**A Study on Creating Digital Twin Foliage Representation Through Computer Vision, Aerial Image Analysis and Machine Learning techniques to enhance the Network Planning and Deployment**

Dissertation Proposal

Submitted to National University

National University

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OFDATA SCIENCE

by

HASHIM SHAIK

San Diego, California

May 2024

**Table of Contents**

[Chapter 1: Introduction 1](#_Toc165902208)

[Statement of the Problem 3](#_Toc165902209)

[Purpose of the Study 4](#_Toc165902210)

[Introduction to Theoretical or Conceptual Framework 5](#_Toc165902211)

[Introduction to Research Methodology and Design (Nature of the Study) 8](#_Toc165902212)

[Research Questions 10](#_Toc165902213)

[Hypotheses 11](#_Toc165902214)

[Significance of the Study 13](#_Toc165902215)

[Definitions of Key Terms 14](#_Toc165902216)

[Summary 17](#_Toc165902217)

[Chapter 2: Literature Review 20](#_Toc165902218)

[Key Areas in Building Digital Twin Representation of Foliage 22](#_Toc165902219)

[Digital Twin Representation of Foliage Theoretical Framework 38](#_Toc165902220)

[Data Ethics and Legal Frameworks in the study of Digital Twin Representation of Foliage 44](#_Toc165902221)

[Summary 49](#_Toc165902222)

[Chapter 3: Research Methodology 52](#_Toc165902223)

[Research Methodology and Design Process Diagram 52](#_Toc165902224)

[Research Methodology and Design (Nature of the Study) 52](#_Toc165902225)

[Population and Sample 53](#_Toc165902226)

[Materials or Instrumentation 55](#_Toc165902227)

[Operational Definitions of Variables 55](#_Toc165902228)

[Study Procedures 56](#_Toc165902229)

[Data Analysis 60](#_Toc165902230)

[Assumptions 62](#_Toc165902231)

[Limitations 62](#_Toc165902232)

[Delimitations 62](#_Toc165902233)

[Summary 63](#_Toc165902234)

[References 65](#_Toc165902235)

[Appendix A Annotated Bibliography 86](#_Toc165902236)

[Appendix B Topic Description and Supporting Literature 97](#_Toc165902237)

[Appendix C GitHub Details 99](#_Toc165902238)

**List of Tables**

Begin list of tables here…

**List of Figures**

[**Figure 1** *Smart City Digital Twin: Urban Planning and Green Spaces Integration* 2](#_Toc165902043)

[**Figure 2** *Digital Twin Representation of Foliage - Example* 8](#_Toc165902044)

[**Figure 3** *Flowchart: Digital Twin Representation of Foliage (AI-Driven Foliage Detection Using Machine Learning and Computer Vision)* 9](#_Toc165902045)

[**Figure 4** *Obstacles in 5G Millimeter-Wave Deployment* 24](#_Toc165902046)

[**Figure 5** *Vegetation Loss Simulation Results* 25](#_Toc165902047)

[**Figure 6** *Illustration of Collection of Data in Lidar and UAV* 27](#_Toc165902048)

[**Figure 7** *Elements of the Digital Twin Ecosystem* 31](#_Toc165902049)

[**Figure 8** *Some of the Key Ethical Considerations* 44](#_Toc165902050)

# Chapter 1: Introduction

In today’s rapidly advancing telecommunications landscape, the emergence of millimeter-wave (mmW) networks is a pivotal development aimed at fulfilling the surging demand for greater bandwidth, enhanced throughput, and minimized latency (Abdullah et al., 2020; Hong et al., 2021). This progression is vital for the progression of 5G wireless networks to meet the escalating needs of the mobile industry. However, mmW networks are challenged by issues like signal scattering, atmospheric absorption, and the obstruction caused by foliage and building structures; these are critical to navigating the successful roll-out of 5G networks, ensuring optimal coverage and data speeds (Barb et al., 2022; Bose et al., 2024; N. A. Khan & Schmid, 2024; Pradeep et al., 2021; Y. Zhang et al., 2019).

The process of accurately capturing foliage data, crucial (Lai et al., 2023) for the deployment of these high-frequency networks like mmW, has traditionally relied on costly and time-consuming methods such as Light Detection and Ranging (LiDAR) and unmanned aerial vehicles (UAVs) (Q. Chen et al., 2022; X. Deng et al., 2022; Mazzacca et al., 2022; Rogers et al., 2020). To adapt to the dynamic nature of our surroundings, the need for frequent updates of data renders traditional methods less viable for continuous application (Gaspari et al., 2022). However, the rise of digital twin technology offers a groundbreaking alternative (Attaran & Celik, 2023; Shahat et al., 2021).

Digital twins have become a game-changing strategy with applications spanning urban development and industrial operations, notably in the planning of wireless networks (T. Deng et al., 2021; Gabriele et al., 2023). One of the most compelling uses of digital twin technology is in the detailed modeling of foliage or vegetation. By constructing a virtual representation of the natural environment, focusing on the variety, distribution, and properties of plant life, digital twins afford network planners, environmental experts, and other stakeholders a nuanced understanding of how vegetation influences signal behavior, including path loss and network coverage, in high frequency mmW networks integral to 5G and 6G technology (L. U. Khan et al., 2022; Kuruvatti et al., 2022).

The sample image (Figure 1) depicting a digital twin representation of an urban area has been created for illustration purposes. It includes 3D models of buildings, streets, and foliage, presenting a clean and simplified city landscape that could be used for architectural visualization or city planning simulation. The design conveys a modern and futuristic tone, indicative of advanced urban planning and smart city concepts (T. Deng et al., 2021; Shahat et al., 2021).

**Figure 1**

*Smart City Digital Twin: Urban Planning and Green Spaces Integration*

A city with many tall buildings

Description automatically generated

*Note.* This image was generated with the assistance of Artificial Intelligence (AI).

## Statement of the Problem

Connected devices are becoming more common, and users are demanding higher bandwidth, throughput, and lower latency. This led to the development of mmW networks. The mmW band suffers from scattering, atmospheric absorption, canopy (or foliage), and building facades. Implementation of 5G requires mmW band propagation channel optimization (Farooq & Lokam, 2023; Pradeep et al., 2021; Y. Zhang et al., 2019).

Accurate modeling of foliage’s channel propagation is vital for wireless network design, particularly in diverse environments like rural, suburban, and urban settings. The blockage effects of foliage, especially at mmW frequencies, can be severe because of the comparable size of leaves and branches to the transmitted signal wavelength. Overcoming these challenges is crucial to developing reliable channel propagation models that effectively consider foliage’s impact on wireless communication systems (Anzum, 2021; Chikhale et al., 2022; Lai et al., 2023). Network operators must consider all these factors while deploying mmW technologies (5G, 6G) to improve user coverage and throughput.

Currently, foliage data is acquired using costly methods such as UAVs and LiDAR, requiring substantial physical effort (Q. Chen et al., 2022; Hematang et al., 2022; Mazzacca et al., 2022; Shen et al., 2023; Suhaizad et al., 2023). The continuous growth and transformation of foliage necessitates regular data collection to keep information current. The impracticality of repeating these tasks for regular foliage updates becomes clear because of their high cost, labor, and resource intensity. A more cost-effective and efficient approach involves leveraging Google Street View and satellite images in conjunction with state-of-the-art computer vision and machine learning models for object detection (Aikoh et al., 2023; Sun et al., 2023; Y. Zhao et al., 2023), presenting a promising way forward to address the challenges on collecting foliage or vegetation data.

As foliage is one of the main characteristics impacting the higher frequency like mmW network deployment, this study addresses the problem of providing foliage information by creating a digital twin (DT) of an environment with foliage with which network operators planning to deploy networks with higher frequency can use in their network planning to place the nodes at right locations for better coverage and user experience (Gabriele et al., 2023; Nguyen et al., 2021; Qi & Tao, 2018; Thuvander et al., 2022; D. Zhao et al., 2022).

## Purpose of the Study

The ~~purpose of this~~ study aims to develop a sophisticated digital twin that mirrors the physical environment, particularly integrating detailed foliage information (Lai et al., 2023; Pradeep et al., 2021; Y. Zhang et al., 2019). The digital twin will serve as a critical tool for network operators, enabling them to estimate the path loss attributed to foliage within the context of high frequency mmW network planning (Lai et al., 2023). Such estimations are pivotal for optimizing network performance and reliability in environments where vegetation can significantly impact signal propagation (Farooq & Lokam, 2023; Pradeep et al., 2021; Y. Zhang et al., 2019).

In order to accomplish this, a machine learning model based on computer vision will be used, which will be meticulously trained on a large dataset of foliage imagery. This model will employ advanced instance semantic segmentation techniques to identify and categorize foliage or vegetation within images. Through a combination of image segmentation, classification, and object detection methodologies, the study will dissect images into precise regions or objects (J. Chen et al., 2021; He et al., 2018; Sun et al., 2023; Y. Zhao et al., 2023).

This approach enables a pixel-level analysis of each scene, facilitating a deeper understanding of the vegetative elements within the digital twin environment (Jiang et al., 2023; Savelonas et al., 2022). The study will explore the nuanced interactions between vegetation and signal propagation, offering network operators a robust framework for mitigating the adverse effects of foliage on mmW network signals (De Beelde et al., 2023; Pradeep et al., 2021; Y. Zhang et al., 2019). This comprehensive approach representing foliage in a DT aims to bridge the gap between theoretical network planning and the practical challenges posed by natural vegetation, fostering more resilient and efficient communication networks in the face of environmental obstacles.

The study will utilize aerial and street view imagery from broad geographic areas in Philadelphia alongside LiDAR or UAV datasets (Research Natural Areas, 2023; OCM Partners, 2024; Philadelphia Lidar - LAS Files 2022 {2022} - Big Ten Academic Alliance Geoportal, 2022) for validating the model via grid-based assessments and Mean Intersection over Union (MIoU) for segmentation accuracy evaluation (Rezatofighi et al., 2019).

## Introduction to Theoretical or Conceptual Framework

The theoretical framework for employing digital twin technology in enhancing mmW network planning and deployment pivots around the Cross-Industry Standard Process for Data Mining (CRISP-DM) process model (Blume et al., 2020; Hayat Suhendar & Widyani, 2023). This framework is specifically tailored to address the unique challenges posed by foliage in urban and suburban environments, which can significantly impact mmW signal propagation due to its high frequency and susceptibility to attenuation by physical obstacles, such as trees and dense vegetation (Barb et al., 2022; De Beelde et al., 2023; Rogers et al., 2020). The digital twin representation of foliage, built upon the CRISP-DM framework, serves as a foundational tool for simulating and analyzing the interaction between mmW signals and urban foliage, facilitating optimized network infrastructure placement and configuration. This initial phase is crucial for delineating the scope and objectives of the mmW network planning project, with a specific emphasis on understanding how foliage impacts signal integrity and network performance (Lai et al., 2023). The aim is to leverage the digital twin to simulate real-world scenarios, thus enabling network engineers to preemptively identify and mitigate potential signal interference or blockage caused by vegetation. Identifying the specific needs, such as improving telecommunications infrastructure, enhancing urban green spaces, or optimizing environmental conservation efforts, will dictate the direction of the subsequent phases.

The second phase involves an initial data collection and familiarization process. For foliage digital twins, this entails gathering high-resolution aerial and street view imagery (Aikoh et al., 2023), LiDAR data, and any available UAV survey data (Q. Chen et al., 2022). Understanding the types, densities, and heights of foliage within the proposed network area is essential for assessing potential mmW signal attenuation or reflection issues. The following Data collection phase is Data preparation. This phase prepares the data for analysis, which may involve cleaning, selecting subsets, constructing data sets, annotating, and formatting data to suit the modeling needs (Dutta & Zisserman, 2019). Given the complexity of urban environments and the diverse data sources involved, this stage is critical for ensuring that the inputs to the machine learning models are of high quality and appropriately structured for detecting and analyzing foliage (J. Chen et al., 2021; He et al., 2018; Sun et al., 2023; J. Zhang et al., 2021; Y. Zhao et al., 2023). With the data prepared, various modeling techniques are applied to extract patterns and generate the digital twin representation. In the case of foliage, machine learning models such as convolutional neural networks (CNNs) or Mask R-CNN are employed to identify, classify, and analyze foliage from the aerial or street view imagery (J. Chen et al., 2021; He et al., 2018; Sun et al., 2023). This involves training models on annotated datasets, selecting the most effective models, and tuning parameters to optimize accuracy and performance (Rezatofighi et al., 2019).

Before proceeding to full-scale deployment, the models and their representations need to be evaluated against predefined success criteria, such as accuracy, reliability, and usability in practical applications (Rezatofighi et al., 2019). This could involve comparing the digital twin outputs with ground-truth data from LiDAR or UAV surveys and assessing the model's ability to represent foliage in various urban scenarios accurately. The final phase involves integrating the digital twin into the mmW network planning and deployment workflow. This enables planners and engineers to visualize signal propagation in the context of urban foliage, identify optimal equipment placement, and anticipate potential maintenance or signal-boosting requirements. The deployment also includes mechanisms for updating the digital twin with new data, ensuring it remains a relevant and effective tool for mmW network optimization (Rogers et al., 2020). By focusing on the unique challenges of mmW network planning in environments with significant vegetation, the CRISP-DM-based digital twin represents a targeted approach to enhancing network reliability and performance. Through detailed simulation and analysis of foliage interactions with mmW signals, network planners can make informed decisions that optimize coverage and capacity while minimizing interference and attenuation, thereby ensuring robust, high-speed wireless connectivity in urban and suburban settings.

## Introduction to Research Methodology and Design (Nature of the Study)

This section describes the research methodology and design for the study relating to the study problem, purpose, and research questions in constructing a Digital Twin Representation of Foliage. The constructive research design aims to bridge the gap between theoretical computer vision techniques and practical applications in digital twin technology for foliage representation. A constructive research design for the “Digital Twin Representation of Foliage” problem ensures a systematic approach to creating a practical solution that addresses the challenges of accurate foliage representation in digital twin models. Below (Figure 2) is a sample of a Digital twin representation of foliage (Z. Li et al., 2023; S. Song & Qin, 2022; Wilk et al., 2022).

**Figure 2**

*Digital Twin Representation of Foliage - Example*

A city with trees and buildings

Description automatically generated

*Note.* This image was generated with the assistance of Artificial Intelligence (AI).

**Figure 3**

*Flowchart: Digital Twin Representation of Foliage (AI-Driven Foliage Detection Using Machine Learning and Computer Vision)*

A screenshot of a computer

Description automatically generated

For future implementation

The current approach focuses on constructing a digital twin model to represent foliage in various environments, leveraging cutting-edge computer vision and machine learning techniques. The flowchart in Figure 3 outlines the process for constructing a Digital Twin model of foliage, starting with the region of interest as the initial input.

## Research Questions

The research questions aim to drive innovation and practical applications in the field of digital twin technology for foliage representation, fostering interdisciplinary collaboration and knowledge exchange among researchers, practitioners, and stakeholders. By addressing these research questions and hypotheses, the study aims to evaluate the feasibility, accuracy, and practicality of using digital twin technology for foliage representation, offering valuable insights for future research and practical applications in wireless network planning and environmental monitoring.

### RQ1

What extent can a digital twin representation of foliage, created using machine learning, computer vision, and image analysis techniques achieving an MIoU greater than 60% with LiDAR as the ground truth, effectively capture the spatial distribution and characteristics of foliage in natural environments?

### RQ2

What is the cost difference in obtaining foliage information using the integration of image analysis from street view and aerial view images with computer vision techniques compared to traditional survey methods like LiDAR or UAV?

### RQ3

What is the accuracy and performance of Digital Twin models concerning the area of foliage represented, as measured by Mean Intersection over Union (MIoU), compared to traditional LiDAR and UAV datasets?

### RQ4

How can digital twin models incorporating foliage representation contribute to more effective wireless network optimization in smart city environments by identifying suitable node placements?

## Hypotheses

With the use of computer vision-based machine learning methods and image analysis techniques on images collected through aerial (satellite) and street view images, DT models can be generated that provide information about foliage so that mmW networks can be deployed more intelligently and efficiently with better information about foliage.

### H10

There is no significant difference between the spatial distribution and characteristics of foliage as represented by a digital twin and as measured in actual field observations in natural environments.

### H1a

There is a significant difference between the spatial distribution and characteristics of foliage as represented by a digital twin and as measured in actual field observations in natural environments.

### H20

There is no significant cost difference between obtaining foliage information through the integration of image analysis from street view and aerial view images with computer vision techniques and obtaining the same information through traditional survey methods.

### H2a

The integration of image analysis from street view and aerial view images with computer vision techniques significantly reduces the cost of obtaining foliage information compared to traditional survey methods.

### H30

There is no significant difference in the accuracy and performance of Digital Twin models, as measured by the MIoU for the area of foliage represented, compared to traditional LiDAR and UAV datasets.

### H3a

There is a significant difference in the accuracy and performance of Digital Twin models, as measured by the Mean Intersection over Union (MIoU) for the area of foliage represented, compared to traditional LiDAR and UAV datasets.

### H40

There is no significant difference in the effectiveness of wireless network optimization in smart city environments between digital twin models incorporating foliage representation and traditional methods.

### H4a

The Digital twin models incorporating foliage representation significantly contribute to more effective wireless network optimization in smart city environments by identifying suitable node placements compared to traditional methods.

## Significance of the Study

The significance of this study on the digital twin representation of foliage primarily revolves around its pivotal role in advancing network planning and deployment strategies, especially pertinent to the challenges posed by urban environments on telecommunications infrastructure. This research is critical as it provides a novel approach to understanding and mitigating the impact of urban foliage on signal propagation, a significant concern for the deployment of high-frequency networks such as 5G and beyond (Barb et al., 2022; De Beelde et al., 2023; Lai et al., 2023; Pradeep et al., 2021; Y. Zhang et al., 2019). By creating virtual replicas of urban landscapes that accurately reflect the spatial distribution and physical characteristics of foliage (Attaran & Celik, 2023; T. Deng et al., 2021; L. U. Khan et al., 2022; Kuruvatti et al., 2022; Shahat et al., 2021), network engineers and planners can simulate and analyze how vegetation impacts network performance, leading to more informed decision-making and optimized network designs.

The purpose of this study is to provide a data-driven framework that can be used to improve the accuracy of predicting signal interference caused by foliage and there by identifying the suitable locations for mmW node placements which provides increased network coverage and user data connectivity (Abdullah et al., 2020; Pradeep et al., 2021; Y. Zhang et al., 2019). The research methodically applies machine learning and computer vision to create digital twins that serve as a sandbox for testing various network configurations and their interactions with urban greenery. This approach not only improves the reliability of network services in densely vegetated areas but also assists in identifying ideal locations for network infrastructure, minimizing environmental disruption and costs associated with physical trials (Rogers et al., 2020).

A number of studies highlight the potential for digital twins to contribute to more sustainable urban development practices (Attaran & Celik, 2023; T. Deng et al., 2021; Kuruvatti et al., 2022; Shahat et al., 2021). The interaction between urban green spaces and network infrastructure can help planners design strategies that protect and enhance vegetation while ensuring technological advancement. Keeping this balance is essential for smart cities of the future, since connectivity needs to be harmonious with conservation of the environment and aesthetics in the urban context (Pradeep et al., 2021; Y. Zhang et al., 2019). The research enriches the data science literature by highlighting an innovative application of digital twins, grounded in rigorous data analysis and modeling. It advances the telecommunications field by providing a novel tool for addressing one of the key challenges in network deployment (Bose et al., 2024; Pradeep et al., 2021), offering insights that are directly applicable to the design and optimization of next-generation wireless networks.

## Definitions of Key Terms

Here are definitions of key terms for the current study:

### Cross Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM provides a structured approach to data mining projects, ensuring that all necessary steps are followed to achieve successful outcomes (Blume et al., 2020; Hayat Suhendar & Widyani, 2023)

### Digital Twin (DT)

A virtual representation of a physical asset that closely mimics its real-world counterpart, including detailed information about its design, materials, components, and behavior, integrating IoT, AI, and real-time data to enable dynamic analysis and predictive modeling (Angin et al., 2020; Attaran & Celik, 2023; Azad et al., 2019).

### Foliage Representation

In the context of the current study, foliage representation focuses on the accurate digital depiction of vegetative elements and their impact on wireless signal. It is a comprehensive characterization and analysis of foliage in the context of radio signal propagation at 28 GHz in vegetated areas of typical residential environments. This representation involves measuring and analyzing the effects of foliage on signal propagation, discriminating between different propagation components such as scattering paths from trees and reflection paths from buildings and ground, calculating power ratios for these components, and developing models like the modified exponential decay (MED) model and the maximum attenuation (MA) model to quantify foliage attenuation and predict path losses through vegetated areas. Seasonal variations in foliage attenuation to understand how environmental changes impact path loss through vegetation (Ko et al., 2020).

### Path loss

Foliage attenuation is a significant factor affecting path loss in vegetated areas at higher frequencies like 28GHz. The reduction in power density of an electromagnetic wave as it propagates through space and interacts with natural obstacles like foliage (Ko et al., 2020).

### Geospatial Data

Data that provides information about the geographic location or spatial characteristics of objects, features, or events is typically represented by latitude and longitude coordinates (Cureton & Hartley, 2023; Rogers et al., 2020; Suhaizad et al., 2023).

### Hypothesis Testing

A statistical process is used to assess the validity of research hypotheses by evaluating whether observed data is consistent with the proposed hypotheses.

### LiDAR (Light Detection and Ranging)

A remote sensing technology that uses laser light to measure distances, providing highly accurate 3D information about the terrain, objects, and surfaces it interacts with (Q. Chen et al., 2022; X. Deng et al., 2022; Mazzacca et al., 2022).

### ~~Machine Learning Model~~

~~A computational system is trained to perform specific tasks or make predictions based on data, often utilizing algorithms that improve their performance over time (Hayat Suhendar & Widyani, 2023; Kapteyn & Willcox, 2020).~~

### Mask RCNN

A type of machine learning model, specifically a convolutional neural network, is used, for instance, in segmentation in computer vision tasks. It segments objects in images by delineating their boundaries (He et al., 2018; Sun et al., 2023; J. Zhang et al., 2021).

### ~~Precision, Recall, F1-Score, and~~ Intersection Over Union (IoU)

Performance metrics are used to evaluate the accuracy and effectiveness of machine learning models. Precision measures the proportion of true positive predictions, recall measures the ability to identify actual positives, the F1-score is a harmonic mean of precision and recall, and IoU measures the overlap between predicted and actual objects in segmentation tasks (Rezatofighi et al., 2019; J. Song et al., 2022).

### ~~Quality Assurance and Validation~~

~~Ensuring data integrity, accuracy, and reliability through systematic checks and validation procedures to confirm that the collected data and results are trustworthy.~~

### ~~Statistical Analysis~~

~~Applying statistical techniques and tests to analyze and interpret data, identify patterns, and draw meaningful conclusions from datasets.~~

### ~~Spatial Analysis~~

~~Examining the spatial distribution, relationships, and patterns of geographic features or data, often using geographic information system (GIS) tools and methods.~~

### UAV (Unmanned Aerial Vehicle)

An aircraft without a human pilot onboard is often equipped with cameras or other data collection and remote sensing sensors (Q. Chen et al., 2022; Hematang et al., 2022; Luo et al., 2023; Suhaizad et al., 2023; D. Zhao et al., 2022).

### ~~Visualization~~

~~Visual aids, such as charts, maps, graphs, or diagrams, can be used to represent and communicate data and analysis results in a comprehensible and informative manner.~~

## Summary

This constructive research design aims to bridge the gap between theoretical computer vision techniques and practical applications in digital twin technology for foliage representation. This multidisciplinary approach, combining remote sensing, computer vision, machine learning, and digital twin technology, offers a comprehensive method for accurately representing foliage in digital models, which is essential for the effective planning and deployment of next-generation wireless networks. The study seeks to offer a cost-effective, scalable, and accurate tool for urban planners and network engineers (Alkhateeb et al., 2023; Fett et al., 2023; Kuruvatti et al., 2022; Lehtola et al., 2022). The approach is grounded in rigorous data analysis, ethical considerations, and a clear acknowledgment of its scope and limitations, setting a foundation for future advancements in digital twin technology and its applications in smart city development and environmental monitoring.

The current research significance of digital twin representation of foliage, utilizing computer vision image analysis methods, compared to traditional approaches like LiDAR and UAV, stems from its capacity to overcome inherent challenges and constraints in conventional methodologies. Traditional techniques such as LiDAR and UAV surveys are often cost-prohibitive (Rogers et al., 2020), labor-intensive, and require extensive human involvement for data collection and processing (X. Deng et al., 2022; H. Li et al., 2021). Moreover, these methods could be more extensive in their coverage, resolution, and ability to maintain up-to-date information. In contrast, the digital twin representation of foliage harnesses advanced computer vision, AI, and machine learning techniques to analyze aerial and street view imagery. This approach offers several advantages, including cost-effectiveness, scalability, and the potential for real-time or near-real-time data updates (Attaran & Celik, 2023; Mylonas et al., 2021).

By automating foliage detection and analysis, digital twin representation enables swift and accurate data collection, facilitating more efficient network planning, urban development, and other applications. The impetus behind developing a DT representation of foliage arises from the escalating demand for precise and current foliage information across diverse sectors, encompassing telecommunications, urban planning, and environmental conservation. Industry reports and white papers underscore the critical role of digital twin technology in optimizing telecommunications infrastructure and enhancing service quality (Alkhateeb et al., 2023; L. U. Khan et al., 2022; Kuruvatti et al., 2022). A number of government initiatives, including ones aimed at sustainable urbanization and environmental stewardship, emphasize the use of digital twins to inform data-driven decision-making (Angin et al., 2020; T. Deng et al., 2021; Mylonas et al., 2021; Shahat et al., 2021).

In summary, traditional methods, like LiDAR and UAVs, will be significantly displaced by digital twins in urban and city planning. The benefits include improved data integration, faster iterations, sustainability, and smart city applications. AI and computer vision are driving the development of digital twins, which can be used to solve the challenges and limitations of conventional methods, offering more efficient, cost-effective, and scalable solutions.

# Chapter 2: Literature Review

The evolution of wireless communication networks towards higher frequencies, particularly mmW bands, presents unique challenges and opportunities in network design and optimization (Kaur et al., 2020). The demand for higher bandwidth, throughput, and lower latency, driven by the proliferation of connected devices, necessitates the development of mmW networks (Abdullah et al., 2020; Hong et al., 2021). These networks, however, are susceptible to various propagation losses, including scattering, atmospheric absorption, and, notably, the blockage effects of foliage and building facades (Barb et al., 2022; Bose et al., 2024; Ko et al., 2020; P. Zhang et al., 2020). Integrating 5G and the expected deployment of 6G technologies amplify the need for precise mmW band propagation channel optimization to ensure reliable and efficient network performance across diverse environments—rural, suburban, and urban (Anzum, 2021; Chikhale et al., 2022; Farooq & Lokam, 2023; Lai et al., 2023; Pradeep et al., 2021; Y. Zhang et al., 2019).

The current study proposes the development of a sophisticated digital twin model incorporating detailed foliage information to accurately reflect the physical environment. The critical need to estimate the path loss underscores this initiative attributed to foliage within high-frequency mmW network planning, a task that becomes pivotal for optimizing network performance and reliability in vegetation-impacted areas (Farooq & Lokam, 2023; Lai et al., 2023; Pradeep et al., 2021; Y. Zhang et al., 2019).

A computer vision model based on machine learning will be used to accomplish this goal by utilizing a large dataset of foliage imagery. Utilizing advanced instance semantic segmentation techniques, the model aims to identify and categorize foliage within images, employing a combination of image segmentation, classification, and object detection methodologies. This approach facilitates a pixel-level analysis of vegetative elements, enhancing the digital twin environment with a nuanced understanding of the interactions between vegetation and signal propagation. Such an endeavor not only aims to mitigate the adverse effects of foliage on mmW network signals but also strives to bridge the theoretical and practical gaps in network planning faced by natural vegetation challenges (J. Chen et al., 2021; De Beelde et al., 2023; He et al., 2018; Jiang et al., 2023; Savelonas et al., 2022; Sun et al., 2023; Y. Zhao et al., 2023).

The methodology will include the utilization of aerial and street view imagery complemented by LiDAR or UAV datasets from extensive geographic areas, including Philadelphia. These resources will support the model’s validation through grid-based assessments and Mean Intersection over Union (MIoU) metrics for segmentation accuracy evaluation (Research Natural Areas, 2023; OCM Partners, 2024; Philadelphia Lidar - LAS Files 2022 {2022} - Big Ten Academic Alliance Geoportal, 2022; Rezatofighi et al., 2019).

Because of the continuous growth and transformation of foliage, conventional methods of acquiring foliage data, such as UAVs and LiDAR, have high operating costs and require substantial physical effort to maintain current information. The proposed study aims to circumvent these challenges by leveraging more cost-effective and efficient alternatives, such as Google Street View and satellite images, in conjunction with advanced computer vision and machine learning models for object detection (Aikoh et al., 2023; Q. Chen et al., 2022; Hematang et al., 2022; Mazzacca et al., 2022; Shen et al., 2023; Suhaizad et al., 2023; Y. Zhao et al., 2023).

Ultimately, by creating a digital twin of an environment replete with foliage, this study seeks to equip network operators with a critical tool for deploying higher-frequency networks, such as those required for 5G and 6G technologies. This tool is intended to facilitate optimal node placement for enhanced coverage and user experience, addressing one of the main challenges—foliage impact—in mmW network deployment (Gabriele et al., 2023; Nguyen et al., 2021; Qi & Tao, 2018; Thuvander et al., 2022; D. Zhao et al., 2022). This literature review delineates the study’s comprehensive approach, positioning it at the forefront of addressing the complex interplay between natural vegetation and advanced wireless communication technologies.

## Key Areas in Building Digital Twin Representation of Foliage

In addressing the development of a digital twin representation of foliage for optimizing mmW network planning and performance, there are several key areas that need to be understood in depth. Each sub-heading listed below will delve into existing research, methodologies, technologies, and findings relevant to the creation and application of a sophisticated digital twin that integrates detailed foliage information for network planning purposes. The literature review provides a comprehensive overview of the state of the art in this domain, highlighting gaps that the current study aims to fill.

### Challenges in mmW Network Planning

The current 4G or LTE networks cannot handle increased demand from the user data throughput, coverage, and connected devices (Kaur et al., 2020). Adopting new frequency bands like mmW for the next generation 5G networks offers several advantages over traditional frequency bands below 6 GHz. These benefits include increased bandwidth for higher data rates, enhanced throughput for faster speeds, reduced latency for real-time communication, improved network performance, capacity expansion for IoT applications, spectrum efficiency, support for innovative use cases, and future-proofing networks for evolving technologies. Leveraging mmW technology enables operators to deliver high-speed, low-latency connectivity, accommodate a large number of devices, and pave the way for advanced services and applications in the wireless communication landscape (Barb et al., 2022; Kaur et al., 2020).

The mmW technology faces several challenges in 5G networks, including high propagation losses, limited range, obstruction sensitivity, fading and reflections, equipment and infrastructure requirements, and regulatory considerations. There are several challenges associated with mmW frequencies, including their susceptibility to atmospheric absorption, scattering, and penetration losses through obstacles like buildings and vegetation (Figure 4). The importance of studying foliage or vegetation in understanding path loss and optimizing the deployment of mmW technology in 5G networks. Vegetation plays a crucial role in causing attenuation and signal loss in mmW propagation due to scattering and absorption effects. By analyzing the impact of foliage on signal propagation, researchers can gain insights into the specific challenges posed by vegetation, such as signal penetration through leaves, branches, and tree structures (Abdullah et al., 2020; Barb et al., 2022; Chikhale et al., 2022; Farooq & Lokam, 2023; Y. Zhang et al., 2019).

Overcoming these challenges requires advanced technologies like beamforming and massive MIMO, as well as thorough propagation studies and network planning to optimize performance and ensure successful deployment of mmW networks in 5G systems (Barb et al., 2022; Bose et al., 2024).

**Figure 4**

*Obstacles in 5G Millimeter-Wave Deployment*

A diagram of waves and trees

Description automatically generated with medium confidence

*Note.* (Shah, 2018).

### Role of Foliage in mmW Wireless Network Planning and Deployment

In 5G millimeter-wave (mmWave) networks, foliage or vegetation plays a crucial role by causing signal attenuation through scattering and absorption effects. The presence of leaves, branches, tree trunks, and twigs in foliage environments can obstruct and scatter mmWave signals, leading to path loss and reduced signal quality. Understanding the impact of vegetation on signal propagation is essential for optimizing network planning, predicting coverage, and implementing mitigation strategies to address signal attenuation in foliated areas. Studying foliage characteristics and their effects on path loss enables network operators to enhance the performance and reliability of mmWave networks in outdoor environments with dense vegetation (Barb et al., 2022; Farooq & Lokam, 2023).

Understanding vegetation loss helps in quantifying the path loss experienced by mmW signals in foliated environments, enabling more accurate network planning and optimization strategies. By considering foliage information in propagation models and simulations, network operators can better predict signal coverage, identify areas with high path loss due to vegetation, and implement mitigation techniques to enhance signal quality and reliability in mmW deployments. Therefore, studying foliage characteristics and their effects on path loss is essential for effectively deploying mmW technology in 5G networks and ensuring optimal performance in outdoor environments with dense vegetation (Barb et al., 2022; Farooq & Lokam, 2023).

Foliage depth plays a crucial role in determining the optimal deployment of mmW networks and the coverage area in the presence of foliage obstacles. Vegetation loss rises precisely proportional to the depth of the vegetation, as shown by the graph (Figure 5). The findings suggest that foliage attenuation can pose challenges to mmW or 5G network deployment by limiting signal propagation and coverage range, thereby affecting the overall network performance and user experience. As foliage depth increases, the optimal coverage radius of the mmW network decreases. This decrease in coverage radius is attributed to the increase in losses caused by foliage attenuation (Bose et al., 2024).

The propagation loss due to foliage or vegetation in network planning or deployment is addressed by utilizing foliage loss models, such as the ITU-R model, to quantify the attenuation caused by vegetation based on frequency and foliage depth. By incorporating these models into network planning processes, operators can assess and mitigate the impact of foliage on signal propagation, optimize site selection, adjust antenna height, employ beam steering techniques, and conduct foliage mapping surveys to enhance coverage, reliability, and performance in mmW communication links deployed in vegetated areas (Farooq & Lokam, 2023).

**Figure 5**

*Vegetation Loss Simulation Results*

A graph of different colored lines

Description automatically generated

*Note.* (Barb et al., 2022).

### Traditional Datasets from LiDAR and UAV data

LiDAR technology offers a unique advantage in addressing foliage or vegetation challenges compared to photogrammetry with Unoccupied Aerial Systems (UAS). LiDAR can penetrate through vegetation to capture ground data, making it suitable for creating bare earth models or Digital Terrain Models (DTMs). Figure 6 below shows the collection of data using LiDAR and UAV methods. This capability allows LiDAR to effectively map terrain features even in areas with dense foliage or vegetation cover (Diab et al., 2022; Rogers et al., 2020). The benefits of LiDAR in comprehending foliage or vegetation include its sensitivity to vertical vegetation structure variations, enabling effective analysis in natural resources and forest applications. Additionally, LiDAR offers precise spatial information on vegetation shape and components, facilitating detailed foliage characterization and accurate tree species classification (Diab et al., 2022).

**Figure 6**

*Illustration of Collection of Data in Lidar and UAV*

A diagram of a gps system

Description automatically generated

*Note.* (Airborne Laser Scanning, 2013).

But that said, LiDAR data comes with many challenges. Challenges involve the computational intensity of LiDAR data processing, the complexity of interpreting large LiDAR point cloud datasets in dense vegetation environments, and factors like sensor calibration and data noise affecting data quality and analysis accuracy (Diab et al., 2022). The LiDAR data can be complex and may require specialized knowledge, software tools, and computational resources. LiDAR data processing typically involves several steps, including data cleaning, point cloud classification, feature extraction, and model generation (Rogers et al., 2020).

Given the high volume of point cloud data generated by LiDAR sensors, processing can be computationally intensive and time-consuming. Users need to have a good understanding of LiDAR data structures, processing workflows, and software tools to effectively analyze and interpret the data. The limitations of LiDAR data include the high data requirements, labor-intensive and time-consuming data collection process, and challenges in processing due to factors such as data volume, model complexity, and generalizability (Lu & Jiang, 2024).

As a result of the massive amount of 3D data points, complex geospatial algorithms, and diverse parameter selection involved in tasks like point classification, ground point extraction, and generating final products such as Digital Elevation Models (DEMs) and 3D building models, LiDAR data processing is highly complex (Z. Li et al., 2018; Z. Wang & Menenti, 2021). This complexity is further exacerbated when dealing with large-scale LiDAR data sets that can reach tens of Terabytes in volume, stressing the storage and computational limits of a single computer (Z. Li et al., 2018). The other major significant limitation of LiDAR technology is its cost. Traditional airborne LiDAR systems can be expensive to acquire and operate, making them less accessible for some research projects or applications with budget constraints (Rogers et al., 2020).

Unmanned aerial systems (UAS) offer key benefits for characterizing vegetation complexity, including high spatial and temporal resolution for detailed monitoring, the ability to assess species diversity and map individual species, analyze ecosystem structure for biomass and stand complexity, integrate sensors for 3D information capture, and quantify stand complexity and volume. These advantages make UAS valuable tools for enhancing vegetation monitoring, biodiversity assessment, and ecosystem analysis (Müllerová et al., 2021). This dataset aims to enhance safety by detecting small and flying UAVs, support object detection tasks, facilitate the development of data-driven models for accuracy, ensure universality and robustness by including various objects, and enable benchmarking of advanced detection models. Overall, UAV Data aims to advance research in UAV detection, object detection, and machine learning fields (Zeng et al., 2021).

Like LiDAR, in the case of the UAV system, there are challenges, including regulatory constraints related to UAV operation, the cost of equipment such as specialized sensors, the complexity of data processing and analysis, the need to integrate UAV data with ecological processes, and staying updated on technological advancements. A collaborative approach between plant ecologists and remote sensing professionals will help overcome these challenges and maximize the benefits of UAVs in plant ecology research (Sun et al., 2021).

The integration of UAV-LiDAR and high-resolution optical data enhances the mapping of complex vegetation communities by improving classification accuracy, providing comprehensive data fusion, enabling better species discrimination, increasing spatial resolution, and streamlining data processing workflows. This integration allows for a more detailed and accurate understanding of ecosystem dynamics and species distribution in challenging environments like upland swamps (Banerjee et al., 2019).

### Digital Twins

The concept of digital twins has a rich history that dates to NASA’s Apollo program, where a digital twin of the spacecraft was effectively used during the Apollo-13 mission to simulate and resolve a life-threatening situation (Arnas et al., 2024). This historical event highlighted the power of digital twins in crisis management and problem-solving. Over the years, advancements in sensor technology and computational power have propelled the concept of digital twins forward, enabling their application in various fields beyond aerospace. Today, digital twins play a crucial role in environmental science, medicine, and engineering, revolutionizing how complex systems are understood, monitored, and optimized (Arnas et al., 2024; Kukushkin et al., 2022; J. Song & Le Gall, 2023).

Digital Twins are virtual representations of physical objects or systems that enable seamless data integration between the physical and virtual realms (National Academies of Sciences, Engineering, and Medicine, 2024). Digital Twins find applications in manufacturing, Industry 4.0, production planning, supply chain management, construction, and healthcare (Kukushkin et al., 2022; J. Song & Le Gall, 2023). In manufacturing, they enable predictive maintenance, performance optimization, and customer-centric product design. In healthcare, Digital Twins aid in disease detection, treatment experimentation, and surgical preparation through accurate human body modeling (Attaran & Celik, 2023; T. Deng et al., 2021).

Digital Twins provides companies with the ability to quickly detect and solve physical problems, design and build better products and realize value faster than before. They facilitate real-time monitoring, predictive analytics, automation, and intelligent decision-making, leading to enhanced operational efficiency and innovation. Figure 7 below depicts the digital twin ecosystem.

Digital twins are helping city planning by providing a mirrored digital representation of the city, enabling stakeholders to visualize and plan various aspects effectively. They enhance the visualization and planning of cities by incorporating detailed information on buildings and city elements. Integration of Building Information Modeling (BIM) and Geographic Information System (GIS) applications maximizes the potential of digital twins in city planning. Real-time data updates and efficient 3D model processing reduce the time between data updates and model adjustments, facilitating better decision-making in urban planning processes (Shahat et al., 2021).

**Figure 7**

*Elements of the Digital Twin Ecosystem*

A black and white card with text

Description automatically generated

*Note.* (National Academies of Sciences, Engineering, and Medicine, 2024).

In the telecommunication industry, Digital Twins play a crucial role in network planning and deployment. By representing foliage in Digital Twins, telecom companies can simulate the impact of vegetation on signal propagation, enabling more accurate network planning and optimization. This representation helps in identifying potential signal blockages, optimizing antenna placements, and improving network coverage and performance (Attaran & Celik, 2023). They enable operators to visualize network infrastructure, predict performance, and proactively address issues to ensure optimal network operation (T. Deng et al., 2021). By representing foliage in a digital twin, telecommunication companies can simulate the impact of vegetation on network signals. This representation helps in optimizing antenna placement for better coverage, reducing signal interference, improving network performance, and enhancing overall connectivity in urban environments (T. Deng et al., 2021).

Overall, digital twins facilitate network planning by providing a realistic virtual environment for testing and optimizing network configurations, including foliage modeling, to ensure optimal performance in real-world scenarios (Kuruvatti et al., 2022).

### Machine Learning and Computer Vision in Environmental Modeling

Artificial Intelligence (AI) and Machine Learning (ML) technologies have revolutionized many fields, including the study of vegetation and foliage. This transformation has been driven by the exponential growth of computing power and the widespread availability of low-cost memory, which have helped process large datasets that were previously unmanageable (Sarirete et al., 2022). With this computational leap, researchers and technologists have been able to develop AI/ML computer vision image analysis algorithms that can identify and classify different types of vegetation with remarkable accuracy using satellite imagery, aerial photographs, and ground-level observations (Raihan, 2023; Sarirete et al., 2022).

AI/ML takes advantage of these technological advancements by employing sophisticated models that learn from vast amounts of data, identifying patterns and features that are often imperceptible to the human eye. By utilizing semantic segmentation and object detection techniques, machine learning models can distinguish various species of trees and other vegetation.

Furthermore, AI/ML models can process and analyze data in real time, providing immediate insights into vegetation conditions and facilitating proactive decision-making. With AI/ML advancements and increased computational power and storage, these applications may become even more sophisticated, offering even more sophisticated tools for monitoring and managing the environment.

Machine learning plays a crucial role in identifying foliage or vegetation by enabling the development of predictive models for semantic segmentation (Sun et al., 2023). These models are trained using features such as geometric characteristics and height-based information to distinguish vegetation from other elements in the dataset. By utilizing machine learning algorithms like Random Forest and semantic segmentation, researchers can accurately filter out vegetation and focus on identifying archaeological structures hidden beneath the foliage. The integration of machine learning enhances the efficiency and accuracy of vegetation filtering (Mazzacca et al., 2022).

By processing images captured from drones or satellites, these computer vision and image analysis techniques can analyze various visual features such as color, texture, and shape to differentiate between different types of vegetation. Machine learning algorithms are employed to classify vegetation types, detect changes over time, and monitor environmental conditions. This application of computer vision and image analysis in vegetation identification offers a non-invasive and efficient way to study and manage ecosystems, agricultural lands, and natural resources (He et al., 2018; Savelonas et al., 2022).

### Data Acquisition Techniques for Digital Twin Development

The limitations of UAVs discussed in the article include regulatory constraints that vary by region, making it challenging to obtain permission for UAV operation; the cost of specialized equipment like hyperspectral cameras and LiDAR systems; the complexity of data processing and analysis; the need to extract more ecologically meaningful parameters from UAV data; and the necessity for plant ecologists to stay updated on technological advancements in UAV technology and remote sensing techniques (Sun et al., 2021).

Traditional airborne LiDAR systems can be expensive to acquire and operate, making them less accessible for projects with budget constraints. The cost of LiDAR data collection, equipment, and processing may pose a financial challenge for some users or research endeavors (Rogers et al., 2020).

LiDAR data collection can be limited by factors such as flight endurance, payload capacity, and operational range, especially when using LiDAR sensors mounted on Unoccupied Aerial Systems (UAS). These limitations may impact the efficiency and coverage of LiDAR data collection missions, particularly in terms of the size of the study area that can be effectively surveyed (Rogers et al., 2020).

Processing LiDAR data can be complex and time-consuming, requiring specialized software, expertise, and computational resources. Analyzing and interpreting LiDAR point cloud data involves data cleaning, classification, feature extraction, and model generation, which can be challenging for users without the necessary skills or resources (Rogers et al., 2020). These limitations underscore the practical challenges associated with utilizing LiDAR data for geospatial applications and emphasize the importance of addressing cost, collection, and processing constraints to effectively leverage the benefits of LiDAR technology in research and analysis (Rogers et al., 2020).

The use of digital twins in the telecommunications sector has become essential for providing a detailed model of physical network components and behaviors. It allows operators to monitor, simulate, and manage their networks efficiently, resulting in heightened efficiency, fewer outages, and better strategic decisions (L. U. Khan et al., 2022). In the field of network design and vegetation analysis for advanced 5G, 6G, and mmW communications, digital twins offer an excellent means of facilitating accurate modeling and simulations. Moreover, they facilitate rapid and cost-effective experimentation and accelerate the deployment of new technologies, making them a crucial component of an effective network planning process (Kuruvatti et al., 2022; Nguyen et al., 2021).

In the context of representing foliage in a digital twin for network planning, the virtual model can incorporate detailed information about the surrounding environment, including trees, buildings, and other obstacles that may impact signal propagation and coverage. By accurately modeling foliage in the digital twin, network planners can simulate how radio waves interact with different types of vegetation and optimize network design to mitigate signal attenuation and interference caused by foliage (L. U. Khan et al., 2022; Kuruvatti et al., 2022).

Furthermore, by leveraging advanced computer vision, image analysis, and machine learning algorithms, digital twins can analyze the impact of foliage on network performance, predict potential coverage gaps or signal blockages, and recommend strategic placement of network elements such as antennas and base stations to optimize signal propagation and ensure reliable connectivity in foliage-rich environments (L. U. Khan et al., 2022).

Overall, integrating foliage representation in digital twins for telecommunications network planning can help operators enhance network coverage, capacity, and quality of service by proactively addressing challenges posed by natural elements like trees and vegetation (Gabriele et al., 2023; L. U. Khan et al., 2022).

The representation of foliage or vegetation in Digital Twins is a key aspect of urban development projects. By utilizing automated procedural workflows, developers can create realistic representations of vegetation in Virtual Reality (VR) environments, allowing stakeholders to visualize and interact with green structures in urban settings. This approach enhances the engagement of citizens in the planning process, especially in areas where new developments are built in nature or forest areas (Thuvander et al., 2022).

Furthermore, Digital Twins is increasingly being used in networking planning and deployment processes. By integrating real-time sensor data, physical models, and simulations into Digital Twins, stakeholders can gain valuable insights into network performance, optimize resource allocation, and improve decision-making in urban infrastructure development. The use of Digital Twins in networking planning and deployment allows for more efficient and effective management of resources, leading to sustainable and smart city developments (Thuvander et al., 2022).

Using low-cost satellite or aerial imagery data sources like Google Imagery, applying computer vision, image analysis, and semantic segmentation machine learning techniques to extract the relevant information related to foliage or vegetation from the imagery, we can build the Digital twin representation of foliage.

### Research Gaps, Current Implementations and Future Directions

As foliage or vegetation is one of the factors impacting the advanced mmW, 5G, and 6G technologies deployment, current methods of identifying foliage for network planning, such as LiDAR and UAV, are noted for their high cost and labor-intensive nature, which limits the flexibility for frequent updates. These methods also struggle to adapt quickly to changes in foliage due to seasonal variations or environmental shifts, posing significant challenges in maintaining up-to-date data for accurate network planning.

There is a promising avenue for future research involving the use of advanced computer vision and image analysis techniques on more accessible and cost-effective data sources like satellite, aerial images, and Google Street View. The development of digital twins that incorporate detailed, regularly updated foliage information can significantly enhance network planning. Such digital twins would allow for more precise assessments of foliage’s impact on signal propagation and path loss, leading to improvements in network coverage and throughput (J. Song & Le Gall, 2023).

The current DTs can be expanded in the future to include various other components that are affecting mmW, 5G, and 6G technologies, like street furniture (stop signs, sign boards, etc.) and building facades.

### Databases Accessed and Other Sources

In the current research study, to ensure a comprehensive and up-to-date review of the literature, the following sources, databases, and search engines are used: 1) NU library, 2) IEEE Xplore, 3) Google Scholar, 4) Science Direct, and 5) Springer Link.

The following sources are used to collect the aerial (satellite) images and street view images:

1. Maps Static API from Google (Google Maps Platform Documentation | Maps Static API | Google for Developers) (Google for Developers Maps Static API, n.d.).
2. Street View Static API from Google (Google Maps Platform Documentation | Street View Static API | Google for Developers) (Google for Developers Street View Static API Overview, n.d.).
3. Google Earth (<https://earth.google.com/web/)>
   1. The following open source, Open Street Map (OSM), Java Open Street Map (JOSM), is used to collect information on streets, roads, and building outlines.
4. Open Street Map (<https://www.openstreetmap.org/>)
5. Java Open Street Map (<https://josm.openstreetmap.de/>)

## Digital Twin Representation of Foliage Theoretical Framework

A digital twin refers to a digital replica of physical entities and processes, integrating IoT, AI, and real-time data to enable dynamic analysis and predictive modeling (Caprari et al., 2022; Huang et al., 2021). In the context of this study, using satellite and street view imagery, AI/ML computer vision, and image analysis techniques, the extracted foliage representation focuses on the accurate digital depiction of vegetative elements and their impact on wireless signal propagation. Using the Digital Twin model, network planners can describe how electromagnetic waves reduce in power density as they propagate through space and interact with natural obstacles (Almasan et al., 2022). The digital twin of an environment, including detailed foliage information, allows for the simulation of path loss due to foliage (Sabri et al., 2015), aiding in the optimization of network planning and performance. Enhanced foliage representation within a digital twin framework directly relates to more accurate estimations of network signal degradation and assists in designing more effective communication infrastructures by identifying suitable network node placement locations.

### Assumptions and Propositions

Integrating detailed and dynamic foliage information into digital twins can significantly enhance the accuracy of simulations reflecting real-world environmental impacts on signal propagation. Leveraging enhanced computational power and advanced machine-learning techniques improves the precision with which digital twins represent and process complex environmental data. This approach is grounded in the assumption that richer environmental data enables more accurate simulations of physical phenomena, particularly in how they affect technologies such as communication signals. Further, studies support the proposition that detailed environmental modeling enhances predictive accuracy in digital twin applications (Correia et al., 2023; Kang & Mo, 2024; Thelen et al., 2022).

### Origin and Development of the Framework

The theoretical framework for the digital twin representation of foliage is a sophisticated blend of environmental science, network engineering, and digital simulation technologies. The origins of this framework can be traced back to the broader concept of digital twins, which first emerged in the early 2000s. Notably, the idea was pioneered by Michael Grieves in 2003 during his tenure at the University of Michigan, initially focusing on manufacturing and aerospace sectors (VanDerHorn & Mahadevan, 2021). In these fields, digital twins served as virtual replicas of physical systems, providing ground for simulations that could predict mechanical wear and tear and process optimization without the risks and costs associated with physical testing.

Historically, digital twins have been associated with NASA's Apollo programs. In the Apollo 13 mission, NASA effectively used a digital twin of the spacecraft to simulate and resolve a life-threatening situation. It highlights the importance of using digital twins to manage crises and solve problems (Arnas et al., 2024). Over the years, this concept expanded into industrial applications, focusing on the predictive maintenance and operation of complex systems. More recently, the adaptation of digital twins in network planning, particularly integrating environmental factors such as foliage, is seen as an evolution in the precise design and deployment of telecommunications networks.

Over the years, the scope of digital twins has broadened significantly beyond its initial industrial confines. This expansion has been driven by technological advancements on the Internet of Things (IoT), artificial intelligence (AI), and big data analytics. These technologies have enabled the creation of more complex and dynamic digital twins capable of real-time data analysis and simulation. As such, the concept has been increasingly applied in fields like environmental science and network planning, marking a significant evolution from its industrial roots.

The intersection of digital twins with environmental science and network engineering reflects an increasingly interdisciplinary approach to technological applications. In environmental science, digital twins are used to model complex ecosystems or urban environments, allowing for detailed simulations of natural processes and human interactions. This capability is particularly valuable for studying the impact of various factors on foliage and vegetation, such as climate change, urban expansion, and ecological conservation efforts.

In the context of network engineering, digital twins facilitate the planning and optimization of telecommunications networks. They help in understanding how natural elements like foliage can affect signal propagation, particularly in the development of advanced communication technologies such as 5G and 6G networks. By simulating different scenarios, network engineers can anticipate potential issues and plan infrastructure, accordingly, enhancing network reliability and performance (Hui et al., 2023).

The convergence of these disciplines within the digital twin framework allows for a more comprehensive analysis and a better understanding of the interactions between technology and the environment. This interdisciplinary approach not only broadens the applicative landscape of digital twins but also enhances their practicality in tackling complex, real-world problems.

As the application of digital twins continues to evolve, they are expected to play a crucial role in sustainable development and smart city initiatives. By integrating environmental data with digital simulation and network planning, digital twins offer a powerful tool for policymakers, scientists, and engineers to make informed decisions that balance technological advancement with environmental conservation and societal needs. This ongoing development underscores the dynamic nature of the digital twin framework, highlighting its potential to adapt and grow in response to new challenges and technological opportunities (Alkhateeb et al., 2023; Hu et al., 2021; L. U. Khan et al., 2022; Nguyen et al., 2021).

### Relevant Studies

Several studies have explored the use of digital twins for urban planning and environmental management (Caprari et al., 2022; Schrotter & Hürzeler, 2020). For instance, the integration of digital twins with urban greenery to assess environmental benefits highlights a similar application, though not specifically targeting telecommunications.

AI technologies enhance the capabilities of digital twins in smart manufacturing by enabling efficient data analysis, improving predictive maintenance, optimizing production processes, enhancing decision-making, and facilitating personalized production. By leveraging AI algorithms, digital twins can analyze data effectively, predict maintenance needs accurately, optimize processes in real time, provide actionable insights for decision-making, and customize production to meet specific customer requirements (Huang et al., 2021).

DTs with foliage facilitate assessing path loss (Ko et al., 2020; P. Zhang et al., 2020) and thereby facilitate network planning, allowing for resource reservation reconfiguration to adapt to dynamic network conditions. Overall, DTs enhance network planning by providing granular insights, adaptability, and intelligence to optimize resource management and improve network performance in advanced networks like mmW/5G/6G (Hui et al., 2023; Zhou et al., 2022).

### Alternative Frameworks

While GIS is potent for spatial analysis and has been used extensively for environmental data integration, it lacks the dynamic and predictive capabilities of digital twins (Chaminé et al., 2021). Traditional simulation models offer detailed analyses but do not support the real-time updating and integration capabilities of digital twins (Angin et al., 2020; Capecchi et al., 2023; Hematang et al., 2022; Rogers et al., 2020; Z. Wang & Menenti, 2021).

The digital twin framework was chosen due to its ability to dynamically integrate and analyze complex datasets, including real-time environmental changes, which is essential for accurate network planning and adaptation in response to environmental changes (Hui et al., 2023).

The Digital Twin, at its core, replicates the physical network and uses machine learning techniques to process network state descriptions and output relevant metrics. Unlike traditional network modeling tools, DTs take a data-driven approach by training models directly with real network data, enabling real-time operation and accurate representation of complex network environments. This approach allows for more efficient network control and management at shorter timescales, making DTs a more advanced and dynamic solution for modern communication networks (Almasan et al., 2022).

### Framework Relevance to the Study

The digital twin framework guides this study by providing a structured approach to simulate and analyze the impact of foliage on network performance. This directly relates to and supports the development of the problem statement (the impact of foliage on mmW networks), purpose statement (development of a digital twin for network planning), and research questions (What extent can a digital twin representation of foliage capture the characteristics and spatial distribution of foliage in natural settings?).

By utilizing the digital twin framework, this study aims to bridge the gap between theoretical network optimization and practical, real-world applications, ensuring that network deployments are both resilient and efficient against environmental challenges (Almasan et al., 2022; Hui et al., 2023; J. Song & Le Gall, 2023). DTs can leverage the power of data-driven modeling to enhance network performance evaluation and drive innovation in the field of networking (Hui et al., 2023).

## Data Ethics and Legal Frameworks in the study of Digital Twin Representation of Foliage

In creating digital twins, Google Street View and Aerial imagery are widely used. There are several ethical considerations to be considered. Collection, processing, and dissemination of imagery that may include identifiable features of individuals’ properties are central to these concerns. The ethical use of data raises issues of privacy, autonomy, and control over personal information.

***Data Ethics and Privacy Concerns in the Collection and Use of Google Imagery Data***

Images captured by Google’s aerial and street view images provide detailed visual information about private residences, businesses, and public places, sometimes including images of identifiable individuals. In order to handle this data ethically, it must take individuals’ privacy rights into account, ensuring that imagery is used in a way that does not violate privacy or misuse it. Some of the key ethical considerations outlined in Figure 8.

**Figure 8**

*Some of the Key Ethical Considerations*

A diagram of information on a computer

Description automatically generated with medium confidence

Privacy is a priority for Google. In order to protect the privacy of individuals, Street View imagery is blurred to obscure faces and license plates. Consider privacy concerns when using Street View imagery provided by users and adhere to Google’s guidelines (Google-Contributed Street View Imagery Policy, n.d.).

***Transparency and Consent***

Transparency in how data is collected, what it is used for, and who has access to it is vital. For imagery used in digital twins, stakeholders should be informed about the data’s scope and purpose. Most Google’s mapping products (such as Google Maps, Google Earth, and Street View) do not require a user request. The use of these products is permitted if you adhere to their Terms of Service and guidelines (Brand Resource Center | Products and Services - Geo Guidelines, n.d.).

Adding annotations to maps (points, lines, labels) will allow you to personalize them. However, it is important not to alter the appearance of Google Maps, Google Earth, or Street View in any significant way. The product interface colors cannot be changed, or attributions cannot be removed (Brand Resource Center | Products and Services - Geo Guidelines, n.d.)

***Data Security***

Ensuring the security of data to prevent unauthorized access and use is a crucial ethical obligation. This includes implementing strong data encryption, access controls, and regular audits to safeguard the data throughout its lifecycle. In the current study, only objects of interest (foliage/vegetation) coordinates like latitude and longitude are stored.

***Social Justice Implications***

The use of aerial and street view imagery in the context of digital twins necessitates a careful consideration of social justice implications, including:

Equity. It is critical to ensure that the advantages derived from digital twins, such as enhanced urban planning and more effective environmental monitoring, are equitably shared among various socio-economic groups. This involves proactive measures to ensure that no community is left behind or disproportionately benefits from the technology.

Non-discrimination: Measures must be in place to prevent the misuse of data in ways that could discriminate against vulnerable or marginalized groups. For example, data collection and analysis methods should be designed to avoid unintentionally targeting or excluding particular demographics based on factors like location, economic status, race, or ethnicity.

**Community Impact.** The process of data gathering should include an evaluation of its impact on local communities. This involves not only assessing potential disruptions or negative effects but also actively engaging with community stakeholders. This engagement is crucial for addressing any concerns, securing community support, and ensuring that the project aligns with the interests and needs of those it affects.

***Regulatory Compliance***

Complying with local and international laws regarding data privacy, such as General Data Protection Regulation (GDPR) in Europe or Californica Consumer Privacy Act (CCPA) in California, is essential. These regulations provide a framework for addressing many of the ethical concerns related to privacy and data protection.

**Ethical Guidelines and Frameworks.** EU-U.S. and Swiss-U.S. Data Privacy Frameworks: According to Data Privacy Framework (DPF) (Data Privacy Framework) certification (Data Privacy Framework, n.d.), Google adheres to the EU-U.S. and Swiss-U.S. DPF and the UK Extension to the EU-U.S. DPF, as outlined by the US Department of Commerce regarding the collection, use, and retention of personal information from the EEA, Switzerland, and the UK. According to the DPF Principles, Google LLC (and its wholly owned US subsidiaries, unless explicitly excluded) adheres to them. Any personal information shared by Google with third parties under the Onward Transfer Principle for external processing on our behalf remains Google’s responsibility (Privacy and Terms – Google, n.d.).

**Belmont Report.** Belmont Report’s principles (respect for persons, beneficence, and justice) provide a moral framework that emphasizes the protection of individuals and the equitable distribution of benefits and burdens (Nagai et al., 2022).

**Association for Computing Machinery (ACM).** ACM Code of Ethics and IEEE Standards offer guidelines and best practices for professionals handling data, ensuring ethical decision-making that aligns with broader societal values (The Code Affirms an Obligation, n.d.).

***Transfer of Data***

As Google maintains servers around the world, information may be processed on servers located outside of your country of residence. There are differences in data protection laws among countries, with some providing more protection than others. The Privacy Policy applies the same protections regardless of where your information is processed (Privacy and Terms – Google, n.d.). Additionally, Google complies with certain legal frameworks related to data transfer (Data Transfer Frameworks – Privacy and Terms – Google, n.d.), such as those described in: a) European Commission adequacy decisions (Data Protection Adequacy, n.d.), b) UK adequacy regulations (A Guide to International Transfers, 2024), and c) Swiss adequacy decisions (Federal Data Protection and Information Commissioner (FDPIC), n.d.).

***Addressing Bias and Ensuring Fairness***

The algorithms that process these images should be designed so that biases do not persist or become more pronounced. In order to ensure algorithms, perform equally across a variety of geographic and demographic scenarios, diverse data sets must be used during the training and validation of models (Bromm et al., 2020).

To mitigate bias and ensure fairness, it is critical that the algorithms used to process imagery from platforms such as Google’s aerial and street view are carefully designed to avoid the perpetuation or exacerbation of existing biases. This involves employing algorithms that perform consistently across various geographic and demographic contexts, necessitating the use of diverse datasets during the training and validation phases of model development.

Furthermore, integrating social justice principles with data ethics requires an integrated approach that extends beyond mere legal compliance to consider the wider societal impacts and the rights of individuals. This comprehensive approach ensures that the use of technologies like Google’s aerial and street view imagery not only advances in terms of technical capabilities but also aligns with higher standards of ethical responsibility and social (Fairness: Types of Bias, n.d.; Overview – Google Earth, n.d.; “Responsible AI at Google,” n.d.).

By adopting such rigorous frameworks, researchers and technologists can work toward the development of technologies that are not only innovative but also inclusive and fair, promoting a more equitable distribution of technology’s benefits.

## Summary

The theoretical framework for digital twin representation of foliage, emphasizing the concept of a digital twin as a virtual replica of physical entities and processes that integrates IoT, AI, and real-time data for dynamic analysis and predictive modeling. The study aims to utilize satellite and street view imagery, AI/ML computer vision, and image analysis techniques to extract accurate foliage representations within the digital twin model, enabling a deeper understanding of how vegetation impacts wireless signal propagation (Kukushkin et al., 2022; J. Song & Le Gall, 2023).

Digital twins contribute to more efficient engineering processes by optimizing operations, enabling predictive maintenance, facilitating resource management, supporting design, and testing through virtual prototyping, and enhancing data-driven decision-making. By creating virtual representations of physical assets, engineers can monitor performance, anticipate maintenance needs, analyze resource usage, streamline design processes, and make informed decisions based on real-time insights. This technology helps improve productivity, reliability, and cost-effectiveness in engineering operations (National Academies of Sciences, Engineering, and Medicine, 2024; Shahat et al., 2021).

The research further explores the significance of the study in advancing network planning and deployment strategies, particularly in smart city environments. By creating virtual replicas of urban landscapes that accurately reflect the spatial distribution and physical characteristics of foliage, network engineers and planners can simulate and analyze how vegetation influences network performance, resulting in more informed decision-making and optimized network designs (Barb et al., 2022; Bose et al., 2024; Farooq & Lokam, 2023). Through machine learning and computer vision, digital twins are created to test a variety of network configurations as they interact with urban greenery, ultimately improving network reliability in densely vegetated areas by applying machine learning and computer vision (Raihan, 2023; Sarirete et al., 2022; Sun et al., 2023).

Additionally, the research delves into ethical and legal frameworks for digital twin foliage representation. To ensure compliance with data protection laws, intellectual property rights, and privacy laws, it emphasizes the importance of ethical considerations in data collection, analysis, and utilization within the context of digital twin technology (Data Privacy Framework, n.d.; Data Transfer Frameworks – Privacy and Terms – Google, n.d.; Privacy and Terms – Google, n.d.). Considering these ethical and legal considerations, the study aims to promote responsible and sustainable technology use in developing digital twins of foliage.

With a focus on foliage, the research combines advanced technologies and environmental understanding to create a nuanced digital replica of natural environments. Using digital twin models, network planners, environmental scientists, and stakeholders can better understand how foliage influences wireless signal propagation dynamics. A special emphasis is placed on modern connectivity challenges and the deployment of millimeter-wave (mmW) networks in this study, which will contribute greatly to wireless communication technology advancement (Barb et al., 2022; Farooq & Lokam, 2023).

In summary, literature review provides a comprehensive overview of the theoretical foundations, technological aspects, and ethical considerations essential for constructing a digital twin representation of foliage. By synthesizing existing literature and theoretical frameworks, the study sets the stage for the subsequent research methodology and design, laying the groundwork for the innovative exploration of digital twin technology in the context of foliage representation and wireless signal propagation.

Checklist:

Briefly restate the key points discussed in the chapter. Review the headings and/or table of contents to ensure all key points are covered.

Highlight areas of convergence, divergence, and gaps in the literature supporting the study's need. This discussion should logically lead to Chapter 3, where the research methodology and design will be discussed.

# Chapter 3: Research Methodology

Begin writing here…

Checklist:

Begin with an introduction and restatement of the problem and purpose sentences verbatim.

Provide a brief overview of the contents of this chapter, including a statement that identifies the research methodology and design.

Include a detailed Process diagram that identifies the stages of your research study; be specific and use a diagramming tool to create it. The same diagram will be expanded and refined in Chapter 4.

**Devote approximately 30-50 pages to this chapter.**

## Research Methodology and Design Process Diagram

Include a diagram created with a diagramming tool that describes the process of your study and depicts all the details of all stages of your plan. Small diagrams that refer to components of your product, algorithm, or design can be added below.

## Research Methodology and Design (Nature of the Study)

Begin writing here…

Checklist:

Describe the different stages of your study plan. The description should be thorough, and we suggest that you have already cleaned and preprocessed your data before writing this section. (Requirements)

Describe the research methodology and design. Elaborate upon their appropriateness about the study problem, purpose, and research questions.

Identify alternative methodologies and designs and indicate why they were less appropriate than the ones selected. Do not simply list and describe research methodologies and designs in general.

## Population and Sample

Begin writing here…

Checklist:

Describe the population, including the estimated size and relevant characteristics. Be as exact as possible, provide complete explanations and sources, and provide the predictive model, equation, and method in case of prediction.

Explain why the population is appropriate, given the study problem, purpose, and research questions.

Identify the sampling frame and support why this represents the population appropriately without introducing bias or issues.

Describe the Random Sampling method used and explain how this doesn’t introduce bias and issues. Explain and support the choice of the specific sampling method in the case of stratification, and provide details of calculations and identification of strata. Do not use Convenience sampling. In the case of census sampling, refer to the purpose and goals.

Provide mathematical equations that measure the appropriate Sample size formulas and explain why this is appropriate for the given process and population.

Describe the sample that will be (proposal) or was (manuscript) obtained.

Explain why the sample is appropriate, given the study problem, purpose, and research questions.

In the case of the Frequentists approach for your evaluation ( when applicable):

In the case of the Data Synthesis approach for your evaluation, include Meta-analysis and/or integrative analysis.

\*\*\*Please note that Surveys as a Data Collection method require secondary approval by your APD at the beginning of Chapter 1. In the case of Data Collection, describe how the participants will be (proposal) or were (manuscript) recruited (e.g., email lists from professional organizations, flyers) and/or the data will be (proposal) or were (manuscript) obtained (e.g., archived data, public records) with sufficient detail so the study could be replicated.

In the case of Data Acquisition, describe how and from where you acquired the data.

Explain why your number of observations is sufficient, given your specific DS methodology and the number of variables in your model.

In the case of augmentation techniques, provide support and validation of your choices.

Refer separately to processes and population constraints for the sample size and how these are satisfied.

Document the programming processes for data acquisition (e.g., Web scraping, public domain, commercial database), including, but not limited to, the programs, platforms, versions, and algorithms used.

☒ Finally, provide the Code used for the data collection in your Appendix, share the code and your data with your committee in a shared cloud folder, and prepare your GitHub folder.

## Materials or Instrumentation

Begin writing here…

Checklist:

In the rare case of survey Data Collection, describe the instruments (e.g., questionnaires) that will be (proposal) or were (manuscript) used, including information on their origin and evidence of their reliability and validity. OR, as applicable, describe the materials (e.g., lesson plans for interventions, webinars, archived data, etc.).

Describe in detail any software, platform, dataset, libraries, or programming language you will use in your research.

☐ If instruments or materials that another researcher developed are used, include evidence in the appendix that permission was granted to use the instrument(s) and/or material(s) and refer to that fact and the appendix in this section. Material provided under Creative Commons and Open Sourced must not be included.

If instruments or materials that another licensed researcher developed are used,

include the link of the primary author and instrument here; explain the constraints of the license and the license number. Please include them in the references as well.

## Operational Definitions of Variables

Begin writing here...

### XXX

Text…

Checklist:

For quantitative studies, identify how each variable will be (proposal) or was (manuscript) used in the study. Use terminology appropriate for the selected statistical test (e.g., independent/dependent, predictor/criterion, datatype, values, and levels of measurement).

Base the operational definitions on published research and valid and reliable instruments.

Identify the specific instrument that will be (proposal) or was (manuscript) used to measure each variable.

Refer in detail to the number of observations and variables in the sample and enumerate in a table the variables, data types, description, and level of categories for Categorical Variables and levels of measurement.

Refer in detail to the data source identification, distribution determination, and appropriateness of data by reviewing if the resulting distribution is suitable, including but not limited to discussions of biases and limitations, sampling techniques and methods, and how these ensure the representation of the distribution.

Explain feature engineering techniques, transformations of variables, and artificial variables that were introduced and the method that this was completed, support with documentation and citation, and the validation of this process.

## Study Procedures

Begin writing here…

Checklist:

(Only in case of primary data collection) Describe the exact steps that will be (proposal) or were (manuscript) followed to collect the data, addressing what data as well as how, when, from where, and from whom those data will be (proposal) or were (manuscript) collected in enough detail the study can be replicated.

Describe the exact steps of your study plan:

### Data Collection/Acquisition

Begin here….

### Data Preprocessing

Begin here…

#### Data Cleaning

Include the size, observations, variables, types of variables, missing values, affected variables of the original dataset, the methods of imputation you plan to include, size, and limitations after the cleaning ( if known otherwise, you should consist of those in Chapter 4), etc.…

Detail the cleansing processes, such as outlier detection, handling missing data through imputation and deduplication, and ensuring data quality.

#### Data Preparation

Explain the expected transformations of variables, including normalization, standardization, encoding categorical data, and introducing artificial variables. Document the methods used with supporting literature.

Discuss potential strategies like limiting sample size, bootstrapping, and resampling techniques. Validate these processes with appropriate statistical methods.

Mention the software programs, platforms, and their versions used for data preparation. Include programming scripts or modules in the Appendix.

#### Data Integration

Describe your plan regarding the process and tools for merging data from different digital sources or datasets, addressing issues like schema integration, entity resolution, and data format normalization.

Include the programming module for data integration in the Appendix, if you have already completed this section.

#### Data Feature Engineering

Explain how and why you are creating interaction features between variables.

Discuss the creation of polynomial features and their relevance to the model.

Detail any aggregated features from multiple data sources or grouped observations.

Describe the methods used to scale features and the rationale behind the choice.

Justify the use of log transformation for reducing skewness in continuous variables.

Outline using PCA or other techniques to reduce the feature space.

Discuss methods for selecting the most relevant features (e.g., filter, wrapper, and embedded methods).

If relevant, detail how date and time information is transformed into features useful for prediction (e.g., extracting day of the week, lag features).

For datasets with textual data, describe preprocessing steps like tokenization, stemming, and lemmatization.

Present an analysis of feature importance and how it influenced the feature engineering process.

#### Data Exploration

You should include the data exploration, identifying the similarities, contrasts, and anomalies you observed. Documentation should include graphs, charts, and tables created programmatically. (Do not include ALL different graphs for all variables; instead, focus on the findings that provide new insights, guide your research, and alter your steps or identified constraints).

Examples can be qqplots proving normality and histograms. Parallel boxplots and time series plots identify differences in distributions and boxplots for initial populations with vast amounts of data. 5-number summaries in tables along with standard deviations, recognizing differences in internal variation of comparing populations, loess graphs that compare time series, interactive graphs identifying differences of tendency measures, heatmaps, scatterplot matrices, etc…

Document only the most insightful visualizations and tables that guide the research direction.

### Data Mining

Follow the CRISP-DM model, iterating as necessary, and report only the final results.

Select and justify the choice of algorithms. Compare them with alternatives and support your choice with evidence from literature or analysis findings.

Detail the input, training choices, and ratio for training. Justify these choices with statistical validation.

Discuss what will be included such as decision trees, charts, or diagrams, explaining their relevance.

Discuss any expectations and challenges that may hinder the plan.

Include relevant programming modules in the Appendix.

#### Data Modeling

In this paragraph, discuss the modeling method or methods you plan to include.

Discuss in detail why the requirements of using the specific model are satisfied by your sample. (you may need to support this with evidence, graphs, or diagnostics)

Describe the input selection process, training, testing, the data you used for testing, and the included or excluded variables.

Explain the different approaches you will follow in your research plan.

Include programming modules related to data modeling in the Appendix.

#### Model Validation and Hyperparameter Tuning

Explain what type of validation measures and statistical measures you will include and support your reasoning with past research. Consider methods like cross-validation, ROC curves, precision-recall curves, and confusion matrices. The above are just examples of measures for validation; depending on your research focus, another type of measure may be appropriate. ( For LLMs, Blue Score, Rouge Score, Hallucinations Score etc..)

Discuss hyperparameter tuning strategies you plan to employ, such as grid or random search, and their impact on model performance.

Compare the model with alternative models or algorithms, focusing on error metrics and performance measures that may create future challenges.

Include the programming modules for model validation and hyperparameter tuning in the Appendix.

### Model Evaluation

#### Model Fit and Diagnostics

Discuss how you intend to evaluate the model fit using appropriate statistical tests and diagnostic plots like residual, qqplots, and influence plots.

Discuss the methods to diagnose issues with the model, such as overfitting or underfitting, and the steps taken to address these issues.

Include metrics to be discussed, reviewed, and incorporated to support the model's fit, such as R-squared, adjusted R-squared, F-statistics, and p-values.

For Classification Models: Elaborate on classification metrics such as accuracy, precision, recall, F1 score, and the AUC-ROC curve. Explain the significance of each metric and the scenarios in which they are instrumental. ( consider entropy as well when applicable)

For Regression Models: Discuss regression metrics like MSE, MAE, and R-squared. Provide insight into what these metrics convey about model performance.

Custom Metrics: If any custom metrics are used, describe them and justify their relevance to the project's objectives.

## Data Analysis

Begin writing here…

Checklist:

Describe the strategies that will be (proposal) or were (manuscript) used to code and/or analyze the data, as well as any software that will be (proposal) or was (manuscript) used.

Ensure the data that will be (proposal) or were (manuscript) analyzed can be used to answer the research questions and/or test the hypotheses to address the identified problem.

Use proper terminology associated with each design/analysis (e.g., independent and dependent variables for an experimental design, predictor, and criterion variables for regression).

For quantitative ( Inferential, Causal) studies, describe the analysis that will be (proposal) or was (manuscript) used to test each hypothesis.

For other quantitative (Constructive, Comparative, Mechanistic, Experimental) studies, describe the metrics dictated by the methodologies and include quality control.

**\*\*\*\*\***No qualitative studies are allowed in Data Science, but for text mining data, describe how the data will be (proposal) or were (manuscript) processed and analyzed, including any triangulation efforts. Explain the role of the researcher.

\*\*\*\*\*For mixed methods studies, include all the above. Due to time limitations, we only allow mixed methods studies in Data Science in rare cases. If you are working on a mixed methods study, ensure second-level approval by your Academic Program Director.

The programming module for this part of the study should also be included in the Appendix and shared with your Chair in a University cloud folder and Github.

Perform two-tiered data analysis: initial analysis post-EDA and comprehensive analysis for drawing inferences at the end of the study.

Present findings in written and visual forms, including discussions on graphs, charts, and tables.

Include the programming module used for data analysis in the Appendix.

## Assumptions

Begin writing here…

Checklist:

Discuss the assumptions along with the corresponding rationale underlying them.

## Limitations

Begin writing here…

Checklist:

Describe the study limitations.

Discuss the measures taken to mitigate these limitations.

## Delimitations

Begin writing here…

Checklist:

Describe the study delimitations along with the corresponding rationale underlying them. An example of delimitations is the conditions and parameters set intentionally by the researcher or by the population and sample selection.

Explain how these research decisions relate to the existing literature and theoretical/conceptual framework, problem statement, purpose statement, and research questions.

Further Suggestions

**Ethical Assurances (Primary Data Collection)**

Begin writing here…

Focus on the following ( Ignore if you are using Secondary Data):

**Ethical Considerations (Secondary data)**

Begin writing here…

Focus on the following:

Checklist:

Confirm in a statement the study will (proposal) or did (manuscript) receive approval from the University’s Institutional Review Board (IRB) before data collection.

If the risk to participants is more significant than minimal, discuss the relevant ethical issues and how they will be (proposal) or were (manuscript) addressed.

Describe how confidentiality or anonymity will be (proposal) or was (manuscript) achieved.

Identify how the data will be (proposal) or were (manuscript) securely stored in accordance with IRB requirements.

Describe the role of the researcher in the study. Discuss relevant issues, including biases and personal and professional experiences with the topic, problem, or context. Present the strategies that will be (proposal) or were (manuscript) used to prevent these biases and experiences from influencing the analysis or findings.

In the dissertation manuscript only, include the IRB approval letter in an appendix.

## Summary

Begin writing here…

Checklist:

Summarize the key points presented in the chapter.

Logically lead the reader to the next chapter on the study's findings.

# References

Abdullah, Q., Abdullah, N., Balfaqih, M., Shah, N. S. M., Anuar, S., Almohammedi, A. A., Salh, A., Farah, N., & Shepelev, V. (2020). Maximizing system throughput in wireless powered sub-6 ghz and millimeter-wave 5G heterogeneous networks. *Telkomnika*, 18(3), 1185–1194. https://doi.org/10.12928/TELKOMNIKA.v18i3.15049

Aikoh, T., Homma, R., & Abe, Y. (2023). Comparing conventional manual measurement of the green view index with modern automatic methods using Google Street View and semantic segmentation. *Urban Forestry & Urban Greening*, 80, 127845. https://doi.org/10.1016/j.ufug.2023.127845

*Airborne laser scanning*. (2013, January 18). Geospatial Modeling & Visualization | A Method Store for Advanced Survey and Modeling Technologies. https://gmv.cast.uark.edu/scanning-2/airborne-laser-scanning/

Alkhateeb, A., Jiang, S., & Charan, G. (2023). Real-time digital twins: Vision and research directions for 6G and beyond. *IEEE Communications Magazine*, 61(11), 128–134. https://doi.org/10.1109/MCOM.001.2200866

Almasan, P., Ferriol-Galmes, M., Paillisse, J., Suarez-Varela, J., Perino, D., Lopez, D., Perales, A. A. P., Harvey, P., Ciavaglia, L., Wong, L., Ram, V., Xiao, S., Shi, X., Cheng, X., Cabellos-Aparicio, A., & Barlet-Ros, P. (2022). Network Digital Twin: Context, Enabling Technologies, and Opportunities. *IEEE Communications Magazine*, 60(11), 22–27. https://doi.org/10.1109/MCOM.001.2200012

Angin, P., Anisi, M. H., Göksel, F., Gürsoy, C., & Büyükgülcü, A. (2020). AgriLoRa: a digital twin framework for smart agriculture. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, 11(4), 77–96. https://doi.org/10.22667/JOWUA.2020.12.31.077

Anzum, R. (2021). Factors that affect LoRa Propagation in Foliage Medium. *Procedia Computer Science*, 194, 149–155. https://doi.org/10.1016/j.procs.2021.10.068

Arnas, Yudan, W., & Eny, K. (2024). Unveiling Innovations Across Multidisciplinary Horizons. *International Journal of Technology*, 15(2), 240–246. https://doi.org/10.14716/ijtech.v15i2.6933

Attaran, M., & Celik, B. G. (2023). Digital Twin: Benefits, use cases, challenges, and opportunities. *Decision Analytics Journal*, 6, 100165. https://doi.org/10.1016/j.dajour.2023.100165

Azad, M. M., Carla, C. M., & Scott, D. L. (2019). Leveraging Digital Twin Technology in Model-Based Systems Engineering. *Systems*, 7(1), 7. https://doi.org/10.3390/systems7010007

Banerjee, B. P., Raval, S., Cullen, P. J., & Kumar Singh, S. (2019). *Mapping of complex vegetation communities and species using UAV-LIDAR metrics and high-resolution optical data*. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. https://doi.org/10.1109/IGARSS.2019.8899160

Barb, G., Danuti, F., Ouamri, M. A., & Otesteanu, M. (2022, November). *Analysis of vegetation and penetration losses in 5G mmwave communication systems*. 2022 International Symposium on Electronics and Telecommunications (ISETC) Electronics and Telecommunications (ISETC), 2022 International Symposium on. :1-5 Nov, 2022. https://doi.org/10.1109/ISETC56213.2022.10009963

Barricelli, B. R., Casiraghi, E., & Fogli, D. (2019). A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access*, 7, 167653–167671. https://doi.org/10.1109/ACCESS.2019.2953499

Blume, C., Blume, S., Thiede, S., & Herrmann, C. (2020). Data-Driven Digital Twins for Technical Building Services Operation in Factories: A Cooling Tower Case Study. *Journal of Manufacturing Materials Processing*, 4(4), 97. https://doi.org/10.3390/jmmp4040097

Bose, T., Chatur, N., Mukherjee, M., Verma, S., & Adhya, A. (2024). Traffic-Aware Optimal Multi-Beam Resource Allocation in 5G Networks Impaired by Rain and Foliage. *IEEE Communications Letters*, 1. https://doi.org/10.1109/LCOMM.2024.3357174

*Brand resource center | Products and services - Geo guidelines*. (n.d.). Google - About Google, Our Culture & Company News. https://about.google/brand-resource-center/products-and-services/geo-guidelines/#general-guidelines

Bromm, K. N., Lang, I.-M., Twardzik, E. E., Antonakos, C. L., Dubowitz, T., & Colabianchi, N. (2020). Virtual audits of the urban streetscape: comparing the inter-rater reliability of GigaPan® to Google Street View. *Int J Health Geogr*, 19(1), 31. https://doi.org/10.1186/s12942-020-00226-0

Capecchi, I., Borghini, T., & Bernetti, I. (2023). Automated urban tree survey using remote sensing data, Google Street View images, and plant species recognition apps. *European Journal of Remote Sensing*, 56(1), 2162441. https://doi.org/10.1080/22797254.2022.2162441

Caprari, G., Castelli, G., Montuori, M., Camardelli, M., & Malvezzi, R. (2022). Digital Twin for Urban Planning in the Green Deal Era: A State of the Art and Future Perspectives. *Sustainability*, 14(10). https://doi.org/10.3390/su14106263

Chaminé, H. I., Pereira, A. J. S. C., Teodoro, A. C., & Teixeira, J. (2021). Remote sensing and GIS applications in earth and environmental systems sciences. *SN Applied Sciences*, 3(12), 870. https://doi.org/10.1007/s42452-021-04855-3

Chen, J., Wang, G., Luo, L., Gong, W., & Cheng, Z. (2021). Building Area Estimation in Drone Aerial Images Based on Mask R-CNN. *IEEE Geoscience and Remote Sensing Letters*, 18(5), 891–894. https://doi.org/10.1109/LGRS.2020.2988326

Chen, Q., Gao, T., Zhu, J., Wu, F., Li, X., Lu, D., & Yu, F. (2022). Individual Tree Segmentation and Tree Height Estimation Using Leaf-Off and Leaf-On UAV-LiDAR Data in Dense Deciduous Forests. *Remote Sensing*, 14(12), 2787. https://doi.org/10.3390/rs14122787

Chikhale, D., Munde, M., & Deosarkar, S. (2022). Atmospheric effects and behavior of electromagnetic signals in the millimeter wave range wireless communication. *International Journal of Microwave & Optical Technology*, 17(2), 115–125.

Clancy, R., O’Sullivan, D., & Bruton, K. (2023). Data-driven quality improvement approach to reducing waste in manufacturing. *The TQM Journal*, 35(1), 72. https://doi.org/10.1108/TQM-02-2021-0061

*The code affirms an obligation of computing professionals to use their skills for the benefit of society.* (n.d.). Association for Computing Machinery. https://www.acm.org/code-of-ethics

Correia, J. B., Abel, M., & Becker, K. (2023). Data management in digital twins: a systematic literature review. *Knowl Inf Syst*, 65(8), 3165–3196. https://doi.org/10.1007/s10115-023-01870-1

Cureton, P., & Hartley, E. (2023). City Information Models (CIMs) as precursors for Urban Digital Twins (UDTs): A case study of Lancaster. *Frontiers in Built Environment*, 9. https://doi.org/10.3389/fbuil.2023.1048510

*Data Privacy Framework*. (n.d.). https://www.dataprivacyframework.gov/

*Data protection adequacy for non-EU countries*. (n.d.). European Commission. https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions\_en

*Data transfer frameworks – Privacy and terms – Google*. (n.d.). Privacy and Terms – Google. https://policies.google.com/privacy/frameworks?hl=en

De Beelde, B., Plets, D., & Joseph, W. (2023). Characterization of Vegetation Loss and Impact on Network Performance at V-Band Frequencies. *IEEE Antennas and Wireless Propagation Letters*, 22(3), 596–600. https://doi.org/10.1109/LAWP.2022.3219556

Deng, T., Zhang, K., & Shen, Z.-J., (Max). (2021). A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *Journal of Management Science and Engineering*, 6(2), 125–134. https://doi.org/10.1016/j.jmse.2021.03.003

Deng, X., Tang, G., & Wang, Q. (2022). A novel fast classification filtering algorithm for LiDAR point clouds based on small grid density clustering. *Geodesy and Geodynamics*, 13(1), 38–49. https://doi.org/10.1016/j.geog.2021.10.002

Diab, A., Kashef, R., & Shaker, A. (2022). Deep Learning for LiDAR Point Cloud Classification in Remote Sensing. *Sensors*, 22(20), 7868. https://doi.org/10.3390/s22207868

Dutta, A., & Zisserman, A. (2019). *The VIA Annotation Software for Images, Audio, and Video*. https://doi.org/10.1145/3343031.3350535

*Fairness: Types of bias*. (n.d.). Google for Developers. https://developers.google.com/machine-learning/crash-course/fairness/types-of-bias

Farooq, U., & Lokam, A. (2023). Performance analysis of mmWave/sub-terahertz communication link for 5G and B5G mobile networks. *Frequenz*, 77(11/12), 599–606. https://doi.org/10.1515/freq-2023-0024

Federal Data Protection and Information Commissioner (FDPIC). (n.d.). *Cross-border transfer of personal data*. https://www.edoeb.admin.ch/edoeb/en/home/datenschutz/arbeit\_wirtschaft/datenuebermittlung\_ausland.html

Fett, M., Wilking, F., Goetz, S., Kirchner, E., & Wartzack, S. (2023). A Literature Review on the Development and Creation of Digital Twins, Cyber-Physical Systems, and Product-Service Systems. *Sensors (Basel, Switzerland)*, 23(24). https://doi.org/10.3390/s23249786

Gabriele, M., Cazzani, A., Zerbi, C. M., & Brumana, R. (2023). DIGITAL TWIN TO MONITOR, UNDERSTAND AND PRESERVE THE COMPLEXITY OF MULTI-SCALE NATURAL, AGRICULTURAL, DESIGNED LANDSCAPES AND ARCHITECTURE: BIODIVERSITY CONSERVATION, TRANSFORMATION AND DECLINE AT VILLA ARCONATI SITE AT CASTELLAZZO OF BOLLATE (MI). *International Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 48(M/2), 613–620. https://doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-613-2023

Gaspari, F., Ioli, F., Barbieri, F., Belcore, E., & Pinto, L. (2022). INTEGRATION OF UAV-LIDAR AND UAV-PHOTOGRAMMETRY FOR INFRASTRUCTURE MONITORING AND BRIDGE ASSESSMENT. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, (B2), 995–1002. https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-995-2022

*Google for developers Maps Static API*. (n.d.). Google for Developers. https://developers.google.com/maps/documentation/maps-static/start

*Google for developers Street View Static API overview*. (n.d.). Google for Developers. https://developers.google.com/maps/documentation/streetview/overview

*Google maps platform terms of service | Google cloud*. (n.d.). Google Cloud. https://cloud.google.com/maps-platform/terms/

*Google-contributed Street View imagery policy*. (n.d.). Google Maps Street View. https://www.google.com/streetview/policy/

*A guide to international transfers*. (2024, March 24). https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/international-transfers/international-transfers-a-guide/#adequacy

Hayat Suhendar, M. T., & Widyani, Y. (2023, September 7–8). *Machine Learning Application Development Guidelines Using CRISP-DM and Scrum Concept* [Conference session]. 2023 IEEE International Conference on Data and Software Engineering (ICoDSE), Toba, Indonesia, Indonesia. https://doi.org/10.1109/ICoDSE59534.2023.10291438

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). Mask R-CNN. *arXiv.org*. https://doi.org/10.48550/arXiv.1703.06870

Hematang, F., Murdjoko, A., Hendri, H., & Tokede, M. (2022). Application of Unmanned Aerial Vehicle (UAV) For Estimation of Tree Height in Heterogeneous Forest. *Biosaintifika: Journal of Biology & Biology Education*, 14(2), 168–179. https://doi.org/10.15294/biosaintifika.v14i2.35637

Hong, W., Jiang, Z. H., Yu, C., Hou, D., Wang, H., Guo, C., Hu, Y., Kuai, L., Yu, Y., Jiang, Z., Chen, Z., Chen, J., Yu, Z., Zhai, J., Zhang, N., Tian, L., Wu, F., Yang, G., Hao, Z., & Zhou, J. Y. (2021). The Role of Millimeter-Wave Technologies in 5G/6G Wireless Communications. *IEEE Journal of Microwaves*, 1(1), 101–122. https://doi.org/10.1109/JMW.2020.3035541

Hu, W., Fang, J., Zhang, T., Liu, Z., & Tan, J. (2023). A new quantitative digital twin maturity model for high-end equipment. *Journal of Manufacturing Systems*, 66, 248–259. https://doi.org/10.1016/j.jmsy.2022.12.012

Hu, W., Zhang, T., Deng, X., Liu, Z., & Tan, J. (2021). Digital twin: a state-of-the-art review of its enabling technologies, applications and challenges. *Journal of Intelligent Manufacturing and Special Equipment*, 2(1), 1–34. https://doi.org/10.1108/JIMSE-12-2020-010

Huang, Z., Shen, Y., Li, J., Fey, M., & Brecher, C. (2021). A Survey on AI-Driven Digital Twins in Industry 4.0: Smart Manufacturing and Advanced Robotics. *Sensors*, 21(19), 6340. https://doi.org/10.3390/s21196340

Hui, L., Wang, M., Zhang, L., Lu, L., & Cui, Y. (2023). Digital Twin for Networking: A Data-Driven Performance Modeling Perspective. *IEEE Network*, 37(3), 202–209. https://doi.org/10.1109/MNET.119.2200080

Jahnke, N., & Otto, B. (2023). Data Catalogs in the Enterprise: Applications and Integration. *Datenbank Spektrum*, 23(2), 89–96. https://doi.org/10.1007/s13222-023-00445-2

Jiang, B., An, X., Xu, S., & Chen, Z. (2023). Intelligent Image Semantic Segmentation: A Review Through Deep Learning Techniques for Remote Sensing Image Analysis. *Journal of the Indian Society of Remote Sensing*, 51(9), 1865–1878. https://doi.org/10.1007/s12524-022-01496-w

Kang, T. W., & Mo, Y. (2024). A comprehensive digital twin framework for building environment monitoring with emphasis on real-time data connectivity and predictability. *Developments in the Built Environment*, 17, 100309. https://doi.org/10.1016/j.dibe.2023.100309

Kapteyn, M. G., & Willcox, K. E. (2020). From physics-based models to predictive digital twins via interpretable machine learning. *arXiv preprint arXiv:2004.11356*. https://doi.org/10.48550/arXiv.2004.11356

Kaur, K., Kumar, S., & Baliyan, A. (2020). 5G: a new era of wireless communication. *International Journal of Information Technology*, 12(2), 619–624. https://doi.org/10.1007/s41870-018-0197-x

Khan, L. U., saad, W., Niyato, D., Han, Z., & Hong, C. S. (2022). Digital-Twin-Enabled 6G: Vision, Architectural Trends, and Future Directions. *IEEE Communications Magazine*, 60(1), 74–80. https://doi.org/10.1109/MCOM.001.21143

Khan, N. A., & Schmid, S. (2024). AI-RAN in 6G Networks: State-of-the-Art and Challenges. *IEEE Open Journal of the Communications Society*, 5, 294–311. https://doi.org/10.1109/OJCOMS.2023.3343069

Ko, J., Hur, S., Noh, Y., Whang, K., Park, J., Park, D., & Cho, D. (2020). Measurements and Analysis of Radio Propagation at 28 GHz in Vegetated Areas of Typical Residential Environments. *IEEE Transactions on Antennas and Propagation IEEE Trans. Antennas Propagat. Antennas and Propagation, IEEE Transactions on.*, 68(5), 4149–4154. https://doi.org/10.1109/TAP.2020.2968801

Kukushkin, K., Ryabov, Y., & Borovkov, A. (2022). Digital Twins: A Systematic Literature Review Based on Data Analysis and Topic Modeling. *Data*, 7(12), 173. https://doi.org/10.3390/data7120173

Kuruvatti, N. P., Habibi, M. A., Partani, S., Han, B., Fellan, A., & Schotten, H. D. (2022). Empowering 6G communication systems with digital twin technology: A comprehensive survey. *IEEE Access*, 10, 112158–112186. https://doi.org/10.1109/ACCESS.2022.3215493

Lai, C., Senic, D., Gentile, C., Senic, J., & Golmie, N. (2023). Raytracing Digital Foliage at Millimeter-Wave: A Case Study on Calibration Against 60-GHz Channel Measurements on Summer and Winter Trees. *IEEE Access*, 11, 145931–145943. https://doi.org/10.1109/ACCESS.2023.3345248

Lehtola, V. V., Koeva, M., Elberink, S. O., Raposo, P., Virtanen, J.-P., Vahdatikhaki, F., & Borsci, S. (2022). Digital twin of a city: Review of technology serving city needs. *International Journal of Applied Earth Observation and Geoinformation*, 114, 102915. https://doi.org/10.1016/j.jag.2022.102915

Li, H., Ye, C., Guo, Z., Wang, L., Wei, R., & Li, J. (2021). A Fast Progressive TIN Densification Filtering Algorithm for Airborne LiDAR Data Using Adjacent Surface Information. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 12492–12503. https://doi.org/10.1109/JSTARS.2021.3131586

Li, Z., Hodgson, M. E., & Li, W. (2018). A general-purpose framework for parallel processing of large-scale LiDAR data. *International Journal of Digital Earth*, 11(1), 26–47. https://doi.org/10.1080/17538947.2016.1269842

Li, Z., Wu, B., Li, Y., & Chen, Z. (2023). Fusion of aerial, MMS and backpack images and point clouds for optimized 3D mapping in urban areas. *ISPRS Journal of Photogrammetry & Remote Sensing*, 202, 463–478. https://doi.org/10.1016/j.isprsjprs.2023.07.010

Lu, D., & Jiang, X. (2024). A brief overview and perspective of using airborne Lidar data for forest biomass estimation. *International Journal of Image and Data Fusion*, 15(1), 1–24. https://doi.org/10.1080/19479832.2024.2309615

Luo, S., Liang, Y., Luo, Z., Liang, G., Wang, C., & Wu, X. (2023). Vision-Guided Object Recognition and 6D Pose Estimation System Based on Deep Neural Network for Unmanned Aerial Vehicles towards Intelligent Logistics. *Applied Sciences*, 13(1). https://doi.org/10.3390/app13010115

Mazzacca, G., Grilli, E., Cirigliano, G. P., Remondino, F., & Campana, S. (2022). Seeing among foliage with lidar and machine learning: Towards a transferable archaeological pipeline. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-2-W1-2022, 365–372. https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-365-2022

Müllerová, J., Gago, X., Bucas, M., Company, Jaume, Estrany, J., Fortesa, J., Manfreda, S., Adrien, M., Mokros, M., Paulus, G., Tiskus, E., Tsiafouli, M., & Kent, R. (2021). Characterizing vegetation complexity with unmanned aerial systems (UAS) -A framework and synthesis. *Ecological Indicators*, 131, 108156. https://doi.org/10.1016/j.ecolind.2021.108156

*MWC23 – Intelligent 5G mmWave network planning*. (2023, February 27). Wireless Technology & Innovation | Mobile Technology | Qualcomm. https://www.qualcomm.com/videos/mwc23-intelligent-5g-mmwave-network-planning

Mylonas, G., Kalogeras, A., Kalogeras, G., Anagnostopoulos, C., Alexakos, C., & Muñoz, L. (2021). Digital twins from smart manufacturing to smart cities: A survey. *IEEE Access*, 9, 143222–143249. https://doi.org/10.1109/ACCESS.2021.3120843

Nagai, H., Nakazawa, E., & Akabayashi, A. (2022). *The creation of the Belmont Report and its effect on ethical principles: a historical study*. https://doi.org/10.1007/s40592-022-00165-5

National Academies of Sciences, Engineering, and Medicine. (2023). *Foundational research gaps and future directions for digital twins*. The National Academies Press. https://doi.org/10.17226/26894

National Academies of Sciences, Engineering, and Medicine. (2024). *Foundational Research Gaps and Future Directions for Digital Twins.* https://doi.org/10.17226/26894

Nguyen, H. X., Trestian, R., To, D., & Tatipamula, M. (2021). Digital Twin for 5G and beyond. *IEEE Communications Magazine*, 59(2), 10–15. https://doi.org/10.1109/MCOM.001.2000343

*Northern Research Station | Research Natural Areas. U.S. Department of Agriculture, Forest Service, Northern Research Station.* (2023, September 22). ArcGIS StoryMaps. https://storymaps.arcgis.com/stories/5b1cfad1e7bd49058b2e570ecdc3b50a

*OCM Partners, 2024: 2022 City of Philadelphia Lidar DEM: Philadelphia, PA*. (2024, January 10). Welcome to NOAA | NOAA Fisheries. https://www.fisheries.noaa.gov/inport/item/70174

*Overview – Google Earth*. (n.d.). Google Earth. https://www.google.com/intl/en\_us/earth/

*Philadelphia LiDAR - LAS files 2022 {2022} - Big Ten academic alliance Geoportal*. (2022). Big Ten Academic Alliance Geoportal. https://geo.btaa.org/catalog/pasda-7154

*Policies for aerial view API*. (n.d.). Google for Developers. https://developers.google.com/maps/documentation/aerial-view/policies

*Policies for street view static API*. (n.d.). Google for Developers. https://developers.google.com/maps/documentation/streetview/policies

Pradeep, T., Shukla, N. K., Animesh, T., & Shiv, P. (2021). *Investigating the Effect of Rain, Foliage, Atmospheric Gases, and Diffraction on Millimeter (mm) Wave Propagation for 5G Cellular Networks*. https://doi.org/10.1007/978-981-16-3346-1\_42

*Privacy policy – Privacy and terms – Google*. (n.d.). Privacy and Terms – Google. https://policies.google.com/privacy?hl=en-US?hl=en#infosharing

QGIS Development Team. (2021). *QGIS Geographic Information System* [Computer software]. QGIS Association. http://www.qgis.org

Qi, Q., & Tao, F. (2018). Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. *IEEE Access*, 6, 3593. https://doi.org/10.1109/ACCESS.2018.2793265

Raihan, A. (2023). Artificial intelligence and machine learning applications in forest management and biodiversity conservation. *Natural Resources Conservation and Research*, 6(2), 3825. https://doi.org/10.24294/nrcr.v6i2.3825

Responsible AI at Google research: Perception fairness. (n.d.). https://blog.research.google/2023/08/responsible-ai-at-google-research.html

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). *Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression*. http://arxiv.org/abs/1902.09630

Rogers, S. R., Manning, I., & Livingstone, W. (2020). Comparing the Spatial Accuracy of Digital Surface Models from Four Unoccupied Aerial Systems: Photogrammetry Versus LiDAR. *Remote Sens.*, 12(17). https://doi.org/10.3390/rs12172806

Sabri, N., Aljunid, S. A., Salim, M. S., Fouad, S., & Kamaruddin, R. (2015). Wireless Sensor Network Wave Propagation in Vegetation. In F. L. Gaol, K. Shrivastava, & J. Akhtar (Eds.), *Recent Trends in Physics of Material Science and Technology.* (pp. 283–298). Springer Series in Materials Science. https://doi.org/10.1007/978-981-287-128-2\_18

Saravanan, S. K., Muthusenthil, B., & Gurusubramani, S. (2022). *A review of digital twin leveraging technology, concepts, tools, and industrial applications*. 2022 1st International Conference on Computational Science and Technology (ICCST) Computational Science and Technology (ICCST). https://doi.org/10.1109/ICCST55948.2022.10040359

Sarirete, A., Balfagih, Z., Brahimi, T., Lytras, M. D., & Visvizi, A. (2022). Artificial intelligence and machine learning research: towards digital transformation at a global scale. *Journal of Ambient Intelligence and Humanized Computing*, 13(7), 3319–3321. https://doi.org/10.1007/s12652-021-03168-y

Savelonas, M. A., Veinidis, C. N., & Bartsokas, T. K. (2022). Computer Vision and Pattern Recognition for the Analysis of 2D/3D Remote Sensing Data in Geoscience: A Survey. *Remote Sensing*, 14(23), 6017. https://doi.org/10.3390/rs14236017

Schrotter, G., & Hürzeler, C. (2020). The Digital Twin of the City of Zurich for Urban Planning. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, Preprints*, 88(1), 99–112. https://doi.org/10.1007/s41064-020-00092-2

Shah, R. (2018, November 9). *Psst, says 5G... Wanna see what my new antenna tech challenge looks like?* Semiconductor Engineering. https://semiengineering.com/psst-says-5g-wanna-see-what-my-new-antenna-tech-challenge-looks-like/

Shahat, E., Hyun, C. T., & Yeom, C. (2021). City Digital Twin Potentials: A Review and Research Agenda. *Sustainability*, 13(6). https://doi.org/10.3390/su13063386

Sharma, A., Kosasih, E., Zhang, J., Brintrup, A., & Calinescu, A. (2022). Digital Twins: State of the art theory and practice, challenges, and open research questions. *Journal of Industrial Information Integration*, 30. https://doi.org/10.1016/j.jii.2022.100383

Shen, Y., Huang, R., Hua, B., Pan, Y., Mei, Y., & Dong, M. (2023). Automatic Tree Height Measurement Based on Three-Dimensional Reconstruction Using Smartphone. *Sensors 2023*, 23(16), 7248. https://doi.org/10.3390/s23167248

Somanath, S., Naserentin, V., Eleftheriou, O., Sjölie, D., Wästberg, B. S., & Logg, A. (2023, May 3). *On procedural urban digital twin generation and visualization of large scale data*. From Oaister®, Provided by the OCLC Cooperative. http://arxiv.org/abs/2305.02242

Song, J., & Le Gall, F. (2023). Digital twin standards, open source, and best practices. In N. Crespi, A. T. Drobot, & R. Minerva (Eds.), *The Digital Twin* (pp. 497–530). Springer Link. https://doi.org/10.1007/978-3-031-21343-4\_18

Song, J., Zhao, Y., Song, W., Zhou, H., Zhu, D., Huang, Q., Fan, Y., & Lu, C. (2022). Fisheye Image Detection of Trees Using Improved YOLOX for Tree Height Estimation. *Sensors*, 22, 3636. https://doi.org/10.3390/s22103636

Song, S., & Qin, R. (2022). A NOVEL INTRINSIC IMAGE DECOMPOSITION METHOD TO RECOVER ALBEDO FOR AERIAL IMAGES IN PHOTOGRAMMETRY PROCESSING. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2022, 23–30. https://doi.org/10.5194/isprs-annals-V-2-2022-23-2022

Suhaizad, L. S., Khalid, N., & Abu Sari, M. Y. (2023). Tree Height and Crown Extraction From UAV-Based Multispectral Imagery. *International Journal of Geoinformatics*, 19(5), 61–68. https://doi.org/10.52939/ijg.v19i5.2661

Sun, Z., Wang, X., Wang, Z., Yang, L., Xie, Y., & Huang, Y. (2021). UAVs as remote sensing platforms in plant ecology: review of applications and challenges. *J Plant Ecol*, 14(6), 1003–1023. https://doi.org/10.1093/jpe/rtab089

Sun, Z., Xue, B., Zhang, M., & Schindler, J. (2023). *An Improved Mask R-CNN for Instance Segmentation of Tree Crowns in Aerial Imagery*. 2023 38th International Conference on Image and Vision Computing New Zealand (IVCNZ), Image and Vision Computing New Zealand (IVCNZ). https://doi.org/10.1109/IVCNZ61134.2023.10343827

Thelen, A., Zhang, X., Fink, O., Lu, Y., Ghosh, S., Youn, B. D., Todd, M. D., Mahadevan, S., Hu, C., & Hu, Z. (2022). A comprehensive review of digital twin — part 1: modeling and twinning enabling technologies. *Struct Multidisc Optim*, 65(12), 354. https://doi.org/10.1007/s00158-022-03425-4

Thuvander, L., Somanath, S., & Hollberg, A. (2022). PROCEDURAL DIGITAL TWIN GENERATION FOR CO-CREATING IN VR FOCUSING ON VEGETATION. *International Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 48(4/W5), 189–196. https://doi.org/10.5194/isprs-archives-XLVIII-4-W5-2022-189-2022

Tran, B.-H., Aussenac-Gilles, N., Comparot, C., & Trojahn, C. (2022). Semantic Integration of Raster Data for Earth Observation on Territorial Units. *ISPRS International Journal of Geo-Information*, 11(2), 149. https://doi.org/10.3390/ijgi11020149

Trevisan, L. R., Brichi, L., Gomes, T. M., & Rossi, F. (2023). Estimating Black Oat Biomass Using Digital Surface Models and a Vegetation Index Derived from RGB-Based Aerial Images. *Remote Sens*, 15(5). https://doi.org/10.3390/rs15051363

VanDerHorn, E., & Mahadevan, S. (2021). Digital Twin: Generalization, characterization and implementation. *Decision Support Systems*, 145, 113524. https://doi.org/10.1016/j.dss.2021.113524

Vrabič, R., Erkoyuncu, J. A., Butala, P., & Roy, R. (2018). Digital twins: Understanding the added value of integrated models for through-life engineering services. *Procedia Manufacturing*, 16, 139–146. https://doi.org/10.1016/j.promfg.2018.10.167

Wang, W., Xiao, L., Zhang, J., Yang, Y., Tian, P., Wang, H., & He, X. (2018). Potential of Internet street-view images for measuring tree sizes in roadside forests. *Urban Forestry & Urban Greening*, 35, 211–220. https://doi.org/10.1016/j.ufug.2018.09.008

Wang, Z., & Menenti, M. (2021). Challenges and Opportunities in Lidar Remote Sensing. *Frontiers in Remote Sensing*, 2, 641723. https://doi.org/10.3389/frsen.2021.641723

Waqar, A., Othman, I., Almujibah, H., Khan, M. B., Alotaibi, S., & Elhassan, A. A. M. (2023). Factors Influencing Adoption of Digital Twin Advanced Technologies for Smart City Development: Evidence from Malaysia. *Buildings*, 13(3). https://doi.org/10.3390/buildings13030775

Weil, C., Bibri, S. E., Longchamp, R., Golay, F., & Alahi, A. (2023). Urban Digital Twin Challenges: A Systematic Review and Perspectives for Sustainable Smart Cities. *Sustainable Cities and Society*, 99, 104862. https://doi.org/10.1016/j.scs.2023.104862

Wilk, Ł., Mielczarek, D., Ostrowski, W., Dominik, W., & Krawczyk, J. (2022). SEMANTIC URBAN MESH SEGMENTATION BASED ON AERIAL OBLIQUE IMAGES AND POINT CLOUDS USING DEEP LEARNING. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, (B2), 485–491. https://doi.org/10.5194/isprs-archives-XLIII-B2-2022-485-2022

Yang, E., Wang, M., Cheng, H., Liu, R., & Chen, F. (2022). *A Method to Improve the Precision of 2-Dimensioanl Size Measurement of Objects through Image Processing*. 2022 21st International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES) DCABES Distributed Computing and Applications for Business Engineering and Science (DCABES), 2022 21st International Symposium on, DCABES, 197–200. https://doi.org/10.1109/DCABES57229.2022.00010

Zeng, Y., Duan, Q., Chen, X., Peng, D., Mao, Y., & Yang, K. (2021). UAVData: A dataset for unmanned aerial vehicle detection. *Soft Computing: A Fusion of Foundations, Methodologies and Applications*, 25(7), 5385–5393. https://doi.org/10.1007/s00500-020-05537-9

Zhang, J., Cosma, G., & Watkins, J. (2021). Image Enhanced Mask R-CNN: A Deep Learning Pipeline with New Evaluation Measures for Wind Turbine Blade Defect Detection and Classification. *Journal of Imaging*, 7(3). https://doi.org/10.3390/jimaging7030046

Zhang, P., Yang, B., Yi, C., Wang, H., & You, X. (2020). Measurement-Based 5G Millimeter-Wave Propagation Characterization in Vegetated Suburban Macrocell Environments. *IEEE Transactions on Antennas and Propagation IEEE Trans. Antennas Propagat. Antennas and Propagation, IEEE Transactions on.*, 68(7), 5556–5567. https://doi.org/10.1109/TAP.2020.2975365

Zhang, Y., Anderson, C. R., Michelusi, N., Love, D. J., Baker, K. R., & Krogmeier, J. V. (2019). Propagation Modeling Through Foliage in a Coniferous Forest at 28 GHz. *IEEE Wireless Communications Letters*, 8(3), 901–904. https://doi.org/10.1109/LWC.2019.2899299

Zhao, D., Li, X., Wang, X., Shen, X., & Gao, W. (2022). Applying Digital Twins to Research the Relationship Between Urban Expansion and Vegetation Coverage: A Case Study of Natural Preserve. *Frontiers in Plant Science*, 13, 840471. https://doi.org/10.3389/fpls.2022.840471

Zhao, W., Persello, C., & Stein, A. (2023). Semantic-aware unsupervised domain adaptation for height estimation from single-view aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 372–385. https://doi.org/10.1016/j.isprsjprs.2023.01.003

Zhao, Y., Cheng, D., Shen, S., Cai, D., & Lyu, X. (2023). *Improved Mask R-CNN for Disturbed Area Extraction in Construction Projects from High-Resolution Satellite Imagery*. 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), Artificial Intelligence and Big Data (ICAIBD). https://doi.org/10.1109/ICAIBD57115.2023.10206407

Zhou, C., Gao, J., Li, M., Sherman Shen, X., & Zhuang, W. (2022). Digital Twin-Empowered Network Planning for Multi-Tier Computing. *Journal of Communications and Information Networks*, 7(3), 221–238. https://doi.org/10.23919/JCIN.2022.9906937

# 

# Appendix A Annotated Bibliography

Chikhale, D., Munde, M., & Deosarkar, S. (2022). Atmospheric Effects and Behavior of Electromagnetic Signals in the Millimeter Wave Range Wireless Communication. *International Journal of Microwave & Optical Technology*, *17*(2), 115–125.

The article provides valuable insights into the impact of foliage on the efficient deployment of millimeter-wave communication systems. Several references and studies are cited to underscore the significance of considering foliage effects in the design and deployment of millimeter-wave networks. Here are some key points discussed in the article are related to foliage and its impact on millimeter-wave deployment: (a) Atmospheric constituents, such as free space loss, rain attenuation loss, gaseous loss, foliage loss, humidity, cloud, fog, and penetration loss, significantly impact the propagation of electromagnetic signals in the millimeter wave range. (b) The presence of foliage is acknowledged as a significant factor causing attenuation in the millimeter wave range, particularly in non-line-of-sight communication scenarios. (c) The need for accurate measurements and analysis of foliage effects is emphasized to suggest new and improved models for designing and developing millimeter-wave communication systems. (d) Studies on foliage attenuation at specific frequencies, such as 35GHz, are referenced, indicating the importance of considering frequency-dependent foliage effects in millimeter-wave deployment. (e) The impact of foliage on signal attenuation in the millimeter band is highlighted, emphasizing the need for effective modeling and measurement of foliage effects to assess and mitigate signal losses accurately.

Overall, the authors underscore the critical role of foliage in the efficient deployment of millimeter-wave communication systems. It emphasizes the need for comprehensive understanding, measurement, and modeling of foliage effects to ensure the reliable and effective operation of millimeter-wave networks, especially in non-line-of-sight scenarios.

Farooq, U., & Lokam, A. (2023). Performance analysis of mmWave/sub-terahertz communication link for 5G and B5G mobile networks. *Frequenz*, *77*(11/12), 599–606. https://doi.org/10.1515/freq-2023-0024

The article provides a comprehensive overview of the evolution of mobile wireless communication to the 5G revolution, High-frequency communications, cellular communications, the architecture, and emerging technologies of 5G networks, and a survey on millimeter-wave communications for fifth-generation wireless networks. Additionally, it discusses the potential of millimeter-wave (mmWave) and terahertz spectrum for 6G wireless with detailed references cited.

The article sets the stage for the subsequent analysis by establishing the context of the study within the broader landscape of wireless communication technologies, including the challenges and opportunities associated with the deployment of mmWave/sub-terahertz communication in 5G and B5G mobile networks.

Moreover, the authors underscore the intricate challenges entailed in the deployment of millimeter-wave communications, with a specific focus on the pronounced influence of foliage on the efficacy of mmWave/sub-terahertz communication. A critical aspect addressed is the substantial power loss encountered by mmWave/sub-terahertz signals as they traverse foliated environments. Delving into the nuances, the article introduces a robust mathematical model meticulously crafted to capture the essence of foliage-induced attenuation, thereby elucidating the gravity of its impact on communication links.

This foundational exploration serves as a precursor to the ensuing analysis, shedding light on the precise environmental variables that demand meticulous consideration for the seamless implementation of mmWave/sub-terahertz communication within the realm of 5G and beyond 5G (B5G) mobile networks. By delineating the mathematical intricacies of foliage loss, the article sets a compelling stage for a more detailed examination of the factors critical to optimizing the performance and reliability of mmWave and sub-terahertz communication in the evolving landscape of advanced mobile networks.

Anzum, R. (2021). Factors that affect LoRa Propagation in Foliage Medium. *Procedia Computer Science*, *194*, 149–155. https://doi.org/10.1016/j.procs.2021.10.068

The article emphasizes the impact on the factors that impact the performance of LongRange (LoRa) propagation in foliage medium. The authors highlight the importance of considering environmental factors such as temperature, foliage, and rainfall, as well as LoRa physical parameter settings and the Fresnel zone.

The authors highlight the limited existing research on LoRa propagation in foliage environments, noting that while the performance of LoRa in indoor and outdoor settings has been studied extensively, there needs to be more research specifically focused on foliage propagation. The article references previous studies that have demonstrated a decrease in range and channel quality for non-line of sight foliage propagation, as well as degradation in the Received Signal Strength Indicator (RSSI) with increasing distance between transmitter and receiver in foliage environments.

The article underscores the impact of foliage on LoRa network deployment. It sets the stage for the current study’s experimental analysis of LoRa propagation in a foliage medium, specifically a line of five date palm trees, and its potential contributions to understanding LoRa propagation channel modeling in foliage environments. The article highlights some of the challenges posed by foliage while deploying LoRa technology. (1) Signal Attenuation: Vegetation or foliage can cause significant signal attenuation, which can reduce the range and quality of LoRa propagation. This can be addressed by optimizing the physical layer parameters of LoRa, such as spreading factors, to improve signal strength and quality. (2) Multipath Propagation: Foliage or Vegetation can also introduce multipath propagation, which can cause interference and reduce the accuracy of signal reception. This can be addressed by using directional antennas or by optimizing the placement of LoRa gateways to minimize the impact of multipath propagation. (3). Environmental Factors: Environmental factors such as temperature, humidity, and rainfall can also impact LoRa propagation in vegetation. These factors can be addressed by carefully selecting the location of LoRa gateways and optimizing the physical layer parameters of LoRa to account for these environmental factors.

Overall, while there are challenges associated with using LoRa technology in the presence of foliage or vegetation, these challenges can be addressed through careful network planning and optimization. Being aware of the existence of foliage or vegetation within the deployment area and having a thorough understanding of it contribute to improved planning and optimization of LoRa. This awareness and comprehension make it feasible to attain reliable and high-quality LoRa propagation, especially in environments with foliage or vegetation.

Nguyen, H. X., Trestian, R., To, D., & Tatipamula, M. (2021). Digital Twin for 5G and beyond. *IEEE Communications Magazine*, *59*(2), 10–15. https://doi.org/10.1109/MCOM.001.2000343

Researchers face several challenges in developing 5G networks to their full potential. These challenges include the complexity of technology, infrastructure deployment, interference and spectrum management, standardization and compatibility, security and privacy concerns, energy efficiency, regulatory and policy challenges, and meeting the diverse requirements of various 5G use cases. Addressing these challenges requires collaborative efforts from researchers, industry stakeholders, and policymakers to ensure the successful development and deployment of 5G networks to their full potential.

The article highlights the potential of digital twin (DT) technology in facilitating the smart deployment of 5G and beyond. It emphasizes the integration of DT with the physical environment to visualize and predict the propagation of 5G radio signals, enabling accurate 3D modeling of urban terrains, buildings, and trees. This integration, along with advanced radio propagation models, allows for the accurate prediction of coverage areas for each base station across the city, which is crucial for successful 5G deployment. Furthermore, the DT enables continuous monitoring, testing, and near-real-time optimizations of the 5G network’s performance. It also supports continuous validation and optimization, relaxation of constraints at the initial stages of 5G services, and flexibility for new use cases. The article underscores the pivotal role of DT in addressing challenges and driving the successful deployment of 5G networks.

Qi, Q., & Tao, F. (2018). Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. *IEEE Access*, *6*, 3593. https://doi.org/10.1109/ACCESS.2018.2793265

The article delving into smart manufacturing and Industry 4.0 underscores the pivotal role played by digital twins and big data in the metamorphosis of traditional manufacturing processes into intelligent, data-driven systems. The authors accentuate the significance of the digital twin in achieving cyber-physical integration in manufacturing by constructing a virtual representation of physical assets and processes. This virtual model serves as a dynamic tool for real-time monitoring, analysis, and optimization of the physical system. The research revolves around the framework and characteristics of IT-driven service-oriented smart manufacturing, underscoring the necessity to address the diverse demands of all stakeholders involved in service collaboration.

In the realm of environmental monitoring systems, the article conducts a comprehensive review of existing systems, emphasizing the criticality of real-time data collection and analysis for smart manufacturing. Notably, the work underscores the pivotal role of digital twins in enabling cyber-physical integration and autonomous decision-making within smart manufacturing systems.

The article enumerates key benefits derived from the integration of digital twin technology. Firstly, the digital twin creates a virtual representation of physical assets and processes, amalgamating data from diverse sources, including sensors, Internet of Things (IoT) devices, and other data sources. Secondly, it facilitates real-time monitoring, allowing for the early detection of faults and anomalies. This is made possible by virtual representation, tracking the behavior of physical assets and processes in real-time. Thirdly, the digital twin serves as a tool for optimizing the physical system by simulating various scenarios and predicting outcomes based on different decisions. Finally, it enables closed-loop control by providing real-time feedback to the physical system, allowing for efficient and effective processes.

In summary, the digital twin is instrumental in creating a virtual representation of physical assets and processes; it empowers real-time monitoring, analysis, and optimization, fostering closed-loop control for more efficient and effective processes.

Mazzacca, G., Grilli, E., Cirigliano, G. P., Remondino, F., & Campana, S. (2022). SEEING AMONG FOLIAGE WITH LIDAR AND MACHINE LEARNING: TOWARDS A TRANSFERABLE ARCHAEOLOGICAL PIPELINE. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVI-2-W1-2022*, 365–372. https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-365-2022

The use of airborne laser scanning (ALS) and light detection and range (LiDAR) technologies has become a well-established practice for identifying and mapping archaeological evidence. LiDAR technology allows for the measurement and mapping of items or structures that would otherwise be hidden under vegetation. The ability to filter the returning signal created by the hit vegetation makes it an essential instrument in areas with dense forest or shrub cover. However, the generation of an accurate Digital Terrain Model (DTM) from LiDAR data relies on various factors, making the creation of a DTM a complex procedure that requires numerous assumptions and decisions during the project planning, data acquisition, and subsequent analytic workflow.

The proposed workflow involves a multi-level multi-resolution (MLMR) point cloud semantic segmentation, which uses machine learning algorithms to classify the 3D dataset into different categories, including vegetation and archaeological structures. The workflow is designed to filter out vegetation and detect hidden archaeological structures directly from LiDAR point clouds. The MLMR procedure involves the development of predictive models using a Random Forest algorithm with reduced manually annotated portions of the datasets, including known elements and discriminative features. The workflow also includes the generation of a bare-ground DTM and the use of visualization techniques for anomaly detection.

The use of machine learning algorithms in the proposed workflow provides a fast and accurate way to filter vegetation and detect archaeological evidence in LiDAR point clouds, facilitating the identification and mapping of archaeological elements above ground and anomalies of potential historical interest at ground level. The workflow is designed to be optimal, fast, and transferable, working in different archaeological environments to distinguish vegetated and non-vegetated areas in LiDAR datasets.

The proposed workflow demonstrates significant advantages in processing large LiDAR datasets, facilitating the otherwise difficult manual identification of hidden heritage evidence and saving time in the process execution. The workflow is designed to spot archaeological artifacts both above and below ground in multiple steps and work entirely on the LiDAR point cloud. It offers fast processing of large datasets through a supervised machine learning approach, accurate vegetation filtering in complex environments, and the detection and mapping of above-ground structures directly on the 3D point cloud. Additionally, the output is easily transferrable to a GIS environment for further data processing, and the classification models can be generalized and transferred to different environments.

Weil, C., Bibri, S. E., Longchamp, R., Golay, F., & Alahi, A. (2023). Urban Digital Twin Challenges: A Systematic Review and Perspectives for Sustainable Smart Cities. *Sustainable Cities and Society*, *99*, 104862. https://doi.org/10.1016/j.scs.2023.104862

The article provides a comprehensive review focusing on Urban Digital Twins (UDTs) and their role in sustainable smart cities; a systematic analysis is presented, which outlines the primary challenges and unresolved issues in their implementation. The review underscores a significant research gap, noting that despite the uptick in UDT-related research, there remains an insufficient exploration of the bottlenecks hindering their deployment. The intention of this study is to bridge this gap through a meticulous examination of the associated challenges and issues. The analysis delineates eight principal categories of challenges that are pivotal to the realization of UDTs. These encompass concerns with interoperability and semantics, foundational infrastructure including storage, computation, and network systems, the intricacies of data acquisition and actuation, and the imperative of ensuring data quality and harmonization. Further, the need for robust modeling, simulation, and decision-support systems is recognized, alongside the criticality of data visualization and the display of information. The review also touches upon the necessity for adequate human and capital resources, and the importance of governance, as well as organizational and social considerations. The review draws particular attention to key issues such as the semantics of data and models, the prevalence of missing data, and the overarching need for data quality and effective modeling practices. These elements are identified as significant hurdles that practitioners must surmount to alleviate the delays in UDT implementation. A methodical approach underpins the review, adhering to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) methodology. This structured approach facilitated a targeted literature search and selection process aimed at unearthing relevant articles that address the myriad challenges encountered in the deployment of UDTs.

In summary, the literature review offers a thorough and insightful overview of the obstacles and unresolved matters surrounding the deployment of Urban Digital Twins in the context of sustainable smart cities. The findings provide a comprehensive resource for practitioners, policymakers, and scholars engaged in the domains of urban development and digital twin technologies, fostering a deeper understanding of the field and guiding future research and application efforts.

Sun, Z., Xue, B., Zhang, M., & Schindler, J. (2023). An Improved Mask R-CNN for Instance Segmentation of Tree Crowns in Aerial Imagery. 2023 38th International Conference on Image and Vision Computing New Zealand (IVCNZ), Image and Vision Computing New Zealand (IVCNZ). https://doi.org/10.1109/IVCNZ61134.2023.10343827

The article presents a new and effective method for segmenting individual tree crowns in aerial imagery, with a focus on the dataset of Wellington City of Aotearoa, New Zealand. The paper begins with an introduction to the importance of instance segmentation in computer vision and the challenges of segmenting individual tree crowns in aerial imagery. The authors then introduce the Mask R-CNN architecture and its limitations in extracting sufficient feature information for individual tree crowns.

To address these limitations, the authors propose an improved Mask R-CNN method by introducing an effective backbone structure, ConvNeXt, and a new mask branch to help segment tree crowns from complex backgrounds. The proposed method is evaluated on the Wellington city dataset, and the results show that it outperforms Mask R-CNN and accurately identifies and segments individual tree canopies.

The authors note that many of these methods are based on Mask R-CNN and that the performance of these methods is limited by the complex canopy structure of trees and the similarity of tree crowns to the surrounding background. The authors also discuss the potential real-world applications and implications of this improved instance segmentation method for tree crowns in aerial imagery, including forest management, urban planning, biodiversity modeling, and pest control.

Overall, “An Improved Mask R-CNN for Instance Segmentation of Tree Crowns in Aerial Imagery” presents a significant contribution to the field of computer vision and forestry management. The proposed method addresses the limitations of previous methods and provides a more accurate and efficient way to identify and segment individual tree crowns in aerial imagery. The paper also highlights the potential applications of this method in various fields, emphasizing the importance of accurate and efficient tree crown mapping for sustainable forest management and urban planning.

# Appendix B Topic Description and Supporting Literature

The “A Study on Creating Digital Twin Foliage Representation Through Computer Vision, Aerial Image Analysis and Machine Learning techniques to enhance the Network Planning and Deployment” research emerges at the intersection of advanced technologies and environmental understanding, aiming to create a nuanced digital replica of natural environments with a particular focus on foliage. As connected devices proliferate and 5G technology becomes pervasive, the project responds to the challenges presented by the deployment of millimeter-wave (mmW) networks, specifically addressing issues related to foliage impact on wireless signal propagation (Anzum, 2021; Chikhale et al., 2022; Farooq & Lokam, 2023). In this endeavor, the research delves into the intricacies of representing foliage within a digital twin, encompassing the spatial distribution, diversity, and inherent characteristics of trees and vegetation. These digital twin models serve as powerful tools, empowering network planners, environmental scientists, and various stakeholders to comprehend how foliage influences the complex dynamics of wireless signal propagation (Nguyen et al., 2021; Qi & Tao, 2018).

The "Digital Twin Representation of Foliage" research embarks on a groundbreaking journey, bridging the realms of technology and environmental comprehension to establish the groundwork for intelligent, efficient, and environmentally conscious network planning. With its focus on modern connectivity challenges, this research is positioned to make a significant contribution to the advancement of wireless communication technologies.

**Objectives**

* Foliage Modeling: Create a comprehensive digital twin of natural environments, emphasizing the spatial distribution and characteristics of foliage.
* Network Optimization: Provide network planners and environmental scientists with a robust tool for intelligent and efficient network planning, ultimately enhancing user coverage and throughput.
* Community Enrichment: Contribute to intelligent urban planning and environmental awareness by extending the digital twin to encompass additional elements like building facades and street infrastructure.
* Wireless Signal Dynamics: Uncover insights into how foliage influences wireless signal propagation, path loss, and coverage, especially in high-frequency networks like millimeter-wave (mmW) networks designed for technologies such as 5G.

**Outcomes**

* Accurate Foliage Representation: A sophisticated digital twin that precisely represents various foliage types, distributions, and characteristics.
* Enhanced Network Planning: Improved efficiency in the planning and deployment of 5G mmW networks, leading to superior coverage and reduced signal interference.
* Holistic Community Representation: Expansion of the digital twin to include additional urban elements, contributing to intelligent urban planning and enriched community experiences.

# Appendix C GitHub Details

GitHub repository contains all the supporting documents related to chapter 1. It contains all the articles referenced in chapter 1.

Link to GitHub repository: https://tinyurl.com/DigitalTwin-nu