

Article

Hybrid CNN Vision Transformer Framework with Grad-CAM and SHAP Analysis for Urban Change Detection

Abdulmajid A. Alnoamani * and Tawfiq Hasanin 

Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; thasanin@kau.edu.sa

* Correspondence: aalnoamani0004@stu.kau.edu.sa

Abstract

To track land use and land cover transformation in Makkah, techniques that allow steep relief, spectral confusion, and dense sacred–commercial mosaics, and can be justified in terms of planning, should be used. Satellite images are tedious and prone to uneven labeling on mixed-pixel boundaries, particularly in urban regions and Haram borders. Using multi-temporal Landsat-8 data (2013 and 2024), a hybrid deep learning architecture comprising U-Net, DenseNet201, and a Vision Transformer was trained. U-Net retained the geometry of the boundaries, DenseNet201 reinforced feature transfer across heterogeneous textures, and the transformer modeled long-range context. Gradient-weighted Class Activation Mapping (Grad-CAM) was used to incorporate interpretability during spatial attention mapping, and Shapley Additive exPlanations (SHAP) during spectral topographic attribution, after which paired class-level statistical tests were performed. Modern residential increased from 15% to 20% (180 million to 240 million m²); roads from 5% to 10% (60 million to 120 million m²); industrial facilities from 3% to 5% (36 million to 60 million m²). The vegetation expanded by 1 to 5% (an addition of 48 million m²), and agriculture declined by 2 to 1% (a loss of 12 million m²). Its tension with urban development and preservation of productive land was growing. The proposed U-Net–DenseNet201–ViT hybrid system achieved over 98% overall accuracy on the test data for both study years, with kappa coefficients of 0.978 and 0.981 for 2013 and 2024, respectively. Grad-CAM identified attention focused on development fronts and transport corridors, whereas SHAP identified SWIR, thermal response, and slope as the main drivers. Significant class-level gains were statistically validated ($p < 0.01$), confirming an interpretable and auditable account of land transformation in Makkah.

Keywords: LULC; hybrid system; CNN; ViT; remote sensing; grad CAM; SHAP



Academic Editor: Enrico Corrado
Borgogno Mondino

Received: 19 April 2026

Revised: 23 June 2026

Accepted: 26 June 2026

Published: 1 July 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Saudi Arabia's Vision 2030 has accelerated urban development in Makkah by intensifying infrastructure expansion, redevelopment, and enhancements to pilgrimage capacity. Consequently, urban planning in Makkah faces the challenge of balancing rapid infrastructure growth, environmental sustainability, and the preservation of religious and historical heritage sites, while supporting the city's role as a model for sustainable development in sacred cities [1]. The LULC classification process has evolved over time due to technological and computational advances in remote sensing. LULC classification was initially carried out manually by experts interpreting aerial photographs, and later by interpreting satellite imagery through time-consuming, subjective decision-making that introduced inherent inconsistency. Alongside the LULC classification, other associated disciplines grew with

the system. Early methods were based on expert interpretation of aerial photography and were slow and subjective. Digital image analysis introduced semi-automatic, pixel-based classifiers, maximum-likelihood classifiers, and support vector machines, among others, to accelerate the process; however, they encounter difficulties in complex cityscapes with patches [2]. The incorporation of AI into LULC classification has revolutionized traditional remote sensing analysis by providing powerful computational models that can learn complicated spatial and spectral patterns. Deep learning techniques, such as CNNs, are considered the most effective tools for identifying and classifying fine-grained land cover characteristics within complex urban landscapes, achieving higher accuracy [3]. These models are effective at processing high-dimensional satellite images and learning hierarchical features that traditional methods cannot capture. Moreover, AI-based solutions enable automatic, dynamic detection of land use changes over time, thereby facilitating sustainable urban planning, environmental impact assessment, and policymaking [4]. Machine learning, a subset of AI, aims to develop algorithms that learn patterns and generate data-driven predictions or decisions without being explicitly programmed for these tasks [5]. Based on the statistical theory and computational design, ML algorithms achieve efficacy through repeated exposure to adaptively adjusted data. In the geospatial domain, machine learning methods have been widely used for LULC classification, anomaly detection, and change monitoring, particularly when multifrequency, multi-temporal stacks of remote sensing satellite imagery are available [6]. An ANN is an algorithm that has also been successfully used to classify different land types by detecting their spectral, spatial, or textural properties [7].

However, these approaches are often limited in highly heterogeneous urban environments, where fine spatial differences, mixed pixels, and spectrally similar land cover classes complicate reliable classification. The CNNs have outperformed classical machine learning methods for LULC mapping by learning hierarchical spatial–spectral features directly from remote sensing imagery [8]. In urban land classification, CNN-based models can extract subtle spatial, spectral, and textural patterns that are difficult to represent using traditional handcrafted features [9]. Recent deep learning and attention-based models have improved LULC mapping by learning spatial–spectral representations from remote sensing imagery and, where available, by incorporating temporal information and terrain-derived predictors such as elevation and slope [10].

Since hybrid CNN-ViT systems are used in LULC classification, they enable precise and scalable monitoring of land use changes over time [11]. They are essential to urban planning, environmental conservation, and resource management, as documented in studies of rapidly changing cities such as Makkah, Saudi Arabia. DenseNet201 and U-Net show how single-network architectures can successfully extract both spatial and spectral characteristics in urban and natural settings. Architectures (e.g., U-Net, DenseNet201) that capture both global semantic context and local spatial precision are essential for recognizing diverse urban types, mountain areas, and subtle land transitions [12]. Thus, large-scale deep learning in geospatial studies can enable more in-depth analysis and improved classification, such as for the dynamically evolving urban landscape in Makkah [13].

This study proposes a hybrid U-Net-DenseNet201-ViT architecture for multiclass land use mapping in the holy, arid, and topographically uneven area of Makkah, combining U-Net, spatial fidelity, DenseNet201, hierarchical texture semantics, and a Vision Transformer for global contextual reasoning. We identified its decision logic through a systematic, spatially resolved visualization of spatial attention with Grad-CAM, and quantified the contributions of spectral and topographic predictors of urban frontiers and infrastructural edges, such as SWIR reflectance and slope, to each classification using SHAP analysis. These instruments go beyond the black box, achieving high accuracy through physically

explainable reasoning. This explainable pattern was then statistically tested at the class level using paired *t*-tests ($p < 0.01$), which demonstrated substantial superiority and provided an auditable criterion for remote sensing in sensitive cityscapes.

The main contributions of this study are as follows:

- **Interpretable application-specific hybrid framework:** We implemented and evaluated a U-Net–DenseNet201–ViT hybrid architecture for fine-grained multiclass LULC mapping in Makkah’s arid and topographically complex urban environment. The contribution of this study does not lie in claiming a previously unseen network combination, but in adapting this hybrid architecture to the Makkah urban–mountain context and integrating Grad-CAM and SHAP to provide spatial and feature-level interpretability.
- **Spatial attention analysis:** Grad-CAM was used to visualize the image regions that contributed most strongly to the model’s class predictions. The resulting activation maps help assess whether the model attends to plausible spatial cues, such as development edges, transport corridors, and urban–rock interfaces. These maps are therefore interpreted as model attention evidence rather than direct proof of real-world land transformation.
- **Feature attribution analysis using SHAP:** SHAP was used to quantify the relative contributions of spectral bands, spectral indices, thermal response, and DEM-derived topographic variables to the model’s class predictions. The detailed feature importance results are reported in the Results Section.
- **Statistical justification for the classification-level performance:** Rigorous paired *t*-tests ($p < 0.01$) were performed between our model and other groups for each land cover category. This statistical evidence confirms the strong consistency of the proposed structure’s superiority, demonstrating that its robustness is not a matter of chance but rather a consistent pattern.
- **A verifiable analytical framework:** Considering all these contributions—an interpretable model, spatial (Grad-CAM) and feature attribution (SHAP), and the statistical basis—the output is more than a classification map—it is a verifiable scientific explanation of urban transformation. This framework provides planners with a clear, evidence-based understanding of governance in cities subject to topographic and cultural constraints.

The remainder of this paper is organized as follows. Section 2 reviews related LULC studies, Section 3 describes the data and proposed framework, Section 4 presents the classification and interpretability results, Section 5 discusses the findings, and Section 6 concludes the study.

2. Related Work

Previous LULC studies in Saudi Arabia and arid urban environments provide an important methodological foundation, but they differ considerably in classification granularity, interpretability, and study-area specificity. Early remote sensing studies, such as Alghamdi and Moore [14], Alqurashi and Kumar [15], and Alqurashi et al. [16], relied mainly on Landsat imagery, GIS-based change detection, vegetation indices, and pixel-based classification to quantify urban growth and land cover transitions.

AlKheder et al. [17] applied deep learning to Saudi urban contexts and reported 91% accuracy in distinguishing selected urban structures, such as mosques and commercial buildings. However, the study remained narrow in classification granularity because it excluded broader LULC classes such as vegetation, old buildings, agriculture, bare land, and mountainous terrain. Alsalman et al. [18] quantified Makkah’s urban expansion by 38% between 2000 and 2020, especially near the Haram boundary, but did not use deep

learning for fine-grained classification. Baik et al. [19] provided a Vision 2030 policy-oriented interpretation, yet lacked computational validation through a reproducible classification model.

These studies were valuable for documenting long-term urban expansion and environmental change; however, their classification schemes were generally coarse, focusing on broad categories such as built-up land, vegetation, bare surfaces, and general urban expansion. Such granularity is insufficient for Makkah, where modern residential areas, old buildings, roads, industrial facilities, bare land, bare rock, vegetation, agriculture, and mountain terrain coexist within a compact and spectrally complex urban–mountain landscape.

Sumit et al. [20] examined several CNN architectures, such as various R-CNN and LSTM-based ones, using data-poor settings. They proved the effectiveness of CNN in LULC classification without considering the religious, topographic, and infrastructural features of Makkah.

A second group of studies attempted to improve classification and prediction through object-based analysis, cellular automata, and machine learning frameworks. For example, Alqurashi et al. [21] applied OBIA–MC–CA modeling to predict urban growth in Saudi cities, while Li et al. [22] reviewed the integration of machine learning and remote sensing for urban sustainability. These approaches improved spatial modeling and forecasting compared with traditional pixel-based methods. Nevertheless, they often depend on hand-crafted features or predefined spatial rules, which limit their ability to learn hierarchical texture patterns and subtle class boundaries. This is particularly problematic in arid cities, where bare rock, bare land, mountain slopes, and built-up surfaces may show similar spectral responses.

Recent deep-learning studies have advanced LULC classification by introducing CNN-based, attention-based, and multimodal architectures. Said et al. [23] used CNNs for desertification detection in the Makkah region and reported high performance, but the task was closer to binary environmental mapping than fine-grained multiclass urban classification. Tong et al. [24] demonstrated the usefulness of transferable deep models for high-resolution land cover classification, while Boulila et al. [25] proposed a CNN–LSTM approach for urban expansion prediction.

More recently, advanced models, such as OBIA–SVM in Alqurashi et al. [26] or self-attention CNNs in Albarakati et al. [27], have incorporated self-attention mechanisms for agricultural LULC classification, and Chen et al. [28] explored optical–SAR fusion for land cover classification. Although these studies show the growing importance of deep and attention-based models, most of them emphasize predictive accuracy without sufficiently explaining the internal reasoning behind the classification decisions. In Imam et al. [29], although these were more topologically general, they varied in their lack of deep learning refinement and per-class performance metrics, which are important for accurate urban landform monitoring.

From a comparative perspective, the reviewed literature reveals three main limitations. First, classification granularity remains limited in many studies, especially when distinguishing visually similar urban and natural subclasses. Second, interpretability is rarely integrated into the modeling pipeline; few studies combine spatial attention visualization with feature attribution analysis to explain where and why a model assigns a specific LULC class. Third, study-area specificity is often insufficient, as most models are designed for general urban, agricultural, or arid environments rather than for the sacred, densely redeveloped, and topographically uneven context of Makkah.

This study addresses these gaps by proposing a Makkah-specific hybrid U-Net–DenseNet201–ViT framework for fine-grained multiclass LULC classification. U-Net pre-

serves local boundary geometry, DenseNet201 extracts hierarchical texture semantics, and the Vision Transformer captures long-range contextual relationships across the urban–mountain mosaic. In addition, Grad-CAM is used to visualize spatial attention, while SHAP quantifies the spectral and topographic drivers of classification decisions. Thus, the proposed framework goes beyond conventional accuracy reporting by combining multiclass classification, explainable AI, and statistical validation to provide an auditable interpretation of land transformation in Makkah.

3. Materials and Methods

3.1. Study Area and Geospatial Reference Framework

Located in the western part of the Kingdom of Saudi Arabia ($21^{\circ}25'21''$ N $39^{\circ}49'24''$ E), Makkah is the holiest city in Islam and a spiritual center for more than 1.8 billion Muslims worldwide. The physiography of a city results from complex interactions between natural and human-made factors. Its landscape consists of cliffs (including the cave Hira, where it is believed that the Islamic prophet Muhammad received his first revelations), valleys, and deserts. It is centered on Al-Masjid Al-Haram, the Sacred Mosque, and is surrounded by densely built-up urban quarters characterized by intricate spatial configurations resulting from rapid urban growth. Makkah exhibits two distinct geomorphological and urban features, which complicate land cover classification, particularly in distinguishing natural rocky outcrops, vegetation, and man-made structures in satellite imagery. To address these issues, this study used high-resolution remote sensing data and classification methods tailored to the study area's particular landform [30], as shown in Figure 1.

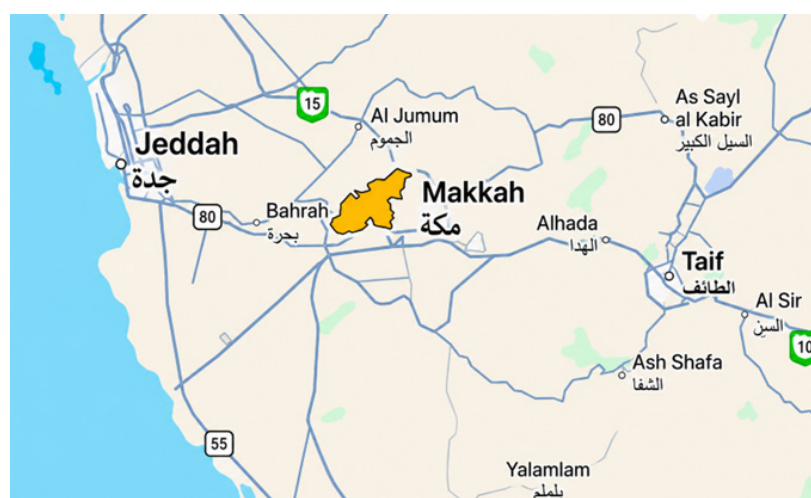


Figure 1. Geographic location and cartographic reference of the Makkah study area, Saudi Arabia. The study area is centered approximately at $21^{\circ}25'21''$ N, $39^{\circ}49'24''$ E and covers about 1200 km². The map is displayed in the WGS84 geographic coordinate system for visualization, whereas all analytical spatial layers were projected to WGS84/UTM Zone 37N. The figure includes the study-area boundary, regional inset, scale bar, north arrow, legend, and coordinate/grid reference. Basemap information was derived from OpenStreetMap contributors and Natural Earth administrative boundary layers.

All spatial datasets were projected in the World Geodetic System 1984 (WGS84) using Universal Transverse Mercator (UTM) Zone 37N to ensure spatial consistency across multi-temporal datasets. The study area covers approximately 1200 km² with an elevation of 1800 m above sea level, derived from Shuttle Radar Topographic Mission (SRTM) data. The complex topography, characterized by steep mountain areas (e.g., Jabal Al-Nour) and dense urban fabric, introduces spectral confusion between natural and anthropogenic surfaces, necessitating terrain-aware classification strategies.

For cartographic consistency, Figure 1 were prepared using the WGS84 geographic coordinate system for display and WGS84/UTM Zone 37N for spatial analysis. The approximate centroid of the study area is 21°25'21" N, 39°49'24" E. Basemap information was derived from OpenStreetMap contributors, while administrative boundaries and reference layers were obtained from municipal GIS records and supporting geographic datasets.

3.2. Remote Sensing and Geospatial Data Sources

This remote sensing and geospatial study examined environmental and urban parameters relevant to LULC classification. Therefore, LULC classification was conducted in the city of Makkah Al-Mukarramah with integrity and scientific validity. The study spans the period from 2013 to 2024 and covers an area of approximately 1200 km². This temporal range was selected because it captures a period of major urban transformation in Makkah, including rapid demographic growth, pilgrimage-related infrastructure expansion, large-scale redevelopment, and strategic urban projects associated with Vision 2030. Therefore, the 2013–2024 interval provides a suitable temporal window for assessing both gradual urban expansion and rapid land use/land cover transitions. The selection of this particular range was based on extensive urban transformations driven by considerable demographic expansion within the city, coupled with infrastructure developments for religious purposes and other strategic initiatives under Vision 2030. Specifically, the period 2013–2024 captures major infrastructure expansions associated with large-scale urban redevelopment projects and pilgrimage-related capacity enhancements. This temporal window enables the detection of both gradual urban sprawl and rapid land use transitions, which are essential for evaluating the robustness of the proposed classification framework under dynamic spatial conditions.

Multi-temporal Landsat data were obtained from the United States Geological Survey (USGS) Earth Explorer platform. Landsat 8 OLI/TIRS imagery was used for both 2013 and 2024 to ensure sensor consistency between the two dates and to avoid cross-sensor differences in spectral, spatial, and radiometric characteristics. Using the same sensor and the same WRS Path/Row improved the comparability of the multi-temporal LULC classification.

For reproducibility, the Landsat scene metadata were specified in greater detail. The 2013 classification was based on the Landsat 8 OLI/TIRS Collection 2 Level-2 Surface Reflectance product LC08_L2SP_169045_20131208_20200912_02_T1, acquired on 8 December 2013 over WRS Path/Row 169/045. The 2024 classification was based on the corresponding Landsat 8 OLI/TIRS Collection 2 Level-2 Surface Reflectance product LC08_L2SP_169045_20241206_20241210_02_T1, acquired on 6 December 2024 over the same WRS Path/Row 169/045. Both scenes were selected under cloud-free, visually clear atmospheric conditions, which improved image clarity and reduced cloud-related uncertainty during classification.

The input multispectral bands used for LULC classification have a spatial resolution of 30 m. The products used in this study are Landsat Collection 2 Level-2 Surface Reflectance science products. This processing level was selected because it provides atmospherically corrected surface reflectance, which is more suitable for multi-temporal LULC comparison than uncorrected digital numbers or top-of-atmosphere reflectance. Therefore, the subsequent preprocessing in this study focused on co-registration, masking, normalization, enhancement, spatial subsetting, and DEM-based topographic adjustment rather than replacing the USGS atmospheric correction workflow. Subsequent preprocessing, therefore, focused on co-registration, masking, normalization, enhancement, and topographic adjustment rather than replacing the USGS Level-2 correction workflow. Landsat 8 OLI/TIRS

records data with a native 12-bit radiometric resolution, while Collection 2 Level-2 products are distributed as scaled 16-bit integer data.

Figure 2 illustrates the Landsat 8 OLI spectral bands used as input variables for LULC classification. Bands 1–7 provide complementary spectral information for distinguishing different land cover types. Band 1 (coastal/aerosol) contributes to atmospheric correction and improved surface reflectance characterization. Bands 2 (blue), 3 (green), and 4 (red) represent the visible spectrum and capture variations in surface color and reflectance properties. Band 5 (near-infrared, NIR) is highly sensitive to vegetation structure and health, enabling effective discrimination between vegetated and non-vegetated areas. Bands 6 (SWIR1) and 7 (SWIR2) are particularly useful for identifying differences in soil composition, rock formations, built-up materials, and moisture conditions. The inclusion of these multispectral bands is especially important in Makkah, where urban areas, barren land, rocky mountains, transportation networks, and sparsely vegetated surfaces often exhibit similar visual characteristics. By combining visible, NIR, and SWIR information, the proposed framework can capture both spectral and spatial variability, thereby improving the reliability of LULC classification. The labeled image shown in Figure 2 represents the reference LULC map used for supervised training and performance evaluation.

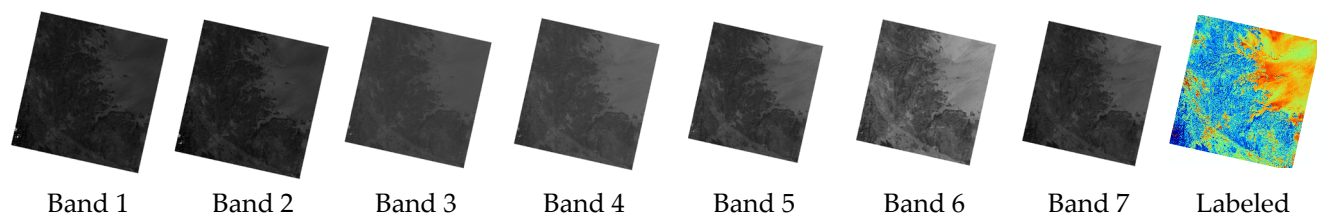


Figure 2. Landsat 8 OLI multispectral input feature stack and labeled reference layer used for supervised LULC classification in Makkah. Bands 1–7 correspond to coastal/aerosol, blue, green, red, near-infrared, SWIR1, and SWIR2, respectively. The bands were derived from USGS Landsat 8 Collection 2 Level-2 Surface Reflectance products and have a spatial resolution of 30 m. The labeled reference layer was used for model training, validation, and testing. These spectral predictors support the discrimination of built-up areas, roads, bare land, bare rock, mountainous terrain, vegetation, and agricultural surfaces.

The Landsat 8 Collection 2 Level-2 Surface Reflectance products were atmospherically corrected using the USGS Level-2 processing workflow; preprocessing in this study focused on masking, co-registration, normalization, and topographic adjustment. Both scenes were acquired in December under cloud-free conditions. Same-month acquisition was used to maintain seasonal consistency between the two study years and to reduce phenological and moisture-related variability in spectral responses. This choice is particularly important in the arid environment of Makkah, where vegetation cover is sparse and highly sensitive to short-term rainfall and seasonal fluctuations. Therefore, December scenes were preferred to minimize seasonal bias and improve the reliability of multi-temporal LULC comparison.

The images in this selection contain the suite of spectral bands in the visible, NIR, and SWIR regimes necessary to delineate built-up areas. Moreover, this dataset (DEM) was used for topographic correction, slope, and elevation to separate mountain areas (e.g., Jabal Al-Nour and Jabal Thawr) from adjacent urban and agricultural areas, utilizing Shuttle Radar Topographic Mission data. Ground-truth data was essential for both training and validating classification results. Such reference data were obtained from a field survey, supplemented by historical land use maps and municipal records from the Makkah Municipality archives. Incorporating fieldwork enabled the classification models to reflect real-world conditions and local knowledge, as in Figure 3.



Figure 3. Spatial distribution of ground-truth sample locations used for training, validation, and testing in the Makkah study area. The map is projected in WGS84/UTM Zone 37N and includes the study-area boundary, scale bar, north arrow, legend, and coordinate/grid reference. Ground-truth samples were derived from field surveys, historical land use maps, and Makkah Municipality records. Basemap layers were obtained from OpenStreetMap contributors, and all spatial layers were co-registered to the same coordinate reference system before analysis.

Table 1 shows the distribution of samples used for training, validation, and testing for the land classification categories for Makkah city. The number of samples to be collected was determined based on the actual area in 2013 and 2024. The samples were divided as follows: 64% for training, 16% for verification, and 20% for testing, to ensure the reliability of the results. The total sample size for the two periods (2013 and 2024) is approximately 6250 pixels, with an average of 900 m² per pixel.

Table 1. Distribution of labeled reference samples used for training (64%), validation (16%), and testing (20%) across the nine LULC classes for Makkah in 2013 and 2024. Values represent the number of labeled Landsat 8 pixels assigned to each class and dataset subset. Each pixel has a spatial resolution of 30 m × 30 m, equivalent to approximately 900 m².

Period Class	2013				2024			
	Testing	Train	Val	2013 Total	Testing	Train	Val	2024 Total
Bare Land	500	1600	400	2500	396	1266	317	1979
Bare Rock	313	1000	249	1562	250	800	200	1250
Modern Residential	188	600	150	938	250	800	200	1250
Roads	63	200	49	312	125	400	100	625
Mountain Areas	63	200	49	312	54	174	43	271
Industrial Facilities	38	120	30	188	63	200	49	312
Vegetation	12	40	11	63	63	200	49	312
Old Buildings	50	160	40	250	38	120	30	188
Agriculture	25	80	20	125	13	40	10	63
Total	1252	4000	998	6250	1252	4000	998	6250

Table 2 provides a clear and concise mapping between visible colors in satellite imagery, their corresponding spectral or geographic interpretations, and the proposed LULC classifications for the study area. Each color is associated with specific urban or natural features, enabling more accurate visual identification and classification of land cover types. The distinct urban built-up categories—standard residential, modern residential, and industrial/commercial—indicate the model’s responsiveness to urban diversity. The differ-

entiation of natural features (bare rock, bare land, and sparse vegetation) enabled a higher level of environmental analysis. In addition, the color scheme enhances the interpretability of the classification results and improves the overall clarity of LULC differentiation.

Table 2. Spectral and visual interpretation of LULC classes based on Landsat 8 visible, NIR, and SWIR band combinations. The table links apparent image color, spectral/geographic interpretation, and final LULC class labels to support transparent visual interpretation and classification consistency. The table is used as an interpretation guide and does not replace the supervised classification procedure.

Classes	Visible Color	Spectral/Geographic Interpretation	Proposed LULC Classification
Modern residential	Light blue/bluish/cyan	Urban areas (concrete surfaces or standard residential buildings).	Built-Up_1 (modern built-up residential surfaces)
Roads	Bright linear features/high-albedo corridors	High-density urban zones, newly constructed or modern residential.	Built-Up_2 (transport corridors and road networks)
Industrial facilities	Bright red (concentrated SW)	Industrial facilities, metal rooftops, or newly developed urban zones.	Built-Up_3 (industrial/commercial)
Mountain areas	Dark brown/gray rugged terrain	Rugged mountainous terrain, highlands, or exposed rocky areas.	Bare rock (steep mountainous terrain)
Bare land	White/very light gray	Bare land, sabkhas (salt flats), or flat saline deposits.	Bare land
Vegetation	Managed green cover	Major roads or road networks.	Natural or managed green cover
Consistent agriculture	Very few dark green patches	Possibly weak vegetation or limited agricultural activity.	Sparse vegetation/agriculture

The data-selection strategy was designed to ensure temporal comparability, spatial consistency, and reproducibility. Landsat 8 OLI/TIRS imagery was used for both 2013 and 2024 to maintain sensor consistency and avoid cross-sensor differences. The same WRS Path/Row was used for both dates to ensure comparable spatial coverage. December scenes were selected because they were cloud-free and acquired in the same season, reducing phenological and moisture-related variability in the arid environment of Makkah. Landsat Collection 2 Level-2 Surface Reflectance products were selected because they provide atmospherically corrected reflectance values suitable for multi-temporal LULC analysis. The 30 m multispectral bands were used because they provide consistent coverage across the full study area and include visible, NIR, and SWIR information required to distinguish built-up areas, roads, bare land, bare rock, mountainous terrain, vegetation, and agriculture. SRTM DEM data were incorporated to derive elevation and slope information, which is necessary for separating mountainous terrain from spectrally similar urban and bare surfaces.

3.3. Spatial Preprocessing, Co-Registration, and Topographic Normalization

All satellite imagery was preprocessed to maximize spatial, radiometric, and geometric consistency prior to classification. The Landsat 8 OLI/TIRS images used in this study were obtained as USGS Collection 2 Level-2 Surface Reflectance products. Therefore, atmospheric correction from Top-of-Atmosphere reflectance to Surface Reflectance had already been performed within the USGS Level-2 processing chain. No additional author-side conversion from raw digital numbers or top-of-atmosphere reflectance to surface reflectance was conducted using metadata calibration coefficients [31]. The preprocessing carried out in this study was limited to applying the official Level-2 scale factors to the delivered scaled

integer values, QA-based masking of cloud, shadow, and low-quality pixels, co-registration checking between the 2013 and 2024 scenes, band stacking, normalization, spatial subsetting to the Makkah administrative boundary, and DEM-based terrain-related normalization and feature derivation. All datasets were projected to a common coordinate reference system, WGS84/UTM Zone 37N, to ensure spatial consistency across the multi-temporal analysis. The spatial subsetting step clipped the imagery to the legal administrative boundary of Makkah City (1200 km²), as defined by GIS shapefiles (Figure 4). This spatial subsetting was used to constrain the analysis area of interest and reduce computational costs. Band Composition and Normalization: Multispectral bands were stacked and normalized to produce input datasets that met the deep learning algorithm's requirements. This guarantees that the feature representations are invariant and can facilitate convergence during model training [32].

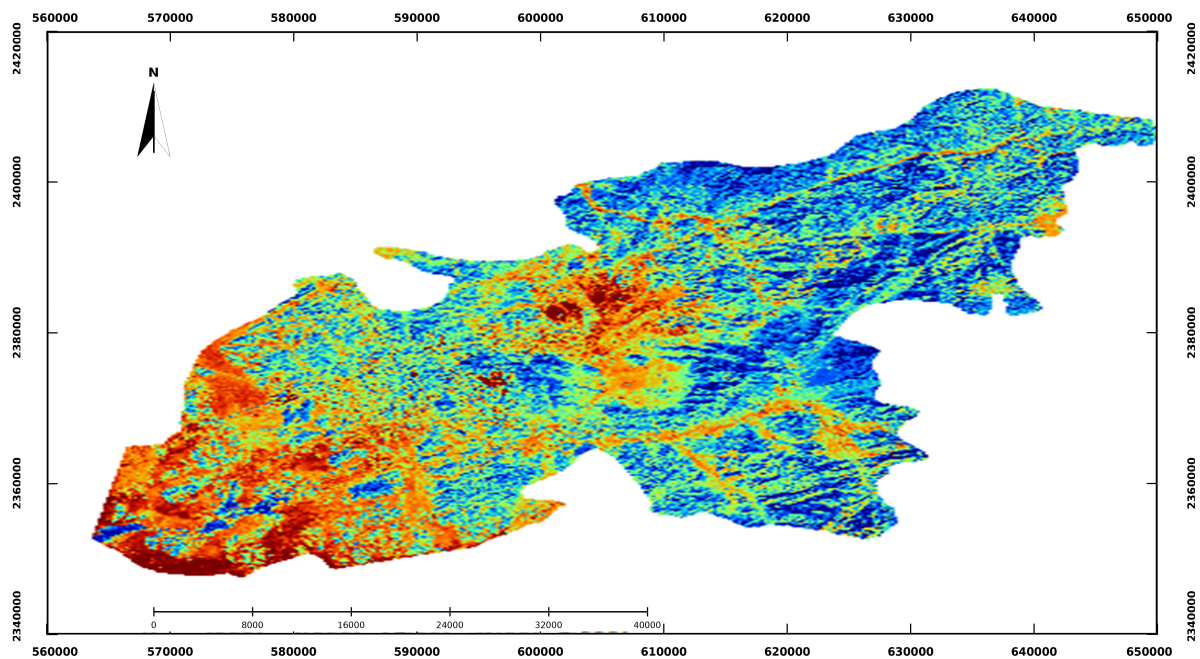


Figure 4. Spatial subsetting of Landsat 8 imagery to the administrative boundary of Makkah City. The clipping process restricted the analysis to the approximately 1200 km² study area, reduced computational complexity, and ensured that all LULC statistics were calculated within a consistent spatial boundary. Coordinate reference system: WGS84/UTM Zone 37N. Area unit: km².

Topographic normalization was additionally applied using DEM-derived slope and aspect to reduce illumination effects caused by terrain variability. This correction is particularly important in mountain areas, where shadowing and slope-induced reflectance distortions can significantly affect classification accuracy.

For image enhancement techniques, the satellite images used in the analysis via CNN-based methods were enhanced. We used a Scalable Linear Contrast enhancement technique to enhance the visibility and interpretability of surface features; we used the Scalable Linear Contrast (SLC) enhancement technique. It also spreads pixel values across its full dynamic range and preserves relative contrast between neighboring pixels, thereby enhancing small features such as narrow roads, vegetation boundaries, and construction sites without distorting or clipping the original radiometric characteristics [33]. For labeling and training dataset preparation, the training datasets were prepared through a detailed manual interpretation of high-resolution imagery overlaid with GIS layers [34]. Each pixel was classified into one of the following land cover classes: agriculture, bare land,

bare rock, industrial facilities, modern residential, mountain areas, old buildings, roads, and vegetation.

Before applying the Laplacian operator, the Landsat surface reflectance bands were first subjected to quality masking using the QA layers to remove cloud, shadow, and low-quality pixels. A light 3×3 median filtering step was then applied to suppress isolated high-frequency noise while preserving major land cover boundaries. The Laplacian filter was subsequently used only for edge enhancement, not for denoising. The original spectral bands were retained in the classification input to avoid over-amplifying artificial edges.

3.4. Laplacian Filter for Edge Enhancement in Preprocessing

A Laplacian filter was applied to enhance edge visibility and fine spatial details in satellite images as a preprocessing step prior to feature extraction and classification. A Laplacian filter is used for edge detection, which focuses on rapid intensity changes and computes the second derivative along the spatial dimensions of an image [35]. Its primary role here is to underline structural boundaries such as roads, building ends, and natural terrain transitions that are key to correctly delineating LULC in extremely heterogeneous urban landscapes such as Makkah [36]. Given the spectral diversity and morphological heterogeneity of the Makkah region, which comprises dense residential areas, clustered industries, rocky mountain areas, and fragmented vegetation, a Laplacian filter was applied to enhance contrast at boundaries between land cover types. The addition of a Laplacian to the filter bank enhanced spatial content and effectively categorized LULC patterns within the study area over the study period [37]. Figure 5 Comparative mapping of Makkah before and after enhancement. The post-enhanced result illustrates improved boundary contrast after prior quality masking and light denoising, while the Laplacian operator itself is used only to emphasize edges. The Laplacian filter improved the quality of the input spatial data by highlighting small textural differences and edge-based information, thereby enabling efficient feature learning in the hybrid U-Net, DenseNet201, and ViT models. This was particularly important for reducing confusion in classes about visually similar features, such as between bare land and built-up areas or mountain ridges and rocky outcrops.

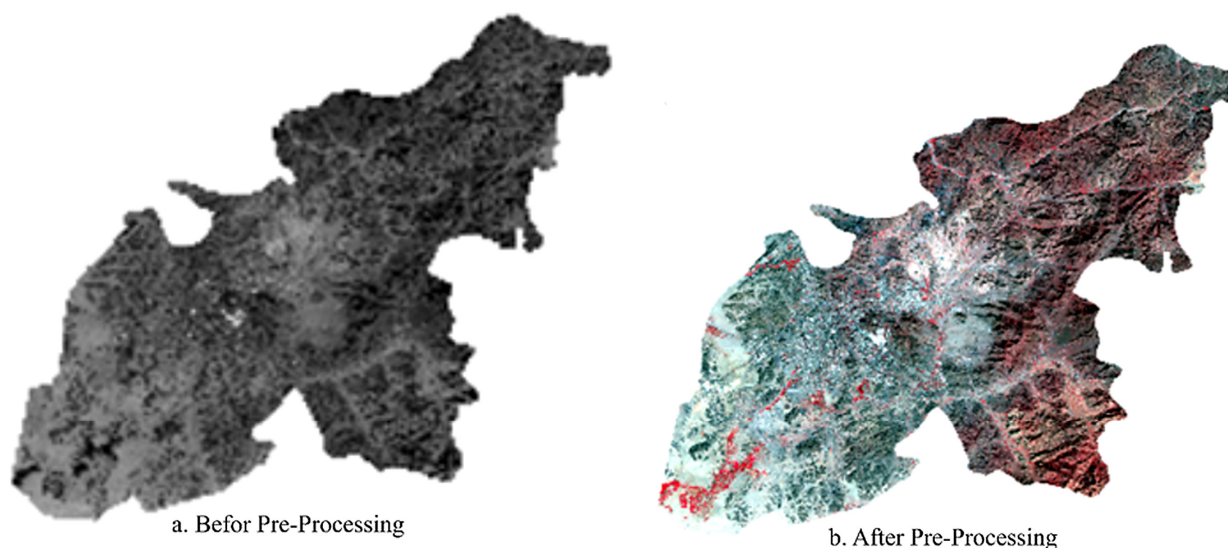


Figure 5. Comparison of Landsat 8 imagery before (a) and after (b) preprocessing using Laplacian edge enhancement. The enhancement was applied after quality masking and light median filtering to emphasize structural boundaries among roads, built-up areas, bare land, and mountainous terrain. The Laplacian operator was used for edge enhancement rather than denoising. Spatial resolution: 30 m; coordinate reference system: WGS84/UTM Zone 37N.

3.5. Training and Validation Sample Quality Assessment

Prior to model training, the quality of the training and validation samples was assessed to ensure the internal consistency of the reference data, their spectral significance, and adequate separation between land use and coverage categories. This step was necessary because the Makkah study area contains a highly variable urban–mountain mosaic, in which several categories might exhibit similar spectral responses at a spatial resolution of 30 m.

First, the homogeneity of the categories was assessed using descriptive statistics for spectral variables and variables derived from the selected topography. For each land use and coverage category, the mean, standard deviation, and coefficient of variation were calculated across the training and validation samples. Categories with low intracategory standard deviations were considered more homogeneous, while samples exhibiting unusually high dispersion were revised to mitigate the effects of mixed pixels, misclassified boundaries, and transition zones. This procedure was particularly important for categories located along rocky urban frontages, road edges, and heterogeneous redevelopment areas.

Second, a pre-training separability assessment was conducted to evaluate whether the selected reference samples could be clearly distinguished in the feature space. Spectrally similar categories were examined, including barren land versus barren rock, barren rock versus industrial structures, modern residential buildings versus older buildings, roads versus barren land, and vegetation versus farmland. This analysis helped verify that the training and validation samples represented distinct categorical fingerprints before being used in supervised classification. Where any overlap was observed, the samples were visually reviewed using multispectral composite imagery, slope/elevation layers derived from a digital elevation model, and reference GIS information.

Third, the band selection strategy was based on the physical significance of each predictor in characterizing land use and cover in an arid urban environment. Landsat 8 visible bands were used to capture surface brightness and color variations; the near-infrared band was included due to its sensitivity to vegetation structure and photosynthesis. The SWIR1 and SWIR2 bands were selected for their high responsiveness to built materials, exposed soil, dry rock, and moisture variations. Slope and elevation data derived from a digital elevation model (DEM) were also incorporated to differentiate mountainous terrain and bare rock from urban and bare surfaces with similar spectral profiles. Accordingly, the final training profile was selected based on both spectral separability and physical interpretability, rather than random band selection.

This sample quality assessment improved the reliability of the training and validation datasets by reducing classification uncertainty, identifying mixed-pixel issues, and ensuring that the selected bands and auxiliary terrain variables were suitable for distinguishing the nine land use categories in Makkah.

3.6. Data Augmentation Techniques

The limited number of training samples in some LULC classes may reduce classification precision and increase the risk of overfitting during training, particularly in imbalanced datasets. Training patches were harvested from the Makkah City classification map. The labeled samples reported in Table 1 were used as reference sample locations, and image patches were extracted around these labeled pixels from the multispectral Landsat stack and the corresponding LULC label map. Therefore, data augmentation was applied to spatial image patches rather than to isolated single-pixel numerical records [38].

Data augmentation was applied only to the training dataset and not to the validation or test sets. The augmentation strategy served two purposes: increasing the effective diversity of training patches and reducing class imbalance among the nine LULC classes [39]. The

augmentation intensity was class-dependent. Classes with fewer original training samples received more augmented patch variants, whereas classes with sufficient training samples received fewer transformations or were left unchanged. The augmentation operations and parameter types included rotation, horizontal and vertical flipping, cropping/zooming, and spatial translation. For each operation, the same geometric transformation was applied to both the input image patch and its corresponding label patch/mask to preserve spatial alignment and semantic consistency. These transformations preserved the class information while exposing the model to variations in orientation, position, scale, and local spatial context.

Table 3 presents the impact of data augmentation techniques applied to balance the training dataset for land cover classes in Makkah for 2013 and 2024. The table shows the number of training samples per class before and after augmentation. In 2013, the original dataset was highly imbalanced across classes. For example, bare land included 1600 training samples, whereas vegetation included only 40. To mitigate this imbalance, minority classes received more augmented patch variants until the training distribution reached 1600 samples per class. This helped ensure that all land cover types were adequately represented during model training and reduced the risk of model bias toward dominant classes.

Table 3. Effect of data augmentation on class balancing for LULC classification. The table compares the number of training samples before and after augmentation for 2013 and 2024, demonstrating improved class balance and mitigation of imbalance-related bias.

Class	2013		2024	
	Training (Before Aug)	Training (After Aug)	Training (Before Aug)	Training (After Aug)
Bare Land	1600	1600	1266	1266
Bare Rock	1000	1600	800	1200
Modern Residential	600	1600	800	1200
Roads	200	1600	400	1200
Mountain Areas	200	1600	174	1200
Industrial Facilities	120	1600	200	1200
Vegetation	40	1600	200	1200
Old Buildings	160	1600	120	1200
Agriculture	80	1600	40	1200
Total	4000	14,400	4000	10,800

For the 2024 period, a similar class-dependent augmentation approach was adopted. However, because the maximum number of original training samples was associated with bare land (1266 samples), the remaining classes were augmented toward a near-balanced training distribution of approximately 1200 samples per class. The balancing approach resulted in 14,400 training samples for 2013 and 10,800 training samples for 2024. Through this approach, the training dataset became more representative of minority classes, improving the model’s ability to learn underrepresented LULC categories and reducing issues related to class dominance, overfitting, and underfitting.

3.7. Hybrid U-Net-DenseNet201-ViT Model

A heterogeneous environment in Makkah can be precisely mapped only with a multi-class land use/land cover (LULC) model that simultaneously ensures pixel-level boundary accuracy and abstract contextual interpretations of the scene across rugged relief and sizable urban mosaics [40]. To meet both needs, we present a hybrid deep-learning model that combines two computationally efficient local feature extractors, an enhanced U-Net, and a streamlined DenseNet201, with a low-depth ViT as a global dependency model.

The structure is not a straightforward combination; it is an orchestrated one in which geometry-preserving (U-Net), texture-semantic (DenseNet201), and long-range contextual (ViT) features are jointly represented as a single representation for pixel-wise class assignment [41].

3.7.1. Improved U-Net for Local Feature Extraction

The U-Net branch provides spatial support to preserve sharp edges (e.g., road, building, and urban rock edges) [42]. To evaluate the effect of U-Net encoder depth, an ablation analysis was added comparing three- and four-downsampling-block configurations under identical training, validation, augmentation, and optimization settings. The three-block configuration was retained only as an efficiency-oriented choice when it achieved comparable class-level performance to the deeper variant while reducing computational cost [43]. Therefore, the revised manuscript does not assume that a shallower U-Net is intrinsically superior; rather, the selected depth is justified through validation-based comparison, as in Equation (1):

$$F_l = \sigma(W_l * F_{l-1} + b_l), \quad (1)$$

where $*$ denotes convolution, $\sigma(\cdot)$ is ReLU, and F_l is the output feature map. Reducing the number of convolutional blocks from L to L' ($L' < L$) lowers the dominant complexity term as in Equation (2):

$$O(L \cdot k^2 C^2 HW) \rightarrow O(L' \cdot k^2 C^2 HW), \quad (2)$$

with kernel size k , channel width C , and spatial size $H \times W$. In the decoder, resolution is recovered via up-sampling and skip fusion as in Equation (3):

$$F_{\text{dec}}^i = \text{Up}(F_{\text{dec}}^{i+1}) \oplus F_{\text{enc}}^i, \quad (3)$$

where $\oplus$ denotes concatenation. This preserves the localization of the U-Net boundaries while reducing parameter and runtime costs, which are of great importance in processing multi-temporal city-level applications [44].

3.7.2. Optimized DenseNet201 for Dense Texture–Semantic Encoding

DenseNet201 was employed to enhance discrimination among spectrally similar surfaces (e.g., industrial roofs, modern residential, and bare rock) via dense connectivity and [45] feature reuse as in Equation (4):

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]), \quad (4)$$

where $[\cdot]$ denotes channel-wise concatenation, and $H_l(\cdot)$ is a composite of BN–ReLU–Conv. This connectivity improves the gradient flow and reduces redundancy by reusing earlier representations [46].

To reduce computation while preserving the dense-connectivity principle, the depth of the densest block is compressed (e.g., from n layers to n' layers, such as $48 \rightarrow 24$). The resulting computation reduction can be expressed as in Equation (5):

$$O(n \cdot k^2 CHW) \rightarrow O(n' \cdot k^2 CHW) (n' < n), \quad (5)$$

This optimization reduces the most expensive stage of DenseNet201 while maintaining the hierarchical semantics required for robust multiclass separation in arid, high-albedo urban settings.

3.7.3. Enhanced ViT for Global Contextual Modeling

Local CNN features alone can remain ambiguous when distinct classes share similar reflectance patterns but differ in their spatial context [47]. To address this, ViT models long-range dependencies by converting images into sequences of patch tokens [48]. For an input $X \in \mathbb{R}^{H \times W \times C}$ is partitioned into N non-overlapping patches of size $P \times P$, where $N = HW/P^2$. Each patch is flattened and linearly projected into a token as in Equation (6):

$$z_0^i = E \cdot \text{Flatten}(X^i) + E_{\text{pos}}^i, \quad (6)$$

where E is the patch-embedding matrix, and E_{pos}^i is the positional embedding for patch i .

To balance accuracy and efficiency, the transformer encoder depth was reduced to 8 transformer blocks. While maintaining the number of attention heads (h) was expanded to preserve representational diversity [49]. Multi-headed self-attention is computed using the standard query projections as in Equations (7) and (8):

$$Q = ZW_Q, K = ZW_K, V = ZW_V, \quad (7)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V. \quad (8)$$

Increasing h compensates for the reduced depth by enabling parallel attention across multiple subspaces, thereby preserving global reasoning capacity across Makkah's complex urban–mountain structure.

3.7.4. Feature Fusion and Pixel-Level Classification

Feature maps from the three branches—U-Net F_{UNet} , DenseNet201 F_{Dense} , and ViT F_{ViT} —are spatially aligned and concatenated as in Equation (9):

$$F_{\text{fusion}} = [F_{\text{UNet}} \oplus F_{\text{Dense}} \oplus F_{\text{ViT}}]. \quad (9)$$

After integrating the features of the U-Net-DenseNet201-ViT model, the PCA algorithm was applied to select the most important features and remove redundant and irrelevant ones. The key components that explain the large variance in the initial feature space were selected for further investigation [50]. This approach was found to significantly reduce computational complexity and improve classification using optimal spectral characteristics [51]. This facilitated the simultaneous optimization of spatial, textural, and spectral information for distinguishing complex urban and natural patterns within the study area. The important integrated features were then submitted to the classification layer [52].

This fused tensor is processed by a lightweight classification head, adapted from the ViT architecture, consisting of Layer Normalization, flattening, dropout, and a Multi-Layer Perceptron (MLP) with one hidden layer and GELU activation, followed by a final dense layer [31]. A softmax function subsequently assigns a discrete probability distribution over the nine LULC classes to every pixel as in Equation (10):

$$P(y = c \mid x) = \frac{e^{z_c}}{\sum_{j=1}^9 e^{z_j}}, \quad (10)$$

where $z = WF_{\text{fusion}} + b$ is the logit vector.

The model was optimized via end-to-end supervised learning with learning-rate scheduling, early stopping, and augmentation (translation, flipping, zooming, and rotation) to improve generalization to the changing landscapes of Makkah. This integrated approach enables the abstraction of complex urban patterns while accurately delineating class bound-

aries to develop a classification model well adapted to Makkah’s complex geographic and infrastructural landscape [53]. This results in a substantial improvement in identifying and interpreting urban transformation patterns, thereby enabling more reliable land change detection for the decade under review (Figure 6).

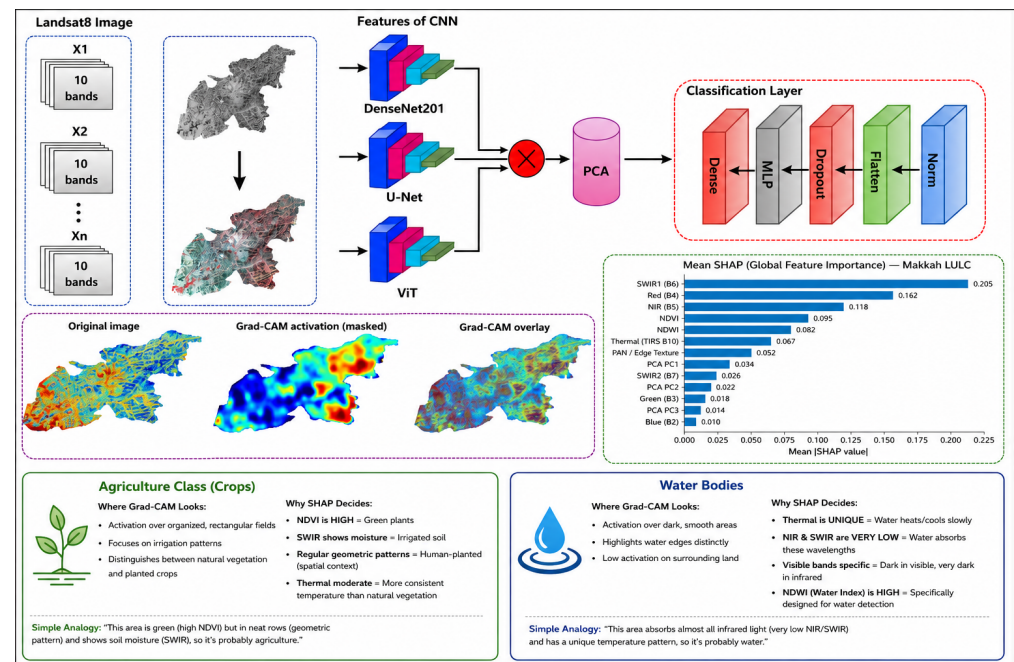


Figure 6. Structure of the proposed hybrid interpretable U-Net–DenseNet201–ViT framework for land use classification and coverage.

4. Results of Experiment

4.1. General Results

The comparative LULC analysis for the Makkah region, covering approximately 1,200,000,000 m², revealed notable spatial differences across land categories between 2013 and 2024. Bare land remained the dominant class, although it decreased from 480,000,000 m² in 2013 to 380,000,000 m² in 2024, a reduction of 100,000,000 m² (8.33 percentage points), as shown in Table 4. Bare Rock also declined from 300,000,000 m² to 240,000,000 m², corresponding to a decrease of 60,000,000 m², or −5.00 percentage points.

Table 4. Classified LULC area distribution for Makkah in 2013 using the proposed U-Net–DenseNet201–ViT model. Area values are reported in square meters (m²), and percentages are calculated relative to the total study area of approximately 1200 km².

Category	Area (m ²)	Percentage	Center	Cumulative
Agriculture	24,000,000	2.00%	26,667	26,667
Bare Land	480,000,000	40.00%	533,333	560,000
Bare Rock	300,000,000	25.00%	333,333	893,333
Industrial Facilities	36,000,000	3.00%	40,000	933,333
Modern Residential	180,000,000	15.00%	200,000	1,133,333
Mountain Areas	60,000,000	5.00%	66,667	1,200,000
Old Buildings	48,000,000	4.00%	53,333	1,253,333
Roads	60,000,000	5.00%	66,667	1,320,000
Vegetation	12,000,000	1.00%	13,333	1,333,333

Urban and infrastructure-related classes expanded during the same period. Modern residential areas increased from 180,000,000 m² in 2013 to 240,000,000 m² in 2024, rep-

representing an absolute increase of 60,000,000 m², a +5.00 percentage-point change, and a relative increase of 33.33%. Roads doubled from 60,000,000 m² to 120,000,000 m², corresponding to a +60,000,000 m² absolute increase, a +5.00 percentage-point change, and a 100.00% relative increase. Industrial facilities increased from 36,000,000 m² to 60,000,000 m², corresponding to an absolute increase of 24,000,000 m², a +2.00 percentage-point change, and a 66.67% relative increase.

Other land cover classes showed contrasting trends. Vegetation expanded from 12,000,000 m² to 60,000,000 m², representing a +48,000,000 m² increase and a +4.00 percentage-point change, while agriculture declined from 24,000,000 m² to 12,000,000 m². Mountain areas decreased slightly from 60,000,000 m² to 52,000,000 m², and old buildings declined from 48,000,000 m² to 36,000,000 m². These changes indicate infrastructure-driven urban expansion, accompanied by a reduction in bare surfaces and agricultural land, together with an increase in possible policy-related greening.

To clarify the percentage-point accounting between 2013 and 2024, the total gain in expanding classes was +16.00 percentage points: modern residential increased by +5.00 pp, roads by +5.00 pp, vegetation by +4.00 pp, and industrial facilities by +2.00 pp. These gains were offset by total losses of −16.00 percentage points across the contracting classes. Bare land declined by −8.33 pp and bare rock by −5.00 pp, while the remaining −2.67 pp came from agriculture (−1.00 pp), mountain areas (−0.67 pp), and old buildings (−1.00 pp). Thus, the revised interpretation avoids the phrase “perfect balance” and instead describes a complete percentage-point redistribution across all changing LULC categories.

The 2024 LULC classification for the Makkah region, covering approximately 1,200,000,000 m², revealed notable spatial differences among land cover categories. Bare land remained the largest class, accounting for 380,000,000 m², or 31.67% of the total study area. Bare rock and modern residential areas each accounted for 240,000,000 m², representing 20.00% of the study area for each class separately. Together, these two classes covered 480,000,000 m², or 40.00% of the total area, reflecting the coexistence of extensive rocky terrain and substantial urban development. Roads covered 120,000,000 m², corresponding to 10.00% of the total area, while industrial facilities and vegetation each occupied 60,000,000 m², or 5.00%. Old buildings accounted for 36,000,000 m², or 3.00%, whereas agriculture represented the smallest class, with 12,000,000 m², or 1.00% of the total area. These results indicate continued urbanization and infrastructure expansion in Makkah, alongside the persistent dominance of bare and rocky surfaces (Table 5).

Table 5. Classified LULC area distribution for Makkah in 2024 using the proposed U-Net-DenseNet201-ViT model. Area values are reported in square meters (m²), and percentages are calculated relative to the total study area of approximately 1200 km².

Category	Area (m ²)	Percentage	Center	Cumulative
Agriculture	12,000,000	1.00%	13,333	13,333
Bare Land	380,000,000	31.67%	422,222	435,555
Bare Rock	240,000,000	20.00%	266,667	702,222
Industrial Facilities	60,000,000	5.00%	66,667	768,889
Modern Residential	240,000,000	20.00%	266,667	1,035,556
Mountain Areas	52,000,000	4.33%	57,777	1,093,333
Old Buildings	36,000,000	3.00%	40,000	1,133,333
Roads	120,000,000	10.00%	133,333	1,266,666
Vegetation	60,000,000	5.00%	66,667	1,333,333

4.2. Projection of Land Use and Land Cover Trends to 2030

The quantitative measure of LULC changes for 2013–2024 provides an empirical basis for assessing the dynamic urban processes at work in Makkah. Generalizing this analysis and predicting future states is not merely an academic exercise; it is central to planning. The annual change rates for 2013–2024 were applied to develop a science-based scenario for the expected LULC pattern in 2030.

The projected area for each LULC category in 2030 was calculated by linearly extrapolating the observed annual change rate from 2013 to 2024. For each LULC class, the annual change rate was calculated as follows:

$$\text{Annual Change Rate} = (\text{Area in 2024} - \text{Area in 2013})/11$$

The projected area for 2030 was then calculated as follows:

$$\text{Projected Area in 2030} = \text{Area in 2024} + (\text{Annual Change Rate} \times 6)$$

where 11 represents the number of years between 2013 and 2024, and 6 represents the number of years between 2024 and 2030. This calculation provides a continuation-of-trends scenario based on the observed class-level changes during the study period. The total study area was held constant at 1200 km² to ensure mass balance. The “Normalized % (2030)” represents the projected percentage of the total area for each category. The “Change vs. 2024” column indicates the absolute change in square kilometers from the 2024 baseline to the 2030 projection. The results are synthesized in Table 6.

Table 6. Projected land use and land cover changes in Makkah by 2030.

Category	Area 2013 (km ²)	Area 2024 (km ²)	Annual Change Rate (km ² /yr)	Predicted 2030 (km ²)	Normalized % (2030)	Change vs. 2024 (km ²)	Trend (2013–2024)
Bare Land	480	380	−9.09	325.5	27.10%	−54.5	▼ Steady Loss
Bare Rock	300	240	−5.45	207.3	17.30%	−32.7	▼ Steady Loss
Modern Residential	180	240	5.45	272.7	22.70%	32.7	▲ Strong Growth
Roads	60	120	5.45	152.7	12.70%	32.7	▲ Strong Growth
Industrial Facilities	36	60	2.18	73.1	6.10%	13.1	▲ Growth
Vegetation	12	60	4.36	86.2	7.20%	26.2	▲ Rapid Growth
Mountain Areas	60	52	−0.73	47.6	4.00%	−4.4	▼ Slight Loss
Old Buildings	48	36	−1.09	29.5	2.50%	−6.5	▼ Decline
Agriculture	24	12	−1.09	5.5	0.50%	−6.5	▼ Decline
Total	1200	1200	—	1200	100.00%	0	—

The 2030 estimates should be interpreted as a linear scenario projection rather than a deterministic forecast. The projection assumes that the annual class-level change rates observed between 2013 and 2024 continue unchanged until 2030, while the total study area remains constant. To reflect uncertainty, the projection was accompanied by an uncertainty envelope derived from classification uncertainty and the limited two-date temporal baseline. Therefore, the projected values are intended to indicate directional tendencies rather than exact future land cover quantities.

The 2030 projection should be interpreted as a trend-based planning scenario derived from the observed LULC changes between 2013 and 2024, a temporal interval of eleven years. This period provides a meaningful basis for estimating the likely direction of future land cover change in Makkah. Nevertheless, as with any scenario-based projection, the estimates are influenced by the assumption that the observed annual change rates continue in a broadly similar direction until 2030. Future development priorities, infrastructure plan-

ning decisions, and environmental management policies may further shape the magnitude and spatial distribution of these changes. Therefore, the 2030 results are best understood as evidence-based planning indicators that describe likely directional tendencies rather than fixed deterministic values.

Based on linear trend extrapolation for 2013–2024, the LULC landscape of Makkah in 2030 is likely to undergo significant changes, indicating intensified and managed urban expansion.

Urbanization and Infrastructure Investment: The most significant future trend will be the continued, substantial expansion of the built environment, including modern residential, which is expected to occupy an additional 32.7 km², bringing the total to 272.7 km² (22.7% of the total area). This is the only land cover type that, in 2030, will be larger in area than in 2024. (the other land cover types will either decrease or remain stable), and thus is likely to be modern residential becomes the largest built class, though bare land remains the single largest class overall. This pattern clearly indicates the ongoing requirement for residential, commercial, and pilgrimage-related infrastructure. Simultaneously, roads will expand by an identical absolute amount (+32.7 km²), thereby reinforcing the enabling role of road networks in urban expansion. Additionally, industrial facilities are likely also to expand (+13.1 km²), which signals the primary stage of development of the diversified urban economic base. The continued and simultaneous growth of all three land cover classes (buildings, roads, and industrial facilities) in Makkah suggests that it has entered an advanced stage of urban-related, intense land cover conversion, in which urban land cover arrangements tend to be more congested and interconnected.

Continued Conversion of Natural Land Cover: Observably, the growth of the built environment will directly correspond to and be equally balanced by the continued conversion of natural and semi-natural land covers. Bare land is projected to be the largest source of converted land, decreasing by 54.5 km² to 325.5 km² (27.1%). Bare rock is also projected to decline significantly by 32.7 km². This systematic conversion of bare substrates is the primary mechanism fueling urban expansion, as identified in the socioeconomic correlates. The stability of mountain areas, with only a minor projected loss, continues to act as a geomorphic constraint, channeling development onto more viable terrain.

Significant Increase in Vegetated Areas: A critically important and positive trend is the projected rapid growth of vegetation, which is expected to increase by 26.2 km² to 86.2 km² (7.2%). This represents a more than seven-fold increase from the 2013 extent. This trend strongly supports the inference from the 2013–2024 analysis that this increase is due to deliberate urban greening initiatives, such as parks, green belts, and landscaping integrated into new developments. This reflects a conscious policy shift towards enhancing urban livability and microclimatic conditions amid intensive development.

Marginalization of Old Buildings and Agriculture: The decline in old buildings and agriculture is projected to continue. The loss of older urban fabric is consistent with high-density urban renewal and redevelopment. The near-disappearance of agriculture (projected to account for only 0.5% of the total area) underscores its non-viability within the competitive land economics of a rapidly expanding global city and a holy site.

To improve the transparency of the 2030 projection, the key assumptions underlying the scenario are summarized in Panel 1. The projection was developed as a scenario-based continuation of observed LULC trends rather than as a deterministic forecast. It assumes that the annual class-level change rates observed between 2013 and 2024 continue linearly until 2030, while the total study area remains constant at approximately 1200 km². Therefore, the projected values should be interpreted as planning-oriented directional indicators rather than exact future land cover quantities.

Panel 1. Assumptions of the 2030 Projection

- Scenario type:** Linear continuation-of-trends scenario, not a deterministic forecast.
- Temporal baseline:** Observed LULC changes from 2013 to 2024.
- Baseline year:** 2024 classified LULC map.
- Projection horizon:** Six years, from 2024 to 2030.
- Area constraint:** Total study area fixed at approximately 1200 km².
- Class consistency:** Same LULC classes, spatial resolution, and classification scheme.
- Excluded shocks:** No abrupt policy shifts, major environmental disturbances, or exceptional development shocks were explicitly modeled.
- Interpretation:** Values indicate directional planning tendencies, not exact future quantities.

4.3. Overall Accuracy and Kappa Coefficient

To assess the performance of the evaluated deep learning-based classification strategies for LULC mapping in Makkah Al-Mukarramah, overall accuracy (OA) and Cohen’s Kappa coefficient (κ) were used as the main accuracy metrics. Cohen’s Kappa was calculated as $\kappa = (P_o - P_e)/(1 - P_e)$, where P_o is the observed agreement, and P_e is the agreement expected by chance. For the proposed U-Net–DenseNet201–ViT model, the reported OA and κ values were harmonized with the year-specific confusion matrices.

As shown in Table 7, the proposed hybrid approach achieved the highest performance among the evaluated methods. Based on the 2013 confusion matrix, the model achieved an OA of 98.32% and $\kappa = 0.978$. For 2024, it achieved an OA of 98.48% and $\kappa = 0.981$. These results indicate a very high level of agreement between the classified outputs and the reference data, confirming the robustness of the proposed hybrid model across both study years.

Table 7. Overall accuracy and kappa coefficient for classification in 2013 and 2024.

Method	2013 Overall Accuracy (%)	2013 Kappa Coefficient	2024 Overall Accuracy (%)	2024 Kappa Coefficient
The proposed hybrid approach (U-Net and DenseNet201 with ViT)	98.32	0.978	98.48	0.981
The ResNet-50 and U-Net with ANN	89	0.855	92.32	0.865
The U-Net and DenseNet201 with ANN	91.98	0.963	94.22	0.866
The ResNet-50 and DenseNet201 with ANN	93.69	0.963	88.32	0.897

Compared with the alternative models, the proposed U-Net–DenseNet201–ViT framework achieved superior performance by combining boundary-preserving spatial features, dense texture–semantic representation, and long-range contextual reasoning. This confirms its suitability for LULC classification in Makkah’s complex urban–mountain environment.

4.4. Confusion Matrix Analysis of the Proposed Systems

Table 8 presents the confusion matrix generated by the proposed U-Net–DenseNet201–ViT hybrid model for the 2013 classification. The diagonal elements represent the correctly classified samples for each LULC class. Overall, 1231 out of 1252 test samples were correctly classified, yielding an overall accuracy of 98.32%.

Table 8. Confusion matrix generated by the U-Net-DenseNet201-ViT hybrid model of classification of land changes in Makkah in 2013.

Actual \ Predicted	Bare Land	Bare Rock	modern residential	Roads	Mountain Areas	Industrial Facilities	Vegetation	Old Buildings	Agriculture	Total Actual	Class Accu (%)
Bare Land	495	3	0	1	0	0	1	0	0	500	99
Bare Rock	3	307	0	0	0	0	1	2	0	313	98.1
Modern Residential	1	0	185	0	0	0	0	2	0	188	98.4
Roads	0	0	0	63	0	0	0	0	0	63	100
Mountain Areas	0	0	0	0	61	0	2	0	0	63	96.8
Industrial Facilities	0	0	0	0	0	36	0	0	2	38	94.7
Vegetation	0	0	0	0	0	0	12	0	0	12	100
Old Buildings	0	0	2	0	0	0	0	47	1	50	94
Agriculture	0	0	0	0	0	0	0	0	25	25	100
Total Predicted	499	310	187	64	61	36	16	51	28	1252	

Bare land was correctly classified in 495 out of 500 samples, corresponding to a class-level accuracy of 99.0%. Bare rock achieved 307 correct classifications out of 313 samples, with six samples misclassified mainly as bare land, vegetation, and old buildings, resulting in a class-level accuracy of 98.1%. Modern residential achieved 185 correct classifications out of 188 samples, with one sample misclassified as bare land and two samples misclassified as old buildings, corresponding to 98.4% class-level accuracy.

Roads, vegetation, and agriculture achieved perfect class-level performance, with 63/63, 12/12, and 25/25 correctly classified samples, respectively. Mountain areas achieved 61 correct classifications out of 63 samples, with two samples misclassified as vegetation. Industrial facilities achieved 36 correct classifications out of 38 samples, with two samples misclassified as agriculture. Old buildings achieved 47 correct classifications out of 50 samples, with two samples misclassified as Modern residential and one sample misclassified as agriculture.

The residual errors mainly occurred between spectrally or spatially similar classes, particularly bare land and bare rock, modern residential and old buildings, and industrial facilities and agriculture. These confusions are expected in Makkah's heterogeneous urban–mountain environment, where built-up surfaces, rocky terrain, old urban fabric, and high-albedo bare surfaces may show similar spectral responses at 30 m spatial resolution.

Table 9 presents the confusion matrix generated by the proposed U-Net–DenseNet201–ViT hybrid model for the 2024 classification. The high diagonal values indicate strong class-level classification performance across all LULC categories. Overall, 1233 out of 1252 test samples were correctly classified, yielding an overall accuracy of 98.48%.

Table 9. Confusion matrix generated by the U-Net-DenseNet201-ViT hybrid model of classification of land changes in Makkah in 2024.

Actual \ Predicted	Bare Land	Bare Rock	Modern Residential	Roads	Mountain Areas	Industrial Facilities	Vegetation	Old Buildings	Agriculture	Total Actual	Class Accu (%)
Bare Land	391	2	0	2	0	1	0	0	0	396	98.7
Bare Rock	0	246	2	0	2	0	0	0	0	250	98.4
Modern Residential	0	1	245	0	0	2	2	0	0	250	98
Roads	0	0	0	123	1	0	1	0	0	125	98.4
Mountain Areas	0	0	0	0	53	0	0	0	1	54	98.1
Industrial Facilities	0	0	0	0	0	62	0	0	1	63	98.4
Vegetation	0	0	0	0	0	0	62	0	1	63	98.4
Old Buildings	0	0	0	0	0	0	0	38	0	38	100
Agriculture	0	0	0	0	0	0	0	0	13	13	100
Total Predicted	391	249	247	125	56	65	65	38	16	1252	

Bare land was correctly classified in 391 out of 396 samples, corresponding to a class-level accuracy of 98.7%. Bare rock achieved 246 correct classifications out of 250 samples, with two samples misclassified as modern residential and two as mountain areas, resulting in a class-level accuracy of 98.4%. Modern residential was correctly classified in 245 out of

250 samples, with one sample misclassified as bare rock, two as industrial facilities, and two as vegetation, corresponding to 98.0% class-level accuracy.

Roads achieved 123 correct classifications out of 125 samples, with two samples misclassified as mountain areas and vegetation. Mountain areas achieved 53 correct classifications out of 54 samples, with one sample misclassified as agriculture. Industrial facilities achieved 62 correct classifications out of 63 samples, with one sample misclassified as agriculture. Vegetation also achieved 62 correct classifications out of 63 samples, with one sample misclassified as agriculture. Old buildings and agriculture were fully correctly classified at the actual-class level, with 38/38 and 13/13 correctly classified samples, respectively.

The limited off-diagonal errors mainly occurred among spectrally or spatially similar classes, particularly bare rock, modern residential, industrial facilities, vegetation, and agriculture. These residual confusions are consistent with the heterogeneous urban–mountain environment of Makkah, where built-up materials, rocky surfaces, sparse vegetation, and small agricultural patches may exhibit overlapping spectral responses at 30 m spatial resolution.

4.5. Performance Analysis of U-Net-DenseNet201-ViT Models

The performance of U-Net, DenseNet201, and ViT models for landform classification in Makkah in 2013, as evaluated using various metrics, was also derived from the confusion matrix. The system performance for the nine landform classes is provided in Table 10. The unique land cover class information, as a representative architectural advantage, was consistently reflected in the accuracy, precision, sensitivity, specificity, and AUC values. Bare land classification was performed by all models with high speed and precision (99% accuracy and 99% sensitivity), which supports the validity of the test for identifying these areas. Specificity and AUC values range from 99.73 to 99.37, indicating that all models have a high probability of correctly identifying bare land relative to other land categories. The classification of bare rock was slightly more difficult, yielding an overall accuracy of 98.1% with sensitivity and specificity of 98.08%. The precision of 99.03% means that the models were highly confident in their positive predictions; occasionally, some rocks overlain by terrain similar to a rock type would be missed, as though predictably correct based on other factors (such as those exhibited by the strong-dominant classes), they looked like these “positive” zones. Here, the U-Net performed best in an unexpected manner because it is well suited to small rocky textures and patterns. Cities exhibit mixed performance across models, with varying levels of confidence. Modern building zones were detected with 98.4% accuracy and sensitivity, and 98.93% precision. Road achieved 100% classification accuracy and sensitivity, but with 98.44% precision due to false positives from linear features in bare land. DenseNet201, when combined with ANN, achieved better classification performance in urban scenes by leveraging its densely connected architecture to capture complex urban patterns. Several notable phenomena concerning natural types are discussed. The mountain areas were identified with an accuracy of 96.4% and a sensitivity of 96.83%, whereas the industrial objects had a lower sensitivity of 94.7%, indicating that this class was among the most difficult. Another uncommon class, vegetation, achieved a perfect sensitivity of 100%, suggesting an issue with classifying some vegetation patches as other green spaces.

Table 10. Performance metrics of U-Net, and DenseNet201 with ViT model for classifying changes in the city of Makkah for the period 2013.

Class	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)
Bare Land	99	99	99	99.73	99.37
Bare Rock	98.1	99.03	98.08	99.68	98.88
Modern Residential	98.4	98.93	98.4	99.81	99.11
Roads	100	98.44	100	100	100
Mountain Areas	96.4	96.83	96.83	100	98.42
Industrial Facilities	94.7	97.3	94.74	100	97.37
Vegetation	100	75	100	99.76	99.88
Old Buildings	94	94	94	99.92	96.96
Agriculture	100	89.29	100	99.82	99.91
Overall (Average)	98.32	95.79	97.89	99.75	98.82

The accuracy of the classification results obtained using the proposed hybrid model, which employs U-Net and DenseNet201 for local deep feature extraction and a ViT for global deep feature extraction and classification, is shown in Table 11, separately for each landform class. This model was used to assess the class of landform changes generated in the city of Makkah in 2024. The system demonstrated high accuracy and generalization across all landform categories. Such averages reveal an excellent capacity for classification on which we can equally count (for the class) to identify true samples as well as to reject false ones. At the class level, classification results were high across all classes; bare land, bare rock, modern residential, and roads had overall classification accuracies of nearly 98% or higher (AUC approximately 98.9–99.2%). Mountain areas and industrial facilities showed slightly lower precision (96.1% and 95.38%, respectively) but higher sensitivity (>98%) and specificity (>99.9%), indicating that the number of true positives was large while that of false positives was low. Vegetation and agriculture had no false negative rates (100% sensitivity). Nevertheless, the accuracies for vegetation (95.38%) and agriculture (81.25%) suggest that there were also some false positives due to spectral confusion or seasonal differences. Old buildings achieved a full score across all categories, representing the model’s best performance in this category.

Table 11. Performance metrics of U-Net, and DenseNet201 with the ViT model for classifying changes in the city of Makkah for the period 2024.

Class	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	AUC (%)
Bare Land	98.7	99.19	98.67	99.77	99.2
Bare Rock	98.4	98.8	98.4	99.8	99.1
Modern Residential	98	99.19	98	99.8	98.9
Roads	98.4	98.4	98.4	99.91	99.2
Mountain Areas	98.1	96.1	98.67	99.92	99.3
Industrial Facilities	98.4	95.38	98.41	99.92	99.2
Vegetation	98.4	95.38	98.41	99.92	99.2
Old Buildings	100	100	100	100	100
Agriculture	100	81.25	100	99.92	99.9
Overall (Average)	98.48	97.5	97.6	99.8	98.7

To further evaluate the contribution of the Laplacian filter, an additional comparison was conducted for the 2024 classification experiment by testing the proposed model with and without the Laplacian-based edge-enhancement step. Without the Laplacian filter, the model achieved an overall accuracy of 94.68%, precision of 94.3%, sensitivity of 94.3%, specificity of 96.9%, and AUC of 95.6%. With the Laplacian filter, the corre-

sponding values increased to 98.48%, 97.5%, 97.6%, 99.8%, and 98.7%, respectively. This represents improvements of +3.80 percentage points in accuracy, +3.20 percentage points in precision, +3.30 percentage points in sensitivity, +2.90 percentage points in specificity, and +3.10 percentage points in AUC. These results indicate that the Laplacian filter improved the model’s classification performance across all reported metrics and did not lead to noise-related performance degradation in the 2024 classification experiment.

To provide a synthetic, category-level accuracy metric derived from confusion matrices, the mean category accuracy (MCA) was calculated as the average accuracy across all categories. Unlike overall accuracy, which can be influenced by dominant categories, the MCA gives equal weight to each land use and coverage category and is therefore suitable for assessing balanced performance across categories. The MCA was calculated by dividing the average value of the main diameter by the total value of each category. Based on the confusion matrices, the proposed model achieved a mean category accuracy of 97.89% for 2013 and 98.72% for 2024, as shown in Table 12. These values confirm that the high performance of the classification was not solely driven by large categories but was consistent across all nine land use and coverage categories.

Table 12. Synthetic classifier-performance metrics derived from the confusion matrices.

Year	Overall Accuracy (%)	Kappa Coefficient	Mean Class Accuracy, MCA (%)	Interpretation
2013	98.32	0.978	97.89	High overall and class-balanced classification performance
2024	98.48	0.981	98.72	High overall and class-balanced classification performance

4.6. Interpreting Spatial Decision Logic with Grad-CAM

Grad-CAM is a special case of deep learning because it enables the interpretation of where the network focuses. Not metaphorically, but spatial. In the Makkah experiments, Grad-CAM was applied to the final convolutional layers of the hybrid CNN pipeline, enabling the projection of class-specific gradients onto the image plane and their display as continuous activation fields. Figure 7 illustrates this logic clearly: the initial multispectral composite, the Grad-CAM activation restricted to the municipal area, and the composite overlay to connect thought with place.

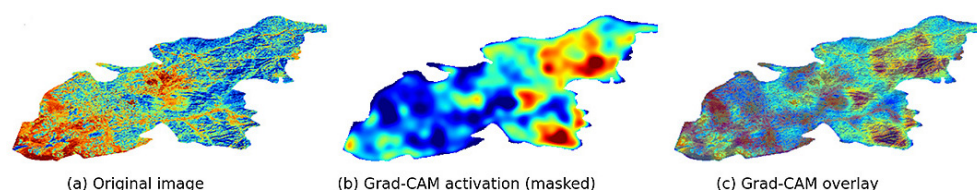


Figure 7. Grad-CAM-based spatial interpretation of LULC classification in the urban–mountain environment of Makkah: (a) the original multispectral composite, (b) class-specific Grad-CAM activation constrained to the municipal boundary, and (c) the fused Grad-CAM overlay.

Grad-CAM activations were aggregated across the test set using a class-wise averaging strategy. For each correctly classified test sample, a class-specific Grad-CAM heatmap was generated from the final convolutional layer of the CNN branch, resized to the input spatial dimensions, normalized to [0, 1], and restricted to the study-area mask. Heatmaps were first averaged within each LULC class to obtain class-level attention maps. The overall Grad-CAM map was then produced by macro-averaging the class-level maps, thereby

reducing the influence of class imbalance and ensuring that minority classes, such as vegetation and agriculture, contributed to the final interpretability assessment.

The first observation is that the model does not attend uniformly, nor does it do so randomly. Areas of high activation are concentrated along urban development lines, industrial belts, and road corridors. The latter are not accidental. They overlap with areas of high SWIR reflectance, high surface temperature, and high edge density, which are spectral-spatial features one would anticipate in dense built-up and industrial land in dry mountain areas. The focus of the model does not concern the large unsaturated units of bare land, even where they are preeminent in terms of area measurements, but rather focuses on interfaces: urban-rock frontiers, road-bare land frontiers, and newly created areas within the pre-existing fabric.

This is a noteworthy behavior that warrants consideration. This demonstrates that the CNN does not merely respond to color or brightness; rather, it is sensitive to coded structural contrast and contextual geometry. Linear elements (roads, corridors, and cut slopes) are always highlighted by edge-sensitive filters, demonstrating the usefulness of Laplacian enhancement and PAN-assisted texture proxies in image preprocessing. Grad-CAM has shown similar spatial selectivity in remote sensing tasks in urban environments, where deep models tend to respond more strongly to semantic boundaries than to interior homogeneity.

What is also indicative is what the model does not consider. Mountain slopes with vegetation, constant NDVI, and low anthropogenic disturbance have low activation, even though these are the most prominent in the eyes. This selective attention is a sign of learned irrelevancy, which is a characteristic of sound representation learning rather than overfitting. Practically, the Grad-CAM maps confirm that the classifier's high accuracy is not due to coincidences between the cues and physical objects, such as heat anomalies, construction materials, spatial regularity, and infrastructural geometry.

To quantitatively support the visual interpretation, Grad-CAM activations were compared with edge density maps derived from the Laplacian-enhanced LULC boundary representation. For each correctly classified test sample, the class-specific Grad-CAM map was resized to the input image dimensions, min-max normalized to [0, 1], and clipped to the Makkah municipal boundary. High-attention pixels were extracted using a top-percentile threshold, and the resulting binary attention masks were compared with binary edge density masks. Intersection-over-union (IoU) and Dice similarity coefficients were then calculated to quantify the degree of spatial overlap between model attention and structurally meaningful land cover boundaries. This analysis provides a quantitative basis for the claim that the model strongly attends to urban-rock interfaces, road corridors, and redevelopment edges.

4.7. Explaining SHAP Spectral and Topographic Contributions: Global Drivers of Local Decisions

Whereas Grad-CAM indicates where the model pays attention, SHAP indicates why the model makes decisions. Figure 8 presents a SHAP analysis of the spatial narrative, decomposing feature influence quantitatively, ranging from global behavior across the entire dataset to the structure of a single prediction.

For SHAP analysis, the background dataset was constructed from the training set only to avoid validation or test-set leakage. A stratified subset of training samples was selected from the nine LULC classes, covering spectral bands, spectral indices, thermal response, and DEM-derived topographic variables. This stratified training subset served as the SHAP reference distribution, and the SHAP expected value was therefore interpreted as the model output relative to a balanced representation of the training-data feature space. All local

SHAPs were computed with respect to this baseline, while global feature importance was obtained by averaging absolute SHAP values across the explained test samples.

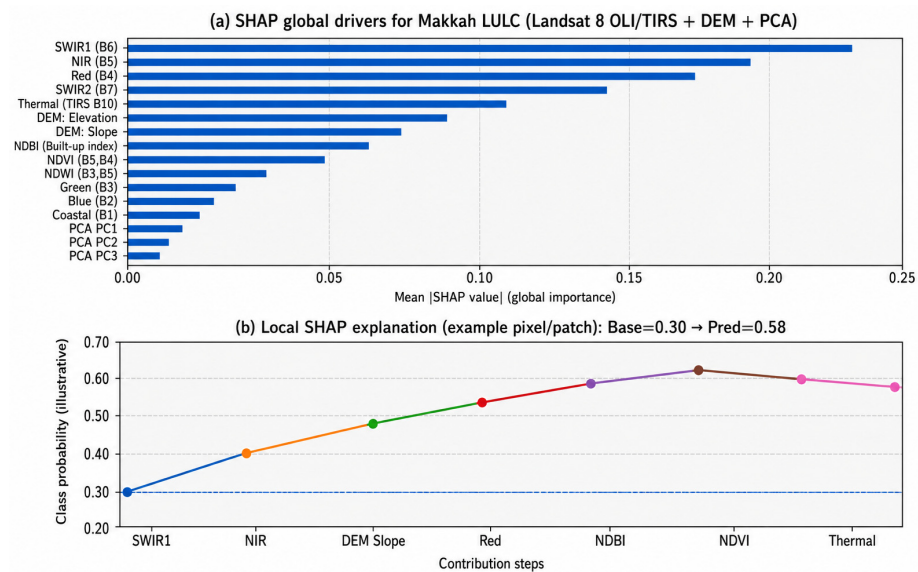


Figure 8. SHAP-based spectral and topographic attribution of model decisions from global feature importance to local prediction dynamics: (a) global feature importance across the Makkah dataset and (b) a curved local explanation tracing the contribution of spectral, thermal, and topographic features in driving a single pixel-level classification.

The ranking of SHAP significance at the global scale is neither trivial nor surprising. The explanation space is dominated by SWIR1, NIR, and Red bands with thermal response and topographic derivatives (slope and elevation) coming second. This hierarchy is a manifestation of the physics of land cover discrimination in arid cities. SWIR frequencies respond to construction materials and bare surfaces; NIR and Red frequencies encode vegetative stress and impenetrable surfaces; and thermal frequencies capture urban heat signatures that identify industrial and dense residential areas. The slope of the DEM is a structural prior that distinguishes developable plains from developed mountain areas.

However, the real worth of SHAP is realized at the local level. As shown in Figure 8, the curved local explanation follows the direction of movement of an individual pixel or a small patch of a sample of a neutral baseline probability to a confident class assignment. Every step curves the curve, occasionally upwards, occasionally back, and not additively. When SWIR1 increases, the prediction is strongly toward built-up land. The NIR reinforces this move. The slope softens slightly, indicating a geomorphic constraint. The momentum is then replenished by the thermal contribution, which swings the decision over the confidence threshold. NDVI and NDWI are moderators, not drivers, and can mitigate overconfidence when residual vegetation is present.

This choreography is important. This demonstrates that spectral, thermal, topographic, and other spectral negotiation evidence drives the model's decision-making. SHAP brings this negotiation to the open. Unlike opaque deep networks, every probability shift in this case is responsible, reproducible, and physically interpretable. Local accuracy and global consistency are the properties that make SHAP more than an explanation tool and transform it into a scientific tool.

In the case of Makkah, the interpretability was operational. Urban growth, as observed in the model, can be traced to increases in SWIR reflectance and surface temperature, whereas vegetation profits are consistent with positive NDVI changes. The explanations

not only justify post hoc predictions but also directly incorporate environmental processes into the model's reasoning.

Collectively, the Grad-CAM and SHAP results were coherent with respect to interpretability. One makes decisions based on space and other variables. One of these is attention, and the other is attribution. The very high spatial and spectral agreement between them strongly indicates that the claimed high classification accuracy is not incidental.

4.8. Synergistic Grad-CAM and SHAP Analysis for LULC Classification

The hybrid U-Net-DenseNet-ViT model achieved an overall accuracy of >97%, demonstrating its effectiveness. However, when deploying it in socio-environmentally relevant areas such as Makkah, it is essential to understand the reasoning behind the predictions. To overcome the black-box constraint inherent to deep learning, this study introduces a combination of Grad-CAM and SHAP into a coherent XAI system. This two-way scheme does not describe the model's behavior but confirms that the scheme's decisions are based on physically interpretable spatial and spectral principles.

Methodological Synergy: Feature Attribution versus Spatial Attention.

The framework employs complementary XAI techniques. Grad-CAM provides spatial interpretability by highlighting, within the image, where the model focuses most for each class prediction, answering the question: Where does the model look? SHAP is a cooperative game-theoretic method that provides feature-level attribution, used to measure the contribution of each input variable (spectral bands, indices, and topography) to a prediction and to explain the decision. Their combination enables a rigorous multimodal audit of the model's internal logic.

Figure 9 illustrates the fundamental novelty of this study by co-interpreting the locations where the model attends (Grad-CAM) and the reasons it assigns particular land cover classes (SHAP). In agriculture, Grad-CAM identifies structured, rectangular field designs and irrigation systems, whereas SHAP supports the conclusion with high NDVI, moisture-sensitive SWIR responses, and intermediate thermal activities, which are indicative of managed cultivation.

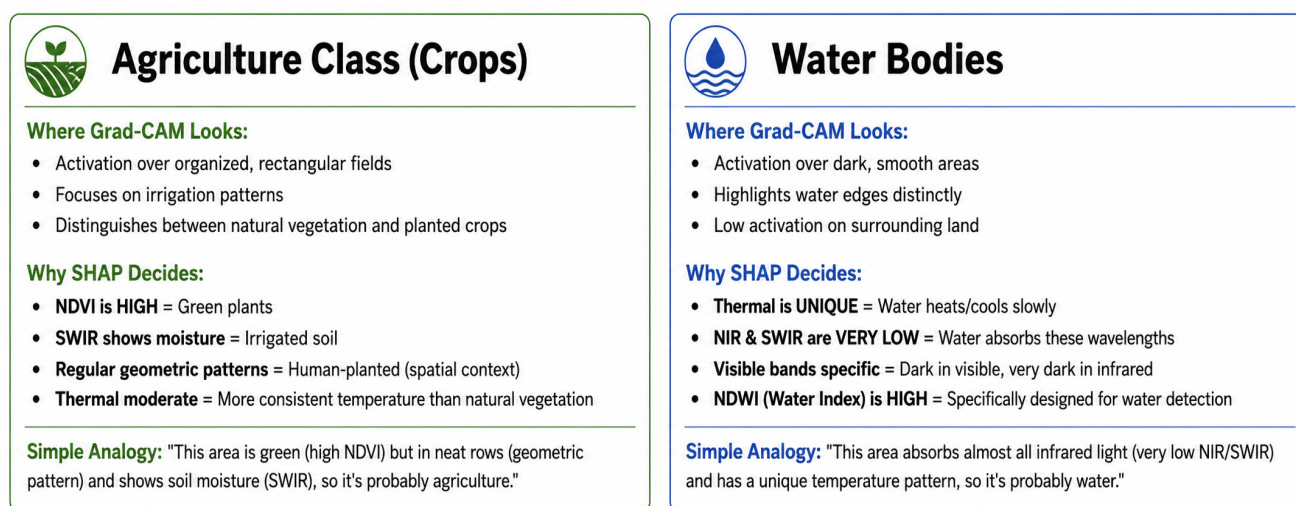


Figure 9. Interpretable AI framework for LULC classification via Grad-CAM and SHAP synergy.

The main interpretability outcome is summarized in the Grad-CAM–SHAP interpretability matrix (Table 13), which links the spatial attention patterns highlighted by Grad-CAM with the feature attribution drivers identified by SHAP for each LULC class. This matrix provides a structured audit of whether the model's spatial focus is consistent with physically meaningful spectral and topographic indicators. Urban classes were mainly

associated with built-up morphology, thermal response, NDBI, and SWIR-related material signatures, whereas vegetation and agriculture were interpreted through NDVI/NIR response, moisture-related information, and spatial pattern differences.

Table 13. Grad-CAM–SHAP consistency matrix for LULC interpretation in Makkah.

LULC Class	Grad-CAM Spatial Focus	Primary SHAP Feature Drivers	Interpretation
Modern Residential	Dense built-up clusters, rooftops, development fronts	Thermal response, NDBI, SWIR-related material signatures	Built-up areas are identified through heat/material signatures and compact urban morphology.
Industrial Facilities	Industrial belts, large roof structures, development edges	Thermal response, SWIR1/SWIR2, NDBI	Industrial surfaces are distinguished by high thermal/material responses and large structural patterns.
Roads	Linear corridors, intersections, transport networks	Thermal response, edge/linear context, SWIR response	Road classification is supported by both spectral response and linear spatial geometry.
Vegetation	Green patches, park edges, wadi corridors	High NDVI/NIR, low thermal response	Vegetation is identified by photosynthetic response and cooler surface conditions.
Agriculture	Structured field patterns, irrigated clusters	NDVI, NDWI, low slope	Agriculture is distinguished from natural vegetation by moisture response, geometric field structure, and terrain suitability.
Bare Land	Homogeneous plains and conversion edges	Low NDVI, dry SWIR response	Bare land is characterized by limited vegetation signal and dry soil reflectance.
Bare Rock/Mountain Areas	Rugged ridges, escarpments, steep slopes	DEM-derived slope, low NDVI, SWIR response	Rocky and mountainous areas are primarily separated through topographic and dry-surface signatures.

Although SHAP provides useful information about the spectral, thermal, and topographic drivers of LULC classification, its interpretation is not always sufficient when different classes share similar feature responses. For example, bare rock and industrial facilities may both exhibit high thermal response, strong SWIR-related material signatures, and elevated NDBI values. In such cases, SHAP alone may indicate similar dominant drivers for both classes, which can make feature-based interpretation ambiguous. The role of Grad-CAM is therefore important because it adds spatial and geometric context to the explanation. Bare rock is generally associated with irregular rocky outcrops, rugged terrain, and natural surface continuity, whereas industrial facilities are associated with large roof structures, rectilinear boundaries, compact artificial layouts, and development-edge patterns. Thus, the combined Grad-CAM–SHAP interpretation reduces the risk of relying only on spectral attribution and helps distinguish classes that share similar thermal or built-up feature responses.

4.9. Statistical Validation of Class-Level Performance

For each LULC class, paired statistical comparisons were conducted to assess performance differences between the proposed U-Net–DenseNet201–ViT model and the competing baseline models. In response to the reviewer’s recommendation, the revised analysis reports the Mean F1 Difference ($\Delta F1$) in addition to the adjusted significance values.

The F1-score was computed from precision and sensitivity as follows:

$$F1 = 2 \times (\text{Precision} \times \text{Sensitivity}) / (\text{Precision} + \text{Sensitivity})$$

The Mean F1 Difference was computed as follows:

$$\Delta F1 = F1_{\text{proposed}} - F1_{\text{baseline}}$$

Positive $\Delta F1$ values indicate that the proposed model achieved higher F1 performance than the corresponding baseline method. Raw p -values were adjusted for multiple comparisons using the Benjamini–Hochberg false discovery rate procedure. Therefore, Tables 13 and 14 now report both adjusted q -values and Mean F1 Difference values, allowing interpretation that considers both statistical significance and the magnitude of practical performance.

Table 14. Paired statistical comparison between the proposed U-Net–DenseNet201–ViT model and alternative hybrid models for Makkah LULC classification in 2013, with Mean F1 Difference ($\Delta F1$).

Class	Proposed Hybrid Model vs. ResNet-50 + U-Net + ANN (p -Value)	Proposed Hybrid Model vs. U-Net + DenseNet201 + ANN (p -Value)	Proposed Hybrid Model vs. ResNet-50 + DenseNet201 + ANN (p -Value)
Bare Land	0.0003	0.0006	0.001
Bare Rock	0.0004	0.0008	0.0012
Modern Residential	0.0006	0.001	0.0015
Roads	0.0005	0.0009	0.0013
Mountain Areas	0.0007	0.0012	0.0017
Industrial Facilities	0.0008	0.0013	0.0018
Vegetation	0.0009	0.0014	0.0019
Old Buildings	0.001	0.0015	0.002
Agriculture	0.0004	0.0007	0.0011
Baseline F1 (%)	89.9	92.65	94.35
Proposed macro-F1 (%)	96.37	96.37	96.37
Mean F1 Difference $\Delta F1$ (%)	6.47	3.72	2.02
Average	<0.001	<0.001	<0.002

Table note. Raw p -values were adjusted for 27 simultaneous class-level comparisons using the Benjamini–Hochberg false discovery rate procedure. All comparisons remained statistically significant after FDR correction at $q = 0.01$; the adjusted q -values ranged from 0.001755 to 0.002000. Bonferroni correction was examined as a more conservative reference threshold but was not used for the final interpretation. Mean F1 Difference ($\Delta F1$) represents the F1-score of the proposed U-Net–DenseNet201–ViT model minus the F1-score of the corresponding baseline model. Positive $\Delta F1$ values indicate a practical performance advantage of the proposed model. Reported q -values were adjusted using the Benjamini–Hochberg false discovery rate procedure.

4.9.1. Interpretation of Class-Level Statistical Results for 2013

The 2013 statistical comparison in Table 14 shows that the proposed U-Net–DenseNet201–ViT model achieved statistically significant improvements over the competing baseline models after Benjamini–Hochberg false discovery rate correction. In addition to the adjusted q -values, the revised table reports the Mean F1 Difference ($\Delta F1$), which quantifies the practical performance margin between the proposed model and each baseline.

The proposed model achieved a macro-F1 score of 96.37% in 2013. The corresponding F1-scores for ResNet-50 + U-Net + ANN, U-Net + DenseNet201 + ANN, and ResNet-50 + DenseNet201 + ANN were 89.90%, 92.65%, and 94.35%, respectively. Therefore, the Mean F1 Differences were +6.47, +3.72, and +2.02 percentage points.

These positive $\Delta F1$ values confirm that the superiority of the proposed model is not only statistically significant but also practically meaningful. The smallest $\Delta F1$ was observed against ResNet-50 + DenseNet201 + ANN, which was the strongest baseline in 2013. This

indicates that the proposed architecture maintained a measurable F1-score advantage even against the closest competing model.

4.9.2. Interpretation of Class-Level Statistical Results for 2024

The 2024 statistical comparison in Table 15 confirms the temporal robustness of the proposed U-Net–DenseNet201–ViT model. All paired comparisons remained statistically significant after Benjamini–Hochberg false discovery rate correction. However, the interpretation was revised to avoid attributing the p -value pattern solely to the overall accuracy level.

Table 15. Paired statistical comparison between the proposed U-Net–DenseNet201–ViT model and alternative hybrid models for Makkah LULC classification in 2024, with Mean F1 Difference ($\Delta F1$).

Class	Proposed Hybrid Model vs. ResNet-50 + U-Net + ANN (p -Value)	Proposed Hybrid Model vs. U-Net + DenseNet201 + ANN (p -Value)	Proposed Hybrid Model vs. ResNet-50 + DenseNet201 + ANN (p -Value)
Bare Land	0.0015	0.002	0.0028
Bare Rock	0.0017	0.0023	0.0031
Modern Residential	0.002	0.0027	0.0035
Roads	0.0018	0.0025	0.0033
Mountain Areas	0.0022	0.0029	0.0037
Industrial Facilities	0.0024	0.0031	0.0039
Vegetation	0.0025	0.0032	0.004
Old Buildings	0.0027	0.0034	0.0042
Agriculture	0.0016	0.0021	0.0029
Baseline F1 (%)	92.95	94.85	89.5
Proposed macro-F1 (%)	97.25	97.25	97.25
Mean F1 Difference $\Delta F1$ (%)	4.31	2.4	7.75
Average	<0.002	<0.003	<0.004

In the revised analysis, adjusted q -values are interpreted together with the Mean F1 Difference ($\Delta F1$). The proposed model achieved a macro-F1 score of 97.25% in 2024. The corresponding F1-scores for ResNet-50 + U-Net + ANN, U-Net + DenseNet201 + ANN, and ResNet-50 + DenseNet201 + ANN were 92.95%, 94.85%, and 89.50%, respectively. Therefore, the Mean F1 Differences were +4.31, +2.40, and +7.75 percentage points.

These results show that the proposed model maintained a positive practical performance margin over all baseline models in 2024. The smallest $\Delta F1$ was observed against U-Net + DenseNet201 + ANN, indicating that this was the closest competing baseline in 2024. Conversely, the largest $\Delta F1$ was observed against ResNet-50 + DenseNet201 + ANN, reflecting its weaker performance in 2024.

Therefore, the higher p -values observed in 2024 should be interpreted in relation to the magnitude and variability of the performance differences, rather than as a direct consequence of the proposed model’s higher overall accuracy. This revised interpretation supports the model’s superiority based on both inferential significance and a practical F1-score improvement.

4.10. Socioeconomic Correlates of Land Cover Change in Makkah (2013–2024)

The Makkah land cover changes between 2013 and 2024, computed from Tables 4 and 5, are a qualitative spatial record of the area’s intensifying socioeconomic change. This analysis provides explanations for these transitions in terms of urban economics, infrastructure construction, and environmental governance, with the implications of each of them properly grounded in the magnitudes of change assessed in this study.

We obtained the following results: modern residential: +5 pp; bare land: −8.33 pp; bare rock: −5 pp. This paired pattern constitutes a process of systematic transformation of undeveloped substrates into a (formal) urban fabric (i.e., previously undevelopable land use substrates). This development is mediated by strong demand for accommodation and services for both the increasing citizen population and pilgrimage flows. The −1 pp decrease in old buildings at the same time clearly signals urban renewal, whereby bygone properties have been rebuilt to current capacity, safety, and amenity standards.

A +2.5 pp increase in roads is significant as it emphasizes the need to invest in additional capacity for the network to reduce the cost of transport and, hence, stimulate development. It is also diagnostically relevant that road expansion goes hand-in-hand with the growth of modern residential (+5 pp): increased accessibility both precedes and enables the increase in urban density and urban overtaking observed by looking at classified maps.

The industrial sector grew two points higher, which was linked to network expansion. This relationship is also in tandem with the policy focus on the strategic location of employment land along accessible corridors. The conversion accounting suggests that this industrial expansion was largely a consequence of growth on previously barren substrates (i.e., greenfield development) rather than competing for existing vegetation or agriculture.

Vegetation increased by +4 pp, from 1% to 5%, a fivefold increase. This increase is most likely due to deliberate landscaping and the incorporation of green infrastructure in new urban developments. The concurrent development of physical assets and plants suggests a co-production of hard and soft services in response to policy objectives for better microclimate and city life.

Agriculture decreased by 1 percentage point between 2013 and 2024. This marginal loss illustrates the strong economic motivation in an urbanizing envelope, where the value of land for agriculture is heavily contested by more profitable uses: urban and industrial.

Mountain areas declined by 8 km², or about −0.67 percentage points. They act as a geomorphic pivot: there is no movement because the phenomenon under study was not based on permutations; it is territorial and should emerge only after cut-off values have been applied to the original diachronic maps. This inflexible category is an unmovable restraint that directs and controls the pattern of urban development.

Between 2013 and 2024, Makkah's land system shows the unmistakable trail of demand-led urbanism. The documented reallocations follow a sequence of development in which social imperatives (housing, facilities) and economic exigencies (access, labor) are the immediate causes of spatial change. The city is more linked, more built up, more landscaped, and less dependent on low-intensity land uses in results that show how the socioeconomic currents have reconfigured the Holy City over the past decade.

5. Discussion and Comparison

The spatial transformation of Makkah has become a distinct remote sensing problem: the convergence of sacred urban structures, challenging arid topography, and unprecedented development driven by pilgrimage demand and Saudi Arabia's Vision 2030 development agenda.

Past studies have consistently failed to answer an important, compounded question: how to develop a sound multiclassification system that is simultaneously accurate and mechanistically interpretable in this complex holy city. They state the aggregate performance cannot be explained by misclassifying sacred and commercial areas or conflating bare rock with industrial rooftops. This indifference to traceable decision-making logic is a deep scientific flaw, in which land use bears a significant social, economic, and religious burden, thereby inhibiting the ability of such models to provide accountable urban governance.

An analysis of the existing literature indicates a trend toward gradual improvement, with gaps that remain unresolved. Initial research provided essential timing points of reference based on Landsat imagery and pixel-based analysis, but its spatial generalization obscured the details and dense tissues around the Haram borders. Later, spectral-index methods, although useful for large-scale detection of vegetation on arid terrain, were too crude to be applied in Makkah, with its mosaic of mixed urban and natural cover. The development of deep learning has constituted a major change, and researchers have shown that convolutional neural networks can differentiate urban structures. But such uses restricted themselves to binary classification or small sets of classes, silently ignoring difficult yet important categories such as older building stocks, sparse agriculture, and certain types of industries, exactly the categories that display subtle urban dynamics. Other threads of research have produced useful policy accounts or suggested complex hybrid and multimodal architectures. However, many were not empirically supported in the context of Makkah and its specific, limited environment, or avoided the core problem of model explainability. They are all linked by their reliance on measures of headline accuracy, such as overall accuracy and Kappa, which can obscure per-class confusion and entirely hide the decision pathways. These methods affirm that change has occurred but are silent on how or why particular, often ill-defined, land cover was categorized, leaving planners with a map on which they must take a leap of faith.

This study presents a new hybrid architecture that combines U-Net, DenseNet201, and a ViT, with a small, explicit architecture design that effectively addresses the dual requirements of accuracy and interpretability. What is new is not just the architectural fusion per se, but also how the systematic design exposes the model's spatial and spectral logic to criticism and, specifically, targets Makkah's spectrally related and contextually distinct land cover.

The importance of this practice is threefold, and it represents a unique form of progress. First, it provides a multiclass, statistically rigorous classification that captures the city's full functional diversity, rather than relying on a single accuracy score, by offering a more specific, class-level performance audit. Second, and most importantly, it considers explainability central to scientific data rather than a post hoc adornment. Third, this interpretability extends not only to interpretability but also to formal statistical and socioeconomic validation, making the model an active analytical framework for understanding urban processes.

The interpretability framework is grounded in complementary principles. Grad-CAM is like a spatial flashlight, helping to understand the places where the model pays the maximum attention. However, the places that are most important in the development of cities and natural bases are not homogeneous areas but their potential interfaces. This displays an acquired interest in structural boundaries and the frontiers of change. Combined with SHAP analysis, it provides a feature attribution ledger and identifies the specific contributions of spectral bands and topographic variables to each classification decision. This grounding of model logic in measurable physical geography ensures that high accuracy is not an artifact of data bias but a result of sound physical reasoning.

The paired *t*-tests at the class level provide strong validation, which is typically not reported in LULC research, and the proposed model is statistically validated with high confidence in performance across all single land cover types ($p < 0.01$). The socioeconomic correlates analysis, as it follows, is an important translation operation that transforms pixel-based accounting of areas into a rational narrative of urban economics. It explains how contemporary building expansion is accompanied by road widening, how industrial expansion follows available routes, and how vegetation growth is a sign of intentional greening rather than climatic happenstance. This produces a causally abundant narrative of spatial transformation.

The tension between urban development and agricultural sustainability is clearly delineated in the model output. The marginalization of agriculture is nearly complete, estimated to be a direct result of excessive competition for land use in a pilgrimage economy. Simultaneously, a fivefold increase in vegetated area is identified as an indicator of corresponding sustainability investments in urban livability and microclimate mitigation. This essential distinction between productive agriculture and amenity greenery can be achieved solely through a detailed, interpretable multiclass scheme.

To address the need for contextual benchmarking, the proposed U-Net–DenseNet201–ViT framework was compared with recent LULC and remote sensing studies published between 2024 and 2026. Priority was given to studies conducted in Makkah and Saudi Arabia, while non-Saudi studies were treated only as methodological references. Because these studies differ in spatial resolution, data source, number of classes, classification objective, and validation protocol, the comparison is interpreted as contextual benchmarking rather than a direct one-to-one accuracy ranking, as shown in Table 16.

Table 16. Comparative benchmarking of the proposed U-Net–DenseNet201–ViT framework against recent LULC studies in Makkah and Saudi Arabia.

Study	Study Area	Data/Resolution	Method	Classes	Reported Performance	Relevance to the Present Study
Aljaddani et al. [54]	Riyadh, Jeddah, Makkah, Madinah, Dammam	Landsat TM-5, OLI-TIRS-8, OLI2-TIRS2-9/30 m	Random Forest, LST, UTFVI, DiD	4	OA: 89–97%	Includes Makkah and confirms urban expansion and thermal impacts, but uses broader four-class LULC mapping. Highly relevant geographically, but focused on flood response rather than fine-grained LULC classification. Useful Saudi urban-growth comparison, but without deep-learning semantic feature extraction. Methodologically close because it uses hybrid learning and Landsat-8 data.
Hidayatulloh et al. [55]	Wadi Ibrahim watershed, Makkah	Landsat + 30 m DEM	RF, CA-ANN, hydrological modeling	LULC flood-risk classes	Average $\kappa = 0.86$	Demonstrates ML effectiveness in Saudi Arabia, but the task is binary and environmentally specific. Provides fine-grained multiclass mapping, high accuracy, interpretability, and statistical validation.
Altamimi et al. [56]	Hail, Saudi Arabia	Landsat 8/30 m	GIS, remote sensing, Markov chain	Broad LULC classes	$\kappa = 0.73$	
Mashraqi et al. [57]	Najran, Saudi Arabia	Landsat 8/30 m	CNN–Random Forest with Ant Colony Optimization	Multiclass LULC	Best OA: 97.56%	
Al-Huqail et al. [58]	Farasan Islands, Saudi Arabia	Landsat 8 SR/30 m	RF, SVM, GB, ensemble ML	2	OA: 92.2%, $\kappa = 0.84$	
Present study	Makkah, Saudi Arabia	Landsat 8/30 m + DEM-derived variables	U-Net–DenseNet201–ViT with Grad-CAM and SHAP	9	OA: 98.32% / $\kappa = 0.978$ (2013); OA: 98.48% / $\kappa = 0.981$ (2024)	

Policy relevance to Saudi Vision 2030: The quantified LULC changes observed between 2013 and 2024 are also relevant to the broader policy directions of Saudi Vision 2030. The expansion of modern residential areas from 180,000,000 m² to 240,000,000 m² (+60,000,000 m²; +5.00 percentage points), roads from 60,000,000 m² to 120,000,000 m² (+60,000,000 m²; +5.00 percentage points), and industrial facilities from 36,000,000 m² to 60,000,000 m² (+24,000,000 m²; +2.00 percentage points) reflects an urban and infrastructure growth pattern consistent with national objectives related to urban development, improved connectivity, economic diversification, and enhanced service capacity. This is particularly relevant for Makkah, where infrastructure and accessibility improvements are closely connected to the goals of improving residents’ quality of life and supporting the pilgrim experience under Vision 2030.

At the same time, vegetation increased from 12,000,000 m² to 60,000,000 m² (+48,000,000 m²; +4.00 percentage points), indicating that the observed urban expansion

was accompanied by measurable growth in green cover. This trend is consistent with the sustainability and quality-of-life dimensions of Vision 2030 and related national greening initiatives. Therefore, the LULC results provide spatially quantified evidence that can support policy monitoring by showing where urban growth, infrastructure expansion, and vegetation gains occurred between 2013 and 2024.

Although the proposed U-Net–DenseNet201–ViT framework achieved strong and consistent performance, a few aspects should be considered. Since the analysis relied on Landsat imagery, the spatial resolution may limit the detection of very small or highly mixed urban features, while the use of two temporal snapshots, 2013 and 2024, may not capture short-term seasonal or annual variations. Some classes with similar spectral or thermal characteristics, such as bare rock and industrial facilities, may also require spatial-context interpretation in addition to SHAP-based feature attribution. The use of Grad-CAM alongside SHAP helped address this issue by adding spatial and geometric evidence to the interpretation. Future work may extend the framework using higher-resolution imagery, additional time points, and external validation in other arid urban regions.

6. Conclusions

Reliable LULC mapping in Makkah requires models that combine spatial precision, contextual reasoning, and transparent decision evidence. The proposed U-Net–DenseNet201–ViT framework achieved strong performance for both 2013 and 2024, with overall accuracy above 97%, confirming its ability to classify complex urban, natural, and infrastructural land cover classes in a topographically constrained environment.

Interpretability analysis strengthened the credibility of these results. Grad-CAM showed that the model focused on urban–rock interfaces, redevelopment edges, and road corridors, which correspond to active land conversion zones. SHAP further indicated that SWIR reflectance, thermal response, and terrain slope were the most influential drivers for distinguishing urban, industrial, and natural classes. Class-level paired statistical tests ($p < 0.01$) confirmed that the observed performance differences were statistically meaningful rather than chance-based.

The detected LULC changes reveal infrastructure-driven urban expansion. Modern residential areas increased from 15% to 20%, roads doubled from 5% to 10%, and industrial facilities rose from 3% to 5%. In contrast, agriculture declined from 2% to 1%, while vegetation increased from 1% to 5%, suggesting policy-related greening rather than productive agricultural expansion.

Beyond Makkah, the framework may be transferable to other rapidly urbanizing sacred, heritage-sensitive, or topographically complex cities, provided that local retraining and calibration are performed to reflect differences in land cover composition, terrain, climate, and planning context.

Future work will extend this framework by testing it in other and topographically complex cities, incorporating higher-resolution multisensor data such as Sentinel-2, SAR, thermal, and UAV imagery, and developing multi-temporal prediction models to support early detection of urban expansion, vegetation change, and infrastructure-driven land conversion.

Author Contributions: Conceptualization, A.A.A. and T.H.; methodology, A.A.A. and T.H.; software, A.A.A. and T.H.; validation, A.A.A. and T.H.; formal analysis, T.H. and A.A.A.; investigation, T.H. and A.A.A.; resources, A.A.A. and T.H.; data curation, T.H. and A.A.A.; writing—original draft preparation, A.A.A.; writing—review and editing, T.H.; visualization, A.A.A. and T.H.; supervision, T.H. and A.A.A.; project administration, T.H. and A.A.A.; funding acquisition, A.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors received financial support from the KAU Endowment (Waqf) and the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia, solely to cover the article processing charge (APC). This support did not fund the research, study design, data collection, analysis, interpretation, or manuscript preparation.

Data Availability Statement: Data supporting this work were obtained from the publicly available Internet at the following link: <https://earthexplorer.usgs.gov/> (accessed on 6 December 2024).

Acknowledgments: The authors acknowledge the KAU Endowment (Waqf) and the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia, for their support toward the publication of this article.

Conflicts of Interest: The authors affirm the absence of any conflicts of interest.

References

1. Amran, Y.H.A.; Amran, Y.H.M.; Alyousef, R.; Alabduljabbar, H. Renewable and sustainable energy production in Saudi Arabia according to Saudi Vision 2030; Current status and future prospects. *J. Clean. Prod.* **2020**, *247*, 119602. [CrossRef]
2. Alshari, E.A.; Gawali, B.W. Development of classification system for LULC using remote sensing and GIS. *Glob. Transit. Proc.* **2021**, *2*, 8–17. [CrossRef]
3. Alshari, E.A.; Gawali, B.W. Artificial Intelligence Techniques for Observation of Earth's Changes. Available online: <https://www.intechopen.com> (accessed on 6 December 2024).
4. Alshahrani, M.; Al-Jabbar, M.; Alshari, E.A.; Senan, E.M.; Ahmed, I.A. Hybrid methods for land use and land cover classification using remote sensing and combined spectral feature extraction: A case study of Najran City, KSA. *Open Geosci.* **2025**, *17*, 19. [CrossRef]
5. Alshari, E.A.; Gawali, B.W. Analysis of Machine Learning Techniques for Sentinel-2A Satellite Images. *J. Electr. Comput. Eng.* **2022**, *2022*, 9092299. [CrossRef]
6. Alshari, E.A.; Gawali, B.W. *Evaluation of the Multispectral Satellites with Object-Based Classifiers for Land Use and Land Cover Classification*; Atlantis Press: Dordrecht, The Netherlands, 2023; pp. 602–625. [CrossRef] [PubMed]
7. Gaur, S.; Singh, R. A Comprehensive Review on Land Use/Land Cover (LULC) Change Modeling for Urban Development: Current Status and Future Prospects. *Sustainability* **2023**, *15*, 903. [CrossRef]
8. Marquez, A.M.; Guevara, E.; Rey, D. Hybrid Model for Forecasting of Changes in Land Use and Land Cover Using Satellite Techniques. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 252–273. [CrossRef]
9. Horry, M.J.; Chakraborty, S.; Pradhan, B.; Shulka, N.; Almazroui, M. Two-Speed Deep-Learning Ensemble for Classification of Incremental Land-Cover Satellite Image Patches. *Earth Syst. Environ.* **2023**, *7*, 525–540.
10. Yaloveha, V.; Podorozhniak, A.; Kuchuk, H.; Garashchuk, N. Performance comparison of CNNs on high-resolution multispectral dataset applied to land cover classification problem. *Radioelectron. Comput. Syst.* **2023**, *106*, 107–118. [CrossRef]
11. Mahendra, H.N.; Pushpalatha, V.; Mallikarjunaswamy, S.; Subramoniam, S.R. A hybrid deep learning model based on vision transformer and convolutional neural networks for land use and land cover classification. *Appl. Soft Comput.* **2026**, *192*, 114775. [CrossRef]
12. Ali, E.; Alshari, A.S.; Alqhatani, A.; Al-Wagih, K.; Olayah, F.; Senan, E.M.; Aljubayri, M. Assessing green area changes in Najran City, Saudi Arabia (2013–2022) using hybrid deep learning techniques. *Open Geosci.* **2025**, *17*, 20250793. [CrossRef]
13. Mirhoseini Nejad, S.M.; Abbasi-Moghadam, D.; Sharifi, A. ConvLSTM-ViT: A Deep Neural Network for Crop Yield Prediction Using Earth Observations and Remotely Sensed Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 17489–17502. [CrossRef]
14. Alghamdi, A.S.; Moore, T.W. Detecting Temporal Changes in Riyadh's Urban Heat Island. *Pap. Appl. Geogr.* **2015**, *1*, 312–325. [CrossRef]
15. Alqurashi, A.F.; Kumar, L. An assessment of the impact of urbanization and land use changes in the fast-growing cities of Saudi Arabia. *Geocarto Int.* **2019**, *34*, 78–97. [CrossRef]
16. Alqurashi, A.F.; Kumar, L. Land Use and Land Cover Change Detection in the Saudi Arabian Desert Cities of Makkah and Al-Taif Using Satellite Data. *Adv. Remote Sens.* **2014**, *3*, 106–119. [CrossRef]
17. AlKheder, S.; AlKandari, A. The impact of dust on Kuwait International Airport operations: A case study. *Int. J. Environ. Sci. Technol.* **2020**, *17*, 3467–3474. [CrossRef]
18. The Impact of Applying Governance Standards on Improving the Performance of Healthcare Organizations: A Case Study of Hospitals in the Kingdom of Saudi Arabia. | EBSCOhost. Available online: https://openurl.ebsco.com/EPDB:gcd:13:17462109/detailv2?sid=ebsco:plink:scholar&id=ebsco:gcd:182070994&crl=c&link_origin=scholar.google.com (accessed on 18 June 2025).

19. Jabal Omar Development Project | Sixsense. Available online: <https://www.sixsense-group.com/en/projects/jabal-omar-development-project> (accessed on 17 June 2025).
20. Sumit; Bisht, S.; Joshi, S.; Rana, U. Comprehensive Review of R-CNN and its Variant Architectures. *Int. Res. J. Adv. Eng. Hub. (IRJAEH)* **2024**, *2*, 959–966. [CrossRef]
21. Alqurashi, A.F.; Kumar, L.; Sinha, P. Urban land cover change modelling using time-series satellite images: A case study of urban growth in five cities of Saudi Arabia. *Remote Sens.* **2016**, *8*, 838. [CrossRef]
22. Li, F.; Yigitcanlar, T.; Nepal, M.; Nguyen, K.; Dur, F. Machine learning and remote sensing integration for leveraging urban sustainability: A review and framework. *Sustain. Cities Soc.* **2023**, *96*, 104653. [CrossRef]
23. Said, Y.; Barr, M.; Saidani, T.; Atri, M. Desertification Detection in Makkah Region based on Aerial Images Classification. *Comput. Syst. Sci. Eng.* **2021**, *40*, 607–618. [CrossRef]
24. Tong, X.Y.; Xia, G.S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [CrossRef]
25. Boulila, W.; Ghadorh, H.; Khan, M.A.; Ahmed, F.; Ahmad, J. A novel CNN-LSTM-based approach to predict urban expansion. *Ecol. Inform.* **2021**, *64*, 101325. [CrossRef]
26. Alqurashi, A.F.; Alharbi, O.A. Changes in Land Use and Land Cover Patterns in Two Desert Basins Using Remote Sensing Data. *Geosciences* **2025**, *15*, 178. [CrossRef]
27. Albarakati, H.M.; Khan, M.A.; Hamza, A.; Khan, F.; Kraiem, N.; Jamel, L. A Novel Deep Learning Architecture for Agriculture Land Cover and Land Use Classification from Remote Sensing Images Based on Network-Level Fusion of Self-Attention Architecture. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 6338–6353. [CrossRef]
28. Chen, W.; Chen, J.G.; Wan, Y.W.; Liu, X.N.; Cai, M.Y.; Xu, J.G.; Cui, H.B.; Duan, M.D. Land Cover Classification Based on Multimodal Remote Sensing Fusion. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2024**, *10*, 35–40. [CrossRef]
29. Imam, A.; Alhaddad, B.; Roca, J. Remote Sensing Efficiency for Urban Analysis of Mecca and Surrounds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B8*, 905–910. [CrossRef]
30. Makkah al-Mukarramah City-Saudipedia. Available online: <https://saudipedia.com/en/article/1148/geography/provinces-and-cities/makkah-al-mukarramah-city> (accessed on 17 June 2025).
31. Kunwar, S.; Ferdush, J. Mapping of Land Use and Land Cover (LULC) using EuroSAT and Transfer Learning. *Rev. Int. Géomatique* **2023**, *33*, 1–13.
32. The Importance of Considering Atmospheric Correction in the Pre-Processing of Satellite Remote Sensing Data Intended for the Management and Detection of Cultural Sites: A Case Study of the Cyprus Area. Available online: https://www.researchgate.net/publication/257067491_The_importance_of_considering_atmospheric_correction_in_the_pre-processing_of_satellite_remote_sensing_data_intended_for_the_management_and_detection_of_cultural_sites_a_case_study_of_the_Cyprus_area (accessed on 17 June 2025).
33. Laban, N.; Abdellatif, B.; Ebied, H.M.; Shedeed, H.A.; Tolba, M.F. Performance Enhancement of Satellite Image Classification Using a Convolutional Neural Network. *Adv. Intell. Syst. Comput.* **2018**, *639*, 673–682. [CrossRef]
34. Khan, M.; Hanan, A.; Kenzhebay, M.; Gazzea, M.; Arghandeh, R. Transformer-based land use and land cover classification with explainability using satellite imagery. *Sci. Rep.* **2024**, *14*, 16744. [CrossRef] [PubMed]
35. Borra, S.; Thanki, R.; Dey, N. Satellite Image Enhancement and Analysis. In *SpringerBriefs in Applied Sciences and Technology*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 13–29. [CrossRef]
36. Ahmed, I.A.; Senan, E.M.; Shatnawi, H.S.A.; Alkhraisha, Z.M.; Al-Azzam, M.M.A. Multi-Models of Analyzing Dermoscopy Images for Early Detection of Multi-Class Skin Lesions Based on Fused Features. *Processes* **2023**, *11*, 910. [CrossRef]
37. Ma, T.; Xu, J.; Zhao, J.; Zhang, Y. Optimized Laplacian image sharpening algorithm based on graphic processing unit. *Phys. A Stat. Mech. Its Appl.* **2014**, *416*, 400–410. [CrossRef]
38. Khalid, A.; Senan, E.M.; Al-Wagih, K.; Al-Azzam, M.M.A.; Alkhraisha, Z.M. Hybrid Techniques of X-ray Analysis to Predict Knee Osteoarthritis Grades Based on Fusion Features of CNN and Handcrafted. *Diagnostics* **2023**, *13*, 1609. [CrossRef] [PubMed]
39. Hamdi, M.; Senan, E.M.; Jadhav, M.E.; Olayah, F.; Awaji, B.; Alalayah, K.M. Hybrid Models Based on Fusion Features of a CNN and Handcrafted Features for Accurate Histopathological Image Analysis for Diagnosing Malignant Lymphomas. *Diagnostics* **2023**, *13*, 2258. [CrossRef] [PubMed]
40. Austin-Gabriel, B.; Monsalve, C.N.; Varde, A.S. Power Plant Detection for Energy Estimation Using GIS with Remote Sensing, CNN & Vision Transformers. 2024. Available online: <http://arxiv.org/abs/2412.04986> (accessed on 9 February 2026).
41. Weng, W.; Zhu, X. INet: Convolutional Networks for Biomedical Image Segmentation. *IEEE Access* **2021**, *9*, 16591–16603. [CrossRef]
42. Borawar, L.; Kaur, R. ResNet: Solving Vanishing Gradient in Deep Networks. *Lect. Notes Netw. Syst.* **2023**, *600*, 235–247. [CrossRef]
43. Ali, E.; Alshari, A.S.; Alqhatani, A.; Al-Wagih, K.; Olayah, F.; Senan, E.M.; Aljubayri, M. Maps analysis of Najran City, Saudi Arabia to enhance agricultural development using hybrid system of ANN and multi-CNN models. *Open Geosci.* **2025**, *17*, 20250898. [CrossRef]

44. Karim, M.; Deng, J.; Ayoub, M.; Dong, W.; Zhang, B.; Yousaf, M.S.; Bhutto, Y.A.; Ishfaq, M. Improved Cropland Abandonment Detection with Deep Learning Vision Transformer (DL-ViT) and Multiple Vegetation Indices. *Land* **2023**, *12*, 1926. [\[CrossRef\]](#)
45. Tejasree, G.; Loganathan, A. A novel multi-class land use/land cover classification using deep kernel attention transformer for hyperspectral images. *Earth Sci. Inform.* **2024**, *17*, 593–616. [\[CrossRef\]](#)
46. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In *Proceedings-30th IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2017; pp. 2261–2269. [\[CrossRef\]](#)
47. Chaudhri, S.N.; Mallikarjuna Rao, Y.; Rajput, N.S.; Subramanyam, M.V. Vision Transformer-Based LULC Classification Using Remotely Sensed Hyperspectral Image. *Lect. Notes Electr. Eng.* **2024**, *1157*, 127–136. [\[CrossRef\]](#)
48. Halawani, H.T.; Alshari, E.A.; Asiri, Y.; Senan, E.M.; Mashraqi, A.M.; Almoaina, O.A. Hybrid deep learning with a random forest system for sustainable agricultural land cover classification using DEM in Najran, Saudi Arabia. *Open Geosci.* **2025**, *17*, 20250902. [\[CrossRef\]](#)
49. Bansal, K.; Tripathi, A.K. Dual level attention based lightweight vision transformer for streambed land use change classification using remote sensing. *Comput. Geosci.* **2024**, *191*, 105676. [\[CrossRef\]](#)
50. Estornell, J.; Martí-Gavliá, J.M.; Sebastiá, M.T.; Mengual, J. Principal component analysis applied to remote sensing. *Model. Sci. Educ. Learn.* **2013**, *6*, 83. [\[CrossRef\]](#)
51. Ahangarha, M.; Rezvan, H.; Zoj, M.J.V.; Youssefi, F. Employing Transfer Learning in Land-use Land-cover for Risk Management. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2024**, *XLVIII-3-W3-2024*, 1–7. [\[CrossRef\]](#)
52. Gadiparthi, M.; Al-Fatlawy, R.R.; Kulkarni, M.; Malathy, V.; Anbunathan, R. Land Cover Classification in High-Resolution Satellite Images Using Vision Transformer and ConvNeXt Approach. In *2nd IEEE International Conference on Data Science and Network Security, ICDSNS*; IEEE: Piscataway, NJ, USA, 2024. [\[CrossRef\]](#)
53. Alzahem, A.; Boulila, W.; Koubaa, A.; Khan, Z.; Alturki, I. Improving satellite image classification accuracy using GAN-based data augmentation and vision transformers. *Earth Sci. Inform.* **2023**, *16*, 4169–4186. [\[CrossRef\]](#)
54. Aljaddani, A.H. Evaluation of the Land Use Land Cover Impact on Surface Temperature and Urban Thermal Comfort: Insight from Saudi Arabia's Five Most Populated Cities (2000–2024). *Urban Sci.* **2026**, *10*, 157. [\[CrossRef\]](#)
55. Hidayatulloh, A.; Elfeki, A.; Bahrawi, J.; Alzahrani, F.; Alamoudi, F.; Elhag, M. Assessing the Impact of Land Use and Land Cover Changes on Flood Hazard in the Wadi Ibrahim Watershed. *Land* **2026**, *15*, 742. [\[CrossRef\]](#)
56. Altamimi, A.; Adenle, Y.A.; Alshuwaikhat, H.M. Spatiotemporal analysis of urban expansion in Hail City, Saudi Arabia, using GIS, remote sensing, and Markov chain model. *Discov. Cities* **2026**, *3*, 74. [\[CrossRef\]](#)
57. Mashraqi, A.M.; Alshari, E.A.; Halawani, H.T.; Senan, E.M.; Asiri, Y.; Alowadhi, B.M. Hybrid deep learning and optimization-based land use and land cover classification for advancing sustainable agriculture in Najran city, Saudi Arabia. *Sci. Rep.* **2025**, *15*, 41808. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Al-Huqail, A.A.; Islam, Z.; Al-Harbi, H.F.; Khan, F. AI-driven mangrove mapping on Farasan Islands, Saudi Arabia: Enhancing the detection of dispersed patches with ML classifiers. *Sci. Rep.* **2025**, *15*, 19285. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.