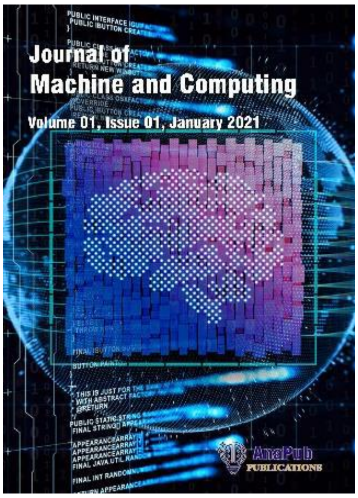
Journal Pre-proof



Advanced Spatial Categorization of Buildings Based on Point-Based Cloud Data Algorithms

Gi Hwan Oh

DOI: 10.53759/7669/jmc202505083
Reference: JMC202505083
Journal: Journal of Machine and Computing.

Received 18 July 2024
Revised form 22 October 2024
Accepted 20 December 2024

Please cite this article as: Gi Hwan Oh, “Advanced Spatial Categorization of Buildings Based on Point-Based Cloud Data Algorithms”, Journal of Machine and Computing. (2025). Doi: https://doi.org/10.53759/7669/jmc202505083

This PDF file contains an article that has undergone certain improvements after acceptance. These enhancements include the addition of a cover page, metadata, and formatting changes aimed at enhancing readability. However, it is important to note that this version is not considered the final authoritative version of the article.

Prior to its official publication, this version will undergo further stages of refinement, such as copyediting, typesetting, and comprehensive review. These processes are implemented to ensure the article's final form is of the highest quality. The purpose of sharing this version is to offer early visibility of the article's content to readers.

Please be aware that throughout the production process, it is possible that errors or discrepancies may be identified, which could impact the content. Additionally, all legal disclaimers applicable to the journal remain in effect.

AnaPub
PUBLICATIONS

Advanced spatial categorization of buildings based on point-based cloud data algorithms

Gi Hwan Oh

Department of Architectural Design, Dong-Seo University, Sasang-gu, Busan, Republic of Korea.
johnoh@dongseo.ac.kr
ORCID ID : 0000-0001-8848-6639

Abstract

There is a tremendous horizontal and vertical growth, where an immediate demand for geospatial tools for precision urban planning and sustainable development is gaining more interest. Acquisition of high resolution, 3D spatial data through Light Detection and Ranging (LiDAR) technology is an exploitable medium. Traditional grid-based LiDAR methods, however, tend to have data loss and lower accuracy. An automated point based classification methodology is introduced to further augment the classification of raw LiDAR data for urban areas in Tamil Nadu. Through spatial characteristics of point height, point density and local plane orientation, the proposed method efficiently classifies LiDAR points into ground, vegetation and building classes. By successfully reconstructing 3D urban models, the study was able to reflect large urban clusters in urban centres and sparse low-rise structures in rural areas. These models demonstrate the spatial relations between urban characteristics, they develop urban patterns and fluctuations in eco balances. Results show the capacity of this approach being potentially applicable to urban planning, smart city development, landslides and flooding management, and ecological conservation. This study aims to contribute to LiDAR's utility for urban analytics by overcoming current limitations of grid-based methods while enhancing classification in complex terrain. This research highlights the importance of LiDAR in making sustainable urban landscapes and beyond, significantly informed by data.

Keywords: Building Classification, LiDAR, Point Cloud. Building Reconstruction, 3D city Model

1. Introduction

Urbanization is happening all over the world and changing the physical, social, and economic landscapes of cities all over the world. In developing countries where rapid industrialization, population growth, and infrastructure development took place, like India, the state of Tamil Nadu is one of the areas of rapid urban growth [1]. Because of its varying geography encompassing coastal plains, hills, and urban centers dense with population, Tamil Nadu represents a distinctive class of challenges and opportunities for urban planning [2]. The ability to capture three-dimensional (3D) urban areas is essential to effective planning, resource management, and realizing sustainable development goals [3].

Light Detection and Ranging (LiDAR) technology has transitioned from a research phenomenon to an emerging technology for the collection of high-resolution 3D spatial data [4]. Together, laser pulses are used within LiDAR systems to create high-resolution point clouds of the Earth's surface, including terrain, buildings, and vegetation [5]. These are critical datasets for land use planning, disaster planning, and environmental management. However, classifying LiDAR data in terms of ground, vegetation, and buildings in terrains with complex topographies is still a challenge [6]. LiDAR data can be challenging to process and analyze due to the high resolutions normally achieved with traditional grid-based approaches, resulting in substantial data loss and consequently inaccurate results, undermining the capabilities of LiDAR data to contribute to detailed urban analysis [7].

These limitations can be overcome by automated, point-based classification methods. Different from grid-based methods, point-based methods consider individual LiDAR points on their basis of spatial features like height, density, and local environment for more accurate classification [8]. The urban landscape of Tamil Nadu, with high-rise clusters in cities such as Chennai and more traditional low-rise structures in rural areas, offers a rich test bed for such methods [9]. Validating LiDAR data is important for the efficient classification of the data to reconstruct a high-accuracy 3D urban model essential for spatial relationship modeling and future growth [10].

2. Literature Review

Urban areas in Tamil Nadu are undergoing a dual transformation: vertical growth within urban centers and horizontal expansion into peri-urban and rural regions [11]. Complex spatial patterns and dynamics are therefore required. For example, stakeholders develop high-rise developments around economic hubs and major roadways as indicators of economic activity and land scarcity. However, in some ecologically sensitive areas, horizontal sprawl is occurring, thus contradicting the development and conservation policies [12]. The 3D models derived from classified LiDAR data provide accurate 3D models that can reveal these patterns, allowing policymakers to manage the trade-off between development and sustainability [13].

With recent advances in LiDAR technology and geographic information systems (GIS), large-scale urban analysis is feasible [14]. LiDAR data on GIS platforms can be integrated with auxiliary datasets like satellite imagery and cadastre maps to perform sophisticated spatial analysis. Such integrations can help urban infrastructure planning, disaster preparedness, and monitoring environmental changes in Tamil Nadu [15]. For example, when hydrological data is tied with accurate digital elevation models (DEMs), it is possible to identify vulnerable areas prone to flooding or landslides [16].

Reconstruction of 3D urban models is also critical to smart city initiatives. There are several smart city projects within Tamil Nadu aimed at enhancing urban living with technology-driven solutions. High-quality geospatial data is required for these projects to optimize urban design, improve public services, and support sustainable development [17]. These goals can be achieved by using LiDAR-based 3D models to provide insights into urban density, building heights, and spatial arrangements, allowing for more efficient planning of transport, utilities, and green spaces [18].

Despite its potential, there have been difficulties in applying LiDAR technology due to the high cost of data acquisition, the requirement of advanced processing tools, and the complexity of accurately classifying urban features [19]. The substantial vegetation overlap, drastic changes in terrain, and extremely dense urban areas contribute to making data classification more difficult [20]. This article deals with these challenges by formulating a robust, automated point-based classification methodology specific to the urban locations chosen [21].

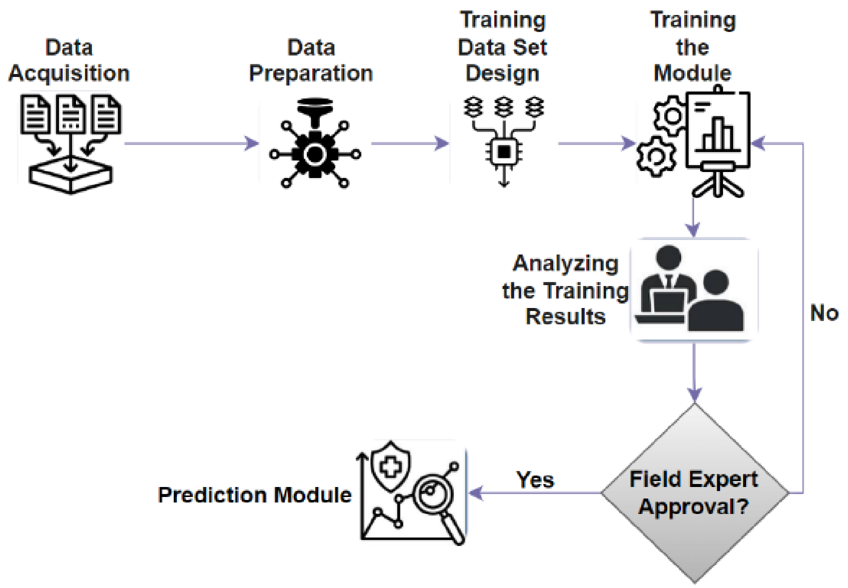
The primary objectives of this study are:

- A point-based classification method is to be developed and implemented to properly classify raw LiDAR data into classes that are ground, vegetation, and building.
- Reconstructing detailed 3D urban models to reflect existing spatial and vertical dynamics of Tamil Nadu's urban areas.
- Analysing the spatial relationships between classified features to provide actionable insights for urban planning and sustainability.

The remainder of this article is organized as follows: the study area and data acquisition process are described in Section 3. The methodology proposed is based on the use of feature selection, rule-based classification, and 3D reconstruction techniques. Results and discussion are presented in section 4 including accuracy of classification, 3D reconstruction performance, and spatial insights. Finally, Section 5 concludes the study with key findings, and provides directions for future research.

3. Materials and methods

The proposed automatic point-based classification methodology for LiDAR data and its corresponding application to 3D building reconstruction workflow are illustrated in Figures 1. Spatial features necessary for discriminating between ground, vegetation, and building classes are selected. Height, local environment, surface-based attributes, eigenvalues, local plane characteristics and point density are key features. Detection routines that accurately classify were defined by careful analysis of these features. The cornerstone of this methodology was a hierarchical rule set which was developed in the TerraScan module of TerraSolid software. The rule set is meant to process sequentially raw LiDAR point cloud



data, using the spatial features selected. Pilot area analysis carefully optimized the parameter selection for the classification routines to achieve high accuracy and to be able to adapt to varying terrain.

Figure1. Step by step stages of point based classification

All LiDAR points are initially assigned to a default class. This data is then subsequently iteratively refined into distinct classes, according to subsequent classification steps as depicted in figure 2. The default class ground points are extracted, then non-ground points are classified into the low vegetation, medium vegetation, and high vegetation classes. The high vegetation class is then recognized and separated from the building points using routines for capturing the planar characteristics of building surfaces.

The results of this classification are then used to infer a 3D building model. The approach integrates ground and building classes with additional attributes including slope and planarity, allowing for accurate and detailed 3D representations of urban structures. The methodology described above is robust and effective in all dimensions and would be highly applicable to many urban landscapes.

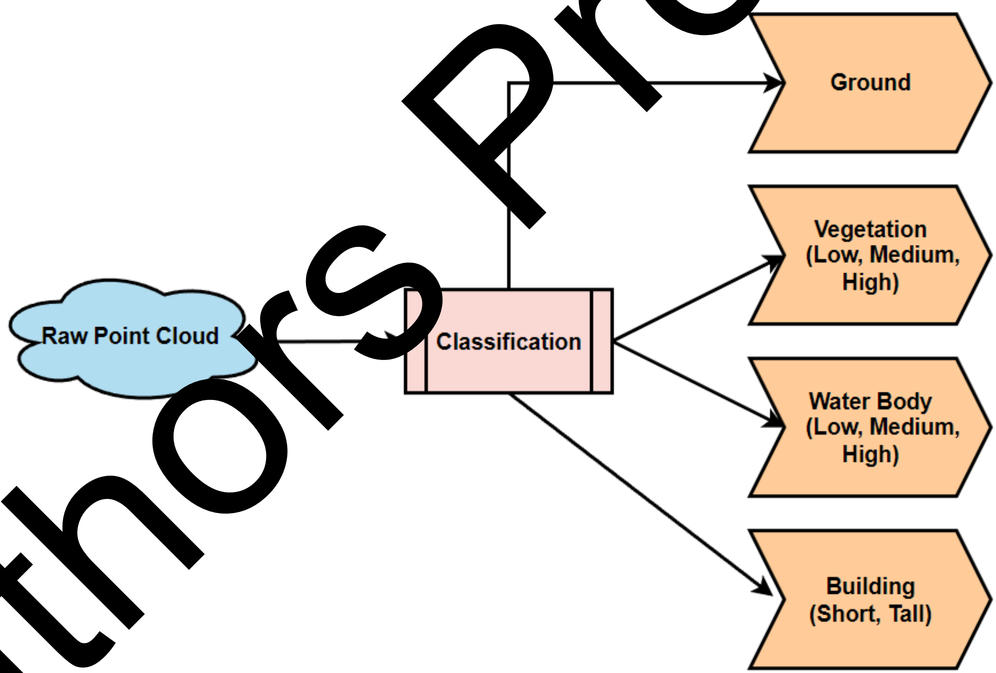


Figure 2. Procedural layout of proposed system

3.1.Study area

This research is done for the study area of Chennai, Tamil Nadu, India, a metropolitan city located along the southeastern coast as two different scenarios as provided in figure 3. Chennai was intentionally chosen as it has a highly varied urban character consisting of highly dense residential areas, commercial zones, industrial areas and pockets of green spaces like parks and mangroves. Chennai, with rapidly urbanizing and varied topography, serves as an ideal test bed for evaluating. A LiDAR dataset, with a point density of 18 points/m^2, was stored in LAS format for data processing. Based on Chennai's unique urban features, the LiDAR points were classified into five primary categories: (i) (Ground): Points representing roads, pavements and bare soil areas; (ii) Buildings: Roofs and other structural surfaces in areas of a city designated for residential, commercial or industrial use, represented by points. (iii) Vegetation: In turn, this component has been subdivided into low vegetation (shrubs and small plants) and high vegetation (trees, mainly mangroves and urban tree cover). (iv) Water Bodies: Some points corresponding to rivers, lakes and coastal waterlines which were significant in Chennai as most of the part was located near Bay of Bengal. (v) Miscellaneous Objects: It also include noise points and objects like utility poles and vehicles. The classification of the program reflects the morphology of Chennai, in terms of an urban morphology, as well as its urban terrain, making sure that the methodology accounts for the problems in densely populated areas and environmental elements such as coastal regions and urban greenery.

(a) (b)

Figure 3. Study site at the selected location

3.2.Building Footprint Extraction

Data preparation and extraction of building boundaries are conducted using ArcGIS software. Results of tests carried out on the UNSW dataset using SVM showed that it misclassified many buildings and roads in sloping terrain. Furthermore, a point and object based classifying tool accessed through ERDAS can be neatly viewed through careful visual inspection of output points appearing to be acceptable. Hence, this study choose the object based classifying tool the choice to build classification and continue footprint extraction on the results of classified buildings points from time series Lidar data sets with 3D urban growth analysis. The 3D urban development analysis using grey level co-occurrence matrix measures and support vector machine classifications required the extraction of building boundaries from Lidar data in preparation for use of NDSM.

4. Results and discussion

A workflow for this classification customized for the urban structures and terrain varying across the city of Chennai is illustrated in Table 1. The classification starts with a density analysis directly applied to the raw LiDAR points to assign default class areas to raw LiDAR points. This step allows points to be segregated by spatial distribution and density of points.

Features related to height, slope, and terrain curvature have been used to detect points which are at the ground level, ignoring the flat terrains, slopes, and uneven ground which are common in Chennai's landscape.

Table 1. Spacial feature for classification

Developed Routines	Slope	Elevation	Density of Vegetation	Urban Structure	Water Body	Segment of Tree
Spatial Features	Terrain and Roughness	Height (Relative and Absolute)	Density based on Point and Varying Height	Planar and Height Properties	Intensity and Depth	Cluster and Height
Classes Adhered	Slope and Ground	Flat and Elevated	Low, Medium and High	Infrastructure along with Building	Low, Medium and High	Short and Tall

Below-surface variances are only in regions containing underground structures or buried utilities and involve depth and elevation anomalies. The parameters used to split vegetation into low, medium and high are height, proximity of the vegetation to the ground and canopy spread, typical to the mix of urban greenery and natural vegetation of Chennai.

For built-up area detection, surface regularity and eigenvalues are used to discriminate buildings from vegetation and open spaces. Finally, the data for this problem is required to classify coastal features and open areas, which are critical for Chennai being a coastal city, to be able to find flat areas with little or no vegetation and areas proximal to water bodies.

Table 2. Various classification and its threshold ranges

Classification	Ground	Vegetation	Water Body	Building
Threshold Range (m)	-0.1~0.3	0.3~1(Low) 1~3(Medium) 3~50(High)	-1~ 0	2.5~100

The elevation thresholds for each class as depicted in Table 2 serve to match Chennai's topographical and structural conditions. Vegetation classes are divided as increasing height thresholds which place various plant types commonly found in urban and peri-urban areas into dendrograms, and ground points are confined to a narrow elevation range. Negative elevation (sub-surface points) indicates features such as buried features or depressions.

Chennai's low-lying coastal zones have specific feature class (0.0 - 1.0 metres). Urban buildings, elevated structures, and rooftops are built-up areas class (>1.50 m). The unique elevations and spatial features of Chennai are captured with these thresholds so that the classification is precise.

4.1.Automatic Point-Based Classification

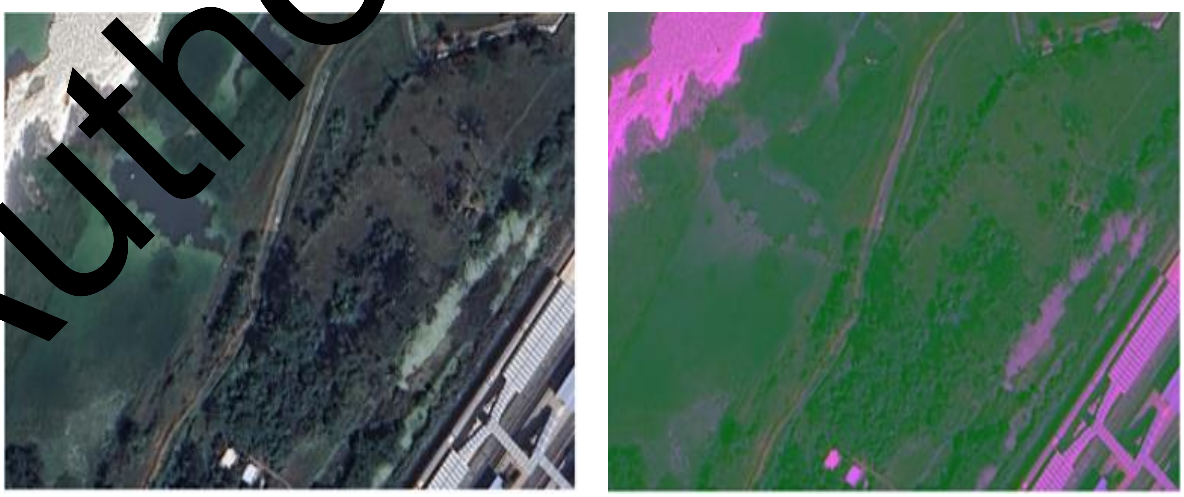
The point-based classification of LiDAR data collected from the Chennai study area was carried out systematically, resulting in the identification of several classes: buildings, low vegetation, medium vegetation, high vegetation, ground, default, subsurface, coastal features. For streamline of classification process and concentration on main objective, vegetation classes (Low, Medium and High vegetation) are grouped into one vegetation class. Ancillary classifications, like default and subsurface points, were also combined into a blended category called 'other features'.

The approach produced a simplified and effective basis for distinguishing critical classes (ground, buildings, vegetation, and others), making subsequent 3D modelling and analysis reliable. The results of the automatic point-based classification of Chennai region are presented in Figure 4, highlighting the use of the proposed spatial features as well as the hierarchal rule sets designed for the idiosyncrasies of the study area. Using this classification as a starting point, terrain mapping and urban structure reconstruction were performed.

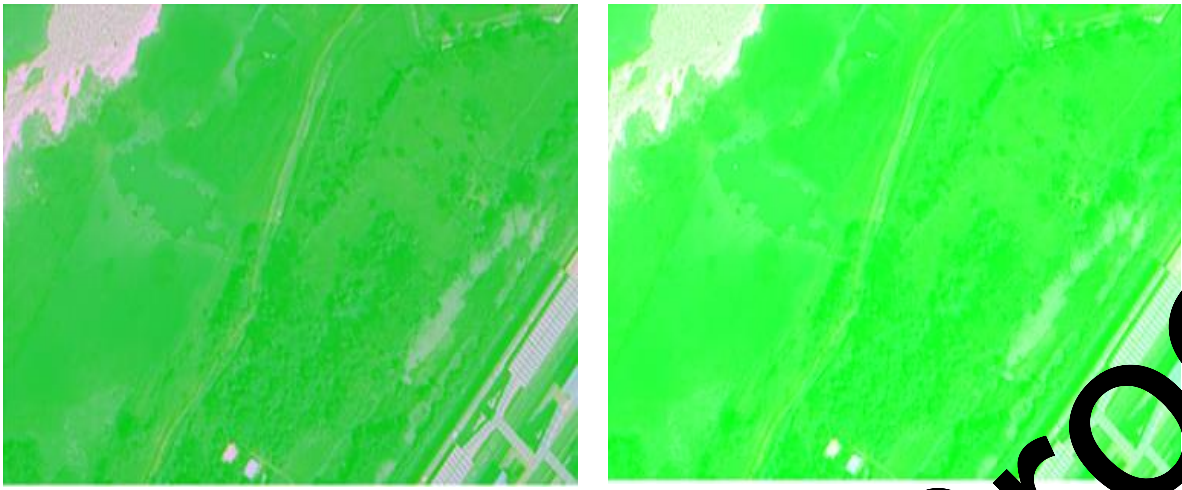
(a) (b) (c) (d)

Figure 4. Study site 1: (a) ortho (b) classified point (rough) (c) classified point (determined ground) (d) classified point (vegetation).

A methodology of LiDAR data classification, which will improve the accuracy of Tamil Nadu urban feature identification. These panel uses the raw satellite or aerial imagery of the area being studied. It is an urban environment with buildings, vegetation and land features: roads, with no classification. The high resolution of the imagery provides a base for the analysis. (b): A basic step of classification is employed to classify the unknown data at this step. The first rough classification groups points into rough classes like vegetation, buildings or ground. But it does not give precise classification as most of the points are misclassified. For instance, some ground points in built-up areas may be mistakenly classified as vegetation due to similar spectral characteristics. (c): As shown, this panel is composed of the output of vegetation-specific classification, in which vegetation LiDAR points are isolated. The adjustment of the parameters is proposed in an automatic point based classification methodology which better separates vegetation from ground points. The result is a more accurate identification of vegetation areas, particularly in regions where rough classification had errors. (d): Here we show a refined classification by means of spatial features, such as height, point density, and local plane orientation, as an example. This method is demonstrated to significantly improve misclassification, specifically between vegetation and ground and buildings in urban and built-up areas. An adjustment is made that better delineates vegetation and non-vegetation classes, as evidenced by the clarity of green, indicating vegetation. Finally, it illustrates the problems with existing urban LiDAR classification techniques and how the proposed automatic point-based approach addresses those problems. The methodology augments the accuracy by incorporating spatial features, and specifically, in urban areas with many complex terrains and structures such as those found in Tamil Nadu. The improved classification provides the fertile ground for generating detailed 3D urban models, which in turn help in urban planning, ecological conservation and disaster management. The success of the study and the consequent improvements observed in the panels (c) and (d) is largely attributed to the way the problem of limitations in grid-based approaches was handled and in the advancement of LiDAR's applicability in urban analytics.



(a) (b)



(c) (d)

Figure 5. Study site 2: (a) ortho (b) classified point (rough) (c) classified point (determined ground) (d) classified point (vegetation).

Stages of vegetation classification in natural and built-up areas are depicted in figure 5. The transition from raw imagery to enhanced classification through a point-based methodology that refines vegetation parameters for more accuracy is illustrated. Here's a detailed explanation of the panels: (a)The image in this panel shows the raw satellite or aerial imagery over the study area. Landscape features, for example water bodies and vegetation, and some built up areas, are included in the dataset. The original image serves as the foundation for subsequent classification processes, showing no distinction between vegetation and other features. (b): On stage where they use a basic classification method to classify the land cover into different land cover types. The classification of the vegetation is attempted to separate the vegetation from areas of non-vegetation, technique but not precise. Misclassification is evident, as some non-vegetation areas are assigned to the vegetation class (e.g., parts of built-up areas and water bodies are misclassified). (c): In this panel, the study isolate vegetation with an improved classification algorithm. The vegetation class is richer that does not depend only on one feature but on spatial features such as the height, point density, and local plane orientation. The green areas correspond to vegetation, but some inaccuracies still exist, especially near the boundaries of mixed land cover zones. (d): In the last stage, the proposed method optimizes the parameters to achieve still higher classification accuracy. Many of the misclassifications present in the rough classification are resolved with the adjustments. The vegetation is now clearly delineated, in which the overlap between non-vegetation and vegetation classes is minimal. In comparison, the green regions are more uniform, they are the true extent of what vegetation exists in the area. By coalescing the spatial features, the approach increases the accuracy of LiDAR based classification for discriminating vegetation from other land cover classes. The benefits of this methodology are particularly pronounced in urban and peri-urban areas in Tamil Nadu with complex terrain and mixed land use, which makes traditional land use classification very difficult. A map of accurate vegetation mapping, as shown here, support ecological conservation, urban planning and sustainable development initiatives.

5. Conclusion

An automated, point based classification methodology to aid in the accuracy of LiDAR data analysis is demonstrated with the results showing how improved classifications can be defined to distinguish vegetation from other land cover types. Using raw satellite imagery as an input, the methodology refines the classification process in stages to overcome the limitations of traditional grid-based approaches. It demonstrate the transition to adjusted vegetation parameters reduces misclassification, particularly in complex built up areas. The ability to handle challenging urban terrains is shown to correctly reassigned misclassified ground points from the rough classification stage. This capability is further emphasised in natural landscapes, as the algorithm isolates vegetation from mixed features such as water bodies and built up structures. It becomes very clear that the use of spatial features including height, point density, and local plane orientation can significantly improve classification accuracy. Overcoming such challenges as overlapping spectral characteristics and terrain complexity, the proposed approach improves the reliability of urban and ecological analysis using LiDAR. Overall, the proposed methodology provides a significant advance to geospatial technology, and it is a very robust tool for urban planning, disaster management and ecological conservation. The study shows that its ability to accurately map and analyse urban and rural parts renders it suitable for sustainable development efforts more generally.

References

1. H. Zhang, J. Chen, and Y. Li, "Urbanization and its environmental implications: A case study of Chennai," *Urban Clim.*, vol. 39, pp. 101045, Apr. 2023.

2. S. Gupta and R. Prasad, "GIS and LiDAR applications in Indian urban centers," *ISPRS Int. J. Geo-Inf.*, vol. 11, no. 7, pp. 420, July 2022.

3. J. Lee, P. Xu, and T. Kim, "Sustainable urban planning with LiDAR and GIS," *Sustain. Cities Soc.*, vol. 87, pp. 104019, Mar. 2023.

4. X. Yu, H. Liu, and W. Gao, "Sliding-window ConvLSTM for real-time predictive maintenance," *Future Gener. Comput. Syst.*, vol. 139, pp. 184–195, Nov. 2023.

5. B. Wang, Z. Luo, and Q. Zhu, "Point-cloud classification for urban landscapes," *Remote Sens.*, vol. 15, no. 3, pp. 1103, Feb. 2023.

6. M. Kumar, R. Raju, and S. Singh, "LiDAR-based urban planning for disaster resilience: A case study of Tamil Nadu," *Nat. Hazards*, vol. 110, pp. 1235–1252, Nov. 2022.

7. L. Chen, D. Zhang, and H. Wang, "High-resolution 3D spatial modeling using LiDAR and automated classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 196, pp. 84–98, Feb. 2023.

8. Y. Liu, X. Ma, and J. Zhao, “Point-based LiDAR classification for heterogeneous urban landscapes,” *IEEE Geosci. Remote Sens. Lett.*, vol. 20, no. 5, pp. 1–5, May 2023.

9. R. Rajan, A. Srinivasan, and N. Krishnan, “3D urban models for sustainable development: Tamil Nadu's urban dynamics,” *J. Urban Plan. Dev.*, vol. 148, no. 6, pp. 05022015, Dec. 2022.

10. J. Li, X. Zhou, and T. Wang, “Validation of LiDAR data for urban growth modeling in mixed topographies,” *Int. J. Remote Sens.*, vol. 44, no. 9, pp. 3502–3517, Aug. 2023.

11. S. Gupta and P. Sharma, “Analyzing dual urban transformations in India using GIS and LiDAR,” *Comput. Environ. Urban Syst.*, vol. 98, pp. 101832, Sept. 2022.

12. M. Patel and V. Desai, “Balancing conservation and development in ecologically sensitive urban zones,” *Land Use Policy*, vol. 124, pp. 10636[illegible], Apr. 2023.

13. F. Wang, J. Wu, and L. Zhang, “Integrating 3D LiDAR and GIS for sustainable urban planning,” *Sustainability*, vol. 15, no. 3, pp. 1740, Mar. 2023.

14. N. Krishnan and V. Subramanian, “GIS-based spatial analysis for smart city development in Tamil Nadu,” *Urban Sci.*, vol. 7, no. 2, pp. 56, May 2023.

15. H. Wang, X. Yu, and K. Chen, “Advanced GIS integration with LiDAR for flood risk management,” *Hydrol. Earth Syst. Sci.*, vol. 27, no. 8, pp. 1029–1042, Aug. 2023.

16. R. Shankar, S. Das, and V. Srinivasan, “Flood risk modeling using LiDAR-derived DEMs in Tamil Nadu,” *J. Hydrol.*, vol. [illegible]9, pp. 129125, Mar. 2023.

17. P. Kumar and S. Arora, “Smart city solutions using LiDAR-based 3D urban analysis,” *IEEE Access*, vol. 11, pp. 57234–57245, June 2023.

18. Y. Zhang, M. Wang, and X. Chen, “Urban density modeling for smart city projects: LiDAR applications,” *Cities*, vol. 139, pp. 103871, Oct. 2023.

19. V. Ra[illegible] and T. Natarajan, “Challenges and prospects of LiDAR data processing in dense urban areas,” *Geocarto Int.*, vol. 38, no. 7, pp. 983–998, July 2023.

20. [illegible] Zhou, J. Fan, and Y. Liu, “Automated point-based classification of LiDAR for urban vegetation overlap,” *Remote Sens. Environ.*, vol. 305, pp. 112029, Nov. 2023.

21. K. Singh, S. Gupta, and R. Reddy, “Developing robust classification methodologies for LiDAR in urban topographies,” *Remote Sens. Appl.: Soc. Environ.*, vol. 30, pp. 101033, Sept. 2023.