GEOMATICS AND ENVIRONMENTAL ENGINEERING • Volume 20 • Number 3 • 2026

https://doi.org/10.7494/geom.2026.20.3.25

Optimizing Built-Up Area Extraction in Semi-Arid Regions Using Sentinel-2A Imagery: A Comparative Analysis of Spectral Indices and PCA-Based Classification in Batna, Algeria

Wahiba Touati[1] *, Mahdi Kalla[1], Lemya Kacha[2]

[1] University of Batna 2, Institute of Earth Sciences and the Universe, Laboratory of Natural Hazards and Territory Planning, Algeria
[2] University of Batna 1, Institute of Architecture and Urbanism, Laboratory of the Child, City and Environment, Algeria
* Corresponding author

Abstract: Accurate detection of built-up areas in semi-arid regions is vital for urban planning and environmental monitoring. However, built-up surfaces and bare soils often produce very similar spectral responses. As a result, this similarity causes confusion in satellite image classification. Additionally, spectral overlap among urban materials, bare soil, and sparse vegetation further complicates detection. This study evaluates several spectral indices, including DBSI, NDTI, NDVI, BRBA, and BSI, combined with Principal Component Analysis (PCA) to enhance built-up area extraction from Sentinel-2A imagery. Images captured during the driest season were selected to maximize spectral contrast. Three classification schemes based on Support Vector Machine (SVM) were tested. The first scheme used DBSI, NDTI, and NDVI. The second used BRBA, NDTI, and NDVI. The third relied on PCA-derived components. The results indicate that the PCA-based approach achieved the highest classification accuracy at 95%. In comparison, the DBSI/NDTI/NDVI combination reached 93%, while the BRBA/NDTI/NDVI scheme achieved 92%. Therefore, PCA helps reduce spectral confusion and enhances the identification of built-up areas in semi-arid environments. Overall, combining multiple spectral indices with dimensionality reduction offers a reliable method for urban analysis using Sentinel-2 imagery.

Keywords: built-up area extraction, Sentinel-2A imagery, semi-arid regions, PCA, spectral indices, SVM classification, dimensionality reduction

Received: June 20, 2025; accepted: February 6, 2026

E-mails and ORCID iDs: wahiba.touati@univ-batna.dz, https://orcid.org/0009-0005-6422-9095 (W.T.); m_kalla1@yahoo.fr, https://orcid.org/0000-0001-6380-8777 (M.K.); lemya.kacha@univ-batna.dz, https://orcid.org/0000-0002-2527-5356 (L.K.).

1. Introduction

Accurate identification of built-up areas plays a key role in urban development, land-use assessment, and long-term sustainability [1, 2]. As cities expand rapidly, especially at their fringes, remote sensing has emerged as an efficient and affordable tool for mapping urban surfaces across large regions and extended timeframes [3]. Unlike official topographic datasets, satellite imagery provides frequent updates, which are essential for monitoring rapid urban change. Moreover, many government databases are unavailable, prohibitively expensive, or outdated in several regions, including the study area. To achieve this, numerous studies have relied on medium-resolution imagery from sensors such as Landsat and Sentinel-2, along with supervised classification methods to detect built-up features [4, 5]. In addition, historical maps and administrative records help trace long-term settlement patterns [6–8]. These methods enable consistent quantification of building footprints, counts, and densities [9–11]. Advances in automated building detection from imagery further facilitate scalable mapping of built-up areas [7]. However, distinguishing these features remains challenging in semi-arid zones, where sparse vegetation and soil reflectance often mimic urban materials [12].

Detection of built-up areas, particularly during dry seasons, has received considerable attention [13]. Many techniques depend on spectral indices that emphasize the contrast between impervious surfaces and surrounding land in dry conditions [14]. To improve this separation, several indices have been developed to better differentiate built-up from non-built-up areas. One of the most widely used is the Normalized Difference Built-up Index (NDBI) [1], which utilizes the difference in reflectance between near-infrared (NIR) and short-wave infrared (SWIR) bands to provide a simple yet effective method for detecting urban areas [15]. However, NDBI has drawbacks, including confusion caused by mixed pixels, prompting the creation of alternative indices. In response, researchers such as Sameen and Pradhan [16] have worked to overcome the limitations of earlier methods. Ghosh et al. [15] evaluated several improved indices, including the Enhanced Built-up and Bareness Index (EBBI) [17], the Index-Based Built-up Index (IBI) [18], the Urban Index (UI) [19], and the Normalized Difference Bareness Index (NDBaI) [20]. In semi-arid environments, however, built-up surfaces frequently resemble dry soil, complicating accurate extraction.

To address this, the Automated Built-Up Extraction Index (ABEI) was introduced, offering improved performance in mapping urban areas using Landsat-8 Operational Land Imager (OLI) data [21]. Consequently, Rasul et al. [22] developed indices such as the Dry Built-up Index (DBI) and the Dry Bare-Soil Index (DBSI) to further enhance the distinction between built-up and bare land [14]. Additionally, Harrak et al. [23] evaluated several indices under dry-season conditions, including the Built-up and Low-Frequency Enhanced Index (BLFEI) [24], the Shortwave Infrared-Based Index for Urban Extraction (SWIRED) [25], and the Built-up Index (BUI) [26]. Their findings indicated that these indices achieved strong

separability even under low-moisture conditions. Aside from single-index methods, many spectral indices are sensitive to seasonal changes and often perform poorly during dry periods because of spectral overlap [27]. Built-up area indices often rely on global settlement masks to distinguish built-up from non-built surfaces [28]. The Global Urban Footprint offers such global built-up layers and is widely used as a reference for calculating and comparing built-up indicators [28, 29].

Multi-index approaches have proven to be more reliable, as they combine the strengths of several indices. For instance, Bouhennache et al. [24] proposed the Built-up Area Extraction Method (BAEM), which integrates outputs from indices like NDVI [30], MNDWI [31], and NDBI. Similarly, Xu [32] introduced thematic combinations of indices to bolster the resilience of built-up detection in diverse environments. Huang and Lu [33] demonstrated that such multi-index techniques, especially when trained on large datasets, are generally more accurate and adaptable than single-index models. Rouibah and Belabbas [34] used Sentinel-2 data to integrate the Normalized Difference Tillage Index (NDTI) [35], Bare Soil Index (BSI) [36], DBSI, and NDVI, applying unsupervised classification to assess built-up areas in semi-arid northeastern Algeria. Their combined strategy enhanced the accuracy of urban-area mapping. Likewise, band-ratio indices such as the Band Ratio for Built-up Area (BRBA) [37], BAEM [38], and the Visible Red Near-Infrared Built-up Index (VrNIR-BI) [39] have demonstrated promising results in arid and dry regions [14]. Furthermore, integrating the Urban Index (UI) [19] with NDBI using multi-temporal Landsat imagery has proven effective for tracking built-up changes over time, even during dry seasons [40]. The development of index-based methods for urban-area detection reflects a shift from basic spectral indices to more advanced, optimized, and multi-source techniques. Indices like ABEI [21] and EBBI [17], along with newer indices developed through optimization strategies, illustrate this evolution. As a result, selecting and combining indices based on the distinct spectral characteristics of dry-season imagery is crucial. This strategy helps improve both the precision and consistency of built-up area extraction. However, accurately identifying urban features in dry climates remains a challenge.

Despite the wide range of available indices, combining multiple spectral indices often introduces redundancy and collinearity within the feature space [41]. Instead of offering complementary information, these mixed compositions may produce overlapping predictors that reduce model efficiency. Such redundancy can negatively affect machine learning algorithms such as SVM, leading to reduced performance. Therefore, selecting relevant features is key to improving the performance of machine learning algorithms across different tasks [42]. However, while combining indices can enhance their descriptive capacity for urban classification, it may also amplify existing noise and errors in individual indices, thereby increasing uncertainty throughout the classification process [43]. A further limitation of this approach is the high computational cost and time needed to select suitable index combinations. When numerous indices are available, determining the most effective

mix, particularly in trichromatic setups, requires extensive testing, often involving supervised classification and accuracy evaluation [44]. This trial-and-error method can significantly prolong analysis, particularly in large or diverse study areas. Moreover, many index combinations yield only minimal improvements, leading to extended computation times without corresponding gains in accuracy [45].

To address these challenges, researchers have employed data reduction techniques such as Principal Component Analysis (PCA), which minimizes redundancy and emphasizes the most informative features in multispectral data [46]. When applied to original bands, PCA is mainly used for band selection and noise removal, helping reduce redundancy and improve the separability of classes [47]. However, limited research has examined the application of PCA to combined spectral indices, particularly in dry-season imagery, leaving this area under-investigated. In 2007, Xu [32] applied PCA to enhance urban area extraction from satellite imagery in China. By combining PCA with three normalized difference indices and spectral mixture analysis, the study improved urban land classification using Landsat data. PCA effectively reduced dimensionality while preserving key features associated with built-up areas. Xu also demonstrated that combining indices like SAVI [48] and MNDWI with PCA increased accuracy by better capturing urban surface characteristics. Similarly, studies by Bhatti and Tripathi [38] and Adeyemi et al. [49] confirmed PCA's value in urban mapping, particularly when used with multiple indices. This growing body of research supports PCA's potential to enhance urban classification by simplifying complex spectral data. For instance, PCA condenses spectral information, making critical features more distinguishable. This simplification improves classification accuracy, and PCA combined with SVM proved effective for urban land classification [50].

This study builds on previous work by proposing a hybrid method that applies PCA to five carefully selected spectral indices, each chosen for its specific role. DBSI, BRBA, and BSI were selected for their proven effectiveness in highlighting built-up features, particularly in semi-arid regions where built-up areas often resemble bare soil spectrally [32, 44]. NDTI was included to reduce confusion between dry soil and urban surfaces, a common challenge in such climates [51]. NDVI was used to identify and exclude vegetation, thereby focusing the classification on non-vegetated areas and enhancing overall accuracy. This research targets built-up area extraction during the driest month (August) using Sentinel-2 data, contributing to the field by introducing an index-based PCA approach specifically tailored to dry-season conditions. While individual indices have shown promise, their combined use as input for dimensionality reduction remains underexplored. Integrating these indices before applying PCA may improve class separation by combining thematic importance with statistical efficiency [52]. Additionally, a comparison with traditional index-based approaches using SVM classification highlights the practical advantages of the proposed method. After testing several trichromatic combinations, two were identified as the most effective for reducing confusion with bare soil

and vegetation: BRBA/NDTI/NDVI and DBSI/NDTI/NDVI. The results demonstrate superior classification accuracy, supporting the robustness of this approach. This study shows that combining multiple index layers, followed by PCA and SVM, outperforms current methods, achieving accuracies above 95%. The goal is to demonstrate the proposed approach's effectiveness and adaptability for reliable built-up area extraction under harsh seasonal and climatic conditions.

2. Materials and Methods

The research followed a structured methodological approach comprising image preprocessing, spectral index computation, dimensionality reduction, and supervised classification. The detailed workflow is presented in Figure 1.

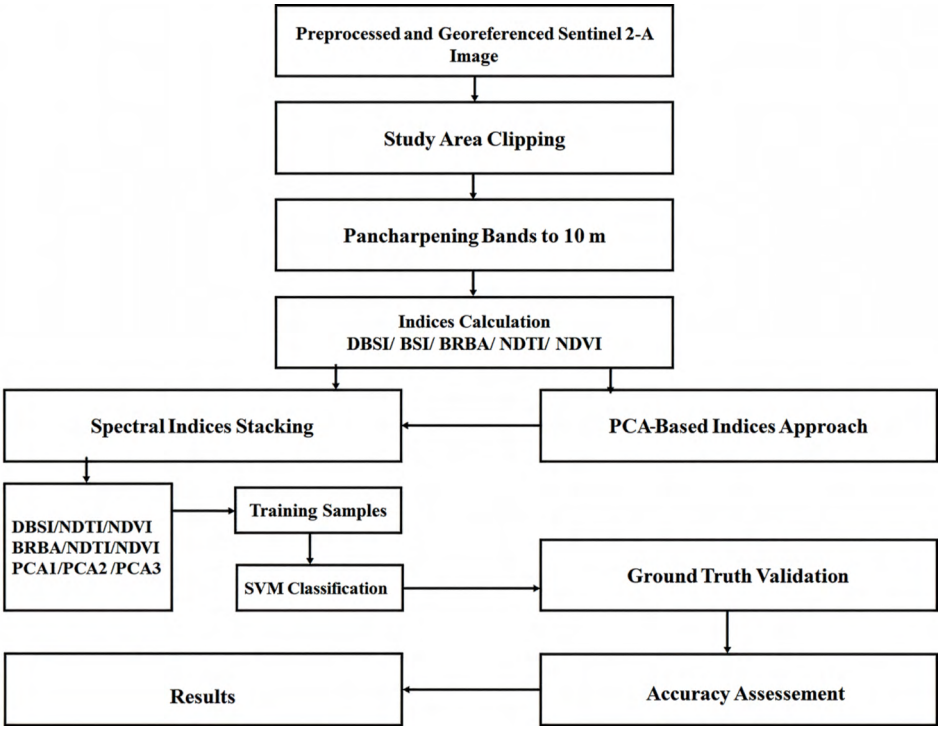

Fig. 1. Schematic representation of the methodology

2.1. Study Area

The study area is located in northeastern Algeria, within the Aurès Mountains, part of the Saharan Atlas range (Fig. 2). The elevation exceeds 980 m above sea level. Batna has become an important city and economic center in the region due to its

strategic location. The area has a semi-arid climate characterized by a distinct dry season. Average temperatures vary significantly throughout the year. In January, the lows are around 4°C, while in July, the highs can reach 35°C. Nighttime temperatures often fall below freezing in winter, while summer daytime temperatures can approach 40°C.

These climatic and topographic conditions significantly affect land use, which includes a wide range of urban zones, dry plains, forests, and mountainous terrains. Batna's built environment is primarily composed of concrete buildings, with rooftops made of concrete, tiles, or metal. The city center is densely developed, whereas the outskirts exhibit a more dispersed settlement pattern. The urban fabric integrates residential, commercial, and industrial zones, along with essential public services. Peri-urban expansion is increasingly noticeable, driven by population growth and economic development. This outward growth creates transitional zones that blend urban and rural characteristics, resulting in mixed land-use patterns.

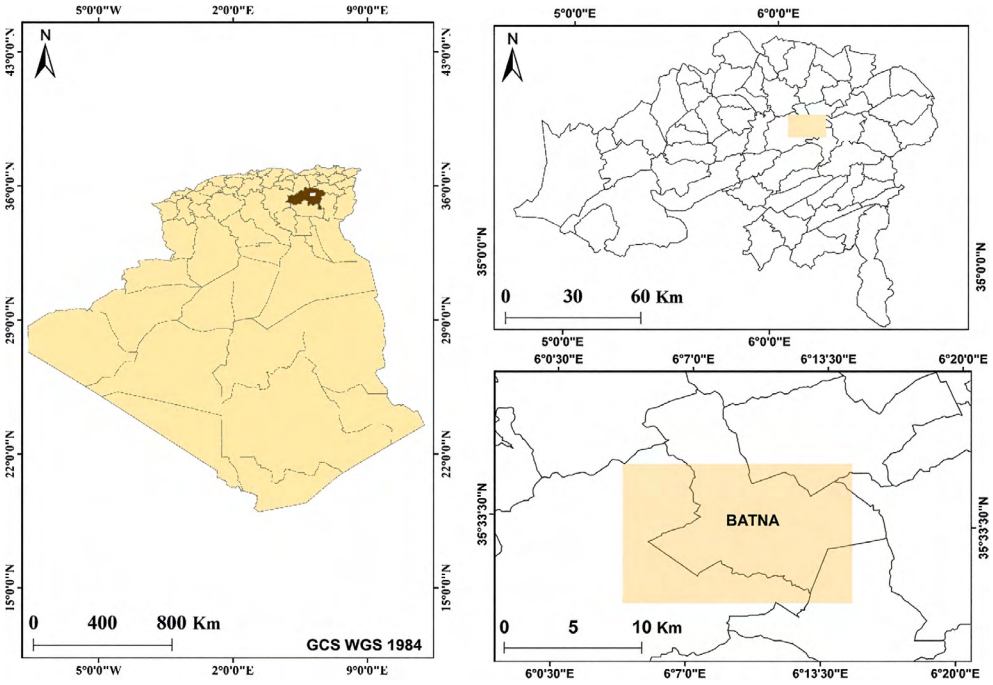

Fig. 2. Geographical location of Batna city

2.2. Data Collection

Sentinel-2A imagery was selected for its high spatial resolution and suitability for distinguishing various land cover classes. This dataset enables detailed analysis of surface characteristics, supporting the recognition of both vegetation types and

urban features [53]. The imagery was acquired from the Copernicus Open Access Hub. To ensure high image quality and minimize atmospheric distortions, scenes captured in August 2024, the driest month of the year, were chosen. This period reduces vegetation interference, enhancing the spectral distinction between built-up areas and exposed soil.

Images with cloud cover below 1% were prioritized, and only pre-orthorectified scenes in the WGS84 UTM Zone 32 coordinate system were used to ensure geometric accuracy. After downloading, all spectral bands were merged into a single stack. The resulting composite image was then clipped to the study area using its vector boundary in ArcGIS 10.8. To enhance spatial detail, pan-sharpening was applied to the multispectral bands using the nearest neighbor resampling method in ENVI 5.6, standardizing all bands to a 10-m spatial resolution.

2.3. Spectral Indices Calculation

Selecting appropriate spectral indices is a critical step in extracting urban areas using remote sensing, as it significantly affects the accuracy and consistency of classification findings. Moreover, both prior research and preliminary tests have demonstrated the effectiveness of these indices in semi-arid environments, supporting their relevance for this study. Accordingly, five spectral indices were selected, as presented in Table 1.

Table 1. Spectral indices calculated for the study

Index	Abbreviation	Formula on Sentinel-2A	Application	Reference
Band Ratio for Built-up Area	BRBA	(float(b4))/(float(b11))	Highlights built-up areas	[37]
Bare Soil Index	BSI	((float(b11) + float(b4)) − (float(b8) + float(b2)))/ ((float(b11) + float(b4)) + (float(b8) + float(b2)))	Enhances the contrast of bare soils	[36]
Dry Bare-Soil Index	DBSI	((float(b11) − float(b3))/ (float(b11) + float(b3))) − ((float(b8) − float(b4))/ (float(b8) + float(b4)))	Enhances the contrast of bare soils	[22]
Normalized Difference Tillage Index	NDTI	(float(b11) − float(b12))/ (float(b11) + float(b12))	Differentiates between bare soil and built-up areas	[35]
Normalized Difference Vegetation Index	NDVI	(float(b8) − float(b4))/ (float(b8) + float(b4))	Identifies vegetated regions	[30]

2.4. Feature Extraction

The spectral profile clearly indicates distinct separability among land cover types based on their reflectance patterns across the selected indices (Fig. 3).

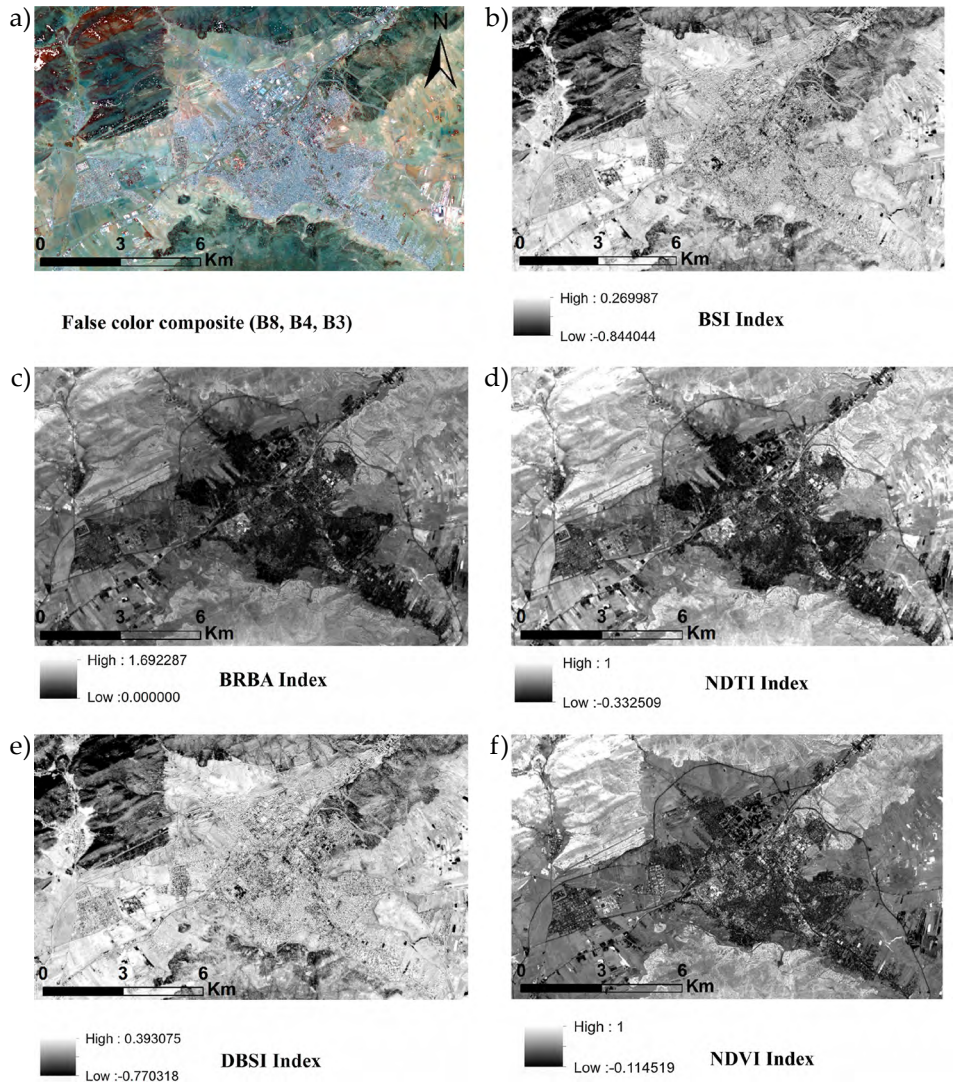

Fig. 3. Derived spectral indices, false color composition: a) false color composite; b) BSI Index; c) BRBA Index; d) NDTI Index; e) DBSI Index; f) NDVI Index

NDVI values are highest for forest (0.7) and grassland (0.6), reflecting dense vegetation cover, while built-up areas and bare soil exhibit much lower values (0.2 and 0.1, respectively). Indices like DBSI and BRBA are effective in distinguishing

non-vegetated surfaces. Bare soil records the highest values (0.7 for both indices), while built-up areas exhibit moderate values ranging from 0.4 to 0.5. NDTI further emphasizes soil exposure, with bare soil reaching values between 0.6 and 0.7, and grassland showing intermediate values around 0.3.

Two indices were consistently included in the band combinations: NDVI, which assists in identifying vegetated and forested areas, and NDTI, which increases the contrast between built-up surfaces and bare soil. Together, these indices facilitate the effective separation of four land cover classes: built-up areas, bare soil, forest, and grassland (Fig. 4).

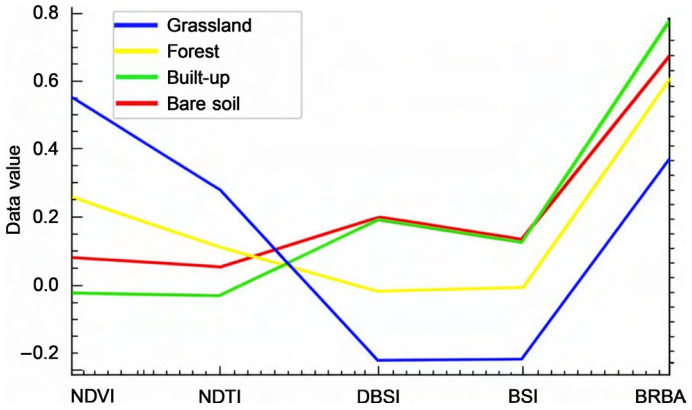

Fig. 4. Spectral profile

To identify which indices best complement NDVI and NDTI for built-up area extraction, a covariance analysis was performed to assess their correlation and overall effectiveness, as shown in Table 2.

Table 2. Correlation statistics of the selected indices

Index	NDVI	NDTI	DBSI	BSI	BRBA
NDVI	1.000000	0.784572	−0.884386	−0.886906	−0.777529
NDTI	0.784572	1.000000	−0.621572	−0.653740	−0.634788
DBSI	−0.884386	−0.621572	1.000000	0.986672	0.435025
BSI	−0.886906	−0.653740	0.986672	1.000000	0.441576
BRBA	−0.777529	−0.634788	0.435025	0.441576	1.000000

This correlation analysis ensured that the chosen indices provided unique information for classification, avoiding redundancy and improving the accuracy of built-up area extraction. NDVI and NDTI demonstrate a strong positive correlation (0.784572), indicating similar behavior in identifying vegetation and soil.

In contrast, DBSI shows a strong negative correlation with NDVI (−0.884386) and a moderate negative correlation with NDTI (−0.621572), whereas it shares a very high positive correlation with BSI (0.986672). This indicates that DBSI captures distinct yet relevant information useful for detecting built-up areas.

Although BRBA is responsive to urban features such as roads and buildings, it shows weak correlations with the other indices, making it less suitable for effective built-up area extraction. The negative correlation between DBSI and NDVI/NDTI underlines DBSI's independent contribution, offering critical information that enhances the separation between built-up and non-built surfaces. While the high correlation between DBSI and BSI (0.986672) suggests overlapping spectral characteristics of urban surfaces, the explained variance (eigenvalues) indicates that DBSI contributes more significantly than BSI, suggesting that BSI is comparatively less effective for this purpose.

Based on covariance analysis and basic statistical measurements (Table 3), numerous index combinations were tested to enhance the visual detection of built-up areas. Using image-derived statistics, visual contrast, and photo-interpretation, two combinations were identified as the most effective: DBSI/NDTI/NDVI and BRBA/NDTI/NDVI. These index sets provided the highest visual distinction between built-up areas and other land-cover types, supporting more accurate and reliable interpretation.

Table 3. Basic statistics of the selected indices for the study area

Index	Min	Max	Mean	St. dev.	Eigenvalue
NDVI	−0.114519	1.000000	0.145029	0.078530	0.016132
NDTI	−0.332509	1.000000	0.082526	0.036750	0.002996
DBSI	−0.770318	0.393075	0.154949	0.076215	0.000542
BSI	−0.844044	0.269987	0.113551	0.049844	0.000073
BRBA	0.000000	1.692287	0.643576	0.063036	0.000042

2.5. Dimensionality Reduction Using Principal Component Analysis (PCA)

The spectral similarity between urban surfaces and bare land in semi-arid areas can lead to classification errors. To reduce this confusion, PCA was applied using ENVI 5.6 to enhance the separation between built-up regions and soils with similar spectral features. This approach utilizes five selected spectral indices: NDTI, NDVI, DBSI, BRBA, and BSI (Fig. 5). Although each index highlights different features, their combination provides a more comprehensive representation of surface conditions.

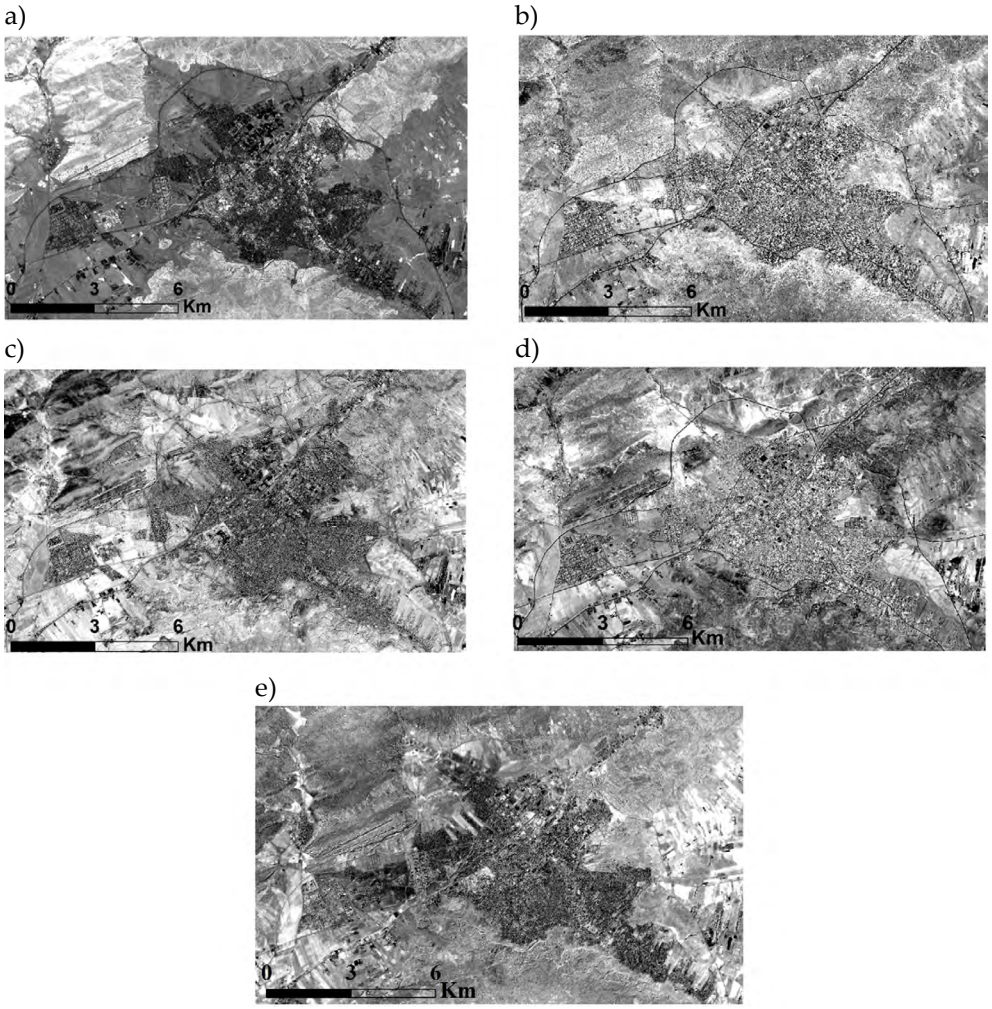

Fig. 5. Derived PCA of selected indices:
a) PCA1; b) PCA2; c) PCA3; d) PCA4; e) PCA5

Although the indices already reduce dimensionality, they remain partially correlated due to similar spectral characteristics of built-up areas and bare soils. PCA was therefore used to eliminate redundancy and combine the most informative content into fewer independent components. Crucially, the goal of applying PCA was not to test multiple configurations but to optimally combine all selected indices and avoid manually testing various trichromatic visualizations. The method retains only the components that most effectively address urban discrimination. Full-band PCA was initially tested but intentionally excluded due to its poorer performance in these environmental conditions.

By extracting three principal components (PC1 to PC3), the classification process was streamlined, enhancing speed and reducing issues of multicollinearity. Together, these components account for more than 92% of the total variance, preserving essential information from each index while eliminating overlap. Specifically, PC1 accentuates contrasts between built-up areas and bare soils within the SWIR bands, PC2 reflects the interaction between vegetation and impervious surfaces, and PC3 filters out short-term noise.

This reduction in data dimensions enhances the clarity of distinguishing features, thereby increasing classification accuracy during the validation process. The three principal components contribute to class separation as follows:

- PC1: Captures urban-bare soil contrasts;
- PC2: Separates vegetated (forests, active agriculture) from non-vegetated classes (urban, bare land);
- PC3: Suppresses noise from seasonal variations and temporary soil moisture in agricultural fields.

The interpretation of PCA components involved examining eigenvectors to determine each input index's contribution and visually inspecting the PCA output images to evaluate which surface features were highlighted. This method provides an approximate yet reliable understanding of the main spectral patterns for each component.

2.6. Classification Method

In this study, classification was performed using Support Vector Machines (SVMs) to separate satellite imagery into built-up and non-built-up classes. SVM was chosen for its demonstrated accuracy in remote sensing applications. First, three different feature combinations were prepared. For each combination, the corresponding three bands were stacked into a single dataset: (1) DBSI, NDTI, and NDVI; (2) BRBA, NDTI, and NDVI; and (3) the first three principal components (PCA1, PCA2, PCA3). For the third dataset, PCA was applied to reduce dimensionality before classification. Finally, SVM was used to classify all prepared datasets, producing built-up and non-built-up maps.

2.7. Validation Using Ground Truth Data

For this classification process, high-resolution images from Google Earth as of May 2, 2024, were used to derive the training samples, as they provide reliable visual information. Their positional accuracy was verified against the official cadastral orthophoto and assessed using stable features such as road intersections and building corners. The residual misalignment was less than one Sentinel-2A pixel, confirming the images' suitability for quality control. Training samples were manually delineated to ensure accurate spatial outlines for both built-up and non-built-up

categories. The built-up class included features such as buildings, roads, industrial sites, military zones, and other impervious surfaces. In contrast, the non-built-up category encompassed vegetation, bare soil, and forested areas.

High-resolution pure polygons were delineated using Google Earth imagery. The spectral values for training were derived solely from Sentinel-2A pixels entirely within these polygons, with a safety margin applied to prevent edge mixing. To enhance the reliability of these samples, additional resources, including land use data, expert judgment, and local field knowledge, were consulted. Although the manual sampling approach required considerable time, it provided a level of accuracy that automated methods may not achieve. Pure training polygons were created in varied shapes and sizes to reflect land-cover uniformity and to align with the spatial scale of the analysis, reducing spectral ambiguity and ensuring representative coverage across the study area.

3. Results

The PCA results (Fig. 6) reveal that the first three components (PC1–PC3) contain the highest proportion of variance within the dataset. PC1 shows the largest eigenvalue, followed by PC2 and PC3. In contrast, PC4 and PC5 show minimal eigenvalues (0.000072 and 0.000028, respectively), reflecting a negligible contribution to the overall variance.

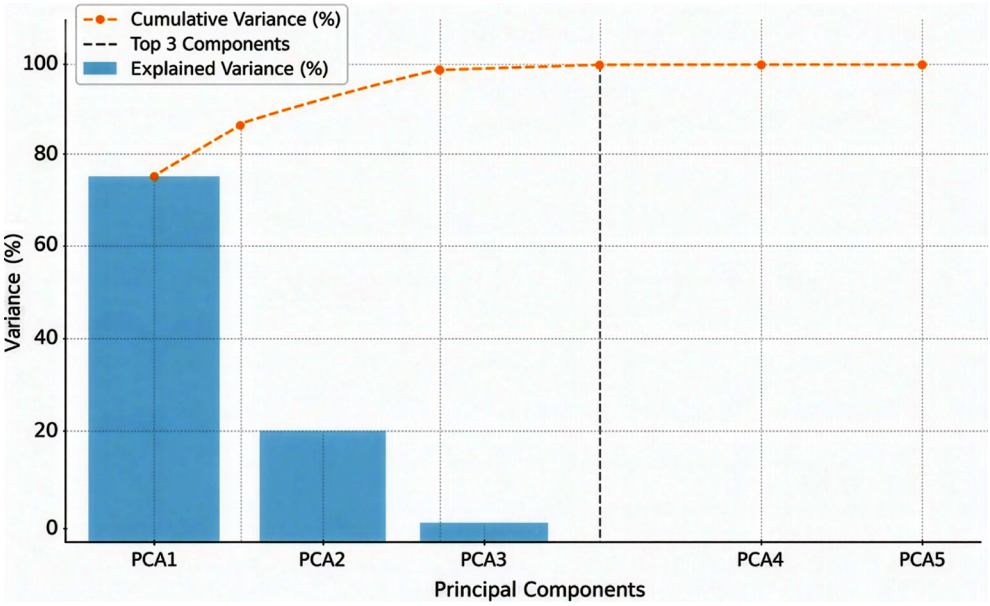

Fig. 6. Scree plot with cumulative variance

The first three components account for 96.69% of the total variance, as shown by both the scree plot (Fig. 6) and the explained variance table (Table 4). Therefore, these components are sufficient for accurate data representation, with no significant loss of meaningful information.

Table 4. PCA component statistics of selected indices

PCA Components	Eigenvalue	Explained Variance [%]	Cumulative Variance [%]
PCA1	0.011235	72.37	72.37
PCA2	0.003338	21.51	93.88
PCA3	0.000435	2.81	96.69
PCA4	0.000072	0.46	97.15
PCA5	0.000028	0.18	97.33

Based on these results, the comparative analysis of the three classification approaches revealed notable differences in their performance for urban-area extraction, as presented in Table 5.

Table 5. Accuracy assessment

Used dataset	DBSI/NDTI/NDVI [%]	BRBA/NDTI/NDVI [%]	PCA-based index [%]
Overall accuracy	93.9856	92.6917	95.2130
Kappa coefficient	0.8796	0.8536	0.9042

The PCA-based classification using the first three principal components (PCA1, PCA2, PCA3) achieved the highest overall accuracy at 95.21%, accompanied by a strong kappa coefficient of 0.9042, indicating excellent agreement with the reference data. The trichromatic combination of DBSI/NDTI/NDVI followed closely, with an overall accuracy of 93.99% and a kappa value of 0.8796. In comparison, the BRBA/NDTI/NDVI trichromatic method demonstrated slightly lower performance, with an accuracy of 92.69% and a kappa coefficient of 0.8536.

The locations of the accuracy assessment points are shown in Figure 7, indicating the exact positions used for validation across all land cover classes

An examination of class-specific accuracies (Table 6) reveals that the non-built-up category exhibited stable producer's accuracy across all methods, ranging narrowly between 95.82% and 95.90%. However, its user's accuracy improved progressively, increasing from 91.25% with the BRBA/NDTI/NDVI method to 94.90% using the PCA123 approach.

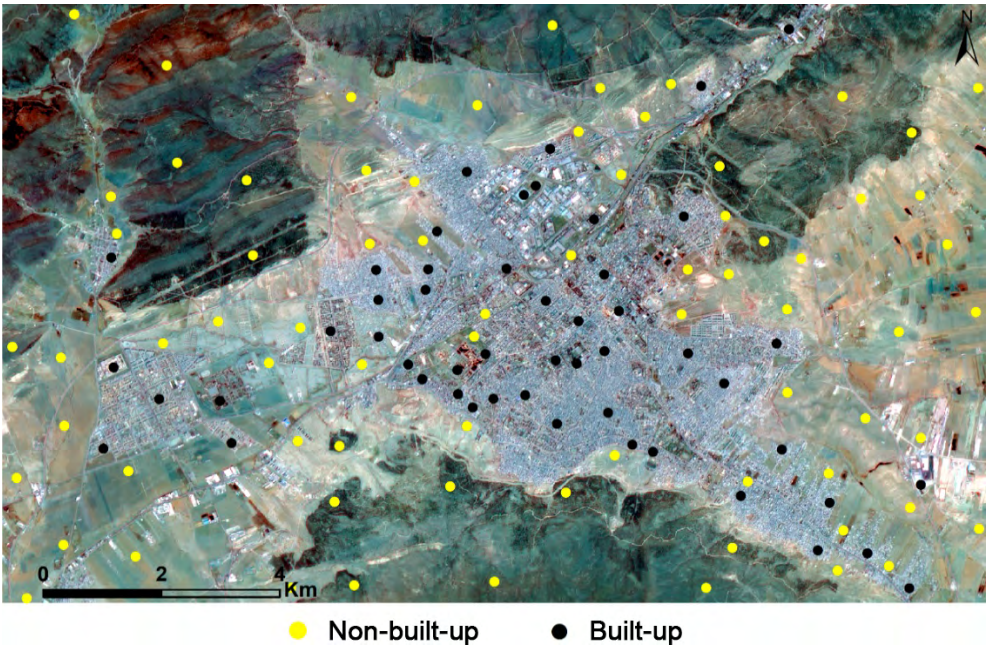

Fig. 7. Accuracy assessment points

Table 6. Study area map accuracy results

Composite Band/Index	Class	Commission Error [%]	Omission Error [%]	Producer's Accuracy [%]	User's Accuracy [%]
DBSI/NDTI/NDVI	Non-built-up Built-up	6.81 5.14	4.76 7.33	95.24 92.67	93.19 94.86
BRBA/NDTI/NDVI	Non-built-up Built-up	8.75 5.65	5.14 9.60	94.86 90.40	91.25 94.35
PCA-based index	Non-built-up Built-up	5.10 4.45	4.18 5.43	95.82 94.57	94.90 95.55

Similarly, the built-up class showed improvements in producer's accuracy, rising from 92.67% in the DBSI/NDTI/NDVI method to 94.57% with the PCA-based classification. The highest user's accuracy for the built-up class was also achieved by PCA1, PCA2, and PCA3 at 95.55%.

Error analysis further supports these findings. Commission errors for the non-built-up class decreased from 8.75% to 5.10%, while omission errors for the built-up class dropped from 9.60% to 5.43% when applying PCA1, PCA2, and PCA3. This notable reduction in classification errors highlights the superior discriminatory power of PCA-based dimensionality reduction compared to direct spectral index combinations (Fig. 8).

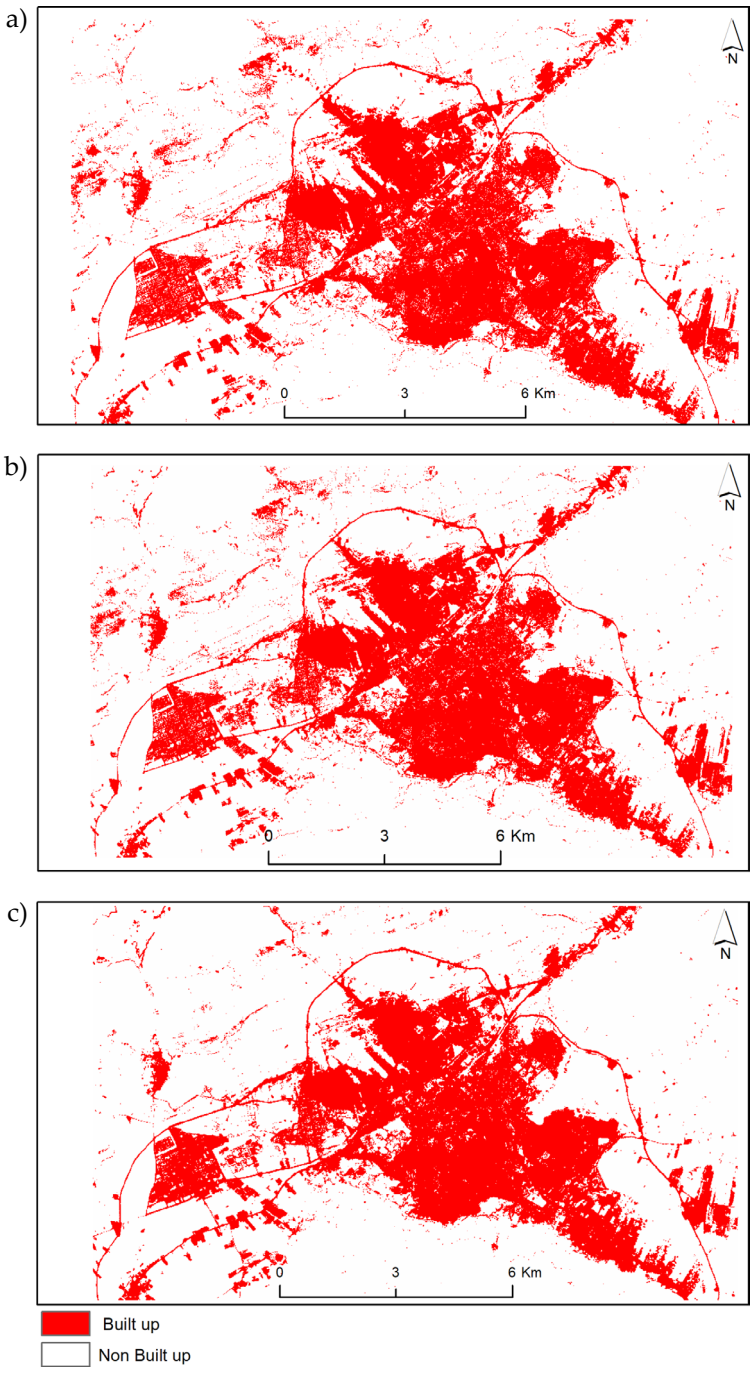

Fig. 8. Binary images of the SVM classifier: a) multi-index BRBA/NDTI/NDVI; b) multi-index DBSI/NDTI/NDVI; c) PCA-based approach

4. Discussion

This study compared three classification methods for extracting built-up areas using Sentinel-2A imagery in a semi-arid zone. The methods included two spectral index combinations, DBSI/NDTI/NDVI and BRBA/NDTI/NDVI, as well as a PCA-based approach applied to five selected indices.

The classification results revealed notable differences in performance, with the PCA-based method consistently achieving the highest accuracy, as presented in Table 5. PCA achieved the highest overall accuracy (95.21%) and kappa coefficient (0.9042), indicating strong agreement with ground truth. The built-up class also recorded a producer's accuracy of 94.57%, surpassing the index-based approaches. These findings highlight PCA's effectiveness in capturing spectral variability, reducing overlap, and enhancing separability between built-up and non-built-up areas.

To further evaluate classification accuracy, commission and omission errors were analyzed. For the non-built-up class, the PCA method exhibited a commission error of 5.10% and an omission error of 4.18%. For the built-up class, commission and omission errors were 4.45% and 5.43%, respectively. Although these figures were close to those of the index-based approaches, the PCA method still outperformed them, as reflected in its higher overall accuracy and kappa score.

Area estimation further supported the classification results (Table 7). The PCA-based method indicated 44.13 km^2 of built-up land, representing 26.10% of the total area. In comparison, the DBSI/NDTI/NDVI method produced results similar to those of PCA, demonstrating its effectiveness for built-up area extraction. The DBSI/NDTI/NDVI combination outperformed the BRBA-based method.

Table 7. SVM classification for built-up area

Composite Band/Index	Built-up [km^2]	Non-built-up [km^2]	Built-up Area [%]	Non-built-up Area [%]
DBSI/NDTI/NDVI	44.1942	124.9329	26.14	73.86
BRBA/NDTI/NDVI	43.1967	125.9304	25.55	74.45
PCA-based index	44.1287	124.9984	26.10	73.90

This finding suggests that the first group of indices delineates built-up surfaces more clearly. However, in mountainous regions, it occasionally confuses built-up areas with rocky outcrops or bare soil. Sparse vegetation and rugged terrain contribute to these errors, as exposed rock and compacted earth often reflect light similarly to artificial materials.

Conversely, the BRBA/NDTI/NDVI method extracted 43.20 km^2 of built-up land, slightly less than the DBSI-based result. This lower figure indicates that BRBA is less

reliable in complex or mixed land covers, leading to a higher rate of misclassified pixels. The built-up share was 25.55%, suggesting that this method may underestimate urban areas in such environments. The better performance of DBSI/NDTI/NDVI results from combining indices that target built-up features (DBSI), soil types (NDTI), and vegetation masking (NDVI). These findings align with those of Ettehadi Osgouei et al. [51], who also demonstrated the effectiveness of NDTI in differentiating built-up areas from bare soils within Mediterranean urban environments.

Nonetheless, even with this improved combination, some confusion persists in mountainous terrain. Features such as rock or dry soil can still be misidentified as built-up land due to spectral similarities. On the other hand, the PCA approach identified 44.13 km^2 of urban area, slightly higher than the DBSI/NDTI/NDVI result. With 26.10% built-up and 73.90% non-built-up, PCA provided a clearer distinction between land classes. This clarity was particularly evident in bare regions with little vegetation. PCA performs well in these areas by reducing data complexity and capturing key spectral patterns associated with materials and surface textures. Unlike individual indices, PCA detects subtle differences between urban and natural surfaces, such as concrete versus weathered rock, leading to better results in challenging settings.

Moreover, PCA results showed well-defined built-up zones, clearly distinct from bare soils and rocks (Fig. 8). In contrast, index-based maps suffered from mixed pixels. PCA extracts essential data from multiple indices, removes redundant signals, and enhances the visibility of urban features, similar to the conclusions drawn by Xu [32], who underscores the utility of thematic index combinations and PCA-based methods for advanced urban feature extraction. These results suggest that PCA could also support future deep learning applications aimed at refining classification tasks. Applying PCA to NDVI, NDTI, DBSI, BRBA, and BSI proved highly effective.

Combining all five indices allowed the method to capture complementary signals. The method was particularly successful in bare zones, where minimal vegetation often leads to misclassification. Additionally, BSI, which was excluded from the traditional trichromatic composites, added value within the PCA framework. Its brightness component contributed additional variance not captured by the other indices. By reducing dimensionality, PCA isolated key spectral patterns linked to materials and textures. Although materials like concrete, asphalt, or metal exhibit significant spectral variations, these differences are often masked by visually similar features, such as dry rock or compacted ground. PCA highlighted these distinctions, resulting in more accurate urban mapping.

In summary, combining multiple indices through PCA not only increased accuracy but also improved urban detection by leveraging the strength of each index while minimizing its individual limitations. These results demonstrate that PCA effectively extracts the most informative spectral features, reduces noise, and enhances the distinction between built-up and non-built-up areas. Therefore, applying PCA

to all five indices represents a more robust classification approach than relying on fixed index combinations alone. Future research should explore integrating PCA with deep learning models and higher-resolution data to further improve urban area detection.

5. Conclusion

This study focused on detecting built-up areas in Batna, Algeria, a semi-arid region, using Sentinel-2A imagery. Batna was selected for its diverse landscape, which includes urban settlements, farmland, forest, and bare soil, conditions that complicate land cover classification. Additionally, the region's semi-arid climate causes seasonal shifts in vegetation and soil reflectance, necessitating careful selection of both spectral indices and classification methods.

To address these classification challenges, several approaches were examined. Initially, well-known spectral indices such as DBSI, NDTI, NDVI, BRBA, and BSI were tested, as they are commonly used to detect built-up areas. A covariance-based statistical analysis identified DBSI, NDTI, and NDVI as the most effective combination, which was subsequently evaluated using the SVM algorithm. This trio delivered strong classification results, with the highest accuracy reaching 93%. However, further refinement was still necessary.

To enhance classification results, PCA was employed to reduce data redundancy and increase spectral contrast. The PCA123 configuration, consisting of the first three principal components, was tested and achieved the highest classification accuracy of 95.21%. This result demonstrates the benefits of dimensionality reduction for distinguishing built-up areas from other land-cover features.

Throughout the study, multiple classification strategies and input configurations were systematically compared, leading to consistent improvements in accuracy. While the combination of BRBA, NDTI, and NDVI was also tested, it yielded weaker results. Ultimately, the combination of PCA with well-chosen spectral indices proved most effective for mapping built-up areas in Batna's semi-arid environment.

The findings offer valuable insights for urban monitoring in arid and semi-arid zones, where the spectral similarity between man-made and natural surfaces poses challenges for accurate mapping. Nevertheless, some limitations were observed, particularly in areas where different land covers overlap spectrally, or because of the moderate spatial resolution of Sentinel-2A.

Looking ahead, future studies should explore the integration of PCA with advanced classifiers, such as deep learning models, and assess the effectiveness of higher-resolution satellite data. Applying this method across diverse climate regions and urban forms could improve its reliability and extend its applicability to broader-scale urban land cover mapping efforts.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT Author Contribution

W. T.: methodology, conceptualization, software, validation, analysis, investigation, resources, data curation, writing – original draft preparation, visualization.

M. K.: methodology, validation, conceptualization, supervision, review, and editing.

L. K.: validation, review, and editing.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data have been deposited on Zenodo and are available via DOI: https://doi.org/10.5281/zenodo.17968634.

Use of AI-assisted tools

During the preparation of this manuscript, the authors used AI-assisted tools, specifically Grammarly and QuillBot, to improve the clarity, grammar, and overall language quality of the English text. These tools were not used to generate scientific content, perform data analysis, or interpret results. All scientific content and interpretations were solely developed and verified by the authors.

References

[1] Zha Y., Gao J., Ni S.: *Use of normalized difference built-up index in automatically mapping urban areas from TM imagery*. International Journal of Remote Sensing, vol. 24(3), 2003, pp. 583–594. https://doi.org/10.1080/01431160304987.

[2] Chen Y., Yao S., Hu Z., Huang B., Miao L., Zhang J.: *Built-up area extraction combing densely connected dual-attention network and multiscale context*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, 2023, pp. 5128–5143. https://doi.org/10.1109/JSTARS.2023.3281363.

[3] Wellmann T., Lausch A., Andersson E., Knapp S., Cortinovis C., Jache J., Scheuer S., Kremer P., Mascarenhas A., Kraemer R., Haase A., Schug F., Haase D.: *Remote sensing in urban planning: Contributions towards ecologically sound policies?* Landscape and Urban Planning, vol. 204, 2020, 103921. https://doi.org/10.1016/j.landurbplan.2020.103921.

[4] John S., Varghese A.O.: *Analysis of support vector machine and maximum likelihood classifiers in land cover classification using Sentinel-2 images*. Proceedings of the Indian National Science Academy, vol. 88(2), 2022, pp. 213–227. https://doi.org/10.1007/s43538-022-00078-1.

[5] Saha J., Khanna Y., Mukhopadhyay J., Aikat S.: *From supervised to unsupervised learning for land cover analysis of Sentinel-2 multispectral images*, [in:] *IGARSS 2020 – 2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 2020*, IEEE, 2020, pp. 1965–1968. https://doi.org/10.1109/IGARSS39084.2020.9323985.

[6] Leyk S., Uhl J.H.: *Data descriptor: HISDAC-US, historical settlement data compilation for the conterminous United States over 200 years*. Scientific Data, vol. 5, 2018, 180175. https://doi.org/10.1038/sdata.2018.175.

[7] Uhl J.H., Leyk S., Chiang Y.Y., Duan W., Knoblock C.A.: *Automated extraction of human settlement patterns from historical topographic map series using weakly supervised convolutional neural networks*. IEEE Access, vol. 8, 2020, pp. 6978–6996. https://doi.org/10.1109/ACCESS.2019.2963213.

[8] Sirko W., Kashubin S., Ritter M., Annkah A., Bouchareb Y.S.E., Dauphin Y., Keysers D., Neumann M., Cisse M., Quinn J.: *Continental-scale building detection from high resolution satellite imagery*. arXiv, 2021. http://arxiv.org/abs/2107.12283.

[9] Szubert P., Kaim D., Kozak J.: *Dataset of building locations in Poland in the 1970s and 1980s*. Scientific Data, vol. 11(1), 2024, 341. https://doi.org/10.1038/s41597-024-03179-2.

[10] Szubert P., Kaim D., Kozak J.: *Changes in building numbers and density in Poland since the 1970s*. Journal of Maps, vol. 21(1), 2025, 2492622. https://doi.org/10.1080/17445647.2025.2492622.

[11] Uhl J.H., Leyk S., Li Z., Duan W., Shbita B., Chiang Y.Y., Knoblock C.A.: *Combining remote-sensing-derived data and historical maps for long-term back-casting of urban extents*. Remote Sensing, vol. 13(18), 2021, 3672. https://doi.org/10.3390/rs13183672.

[12] Taghdis S., Farpoor M.H., Mahmoodabadi M.: *Pedological assessments along an arid and semi-arid transect using soil spectral behavior analysis*. CATENA, vol. 214, 2022, 106288. https://doi.org/10.1016/j.catena.2022.106288.

[13] Masiero É., Christoforo A.L., Kowalski L.F., Fernandes M.E.: *Urban morphology and prediction models of microclimatic phenomena in dry atmospheric context*. Revista Brasileira de Climatologia, vol. 31, 2022, pp. 259–284. https://doi.org/10.55761/abclima.v31i18.15707.

[14] Almohamad H., Alshwesh I.O.: *Evaluation of index-based methods for impervious surface mapping from Landsat-8 to Cities in dry climates; A case study of Buraydah City, KSA*. Sustainability, vol. 15(12), 2023, 9704. https://doi.org/10.3390/su15129704.

[15] Ghosh D.K., Mandal A.C., Majumder R., Patra P., Bhunia G.S.: *Analysis for mapping of built-up area using remotely sensed indices – a case study of Rajarhat block in Barasat Sadar sub-division in West Bengal (India).* Journal of Landscape Ecology, vol. 11(2), 2018, pp. 67–76. https://doi.org/10.2478/jlecol-2018-0007.

[16] Sameen M.I. Pradhan B.: *A novel built-up spectral index developed by using multiobjective particle-swarm-optimization technique*. IOP Conference Series: Earth and Environmental Science, vol. 37(1), 2016, 012006. https://doi.org/10.1088/1755-1315/37/1/012006.

[17] As-syakur A.R., Adnyana I.W.S., Arthana I.W., Nuarsa I.W.: *Enhanced built-up and bareness index (EBBI) for mapping built-up and bare land in an urban area*. Remote Sensing, vol. 4(10), 2012, 10. https://doi.org/10.3390/rs4102957.

[18] Xu H.: *A new index for delineating built-up land features in satellite imagery*. International Journal of Remote Sensing, vol. 29(14), 2008, pp. 4269–4276. https://doi.org/10.1080/01431160802039957.

[19] Kawamura M., Jayamana S., Tsujiko Y.: *Relation between social and environmental conditions in Colombo, Sri Lanka and the urban index estimated by satellite remote sensing data*. International Archives of the Photogrammetry and Remote Sensing Data, vol. XXXI-B7, 1996, pp. 321–326.

[20] Zhao H.-M., Chen X.: *Use of normalized difference bareness index in quickly mapping bare areas from TM/ETM+ imagery*, [in:] *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05., Seoul, 2005*, IEEE, 2005, pp. 1666–1668. https://doi.org/10.1109/IGARSS.2005.1526319.

[21] Firozjaei M.K., Sedighi A., Kiavarz M., Qureshi S., Haase D., Alavipanah S.K.: *Automated built-up extraction index: A new technique for mapping surface built-up areas using Landsat 8 OLI imagery*. Remote Sensing, vol. 11(17), 2019, 1966. https://doi.org/10.3390/rs11171966.

[22] Rasul A., Balzter H., Ibrahim G.R.F., Hameed H.M., Wheeler J., Adamu B., Ibrahim S., Najmaddin P.M.: *Applying built-up and bare-soil indices from Landsat 8 to cities in dry climates*. Land, vol. 7(3), 2018, 81. https://doi.org/10.3390/land7030081.

[23] Harrak Y., Rachid A., Aguejdad R.: *Evaluation of spectral indices and global thresholding methods for the automatic extraction of built-up areas: An application to a semi-arid climate using Landsat 8 imagery*. Urban Science, vol. 9(3), 2025, 78. https://doi.org/10.3390/urbansci9030078.

[24] Bouhennache R., Bouden T., Taleb-Ahmed A., Cheddad A.: *A new spectral index for the extraction of built-up land features from Landsat 8 satellite imagery*. Geocarto International, vol. 34(14), 2019, pp. 1531–1551. https://doi.org/10.1080/10106049.2018.1497094.

[25] Capolupo A., Monterisi C., Tarantino E.: *Landsat images classification algorithm (LICA) to automatically extract land cover information in Google Earth engine environment.* Remote Sensing, vol. 12(7), 2020, 1201. https://doi.org/10.3390/rs12071201.

[26] Kaimaris D., Patias P.: *Identification and area measurement of the built-up area with the built-up index (BUI)*. International Journal of Advanced Remote Sensing and GIS, vol. 5(1), 2016, pp. 1844–1858. https://doi.org/10.23953/cloud.ijarsg.64.

[27] Kato A., Carlson K.M., Miura T.: *Assessing the inter-annual variability of vegetation phenological events observed from satellite vegetation index time series in dryland sites*. Ecological Indicators, vol. 130, 2021, 108042. https://doi.org/10.1016/j.ecolind.2021.108042.

[28] Esch T., Heldens W., Hirner A., Keil M., Marconcini M., Roth A., Zeidler J., Dech S., Strano E.: *Breaking new ground in mapping human settlements from space – the Global Urban Footprint*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 134, 2017, pp. 30–42. https://doi.org/10.1016/j.isprsjprs.2017.10.012.

[29] Marconcini M., Metz-Marconcini A., Üreyen S., Palacios-Lopez D., Hanke W., Bachofer F., Zeidler J., Esch T., Gorelick N., Kakarla A., Paganini M., Strano E.: *Outlining where humans live, the World Settlement Footprint 2015*. Scientific Data, vol. 7(1), 2020, 242. https://doi.org/10.1038/s41597-020-00580-5.

[30] Rouse J.W., Haas R.H., Schell J.A., Deering D.W.: *Monitoring vegetation systems in the Great Plains with ERTS*, [in:] *Third Earth Resources Technology Satellite-l Symposium. Volume I: Technical Presentations Section A. Goddard Space Flight Center: Washington, D.C, December 10–14, 1973*, Paper A-20, National Aeronautics and Space Administration, Washington 1974, pp. 309–317.

[31] Xu H.: *Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery*. International Journal of Remote Sensing, vol. 27(14), 2006, pp. 3025–3033. https://doi.org/10.1080/01431160600589179.

[32] Xu H.: *Extraction of urban built-up land features from Landsat imagery using a thematicoriented index combination technique*. Photogrammetric Engineering & Remote Sensing, vol. 73(12), 2007, pp. 1381–1391. https://doi.org/10.14358/PERS.73.12.1381.

[33] Huang X., Lu Q., Zhang L.: *A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas*. ISPRS Journal of Photogrammetric Engineering and Remote Sensing, vol. 90, 2014, pp. 36–48. https://doi.org/10.1016/j.isprsjprs.2014.01.008.

[34] Rouibah K., Belabbas M.: *Applying multi-index approach from Sentinel-2 imagery to extract urban area in dry season (semi-arid land in North East Algeria)*. Revista de Teledetección, no. 56, 2020, pp. 89–101. https://doi.org/10.4995/raet.2020.13787.

[35] van Deventer A.P., Ward A.D., Gowda P.H., Lyon J.G.: *Using thematic mapper data to identify contrasting soil plains and tillage practices*. Photogrammetric Engineering and Remote Sensing, vol. 63(1), 1997, pp. 87–93.

[36] Rikimaru A., Miyatake S.: *Development of forest canopy density mapping and monitoring model using indices of vegetation, bare soil and shadow*, [in:] *Proceedings of the 18th Asian Conference on Remote Sensing: ACRS 1997: 20–24 October, 1997, Kuala Lumpur, Malaysia*, Malaysian Centre for Remote Sensing, Kuala Lumpur 1997. https://acrs-aars.org/proceeding/ACRS1997/Papers/FR97-5.htm.

[37] Waqar M.M., Mirza J. F., Mumtaz R., Hussain E.: *Development of new indices for extraction of built-up area & bare soil from Landsat data*. Open Access Scientific Reports, vol. 1(1), 2012, pp. 1–4.

[38] Bhatti S.S., Tripathi N.K.: *Built-up area extraction using Landsat 8 OLI imagery*. GIScience & Remote Sensing, vol. 51(4), 2014, pp. 445–467. https://doi.org/10.1080/15481603.2014.939539.

[39] Estoque R.C., Murayama Y.: *Classification and change detection of built-up lands from Landsat-7 ETM+ and Landsat-8 OLI/TIRS imageries: A comparative assessment of various spectral indices*. Ecological Indicator, vol. 56, 2015, pp. 205–217. https://doi.org/10.1016/j.ecolind.2015.03.037.

[40] Ichsan Ali M., Hafid Hasim A.H.H., Raiz Abidin M.: *Monitoring the built-up area transformation using urban index and normalized difference built-up index analysis*. International Journal of Engineering, vol. 32(5), 2019, pp. 647–653. https://doi.org/10.5829/ije.2019.32.05b.04.

[41] Feng J., Zhang J., Zhang Y.: *A multiview spectral–spatial feature extraction and fusion framework for hyperspectral image classification*. IEEE Geoscience and Remote Sensing Letters, vol. 19, 2022, 5504805. https://doi.org/10.1109/LGRS.2021.3066613.

[42] Heidary K., Atluri V., Bland J.: *Performance evaluation of machine learning algorithms in reduced dimensional spaces*. Journal of Cyber Security, vol. 6(1), 2024, pp. 69–87. https://doi.org/10.32604/jcs.2024.051196.

[43] Ikokou G.B., Smit J.L.: *Optimizing the selection of spatial and non-spatial data for higher accuracy multi-scale classification of urban environments*, [in:] Wade S. (ed.), *Earth Observations and Geospatial Science in Service of Sustainable Development Goals*, Springer, Cham 2019, pp. 161–169. https://doi.org/10.1007/978-3-030-16016-6_14.

[44] Dib S., Souiher N., Bengusmia D.: *Extraction of urban areas using spectral indices combination and Google Earth Engine in Algerian Highlands (case study: cities of Djelfa, Messaad, Ain Oussera)*. Anuário do Instituto de Geociências, vol. 45, 2022, 44537. https://doi.org/10.11137/1982-3908_2022_45_44537.

[45] Montero D., Aybar C., Mahecha M.D., Martinuzzi F., Söchting M., Wieneke S.: *A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research*. Scientific Data, vol. 10(1), 2023, 197. https://doi.org/10.1038/s41597-023-02096-0.

[46] Jensen J.R.: *Introductory Digital Image Processing: A Remote Sensing Perspective* (3rd ed.). Prentice Hall, Upper Saddle River 2005.

[47] Sehgal S., Ahuja L., Bindu M.H.: *A novel approach for band selection using virtual dimensionality estimate and principal component analysis for satellite image classification*. International Journal of Intelligent Information Technologies (IJIIT), vol. 18(2), 2022, pp. 1–16. https://doi.org/10.4018/IJIIT.296272.

[48] Huete A.R.: *A soil-adjusted vegetation index (SAVI)*. Remote Sensing of Environment, vol. 25(3), 1988, pp. 295–309. https://doi.org/10.1016/0034-4257(88)90106-X.

[49] Adeyemi A., Ramoelo A., Cho M., Masemola C.R.: *Spectral index to improve the extraction of built-up area from WorldView-2 imagery*. Journal of Applied Remote Sensing, vol. 15(2), 2021, 024510. https://doi.org/10.1117/1.JRS.15.024510.

[50] Hazaymeh K., Mosleh M.K., Al-Rawabdeh A.M.: *A combined PCA-SIs classification approach for delineating built-up area from remote sensing data*. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, vol. 87(3), 2019, pp. 91–102. https://doi.org/10.1007/s41064-019-00071-2.

[51] Ettehadi Osgouei P., Kaya S., Sertel E., Alganci U.: *Separating built-up areas from bare land in Mediterranean cities using Sentinel-2A imagery*. Remote Sensing, vol. 11(3), 2019, 345. https://doi.org/10.3390/rs11030345.

[52] Arsa D.M.S., Sanabila H.R., Rachmadi M.F., Gamal A., Jatmiko W.: *Improving principal component analysis performance for reducing spectral dimension in hyperspectral image classification*, [in:] *2018 International Workshop on Big Data and Information Security (IWBIS), Jakarta, Indonesia, 2018*, IEEE, 2018, pp. 123–128. https://doi.org/10.1109/IWBIS.2018.8471705.

[53] Phiri D., Simwanda M., Salekin S., Nyirenda V.R., Murayama Y., Ranagalage M.: *Sentinel-2 data for land cover/use mapping: A review*. Remote Sensing, vol. 12(14), 2020, 14. https://doi.org/10.3390/rs12142291.