Integration of remote sensing and machine learning for transitional land cover mapping and urban spatial feature analysis in Da Nang

Tích hợp viễn thám và học máy để lập bản đồ lớp phủ đất chuyển tiếp và phân tích các đặc điểm không gian đô thị tại Đà Nẵng

Hoang Nhat Duc[a,b*]
Hoàng Nhật Đức[a,b*]

[a]Institute of Research and Development, Duy Tan University, Da Nang, 550000, Viet Nam
[a]Viện Nghiên cứu và Phát triển Công nghệ cao, Đại học Duy Tân, Đà Nẵng, Việt Nam
[b]Faculty of Civil Engineering, School of Engineering and Technology, Duy Tan University, Da Nang, 550000, Viet Nam
[b]Khoa Xây dựng, Trường Công nghệ và Kỹ thuật, Đại học Duy Tân, Đà Nẵng, Việt Nam

(Date of receiving article: 12/05/2025, date of completion of review: 18/08/2025, date of acceptance for posting:)

Abstract

Da Nang city has currently experiencing rapid land use transitions as a result of urban expansion and infrastructure development. This study proposes an advanced framework integrating machine learning, remote sensing, and statistical analysis to evaluate the transitional land cover in 2024. Random forest model coupled with Sentinel-2 remote sensing data is used to recognize transitional land cover. Herein, the land cover in the study area is classified into two categories: “stable land cover” and “transitional land cover.” Labeling is performed using Sentinel-2 imagery, with supplementary interpretation from Google Earth Pro. The machine learning model achieves satisfactory classification performance, with an accuracy of 87% and an F1 score of 0.87. Subsequently, statistical analyses are employed to reveal insights into the spatial distribution and characteristics of the transitional land cover in the study area. The spatial features of distance to coastlines, distance to industrial zones, distance to roads, and distance to rivers are used to quantify infrastructure and environmental influences on the spatial distribution of the variable of interest. The logistic regression analysis indicates that distance to roads is the most influential factor. Spatial autocorrelation analysis revealed a highly significant and positive spatial clustering of transitional land cover, with recognizable clusters across districts. These clusters align with major infrastructure developments and current status of urban development in the study area. The findings of this study can be useful for supporting urban management and land use monitoring in Da Nang.

Keywords: urban planning; transitional land cover; machine learning; remote sensing; statistical analysis.

Tóm tắt

Thành phố Đà Nẵng hiện đang trải qua quá trình chuyển đổi sử dụng đất nhanh chóng do đô thị hóa và phát triển hệ thống cở sở hạ tầng. Nghiên cứu của chúng tôi đề xuất một phương pháp tiên tiến kết hợp học máy và phân tích dữ liệu viễn thám để đánh giá lớp phủ đất chuyển tiếp tại Đà Nẵng vào năm 2024. Mô hình Random Forest kết hợp dữ liệu vệ tinh Sentinel-2 được sử dụng để nhận diện lớp phủ đất chuyển tiếp. Lớp phủ đất trong khu vực nghiên cứu được phân thành hai loại: “lớp ổn định” và “lớp chuyển tiếp”. Việc gán nhãn được thực hiện dựa trên quan sát ảnh vệ tinh Sentinel-2, kết hợp với phân tích ảnh trong Google Earth Pro. Mô hình học máy đạt hiệu quả phân loại ở mức tốt, với độ chính xác 87% và chỉ số F1 đạt 0,87. Phân tích thống kê được tiến hành để khảo sát các đặc điểm phân bố không gian của lớp phủ đất chuyển tiếp. Các đặc trưng không gian về khoảng cách đến đường bờ biển, khoảng cách đến khu công nghiệp, khoảng cách đến đường giao thông, và khoảng cách đến sông ngòi được sử dụng để khảo sát ảnh hưởng của các yếu tố cơ sở hạ tầng và môi trường đối với phân bố không gian của biến số được nghiên cứu. Phân tích hồi quy logistic cho thấy khoảng cách đến đường là yếu tố ảnh hưởng mạnh nhất. Kết quả phân tích tương quan không gian cho thấy sự tập trung rõ rệt của các khu vực bề mặt đất chuyển tiếp. Điều này phản ánh sự thay đổi bề mặt đất do các dự án hạ tầng trọng điểm và phát triển đô thị. Nghiên cứu của chúng tôi có ý

**Corresponding author:* Hoang Nhat Duc
Email: hoangnhatduc@duytan.edu.vn

nghĩa thiết thực trong việc hỗ trợ công tác quản lý và giám sát hiện trạng bề mặt đất đô thị tại Đà Nẵng.
Từ khóa: quy hoạch đô thị; lớp phủ đất chuyển tiếp; học máy; viễn thám; phân tích thống kê.

1. Introduction

Rapid urbanization and changes in land use, especially the increase in construction land, are common trends in many regions of Vietnam. Land use and land cover maps provide crucial information for effective land management and planning [1]. Moreover, monitoring economic activities plays an important role in understanding economic conditions and supporting the decision-making processes aimed at sustainable city development [2]. Since economic activities gradually alter the land surface, data that illustrate these land use changes significantly facilitates frequent and large-scale observation of construction. This approach can greatly enhance the understanding of the on-going economic circumstances.

According to [2], changes in the intensity of economic activity can be observed via land cover data. Socioeconomic factors are typically the causes of these changes and land cover patterns directly reflect economic processes. [5] has demonstrated a strong correlation between land use change patterns and increases in Gross Domestic Product. Moreover, the relationship between economic growth and land cover is recognized to be complex and involves mutual influence, rather than a simple one-way effect. Economic activities have caused significant transformations to the land surface. The reason is that as economies and populations grow, changes in land use and land cover have accelerated. Concurrently, shifts in land use and land cover have substantial effects on the economy. It is because land serves as a fundamental production resource, and its use is crucial for supporting economic growth.

Traditional data collection methods such as mapping and ground surveying are time-consuming and costly [2]. Additionally, the information is not updated frequently and is difficult to access. For large-scale areas, high-resolution land cover mapping takes a massive amount of data for classification in traditional geospatial techniques, so choosing Google Earth Engine (GEE) for classification makes it simple to categorize an entire large region [1]. Analyses based on remotely-sensed images provide an up-to-date and realistic presentation of the land surface. The spatial map of land use and land cover as well as other earth surface features can be done quickly and more accurately because of the use of GEE, remote sensing, geographic information system technology, and machine learning [6].

Sentinel-1 and Sentinel-2 imagery has become an essential resource for automated urban mapping and monitoring [4,8]. They enable accurate extraction of built-up areas and detailed detection of land cover changes. A two-step framework for the automated extraction of built-up areas using Sentinel-1 and Sentinel-2 data was introduced in [7]. In the first step, the system automatically samples and labels training data for built-up and non-built-up classes from the Sentinel-1 and Sentinel-2 imagery for a specified area of interest. In the second step, a cross-fusion neural network uses these training samples to generate a built-up map covering the entire study area. Sentinel-2 imagery was employed for monitoring urban land cover changes in [10]; this study relied on remote sensing data to enable detailed tracking of urban transformations.

Based on these motivations, this study proposes a framework that combines machine learning, Sentinel-2 remote sensing data, and statistical analyses to map and analyze the pattern of transitional land cover. The mainland region of Da Nang City in 2024 is chosen as the study area. This region is selected because it has been undergoing rapid transformation in recent years. Large-scale infrastructure projects, new residential and commercial developments, and the expansion of tourism facilities have accelerated the pace of urbanization.

In summary, this study aims to answer the following research questions: (i) What are the current locations of transitional land cover in Da Nang? (ii) Is there any spatial correlation between the locations of this land cover type and the spatial features in the city? (iii) What are the most influential variables in land transition within the study area? The rest of the paper is organized as follows: the second section reviews the research methods and materials used. The third section presents the research results, followed by the conclusion section.

2. Research method and materials

2.1. General description of the study area and remote sensing data

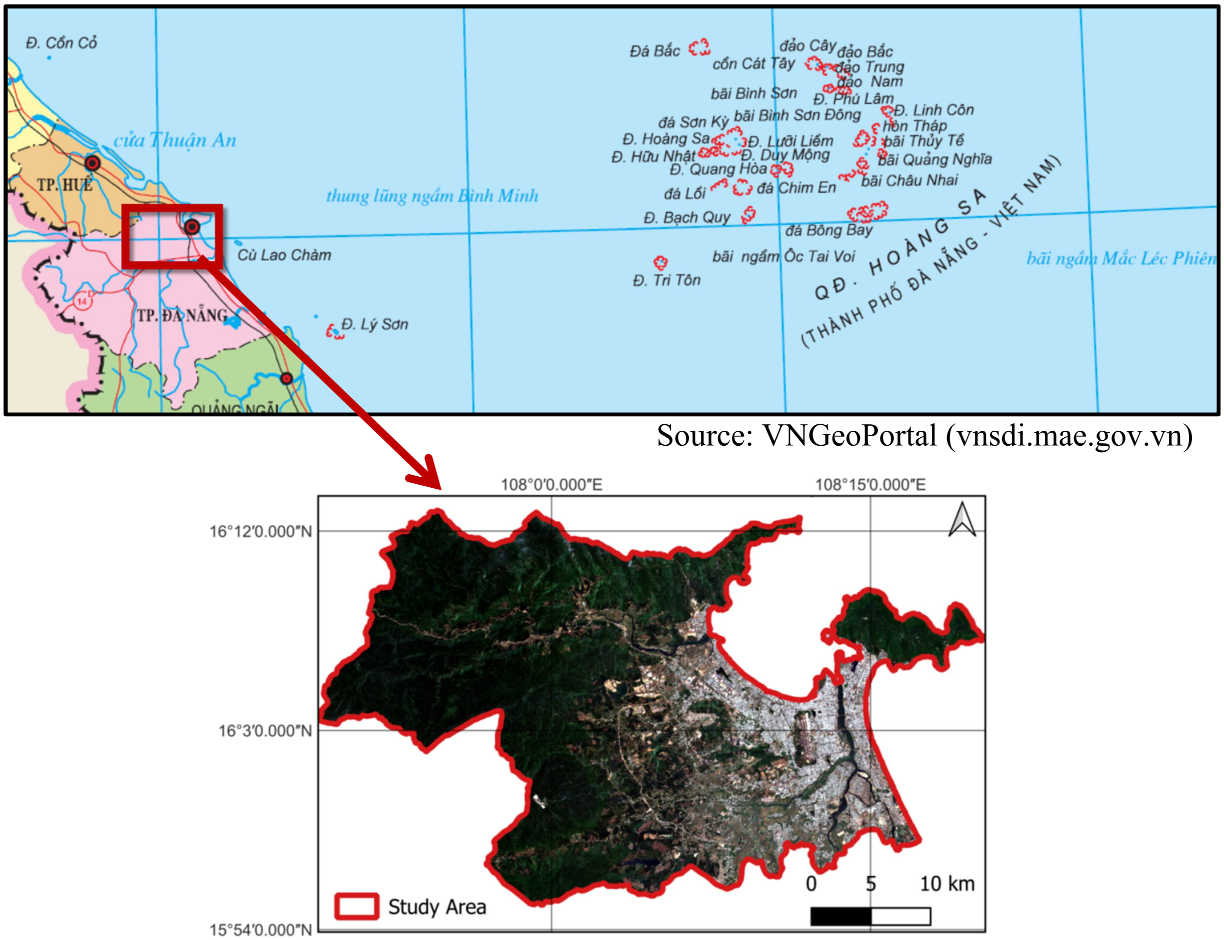
Da Nang is a city in central Vietnam known for its varied geography, which includes mountains, coastal plains, and marine areas. The city lies between the East Sea to the east and the Truong Son Mountains to the west. In 2024, Da Nang comprised the districts of Hai Chau, Thanh Khe, Son Tra, Ngu Hanh Son, Lien Chieu, Cam Le, Hoa Vang, and the Hoang Sa Islands in the East Sea. It is important to note that an administrative reform was enacted in mid-2025 by the Vietnamese government. Following this reform, Da Nang was merged with the neighboring Quang Nam Province to form a newly enlarged Da Nang City. For the most current provincial boundary of Vietnam, readers are guided to VNGeoPortal at https://vnsdi.mae.gov.vn/home.

The study area of the current work (see Figure 1) is restricted to the mainland region of Da Nang in 2024. This region is undergoing a period of rapid urban expansion, dynamic urban development, and major infrastructure projects. In recent years, large-scale urban projects have been carried out in the study area, including various residential, commercial, and tourism facilities. The study area is also attracting significant investment in industrial zones, technology parks, and innovation centers. These projects are reshaping the city’s urban landscape and facilitating robust economic growth.

This study relies on Sentinel-2 satellite to provide medium-resolution multispectral imagery of the land surface. Sentinel-2 offers frequent global coverage with a revisit time of five days. These characteristics make Sentinel-2 data highly valuable for land use and land cover classification. The Sentinel-2 data retrieved from Google Earth Engine (GEE) for this study includes images collected between January 1, 2024, and December 31, 2024, with the cloud coverage of less than 25%. The used spectral bands of the Sentinel-2 satellite are summarized in Table 1. For more details on Sentinel-2 imagery, readers are guided to the Copernicus and ESA documentation at https://sentinels.copernicus.eu/. It is noted that the B1, B9, and B10 bands are excluded because their coarse spatial resolution is not suitable for accurate land cover mapping. These bands are employed by the machine learning algorithm to classify the land cover in the study area into two distinctive labels: ‘stable land’ (the negative class) and ‘transitional land’ (the positive class).

2.2. Random forest classifier

The Random Forest (RF) classifier is a robust supervised machine learning algorithm that excels at pattern classification tasks. It operates by constructing an ensemble of decision trees, each trained on a random subset of the data and a random subset of input features, a process known as bagging [9]. The final prediction is made by aggregating the outputs of all trees, typically through majority voting for classification problems. This ensemble approach reduces overfitting and improves generalization compared to single decision trees. This feature makes RF particularly reliable for multivariate and nonlinear datasets [3]. The main advantages of RF are high accuracy and ability to handle high-dimensional data. Therefore, RF is especially suitable for land cover classification using spectral bands from Sentinel-2 satellite imagery. This study implements RF in Google Earth Engine’s code editor.



Source: VNGeoPortal (vnsdi.mae.gov.vn)

Source: European Union/ESA/Copernicus

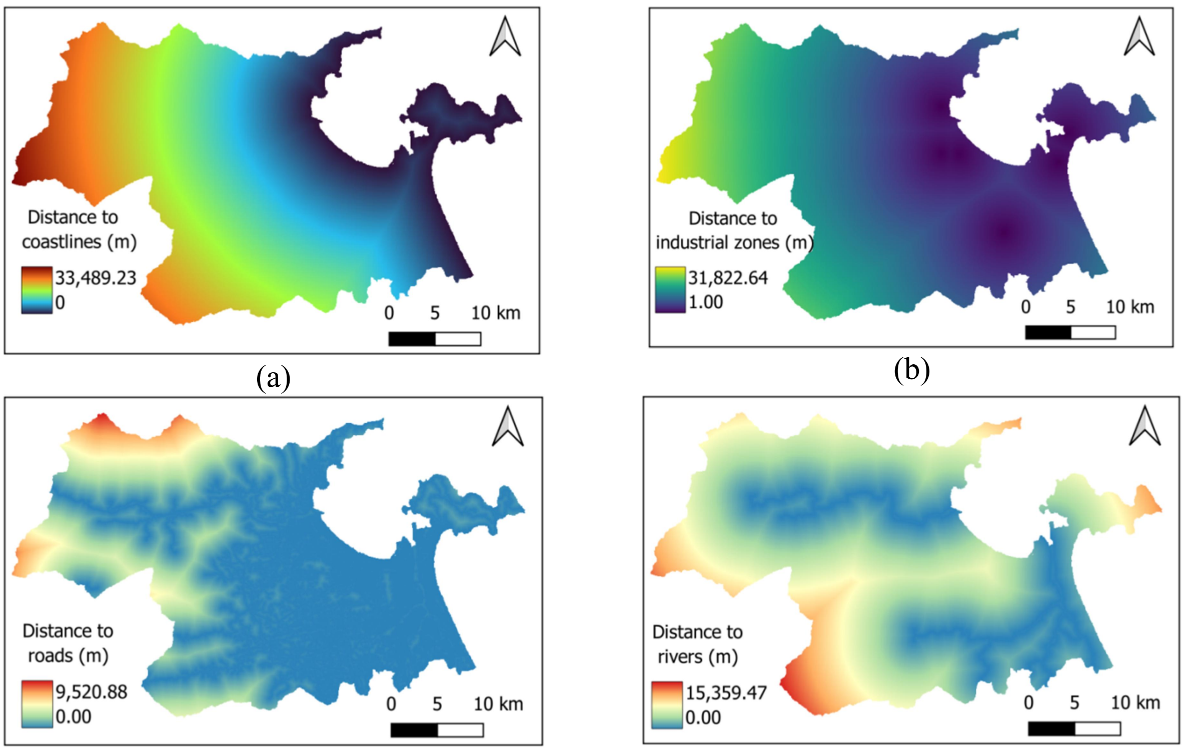
Figure 1. The study area (true color composite Sentinel-2 imagery)

Table 1. The employed spectral bands of the Sentinel-2 used for land cover mapping

Spectral band	Resolution (m)	Wavelength (nm) (S2A / S2B)	Description
B2	10	496.6 (S2A) / 492.1 (S2B)	Blue
B3	10	560 (S2A) / 559 (S2B)	Green
B4	10	664.5 (S2A) / 665 (S2B)	Red
B5	20	703.9 (S2A) / 703.8 (S2B)	Red Edge 1
B6	20	740.2 (S2A) / 739.1 (S2B)	Red Edge 2
B7	20	782.5 (S2A) / 779.7 (S2B)	Red Edge 3
B8	10	835.1 (S2A) / 833 (S2B)	NIR
B8A	20	864.8 (S2A) / 864 (S2B)	Red Edge 4
B11	20	1613.7 (S2A) / 1610.4 (S2B)	SWIR 1
B12	20	2202.4 (S2A) / 2185.7 (S2B)	SWIR 2

2.3. Spatial features

The spatial features of distance to coastlines (Figure 2a), distance to industrial zones (Figure 2b), distance to roads (Figure 2c), and distance to rivers (Figure 2d) are used in this study for analyzing the spatial distribution of transitional land cover types in Da Nang. It is noted that the maps in this study uses the Universal Transverse Mercator (UTM) projection to the World Geodetic System 1984 (WGS84) within zone 49N. These distance metrics capture the proximity of land parcels undergoing transformation from rural or agricultural uses to urban land. This information can provide insights into the underlying drivers and patterns of urban development. Proximity to coastlines often reflects the attractiveness of waterfront areas for residential, commercial, or tourism-related urban expansion, while closeness to industrial zones may implies the influence of industrial development. The distance to roads is also a key indicator of infrastructure development. It is because areas near major transportation corridors are more accessible and thus more likely to experience urbanization and investment. Similarly, proximity to rivers can affect land suitability and real estate value.



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

Figure 2. Spatial features: (a) Distance to coastlines, (b) distance to industrial zones, (c) distance to roads, and (d) distance to rivers

3. Result and discussion

The RF model implemented in GEE employs 100 trees in the ensemble with a maximum number of leaf nodes of 50 and a bagging fraction of 0.7. The model's performance is summarized in Table 2. RF is used to classify land cover in the study area into stable and transitional land areas. In the training phase, the model achieved exceptional results with 97% Classification Accuracy Rate (CAR), precision, recall, and F1-score. When applied to the testing dataset, the model maintained good performance with 87% across all metrics. The reduction of accuracy in the testing phase compared to that in the training phase indices transitional land cover recognition is indeed a complex classification task. Notably, the balanced precision and recall values in the testing phase indicate that the model is equally effective at identifying both transitional and stable land areas without significant bias toward either class. With an F1-score of

0.87 in the testing phase, the model successfully balances precision and recall. The resulting transitional land cover map of the study area is presented in Figure 3.

Table 2. Performance of the RF model

Phase	CAR	Precision	Recall	F1-Score
Training	0.97	0.97	0.97	0.97
Testing	0.87	0.87	0.87	0.87

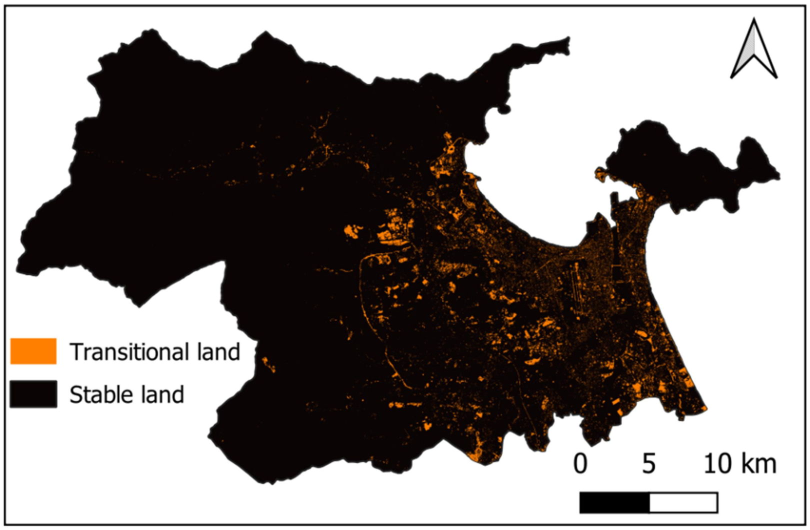


Figure 3. Transitional land cover map in 2024

Table 3. Statistical description of the spatial features

Spatial variables	**Mean**	**Std**	**Max**	**Min**
Distance to coastlines	5550.64	4499.84	25953.94	0.13
Distance to industrial zones	2301.46	1808.05	14436.56	0.42
Distance to rivers	4830.45	3151.34	23918.84	16.06
Distance to roads	53.74	122.45	3353.59	0.00

Table 4. Computed percentiles for the spatial variables

Variable	**50th percentile**	**60th percentile**	**70th percentile**	**80th percentile**	**90th percentile**
Distance to coastlines	4663.25	6027.09	7364.62	9071.95	12426.08
Distance to industrial zones	4368.33	4963.31	6125.15	7329.03	9164.25
Distance to rivers	1927.70	2482.77	3155.59	4009.51	4969.93
Distance to roads	26.05	35.62	49.96	73.31	114.66

The statistical analysis of transitional land cover in Da Nang reveals distinct spatial patterns linked to key geographic features. Transitional areas exhibit a mean distance of 5,550 meters from coastlines, with significant variability (standard deviation: 4,499 meters), indicating both coastal and inland transitions. Proximity to industrial zones shows a mean distance of 2,301 meters, but the 90th percentile (9,164 meters) highlights expansion in peripheral regions. Rivers influence transitions moderately, with a mean distance of 4,830 meters, though 50% of transitions occur within 1,927 meters (50th percentile); this fact suggests that river-adjacent development is commonly observed in the region. Roads emerge as the strongest correlate, with a mean distance of 53 meters and 50% of transitions within 26 meters; this fact underscores the role of infrastructure in driving land use changes in Da Nang. The 90th percentile for road proximity (roughly 115 meters**)** further confirms that most transitional areas cluster near transportation networks. Coastal transitions display the widest spread, with 10% of cases exceeding 12,426 meters from shorelines; this result, to some degree, reflects urbanization processes in the hinterland.

Table 5. Logistic model's weights

Spatial features	Distance to coastlines	Distance to rivers	Distance to industrial zones	Distance to roads
Weight	-0.195275	-0.35405	-0.595117	-5.791796

This study uses a logistic regression model to investigate how spatial features influence transitional land cover. For training, 5,000 data points were sampled from transitional land cover areas and 5,000 from stable land areas in 2024. The input features include four spatial variables, which were normalized using Z-Score standardization. The whole dataset (10,000 samples) was randomly split into a training set (70%) and a testing set (30%). After training the model over 100 iterations, it achieved a CAR of 78.84% on the training data and 78.63% on the testing data. The logistic model's weight is reported in Table 5, which can be assessed to reflect the influence of each spatial variable. Since all weights are negative, it can be seen that as distance to these variables increase, the likelihood of transitional land cover decreases. In other words, transitional areas tend to be closer to roads, industrial zones, rivers, and coastlines. Via the model's weights, it is observable that distance to roads is the dominating factors with a weight of -5.792. Industrial zones and rivers show moderate influence and coastlines have the lowest contribution. This result can be explained by the fact that the road network serve as a key infrastructure for land development. Hence, proximity to roads is critical for the appearance of transitional areas.

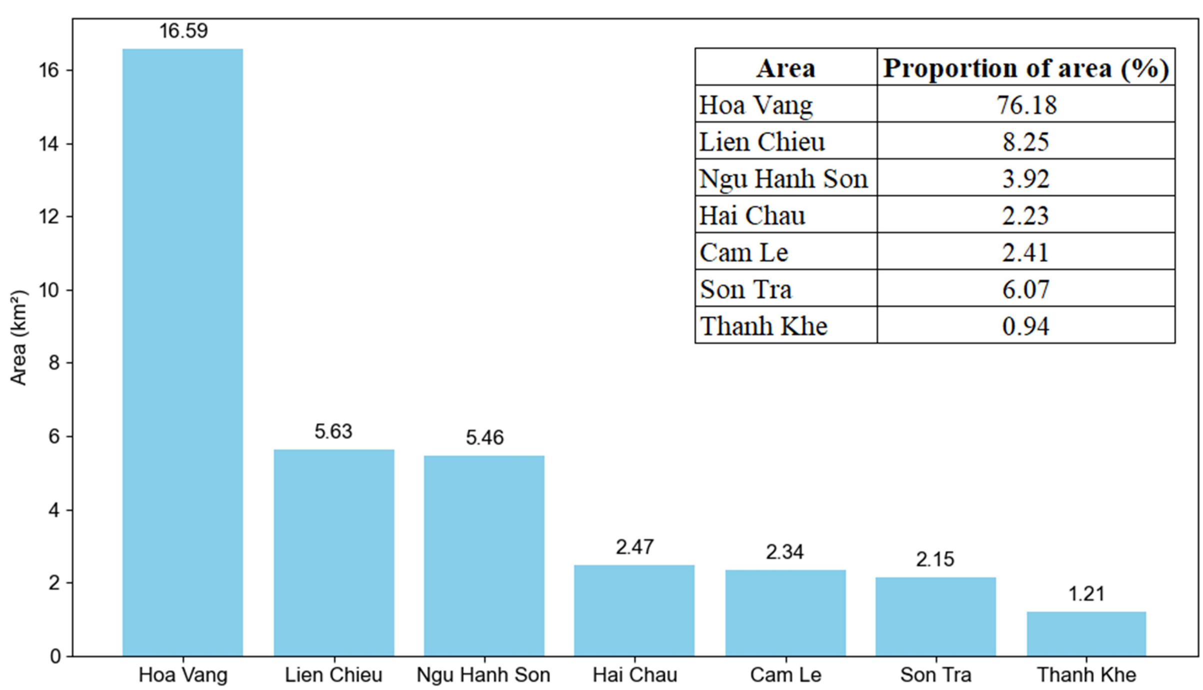


Figure 4. District-wise allocations of transitional land cover in the study area

Furthermore, the allocations of the area of transitional land cover across the districts of Da Nang city reveal significant spatial variation (refer to Figure 4). Hoa Vang stands out with the largest area of transitional land. This district accounts for 16.59 km^2, which is substantially higher than other districts. This suggests that Hoa Vang is experiencing the most extensive conversion of natural land surface to urban areas or for infrastructure developments. This result can be explained by the fact that Hoa Vang consists of larger land reserves and is experiencing intensive urban expansion. Lien Chieu and Ngu Hanh Son districts also have notable areas of transitional land cover, at 5.63 and 5.46 km^2, respectively. These outcomes point out active urban development in these zones. In contrast, the districts of Cam Le, Hai Chau, Son Tra, and Thanh Khe have smaller areas of transitional land cover; the areas range from 1.21 to 2.47 km^2. This pattern, to some degree, reflects the more urbanized and spatially constrained nature of these central districts. Overall, the allocations of the area of transitional land cover reveal that the urban growth in Da Nang is concentrated in the peripheral districts, particularly Hoa Vang. Meanwhile, the central districts are experiencing less intensive land transition.

The logistic regression model identifies distance to roads as the dominant predictor of transitional land cover, with a standardized coefficient far exceeding other variables. This finding may point out important implications for urban planning in the study area. Notably, transitional zones exhibit strong clustering near road networks, as evidenced by the standard deviation (Std) of 1,342.24 meters for road proximity. For every Std increase in distance from roads, the log-odds of land being transitional decrease by 5.79 units. Consequently, areas within 1 unit of Std (roughly 1.3 km) of roads are more likely to experience transition. This fact reflects Da Nang's infrastructure-driven urbanization patterns.

Herein, the Moran's I test is utilized to investigate the spatial autocorrelation of the transitional land cover in Da Nang. This test reveals a significant clustering pattern in the spatial distribution of transitional land cover across Da Nang City (Moran's I = 0.3584). This positive spatial autocorrelation (Moran's I > 0) indicates that areas of transitional land cover tend to aggregate near each other, rather than being randomly distributed. The extremely high z-score (426.48) and near-zero p-value provide robust statistical evidence against spatial randomness, confirming a non-uniform urban transformation process. Compared to the expected value under randomness, the observed outcome suggests moderate clustering intensity. These results imply that transitional zones are influenced by localized drivers such as road networks, industrial expansion, or coastal development policies. These spatial features help create spatially correlated zones of land cover transition.

4. Concluding remarks

This study has carried out an investigation into the current spatial distribution of transitional land cover in the mainland region of Da Nang City. Via the analysis outcomes, it can be shown that Da Nang's rapid urbanization has caused significant land cover transitions, particularly in peri-urban and infrastructure-adjacent areas. By integrating machine learning, remote sensing, and spatial statistics, the proposed framework effectively maps transitional land cover with 87% accuracy using Sentinel-2 data and RF classification.

Logistic regression identifies road proximity as the strongest predictor of transitions, with the corresponding weight of -5.79; this weight indicates a sharp decline in transitional likelihood as distance from roads increases. Additionally, spatial autocorrelation analysis via the Moran's I test helps confirm clustered transition patterns, which align with infrastructure projects and industrial zones. Notably, the district of Hoa Vang exhibits the largest transitional areas (16.59 km²); this fact reflects concentrated urban expansion in peripheral regions, while central districts show limited transitions. These findings also accentuate the role of road networks as the main driver for land development and highlight the underlying pattern of Da Nang’s growth.

The research findings provide valuable insights for the task of urban planning. Transitional land cover’s strong correlation with roads and industrial zones suggests that infrastructure investments directly shape land use changes. The clustering of transitions in Hoa Vang and Lien Chieu districts may necessitate the need for targeted zoning policies to balance growth with environmental protection. While this study provides critical insights into Da Nang’s transitional land cover status, several extensions could further improve its impact. First, mapping of the temporal change in transitional land cover can help evaluate the urban development trend in the region. Second, incorporating socioeconomic data, such as population density and government policy, can improve the logistic regression model and reveal causal drivers beyond the currently employed spatial proximity.

References

[1] Arpitha, M., Ahmed, S.A., Harishnaika, N. (2023). “Land use and land cover classification using machine learning algorithms in Google Earth Engine”. *Earth Science Informatics* (16), 3057–3073. DOI: 10.1007/s12145-023-01073-w

[2] Chen, C., He, X., Liu, Z., Sun, W., Dong, H., Chu, Y. (2020). “Analysis of regional economic development based on land use and land cover change information derived from Landsat imagery”. *Scientific Reports* (10), 12721. DOI: 10.1038/s41598-020-69716-2.

[3] Hung, D.V., Dieu, T.B., Linh, T.X, Duc, H.N. (2019). “Enhancing the accuracy of rainfall-induced landslide prediction along mountain roads with a GIS-based random forest classifier”. *Bulletin of Engineering Geology and the Environment* (78, 2835–2849. DOI: 10.1007/s10064-018-1273-y.

[4] De Fioravante, P., Luti, T., Cavalli, A., Giuliani, C., Dichicco, P., Marchetti, M., Chirici, G., Congedo, L., Munafò, M. (2021). “Multispectral Sentinel-2 and SAR Sentinel-1 integration for automatic land cover classification”. *Land* (10), 611. DOI: 10.3390/land10060611.

[5] He, C., Huang, Z., & Wang, R. (2014). “Land use change and economic growth in urban China: A structural equation analysis”. *Urban Studies* (51), 2880–2898. DOI: 10.1177/0042098013513649.

[6] Duc, H.N., Duc, T.V., Canh H.T. (2025). “From data to insights: Modeling urban land surface temperature using geospatial analysis and interpretable machine learning”. *Sensors* (25), 1169. DOI: 10.3390/s25041169.

[7] Li, Y., Matgen, P., Chini, M. (2025). “Extraction of built-up areas using Sentinel-1 and Sentinel-2 data with automated training data sampling and label noise robust cross-fusion neural networks”. *International Journal of Applied Earth Observation and Geoinformation* (139), 104524. DOI: 10.1016/j.jag.2025.104524.

[8] Petrushevsky, N., Manzoni, M., Monti-Guarnieri, A. (2022). “Fast urban land cover mapping exploiting Sentinel-1 and Sentinel-2 data”. *Remote Sensing* (14), 36. DOI: 10.3390/rs14010036.

[9] Svoboda, J., Štych, P., Laštovička, J., Paluba, D., Kobliuk, N. (2022). “Random forest classification of land use, land-use change and forestry (LULUCF) using Sentinel-2 data—A case study of Czechia”. *Remote Sensing* (14), 1189. DOI: 10.3390/rs14051189.

[10] Wang, N., Li, W., Tao, R., Du, Q. (2022). “Graph-based block-level urban change detection using Sentinel-2 time series”. *Remote Sensing of Environment* (274), 112993. DOI: 10.1016/j.rse.2022.112993.