Building Classification: A Comprehensive Dataset and DenseNet201-Based Approach

Abstract

As urban areas expand, accurately classifying diverse building types is critical for sustainable urban planning, resource allocation, and disaster response. However, real-world challenges such as irregular building orientations, shadows, and geographic diversity hinder multi-class building classification. This study addresses these complexities by leveraging DenseNet-201, a robust deep learning architecture, to classify buildings across seven categories: hospitals, high-rise buildings, commercial structures, schools, single-unit residential buildings, industrial facilities, and multi-unit residential buildings. We utilize the USA Building dataset, comprising 15,329 high-resolution satellite images spanning all 50 U.S. states, ensuring broad geographic representation and diverse urban conditions. Our model achieved 84.39% classification accuracy, outperforming existing methods while demonstrating resilience to distortions caused by angular layouts and shading. Enhanced with precise geolocation tagging, the framework enables detailed spatial analysis of regional building distributions, providing actionable insights for urban planners and policymakers. The integration of a user-friendly interface further allows for customizable model training, making the system adaptable to specific regional needs. This combination of a comprehensive dataset, advanced classification architecture, and practical tools represents a significant advancement in automated building classification technology. Applications of this framework span zoning regulations, infrastructure planning, environmental impact studies, and emergency response, offering a data-driven approach to urban management. By improving the capacity to classify complex urban structures, this work supports the development of resilient, inclusive, and sustainable cities. Ultimately, this research bridges the gap between machine learning innovation and real-world urban challenges, empowering stakeholders to make informed decisions for dynamic and equitable urban growth.

1. Introduction

Buildings shape our cities, defining skylines and influencing how we interact with our urban environments. Accurately identifying and classifying these structures from satellite imagery is crucial for urban planning, disaster response, and land-use management (Farajzadeh et al., 2023; Wu et al., 2020). Classification typically involves sorting buildings into categories (e.g., single-family, commercial, industrial) based on size, shape, and context (Ithape et al., 2023). Despite advances in satellite technology providing high-resolution images of urban areas, interpreting this data accurately remains a challenge. Traditional methods often struggle in complex urban environments where buildings can look similar from above; shadows can obscure important features; and densely built areas can make individual structures hard to distinguish (Erdem & Avdan, 2020; Vasavi et al., 2023; Alsabhan et al., 2022). Variations in architectural styles and construction materials compound these challenges, underscoring the need for robust classification methods (Atwal et al., 2022).

The task of accurately classifying buildings presents many interconnected challenges in today's rapidly evolving urban landscapes. Cities are dynamic, with buildings coming in countless shapes, sizes, and uses, often changing their purposes over time as communities grow and adapt (Ithape et al., 2023). Modern buildings frequently serve multiple functions, such as apartment complexes with ground-floor retail spaces or office buildings that include both commercial and residential areas, making them particularly difficult to categorize. Traditional field surveys where people walk through cities to document structures can provide detailed information but are

increasingly impractical, requiring significant time and resources that quickly become outdated (Adha et al., 2022; Hu et al., 2023).

This is where satellite technology has transformed our approach to building classification. Modern satellites can capture detailed images of entire cities in minutes, providing regular updates that reveal new construction, demolition, and changes in land-use patterns (Abburu & Golla, 2015; Vasavi et al., 2023). These high-resolution images offer detailed insights into building layouts with minimal effort compared to traditional surveying, making them indispensable for monitoring urban development (Alsabhan et al., 2022). Nevertheless, accurately isolating buildings from satellite images remains challenging. Trees can overlap or cast shadows on buildings, roads may look similar to rooftops, and natural features like water or terrain can produce reflections and shadows that obscure critical details (Erdem & Avdan, 2020; Vasavi et al., 2023; Alsabhan et al., 2022). The challenge becomes even more complicated when considering the variety of buildings themselves from sprawling industrial complexes to tightly packed urban homes, each with unique roof designs, orientations, and architectural features (Huang et al., 2022; Ji et al., 2019). Shadows cast by tall buildings can hide smaller structures, and different times of day or seasonal changes can alter how buildings appear, demanding sophisticated methods that can reliably distinguish these variations (Reda & Kedzierski, 2020; Lloyd et al., 2020).

While researchers have made significant progress in developing methods to overcome these obstacles, current approaches still face important limitations. Many existing studies focus on classifying only a few building types, which does not reflect the rich diversity of real cities (Alsabhan et al., 2022; Vasavi et al., 2023). Accurate boundary detection remains especially challenging in dense, shadowed, or region-specific urban environments (Huang et al., 2022; Lloyd et al., 2020). Additionally, researchers lack access to comprehensive datasets that include a broad array of building types from varied cultural and architectural contexts. This constraint hinders the development of universally robust methods (Dimassi et al., 2021; Ji et al., 2019). Larger, well-structured datasets are therefore critical for advancing the reliability and global applicability of building classification.

Given these significant challenges and limitations, our paper aims to provide a thorough analysis of existing methods and chart a path for future developments. We examine how different researchers have approached the complex task of classifying buildings from satellite images, focusing on what has worked well and where improvements are needed. Our analysis highlights recurring issues such as handling building shadows, identifying complex shapes, and creating systems robust across different cities and countries (Huang et al., 2022; Erdem & Avdan, 2020).

In response to these challenges, our research makes three key contributions to advance the field of building classification:

1. We introduce a high-resolution satellite imagery dataset (15,329 images) specifically designed to address gaps in existing resources. It emphasizes challenging urban scenarios, irregular architectures, occluded structures, and densely packed areas, providing a robust foundation for training models on real-world complexities.
2. Leveraging DenseNet201's efficiency in processing intricate spatial hierarchies, our method achieves accurate classification across seven building types (e.g., hospitals, high-rises, industrial facilities). This granularity bridges a critical gap in urban classification systems, which often oversimplify structural diversity.
3. To maximize real-world applicability, we integrate coordinate mapping for geolocalization and a user-friendly interface for model customization. These tools empower urban planners and responders to adapt the system to region-specific needs, enhancing scalability across contexts.

By addressing the limitations of current methods and providing new tools and approaches, we aim to improve the accuracy and reliability of building classification from satellite imagery. Our work has the potential to benefit urban planners, emergency responders, and other professionals who rely on accurate building information, ultimately contributing to better-informed decisions and more efficient management of urban environments.

2. Related Work

As urban landscapes continue to evolve, accurately identifying and classifying buildings from satellite imagery has become an essential task for urban planning, disaster management, and infrastructure development. In practice, this involves going beyond mere building detection such as distinguishing between a building and a non-building and moving toward more nuanced classifications that reflect diverse building types, including single-unit residential homes, multi-unit residential complexes, commercial offices, industrial facilities, schools, hospitals, and high-rise structures. Current research efforts have often focused on limited building categories, covered relatively small geographic areas, or struggled to manage architectural complexity. To address these gaps, our work aims to introduce a comprehensive approach that classifies multiple building types across a wide range of environments and urban forms.

This research is grounded in the theoretical frameworks of remote sensing, urban morphology, and geospatial informatics. From a remote sensing perspective, satellite imagery is not merely a passive record of the Earth's surface; it is a dynamic data source that, when analyzed with advanced machine learning models, can reveal patterns of urban development and land use. Urban morphological theory emphasizes understanding the form, structure, and evolution of cities. Within this context, building classification serves as a critical tool for interpreting how various socio-economic, environmental, and cultural factors shape the built environment. By situating our study at the intersection of these frameworks, we aim to advance the theoretical understanding of urban complexity through more accurate and detailed building classification.

Early approaches frequently relied on specialized datasets that contained annotated aerial or satellite images, often focusing on a single region or a narrow set of building types. For example, Erdem and Avdan [1] used the Inria Aerial Image Labeling Dataset to classify buildings in large-format images (5,000×5,000 pixels) from Chicago, demonstrating the potential of a modified U-Net architecture with skip connections. Their work reached an accuracy of about 87.69%, yet struggled in densely constructed areas where building footprints overlapped, making it difficult to capture complex geometries fully.

Building upon such detection-oriented methods, Vasavi et al. [2] expanded the classification task using high-resolution imagery (12,025×5,878 pixels) from Nashik, India, collected via the SASPlanet tool. Employing a U-Net with a ResNet-34 backbone, they classified buildings into categories like residential, industrial, and holy places with an accuracy of 89%. While this represented progress in moving beyond binary classification, challenges remained in consistently distinguishing similar building types particularly in complex urban layouts.

Other dataset-driven efforts have traded off granularity for broad land-cover categorization. Helber et al. [5], for instance, introduced the EuroSAT dataset, leveraging Convolutional Neural Networks (CNNs) such as ResNet-50 and GoogleNet to classify 27,000 images (64×64 pixels) into general land-use classes. They achieved an accuracy of 98.57%, but their focus was on broad categories rather than detailed building distinctions. Similarly, Atwal et al. [6] incorporated OpenStreetMap (OSM) data from various U.S. counties, utilizing decision tree classifiers to categorize buildings into multiple classes. Although they reached a 98% accuracy rate, OSM's

inconsistent annotations and missing attributes led to reliability issues, and the building categories remained relatively coarse.

Deep learning architectures, especially those inspired by encoder-decoder frameworks, have profoundly influenced building classification. U-Net models are common, as seen in Erdem and Avdan's study [1] and extended by Vasavi et al. [2] and Alsabhan et al. [3], who tested U-Nets with backbones like ResNet-50 and VGG16. While these models captured both global context and fine-grained details, their focus often remained on binary building extraction rather than robust multiclass classification.

Researchers have pursued increasingly nuanced approaches. Dimassi et al. [9] introduced the Beirut Buildings Type Classification dataset (17,033 images) to distinguish residential from non-residential buildings, achieving a 94.8% accuracy using a RexNet model. Ji et al. [10] employed a Siamese U-Net on 220,000 images from Christchurch, New Zealand, showing that large datasets could improve performance. Yet they faced difficulties with shadows, complex backgrounds, and mixed land-use areas issues that simple category sets and localized datasets cannot adequately address.

Other advanced architectures have shown promise in this field. Reda and Kedzierski [7] applied a Faster Edge Region Convolutional Neural Network (FER-CNN), achieving 93% accuracy across six building categories but struggling with small structures and overlapping shadows. Huang et al. [11] experimented with state-of-the-art models like Mask R-CNN and SOLOv2 on images from Beijing and Munich, though they still encountered issues with diagonal rooftops and merged building clusters.

Alternative data sources and novel neural network architectures have also been explored in the field. Kusz et al. [12] utilized LiDAR data to classify 93,440 images from Hamilton County, Indiana, distinguishing residential from non-residential buildings using U-Net. While effective for common structures, their model underperformed on large, atypical building forms like warehouses, likely due to their rarity in the training set. Hang and Cai [13] focused on rooftop shapes and sizes for classifying industrial and residential buildings from Gaofen-2 imagery, but their limited consideration of additional building attributes such as height or function restricted the model's real-world applicability.

A significant challenge in building classification lies in ensuring models can generalize across diverse geographical contexts and architectural styles. Many studies have focused on single regions or cities such as Chicago [1], Nashik [2], Guangdong [3], Beirut [9], or Christchurch [10] making it difficult for their methods to adapt to different climates, cultural norms, and construction materials. While some research has attempted broader coverage, the lack of a truly national-scale or multinational-scale dataset has hindered efforts to develop models that handle wide-ranging urban morphologies.

Architectural complexity further complicates these efforts. Buildings vary greatly in shape, orientation, color, and roofing materials. Shadows and occlusions distort visible features, and high-density areas create challenges in differentiating one structure from another. While sophisticated models like FER-CNN or Mask R-CNN can address some complexities, fully capturing the subtleties of diverse building types remains an ongoing issue.

Across this body of research, two critical gaps emerge. First, most existing studies emphasize either binary classification or a small number of broad building categories, overlooking the nuanced differences among various building types. Second, few datasets provide nationwide or globally representative samples that incorporate the full spectrum of urban environments. Without this diversity, it is challenging for models to learn robust patterns that remain reliable across different regions, climates, and architectural styles.

Our study directly addresses these gaps by offering a more comprehensive classification scheme that includes seven distinct building types. Unlike previous research limited by regional scope or restricted building categories, our dataset spans all 50 U.S. states, ensuring that a wide variety of architectural styles, environmental conditions, and urban forms are represented. By integrating a DenseNet-201 architecture an advanced deep learning model known for efficiently reusing learned features and capturing subtle details we aim to create a system capable of managing the full complexity of building classification tasks. DenseNet-201 layers are densely connected CNN components that enhance feature propagation and minimize gradient vanishing problems, making them suitable for distinguishing finer structural characteristics of different building classes.

Building on previous research, our primary question is: How can we accurately classify a diverse range of building types across large and varied geographic areas using high-resolution satellite imagery and advanced deep learning models? We hypothesize that a comprehensive dataset covering multiple U.S. regions, combined with a DenseNet-201-based classification approach, will outperform earlier methods that focused on limited categories and geographic scopes. By bridging the gaps left by previous studies, our research aims to provide deeper insights into urban morphology, improve resource allocation for urban planners, and offer a valuable tool for emergency responders and infrastructure managers ultimately contributing to a more holistic understanding of the built environment.

Figure 1: Building Classification Architecture
(A diagram illustrating the architecture and components of a building classification system, including key modules such as DenseNet201, Feature Extraction, Dataset, and Training.)

Dataset	Image Source	Year of publication	class	Study area	channels	Image Size(Pixel)	GSD(m/pixel)	Number of Labeled Pixels (billion)	coverage	Accuracy
Atwal et al. (Atwal et al., 2022)	OpenStreet Map (OSM) (*OpenStreet Map*, n.d.)	2022	Non-residential, residential	Fairfax County in Virginia, Mecklenburg County in North Carolina, and the City of Boulder in Colorado					United Sates	98%
Li et al. (Li et al., 2021)		2021	New and old rural buildings	rural Xinxing County, Guangdong Province, China	RGB	900 × 900 to 1024 × 1024 pixels	0.26		china	
Ji et al. (Ji et al., 2019)	Wuhan Dataset (WHU)	2019	countryside, residential, culture, and industrial area	Christchurch, New Zealand	RGB	512 × 512	0.075-2.7	57.67	New Zealand	
Dimassi et al. (Dimassi et al., 2021)	Beirut Buildings Type Classification (BBTC)	2021	Residential, Non-residential	Beirut city	RGB		1.19		Lebanon	94.8%

Erdem & Avdan (Erdem & Avdan, 2020)	Inria Aerial Image	2020	Buildings	Chicago	RGB	5000 x 5000	0.3	67.875	United States	87.69%
Hang & Cai (Hang & Cai, 2020)	Gaofen-2	2020	industrial and residential building roofs	Changchun		256× 256			china	
Helber et al. (Helber et al., 2019)	EuroSAT	2019	AnnualCrop, Forest, HerbaceousVegetation, Highway, Industrial, Pasture, SeaLake PermanentCrop, Residential, River	34 European countries		64×64	10	0.11	global	98.57%
X. Huang et al. (X.	SuperView and Gaofen-2 satellites	2022	residential, com	Beijing in China and Munich		600× 600		5.38	China and Germany	

Huang et al., 2022)			mercial, industrial, public and other	in Germany						
Lloyd et al.(Lloyd et al., 2020)	Maxar Technologies building footprints, OpenStreet Map (OSM) building footprints, highways, Democratic Republic of the Congo (COD)-building points for Kinshasa and North Ubangi, Nigeria (NGA)-household survey data, Democratic Republic of the Congo-household survey data, Global Man-made Impervious Surface (GMIS) Dataset from Landsat, v1	2021	residential or nonresidential	Congo, Nigeria					Nigeria	93%
Reda & Kedzierski, 2020(Reda &	WorldView-2 and Pléiades	2020	shopping center, bloc	western part of Warsaw(Poland)	RGB	512 × 512	0.5 m	0.13	Poland	93%

Kedzierski, 2020)			k of flats, church, terraced houses, single-family house and garage							
Kusz et al. (Kusz et al., 2021)	LIDAR	2021	Non-residential, residential	Hamilton County, Indiana	RGB	256x256		6.12	United Sates	

Table 1: Previously published building classification studies using satellite imagery

3. Methodology

Our research aimed to develop a robust deep learning model capable of classifying various building types commercial, single-unit residential, multi-unit residential, industrial, high-rise, hospital, and school using high-resolution satellite imagery. The central objective was to determine whether a convolutional neural network (CNN), trained on a large and geographically diverse dataset, could reliably distinguish these categories, even amidst differences in architectural style, environmental context, and geographic region.

We acquired 512×512-pixel images at approximately 0.15 m/pixel resolution, ensuring that subtle details such as rooftops, building footprints, and adjacent land use patterns were preserved. This level of detail is crucial because it allows the model to pick up on fine-grained visual cues that often differentiate one building class from another. Figure 1 provides a bar chart illustrating the total number of images collected for each building category, and Table 2 (presented below) offers a comprehensive breakdown of the number of images collected per building class and per state, confirming our dataset's broad thematic and spatial coverage.

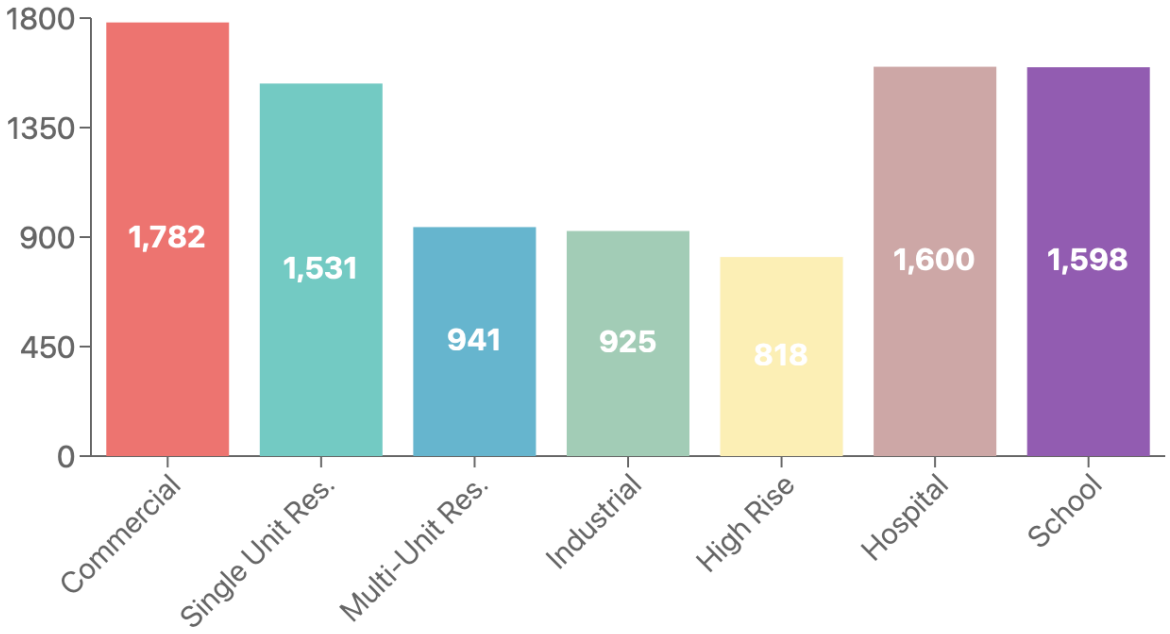

Figure 2: Total Counts per Building Category
(A bar chart showing image counts for each building class, ensuring a strong representation of commercial, hospital, and other key categories.)

State	Commercial	Single-Unit	Multi-Unit	Industrial	High-Rise	Hospital	School	Total
AR	0	0	8	85	19	40	95	247
AZ	20	23	5	32	28	43	20	171
CA	334	100	192	12	6	46	64	754
CO	16	2	2	0	0	42	5	67
CT	5	7	5	45	40	40	33	175
DC	10	5	5	0	5	0	19	44
DE	0	0	0	38	12	16	21	87
FL	67	43	32	19	34	45	20	260
GA	92	110	79	24	29	40	17	391
HI	137	139	60	0	0	31	12	379
IA	7	6	8	0	0	40	1	62
ID	27	51	12	13	0	40	1	144
IL	226	88	76	38	66	44	44	582
IN	4	6	1	35	23	41	6	116
KS	50	55	28	32	3	41	10	219
KY	0	0	0	0	0	41	6	47
LA	0	0	0	0	0	43	1	44
MA	21	11	15	0	0	40	25	112
MD	62	87	46	5	27	42	41	310
ME	26	36	5	5	0	40	2	114
MI	0	0	0	21	20	41	6	88

MN	17	17	4	2	0	41	1	82
MO	9	63	7	76	31	41	185	412
MT	25	46	3	4	0	40	3	121
NC	0	0	0	90	39	40	216	385
NE	250	40	89	5	2	41	24	451
NJ	0	0	0	49	45	41	286	421
NM	42	66	17	34	5	40	27	231
NV	52	73	21	21	2	41	7	217
NY	139	36	130	26	157	40	66	594
OH	14	5	13	0	0	44	3	79
OK	39	22	6	2	0	40	4	113
OR	15	19	2	13	0	41	0	90
PA	58	41	31	12	44	42	14	242
RI	0	0	0	43	6	20	18	87
SC	18	30	15	8	3	43	12	129
SD	0	97	0	30	0	40	7	174
TN	0	16	0	0	0	41	14	71
TX	0	172	0	34	76	45	37	364
UT	0	19	0	2	0	40	1	62
WI	0	0	24	70	96	43	224	457

Table 2: Number of Images Collected per Building Class and State
(This table details the distribution of collected imagery across multiple U.S. states, indicating the geographic diversity and comprehensive coverage achieved.)

To further illustrate geographic diversity, Figure 3 displays a U.S. map where states are color-coded by the number of collected images. This visual confirms that data were sampled from all 50 states, encompassing dense urban centers, sprawling suburbs, industrial hubs, and more remote rural areas. Volunteers spent approximately two months identifying suitable sites to ensure maximal diversity. Images were organized into class-specific directories and meticulously documented with associated metadata, such as geographical coordinates and date of acquisition.

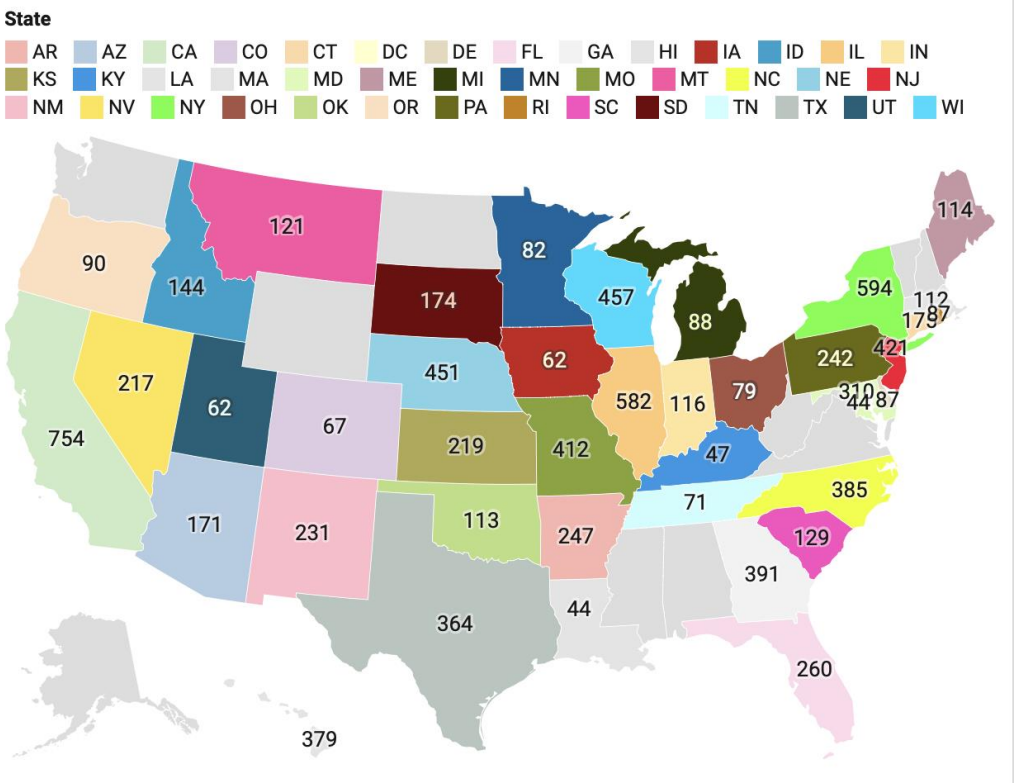

Figure 3: Geographic Distribution of Collected Images by State
(A U.S. map visualization indicating the number of images collected per state, demonstrating broad geographic coverage across rural, suburban, and urban settings.)

To ensure reliable ground truth labels, annotators used Label Studio to draw polygons around buildings and assign them to the predefined categories. A multi-step quality control process where a second annotator reviewed each image, and disagreements were resolved through consensus ensured high annotation fidelity. We measured inter-annotator consistency with Cohen's Kappa (κ=0.85)

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

Where popo is the observed agreement and pepe the expected agreement by chance. Achieving κ=0.85κ=0.85 indicated strong labeling consistency. To provide a visual sense of the annotation quality, Figure 3 shows a sample annotated image, highlighting how building footprints were delineated and classified.

Single-unit residential

school

Multi-unit residential

Industrial

High building

Hospital

Commercial

Figure 4: Example of an Annotated Satellite Image
(This figure presents a single annotated image with polygons drawn around buildings and labels assigned to each structure, illustrating the rigorous annotation procedure.)

Before training, images were verified or resized (using bilinear interpolation) to a uniform dimension of 512×512 pixels. All pixel intensity values were normalized to [0,1]:

$$I_{norm}(x, y) = \frac{I(x, y)}{255.0}$$

This normalization stabilized training by ensuring consistent input scales. To avoid data redundancy, we computed MD5 hashes for each file and removed duplicates:

$$\text{MD5}(I) = \sum_{j=1}^{L} f(b_j) \backslash mod\ 2^{128}$$

Where bjbj are the bytes of the image file and ff is the compression function in MD5.

Addressing class imbalance was critical. We identified over- and under-represented classes and employed a combination of under sampling large classes and data augmentation for minority classes. Augmentations included random flips, rotations (±15°), zooms (90–110%), and brightness/contrast adjustments:

$$I' = T(I), \quad T \in \{\text{Flip,Rotate,Zoom,Brightness,Contrast}\}$$

These transformations were applied on-the-fly during training, increasing data variability and enhancing model robustness. After balancing, we split the dataset into training (80%), validation (10%), and test (10%) sets, maintaining class distributions across subsets and ensuring that images from the same geographic area did not appear in multiple sets. Table 3 shows the final number of

images per class for each subset, confirming that all classes were well represented and appropriately balanced.

Class	Training Set	Validation Set	Test Set	Total
Commercial	1,426	178	178	1,782
Single-Unit	1,225	153	153	1,531
Multi-Unit	753	94	94	941
Industrial	740	93	93	925
High-Rise	654	82	82	818
Hospital	1,280	160	160	1,600
School	1,278	160	160	1,598
Total	7,356	920	919	9,195

Table 3: Number of Images per Class in Training, Validation, and Test Sets
(This table presents the class-wise distribution after splitting the dataset, ensuring balanced and fair evaluation.)

Prior to finalizing our model selection, we conducted exploratory experiments on several well-established convolutional neural network (CNN) architectures to assess their classification performance on our curated dataset. We evaluated DenseNet169, DenseNet121, MobileNetV2, InceptionV3, VGG19, and EfficientNetV2, recording the highest validation accuracy each could attain. Figure 5 presents these preliminary results, confirming that DenseNet201 exhibited a significant advantage over other models, achieving an accuracy of 0.8439, substantially higher than the next best architecture. Given this substantial margin of improvement, DenseNet201 was selected as the backbone for subsequent fine-tuning and optimization endeavors.

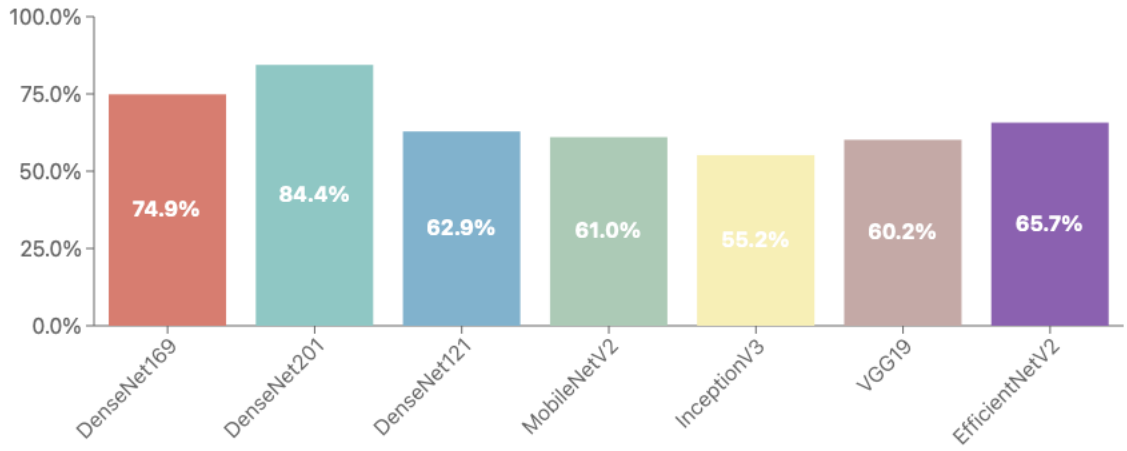

Figure 5: Model Performance Comparison

(A bar chart showing the performance metrics of different neural network models, with DenseNet201 achieving the highest accuracy at 84.39%.)

We chose DenseNet-201 as our base model because its densely connected layers mitigate vanishing gradients and encourage feature reuse, resulting in richer and more effective representations. Each convolutional layer applies:

$$\boldsymbol{h}^{(l)} = \sigma\left(\boldsymbol{W}^{(l)} * \boldsymbol{h}^{(l-1)} + \boldsymbol{b}^{(l)}\right)$$

Where σσ is a nonlinear activation (e.g., ReLU). DenseNet's skip connections concatenate features from all previous layers, enabling efficient gradient flow and better utilization of learned features.

We initialized DenseNet-201 with ImageNet-pretrained weights to capitalize on generic low-level visual features like edges and textures. To adapt the model to our building classification task, we froze the initial layers and fine-tuned only the latter layers. We replaced the top layers with a Global Average Pooling (GAP) layer, a fully connected layer of 768 units with L2 regularization (λ=0.001λ=0.001):

$$\mathcal{L}_{reg} = \lambda \sum_{i,j} W_{ij}^2$$

and a dropout layer (0.5) to minimize overfitting. A final softmax layer yielded class probabilities:

$$\widehat{y_c} = \frac{\exp(z_c)}{\sum_k \exp(z_k)}$$

Figure 6 illustrates the DenseNet-based model architecture, highlighting dense blocks, the GAP layer, and the final classification layers specifically tailored for building types.

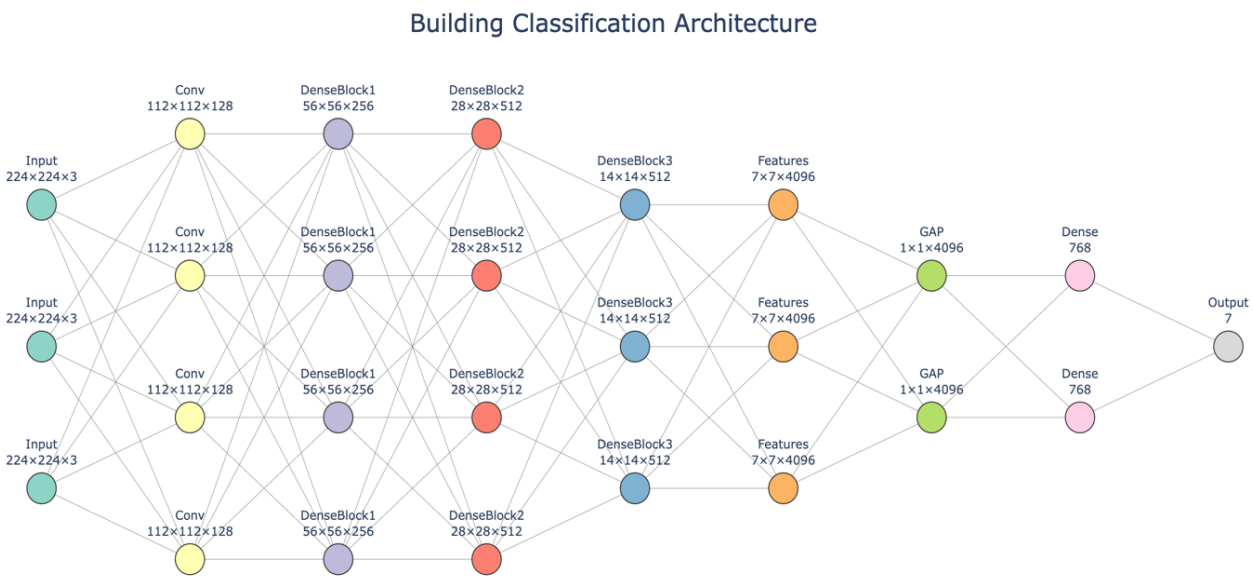

Figure 6: DenseNet-201-Based Building Classification Architecture
(A schematic depicting dense connectivity, global average pooling, and the final classification layers, adapted from the ImageNet-pretrained DenseNet-201 model.)

We used TensorFlow (2.x) with Keras APIs, GPU acceleration, and mixed-precision training for computational efficiency. Reproducibility was ensured by fixing random seeds for Python, NumPy, and TensorFlow. The primary loss function was sparse categorical cross-entropy:

$$\mathcal{L}_{\mathcal{CE}}(y, \hat{y}) = -\sum_{c=1}^{C} y_c \log(\hat{y}_c)$$

Class weights wcwc further addressed residual imbalance:

$$\mathcal{L}_{weighted}(y, \hat{y}) = -\sum_{c=1}^{C} w_c y_c \log(\hat{y}_c)$$

The Adam optimizer started with η=10−4η=10−4. Adam's updates are defined as:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\widehat{m_t} = \frac{m_t}{1 - \beta_1^t}$$

$$\widehat{v_t} = \frac{v_t}{1 - \beta_2^t}$$

$$W_t = W_{t-1} - \eta \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \epsilon}$$

We employed early stopping to prevent overfitting, halting training if validation loss did not improve for three epochs. ReduceLROnPlateau reduced the learning rate by a factor of 0.2 after two consecutive stagnations in validation loss. We trained for up to 20 epochs with a batch size of 32. Table 4 lists the key hyperparameters, such as learning rate, batch size, L2 regularization, and patience thresholds for callbacks.

Hyperparameter	**Value**	**Notes**
Optimizer	Adam	Default beta1=0.9, beta2=0.999
Initial Learning Rate	1e-4	Reduced upon plateau
Batch Size	32	Balanced memory usage and speed
Number of Epochs	Up to 20	Early stopping applied (patience=3)
Loss Function	Sparse Categorical Cross-Entropy	Suitable for integer labels
Dropout Rate	0.5	Applied to fully connected layer
L2 Regularization	λ = 0.001	Applied to dense layer
Learning Rate Scheduler	ReduceLROnPlateau	Factor=0.2, patience=2
Number of Unfrozen Layers	201 layers	Layers after the 500th in DenseNet-201

Table 4: Hyperparameters and Settings Used During Model Training
(This table summarizes chosen hyperparameters, including optimizer settings, learning rate schedules, batch size, and regularization parameters, providing a clear snapshot of the training configuration.)

We evaluated the final model on the test set, which the model never encountered during training or validation. We computed accuracy:

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} I(\hat{y}_i = y_i)}{N}$$

and per-class precision, recall, and F1-scores:

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c}$$

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c}$$

$$F1_c = 2 \cdot \frac{\text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}$$

We also generated a confusion matrix to visualize class confusions. These performance metrics provided both an overall measure of model quality and a detailed per-class assessment, guiding future refinements in data collection, augmentation, or model architecture.

The entire pipeline data loading, annotation, preprocessing, balancing, splitting, model construction, training, and evaluation was orchestrated through well-documented Python scripts. Python's os and hashlib helped manage files and detect duplicates, while Pillow (PIL) handled image resizing and format conversions. The tf.data API streamlined the input pipeline, and functions like parse function(filename, label) ensured consistent decoding, normalization, and resizing of input imagery. On-the-fly augmentation was applied to training batches, while validation and test sets remained unchanged for fair performance evaluation. Logging and seed-setting ensured reproducibility and traceability, enabling other researchers to replicate or extend this study.

In summary, this methodology seamlessly integrated rigorous data preparation, balanced preprocessing, a sophisticated DenseNet-201 architecture, and carefully tuned training procedures. The figures and tables including distributions of categories (Figure 2), geographic coverage (Figure 3), annotation samples (Figure 4), model architecture (Figure 6), and detailed tabular summaries (Tables 2–4) together with the provided formulas, ensured a transparent, comprehensive, and reproducible framework for satellite imagery-based building classification.

4. Results

After completing the training and validation processes described above, the final model demonstrated strong classification performance on the held-out test set. Throughout training, we monitored the model's accuracy and loss on both the training and validation subsets, ensuring that improvements were consistent and that the model was not overfitting.

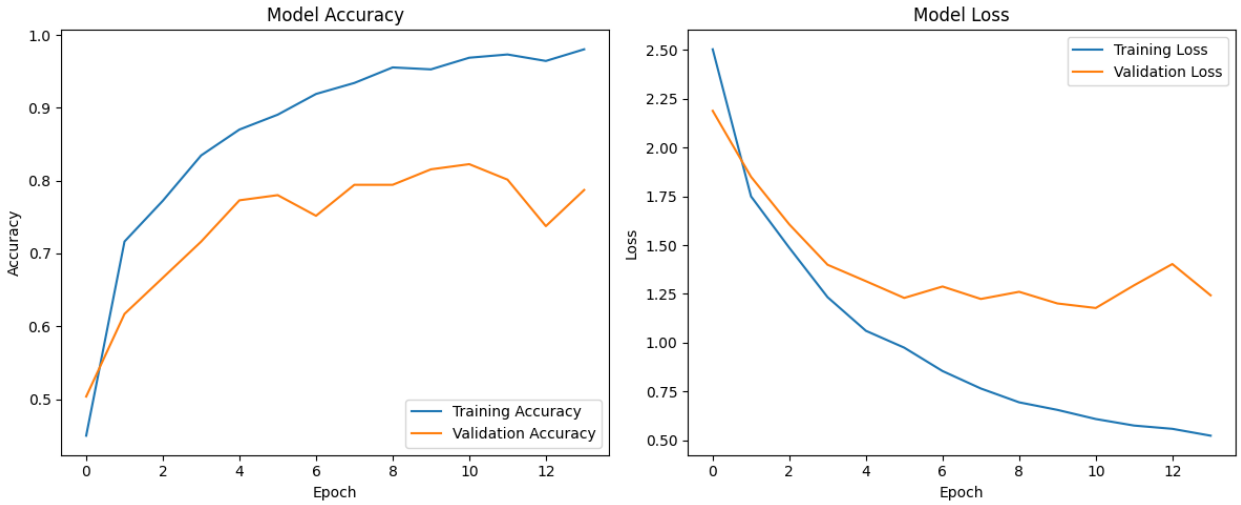

Figure 7: Training and Validation Accuracy and Loss Curves
(This figure displays the model's accuracy (left plot) and loss (right plot) over successive epochs. The training accuracy steadily increases and converges above 95%, while the validation accuracy peaks near 82%. The corresponding loss curves show a decreasing trend, suggesting effective learning and gradual refinement of model parameters.)

During training, the initial epochs established a solid feature representation, reflected by a rapid rise in accuracy and a steep decline in loss. As training progressed, validation accuracy varied slightly but remained relatively stable at a high level, indicating that the model successfully generalized to unseen validation data without severe overfitting. By applying early stopping and learning rate reduction strategies, we halted training at an optimal point, maintaining model weights that achieved a favorable balance between bias and variance.

Upon evaluation on the independent test set of 141 images, the final model achieved an accuracy of 84.40%. The class distribution in this test set was balanced across the seven building classes, ranging from 20 to 21 images per class, mirroring the balanced approach used during training and validation. This balanced evaluation ensured that the model's performance metrics reliably reflected its capability to classify multiple building types.

Confusion Matrix and Classification Report
To gain deeper insights into class-specific performance, we generated a confusion matrix and a

comprehensive classification report. The confusion matrix (Figure 8) visualizes the distribution of correct and incorrect predictions across all classes, highlighting that the model consistently identified classes such as “High” and “Single” with high precision and recall. While some misclassifications occurred particularly among classes like “Commercial” and “Multi” these were relatively infrequent.

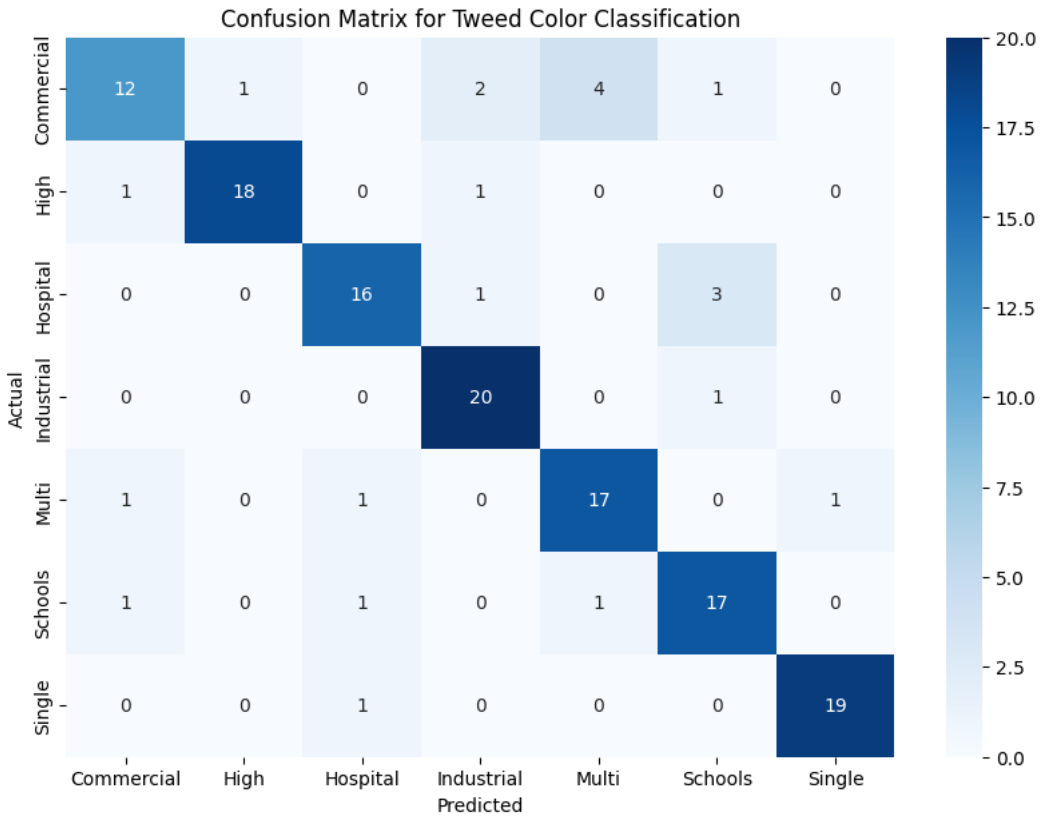

Figure 8: Confusion Matrix for Test Set Predictions
(Entries represent the count of predictions made by the model for each actual class. The diagonal elements indicate correct classifications, while off-diagonal entries show confusions between classes.)

The classification report (Table 3) further quantifies the model’s performance by listing precision, recall, and F1-score for each class. These metrics confirmed robust performance across most classes, with several categories (e.g., “High”, “Single”, and “Hospital”) demonstrating particularly high F1-scores. Classes with slightly lower precision or recall generally suffered from a small number of misclassifications, which might be addressed in future work through targeted data augmentation or refinement of the model’s decision boundaries.

Classification	Precision	Recall	F1-Score	Support
Commercial	0.80	0.60	0.69	20
High	0.95	0.90	0.92	20
Hospital	0.84	0.80	0.82	20

Classification	Precision	Recall	F1-Score	Support
Industrial	0.83	0.95	0.89	21
Multi	0.77	0.85	0.81	20
Schools	0.77	0.85	0.81	20
Single	0.95	0.95	0.95	20
accuracy	-	-	0.84	141
macro avg	0.85	0.84	0.84	141
weighted avg	0.85	0.84	0.84	141

Final Test Accuracy: 84.3972%

Table 3: Classification Report for Each Building Class
(Precision, recall, and F1-score are reported for each class, along with the macro and weighted averages. High F1-scores across multiple classes indicate that the model handles diverse building categories effectively.)

In summary, the final model achieved a strong test accuracy (84.40%) and exhibited balanced, high-quality predictions across the majority of classes. The training curves and early stopping criteria ensured that the model did not overfit, while the confusion matrix and classification report confirm that the model learned to distinguish building types with a high degree of specificity and generality. This performance suggests the potential for deploying such a model in practical applications related to urban planning, infrastructure assessment, and geographic analysis.

5. Discussion

The presented study demonstrates that a deep learning-based approach, specifically utilizing a DenseNet-201 model, can effectively classify multiple building types from high-resolution satellite imagery. By achieving a test accuracy of approximately 84%, our results highlight the model's potential for assisting in a variety of geospatial applications, such as urban planning, infrastructure assessment, environmental monitoring, and disaster response. The consistent differentiation among categories ranging from high-rise and hospital structures to single-unit residential and industrial buildings underscores the ability of advanced CNN architectures to extract meaningful features even from overhead perspectives.

Our findings align with the broader trend of using deep learning methods to automate the interpretation of remote sensing data. As more satellite imagery becomes publicly available and increasingly detailed, building classification models can serve as key analytical tools, helping researchers and practitioners identify spatial patterns, track urban growth, or understand the distribution and characteristics of built environments across regions. Furthermore, the relative ease with which the DenseNet-201 model adapted to this task, after data balancing and augmentation, emphasizes the importance of curating well-prepared datasets. Balanced training sets and carefully chosen augmentation strategies ensured that the model did not become overly biased toward dominant classes or unrepresentative samples.

An interesting aspect of our results is the variability in class-wise performance. While "High" and "Single" categories exhibited particularly robust metrics (including high F1-scores),

some classes, such as “Commercial” and “Multi”, faced more frequent misclassifications. This suggests that certain building types may be more visually distinctive, allowing the model to latch onto characteristic architectural cues (e.g., building height, shape, roof patterns), whereas others may share overlapping features or appear in heterogeneous contexts that impede straightforward classification. The subtleties in building design, environmental context (e.g., vegetation, shadows, adjacent structures), and roof materials can all influence a model’s predictive confidence. Future work might include incorporating additional data modalities such as multispectral bands, LiDAR elevation data, or nighttime light intensity to provide richer information and improve discriminative power where pure RGB imagery falls short.

While our chosen model architecture and training protocols proved effective, certain limitations warrant discussion. First, this study relied on static images collected at a given time, without considering temporal data that might help differentiate building types based on seasonal or long-term changes in land use. Additionally, external contextual data, such as zoning information or local infrastructure databases, could complement purely visual signals, providing more holistic cues for classification.

Another consideration involves the underlying labeling process. Although annotations were meticulously reviewed and validated, building classifications can sometimes be ambiguous or subjective. Enhancing the annotation pipeline, perhaps through integrating crowdsourced input or refining the class definitions to reduce overlap, could yield even more reliable ground truths. Expanding class categories or subdividing existing classes into finer-grained categories (e.g., differentiating various commercial subtypes) might also help the model learn more nuanced distinctions, although care must be taken to maintain sufficient training examples for each category.

The reliance on a single deep learning architecture offers another avenue for future improvements. While DenseNet-201 provided strong results, experimenting with other advanced architectures such as Vision Transformers or hybrid CNN-LiDAR approaches could uncover incremental gains in accuracy or speed. In addition, techniques like transfer learning from domain-specific satellite datasets or self-supervised pretraining using massive amounts of unlabeled imagery may further bolster performance and reduce the dependency on large, labeled datasets.
Data availability and domain generalization also remain key challenges. Our approach assumed that images from one geographic region are representative of other areas, but local differences in building materials, styles, and densities may limit model transferability. Testing the model’s robustness across multiple countries, climates, or cultural contexts would be a natural next step. Similarly, implementing domain adaptation methods could help the model generalize better when confronted with imagery from regions not represented in the training data.

In terms of broader impacts, our findings suggest that automated building classification models can play a role in numerous societal and environmental applications. For instance, urban planners could use these tools to monitor growth patterns, identify illegal constructions, or inform transportation and utility network expansions. Emergency responders might leverage the model’s output to quickly assess building distributions in disaster-hit areas and prioritize rescue or recovery operations. Environmental researchers could track changes in the built environment as proxies for economic development or land-use changes over time.

Several directions remain open for future research. Incorporating temporal data, as mentioned, could clarify whether certain building types exhibit distinctive seasonal patterns (e.g., the appearance of certain materials under varying light or vegetation conditions). Integrating other data sources such as topographic maps, cadastral data, or building footprint polygons could enrich model inputs. Expanding model interpretability methods would also be beneficial, helping stakeholders understand which features the model deems important and why certain

misclassifications occur. Additionally, exploring novel loss functions or active learning strategies to select the most informative samples for annotation might further improve model performance and reduce the labor required for dataset curation.

In conclusion, this study provides strong evidence that deep learning models like DenseNet-201 can reliably classify building types from overhead imagery. While limitations and avenues for enhancement remain, the promising performance and adaptability demonstrated here pave the way for future research and practical implementations. Integrating more data sources, refining class definitions, and exploring advanced modeling techniques can propel building classification to more accurate, generalizable, and context-aware frameworks that support diverse geospatial decision-making tasks.

6. Conclusion and Future Work

This research confirms that a carefully curated collection of high-resolution satellite images can effectively train a convolutional neural network to categorize buildings into multiple distinct classes. By capitalizing on DenseNet-201's densely connected structure, the model demonstrated considerable accuracy in extracting and distinguishing visual cues such as roof geometry, building height, and surrounding contexts even when structures shared some architectural features. The promising performance of this approach emphasizes the utility of automated building classification for a variety of large-scale geospatial tasks, including urban development monitoring, resource allocation, and crisis response planning.

Despite these positive outcomes, overlapping characteristics between some building types indicate potential areas for advancement. Future efforts could integrate supplementary data sources (such as LiDAR, multispectral images, or building footprint databases) and adopt more sophisticated labeling strategies to mitigate ambiguity in boundary cases. In addition, investigating the temporal dimension of satellite imagery, refining data augmentation methods, and exploring interpretability techniques would further bolster the model's robustness and transparency.

Altogether, this study provides a solid base from which researchers can pursue deeper insights into how and why certain buildings are misclassified, as well as explore the broader applicability of machine learning in remote sensing. By building on the insights presented here expanding data diversity, refining annotations, and adopting emerging deep learning innovations the field can move toward more accurate, context-aware, and operationally valuable tools for assessing built environments worldwide.

References

1. [1]Vasavi, K. P., Babu, G., & Manjula, S. R. (2023). Building Classification Using Deep Learning Techniques for Remote Sensing Images. *International Journal of Engineering Research & Technology*, 12(01), 549–554.
2. [2]Li, Y., Chen, L., Liu, S., & Guo, Z. (2021). A Modified U-Net Model for Rapid Detection of Rural Buildings from High-Resolution Remote Sensing Images. *Remote Sensing*, 13(9), 1817. https://doi.org/10.3390/rs13091817
3. Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7), 2217–2226. https://doi.org/10.1109/JSTARS.2019.2918242
4. Atwal, M., Lee, S., Rushing, G., Shrestha, P., & Clubb, K. (2022). Classification of Building Types Using Data-Fusion of Volunteered Geographic Information and

Authoritative Data. *ISPRS International Journal of Geo-Information*, 11(4), 223. https://doi.org/10.3390/ijgi11040223

5. Reda, S., & Kedzierski, M. (2020). FER-CNN: A Faster Edge Region CNN for Building Detection in Satellite Imagery. *Remote Sensing*, 12(13), 2103. https://doi.org/10.3390/rs12132103
6. Lloyd, C. T., Sorichetta, A., & Tatem, A. J. (2020). High Resolution Global Gridded Data for Use in Population Studies. *Scientific Data*, 7(1), 1–11. https://doi.org/10.1038/s41597-020-00762-5
7. Dimassi, S., Mhale, L., & Maatouk, M. (2021). Beirut Buildings Type Classification Dataset. *Data in Brief*, 35, 106854. https://doi.org/10.1016/j.dib.2021.106854
8. Ji, S., Wei, S., & Lu, M. (2019). Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Dataset. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 574–586. https://doi.org/10.1109/TGRS.2018.2858817
9. Huang, X., Li, Z., Zhang, S., & Zhu, X. X. (2022). A Large-Scale Dataset for Object Detection in Very High-Resolution Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–12. https://doi.org/10.1109/TGRS.2022.3141486
10. Kusz, K., Wan, B., Duarte, F., & Ratti, C. (2021). Building Classification for the United States Using Lidar Data. *Remote Sensing*, 13(8), 1544. https://doi.org/10.3390/rs13081544
11. Hang, Y., & Cai, S. (2020). Roof Classification in Satellite Imagery Using Deep Learning. *Proceedings of the 2020 International Conference on Artificial Intelligence and Computer Engineering*, 140–144. https://doi.org/10.1145/3427364.3427406