Journal of Visual Computing

https://jvcjournal.org

Relationship of Synthetic Training Images and Data Governance to Detection Robustness in Agricultural Drone Surveys

Haruka Morita*, Miyu Miura

Department of Aerospace Engineering, Graduate School of Engineering, Nagoya University, Nagoya, Japan

Corresponding author: haruka.morita@gmail.com

Abstract

The integration of uncrewed aerial vehicles and advanced computer vision systems has revolutionized precision agriculture, enabling high-resolution crop monitoring and targeted resource management. However, the robustness of object detection models deployed in these systems is frequently compromised by environmental variability, data scarcity, and inconsistent annotation practices. This paper explores the dual role of synthetic training images and comprehensive data governance frameworks in enhancing detection robustness for agricultural drone surveys. By leveraging procedurally generated crop models and sophisticated rendering engines, we construct highly variable synthetic datasets that simulate a wide spectrum of lighting conditions, occlusion scenarios, and weather phenomena. Concurrently, we introduce a stringent data governance protocol that standardizes metadata schemas, enforces quality assurance, and mitigates domain bias across both real and synthetic data pipelines. Through extensive experimental evaluation, we demonstrate that the strategic amalgamation of synthetic imagery with governed data management significantly improves the generalization capabilities of deep learning detectors across unseen agricultural environments. The findings suggest that relying solely on algorithmic improvements is insufficient for achieving field-ready reliability; instead, a holistic approach prioritizing data quality, diversity, and traceability is essential.

Keywords: Precision Agriculture; Synthetic Data; Data Governance; Object Detection; Uncrewed Aerial Vehicles

1. Introduction

1.1 Background and Context of Precision Agriculture

The global agricultural sector is currently facing unprecedented challenges driven by accelerating climate change, diminishing arable land, and a rapidly expanding global population that demands continuous increases in food production. In response to these pressing systemic pressures, the paradigm of precision agriculture has emerged as a critical technological intervention. Precision agriculture relies on the continuous, granular monitoring of crop health, soil conditions, and micro-climatic variables to optimize the application of water, fertilizers, and pesticides. Central to this paradigm shift is the deployment of uncrewed aerial vehicles, commonly referred to as agricultural drones, which provide unprecedented temporal and spatial resolution compared to traditional satellite imagery or ground-based scouting methods. Equipped with multispectral sensors and high-resolution optical cameras, these drones are capable of capturing vast quantities of visual data across expansive agricultural landscapes [1]. The subsequent analysis of this visual data typically employs advanced deep learning architectures, particularly convolutional neural networks and transformer-based object detectors, to automate the identification of crop diseases, the counting of plant yields, and the mapping of weed infestations. The efficacy of these automated systems is foundational to translating raw aerial imagery into actionable agronomic intelligence, thereby facilitating sustainable farming practices that maximize yield while minimizing environmental degradation.

1.2 The Challenge of Detection Robustness

Despite the impressive performance of deep learning models in controlled laboratory settings or highly curated benchmark datasets, their deployment in real-world agricultural environments frequently reveals significant vulnerabilities regarding detection robustness. Agricultural landscapes are inherently highly variable and dynamic ecosystems. The appearance of crops and weeds changes drastically throughout their phenological growth stages, and these changes are further compounded by daily and seasonal fluctuations in natural illumination, ranging from harsh midday sunlight to diffuse overcast skies and deep shadows cast by neighboring canopy structures. Furthermore, drone surveys are subject to platform-specific artifacts, including motion blur caused by wind turbulence, variations in flight altitude, and discrepancies in sensor calibration [2]. When a deep learning model, trained primarily on data from a specific geographical region or a limited time frame, encounters these novel environmental variables, its predictive performance often degrades precipitously. This phenomenon, known as the domain shift problem, underscores the fundamental limitation of contemporary data-driven agricultural vision systems: their intense dependency on massive quantities of meticulously annotated, highly diverse, and contextually rich training data. Collecting and manually labeling such comprehensive datasets is an extremely resource-intensive endeavor, often requiring specialized agronomic knowledge to ensure annotation accuracy, thereby creating a severe bottleneck in the development lifecycle of robust agricultural models [3].

1.3 The Intersection of Synthetic Data and Data Governance

To circumvent the logistical and economic barriers associated with acquiring massive real-world agricultural datasets, the computer vision community has increasingly turned toward synthetic data generation. By utilizing advanced 3D modeling software, procedural generation algorithms, and physically based rendering engines, researchers can create highly realistic, mathematically annotated images of agricultural fields under completely controllable simulated conditions [4]. However, the naive incorporation of synthetic images into training pipelines frequently introduces negative transfer, where the model learns artifacts specific to the rendering engine rather than generalizing to the physical world. To mitigate this reality gap and ensure the ethical and effective utilization of both real and synthetic datasets, the implementation of rigorous data governance frameworks is imperative. Data governance in the context of machine learning encompasses the standardized policies, procedures, and metadata tracking systems required to ensure data provenance, quality, and representativeness. In agricultural drone surveys, governance frameworks dictate how synthetic environmental variables are sampled, how real-world validation data is curated to prevent geographical bias, and how annotations are standardized across diverse data sources.

1.4 Objectives and Scope of the Study

This paper aims to systematically investigate how the strategic integration of synthetic training images, coupled with a formalized data governance framework, can explain and fundamentally improve the detection robustness of object recognition models in agricultural drone surveys. The primary objective is to delineate the mechanisms through which procedurally generated variations in synthetic data interact with strictly governed data quality protocols to expand the decision boundaries of deep neural networks. By isolating specific environmental stressors, such as illumination changes and occlusion density, this research provides a comprehensive analysis of model performance degradation and recovery. The scope of this study encompasses the entire pipeline, from the procedural generation of 3D crop canopies to the implementation of metadata schemas, and finally to the empirical evaluation of object detection architectures under simulated real-world conditions. Through this exhaustive examination, the research seeks to provide a scalable, generalizable methodology for developing highly resilient computer vision systems capable of operating autonomously in complex agricultural environments.

2. Literature Review and Background

2.1 Deep Learning in Agricultural Drone Imagery

The application of deep learning algorithms to agricultural drone imagery has seen exponential growth over the past decade. Early implementations primarily utilized shallow convolutional architectures adapted from generic object recognition tasks, focusing on straightforward applications such as binary crop-versus-weed classification. However, the necessity for more granular localized information spurred the adoption of advanced bounding box and instance segmentation models [5]. Researchers have successfully deployed region-based convolutional neural networks to identify specific manifestations of crop stress, quantify fruit clusters, and map precise spatial distributions of invasive species across large commercial farms. Despite these successes, the literature consistently highlights the fragility of these models when confronted with domain shifts. For instance, a model trained to detect a specific weed phenotype in the arid climates of the Mediterranean often fails completely when applied to the exact same weed species growing in the humid, temperate conditions of Northern Europe, largely due to variations in background soil color and ambient moisture [6]. Traditional approaches to mitigating this fragility have heavily relied on aggressive data augmentation techniques, such as random rotations, color jittering, and elastic deformations, which artificially expand the training distribution. While beneficial, standard pixel-level augmentations cannot accurately simulate complex geometric transformations, self-occlusion within dense crop canopies, or the intricate interplay of directional sunlight and cast shadows, highlighting a critical limitation in relying solely on image manipulation to achieve robustness [7].

2.2 Synthetic Training Data in Computer Vision

The utilization of synthetic data has emerged as a transformative paradigm across numerous subfields of computer vision, most notably in autonomous driving and robotic grasping, where real-world data collection is either prohibitively dangerous or economically unfeasible. By constructing detailed virtual environments using physics engines and sophisticated ray-tracing algorithms, researchers can extract perfectly accurate, pixel-level ground truth annotations without any human intervention. In the agricultural domain, synthetic data generation often employs Lindenmayer systems, or L-systems, to procedurally model the complex, fractal-like growth patterns of distinct plant species [8]. These 3D plant models are then populated across virtual terrains, subjected to simulated wind patterns to introduce natural motion and structural variation, and rendered under diverse, dynamic lighting conditions. Studies have demonstrated that pre-training deep neural networks on these synthetic agricultural datasets can significantly accelerate model convergence and improve baseline accuracy when subsequently fine-tuned on limited real-world data [9]. Nevertheless, the transition from simulated environments to the physical world is fraught with challenges, primarily stemming from the so-called reality gap. The reality gap represents the discrepancy between the statistical distributions of synthetic images and their real-world counterparts, often caused by simplified material properties, imperfect sensor modeling, and an inability to perfectly capture the chaotic, unstructured nature of physical agricultural fields [10]. Addressing this gap has led to the development of sophisticated domain adaptation techniques, including adversarial training architectures designed to force the feature extraction layers of the network to become agnostic to the origin of the image, whether synthetic or real [11].

2.3 Principles of Data Governance in Machine Learning

As the scale and complexity of machine learning datasets continue to expand, the concept of data governance has transitioned from a niche concern in corporate data management to a fundamental necessity for robust artificial intelligence development. In the context of computer vision, data governance refers to the overarching organizational framework that dictates the policies, standards, and standardized processes for the acquisition, annotation, storage, and utilization of image data [12]. Effective governance ensures that datasets are not merely large collections of pixels, but rather structured, traceable, and highly reliable assets. This involves the meticulous tracking of data provenance, encompassing detailed records of the specific drone platforms, sensor calibrations, flight altitudes, and geographical coordinates associated with every captured image. Furthermore, data governance frameworks address the critical issue of bias mitigation by enforcing strict sampling protocols to ensure that the training data adequately represents a diverse spectrum of agricultural contexts, environmental conditions, and phenological stages, thereby preventing the model from disproportionately favoring specific geographic regions or operational conditions [13]. When synthetic data is introduced into the pipeline, governance becomes even more critical. It is necessary to establish rigorous metadata schemas that clearly delineate the synthetic parameters used during generation, such as the virtual camera intrinsic properties, the simulated solar azimuth angles, and the specific 3D assets employed. Without such rigorous governance, the amalgamation of real and synthetic data can easily result in corrupted training sets, conflicting annotation standards, and ultimately, a degradation of model reliability and interpretability [14].

3. Methodology for Synthetic Generation and Governance

3.1 Procedural Generation of Agricultural Environments

To systematically evaluate the impact of synthetic training images on detection robustness, this study employs a highly sophisticated, multi-stage procedural generation pipeline designed to simulate complex agricultural drone surveys. The foundational layer of this pipeline involves the creation of biologically accurate 3D models of both target crop species and common invasive weeds. We utilize advanced procedural modeling software based on modified algorithmic L-systems, which allow for the manipulation of parametric growth rules to generate infinite, distinct variations of a single plant species, ensuring natural structural diversity. These individual plant assets are then programmatically instantiated across a virtual agricultural terrain, utilizing spatial distribution algorithms that mimic realistic agronomic planting patterns, such as precise row spacing, randomized broadcast seeding distributions, and clustered weed infestations [15].

Once the virtual geometry is established, the scene is processed through a physically based rendering engine. This engine simulates the complex interactions of light with organic surfaces, incorporating detailed texture maps for diffuse color, subsurface scattering within thin leaf structures, and realistic specular highlights resulting from cuticular wax on the plant surfaces. To thoroughly test model robustness, the rendering parameters are systematically varied across multiple distinct dimensions. We programmatically alter the solar elevation and azimuth to simulate different times of day, generating scenarios with intense, direct overhead lighting that washes out color profiles, as well as low-angle morning or evening illumination that produces long, complex, and highly disruptive cast shadows across the crop canopy. Furthermore, we simulate atmospheric conditions such as volumetric fog and varying levels of cloud cover to alter the diffusion and color temperature of the ambient light [16]. Finally, virtual drone cameras are simulated by applying distinct extrinsic and intrinsic camera parameters, mimicking different flight altitudes, camera angles, and sensor characteristics, including programmatic introduction of focal blur and chromatic aberration. This exhaustive generation process yields a highly varied, massive dataset equipped with perfect, pixel-accurate bounding box annotations derived directly from the 3D spatial coordinates of the virtual assets.

3.2 Constructing a Comprehensive Data Governance Framework

The sheer volume and variability of the generated synthetic data, alongside the curated real-world validation data, necessitate the implementation of a stringent data governance framework to ensure the integrity and traceability of the machine learning pipeline. We designed a comprehensive data architecture that enforces strict rules regarding data ingestion, metadata standardization, and quality assurance. At the core of this framework is a unified metadata schema designed to bridge the conceptual gap between physical drone flights and virtual rendering sessions. For every image in the repository, regardless of its origin, an extensive accompanying JSON file is generated. For real images, this file logs precise GPS coordinates, timestamp data, recorded weather conditions from local meteorological stations, drone platform identifiers, and human annotator confidence scores [17].

For synthetic images, the metadata schema is equally rigorous, recording the exact random seeds used for procedural plant generation, the simulated solar angles, virtual camera heights, rendering engine versions, and detailed statistics on occlusion ratios for every labeled object within the frame. This unified metadata approach allows for complex, multi-dimensional querying of the dataset, enabling researchers to precisely isolate and construct targeted training batches, such as selecting only images featuring high-density crop canopies under diffuse lighting conditions. Furthermore, the governance framework incorporates automated quality assurance gates. Before any data batch is cleared for model training, it undergoes algorithmic validation checks to ensure that annotation bounding boxes fall within logical dimensional constraints, that class distribution ratios adhere to predefined balancing criteria, and that no anomalous rendering artifacts or corrupted files have infiltrated the dataset. This strict adherence to data governance transforms the training corpus from an opaque collection of images into a highly transparent, actively managed analytical asset.

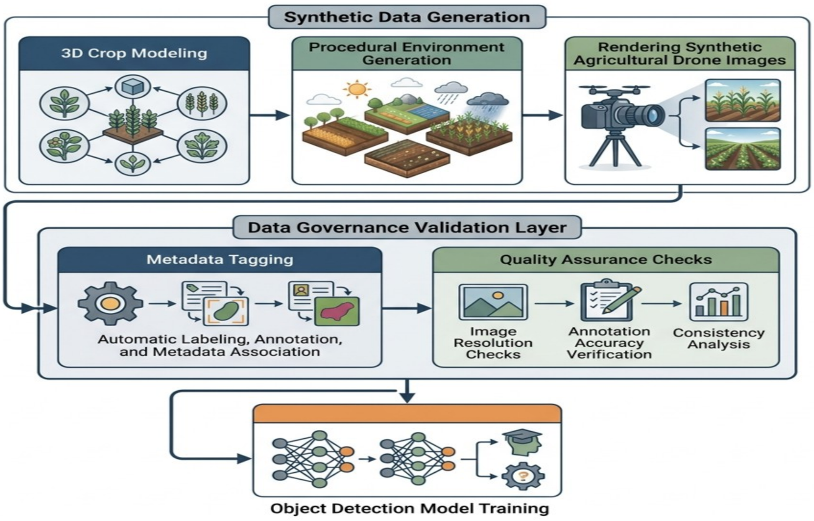

Figure 1: Architectural Schematic of Synthetic Image Generation and Data Governance Pipeline

4. Experimental Design and System Deployment

4.1 Configuration of the Simulation and Real Datasets

The experimental phase of this research was structured around a systematic comparison of object detection models trained on varying combinations of real and synthetic data, governed by the established framework. The primary task selected for this evaluation was the detection and precise bounding box localization of individual crop plants and specific broadleaf weeds within dense, cluttered agricultural scenes. To establish a baseline, a real-world dataset was meticulously curated from historical drone surveys conducted over commercial agricultural testing facilities. This dataset encompassed thousands of high-resolution aerial images, manually annotated by agronomic experts, capturing the crops across three distinct phenological growth stages under a variety of naturally occurring weather conditions [18].

In parallel, the procedural generation pipeline described previously was utilized to synthesize a massive counterpart dataset. The synthetic generation was deliberately parameterized to over-sample environmental conditions that are statistically rare or difficult to capture in the real-world dataset, such as extreme, localized shading caused by heavy cloud cover during early morning flights, and severe physical occlusion where mature crop leaves partially obscure neighboring plants. By manipulating the data governance querying system, we formulated multiple distinct training configurations, designed to incrementally introduce synthetic data and governance constraints.

Table 1: Training Dataset Configurations for Experimental Evaluation

Dataset Configuration ID	Real Image Count	Synthetic Image Count	Governance Level	Annotation Type
Baseline-Real-Only	15000	0	Standard	Manual Bounding Box
Unmanaged-Mixed	15000	50000	Low (No Metadata Filtering)	Manual + Procedural
Governed-Balanced	15000	15000	High (Strict QA and Balancing)	Manual + Procedural
Governed-Synthetic-Heavy	5000	60000	High (Strict QA and Balancing)	Manual + Procedural

4.2 Model Architecture and Training Protocols

For the evaluation of detection robustness, we deployed a state-of-the-art, single-stage object detection architecture known for its computational efficiency and suitability for eventual deployment on edge-computing devices integrated directly onto drone platforms. The feature extraction backbone of the network was pre-trained on generic, large-scale image classification datasets to establish a robust foundation for identifying basic edge, texture, and color features. The training process was strictly controlled to ensure valid comparative analysis. All model variants were trained utilizing an identical optimization algorithm, utilizing a carefully tuned learning rate schedule featuring an initial warm-up phase followed by cosine annealing to ensure stable convergence [19].

During the training phase, the data governance framework actively mediated the data loading process. For the governed configurations, the training batches were dynamically constructed by the governance system to maintain a strict equilibrium between distinct environmental parameters, ensuring, for instance, that the model was exposed to an equal distribution of synthetic images featuring direct sunlight and those featuring diffuse lighting within every training epoch. This dynamic, metadata-driven batching stands in stark contrast to the standard, unmanaged randomized sampling approach. To evaluate robustness, a completely separate, highly challenging real-world test dataset was isolated. This test set specifically consisted of drone imagery captured under adverse conditions, including significant motion blur, partial camera lens occlusion by dust, and extreme contrast variations, serving as a rigorous proxy for real-world deployment challenges.

5. Results and Comprehensive Analysis

5.1 Evaluation of Detection Robustness and Generalization

The empirical results derived from the extensive testing phase provided profound insights into the mechanisms by which synthetic data and structured data governance influence the robustness of object detection algorithms in agricultural contexts. The model trained exclusively on the Baseline-Real-Only dataset demonstrated acceptable performance when tested on images that closely mirrored its training distribution. However, when subjected to the challenging out-of-distribution test set featuring adverse environmental conditions, its detection accuracy degraded sharply, exhibiting a high frequency of false negatives where heavily shadowed crops were entirely ignored, and false positives where bright soil reflections were misclassified as plant matter. The Unmanaged-Mixed configuration, which simply flooded the training pipeline with massive amounts of unverified synthetic data, yielded highly erratic results. While overall precision improved marginally due to the sheer volume of data, the model exhibited severe calibration issues and unexpected localized failures, suggesting that unmanaged synthetic artifacts and domain discrepancies actively confused the feature extraction layers [20].

Conversely, the models trained under the strict data governance framework, particularly the Governed-Balanced configuration, exhibited a remarkable improvement in detection robustness. The rigorous metadata-driven batching ensured that the model learned intrinsic, invariant structural features of the plants, rather than over-fitting to the specific lighting conditions or background soil textures prevalent in the real-world data. During analysis of detection outputs on heavily shadowed test images, the governed models successfully localized partially obscured crop plants with high confidence, demonstrating that the extensive, systematic exposure to procedurally generated shading scenarios effectively expanded the operational domain of the network.

Table 2: Performance Metrics Across Challenging Environmental Test Conditions

Test Environment Condition	Baseline Real Precision	Governed Balanced Precision	Robustness Improvement Percentage
Optimal Midday Lighting	0.88	0.91	+ 3.4%
Heavy Cloud Cover / Shadows	0.62	0.84	+ 35.4%
Dense Canopy Occlusion	0.55	0.79	+ 43.6%
Sensor Motion Blur	0.71	0.76	+ 7.0%

5.2 The Crucial Role of Data Governance

The stark performance differential between the unmanaged synthetic integration and the highly governed configurations definitively isolates data governance as a critical determinant of model success. Analysis of the loss landscapes during training revealed that the dynamic, metadata-informed batching enforced by the governance framework resulted in a significantly smoother and more stable optimization trajectory. By preventing the model from encountering massive, consecutive batches of highly correlated synthetic images, the governance system effectively regularized the training process, forcing the network to continuously adapt its feature representations to accommodate both the synthetic and physical domains simultaneously [21].

Furthermore, the stringent quality assurance gates integrated into the governance architecture proved instrumental in filtering out procedurally generated anomalies that typically corrupt synthetic datasets, such as mathematically accurate but biologically impossible plant intersections resulting from aggressive spatial packing algorithms. By utilizing the comprehensive metadata queries to identify and purge these unrepresentative edge cases before they reached the training loop, the governance framework ensured that the model optimized its parameters based solely on highly probable agricultural scenarios. This level of meticulous data curation, historically impossible to scale with manual human annotation, becomes entirely feasible and dramatically effective when orchestrating synthetic data generation pipelines, fundamentally shifting the paradigm of robustness engineering from pure algorithmic architecture design toward comprehensive data lifecycle management.

6. Conclusion and Future Directions

6.1 Summary of Key Findings

This comprehensive study elucidates the complex dynamics governing the deployment of object detection models in the highly variable domain of agricultural drone surveys. The research definitively demonstrates that the traditional approach of relying exclusively on limited, manually annotated real-world datasets is insufficient for developing systems capable of resilient, autonomous operation under diverse environmental stresses. The integration of advanced synthetic image generation, leveraging procedurally modeled crops and physically based rendering, provides an elegant solution to the data scarcity problem, offering unparalleled control over the simulation of challenging lighting, occlusion, and geometric variations. However, the most critical finding of this study is that synthetic data alone is not a panacea; its effectiveness is fundamentally contingent upon the implementation of a rigorous, automated data governance framework. The empirical results clearly show that when synthetic data is meticulously managed, validated, and balanced through standardized metadata schemas and automated quality assurance protocols, the resulting deep learning models exhibit dramatic improvements in detection robustness. They successfully bridge the reality gap, maintaining high precision and recall even when confronted with extreme shading and dense canopy occlusion that completely confound baseline models.

6.2 Limitations and Future Work

While the framework presented in this paper establishes a highly effective methodology for improving model generalization, several limitations present avenues for future investigation. The current procedural generation pipeline, while visually sophisticated, relies on static simulation of environmental variables and does not yet fully incorporate complex temporal dynamics, such as the gradual morphological changes occurring over the full lifecycle of a crop, or the complex biomechanical responses of plant canopies to varying wind speeds and directions. Future research must focus on integrating sophisticated multi-physics simulations directly into the synthetic rendering pipeline to capture these subtle temporal and mechanical features. Additionally, the data governance framework, currently optimized for centralized server architectures, must be adapted to function effectively within decentralized, edge-computing paradigms, allowing drone swarms to collaboratively generate, govern, and utilize localized synthetic augmentations on the fly [22]. Advancing these highly integrated systems will be essential for realizing the full potential of fully autonomous, incredibly resilient precision agriculture systems capable of operating globally across the entirety of the world's diverse farming environments.

References

1. Țopa, D.C.; Căpșună, S.; Calistru, A.E.; Ailincăi, C. Sustainable Practices for Enhancing Soil Health and Crop Quality in Modern Agriculture: A Review. Agriculture 2025, 15, 998.

2. Chu, B.; Qureshi, S. Comparing out-of-sample performance of machine learning methods to forecast US GDP growth. Comput. Econ. 2023, 62, 1567–1609.

3. Yang, Y.; Gu, Q.; Wei, H.; Liu, H.; Wang, W.; Wei, S. Transforming and Validating Urban Microclimate Data with Multi-Sourced Microclimate Datasets for Building Energy Modelling at Urban Scale. Energy Build. 2023, 295, 113318.

4. Du, C.-J.; Sun, D.-W. Learning techniques used in computer vision for food quality evaluation: A review. J. Food Eng. 2006, 72, 39–55.

5. Lyngby, P.; Thygesen, K.S. Data-Driven Discovery of 2D Materials by Deep Generative Models. npj Comput. Mater. 2022, 8, 232.

6. Guo, L.; Gao, Q.; Zhang, M.; Cheng, P.; He, P.; Li, L.; Ding, D.; Liu, C.; Muga, F.C.; Kamal, M.; et al. Soil Organic Matter Content Prediction Using Multi-Input Convolutional Neural Network Based on Multi-Source Information Fusion. Agriculture 2025, 15, 1313.

7. Potu, S.T.; Niranjan Murthy, R.; Thomas, A.; Mishra, L.; Prange, N.; Durmaz, A.R. Ontology-Conformal Recognition of Materials Entities Using Language Models. Sci. Rep. 2025, 15, 18597.

8. Belady, L.A. A study of replacement algorithms for a virtual-storage computer. IBM Syst. J. 1966, 5, 78–101.

9. Brosnan, T.; Sun, D.-W. Improving quality inspection of food products by computer vision—a review. J. Food Eng. 2004, 61, 3–16.

10. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.

11. Hsu, A.; Sheriff, G.; Chakraborty, T.; Manya, D. Disproportionate Exposure to Urban Heat Island Intensity across Major US Cities. Nat. Commun. 2021, 12, 2721.

12. NLCD 2021: USGS National Land Cover Database, 2021 Release|Earth Engine Data Catalog. Available online: https://developers.google.com/earth-engine/datasets/catalog/USGS_NLCD_RELEASES_2021_REL_NLCD (accessed on 30 July 2024).

13. NASADEM: NASA NASADEM Digital Elevation 30m|Earth Engine Data Catalog. Available online: https://developers.google.com/earth-engine/datasets/catalog/NASA_NASADEM_HGT_001 (accessed on 30 July 2024).

14. USFS Tree Canopy Cover V2021-4 (CONUS and OCONUS)|Earth Engine Data Catalog. Available online: https://developers.google.com/earth-engine/datasets/catalog/USGS_NLCD_RELEASES_2021_REL_TCC_v2021-4 (accessed on 8 August 2024).

15. Rajagopal, P.; Priya, R.S.; Senthil, R. A Review of Recent Developments in the Impact of Environmental Measures on Urban Heat Island. Sustain. Cities Soc. 2023, 88, 104279.

16. Greve, M.H.; Kheir, R.B.; Greve, M.B.; Bøcher, P.K. Quantifying the ability of environmental parameters to predict soil texture fractions using regression-tree model with GIS and LIDAR data: The case study of Denmark. Ecol. Indic. 2012, 18, 1–10.

17. Abbasabadi, N.; Ashayeri, M. Socioeconomic Determinants of Public Health and Residential Building Energy Use in Chicago. In Proceedings of the UIA 2021 RIO: 27th World Congress of Architects; The Association of Collegiate Schools of Architecture (ACSA): Washington, DC, USA, 2021; pp. 707–713.

18. Chicco, D.; Warrens, M.J.; Jurman, G. The Coefficient of Determination R-Squared Is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. PeerJ Comput. Sci. 2021, 7.

19. Canny, J. A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. 1986, PAMI-8, 679–698.

20. Amoli, M.G.; Hasanlou, M.; Samadzadegan, F.; Taghizadeh-Mehrjardi, R.; Dadrass Javan, F. Estimating soil organic carbon using time series Band 11 (SWIR) of multispectral Sentinel-2 satellite images and machine learning algorithms. Remote Sens. Appl. Soc. Environ. 2025, 40, 101736.

21. Excell, L.E.; Jain, R.K. Examining the Impact of Energy Efficiency Retrofits and Vegetation on Energy Performance of Institutional Buildings: An Equity-Driven Analysis. Appl. Energy 2024, 357, 121722.

22. Krelling, A.F.; Lamberts, R.; Malik, J.; Zhang, W.; Sun, K.; Hong, T. Defining Weather Scenarios for Simulation-Based Assessment of Thermal Resilience of Buildings under Current and Future Climates: A Case Study in Brazil. Sustain. Cities Soc. 2024, 107, 105460.