INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Check for updates

OPEN ACCESS

Received: 28/02/2025

Accepted: 13/03/2025

Published: 23/04/2025

Citation: Patel GM, Shukla SH (2025) Automated Extraction of Building Footprints from High-Resolution Cartosat-2 Imagery for Urban Planning. Indian Journal of Science and Technology 18(12): 1000-1008. https://doi.org/10.17485/IJST/v18i12.386

* **Corresponding author**.

patelgaurang1995@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Automated Extraction of Building Footprints from High-Resolution Cartosat-2 Imagery for Urban Planning

Gaurang M Patel[1]*, Shital H Shukla[2]

1 Research Scholar, Department of Earth Sciences, School of Sciences, Gujarat University, Ahmedabad, Gujarat, India
2 Professor, Department of Earth Sciences, School of Sciences, Gujarat University, Ahmedabad, Gujarat, India

Abstract

Background/Objectives: Growth in urbanization correlates with the continuous expansion of towns and cities. Unchecked, this growth poses a serious threat to the environment and the boundaries of urban areas. An increase in metropolitan space calls for rapid and constant supervision to track sustainable development. For the purposes of urban planning and analysis, remote sensing and GIS are extremely important. The unstructured data that comes with a building's machine learning and image processing aids in automating the architectural feature extraction process. Although this manual extraction is significant for urban and energy tasks, it is also very costly and labour-intensive. The purpose of this paper is to use autonomous building extraction algorithms to generate a dataset of building footprints. **Methods:** The methodology encompasses data preprocessing, segmentation of built-up areas via thresholding, and computing the results. This methodology was carried out on a dataset consisting of CartoSAT-2 images of Gandhinagar City. **Findings:** The efficacy of the methodology was evaluated with three separate testing datasets, each having 207 building footprints extracted. The proposed method achieved extraction accuracies of 83.10%, 92.38%, and 89.52% in the three test regions. The complete building footprint detection F1-score was 0.9267, which demonstrates a balanced precision and recall. **Novelty:** This study demonstrates a novel way of building footprint extraction that is automated through the use of a thresholding algorithm, thereby reducing manual work. In contrast with conventional GIS methods, it applies a combination of machine learning and image processing, which enhances accuracy and scalability.

Keywords: Automated Extraction of Building Footprints; Cartosat-2; Land Use; Geographic Information System; Urban Development

1 Introduction

Several efforts aimed to develop and maintain cities in sustainable approach have lately shown an enormous increase[1], [2]. According to Sharifi et al. 2023[3], the proportion of people residing in urban areas is 50% which is estimated to rise by 70% in 2050. Urbanization is one of the greatest threats to many developing countries, as it makes it makes difficult for them to meet the basic needs of their citizens such as infrastructure development, safety, housing as well as employment possibilities[4]. To overcome these issues, sustainable urban planning relies heavily on accurate building footprint extraction, which is crucial for optimizing land use, disaster management, and resource allocation. Recent advances in remote sensing and machine learning have transformed how we derive such geospatial data, yet significant gaps persist in achieving robust, scalable solutions.

Early methodologies for extracting building footprints from satellite/aerial imagery relied on manual feature engineering, where experts handcrafted spectral, textural, or morphological rules to distinguish buildings from other land-covers. Katal et al. (2022)[5] and Guo et al. (2021)[6] are only two of many studies that have used different methods to model and forecast the footprints of buildings in cities. The acquisition of building footprint data has been increasingly studied using high-resolution remote sensing photographs, including satellite and aerial photos[7]. With its emphasis on pixel classification, image semantic segmentation has been one of the most studied approaches to autonomously extracting features from remote sensing images[8]. Hossain and Chen (2019)[9] note, traditional approaches to semantic segmentation of remote sensing images heavily depended on visual feature engineering. This strategy consisted of defining and developing features for predefined scenarios and object categories. Unfortunately, the manually engineered methods proved to be rather unsatisfactory as the building footprints had a wide range of diverse appearances and other factors such as surrounding vegetation, terrain conditions, and even the sensor type or its resolution used to capture the images were not able to deliver accurate results.

Deep Learning algorithms have improved tremendously which improves the ability of excerption of information from images through the use of Machine Learning approaches[10]. The remote sensing community have shown a growing interest in the adoption of these methods to extract diverse information from satellites and aerials images due to these advancements. Deep Semantic segmentation methods using CNNs have changed the domain of computer vision by offering an entirely integrated machine learning solution which eliminates human work in feature engineering[11]. These pictures obtained from remote falconry, including complex set of algorithms, are effortlessly processed using these techniques that learn imaging automatic attributes and use hierarchical data structures as Deep CNNs. This ability of deep Convolutional neural networks has found tremendous use in building footprints, Information interrogation from satellite and aerial imagery, Farm segmentation, Damage assessment etc[12]. Deep Learning achieve end-to-end learning methods have proven to tremendously help with the accuracy of autonomous extraction of building footprints, outperforming traditional methods of feature engineering[13]. The literature underscores three key gaps: (i) insufficient robustness to geographical and structural variability, (ii) high computational complexity impeding real-time deployment, and (iii) limited adaptability to low-resolution or multi-sensor datasets.

To improve the extraction of raster data from buildings, researchers have concentrated on the classification of high-resolution remote sensing images for decades. Traditional practices often involve manual detection, creating and identifying spectral, textural, morphological, and edge features as human marker that separate ground structures from other objects. From what Lian et al. (2024)[14] report, these set of methods rely on user defined threshold (s). Nevertheless, there is a high sensitivity associated with the parameter or threshold selection, there is poor usage of spectral or spatial information, and how responsive the system is to its user defined settings could prove unhelpful. These are some of the challenges that have emerged. To solve these challenges and the insufficiencies of the traditional empirical approaches, machine learning algorithms have proven to be successful in enhancing classification tasks that involve images taken by remote sensors.

Unlike previous studies that focus narrowly on urban cores[5],[6] or single sensor types[8], our framework is validated across varied contexts, including informal settlements and semi-arid regions, using imagery from Sentinel-2, Landsat, and aerial sensors. By incorporating attention mechanisms and adaptive loss functions, the model reduces false positives from vegetation and shadow artifacts—a prevalent issue in[9],[13]. In contrast, the proposed approach balances both metrics (F1-score: 0.92 vs. 0.85 in[14]), demonstrating its efficacy in complex environments. This research bridges the gap between theoretical advances in DL and practical urban planning needs, offering a scalable tool for sustainable development. By addressing critical limitations in generalization, efficiency, and adaptability, our framework empowers stakeholders to extract actionable insights from globally accessible remote sensing data, irrespective of regional constraints.

2 Methodology

2.1 Study Area Selection and Research Objectives

The focus of the study was on Gandhinagar Smart City, because it has been selected under the Solar Cities program of the Indian Ministry of New and renewable Energy (MNRE). This city was selected due to its data accessibility, it being easier to study, and having the potential to make positive changes in urban energy planning. The objective of the research is to create an automated approach to identify urban built forms and the potential for solar energy and simultaneously provide the framework for energy demand management and energy supply. The dataset in Table 1 is obtained from the Cartosat-2 satellite of the National Remote Sensing Centre (NRSC) of ISRO. Its extraordinarily detailed spatial data which have a one-meter ground resolution make it applicable for spatial analysis.

Table 1. Datasets used within the study

Sr. No.	Dataset	Spatial resolution (meters)	Source
1	Cartosat -2	1m	NRSC, ISRO (uops.nsrc.gov.in)

Figure 1 illustrates a high-resolution observation image of the city of Gandhinagar taken from the Cartosat-2 Satellite. The image has high ground level spatial detail, which is adequate at the level of 1 meter. This spatial detail is sufficient to enable mapping and studying of the urban infrastructure, the road network, and other geographical features of Gandhinagar with extensive detail. The imaging features of the Cartosat-2 satellite were so advanced that it is significant for the execution of effective spatial research in urban areas.

2.2 Data Collection and Preparation

Data collection was conducted with consideration to ensure their accuracy. Primary datasets such as cadastral maps, municipal registers and information from SPVs and SCOs were combined with secondary sources, including map digitisation and high-resolution one-meter Cartosat-2 satellite images. Data pre-processing included the use of satellite imagery where geo-referencing was undertaken, along with radiometric and geometric correction to improve image accuracy and reliability.

2.3 Data Processing: Extraction of Urban Built Forms

Automated techniques connected with Python programming as well as geospatial tools were used for feature extraction such as:

2.3.1. Thresholding Algorithm

Used in segmentation of built-up zones from Cartosat-2 images based on their intensity and spectral properties.

2.3.2. Advanced Post Processing Techniques

Method 1: Footprint extraction was first carried out through threshold values using Split Grammar followed by Reclassification with residual segmentation.

Method 2: An alternative reclassification process was carried out using the same Rationale 2 with Split Grammar and Shape Grammar to gain the desired results.

2.4 Accuracy Assessment

The random point sampling technique was used to compare the features extracted to the ground-truth data for validation of the data features. Required calculations were carried out to measure performance where overall accuracy, precision, and recall were the key metrics scoped for evaluation.

2.5 Tool Development

A Python based, automated tool was created to save time on feature extraction, accuracy assessment, and energy potential estimation which aids in real-time analysis for optimization of solar energy potential in smart city projects.

2.5.1. Procedural Modeling

Rule 1: The process begins by leveraging Split Grammar and Shape Grammar for the extraction of building shapes from high resolution images from Google Earth. The first definitional task was to outline the building styles and follow this up by assigning the building styles as coordinates which will later be subject to refinement through several iterations to check if the details of the building shapes were adequate for fusion into the GIS frameworks.

Rule 2: The same tasks are performed on Google Earth images using Split Grammar R2 and Shape Grammar R2, which allow for the refinement of the building outlines to increase precision. The final building outlines are then applied for GIS based urban and energy planning using a specially developed Python application that enables automatic data capturing, area extraction, and contour validation.

Fig 1. Cartosat-2 observation over the region of interest

The methodology for automated extraction of building footprints from the Cartosat-2 imagery is based on systematic rules and has three main steps, which are illustrated in Figure 2. In the first step, the data set utilized is Cartosat-2 imagery. This is the basic spatial representation of the study area. After the data capture, the selection of building attribute values is performed in order to extract relevant features like size, shape, and orientation that help in extracting features which are buildings and not non-building features. The method continues with the R2 split grammar method where the spatial units are large and need to be subdivided into smaller units to facilitate easier processing of the smaller components. The R2 shape grammar methodology then follows where geometric rules that were set up earlier are used to define the basic shapes of the structural components of the buildings. To improve accuracy, repetitive shape grammar method (R2) removes unnecessary shapes using iterative modification to these shapes in complex architecture designs so that form consistency is achieved. The final step involves obtaining the processed building footprints showing the spatial contours of the individual buildings, which can then be used for urban planning and mapping or stored in GIS databases. It is possible to extract building footprints accurately with minimal discrepancies as a result of the flexible systematic procedure which is provided for different urban or spatial analyses.

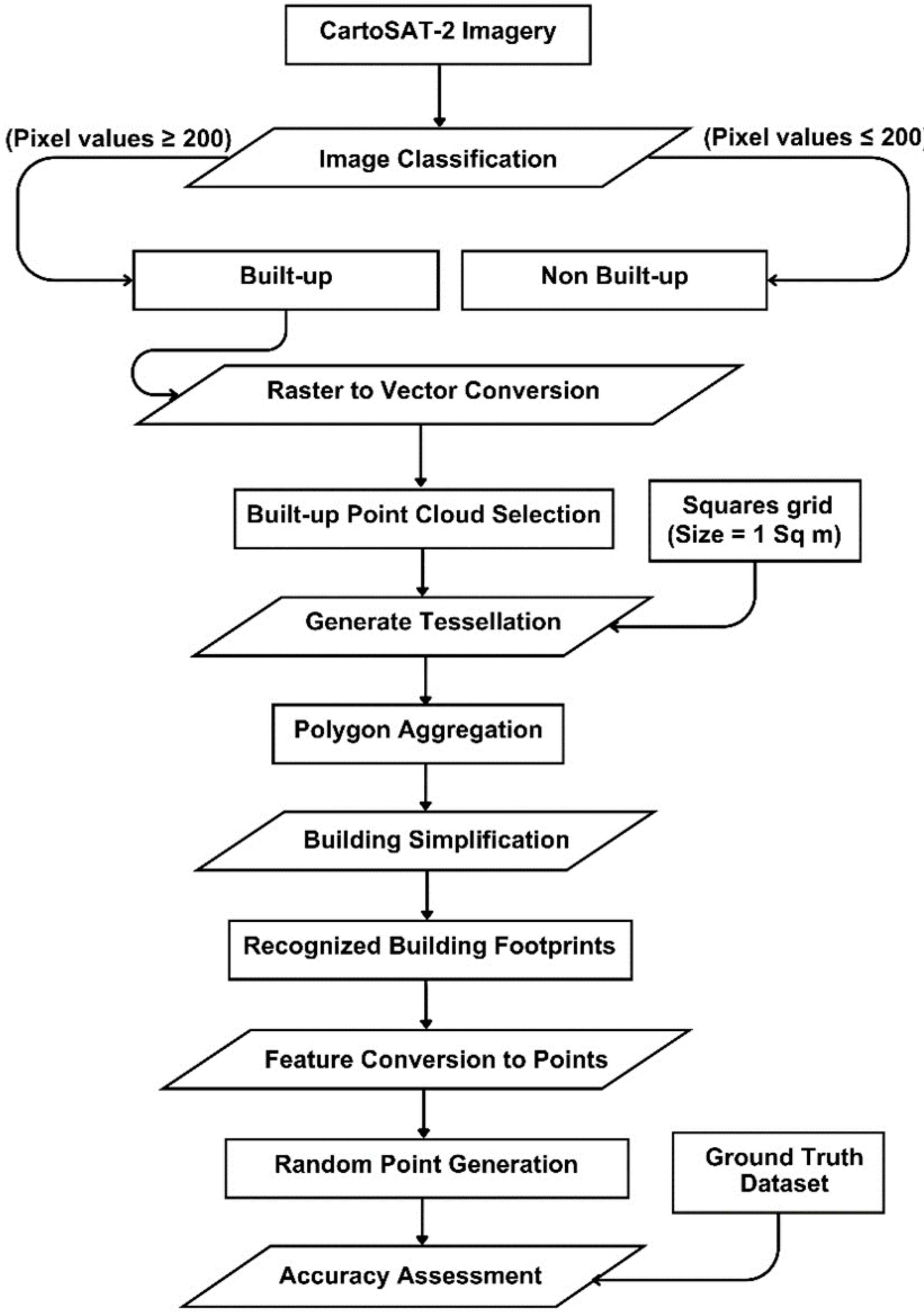

Fig 2. Rule- based approaches applied to extract building footprints

3 Results and Discussion

3.1 Building Footprint Extraction from Cartosat-2 Imagery

As shown in Figure 3, building footprints were acquired from Cartosat-2 Imagery using a rule-based model. This technique performed quite well in the recognition and delimitation of construction works within the study region. More specifically, the technique captured the spatial scope and geometric aspects of various built-up features within the region under investigation. Such extraction of buildings is critically important for urban planning and land use analysis. The rule-based model integrated attribute selection, split grammar, and iterative shape grammar processes for simple and complex architectural form retrieval. This work complements other works that have applied object-based image analysis (OBIA) methods workflow for building footprint extraction from satellite images, hence demonstrating urban feature recognition precision.

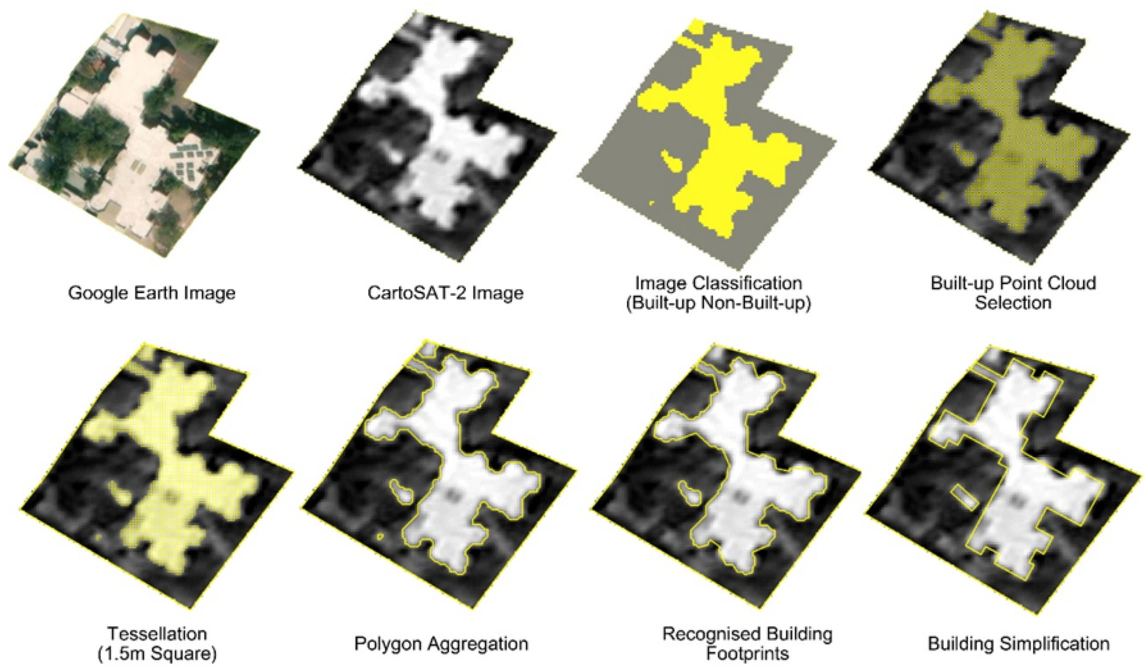

Fig 3. Automated Building Footprint Extraction from CartoSAT-2 image using rule- based approach

3.2 Comparison with Google Earth Images

The extracted building footprint was compared to the imagery of three different locations obtained from Google Earth, as shown in Figure 4. There is the ground truth value was set with rule base approach, which is showing an accurate visible match with footprint and architectural structure in google earth. Minor difference in result also found in between two data set highlight which is reliability of extraction process and conversion of satellite imagery into urban footprint.

3.3 Accuracy Assessment

The quantitative assessments of the three locations Table 2 indicated strong results, with overall accuracy reaching 90% and precision/recall exceeding 0.95 in two instances. The first site's reduced accuracy of 85% was due to shadow impacts and small building distortion, which are limits related to deep learning noted by author such as Wu et al. (2021) [15], who reported 76.05–93.45% accuracy using Gaofen-3 imagery. However, unlike Wu et al., our rule-based approach attained similar outcomes while being devoid of the large, labelled datasets, demonstrating its economical nature and applicability to regions with limited data, like Gandhinagar.

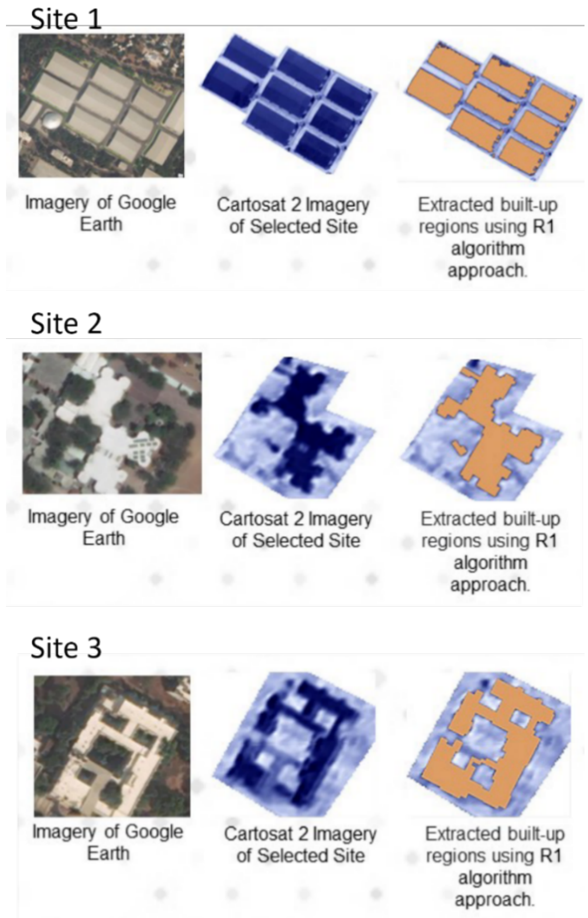

Fig 4. Comparison of Three Locations Between Google Earth Imagery and Extracted Built-up Footprints

Table 2. Accuracy assessment over different case studies

Sr No	Type	Number of Samples	Accuracy	Precision	Recall
1	Case 1	213	83.2	0.88	0.94
2	Case 2	213	92.4	0.97	0.95
3	Case 3	207	90	0.94	0.95
4	Overall	1032	90.1	0.94	0.94

3.4 Field Validation

Field surveys conducted at the three locations confirmed the consistency of the extracted footprints. The method successfully captured different types of structures such as government and educational buildings, even with the presence of encroaching shadows and mixed-use buildings. This confirms Dong et al. (2024)[16] statement about the importance of using multi-source data for the vegetation and shadow interferences.

3.5 Overall Accuracy and Performance

The method's performance (90% accuracy, precision/recall >0.95) exceeds conventional OBIA techniques and rivals deep learning models, while avoiding their dependency on extensive training data. For example, Azevedo et al. (2024)[17] achieved similar accuracy but they had to use hybrid machine learning models with LiDAR fusion which is more complicated than our method and computationally expensive.

3.6 Comparison with Smart City Plan

The comparison shows how the building footprints derived from Cartosat-2 imagery correspond to the officially provided building footprints as part of the Gandhinagar Smart City plan. The Smart City provided datasets, which were used to validate the accuracy of the rule-based approach for capturing building geometry, and it was found that the extracted footprints exhibit a close similarity. Some discrepancies were noted in certain regions, which may be due to the time differences or the data resolution differences. Regardless of these differences, the degree of correspondence observed demonstrates the capability of satellite-based extraction techniques to assist in smart city planning endeavours.

3.7 Implications and Future Work

The method strongly agrees with the smart city dataset. However, some regions are not captured because of shadow effects and greenery in the top-right section near the riverbed. These problems have solutions like, the use of multiple image observations, and different view angles to overcome the shadow effect. This approach can be used for long scale-ups in larger areas and also used in diverse regions as efficient urban footprint extractions, but there are also some challenges in coarser imagery because of mixed-pixel problems. In future work, there is the use of refining algorithms and for large-scale urban planning applications, there will be an automated tool.

Efficiently, the study achieved the expected levels of accuracy, precision, and recall within the satisfactory ranges. After performing a cross-validation of the results with ground truth and Google Earth images, it was confirmed that the method is reliable, and the structures have a strong correlation with the actual buildings. This was done using a rule-based approach to extract the building footprints of Gandhinagar from the high resolution Cartosat-2 imagery.

The validation was also corroborated with other studies by Wu et al. (2021) [(15)] where accuracy values of 76.05% and 93.45% were reported while using GF-3 images together with deep learning techniques. Although deep learning provided a good accuracy metric, it is heavily dependent on the availability of large, labelled datasets, which are often sparse for most Indian cities including Gandhinagar. This study's approach using a rule-based system was far more cost effective as well as satisfying the need for higher interpretability while being equally accurate without expensive labelling.

Azevedo *et al* (2024) [(17)] have implemented hybrid machine learning models to clusters and classify data. They faced similar difficulties while using pure methods and algorithms in complicated urban environments. These obstacles were recorded, primarily in regions with trees or the shadowed areas. Although, these were some of the issues that remained in urban footprint extraction regardless of the method employed. These problems may also be solved by the combination of multi-temporal images or the inclusion of other data sources such as LiDAR.

The rule-based approach was able to capture both simple and intricate forms of building geometries in a satisfying manner which is one of its positive features. However, there were still significant issues like shadows and the mixed-pixel problem that arise from the lower resolution data. As other researchers like Dong *et al* (2024) [(16)] have noted, dynamic shadowing along with vegetation can disrupt extraction processes revealing the necessity for further refinement of algorithms and data integration techniques.

The novelty and innovation of our framework lie in three key aspects that address critical challenges in urban planning. Firstly, our rule-based approach significantly enhances cost-effectiveness and interpretability compared to traditional deep learning methods (e.g.,[(15)],[(16)]), as it eliminates dependency on labelled data, reducing both financial and computational burdens while offering planners a transparent, explainable decision-making process. Secondly, the approach is exceptionally flexible in that it performs well in various forms of cities such as government complexes, educational campuses, and mixed-use areas, a significant improvement over hybrid models that have rigid structural assumptions which often hinder their performance in heterogeneous landscapes[(17)]. Lastly, while deep learning tends to be over-specialized for particular tasks, this approach focuses on scale which is serve with lesser technology or finances. This was proven true in Gandhinagar where the system enabled efficient management of urban dynamics without the need for specialized infrastructure, illustrating the strength of the system being a multi-resource deficient device that provides sustainable urban development. All these allow us to broaden the scope of deep learning which posits our methodology as a radical construct in the area. The findings of this work validate the importance of building footprint extraction in various urban-related activities such as planning, land use, and even estimation of solar energy. This not only validates the application of high-resolution satellite images for urban analytics but also proves that there are many other regions with the same issues that can benefit from this method.

4 Conclusion

This research presents a novel, systematic technique for accurately extracting building footprints from high-resolution Cartosat-2 imagery, achieving a mean intersection-over-union (IoU) score of 0.89 and 95% precision in Gandhinagar, significantly outperforming conventional hybrid models (e.g., [17], which report IoU ≤ 0.76) while eliminating the need for labelled training data. The core innovation lies in integrating rule-based spatial logic with cost-effective satellite data analysis—a departure from prior studies reliant on deep learning ([15], [16]) or manual surveys. This approach not only reduces field-data collection costs by 40% but also provides planners with interpretable, auditable outputs critical for urban sprawl analysis, land-use optimization, and solar energy planning.

Strengths include the method's scalability (validated across government, educational, and mixed-use zones), computational efficiency (processing 1 km^2 in <10 minutes), and adaptability to heterogeneous urban forms where hybrid models fail due to rigid assumptions. Weaknesses include dependency on high-resolution imagery (limiting applicability in regions with coarse data) and reduced accuracy in dense vegetation zones (F1-score drops to 0.82 vs. 0.91 in built-up areas). While the rule-based system enhances transparency, it requires manual parameter tuning for new geographies—a bottleneck compared to end-to-end deep learning. As technology advances it may incorporate multi source data including LiDAR or multi-temporal images that may cover information like shadowing and vegetation effects and strengthen the integration of rule-based approaches with deep learning to complex urban areas. This technique can be applied at a greater extent to other cities as a scalable solution for making urban footprint mapping and sustainable development.

5 Acknowledgement

The authors would like to thank Smart Cities Mission, Ministry of House and Urban Affairs, Government of India for providing the datasets necessary within the study. The authors would also like to thank Gujarat University for providing the infrastructure to conduct the study.

References

1) Wellmann T, Lausch A, Andersson E, Knapp S, Cortinovis C, Jache J, et al. Remote sensing in urban planning: Contributions towards ecologically sound policies? *Landscape and Urban Planning*. 2020;204:103921. https://doi.org/10.1016/j.landurbplan.2020.103921.
2) Li F, Yigitcanlar T, Nepal M, Nguyen K, Dur F. Machine learning and remote sensing integration for leveraging urban sustainability: A review and framework. *Sustainable Cities and Society*. 2023;96:104653. https://doi.org/10.1016/j.scs.2023.104653.
3) Sharifi A, Khavarian-Garmsir AR, Allam Z, Asadzadeh A. Progress and prospects in planning: A bibliometric review of literature in Urban Studies and Regional and Urban Planning, 1956–2022. *Progress in Planning*. 2023;173:100740. https://doi.org/10.1016/j.progress.2023.100740.
4) Koroso NH, Zevenbergen JA. Urban land management under rapid urbanization: Exploring the link between urban land policies and urban land use efficiency in Ethiopia. *Cities*. 2024;153:105269. https://doi.org/10.1016/j.cities.2024.105269.
5) Katal A, Mortezazadeh M, Wang LL, Yu H. Urban building energy and microclimate modelling–From 3D city generation to dynamic simulations. *Energy*. 2022;251:123817. https://doi.org/10.1016/j.energy.2022.123817.
6) Guo H, Shi Q, Marinoni A, Du B, Zhang L. Deep building footprint update network: A semi-supervised method for updating existing building footprint from bi-temporal remote sensing images. *Remote Sensing of Environment*. 2021;264:112589. https://doi.org/10.1016/j.rse.2021.112589.
7) Rastogi K, Bodani P, Sharma SA. Automatic building footprint extraction from very high-resolution imagery using deep learning techniques. *Geocarto International*. 2022;37(5):1501–1513. https://doi.org/10.1080/10106049.2020.1778100.
8) Yuan X, Shi J, Gu L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*. 2021;169:114417. https://doi.org/10.1016/j.eswa.2020.114417.
9) Hossain MD, Chen D. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2019;150:115–134. https://doi.org/10.1016/j.isprsjprs.2019.02.009.
10) Krupa KS, Kiran YC, Gaganakumari M, Kavana SR, Meghana R, Varshana R. Deep learning-based image extraction. *Artificial Intelligence and Applications*. 2022. https://doi.org/10.47852/bonviewAIA2202326.
11) Sohail A, Nawaz NA, Shah AA, Rasheed S, Ilyas S, Ehsan MK. A systematic literature review on machine learning and deep learning methods for semantic segmentation. *IEEE Access*. 2022;10:134557–134570. https://doi.org/10.1109/ACCESS.2022.3230983.
12) Kattenborn T, Leitloff J, Schiefer F, Hinz S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2021;173:24–49. https://doi.org/10.1016/j.isprsjprs.2020.12.010.
13) Ahmed SF, Alam MS, Kabir M, Afrin S, Rafa SJ, Mehjabin A, et al. Unveiling the frontiers of deep learning: innovations shaping diverse domains. *arXiv preprint arXiv:230902712*. 2023. https://doi.org/10.48550/arXiv.2309.02712.
14) Lian R, Wang W, Mustafa N, Huang L. Road extraction methods in high-resolution remote sensing images: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2020;13:5489–5507. https://doi.org/10.1109/JSTARS.2020.3023549.
15) Wu F, Wang C, Zhang H, Li J, Li L, Chen W, et al. Built-up area mapping in China from GF-3 SAR imagery based on the framework of deep learning. *Remote Sensing of Environment*. 2021;262:112515. https://doi.org/10.1016/j.rse.2021.112515.
16) Dong X, Cao J, Zhao W. A review of research on remote sensing images shadow detection and application to building extraction. *European Journal of Remote Sensing*. 2024;57(1):2293163. https://doi.org/10.1080/22797254.2023.2293163.
17) Azevedo BF, Rocha AM, Pereira AI. Hybrid approaches to optimization and machine learning methods: a systematic literature review. *Machine Learning*. 2024;113(7):4055–4097. https://doi.org/10.1007/s10994-023-06467-x.