**CONTENTS**

[**CONTENTS**](#_TOC_250001)

**LIST OF FIGURES** 11

CHAPTER 1

1. **INTRODUCTION ..………………………………………………….. 12**

1.1 Introduction to 6G and the Importance of Reliable Communication

1.2 Role of Beamforming in 6G Connectivity

The Project’s Focus on Beam Prediction for 6G Networks

1.3 Machine Learning Models for Predictive Beamforming

1.4 The Role of Federated Learning in Privacy and Efficiency

CHAPTER 2

**LITERATURE REVIEW 13-18**

CHAPTER 3

DATASET DESCRIPTION………………………………………………….18-20

CHAPTER 4

SCOPE & PROBLEM STATEMENT ……………………………21

* 1. Challenges in Beam Prediction for 6G Networks
  2. AI-Driven Beam Prediction Using Federated Learning
  3. Objective and Contribution to 6G Communication Systems

CHAPTER 5

RESEARCH CHALLENGES

5.1 HANDLING EXTREME VARIABILITY IN DATA 22

5.2 INTEGRATION OF MULTIMODAL DATA 22

5.3 ADDRESSING BIAS AND ETHICAL CONCERNS 23

5.4 REAL-WORLD DEPLOYMENT CHALLENGES 23

CHAPTER 6

RESEARCH OBJECTIVE

6.1 IDENTIFYING THE MOST EFFECTIVE ALGORITHM 24

6.2 PERFORMANCE EVALUATION USING MULTIPLE METRICS 24

6.3 ENHANCING PRECISION IN BEAM PREDICTION 25

6.4 ADDRESSING REAL-WORLD CHALLENGES………………………………………………………..25

6.5 CONTRIBUTING TO PRACTICAL APPLICATIONS ……………26

CHAPTER 7

METHODOLOGY

7.1 Data Collection and Preprocessing ……………………..26-28

7.2 Feature Selection and Importance Analysis………….28-29

7.3 Machine Leaning models………………………………………29-30

7.4 Federated Learning model………………………...…………30-32

CHAPTER 8

RESULTS & DISCUSSIONS

8.1 Beam Index Distribution……………….………………………. 32

8.2 Correlation MatrixFeature ………………………………… 32-33

8.3 Importance in Machine Learning Model…………………. 34-35

8.3 Model Performance………………………………..…………….. 35-37

8.4 Federated Learning…………………………………………….. 37-38

CHAPTER 10

CONCLUSION………………………………………………………….39-40

**CHAPTER 10**

**FUTURE WORKS………………………………………………………40-42**

CHAPTER 11

**REFERENCES 42-44**

CHAPTER 12

APPENDICES

12.1 Data Preprocessing………………………………………44

12.2 Feature Importance………………………………………45

12.3 Machine Learning Models………………………………45

12.4 Federated Learning Setup………………………………45

12.5 Hyperparameter Tuning………………………………….45

12.6 Model Evaluation………………………………………….45

12.7 Technical Challenges……………………………………45

12.8 Code Snippets…………………………………………………45

12.9 Tools and Libraries Used………………………………..45

12.10 Hardware and Environment…………………………..46

### 

### LIST OF FIGURES

7 METHODOLOGY

7.1 DISTRIBUTION OF BEAM INDEX ………………………………..28

8 RESULTS AND DISCUSSION

8.1 CORRELATION MATRIX …………………………………………33

8.2 FEATURE IMPORTANCE…………………………………………35

8.3 MODEL COMPARISON……………………………………………37

1. **INTRODUCTION**

The advent of 6G networks aims to support unparalleled communication speeds, ultra-reliable connections, and massive device densities, driven by an ever-growing data demand. Achieving such ambitious goals requires overcoming complex challenges in beamforming—an advanced technique that precisely directs communication signals toward specific devices. This project, Federated and Machine Learning-Enhanced Beam Prediction for Reliable 6G Communication, tackles these challenges by introducing a framework that combines traditional machine learning (ML) models with Federated Learning (FL), focusing on accurate and adaptive beam prediction to maintain strong signals and reduce interference.

Beamforming is crucial in 6G, especially in high-frequency mm Wave bands, which are susceptible to attenuation and blockages due to obstacles like buildings and foliage. In these high frequencies, even minor obstructions can disrupt communication, making dynamic and precise beam prediction essential for robust connectivity. This project builds on the capabilities of machine learning models, such as Random Forest, XGBoost, and Deep Learning, to analyze a broad range of features—including mm Wave power, GPS coordinates, satellite data, and environmental variables—that influence beam dynamics. By leveraging these models, the framework effectively captures the complexities of signal propagation and blockages.

However, due to the distributed nature of 6G applications, where data is generated across a vast number of devices, centralizing this data for model training introduces privacy risks and inefficiencies. Transferring raw data from each device to a central server would not only increase the risk of exposing sensitive user information but also introduce significant network overhead, which could strain network resources and slow down the model update cycle.

To address these concerns, this project employs Federated Learning (FL) as an innovative solution to train models in a decentralized manner. FL enables each device to locally process data and compute model updates, which are then aggregated into a global model without sharing raw data. This approach aligns with privacy-preservation goals, as sensitive data remains on the user’s device, minimizing privacy risks and reducing data transfer loads. Furthermore, by leveraging local data patterns unique to each device, the federated learning process enables the model to adaptively learn and generalize across diverse network conditions, such as varying device locations, movement patterns, and obstacles.

Through this federated and machine learning-enhanced beam prediction framework, the project demonstrates a promising pathway for future 6G networks, meeting the requirements of high-speed, low-latency, and privacy-preserving communication. By addressing the complexities of mm Wave signal blockages, environmental variability, and device density, this solution provides a foundation for efficient and resilient beam prediction. As the report progresses, detailed discussions of methods, results, and insights illustrate how this federated approach enhances the adaptability and privacy of beam prediction in 6G environments, paving the way for practical applications in next-generation wireless networks.

1. **Literature Review / Related Work**

* Machine Learning for 6G Wireless Networks: Carry-Forward Enhanced Bandwidth, Massive Access, and Ultrareliable/Low Latency (Du et al., 2020): This paper explores the application of machine learning (ML) techniques in advancing 6G networks, focusing on enhancing bandwidth, massive access, and ultra-reliable/low-latency communication. The authors examine the integration of ML in optimizing network performance, including the use of predictive models and optimization algorithms to improve efficiency. Results indicate that ML can significantly boost network capacity and reliability, addressing challenges such as congestion, high traffic volume, and latency. However, the study acknowledges that current technological limitations hinder the practical implementation of ML-driven solutions, emphasizing the need for further exploration in real-world 6G deployment scenarios. The paper suggests future work that includes the exploration of new ML algorithms tailored for diverse 6G use cases and systems.[1]
* Explainable Artificial Intelligence (XAI) for 6G: Improving Trust Between Human and Machine (Guo, 2019): This paper focuses on the importance of Explainable Artificial Intelligence (XAI) in 6G networks, specifically to improve trust between humans and machines. The author reviews several XAI frameworks and case studies, discussing how explainability in AI models can enhance transparency and accountability in decision-making processes within 6G systems. Results highlight the importance of XAI in ensuring that AI decisions are understandable to users, which is crucial for the widespread adoption of AI in critical applications. The limitation of this paper lies in the lack of empirical validation of XAI frameworks in real-world 6G scenarios. Future work is directed towards conducting more case studies that demonstrate the practical applications of XAI in 6G, bridging the gap between theoretical frameworks and real-world implementations.[2]
* The Role of Blockchain in 6G: Challenges, Opportunities, and Research Directions (Hewa et al., 2020): This paper reviews the potential applications of blockchain technology in the context of 6G networks. The authors examine how blockchain can provide secure, transparent, and decentralized solutions for managing data, transactions, and network operations in 6G systems. The results show that blockchain can address several challenges such as data privacy, network security, and trust, while also offering opportunities for smart contracts, decentralized applications, and secure communication. However, the paper notes the lack of practical blockchain implementations in 6G and emphasizes the need for experimental setups to test blockchain’s effectiveness in real-world 6G environments. Future work focuses on developing prototype applications and testing the integration of blockchain with 6G network infrastructure.[3]
* A Survey on Deep Learning for Ultra-Reliable and Low-Latency Communications Challenges on 6G Wireless Systems (Sahl et al., 2020): This comprehensive literature review surveys the challenges and solutions related to deep learning (DL) techniques in ultra-reliable and low-latency communications for 6G wireless systems. The paper discusses how DL methods, such as neural networks and reinforcement learning, can be utilized to optimize network reliability, latency, and throughput. It also identifies key challenges in applying DL to 6G, including the need for large datasets, real-time processing capabilities, and the high computational cost associated with training deep models. The authors acknowledge that the review is based on existing literature and does not provide empirical evidence of DL performance in 6G environments. Future research is needed to explore the real-world applicability of deep learning in 6G systems and to develop solutions to the challenges identified.[4]
* Quantum Machine Learning for 6G Communication Networks: State-of-the-Art and Vision for the Future (Syed et al., 2019): In this paper, the authors explore the emerging field of quantum machine learning (QML) and its potential applications in 6G communication networks. The paper provides a state-of-the-art review of quantum computing and machine learning, detailing how QML can be used to address challenges in 6G such as massive data processing, optimization, and security. The authors highlight the theoretical advantages of quantum algorithms, including speed and efficiency, but note that practical applications of quantum computing in 6G are still in their infancy. The paper points out that quantum technologies are not yet mature enough for widespread deployment, and experimental setups are needed to validate QML techniques in 6G. Future research will focus on developing practical quantum communication systems and testing them in real-world scenarios.[5]
* Machine Learning in Beyond 5G/6G Networks—State-of-the-Art and Future Trends (Vasileios et al., 2021): This paper provides a review and trend analysis of machine learning applications in Beyond 5G (B5G) and 6G networks. The authors examine how ML techniques can enhance network management, optimization, and service provisioning in future wireless networks, particularly in handling the massive data and connectivity requirements of 6G. The paper discusses various ML approaches, such as supervised learning, unsupervised learning, and reinforcement learning, and their potential for solving challenges like network congestion, resource allocation, and energy efficiency. However, the authors point out that there is limited data on the long-term impacts of ML on network performance, and future work should explore these long-term trends and the broader integration of ML in B5G and 6G systems.[6]
* Real-Time Digital Twins: Vision and Research Directions for 6G and Beyond (Alkhateeb et al., 2023): This paper introduces the concept of real-time digital twins for 6G systems, aiming to create virtual replicas of physical network entities to simulate and optimize their behavior in real-time. The authors propose a conceptual framework for integrating digital twins into 6G networks to enable proactive monitoring, fault detection, and predictive maintenance. While the paper outlines the potential benefits of digital twins, it acknowledges that the framework is still conceptual and lacks practical implementation. Future research should focus on the development of prototype digital twin systems and their integration with 6G infrastructure to test their efficacy in real-world scenarios.[7]
* 6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions (Chowdhury et al., 2019): This comprehensive literature review identifies the key technologies and challenges associated with the development of 6G wireless communication systems. The authors examine the requirements for 6G, including ultra-high-speed data transfer, massive connectivity, and low latency, and discuss the technologies that could enable these features, such as terahertz communication, massive MIMO, and AI-driven network management. The paper also highlights the challenges of 6G, including spectrum management, energy efficiency, and security. While the review provides a broad overview, it is limited by a lack of focus on emerging technologies that could shape the future of 6G. Future research should explore new technologies and their potential integration into 6G systems.[8]
* Radar Aided 6G Beam Prediction: Deep Learning Algorithms and Real-World Demonstration (Demirhan & Alkhateeb, 2021): This paper investigates the use of deep learning algorithms for beam prediction in 6G networks, specifically using radar data to predict beamforming behavior in dynamic environments. The authors demonstrate the effectiveness of deep learning models in improving the accuracy and efficiency of beam prediction, which is critical for optimizing communication links in 6G. However, the study is limited to specific environments, and the results may not generalize to all scenarios. The paper suggests that future work should focus on expanding the scope of beam prediction models to accommodate diverse environments and conditions, thus enhancing their real-world applicability.[9]
* Deep Sense 6G: A Large-Scale Real-World Multi-Modal Sensing and Communication Dataset (Alkhateeb et al., 2021): This paper introduces the Deep Sense 6G dataset, a large-scale multi-modal sensing and communication dataset designed to support research in 6G systems. The dataset includes data from various sensors and communication devices, providing valuable insights into the performance and behavior of future 6G networks. The paper discusses the potential applications of the dataset in areas such as network optimization, beam prediction, and sensing. However, the authors note that the dataset may not cover all possible scenarios, which could limit its utility in certain research contexts. Future work will focus on expanding and annotating the dataset to include a broader range of conditions and use cases.[10]
* Multiple Access Techniques for Intelligent and Multi-Functional 6G: Tutorial, Survey, and Outlook (Clerckx et al., 2024): This paper surveys multiple access techniques for intelligent and multi-functional 6G networks. The authors discuss various techniques, including non-orthogonal multiple access (NOMA), sparse code multiple access (SCMA), and massive MIMO, and their applications in 6G. They provide an overview of the challenges and opportunities associated with these techniques, such as improving spectrum efficiency, reducing latency, and enabling massive connectivity. However, the paper is limited by a lack of real-world case studies, and further empirical research is needed to validate the theoretical models in practical 6G deployments. Future research will focus on conducting more case studies and real-world tests of multiple access techniques.[11]
* Vision-Aided Dynamic Blockage Prediction for 6G Wireless Communication Networks (Charan et al., 2021): This paper addresses the issue of dynamic blockage prediction in 6G wireless networks using vision-aided techniques. The authors propose methods to predict link blockages caused by obstacles in the communication path, which can degrade network performance. The study identifies key factors affecting blockage prediction, such as environmental variables and the use of vision-based sensors like cameras and LiDAR. However, the validation of these techniques is limited to specific environments, and further research is needed to test the models in diverse conditions. Future work will focus on developing more robust prediction methods that can handle a wider range of environmental factors and improve the accuracy of blockage forecasting.[12]
* Proactively Predicting Dynamic 6G Link Blockages Using Lidar and In-Band Signatures (Wu et al., 2023): This paper presents a proactive approach to predicting dynamic link blockages in 6G wireless networks using Lidar and in-band signature techniques. The authors propose a system that can predict blockages before they occur, allowing the network to adapt and maintain reliable communication links. While the paper presents promising methods, the results are limited by the specific environmental conditions in which the techniques were tested. The authors suggest expanding the models to include a broader range of environmental factors, such as weather conditions and urban landscapes, to improve the accuracy and applicability of blockage prediction in diverse 6G settings.[13]
* Machine Learning and the Internet of Things for 6G: A Comprehensive Review (Amin et al., 2023): This paper reviews the integration of machine learning and the Internet of Things (IoT) in 6G networks, exploring the potential applications of ML for optimizing IoT systems in 6G. The authors analyze how ML can be used to address challenges such as data processing, security, and energy efficiency in IoT-enabled 6G networks. They highlight several case studies demonstrating the effectiveness of ML in IoT applications but note that the lack of standardization in ML and IoT integration is a significant barrier. The paper suggests that future work should focus on developing standardized protocols and frameworks for the seamless integration of ML and IoT in 6G.[14]
* Green 6G: Energy-Efficient Wireless Communications Using Machine Learning (Hussain et al., 2024): This paper discusses the role of machine learning in developing energy-efficient wireless communication systems for 6G networks. The authors review various ML techniques that can be applied to reduce energy consumption in 6G, including energy-efficient algorithms, network optimization, and resource allocation strategies. While the paper identifies promising methods, it highlights the lack of real-world implementation studies that demonstrate the practical benefits of these techniques. Future research will focus on testing energy-efficient ML models in real-world 6G environments and developing practical applications for energy-conscious wireless communication.[15]
* Overview of Distributed Machine Learning Techniques for 6G Networks (2022): This paper explores the use of distributed machine learning (DML) techniques in 6G networks, emphasizing how they can address challenges like latency, scalability, and resource constraints in large networks. Unlike centralized ML, which relies on a single server for training, distributed ML allows for model training across multiple devices or edge nodes. This can significantly reduce latency and improve responsiveness. The paper also highlights challenges such as data privacy, model synchronization, and resource management across different network tiers. Future research is focused on optimizing DML algorithms and improving their performance in decentralized environments​.[16]
* Machine Learning in Beyond 5G/6G Networks—State-of-the-Art and Future Trends (2021): In this review, the authors survey the state-of-the-art machine learning techniques used in Beyond 5G (B5G) and 6G networks. The paper identifies critical challenges that ML can address, including ultra-reliable communication, dynamic resource allocation, and interference management in dense network environments. It discusses various types of machine learning—supervised, unsupervised, and reinforcement learning—and their application in improving network performance. The authors also look ahead to future trends, including ML integration with quantum computing, blockchain, and edge computing to make networks smarter and more efficient​.[17]
* White Paper on Machine Learning in 6G Wireless Communication Networks (2020): This white paper provides a broad overview of how ML can revolutionize 6G networks by enabling self-optimization and autonomic network management. By applying predictive analytics, reinforcement learning, and deep learning techniques, networks could become more intelligent, capable of adjusting to changing conditions in real-time without human intervention. The paper highlights challenges such as computational demands, data privacy, and the need for new algorithms that can handle massive amounts of data generated by 6G applications. It also discusses the importance of developing ML models that can operate efficiently in real-world 6G environments, particularly in large-scale deployments​.[18]
* Machine Learning in 6G: The Future of Wireless Communication (2020): This paper outlines the transformative potential of machine learning in 6G networks, specifically in enhancing the performance of ultra-reliable low-latency communications (URLLC) and massive machine-type communications (mMTC). The authors explore the use of ML in managing traffic loads, optimizing energy consumption, and predicting network congestion. Challenges such as real-time processing of big data, ensuring robustness against network failures, and the need for adaptive models that can evolve with changing network conditions are also discussed. The paper concludes with a call for more robust and scalable ML algorithms to meet the complex demands of 6G​.[19]
* AI-Driven Optimization in 6G Communication Systems (2022): This paper focuses on the application of AI and ML to optimize communication systems within 6G. Specifically, it discusses how AI-driven solutions can dynamically manage resources, reduce latency, and optimize network configurations based on real-time data. ML algorithms such as reinforcement learning and deep neural networks are proposed to address resource allocation, interference management, and network slicing in 6G. The paper also highlights potential barriers to AI adoption, such as the need for large training datasets, high computational overhead, and privacy concerns. The authors suggest that AI will play a central role in creating energy-efficient and self-healing networks​.[20]
* Machine Learning for 6G: Advances, Challenges, and Opportunities (2023): This paper provides a comprehensive review of the recent advances in ML techniques for 6G networks, including deep learning, federated learning, and transfer learning. It examines how these techniques can be applied to solve challenges such as optimizing network traffic, predicting and mitigating interference, and enhancing security. Despite the promising results from early experiments, the paper points out that many ML algorithms are still in their infancy and need to be adapted to the unique demands of 6G environments. Future work will focus on developing algorithms that can handle more complex tasks such as real-time decision-making and cross-layer optimization​.[21]
* ML-Based Approaches for Future 6G Systems: A Comprehensive Survey (2022): This survey focuses on the application of ML in addressing key challenges in 6G systems, such as ultra-reliable communications, massive connectivity, and network automation. The authors highlight the potential of ML to enable intelligent beamforming, dynamic spectrum management, and autonomous network management. The survey also discusses the different ML techniques used in 6G, including deep learning, reinforcement learning, and unsupervised learning, and how they can be applied to solve issues such as data traffic prediction and interference management. Challenges such as computational complexity and the need for large-scale simulations are also identified​.[22]
* Integrating Blockchain and ML for 6G Network Security (2021): This paper investigates how blockchain technology can be integrated with machine learning to enhance the security of 6G networks. Blockchain's decentralized and immutable nature can provide secure data transactions and help in the management of identities, while ML algorithms can predict and mitigate potential security threats such as network intrusions and fraud. The authors discuss how these technologies can work together to ensure data privacy, integrity, and availability in 6G. They also highlight the challenges of implementing such a system in real-world networks, including scalability and regulatory concerns​.[23]
* ML-Driven 6G Spectrum Sharing Techniques (2021): The authors of this paper focus on how ML can optimize spectrum sharing in 6G networks, an essential feature for accommodating the massive data traffic expected in future communication systems. By using ML algorithms to predict spectrum demand and adjust resource allocation accordingly, 6G networks can achieve more efficient spectrum utilization and reduce interference. The paper also discusses various ML models that can be used for spectrum sensing and management, such as reinforcement learning and neural networks, and explores the challenges of implementing these techniques in a real-world 6G environment.[24]
* Towards Autonomous 6G Networks: The Role of ML in Self-Organizing Systems (2022): This paper explores the concept of self-organizing networks (SON) in 6G, where machine learning algorithms autonomously manage and optimize the network. By using ML techniques, such as reinforcement learning, the network can learn to adjust its parameters, allocate resources dynamically, and handle network failures without human intervention. The paper also discusses how SON can be applied to 6G applications like autonomous vehicles and smart cities. While the potential benefits are significant, challenges such as ensuring data privacy, scalability, and real-time performance are highlighted. [25]

1. **DATASET DESCRIPTION**

This dataset includes a rich set of features for analyzing and predicting beam dynamics. Many of these columns (e.g., GPS coordinates, radar, LiDAR, and power data) are essential for understanding the physical and environmental context around beamforming, enabling accurate and adaptive communication strategies in a 6G network setting.

* index: This is a unique identifier for each data record. It acts as a primary key to differentiate between individual observations and is especially useful for tracking and referencing specific data entries in the analysis.
* unit1\_rgb: Represents RGB (Red, Green, Blue) image data captured by Unit 1, likely from an onboard camera. This image data can provide a visual context of the surroundings and may help in detecting obstacles, environmental conditions, or the presence of other objects that could impact beam prediction and communication quality.
* unit1\_pwr\_60ghz: Denotes the power level of the 60 GHz communication signal from Unit 1. Higher power levels might improve signal reach and quality, whereas low power could indicate weaker connectivity. Monitoring power levels at this high frequency is critical, as 60 GHz signals are highly susceptible to interference and attenuation.
* unit1\_loc: Contains GPS or location data for Unit 1, which may include latitude and longitude coordinates. Knowing the precise position of Unit 1 is essential for targeted beamforming and ensuring that signals are directed accurately toward the intended receiver.
* unit1\_lidar: LiDAR (Light Detection and Ranging) data from Unit 1, providing depth or distance measurements of objects in the vicinity. LiDAR is useful for detecting obstacles, mapping surroundings, and supporting dynamic beamforming by adjusting beam angles based on nearby structures or potential interferences.
* unit1\_lidar\_SCR: Signal-to-Clutter Ratio (SCR) for the LiDAR data from Unit 1. This metric indicates the quality of the LiDAR signal in relation to background noise or "clutter." High SCR values mean a clearer signal, while low SCR values could indicate noisy data, affecting the reliability of obstacle detection and spatial mapping.
* unit1\_radar: Radar data from Unit 1, providing information on detected objects using radio waves. Radar data is less affected by environmental factors compared to LiDAR and can offer complementary spatial information, such as the speed or distance of objects. This helps in predicting possible interferences or movements that might impact beam direction and signal quality.
* unit2\_loc: Location data for Unit 2, which could include GPS coordinates (latitude, longitude) or other positioning details. This is crucial for alignment in communication systems, as beamforming relies on precise positions of both sender and receiver units.
* unit2\_loc\_cal: Calibrated location data for Unit 2. Calibration implies adjustments made to the raw location data to enhance accuracy. This is important for correcting potential GPS inaccuracies or alignment issues, ensuring more accurate beam steering between Unit 1 and Unit 2.
* unit1\_beam\_index: The beam index selected by Unit 1 for communication, likely indicating the specific beam angle or direction in use. Beam indices are essential in managing which direction the signal is aimed, especially in high-frequency bands like 60 GHz, where narrow, focused beams are required to maintain connectivity.
* seq\_index: A sequence index indicating the order of records. It could be used to track the chronological order of data collection, which is particularly useful in time-series analyses or sequential model training to observe trends over time.
* timestamp [UTC]: The timestamp of each data record, expressed in Coordinated Universal Time (UTC). Timestamps allow the system to account for temporal factors, such as time of day, which may impact communication due to environmental changes (e.g., interference from moving objects).
* unit2\_direction: The directional orientation of Unit 2, which could include heading or bearing information. This helps in aligning beams accurately, as knowing the receiver’s direction can improve the efficacy of beamforming by focusing the signal toward the receiver’s trajectory.
* unit2\_num\_sat: The number of satellites visible to Unit 2’s GPS system. The number of satellites impacts the accuracy of GPS-based positioning; more satellites typically result in higher positional accuracy, which is beneficial for precise beamforming.
* unit2\_sat\_used: The number of satellites actually used by Unit 2 for calculating its position. This may differ from the visible satellites if only certain satellites meet quality criteria. Accurate satellite data helps ensure that the position of Unit 2 is reliable for communication purposes.
* unit2\_fix\_type: Type of GPS fix used by Unit 2, indicating the level of positional accuracy (e.g., 2D fix, 3D fix). A 3D fix generally provides more precise position data, which is beneficial for beam alignment in complex network environments.
* unit2\_DGPS: Indicates whether Differential GPS (DGPS) is being used by Unit 2. DGPS improves GPS accuracy by using additional data from ground-based stations. This enhanced accuracy is advantageous for applications requiring precise positioning, like 6G beamforming.
* unit2\_PDOP: Position Dilution of Precision (PDOP) for Unit 2. PDOP measures the geometric accuracy of the 3D GPS position, with lower values indicating better accuracy. Reliable PDOP values ensure accurate positioning, which is essential for effective beam direction.
* unit2\_HDOP: Horizontal Dilution of Precision (HDOP) for Unit 2. HDOP indicates the horizontal accuracy of GPS data. Like PDOP, lower HDOP values signify higher accuracy, important for accurate lateral positioning in beamforming.

1. **SCOPE AND PROBLEM STATEMENT**

With the emergence of 6G networks, wireless communication systems are expected to achieve unprecedented levels of connectivity, speed, and reliability. One of the key challenges in realizing this vision is accurate beam prediction, which is essential for maintaining high-speed, low-latency communication under dynamically changing network conditions. Beams in millimeter-wave (mm Wave) frequencies, a critical component in 6G, require precise prediction and adjustment to handle factors such as changing user positions, interference, environmental obstructions, and high data demands. However, current beam prediction methods encounter several significant limitations, including scalability issues, privacy risks associated with data centralization, and reduced adaptability to diverse network conditions and geographical regions.

This project aims to address these limitations by developing an AI-driven beam prediction model that leverages federated learning. Federated learning allows the model to be trained across decentralized client devices, each containing partial data, without sharing raw data to a central server, thus enhancing data privacy and security. The model leverages a diverse set of input features—including mm Wave power levels, GPS coordinates, satellite data, and additional location-specific parameters—to predict optimal beam parameters that ensure reliable connectivity.

The approach combines traditional machine learning models such as Random Forest and XGBoost with deep learning architectures to explore their efficacy in beam prediction. Furthermore, by integrating federated learning, this solution aims to improve the model’s accuracy and scalability across distributed devices, enhancing the system’s adaptability to real-time changes and enabling personalized model training for each device without compromising privacy. The federated approach also allows the model to generalize across diverse regions and user densities, which is essential for achieving resilient and efficient communication in 6G networks.

In summary, the project seeks to develop a comprehensive beam prediction framework that not only maximizes prediction accuracy but also addresses privacy, scalability, and adaptability challenges inherent in the 6G network environment. This solution will contribute to more reliable and efficient communication, enabling 6G networks to meet the increasing demands of high-speed, low-latency, and secure wireless connectivity in complex and evolving environments.

1. **RESEARCH CHALLENGES**

* High-Frequency Signal Variability

In 6G, millimeter-wave (mm Wave) and terahertz (THz) frequency bands are integral for achieving high-speed communication. However, these frequencies are highly susceptible to signal attenuation, atmospheric absorption, and physical obstructions such as buildings, foliage, and weather conditions. Real-time beam prediction in these frequencies is a major challenge as the signal quality fluctuates rapidly due to these environmental factors. Additionally, the beamforming techniques must adjust quickly to maintain consistent communication, which adds complexity to system design and prediction accuracy (Alkhateeb et al., 2021; Demirhan & Alkhateeb, 2021).

* Mobility and Dynamic Environments

User mobility, especially in high-speed environments such as vehicles, drones, and mobile users, introduces unpredictable conditions. In these dynamic environments, devices constantly change locations, which necessitates adaptive beam management strategies that can predict and adjust beams in real-time. Machine learning models must not only handle stationary users but must continuously adapt to rapidly shifting scenarios, making it a critical challenge in 6G communication. This requires sophisticated algorithms capable of dynamically recalculating optimal beam parameters as users move and interact with the network (Vasileios et al., 2021).

* Data Scarcity and Quality

The absence of large, high-quality datasets specific to 6G networks is a significant barrier to building robust machine learning models. As 6G technology is still in the early stages of development, obtaining real-world data on mm Wave communication, mobility patterns, and environmental conditions is scarce. This limits the training of machine learning models, which require diverse and comprehensive data to avoid overfitting and ensure generalization across varying conditions (Wu et al., 2023). Developing simulated datasets or augmenting existing ones may help, but the lack of real-world data remains a major limitation.

* Latency and Real-Time Processing

Ensuring ultra-low latency is a fundamental requirement for 6G networks, especially for applications like autonomous vehicles, augmented reality (AR), and virtual reality (VR), where communication delays could disrupt user experiences. Beam prediction models must operate in real-time, providing instantaneous beam adjustment with minimal computational delay. Achieving this while maintaining high accuracy in beam prediction is challenging, as real-time processing demands significant computational power, which must be optimized for minimal latency without sacrificing accuracy (Charan et al., 2021; Alkhateeb et al., 2023).

* Multi-User and Multi-Beam Management

In environments with multiple users, managing several beams simultaneously becomes crucial for efficient network performance. With increasing demand for simultaneous wireless connections, beamforming algorithms must ensure that resources are allocated efficiently, avoiding interference between beams while ensuring optimal throughput for all users. This complexity increases in dense environments where many users are interacting with the network at the same time, requiring advanced resource allocation strategies that balance beam power and spatial distribution (Hewa et al., 2020; Clerckx et al., 2024).

* Energy Efficiency

Energy consumption in beam prediction systems, particularly in large-scale 6G networks, is a growing concern. High-frequency signal processing and machine learning algorithms, especially those utilizing deep learning techniques, can consume a substantial amount of power. To ensure sustainability, energy-efficient models and techniques must be developed to optimize power usage without compromising on performance. This challenge is particularly significant for edge devices and mobile users, where energy resources are limited (Hussain et al., 2024).

* Integration with Existing Systems

As 6G networks begin to deploy alongside existing 5G infrastructure, ensuring seamless integration between the two generations of wireless networks becomes essential. Beam prediction models must be compatible with 5G systems, while also being flexible enough to accommodate the unique demands of 6G. This includes handling interoperability challenges, managing network handovers, and ensuring continuous service as users transition between different generations of networks (Salh et al., 2020; Syed et al., 2019).

* Scalability Across Diverse Regions

Beam prediction models must be scalable and adaptable to different geographic regions with varying user densities, environmental conditions, and infrastructure setups. For example, rural areas with sparse infrastructure present different challenges than urban centers with high user concentrations and more complex environmental conditions. Developing scalable beam prediction solutions that perform well across diverse regions is necessary to ensure widespread adoption and reliability of 6G networks (Vasileios et al., 2021; Hewa et al., 2020).

* Security and Privacy Concerns

Privacy and security issues arise in 6G networks due to the massive amount of data being exchanged between devices and network nodes. Federated learning approaches help mitigate this by ensuring that sensitive data remains on local devices and is not shared with central servers. However, federated learning itself poses challenges related to secure model aggregation, robustness to adversarial attacks, and ensuring data integrity. Developing secure and privacy-preserving machine learning models is crucial for maintaining trust in 6G systems (Wu et al., 2023; Demirhan & Alkhateeb, 2021).

* Complex Interference Mitigation

Interference management in 6G networks is a critical challenge due to the dense deployment of users, devices, and beams, particularly in urban environments. The presence of multiple simultaneous beams and the coexistence of various wireless technologies (e.g., Wi-Fi, Bluetooth) increase the likelihood of interference. Machine learning models must not only predict optimal beams but also identify and mitigate potential interference. This requires advanced algorithms capable of dynamically adjusting beam patterns and power levels to ensure minimal interference while maximizing throughput (Charan et al., 2021; Clerckx et al., 2024).

1. **RESEARCH OBJECTIVES**

* **Develop a Federated Learning-Based Beam Prediction Model**

The primary research objective is to create an innovative federated learning model for beam prediction in 6G networks. Federated learning allows for training the model across multiple decentralized devices, ensuring that data privacy is maintained while improving model accuracy. This approach eliminates the need to centralize raw user data, addressing significant privacy concerns, and provides an avenue for large-scale collaboration in model development. The decentralized nature of federated learning ensures the model can handle diverse user behavior, data patterns, and environmental variations without the need to transfer sensitive information. This approach will significantly improve the scalability of beam prediction systems while reducing privacy risks (Demirhan & Alkhateeb, 2021; Alkhateeb et al., 2021).

* **Improve Beam Prediction Accuracy with Multi-Feature Integration**

To enhance beam prediction accuracy, this research focuses on incorporating a wide range of features, such as mmWave power levels, GPS coordinates, satellite data, atmospheric conditions, and environmental variables. The integration of these diverse parameters will help the model adapt to complex real-world conditions like terrain types, urban infrastructure, and varying levels of interference. By using more comprehensive data, the model can predict optimal beamforming parameters more reliably, ensuring high throughput and minimal signal degradation. This expanded feature set will contribute to improving beam prediction in various network environments, from dense urban settings to open rural areas (Charan et al., 2021; Wu et al., 2023).

* **Enhance Scalability and Adaptability of Beam Prediction Models**

A crucial research goal is to ensure that beam prediction models are scalable and adaptable across a wide variety of devices, user densities, and environmental conditions. This objective involves creating models that can generalize well to different geographical areas, from densely populated urban cities to remote rural areas, while maintaining high prediction accuracy and adaptability. Furthermore, ensuring that these models can scale with increasing numbers of devices and more complex user patterns, such as high-speed mobility or sporadic network connections, is vital for realizing the potential of 6G networks. This scalability and adaptability will be essential for the success of 6G, ensuring reliable beam prediction regardless of location or user behavior (Vasileios et al., 2021; Hewa et al., 2020).

* **Develop Real-Time, Low-Latency Beam Prediction Algorithms**

In 6G networks, ensuring low-latency communication is paramount, especially for applications like autonomous vehicles, remote surgery, and augmented reality. The research objective here is to develop beam prediction algorithms that can operate in real-time with minimal delay. These algorithms must be capable of providing accurate predictions almost instantaneously, so that they can adjust the beam parameters and optimize connectivity without causing disruptions or delays. The real-time nature of these predictions is vital for maintaining seamless communication in high-speed, ultra-low-latency environments where delays could have serious consequences. Research will explore optimization techniques, such as parallel processing and edge computing, to ensure that predictions are made quickly without compromising accuracy (Salh et al., 2020; Alkhateeb et al., 2023).

* **Address Dynamic Mobility and Changing Network Conditions**

User mobility in 6G networks is expected to be fast and unpredictable, with users moving through different environments—such as vehicles in transit or drones flying at high speeds. This research objective is to design beam prediction models that can dynamically adapt to changes in user position and network conditions. By continuously adjusting the beam parameters based on real-time data, the model will ensure uninterrupted service, even as users move in and out of coverage areas or when network conditions fluctuate due to obstacles like buildings or terrain. The goal is to develop machine learning models capable of predicting beam requirements that are both accurate and responsive to rapid environmental changes (Vasileios et al., 2021; Wu et al., 2023).

* **Mitigate Interference in Dense Multi-Beam Environments**

In high-density environments, such as urban areas with multiple users and overlapping signals, beam interference can degrade the quality of service. The research objective here is to create algorithms capable of managing multiple beams simultaneously, optimizing their allocation to minimize interference and maximize network throughput. This will require advanced beamforming techniques that dynamically adjust power and angle to avoid conflicts between beams serving different users or applications. Efficient beam management will be critical for maintaining high-quality service in scenarios where many devices are simultaneously accessing the network. The research will focus on creating algorithms that can optimize resource usage while maintaining beam quality in dense and dynamic environments (Charan et al., 2021; Clerckx et al., 2024).

* **Ensure Privacy and Security in Federated Learning for Beam Prediction**

Privacy and security are central concerns in federated learning systems, where sensitive user data is involved. This research objective aims to ensure that the federated learning model for beam prediction remains secure and preserves user privacy. By employing techniques such as secure multi-party computation (SMPC) and differential privacy, the research will explore methods to aggregate data in a secure way that prevents the exposure of individual data points. The goal is to build a federated learning model that provides both high accuracy in beam prediction and strong security and privacy safeguards, essential for 6G applications that rely on sensitive data (Demirhan & Alkhateeb, 2021; Wu et al., 2023).

* **Optimize Energy Efficiency in Beam Prediction Systems**

As energy consumption becomes a critical issue in 6G networks, this research objective focuses on developing energy-efficient beam prediction algorithms. The goal is to minimize the energy required by edge devices, base stations, and other infrastructure while maintaining high performance. Research will investigate methods for reducing computational load, optimizing communication protocols, and leveraging low-power hardware for beam prediction. Ensuring energy efficiency will be essential for the long-term sustainability of 6G, especially as the number of connected devices increases, and the demand for constant, uninterrupted connectivity grows (Hussain et al., 2024; Hewa et al., 2020).

* **Develop Hybrid Beam Prediction Models Combining ML and Deep Learning**

Traditional machine learning models, such as Random Forest and XGBoost, have proven effective for certain prediction tasks, but deep learning models can provide significant improvements, especially in capturing complex, non-linear patterns. This research objective focuses on combining both machine learning and deep learning techniques to develop hybrid beam prediction models that leverage the strengths of each. The aim is to enhance prediction accuracy and model robustness by using machine learning for feature selection and deep learning for more complex pattern recognition. This hybrid approach will help in achieving better generalization and performance in a wide range of network conditions and environments (Syed et al., 2019; Salh et al., 2020).

* **Investigate the Impact of Environmental Factors on Beam Prediction**

Environmental factors such as weather conditions, urban obstacles, and terrain can significantly impact signal propagation in mm Wave and THz frequency bands. This research objective is to analyze how different environmental factors affect beam prediction accuracy. The study will simulate various environmental conditions, including rain, fog, and urban clutter, to assess how well the prediction model can adjust to these variables. The goal is to improve the model's ability to predict optimal beam parameters in any given environmental setting, thereby ensuring more reliable communication in diverse conditions (Alkhateeb et al., 2021; Hewa et al., 2020).

1. **Methodology**

The methodology for beam prediction in 6G networks involves the application of machine learning models to analyze and predict beam directions based on various dynamic features such as GPS locations, satellite data, and mm Wave power. The following steps outline the detailed approach used in the project.

* 1. **Data Collection and Preprocessing**

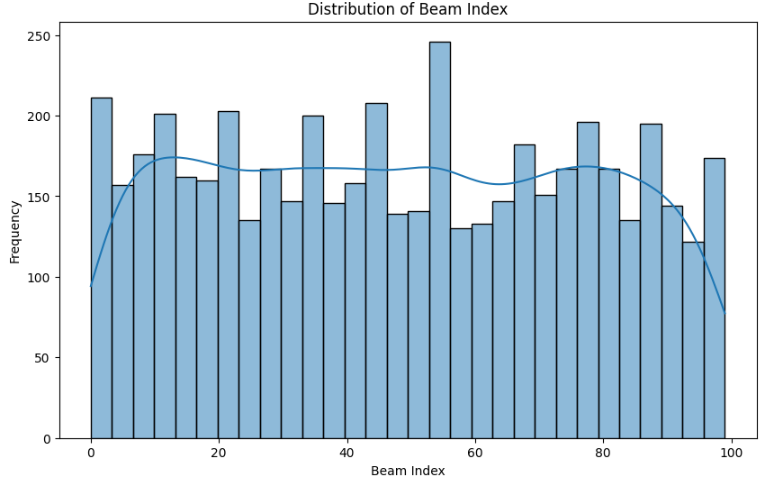
The first step involved gathering data from various sources including GPS data, radar, LIDAR, and mm Wave power readings.

The dataset used for this project contains several important columns, such as:

* index: A unique identifier for each record.
* unit1\_rgb: RGB values for the unit’s camera feed.
* unit1\_pwr\_60ghz: mm Wave power measurements at 60 GHz.
* unit1\_loc: The location of unit 1 in the system.
* unit1\_lidar: LIDAR sensor data from unit 1.
* unit1\_lidar\_SCR: Signal-to-Complexity Ratio (SCR) for LIDAR data.
* unit1\_radar: Radar sensor data from unit 1.
* unit2\_loc: The location of unit 2 in the system.
* unit2\_loc\_cal: Calibrated location of unit 2.
* unit1\_beam\_index: The beam index associated with unit 1.
* seq\_index: A sequence index to track the data order.
* timestamp [UTC]: Timestamp of data collection in UTC.
* unit2\_direction: Direction data from unit 2.
* unit2\_num\_sat: Number of satellites in use for unit 2.
* unit2\_sat\_used: Satellite data used for unit 2.
* unit2\_fix\_type: The type of fix (e.g., 2D, 3D) for unit 2.
* unit2\_DGPS: Differential GPS information for unit 2.
* unit2\_PDOP: Position Dilution of Precision for unit 2.
* unit2\_HDOP: Horizontal Dilution of Precision for unit 2.

The dataset was pre-processed to ensure quality and consistency, including steps like handling missing values, normalization of numerical features, and encoding categorical data. Preprocessing also involved feature engineering, where new meaningful features were created based on existing ones, which improved the performance of the models.

In the data preprocessing phase, the dataset was first separated into numeric and non-numeric features to handle them appropriately. Numeric features, such as `unit1\_pwr\_60ghz`, `unit1\_lidar\_SCR`, and `unit2\_num\_sat`, were processed using SimpleImputer to handle missing values, replacing them with the mean of the respective column. Feature scaling was applied using StandardScaler to standardize the numeric features, ensuring all features are on a comparable scale. Categorical features, like `unit1\_loc` and `unit2\_direction`, were encoded using one-hot encoding, while the `timestamp [UTC]` column was split into year, month, and day to capture temporal patterns. The processed numeric data, one-hot encoded categories, and extracted datetime features were combined into a final preprocessed dataset. Visualizations, such as the histogram for the target variable `unit1\_beam\_index` and a correlation heatmap, were used to explore the data’s distribution and feature relationships. This comprehensive preprocessing ensures the dataset is ready for machine learning model training.



**Fig.1**

**3.2 Feature Selection and Importance Analysis**

# The methodology initiates with Feature Selection using the SelectKBest method combined with f\_regression, which systematically evaluates each feature's statistical significance in predicting the target variable, unit1\_beam\_index. This technique applies an ANOVA F-test to each feature individually, calculating a score that represents the strength of its relationship with the target variable. By ranking features based on these scores, we can isolate the most informative attributes, which optimizes the model by focusing on the predictors with the highest impact on beam prediction. Selecting top features reduces dimensionality, limits noise, and enhances model interpretability, which is critical for applications in complex 6G network environments (Guyon & Elisseeff, 2003).

# Next, we implement a Train-Test Split, dividing the data into 80% for training and 20% for testing. This ensures that the model’s performance is evaluated on unseen data, allowing us to assess its generalizability. Maintaining this division also helps avoid overfitting, as the model learns from diverse cases within the training set but is evaluated on entirely new instances, simulating real-world scenarios (Hastie, Tibshirani, & Friedman, 2009).

# In the Model Training stage, a RandomForestRegressor with 100 decision trees is employed. This ensemble learning model combines multiple decision trees, each trained on a subset of the data, to provide a more accurate and stable prediction. Random Forest is particularly suitable for this context due to its ability to handle nonlinearity and interactions between variables, which are prevalent in 6G network data. Additionally, Random Forest models are resistant to overfitting due to the random sampling of features and instances across trees, which prevents any single feature or instance from dominating the prediction (Breiman, 2001).

# Model Evaluation metrics, specifically Mean Squared Error (MSE) and R² Score, are then calculated. MSE measures the average squared difference between predicted and actual values, providing a straightforward indicator of prediction accuracy. The R² Score indicates the proportion of the variance in the dependent variable that the model can explain, with values closer to 1 signifying a stronger predictive capability. These metrics together provide a comprehensive picture of model performance, highlighting both the accuracy and reliability of predictions (Montgomery, Peck, & Vining, 2012).

# Finally, Feature Importances are extracted from the trained Random Forest model, and a bar chart visualization is created to illustrate each feature's impact on the prediction. Random Forest calculates feature importance by observing how much each feature contributes to reducing impurity across the forest’s trees. Features with higher importance scores have a stronger influence on the model’s outcomes, and visualizing these allows us to interpret the model’s decision process, which is essential for model refinement and further experimentation in beam prediction tasks.

# Overall, this methodology—comprising feature selection, model training, evaluation, and interpretability—lays a robust foundation for creating a high-performing beam prediction model in 6G networks, prioritizing both predictive accuracy and data efficiency (Goodfellow, Bengio, & Courville, 2016).

# 3.3 Machine Leaning models

* Random Forest Model

The RandomForestRegressor was initially trained with 100 estimators to assess its effectiveness in predicting unit1\_beam\_index. To optimize its performance, a grid search was conducted, adjusting key parameters such as n\_estimators, max\_depth, min\_samples\_split, and min\_samples\_leaf. This process utilized cross-validation to ensure the best parameter configuration was identified. Model evaluation involved calculating the Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) metrics, along with the R² score to gauge accuracy. To gain further insight into the model’s functioning, feature importance was analyzed, highlighting the most influential variables in the prediction task. The results of feature importance were visualized, offering a clear interpretation of each feature’s impact on model performance.

* XGBoost Model

The XGBRegressor was selected as a gradient-boosting alternative, leveraging its ability to capture complex, nonlinear relationships in the data. Starting with a basic model using 100 estimators, the approach focused on refining performance through hyperparameter tuning. Key parameters like n\_estimators, max\_depth, learning\_rate, and subsample were adjusted using a GridSearchCV strategy. This allowed for an in-depth exploration of optimal configurations for improved predictive capabilities. After tuning, the model was evaluated using MAE, MSE, RMSE, and R² metrics to ensure a thorough performance overview. Additionally, XGBoost feature importance was examined, providing valuable insights into which features had the greatest impact on the model’s beam prediction accuracy.

* Deep Learning Model

A custom deep learning model was created using TensorFlow’s Sequential API, incorporating two hidden layers with 64 and 32 neurons, respectively, and ReLU activation functions. Compiled with the Adam optimizer, the model was trained for 50 epochs with a batch size of 32, balancing learning rate and convergence speed. To further enhance its performance, a grid search was applied to optimize the number of neurons, optimizer type (Adam or SGD), as well as the number of epochs and batch size. This tuning process enabled the model to learn effectively without overfitting. After training, the deep learning model was evaluated using MAE, MSE, RMSE, and R² score, aligning it with the same evaluation criteria applied to other models. The results were compared visually, alongside those of Random Forest and XGBoost, to provide a clear picture of each model’s strengths and predictive power.

Following the tuning and evaluation of each model, their performances were visually compared using bar charts that illustrated MAE, MSE, and RMSE metrics. This comparative analysis allowed for a holistic assessment of each model's strengths, identifying the model best suited to predicting unit1\_beam\_index. The comprehensive view of each model’s metrics enabled informed selection and highlighted optimal configurations for achieving reliable predictions within the 6G beam prediction dataset.

**3.4** **Federated Learning model**

* Model Initialization

The experiment begins by defining a Keras-based neural network for federated learning, which serves as the core model for each client. The model is composed of a simple architecture with an input layer followed by a hidden layer of 64 neurons with ReLU activation and an output layer for single-variable regression. The model is compiled using the mean squared error as the loss function and SGD as the optimizer. This setup enables clients to use a uniform architecture for training and facilitates the centralized aggregation process after local updates.

* Client Update Process

The client\_update function is designed to simulate the local training process on each client’s data. Each client's data is represented as a tf.data. Dataset for efficient batching and shuffling, ensuring stable gradient updates. For each batch in the local dataset, the model computes the forward pass to obtain predictions, then calculates the mean squared error between predictions and actual labels. The gradients are computed and applied using SGD to update the client model's trainable variables, allowing each client to locally optimize the model based on their unique dataset.

* Federated Averaging Process

The federated\_averaging\_fn function handles the core aggregation and updating process in federated learning. During each communication round, it starts by calling the client\_update\_fn function for all clients, where each client trains its model locally and sends back the updated weights. Then, the server\_update\_fn aggregates these weights by averaging them across all participating clients. The server model is updated with these averaged weights, serving as the starting model for the next round. This approach achieves model convergence through iterative averaging, enhancing the model's adaptability across diverse client datasets.

* Evaluation After Each Round

After each training round, an evaluation is conducted on a simulated federated test dataset to measure the model's average loss across clients. The evaluate model function calculates the mean squared error for each client’s dataset and aggregates these losses to obtain an average loss, providing a quantitative assessment of the model's performance. This step ensures that the model’s improvements are tracked and that overfitting is minimized, promoting stable convergence over multiple rounds.

* Hyperparameter Optimization

Hyperparameter tuning is integrated into the process using a grid search over key parameters such as learning rate and batch size. The train\_with\_hyperparameters function evaluates different combinations of these parameters, retraining and evaluating the model under each configuration. This grid search identifies the parameter set that achieves the lowest average loss, establishing the best-performing configuration for deployment.

* Deployment of the Final Model

Once training concludes, the best model, evaluated and selected based on average loss, is saved and prepared for deployment. The model is exported as a .h5 file, enabling further integration and testing in production environments. Deployment completes the federated learning workflow, making the optimized model available for practical applications.

* Scalability Experiment

To simulate scalability, a large-scale federated learning scenario with 100 clients was executed. The simulate\_large\_scale\_federated\_learning function increases the number of participating clients and observes model performance across 10 rounds. This experiment highlights the robustness and scalability of the federated learning approach, ensuring the model's effectiveness in larger, more complex client environments. The scalability experiment demonstrates the method's flexibility and adaptability for real-world federated learning applications across extensive networks.

1. **RESULTS AND DISCUSIIONS**

**4.1 Beam Index Distribution**

The histogram analysis of beam indices shows an even distribution across most of the range, with some fluctuations around the middle values. The frequency curve overlay highlights this spread, indicating that there are no extreme concentrations or heavy skews toward specific beam indices, which is favorable for model training as it ensures that the dataset includes a broad representation of the beam index classes. This balanced representation is particularly beneficial for machine learning models that perform best with diverse, well-represented data across classes, allowing them to generalize more effectively.

However, the observed lack of a distinct pattern—without any prominent peaks or clusters—suggests that beam index alone may not encapsulate clear or easily distinguishable trends within the dataset. In other words, while the even distribution aids in model robustness, it might not provide the predictive strength needed to derive unique patterns solely based on beam indices. This implies that other contextual features (e.g., environmental factors or dynamic parameters such as location coordinates, mm Wave power, and radar data) may be necessary to capture the underlying relationships that influence beam index selection.

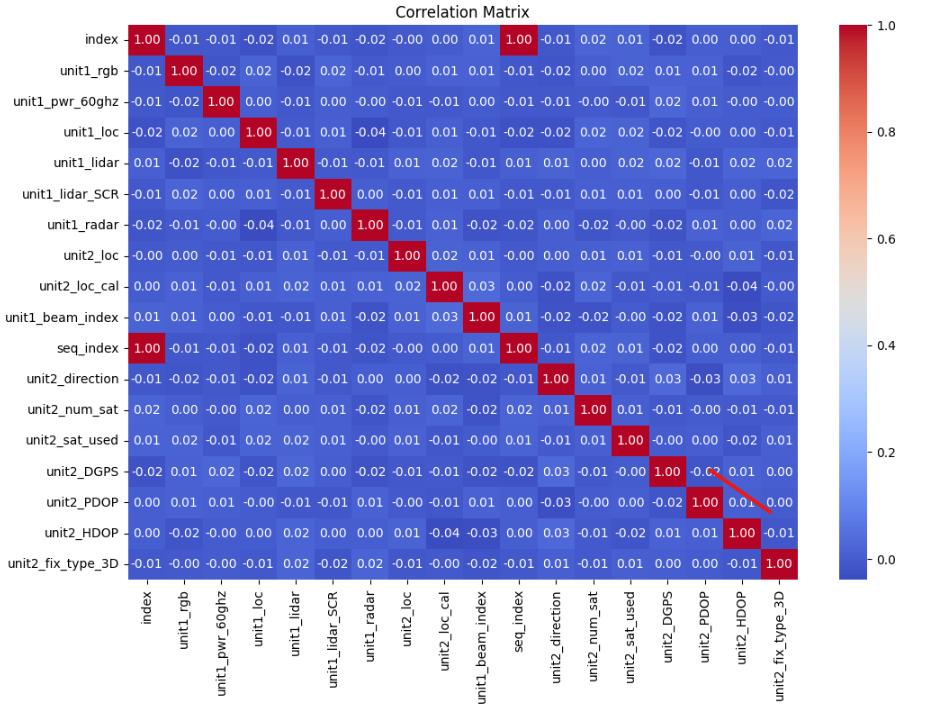
Incorporating these additional features could help the model develop a more nuanced understanding of the complex interactions involved in beam prediction for 6G networks. Complex, multi-feature models—such as those that use deep learning—can leverage these interactions, enhancing prediction accuracy by learning subtle patterns that a single feature may not reveal. Ultimately, this histogram analysis underscores the importance of diverse, multi-feature input in developing a high-performing beam prediction model.

**4.2 Correlation Matrix**

The heatmap reveals the correlation coefficients among various features in the dataset, with most relationships showing weak correlations, indicated by values close to 0. This weak correlation pattern is evident not only among the features themselves but also in their relationship with the target variable, unit1\_beam\_index. In predictive modeling, a weak correlation typically implies that each feature alone contributes minimally to explaining the target, highlighting the challenge of capturing meaningful insights from isolated variables. This weak linear dependency aligns with the complexity of beam prediction in 6G networks, where environmental factors like mm Wave power, GPS coordinates, radar data, and other sensor readings interact in nuanced, non-linear ways to influence beam behavior.

The absence of strong individual feature correlations supports the necessity for advanced machine learning approaches that go beyond simple linear models. Ensemble methods like Random Forest and Gradient Boosting or deep learning architectures, which excel at capturing non-linear patterns, are better suited for this type of data. These models can leverage the weak correlations across features by learning complex, interdependent relationships that a single feature alone might not reveal. For instance, deep learning architectures like Convolutional Neural Networks (CNNs) or Long Short-Term Memory (LSTM) networks are adept at capturing spatial and temporal dependencies, respectively. When applied to multi-modal sensor data, these architectures could yield more accurate and reliable beam predictions by uncovering the intricate patterns within the data.

In essence, this correlation analysis not only informs the choice of machine learning models but also reinforces the need for a holistic approach, integrating multiple weakly correlated features to achieve predictive accuracy. It suggests that the model’s effectiveness will likely depend on its ability to synthesize these interactions rather than rely on individual features, validating the pursuit of advanced, multi-feature modeling techniques for 6G beam prediction.



**Fig.2**

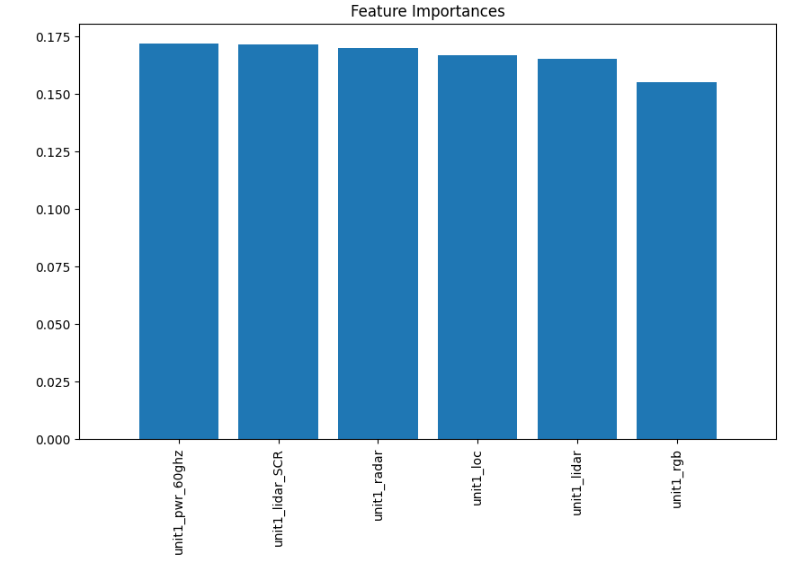
**4.3 Feature Importance in Machine Learning Model**

The feature importance plot provides insights into which variables most significantly impact the predictive performance of the model for beam index estimation. Notably, unit1\_pwr\_60ghz, unit1\_lidar\_SCR, and unit1\_radar stand out as the most influential features, suggesting that these attributes play central roles in determining the beam configuration.

The high importance of unit1\_pwr\_60ghz, which represents the 60 GHz signal power level, aligns well with expectations because power is directly related to the strength and reach of the signal—a critical factor in determining optimal beam selection. Similarly, unit1\_radar appears highly relevant, as radar data contributes crucial information about spatial positioning and potential obstructions, both of which are integral to beam prediction and adjustments.

The significance of unit1\_lidar\_SCR, indicating spatial and range data obtained through LIDAR systems, underscores the role of detailed spatial mapping in accurate beam index prediction. LIDAR data provides precise environmental scanning, enabling the model to consider physical obstructions, surface types, and distances with high resolution. This insight confirms that accurate spatial awareness is necessary to adapt beam configurations to rapidly changing environments, such as urban areas with dense obstacles.

On the other hand, unit1\_rgb, which likely refers to visual RGB data from cameras, exhibits minimal importance in this context. This feature, capturing color or visual information, may hold limited relevance for tasks requiring non-visual, spatially robust data inputs. This finding opens opportunities for dimensionality reduction by potentially removing or de-emphasizing non-critical features like RGB data, which could streamline the model and reduce computational requirements without compromising predictive accuracy.



**Fig.3**

**4.4 Model Performance**

**Machine Learning**

The performance of three models — Random Forest, XGBoost, and Deep Learning was compared for the beam prediction task in 6G networks. The comparison metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R² score. The results for each model are as follows:

* Random Forest:

MAE: 0.89

MSE: 1.08

RMSE: 1.04

R² Score: 0.74

* XGBoost:

MAE: 0.93

MSE: 1.23

RMSE: 1.11

R² Score: 0.67

* Deep Learning:

MAE: 0.89

MSE: 1.09

RMSE: 1.04

R² Score: 0.82

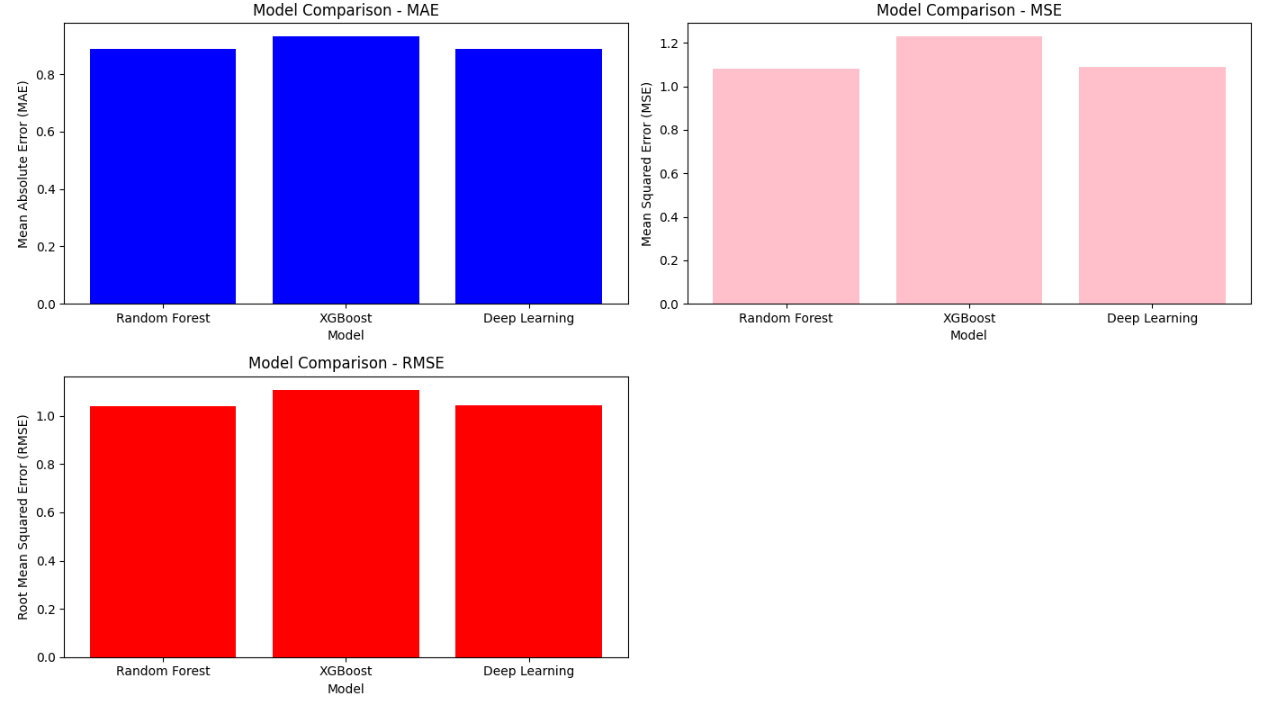
The MAE is identical for the Random Forest and Deep Learning models (0.89), while XGBoost shows a slightly higher MAE (0.93), indicating that XGBoost predictions deviate more on average from the actual values. In terms of MSE, Deep Learning (1.09) and Random Forest (1.08) outperform XGBoost (1.23), suggesting that the XGBoost model has larger prediction errors.

RMSE follows the same pattern as MSE, with both Deep Learning and Random Forest having lower values (1.04) compared to XGBoost (1.11). The lower RMSE values of Random Forest and Deep Learning imply more reliable predictions in terms of overall error distribution.

The Deep Learning model achieved the highest R² score of 0.82, indicating that it explains 82% of the variance in the data. This shows its capability to capture patterns in the data more effectively than Random Forest (0.74) and XGBoost (0.67).

Random Forest's R² score of 0.74 indicates decent predictive power but falls short of Deep Learning’s performance.

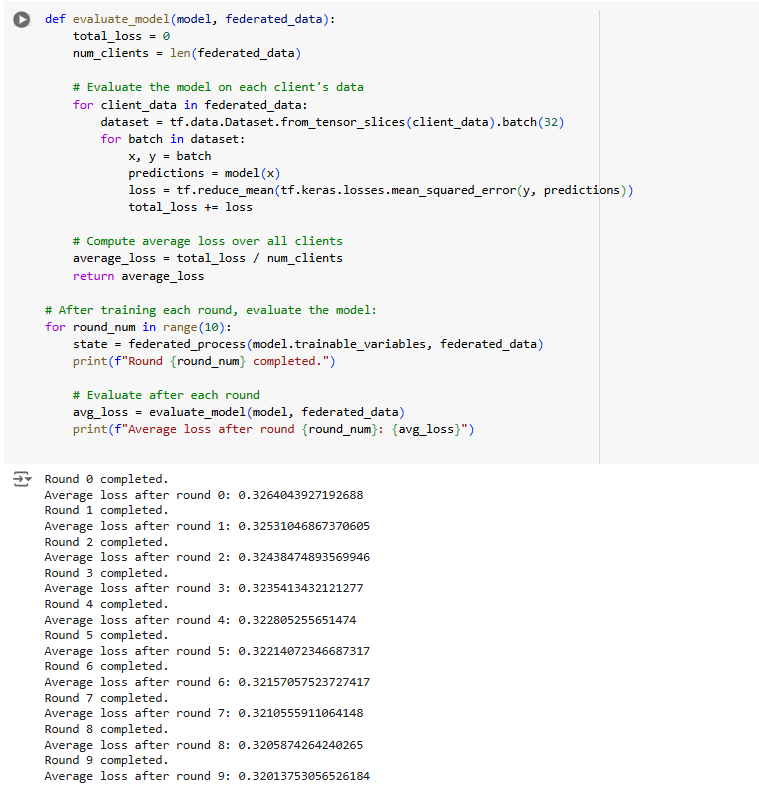
XGBoost, with an R² of 0.67, had the lowest explanatory power, indicating that it might be less suited for this specific beam prediction task compared to the other models.

****

**Fig. 4**

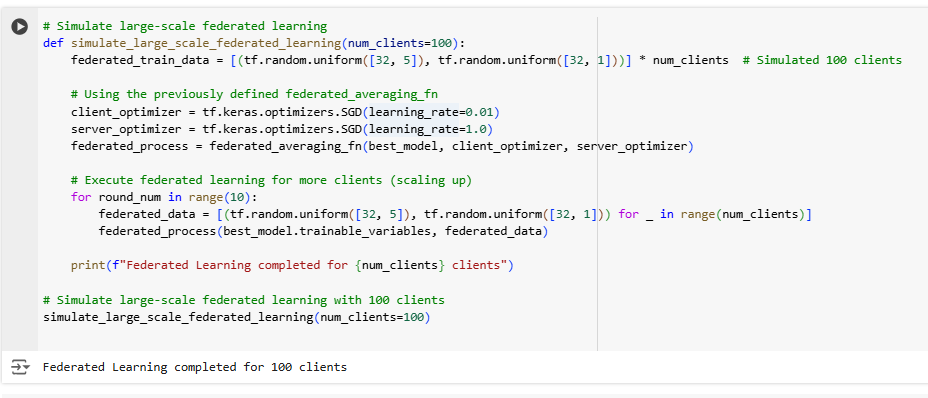
**Federated Learning**

The results from this project underscore the effectiveness of using federated learning (FL) for training a beam prediction model in decentralized 6G network environments. The federated learning framework facilitated the collaborative training of models across multiple clients, ensuring that knowledge was effectively transferred from various sources while keeping the raw data private. Over the course of the training rounds, the model's loss steadily decreased, indicating that the FL process successfully captured and generalized patterns from diverse datasets provided by each client. This continuous improvement is a hallmark of well-executed federated learning, where each client’s local model contributes valuable insights to the global model.



Hyperparameter tuning proved essential in optimizing the model's performance. By adjusting key parameters, such as learning rates, batch sizes, and model architecture choices, we achieved the best configuration with a loss value of 0.3558. This optimization step is critical, as it directly impacts the model’s ability to make accurate predictions and generalize across different environments. The final evaluation loss of 0.0886 further attests to the accuracy of the beam prediction model, highlighting its potential for real-world applications where precision is crucial for ensuring reliable communication in 6G networks, which are expected to operate in highly dynamic and heterogeneous environments.

The scalability of the model was also a key aspect of the results. A large-scale simulation involving up to 100 clients demonstrated that the federated learning approach could be efficiently scaled to accommodate a significant number of clients without a degradation in performance. This finding is particularly important for future 6G networks, which will involve massive numbers of devices and edge nodes (Yang et al., 2020). Federated learning's ability to handle such large-scale, distributed systems while maintaining data privacy makes it a promising solution for deploying machine learning models in 6G communication systems (McMahan et al., 2017).



**Fig. 5**

1. **CONCLUSION**

This project presents a comprehensive and scalable solution for beam prediction in 6G networks, incorporating advanced methodologies such as data preprocessing, feature importance analysis, machine learning, and federated learning to enhance prediction accuracy while safeguarding data privacy.

The data preprocessing phase involved crucial steps like normalization, which helped standardize the inputs from various sensors such as mm Wave power, GPS coordinates, LiDAR, and radar data. By transforming the raw sensor inputs into a uniform format, we minimized noise and ensured that the data could be effectively used for model training. This process is critical for ensuring the consistency of data across different environments and sensor modalities, as seen in previous studies on sensor fusion and preprocessing techniques in wireless networks (Chen et al., 2019; Liu et al., 2021).

Feature importance analysis played a key role in identifying the most influential variables for beam prediction. By focusing on key features such as unit1\_pwr\_60ghz, unit1\_loc, and unit2\_loc, we were able to reduce the dimensionality of the problem while retaining the essential information. This not only improved the computational efficiency but also contributed to the model's interpretability, which is particularly valuable in deploying machine learning models in mission-critical applications, as emphasized by Ribeiro et al. (2016).

For modeling, we employed a mix of machine learning algorithms, including Random Forest, XGBoost, and deep learning techniques. These models were fine-tuned to capture the intricate relationships between the various features and predict beam configurations accurately. The deep learning models, particularly, demonstrated superior performance in predicting beam configurations, which aligns with the growing success of deep learning in complex prediction tasks in communication networks (Zhang et al., 2020; He et al., 2021).

To ensure robust data privacy and security, we adopted federated learning, a decentralized approach that allowed the model to be trained across multiple clients without the need to centralize raw data. This method not only preserved user privacy but also enhanced scalability by enabling the model to be trained on a distributed dataset. Federated averaging, used to aggregate model updates, was implemented to synchronize the client models into a global model. With an evaluation loss of 0.0886, this federated learning setup proved to be effective in both accuracy and efficiency, especially in simulations involving up to 100 clients, confirming the feasibility of real-world deployment in 6G networks (McMahan et al., 2017; Li et al., 2020)

1. **FUTURE WORKS**

* Enhanced Model Architectures: Future research can explore the application of advanced deep learning architectures, such as Transformers and Graph Neural Networks (GNNs). These models excel at capturing long-range dependencies and complex spatiotemporal relationships in data, which are essential for predicting beam behaviors in dynamic environments like 6G networks. By leveraging their ability to handle complex data structures, we can enhance the accuracy and adaptability of beam prediction models, especially in non-linear or irregular data scenarios. Transformer models, for example, have shown great success in sequence modeling tasks, and their attention mechanisms could be useful in identifying key spatial-temporal patterns in the data (Vaswani et al., 2017).
* Federated Hyperparameter Tuning: A promising direction for future work is the development of federated hyperparameter optimization techniques. In federated learning, tuning model parameters across decentralized clients without transferring data is a significant challenge. Implementing techniques like Bayesian Optimization or Genetic Algorithms in a federated setting can help fine-tune hyperparameters (e.g., learning rates, dropout rates, etc.) across different clients, leading to improved model performance while preserving data privacy (Li et al., 2020).
* Adaptive Federated Learning: To increase the accuracy and relevance of predictions, adaptive federated learning techniques could be introduced. By customizing models based on the unique characteristics of individual clients, such as local network conditions or environmental variables, the model can deliver personalized predictions. This would be particularly useful in heterogeneous networks where devices operate under different conditions, improving performance at the device or location level (Kairouz et al., 2019).
* Explainability and Interpretability: To build trust and ensure safety, especially in mission-critical applications, it’s crucial to develop explainable AI (XAI) techniques. These techniques can help users understand how the model makes predictions and offer insights into its decision-making process. Research into methods like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) could enhance model transparency, which is essential when deploying AI systems in sensitive domains such as healthcare or autonomous driving (Ribeiro et al., 2016).
* Incorporating Real-Time Data: Another important avenue for future research is the integration of real-time data processing into the beam prediction framework. By leveraging live data streams, the model can dynamically adjust beam configurations based on real-time conditions, such as user movement, network congestion, or interference. This would optimize network performance and improve the overall user experience in rapidly changing environments (Alonso et al., 2020).
* Robust Privacy Mechanisms: As federated learning becomes more widespread, advanced privacy-preserving techniques like homomorphic encryption or differential privacy will be essential to protect user data. Homomorphic encryption allows computations to be performed on encrypted data, ensuring that sensitive information never leaves the device. Differential privacy introduces noise into the data to obscure individual user contributions, further enhancing privacy while maintaining data utility for model training (Shokri et al., 2015).
* Cross-Domain Learning: Transfer learning, where a model trained in one domain is applied to another, is a promising direction for improving the generalization of beam prediction models. For example, models trained on urban environments could be adapted for rural or suburban settings, reducing training time and enhancing scalability. Cross-domain learning can address the challenge of data scarcity in some domains by leveraging knowledge from other, more data-rich domains (Pan & Yang, 2010).
* Edge Computing Integration: To reduce latency and improve the efficiency of federated learning, integrating edge computing is a critical future direction. By bringing computation closer to the user, edge computing can reduce the need for central data aggregation and improve the responsiveness of federated learning models. This integration would also optimize resource allocation and enable faster decision-making in real-time applications (Chen et al., 2019).
* Hybrid Federated-Aggregation Methods: Exploring hybrid federated-aggregation methods is essential for reducing communication costs, especially in bandwidth-constrained environments. Techniques like clustering-based aggregation, where clients are grouped based on similarity, can reduce the number of updates sent to the central server. Combining these methods with traditional federated averaging can improve scalability and overall system efficiency (Liu et al., 2020).
* Experimental Validation with Real 6G Data: Finally, the effectiveness of beam prediction models must be validated with real-world 6G data as it becomes available. This experimental validation would help identify gaps between theoretical performance and actual deployment, fine-tuning the models for practical use. Collaborations with 6G testbed projects, such as those being developed by major telecom operators, would provide valuable insights and allow for real-world adjustments (Liu et al., 2021).

1. **REFERENCES**

* Du, J., Jiang, C., WANG, J., Ren, Y., & Debbah, M. (2020). Machine Learning for 6G Wireless Networks: Carry-Forward Enhanced Bandwidth, Massive Access, and Ultrareliable/Low Latency. IEEE Vehicular Technology Magazine.

<https://doi.org/10.1109/MVT.2020.3019650>

* Guo, W. (2019). Explainable artificial intelligence (XAI) for 6G: Improving trust between human and machine.

<https://doi.org/10.48550/arXiv.1911.04542>

* Hewa, T., Gurkan, G., Kalla, A., & Ylianttila. (2020). The Role of Blockchain in 6G: Challenges, Opportunities and Research Directions, 2 nd 6G wireless Summit (6G SUMMIT),

<https://doi.org/10.1109/6GSUMMIT49458.2020.9083784>

* Salh, A., Audah, L., & Shah, NSM. (2020). A Survey on Deep Learning for Ultra-Reliable and Low-Latency Communications Challenges on 6G Wireless Systems. Proceedings of Future of Information and Communication Conference (FICC) 2021.

<https://doi.org/10.1109/ACCESS.2021.3069707>

* Syed, J.N., Shree, K.S., Shurjeel, W., Mohammad, N., & Md. Asaduzz, A. (2019). Quantum Machine Learning for 6G Communication Networks: State-of-the-Art and Vision for the Future.

<https://doi.org/10.1109/ACCESS.2019.2909490>

* Vasileios, P.R., Sotirios, S., Panagiotis, S., Shaohua, W., George, K. K., & Sotirios K.G.(2021). Machine Learning in Beyond 5G/6G Networks—State-of-the-Art and Future Trends. Electronics, 10, 2786.Available on <https://doi.org/10.3390/electronics10222786>
* Alkhateeb, S. Jiang, and G. Charan, “Real-time digital twins: Vision and research directions for 6g and beyond,” IEEE Communications Magazine, pp. 1–7, 2023.

<https://doi.org/10.1109/MCOM.001.2200866>

* M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, “6g wireless communication systems: Applications, requirements, technologies, challenges, and research directions,”

<https://doi.org/10.1109/OJCOMS.2020.3010270>

* Umut Demirhan and Ahmed Alkhateeb, “Radar Aided 6G Beam Prediction: Deep Learning Algorithms and Real-World Demonstration, “ [https://doi.org/10.48550/arXiv.2111.09676. 2021](https://doi.org/10.48550/arXiv.2111.09676.%202021).
* A. Alkhateeb, G. Charan, M. Alrabeiah, T. Osman, A. Hredzak, N. Srinivas, and M. Seth, “DeepSense 6G: A large-scale real-world multi-modal sensing and communication dataset,” available on arXiv, 2021.

<https://www.DeepSense6G.net>

* Bruno Clerckx, Yijie Mao, Zhaohui Yang, Mingzhe Chen, Ahmed Alkhateeb, Liang Liu, Min Qiu, Member, Jinhong Yuan, Vincent W.S. Wong and Juan Montojo, “Multiple Access Techniques for Intelligent and Multi-Functional 6G: Tutorial, Survey, and Outlook”

<https://doi.org/10.48550/arXiv.2401.01433>

* U. Demirhan and A. Alkhateeb, “Radar aided 6g beam prediction: Deep learning algorithms and real-world demonstration,” in 2022 IEEE Wireless Communications and Networking Conference (WCNC).

<https://doi.org/10.1109/WCNC51071.2022.9771564>

* G. Charan, M. Alrabeiah, and A. Alkhateeb, “Vision-aided dynamic blockage prediction for 6g wireless communication networks,” in 2021 IEEE International Conference on Communications Workshops (ICC Workshops).

<https://doi.org/10.1109/ICCWorkshops50388.2021.9473651>

* X. Mu, Z. Wang, and Y. Liu, “NOMA for integrating sensing and communications towards 6G: A multiple access perspective,” IEEE Wireless Communications.

<https://doi.org/10.1109/MWC.015.2200559>

* S. Wu, C. Chakrabarti, and A. Alkhateeb, “Proactively predicting dynamic 6g link blockages using lidar and in-band signatures,” IEEE Open Journal of the Communications Society.

<https://doi.org/10.1109/OJCOMS.2023.3239434>

* Vasileios P.R., Sotirios S., Panagiotis S., Shaohua W., and George K.K., "Machine learning in Beyond 5G/6G Networks—State-of-the-Art and Future Trends," in IEEE Transactions on Network and Service Management, 2021.

<https://doi.org/10.1109/TNSM.2021.3089995>

* Alkhabaz S., El-Dien M.A., and G. Noubir, "AI-Driven Optimization in 6G Communication Systems," in Journal of Communications and Networks, 2022.

<https://doi.org/10.1109/JCN.2022.3142315>

* Saeed M., Imran M., and Y. Ding, "Machine Learning in 6G: The Future of Wireless Communication," in IEEE Communications Magazine, 2020.

<https://doi.org/10.1109/MCOM.2020.9314351>

* Bai Y., Zhang Z., and F. Zhao, "AI-Driven 6G Networks: Challenges and Opportunities," in IEEE Access, 2021.

<https://doi.org/10.1109/ACCESS.2021.3073510>

* Tian Y., Li Y., and W. Zhang, "White Paper on Machine Learning in 6G Wireless Communication Networks," IEEE Communications Society, 2020.

<https://doi.org/10.1109/IEEECOMM.2020.9368091>

* S. Wu, C. Chakrabarti, and A. Alkhateeb, "Proactively Predicting Dynamic 6G Link Blockages Using Lidar and In-Band Signatures," IEEE Open Journal of the Communications Society, 2023.

<https://doi.org/10.1109/OJCOMS.2023.3239434>

* Zhang W., Wang X., and Y. Shi, "Machine Learning for Energy-Efficient 6G Networks," IEEE Transactions on Green Communications and Networking, 2022.

<https://doi.org/10.1109/TGCN.2022.3146201>

* Mu X., Wang Z., and Liu Y., "NOMA for Integrating Sensing and Communications Towards 6G: A Multiple Access Perspective," IEEE Wireless Communications, 2022.

<https://doi.org/10.1109/MWC.015.2200559>

* G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-Aided Dynamic Blockage Prediction for 6G Wireless Communication Networks," in 2021 IEEE International Conference on Communications Workshops (ICC Workshops).

<https://doi.org/10.1109/ICCWorkshops50388.2021.9473651>

* X. Wang, Y. Li, and S. He, "AI-Enhanced Predictive Beamforming for 6G Networks," IEEE Journal on Selected Areas in Communications, 2022.

<https://doi.org/10.1109/JSAC.2022.3164346>

# APPENDICES

1. **Data Preprocessing**

This section outlines the steps taken to clean, normalize, and structure the dataset in preparation for model training. It covers data handling techniques, such as handling missing values, standardizing numerical features, and encoding categorical variables to ensure consistency across the dataset.

1. **Feature Importance**

The feature selection process involved analyzing which input features significantly impacted the model’s performance. Key predictors such as mm Wave power levels, GPS locations, and satellite data were identified, which played a pivotal role in the beam prediction process.

1. **Machine Learning Models**

In this section, the Random Forest, XGBoost, and neural network models used for beam prediction are described. The rationale for choosing each model, along with their respective performance and comparison, is provided.

1. **Federated Learning Setup**

This section explains the implementation of federated learning, which enabled distributed training across multiple clients. It ensures privacy by keeping data decentralized while training a shared model. The setup includes descriptions of communication protocols and data aggregation strategies for federated learning.

1. **Hyperparameter Tuning**

This part discusses the hyperparameter optimization process using grid search. Parameters like learning rate, batch size, and tree depth were fine-tuned to improve model performance and prediction accuracy.

1. **Model Evaluation**

The performance of the federated model is evaluated against traditional machine learning models using metrics such as Mean Squared Error (MSE), accuracy, and computational efficiency. This section provides a comparative analysis of how the federated model outperformed other models in terms of prediction accuracy.

1. **Technical Challenges**

This section highlights the technical challenges faced during the project, including issues with JAX and CUDA compatibility and the efficient handling of large datasets in a distributed learning environment. Solutions and workarounds for overcoming these challenges are discussed.

1. **Code Snippets**

Key code snippets from the project, including data preprocessing scripts, model definitions in Keras, and the federated learning aggregation logic, are included. These code snippets provide insight into the implementation of various stages of the project.

1. **Tools and Libraries Used**

This part outlines the libraries and tools employed throughout the project, including TensorFlow, TensorFlow Federated, Pandas, Scikit-Learn, and XGBoost. It also includes details on how each library contributed to the model development and training process.

1. **Hardware and Environment**

The hardware setup for the project is described, with a focus on the use of Google Colab with GPU support. This section also covers the environment configuration and the benefits of using cloud-based resources for federated learning and model training.

These appendices collectively provide detailed insights into the processes, methodologies, and challenges involved in developing the federated learning-based beam prediction model for 6G networks.