The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W15-2025
20th 3D GeoInfo Conference 2025, 2–5 September 2025, Kashiwa, Japan

Predicting Building Height from Footprint and Urban Planning information for Digital Twin Generation

Jue Ma[1,*], Chenbo Zhao[2], Yoshiki Ogawa[3], Yoshihide Sekimoto[4]

[1] Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan - majue@csis.u-tokyo.ac.jp
[2] Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan - zhaocb@csis.u-tokyo.ac.jp
[3] Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan - ogawa@csis.u-tokyo.ac.jp
[4] Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan - sekimoto@csis.u-tokyo.ac.jp

Keywords: Building Height, Urban Digital Twin, Machine Learning, Roof Type

Abstract

Accurate building height data is essential for constructing realistic and analytically useful 3D city models within digital twin systems. However, such data are often incomplete, particularly in small to medium-sized cities. This study presents a machine learning-based approach to predict building heights by integrating multi-source urban data, including building footprints, zoning regulations, and roof-type classifications. To enhance prediction accuracy, we clustered zoning types by height profiles and trained models separately for each group. We evaluated three regression algorithms—Random Forest (RFR), Support Vector Regression (SVR), and XGBoost—using stratified sampling and cross-validation. Among them, RFR achieved the highest overall accuracy, particularly in homogeneous zoning areas (R^2 = 0.67), while SVR showed lower generalizability. The proposed model has been implemented in an open-source "3D City Model Generation Simulator," enabling automated LOD3-level urban model generation using only remote sensing and street-view inputs. This work contributes a scalable and cost-effective solution for urban digital twin applications in data-sparse environments.

1. Introduction

Knowing building height is essential for understanding urban process regimes (Zhu et al., 2019), enabling improved urban management and planning. Despite advancements in global scale datasets, such as those developed by Kamath et al. (2024), comprehensive building-height information for Japanese cities remains notably absent. Most existing studies rely heavily on imagery-based methods, including satellite imagery (Cao and Huang, 2021), street view images (Yan and Huang, 2022) or Synthetic Aperture Radar (SAR) data derived from Sentinel-1 and Sentinel-2 time series (Frantz et al., 2021). For instance, Yang and Zhao (2022) employed spatially-informed Gaussian Process Regression with Sentinel-1 data for major Chinese cities, while Cai et al. (2024) utilized building footprints combined with a self-adaptive buffer for building photon selection methods to improve height estimations. Additionally, how building height is determined in the absence of building height regulatory restrictions has been explored by Chau et al. (2007), emphasizing contextual factors influencing building morphology.

To address these gaps, we propose a novel building-height prediction method specifically tailored to Japanese urban contexts. In contrast to previous studies that primarily measure height for existing buildings, our objective extends to provide building-height estimations necessary for generating hypothetical urban scenarios within urban digital twins. This functionality is essential for simulation urban development processes and evaluating planning policies through digital twin applications.

For the first time, our approach integrates urban planning information, explicitly considering building footprints, local zoning regulations, and detailed roof-type attributes derived directly from building geometry. Incorporating roof-type features predicted from footprints, alongside traditional urban morphological characteristics, our method significantly enhances predictive accuracy. The main contributions of our research are: it fills the critical data gap in global building-height datasets, particularly addressing the absence of detailed Japanese data; and it provides a robust, high-resolution building-height generation model directly applicable to digital twin frameworks, facilitating the generation of realistic 3D urban models.

* Corresponding author

1.1 Methods

1.2 Data

Building data
This study utilizes the PLATEAU dataset, a detailed 3D urban model produced from aerial surveys and other geospatial measurements, developed and maintained by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) in Japan. As of August 2024, when our dataset was acquired, PLATEAU provides publicly accessible 3D urban model data covering 212 municipalities in Japan (Figure 1). We employed Plateau-kit (Ozeki, 2024), an open-source tool designed specifically for accessing and processing PLATEAU data are available in GeoJSON, CityJSON, and Parquet formats; in this research, we utilize the Parquet format. Both our target variable—measured building height—and urban morphological features of buildings are sourced directly from this dataset.

Urban planning data
Urban planning information incorporated in our study primarily refers to the "Use Districts" (Japanese: *youto chiiki*), a foundational zoning regulation in Japan. These districts regulate building usage and parameters such as Floor Area Ratio (FAR) and Building Coverage Ratio (BCR), and are categorized into 12 distinct types covering residential, commercial, and industrial purposes (Ma et al., 2024). Given that zoning regulations significantly influence building height and form in Japanese cities, we regard them as crucial inputs for urban digital twin applications that simulate urban planning

scenarios, and thus involve them in our model. We sourced this zoning regulation data from the MLIT Urban Planning Information Dataset (*Toshi-keikaku Kettei Joho*), which provides nationwide coverage. We found that zoning information for 98 municipalities was already integrated into the PLATEAU dataset, enabling direct usage for our analysis.

Road network data

In addition, the National Digital Road Map Database (DRM) was involved to calculate two urban morphological characteristics—level and width of the adjacent road—as features in our model. These road network data were provided by the Japan Digital Road Map Association.

Roof type data

Finally, we introduced a roof type dataset generated through a roof classification algorithm (Section 1.4). This model integrates satellite imagery with building footprint data to classify roof structures into several typical roof categories, achieving classification accuracy exceeding 95%. Roof type information is uncommon in most building datasets; hence, its inclusion provides valuable architectural insights that enhance the accuracy of our predictive model.

Figure 1. Example of Building Height data used in this study (Shibuya, Tokyo)

1.3 Roof Classification

1.4 Roof Type Classification

To incorporate semantic information on roof structures into the building height prediction model, we defined five roof types: Flat (FL), Stepped-flat (SFL), Folded (FD), Hipped (HP), and Gable (GB). For each building, we extracted the footprint outline at level of detail (LOD) 0 from the CityGML files based on its building ID and then cropped the corresponding aerial image (0.3-meter resolution) provided by PLATEAU. To ensure sufficient image quality for classification, the shorter side of the cropped image was constrained to exceed 40 pixels.

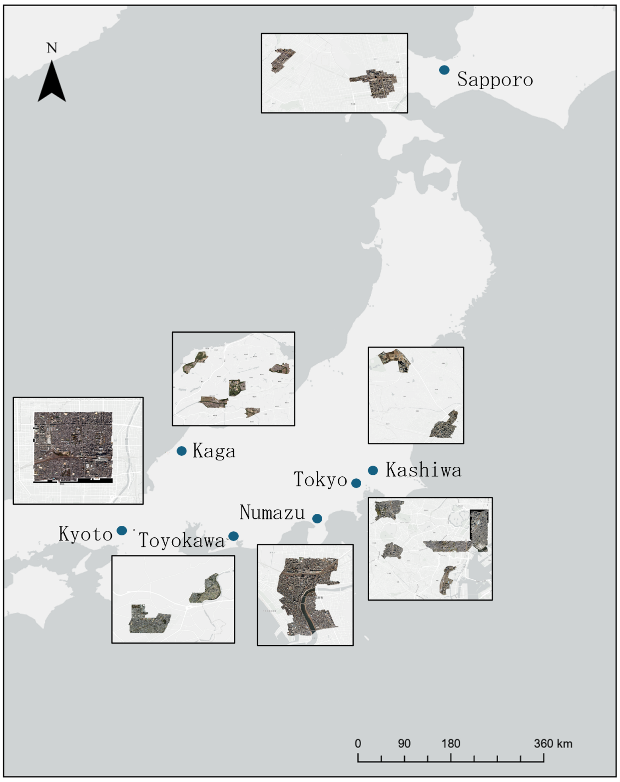

Figure 2. Roof type classification dataset distribution in Japan

The training data for roof classification were collected from various cities across Japan, including Sapporo, Kashiwa, Tokyo, Numazu, Kaga, Toyokawa, and Kyoto. The geographic distribution of training samples is illustrated in Figure 2. And as shown in Table 1, a total of 56,955 buildings were used for roof classification, with detailed breakdowns by both city and roof type. The examples of those different roof type is shown in Figure 3.

Table 1. Number of buildings used for roof classification by city and by roof type

Category	Type	Number of buildings
City	Sapporo	2,753
	Kashiwa	2,495
	Tokyo	14,087
	Numazu	8,022
	Kaga	3,864
	Toyokawa	5,109
	Kyoto	20,625
	Total (city)	**56,955**
Roof type	Flat (FL)	13,771
	Stepped-flat (SFL)	6,027
	Folded (FD)	18,514
	Hipped (HP)	3,609
	Gable (GB)	15,007
	Total (roof)	**56,955**

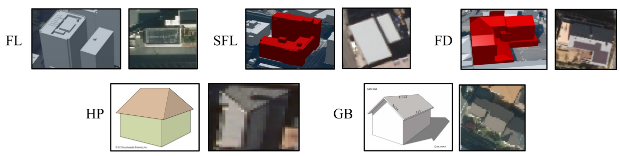

Figure 3. Examples of different roof type

After completing the data annotation, we fine-tuned the EVA-

02-base model (Fang et al. (2024)), which was pre-trained on ImageNet, using our roof classification dataset. EVA-02 is a lightweight yet powerful vision transformer model that achieves state-of-the-art performance in both fine-tuned and zero-shot settings with significantly fewer parameters and training data. Specifically, we split the dataset into 80% training and 20% validation sets, fine-tuned the model for 100 epochs, and achieved an average validation F1 score of 95%. The The confusion matrix is shown in Figure 4.

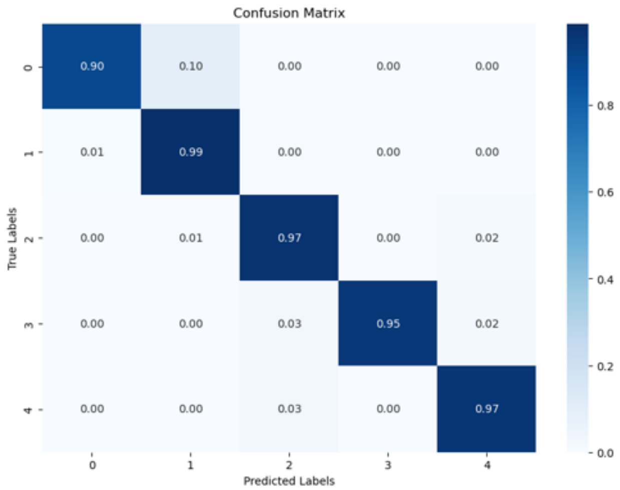

Figure 4. Confusion matrix of roof classification validation

Upon completion of the roof classification training, we performed inference on a total of 2,080,508 buildings across 11 cities in Japan—Sapporo, Tatebayashi, Nerima, Yokosuka, Kaga, Suwa, Toyohashi, Yokkaichi, Himeji, Kurume, and Mashiki. The resulting roof type data were then incorporated into the feature engineering process for building height prediction.

1.5 Feature Engineering

Referring to existing building-height prediction models (Milojevic-Dupont et al., 2020; Stipek et al., 2024), and incorporating knowledge from urban science, we selected a total of 13 predictive features (Table 2). These features cover five categories: building geometric characteristics, neighborhood context, zoning regulations, roof type classification, and road information.

Among these, building geometric characteristics hold predictive information about its height, particularly the footprint area (Biljecki and Sindram, 2017). Neighborhood-level context is captured through the number of adjacent buildings (within a 1-meter buffer) and the number of nearby buildings (within a 25-meter radius), offering additional insight into urban density and configuration. Zoning regulations are encoded through a simplified categorical variable that maps the original 12 Japanese “Use District” types into three broader categories. Roof types, rarely available in public datasets, are integrated via a classification model developed by our research group using remote sensing imagery. Finally, road characteristics such as the classification and width of the nearest road segment are included as proxies for urban accessibility and infrastructure intensity.

Unlike some prior studies that incorporate multi-scale or street-block-level urban features, we deliberately excluded higher-scale spatial attributes. This decision is motivated by our objective: to develop a building-level prediction model that is compatible with fine-grained urban digital twin generation. In this context, where predictions are needed for hypothetical or newly planned structures, block-level or city-wide features may not be computable or even available. Thus, we limited our model input to features that are locally derivable and scalable within a 3D city model framework.

1.6 Data preprocessing

To construct predictive features from raw geospatial datasets, we implemented a comprehensive preprocessing pipeline that includes geometry parsing, feature computation, neighborhood context extraction, roof-type integration, zoning recategorization, and optional road information linkage. In total, 13 features were constructed and used for training predictive models.

Data Cleaning. Building data was read from Parquet files in the PLATEAU dataset. We retained key fields: building ID, geometry, height (`measuredHeight`), zoning category (`districtsAndZonesType`), and roof type label. For data cleaning, we selected cities that simultaneously included complete datasets mentioned above. Buildings with measured heights below 2.4 meters (one floor) and large outlier values–the threshold was set as 40 meters as this value corresponds to high industrial buildings, skyscrapers, and high residential tower blocks (Milojevic-Dupont et al., 2020)-as well as any observations with missing values, were removed to ensure data integrity and predictive reliability.

Geometric Features. Geometry was converted from WKB to CRS EPSG:32654. Based on building footprints, we computed multiple geometric features, including area, perimeter, compactness, number of vertices, length&width&slimness of the minimum bounding rectangle (MBR), and complexity.

Neighborhood Features. To describe local spatial context, neighborhood features such as the number of adjacent buildings and nearby buildings were also extracted. The original zoning categories (*youto chiiki*) were mapped into three aggregated categories based on building use and regulatory stringency, encoded into a new categorical variable `category`.

Zoning-Based Grouping. Due to substantial intra-class variance in building height across Japan’s 12 standard zoning types (*Use Districts*), we applied K-means clustering on statistical descriptors of building height per zoning type (mean, std, max, min, median, skewness). The output clusters were used to partition the dataset into three representative groups. This allowed more tailored model training per cluster, improving accuracy and robustness.

Roof Type Integration. We merged roof-type labels generated by a deep learning model . These labels were matched using building GML IDs. The model achieved over 95% classification accuracy and provides essential structural information absent in most public datasets.

Road Information. Road level and width data from the Japan Digital Road Map (DRM) dataset were matched to each building by computing the nearest road geometry using spatial indexing. This step enabled inclusion of road classification attributes such as `rdclasscd` and `rdwdcd`.

1.7 Machine Learning Methods

To train predictive models for building height, we implemented a supervised machine learning workflow incorporating both

Feature	Description
Compactness	Normalized Perimeter Index (NPI)
Perimeter	The total perimeter length of the building footprint (m)
Vertices	The number of vertices in the polygon
Length of MBR	Length of the minimum bounding rectangle (MBR)
Width of MBR	Width of the minimum bounding rectangle (MBR)
Slimness	Ratio of the length to width of the MBR
Complexity	Shape complexity calculated by perimeter divided by area
Number of adjacent buildings	Number of buildings adjacent to the target building
Number of neighbours	Number of buildings within a fixed radius around the target building
Level of adjacent road	Hierarchy level of the adjacent road
Width of adjacent road	Width (m) of the adjacent road
Roof type	Categorized roof type

Table 2. Features selection

baseline and optimized models. Our approach consists of zoning-aware model training, hyperparameter tuning via cross-validation, and evaluation using multiple error metrics.

Model Selection. Supervised learning methods have been widely applied in building-height prediction tasks (Milojevic-Dupont et al., 2020). We selected three representative regression algorithms based on prior studies and scalability: Random Forest Regression (RFR), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost). Each model was trained on subsets of data from five representative Japanese municipalities (Sendai, Maebashi, Omuta, Chino, Tokushima), totaling 15,036 building samples. The data was randomly split into 70% training and 30% testing subsets.

Stratified Data Splitting. Considering that a substantial proportion of buildings were below 10 meters in height, we applied stratified sampling on the target variable (`measuredHeight`) to ensure balanced height group representation across training and testing splits. Buildings were grouped into five height categories: 0–10m, 10–20m, 20–40m. This ensured the model learned effectively across the full range of urban forms.

Zoning-Based Modeling. The data was divided into three zoning groups according to the previously defined `category` variable. Each group was modeled separately to reflect differences in building regulation and morphology. A combined model was also trained on the full dataset for benchmarking.

Cross-Validation and Optimization. For each model, we performed 5-fold cross-validation using randomized search over a grid of hyperparameters. For RFR, parameters included number of estimators, maximum depth, minimum samples per split/leaf, and feature selection strategy. For SVR, we tuned loss functions, epsilon margins, regularization strength (C), iteration limits, and dual formulation. For XGBoost,

Evaluation Metrics. Model performance was assessed using Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Both bare models (default hyperparameters) and optimized models were tested to compare performance gains. Evaluation was conducted on both full datasets and zoning-specific subsets.

Implementation. All experiments were implemented in Python using `scikit-learn`, `xgboost` and executed on a multi-core computing environment. Feature scaling was applied consistently to all input features using min-max normalization. This framework enabled us to robustly assess the sensitivity of model performance to both hyperparameter configurations and urban regulatory heterogeneity, offering insights into how building morphology and policy context affect predictive accuracy.

2. Results

After integrating multiple data sources, we constructed a comprehensive dataset covering 5 Japanese cities. Due to variations in data availability across cities, not all datasets were fully matched; however, the final sample includes only buildings with complete geometry and attribute records. Figure 5 illustrates the distribution of measured building heights after data cleaning. Subfigure (a) presents city-level variations in building height, with Sendai showing a higher median and broader interquartile range compared to other cities as the only one metropolitan area. Subfigure (b) displays the overall distribution across the entire dataset, where most buildings cluster below 10 meters, indicating a highly skewed height distribution dominated by low-rise structures. This skewness highlights the necessity of stratified sampling in the model training process.

The predictive performance of the three supervised learning algorithms with testing data is presented in Table 3, with the joint plot of predicted values over target values for each data group (Figure 6. Overall, RFR achieved the best predictive performance among the three models, particularly in terms of the coefficient of determination (R^2) and mean absolute error (MAE). Notably, RFR obtained the highest R^2 value of 0.67 in Group 2, which corresponds to the zoning category with the most consistent height distribution. This suggests that the model captures the relationship between building features and height most effectively in more regulated or homogeneous areas. SVR showed relatively high MAE and unstable R^2 values, especially in Group 1, indicating poor generalization for more diverse urban forms. Although the full-dataset performance of SVR was comparable to RFR in terms of MAPE, its predictive reliability was notably lower. XGBoost yielded consistent results across groups but underperformed compared to RFR. Its relatively lower R^2 values and higher errors suggest that while robust.

3. Discussion

3.1 Model Implementation in Digital Twin Generation

In digital twin city systems, building height is a critical component for generating accurate and realistic 3D urban models, directly influencing the fidelity and analytical capabilities of virtual environments. However, in practical scenarios such as urban planning, disaster simulation, or energy analysis, complete building height data is often unavailable, particularly in newly developed areas or small to medium-sized cities lacking comprehensive surveying resources. The building height prediction model proposed in this study addresses this gap by lever-

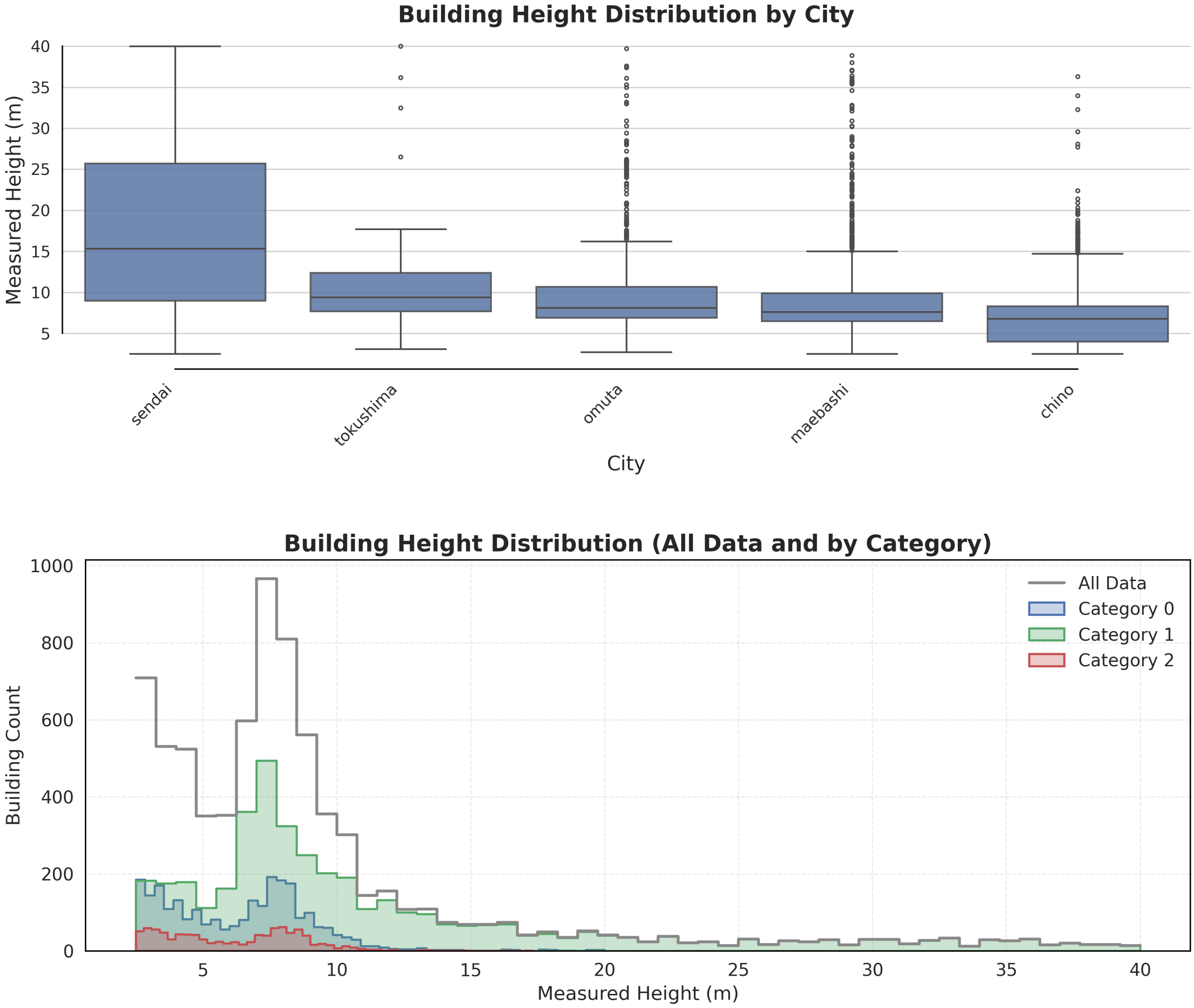

Figure 5. Building height statistics used in this study: (a) distribution by city and (b) full distribution across whole dataset.

aging building footprints, urban zoning regulations, and roof-type attributes to estimate missing height information with high accuracy. This substantially enhances the spatial completeness and usability of digital twin systems.

By integrating our model into digital twin platforms, urban planners can dynamically generate or modify 3D building representations, enabling applications such as policy simulation, development potential assessment, and urban morphology analysis. Furthermore, the model can be deployed without relying on high-precision survey data, making it a cost-effective and scalable solution for resource-constrained municipalities.

In practice, the proposed model has already been implemented as a key component in a branch of our "3D City Model Generation Simulator," where it provides height estimates for generated buildings and contributes to the automated creation of realistic 3D city models in user-defined areas. The simulator has been released as an open-source project on GitHub. By simply providing remote sensing imagery and street-view images as input, users can generate LOD3-level 3D urban models for their regions of interest. The simulator is available as an open-source project on GitHub: https://github.com/Project-PLATEAU/3D-City-Model-Generator.git

References

Biljecki, F., Sindram, M., 2017. Estimating building age with 3D GIS. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, 4, 17–24.

Cai, P., Guo, J., Li, R., Xiao, Z., Fu, H., Guo, T., Zhang, X., Li, Y., Song, X., 2024. Automated Building Height Estimation Using Ice, Cloud, and Land Elevation Satellite 2 Light Detection and Ranging Data and Building Footprints. *Remote Sensing*, 16(2), 263.

Cao, Y., Huang, X., 2021. A deep learning method for building height estimation using high-resolution multi-view imagery over urban areas: A case study of 42 Chinese cities. *Remote Sensing of Environment*, 264, 112590.

Chau, K.-W., Wong, S., Yau, Y., Yeung, A., 2007. Determining Optimal Building Height. *Urban Studies*, 44(3), 591-607. https://doi.org/10.1080/00420980601131902.

Model	Metric	Full dataset	Group 0	Group 1	Group 2
RFR	MAPE	41.23%	26.81%	52.91%	27.97%
	MAE	3.50	1.40	5.28	1.20
	R^2	0.52	0.49	0.53	0.67
SVR	MAPE	40.64%	30.10%	46.83%	32.83%
	MAE	3.91	1.59	6.04	1.43
	R^2	0.35	0.38	0.58	0.63
XGBoost	MAPE	43.48%	30.50%	56.66%	27.43%
	MAE	3.87	1.57	5.65	1.28
	R^2	0.37	0.37	0.46	0.52

Table 3. Comparison of predictive performance for different machine learning models

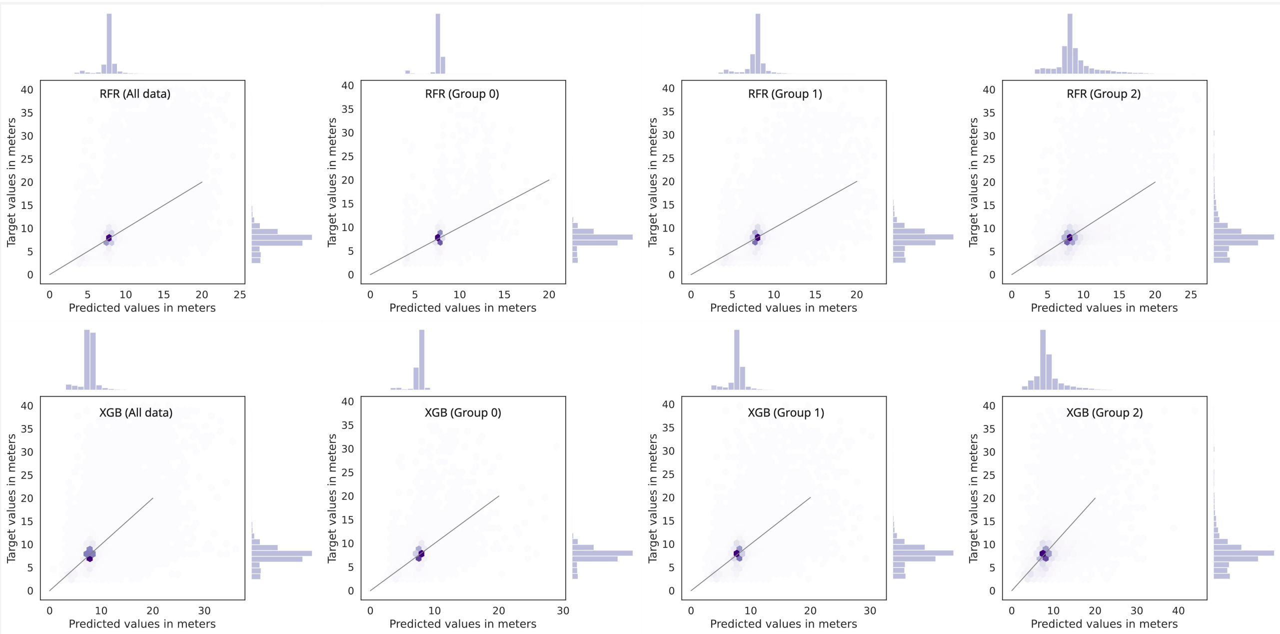

Figure 6. Results of the predictions with test data for each group (Top: Random Forest Regression (RFR), Bottom: Extreme Gradient Boosting

Fang, Y., Sun, Q., Wang, X., Huang, T., Wang, X., Cao, Y., 2024. EVA-02: A visual representation for neon genesis. *Image and Vision Computing*, 149, 105171. https://www.sciencedirect.com/science/article/pii/S0262885624002762.

Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., van der Linden, S., Hostert, P., 2021. National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series. *Remote Sensing of Environment*, 252, 112128.

Kamath, H. G., Singh, M., Malviya, N., Martilli, A., He, L., Aliaga, D., He, C., Chen, F., Magruder, L. A., Yang, Z.-L. et al., 2024. GLObal Building heights for Urban Studies (UT-GLOBUS) for city-and street-scale urban simulations: development and first applications. *Scientific Data*, 11(1), 886.

Ma, J., Shibuya, Y., Pang, Y., Omata, H., Sekimoto, Y., 2024. A cost-and-effect simulation model for compact city approaches: A case study in Japan. *Cities*, 152, 105212. https://www.sciencedirect.com/science/article/pii/S0264275124004268.

Milojevic-Dupont, N., Hans, N., Kaack, L. H., Zumwald, M., Andrieux, F., de Barros Soares, D., Lohrey, S., Pichler, P.-P., Creutzig, F., 2020. Learning from urban form to predict building heights. *PLOS ONE*, 15(12), 1-22. https://doi.org/10.1371/journal.pone.0242010.

Ozeki, K., 2024. Plateaukit: A toolkit for plateau data processing. `https://ozekik.github.io/plateaukit/`. Accessed: 2024-08-25.

Stipek, C., Hauser, T., Adams, D., Epting, J., Brelsford, C., Moehl, J., Dias, P., Piburn, J., Stewart, R., 2024. Inferring building height from footprint morphology data. *Scientific reports*, 14(18651). https://doi.org/10.1038/s41598-024-66467-2.

Yan, Y., Huang, B., 2022. Estimation of building height using a single street view image via deep neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192, 83–98.

Yang, C., Zhao, S., 2022. A building height dataset across China in 2017 estimated by the spatially-informed approach. *Scientific Data*, 9(1), 76.

Zhu, Z., Zhou, Y., Seto, K. C., Stokes, E. C., Deng, C., Pickett, S. T., Taubenböck, H., 2019. Understanding an urbanizing planet: Strategic directions for remote sensing. *Remote Sensing of Environment*, 228, 164–182.