducted in a practical 2.4 GHz Wi-Fi propagation environment. The robot operated in an area containing a complex mock-up stairs structure, which acted as a physical obstacle between the wireless access point and the robot. This structure caused signal attenuation, shadowing, reflection, and multipath propagation. Therefore, the environment was suitable for testing whether multimodal sensing could help predict future wireless link degradation.

Before training, the raw sensor data are preprocessed into suitable forms for deep learning. Camera images



(a) Mock-up stairs experiment setup



(b) Devices on the robot

Fig. 3 Experimental setup for 2.4 GHz Wi-Fi RSSI prediction using a mobile robot equipped with camera, LiDAR, and Wi-Fi antenna.

are resized to reduce computational cost. LiDAR point clouds are converted into voxel grids. A voxel grid is a three-dimensional array that divides space into small cubes, making irregular point cloud data easier to process with neural networks [4]. RSSI outliers were filled by interpolation, and smoothing was applied carefully to the training data to improve prediction stability without excessively distorting the original signal trend. In this framework, RSSI values are fused into the LiDAR voxel data. This design allows the model to include wireless temporal information without adding a separate processing branch for RSSI, thereby keeping the model structure simpler. To train the prediction model, the different data streams must also be synchronized as shown in Fig. 4.

As a result, each training sample consisted of a short sequence of synchronized camera images, RSSI-fused LiDAR voxel grids, and the corresponding future RSSI label. In the reported implementation, four consecutive frames were used as the input sequence, and the target was the RSSI value one second after the last input frame. This dataset design enabled the model to learn both spatial features, such as obstacles and surrounding structures, and temporal features, such as recent changes in wireless signal strength. These features are essential for predicting whether the robot is approaching a weak-signal region and for supporting proactive wireless-aware robot control.

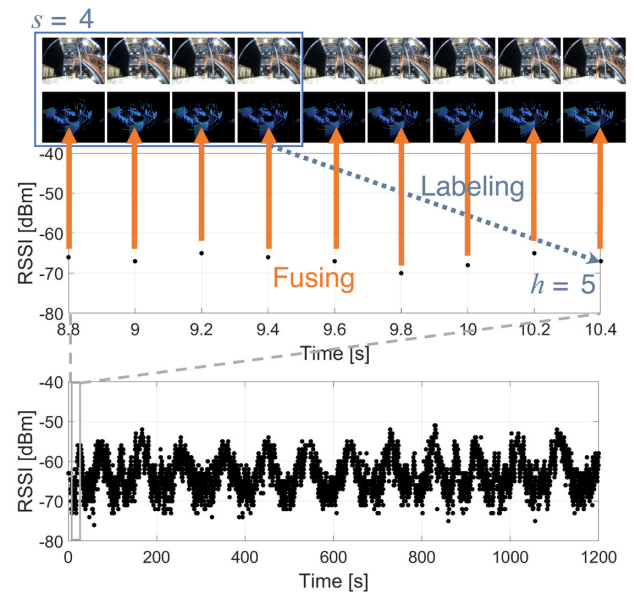


Fig. 4 Example of synchronized multimodal data labeled with future RSSI one second ahead.

4 Prediction Results and Evaluation

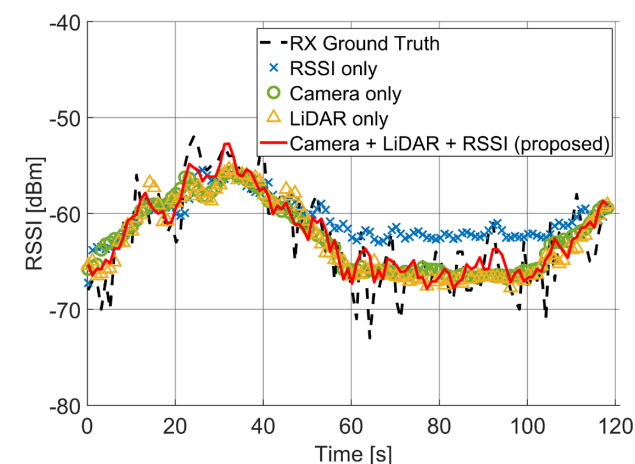
The prediction performance was evaluated by comparing the predicted RSSI with the measured RSSI one second ahead. Several metrics were used: root-mean-squared error (RMSE), accuracy within ± 3 dB, Pearson correlation coefficient, and coefficient of determination. RMSE indicates the average prediction error in dB. Accuracy within ± 3 dB shows the percentage of samples whose prediction error is within a practically acceptable range. The correlation coefficient and coefficient of determination indicate how well the predicted RSSI follows the trend of the measured RSSI.

As shown in Fig. 5, the proposed multimodal model was compared with single-modality models using only RSSI, only camera images, or only LiDAR data. The multimodal model achieved the best overall performance, with an RMSE of 2.34 dB, accuracy within ± 3 dB of 79%, correlation coefficient of 0.88, and coefficient of determination of 0.76.

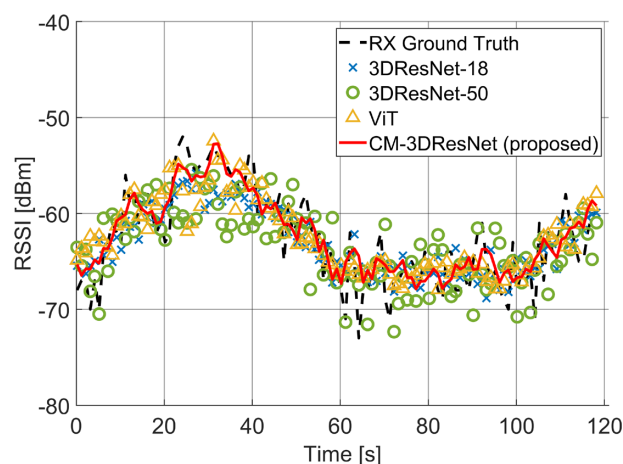
These results show that combining environmental sensing data with wireless signal information is more effective than using only one type of data. In particular, camera and LiDAR data help the model understand surrounding obstacles and structures, while RSSI history helps capture recent wireless signal changes.

The prediction curves show that the proposed multimodal model follows the measured RSSI trend more closely than the single-modality models. This is important for robot operation because the robot does not necessarily need to know every instantaneous signal fluctuation. Rather, it needs to predict the near-future trend of link degradation early enough to take a safe action, such as slowing down, stopping, or avoiding a weak-signal area.

The proposed CM-3DResNet model was also compared with larger neural network architectures, including 3DResNet-18, 3DResNet-50, and Vision Transformer as shown in Fig. 6. Although these large models are widely used in computer vision tasks, they are not always suitable

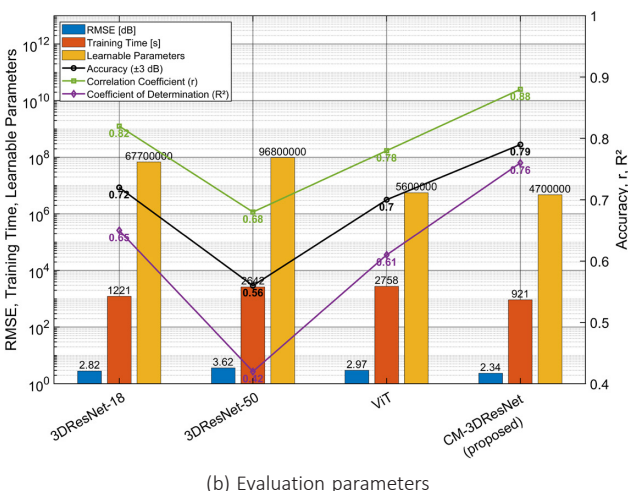


(a) RSSI one-second-ahead predictions and measured values

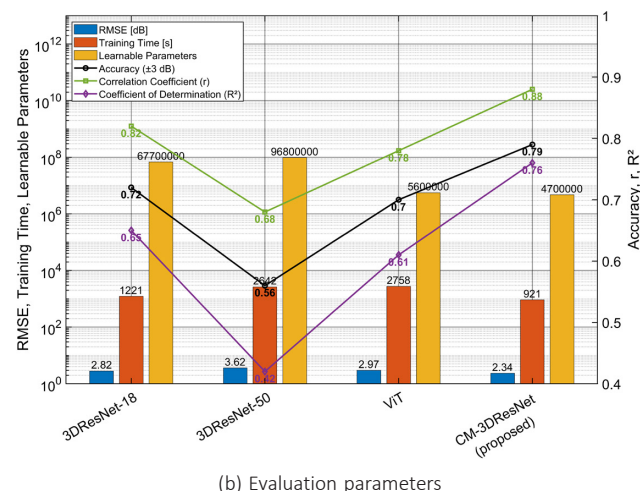


(b) Evaluation parameters

Fig. 5 Comparison of RSSI prediction results between the proposed multimodal model and single-modality models.



(a) RSSI one-second-ahead predictions and measured values



(b) Evaluation parameters

Fig. 6 Comparison of the proposed CM-3DResNet with larger multimodal neural network architectures.

for robot wireless prediction. The proposed CM-3DResNet used only about 4.7 million learnable parameters, while 3DResNet-18 and 3DResNet-50 used about 67.7 million and 96.8 million parameters, respectively. The compact model achieved better prediction accuracy with lower computational complexity, which is important for future edge implementation on mobile robots.

The results suggest that a task-specific compact model is more practical than simply applying a large general-purpose AI model. For mobile robot applications, the model must not only be accurate but also lightweight enough for real-time inference. This is especially important when the prediction module is installed directly on the robot and must work together with sensing, communication, and control functions.

Accordingly, these results demonstrate that multimodal AI can improve future wireless link quality prediction for mobile robots by fusing wireless and environmental information. From the viewpoint of resilient robot operation, the predicted RSSI can be used as early warning information before the robot enters a weak-signal area. This creates a basis for proactive control, such as route adjustment, access point switching, or communication-aware motion planning.

5 Extension Toward Multiple Robots, Quality-Aware and Multi-Band Prediction

The proposed framework can also be extended from a single robot to multiple robots. In practical robot operation, each robot may collect data in a different area or along a different route. Directly gathering all raw camera, LiDAR, and wireless data at a central server can increase communication load and may not be suitable for scalable robot systems. To address this issue, we proposed a federated learning approach [3]. In federated learning, each robot trains a local model using its own data, and only model parameters are shared and aggregated. This makes it possible to improve a common prediction model without exchanging raw sensor data.

This multimodal framework also provides the basis for our quality-aware prediction [5]. In real environments, not all sensor modalities are always equally reliable. Camera data may become less useful under poor lighting, LiDAR data may be affected by occlusion or sparse reflections, and wireless measurements may become unstable in weak-signal regions. A quality-aware extension, as described in Fig. 7, aims to evaluate the reliability of each modality and adaptively use more trustworthy information for prediction.

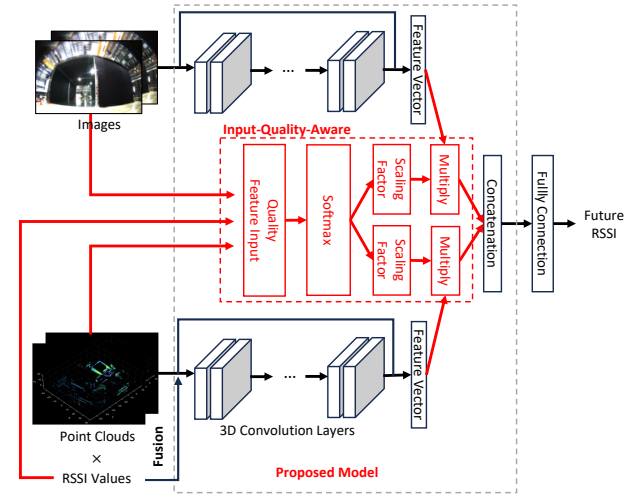


Fig. 7 Concept of quality-aware multimodal wireless link quality prediction.

This direction is expected to improve robustness when the robot operates under changing environmental and wireless conditions.

Another important extension is multi-band prediction. The current validated experiment mainly focuses on 2.4 GHz Wi-Fi, which is widely used and suitable for practical robot experiments. However, future robot systems may use heterogeneous wireless links, such as Wi-Fi, Local 5G, and millimeter-wave communication. These frequency bands have different propagation characteristics. Our previous study has already investigated camera-based received power prediction for 60 GHz millimeter-wave communication [2]. Building on the 2.4 GHz Wi-Fi framework, initial trials have also been conducted for Local 5G at 4.9 GHz [5]. Local 5G is attractive for robot applications because it can provide managed wireless coverage in private, industrial, or experimental environments.

The long-term goal is to develop a multi-band wireless link quality prediction framework that can operate across different wireless systems. Such a framework would learn frequency-dependent propagation characteristics from data and predict the future quality of multiple links. For example, a robot may use Local 5G as the main control link, Wi-Fi as a supplementary link, and millimeter-wave communication for high-capacity data transmission when the line-of-sight condition is favorable. By predicting the future quality of each link, the robot and the network can prepare control actions before degradation occurs. This direction is consistent with the broader concept of AI-native wireless systems. In an AI-native system, AI is not only used as an offline analysis tool, but is integrated into sensing, communication, and control functions. For

mobile robots, this means that environmental sensing, wireless measurement, AI prediction, and proactive control can work together as one system. Quality-aware and multi-band prediction are therefore important next steps toward resilient robot operation in future heterogeneous wireless environments.

6 Conclusion

This research demonstrates the potential of multimodal AI for predicting future wireless link quality and supporting resilient mobile robot operation. By combining camera images, LiDAR point clouds, and 2.4 GHz Wi-Fi signal information, the proposed framework can predict RSSI one second ahead more accurately than single-modality approaches, providing useful prior information for proactive robot and network control. The main validated result is based on 2.4 GHz Wi-Fi, while quality-aware prediction, preliminary Local 5G extension, and previously investigated millimeter-wave prediction indicate a path toward a more general multi-band framework. Future work will focus on improving robustness under unreliable sensor conditions, integrating prediction with real-time robot control, and extending the framework across heterogeneous wireless systems for safer and more adaptive robot operation.

Acknowledgement

This work was supported by the Commissioned Research Fund through F-REI under Grant JPFR23010101.

References

- 1 J. Park, S. Samarakoon, H. Shiri, M. K. Abdel-Aziz, T. Nishio, A. Elgabli, and M. Bennis, "Extreme URLLC: Vision, challenges, and key enablers," arXiv:2001.09683, 2020.
- 2 K. N. Nguyen and K. Takizawa, "Deep learning-based proactive physical layer handover using cameras for indoor environment," Proc. IEEE 21st Consum. Commun. Netw. Conf. (CCNC), Las Vegas, NV, USA, pp.364–367, Jan. 2024. doi: 10.1109/CCNC51664.2024.10454820.
- 3 K. N. Nguyen, K. Takizawa, and T. Nara, "Distributed Multimodal 2.4 GHz Wi-Fi Received Signal Strength Indicator Prediction for Wireless Resilient Robot Operation," IEEE Access, vol.13, pp.171048–171061, Sept. 2025.
- 4 D. Maturana and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Hamburg, Germany, pp.922–928, Sept. 2015. doi: 10.1109/IROS.2015.7353481.
- 5 Nguyen Nam Khanh, 滝沢 賢一, 薄田 一, "ローカル5G信号品質を先読みするフィジカルAI～途切れないロボット遠隔制御に向けて～," ワイヤレスジャパン×ワイヤレス・テクノロジー・パーク (WTP) 2026 特別講演, 東京ビッグサイト, 東京, May 2026.



Nguyen Nam Khanh (ヌエン ナム カーン)

レジリエントICT研究センター
サステナブルICTシステム研究室
主任研究員
Ph.D.

Electromagnetic Wave Scattering, Radio
Wave Propagation, Machine Learning

【受賞歴】

2020年 電子情報通信学会論文賞

2018年 国際学会会議学生論文賞



滝沢 賢一 (たきざわ けんいち)

レジリエントICT研究センター
研究センター長／
サステナブルICTシステム研究室
室長
博士 (工学)

無線通信、信号処理、レジリエントICT基盤技術

【受賞歴】

2021年 電子情報通信学会論文賞

2006年 電子情報通信学会学術奨励賞