



Enhancing Land Use Classification Accuracy Through High-Resolution GIS and Remote Sensing: A Case Study of Urban Sprawl in Baghdad



Rasha Flayyih Hasan

Training and Workshop Center, University of Technology–Iraq, 10066 Baghdad, Iraq

* Correspondence: Rasha Flayyih Hasan (rasha.f.hasan@uotechnology.edu.iq)

Received: 06-20-2025

Revised: 07-30-2025

Accepted: 08-05-2025

Citation: R. F. Hasan, “Enhancing land use classification accuracy through high-resolution GIS and remote sensing: A case study of urban sprawl in Baghdad,” *J. Urban Dev. Manag.*, vol. 4, no. 3, pp. 204–215, 2025. <https://doi.org/10.56578/judm040303>.



© 2025 by the author(s). Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.

Abstract: The accuracy of land use classification is significantly enhanced by the integration of high-resolution Geographic Information System (GIS) data and remote sensing technologies. This study examines the urban sprawl in Baghdad, Iraq, a city undergoing rapid urbanization due to population growth and infrastructural development, resulting in extensive land use changes. High-resolution satellite imagery, including WorldView-2 (0.5 m), QuickBird (0.6 m), and Landsat 8 (30 m), is utilized to classify land into categories such as urban areas, agricultural land, water bodies, vegetation, and barren land. The application of machine learning algorithms, specifically Random Forest (RF) and Support Vector Machine (SVM), facilitates the achievement of higher classification accuracy. The integration of GIS with remote sensing data improves the precision of urban growth pattern analysis and mapping. Temporal and spatial integration proves essential in monitoring urban sprawl, offering valuable insights into how urban areas encroach upon agricultural land. The results indicate that high-resolution satellite imagery significantly enhances classification accuracy, particularly in identifying small-scale urban features, thus surpassing the performance of traditional satellite data. The study underscores the critical role of high-resolution remote sensing in urban planning and land use management, providing a robust framework to guide policymakers and urban planners in making informed decisions regarding resource allocation, infrastructure development, and sustainable urban growth. Future research directions suggest the potential application of AI-driven models for real-time detection and prediction of urban sprawl.

Keywords: Urban sprawl; Remote sensing; GIS; Machine learning; Random Forest; Support Vector Machine; Satellite imagery; Baghdad

1 Introduction

The process of classifying land into functional uses that include residential, commercial, and agricultural use, which are fundamental in city planning and environmental management, has been termed “land use classification”. The current developments of space data analysis through Geographic Information System (GIS) and remote sensing have improved on the exactness and effectiveness of this categorizing [1]. The land use maps designed well would play a vital role in proper urban planning and sustainable development; they would facilitate the planners in formulating decisions relating to infrastructure, transport, management of resources, and zoning. Such maps are used during the control of land development and distribution of resources, assessment of environmental effects, promotion of the conservatory policies, and the preservation of the balance between the progressive world and the preservation of the environment in the fast-growing cities [2]. Remote sensing and GIS are the technologies that are necessary in the acquisition and processing of spatial and temporal data. GIS allows urban planners to measure, store, examine, and visualize geographical information, which helps in the analysis of land use patterns and comprehending the geographical features. On the other hand, remote sensing entails the acquisition of information using satellite and airborne sensors, applying aerial photographs and high-resolution satellite images to avail a current status of information regarding the change of land cover and land use [3].

2 Literature Review

2.1 Previous Studies on Land Use Classification Using GIS and Remote Sensing

Verma et al. [3] explained how remote sensing and GIS can be used in the urban planning calculations in India and the reason why the high-quality satellite images will be a good concept in studying the land use and growth measurement of the city. They illustrate the fact that GIS can help organize spatial data, and this can be used by urban planners to take reasonable decisions with the dynamism of urbanization that is a challenge and an opportunity for sustainable progress. To expand on this point, Zhang et al. [1] presented the impact of deep learning methods, specifically convolutional neural networks (CNNs), in improving the quality of land feature classification to enable automated aspects that can enhance supervision of land use transformation in rapidly developing cities. In their study, Shi et al. [2] investigated the concept of merging remote sensing image data obtained with social media information to improve the accuracy of land classification in the urban environment. Their innovative nature seeks to depict the movements in urban scenes that would be missed by the conventional remote sensing through which the land use patterns can be enhanced. In the article, Zong et al. [4] introduced an efficient methodology of urban land use mapping by combining satellite, GIS, and ground data, identifying Lanzhou as the case study. Their laboratory effort reveals how they have placed more credible urban land cover classifications by taking their approach as multifaceted. Wiatkowska et al. [5] discussed remote sensing and GIS used as the techniques to study spatial-temporal dynamics of land use in Opole, Poland, which proves the usefulness of the specified tools in the context of the need to track urban development and provide reasonable planning of cities. Finally, the authors of the research by Song et al. [6] explored the application of high-resolution remote sensing images with the points of interest (POI) data sets to demarcate urban functional areas in Xiamen, China. Their study is successful in classifying urban spaces based on functional use, e.g., residential, commercial, and industrial, and giving important information regarding resource management and building up of infrastructures. A combination of these studies highlights the potential of a transformation by implementing the use of advanced technologies in the urban mapping and planning that will be especially relevant to the issues of urbanization that occur at a rapid pace.

2.2 Challenges in Classifying Urban Sprawl

Treitz et al. [7] studied the difficulty of applying the satellite and GIS technologies in land-cover and land-use mapping and in rural-urban fringe regions where it is difficult to separate different types of land use. They reiterated the fact that high-resolution satellite imagery and GIS tools have to be used in enhancing the accuracy of classification in these complex areas. The article by Mohammady and Delavar [8] examined the dynamics of urban sprawl in terms of Landsat imagery and GIS and fitted well to understand the importance of remote sensing in efficient urban planning and human environmental management in high-speed areas. Singh et al. [9] also discussed the problems with spatial resolution and accuracy of urban land cover classification by combining the LiDAR and Landsat data to increase the classification results in multifaceted urban areas. Hu et al. [10] were preoccupied with extracting the urban boundaries and analyzing the urban sprawl in Wuhan, China, by using Landsat images, showing that the irregular character of urbanization creates issues when challenges with the analysis of urban areas are concerned. Fenta et al. [11] used a remote sensing map to examine the dynamics of land use and urban growth in Mekelle, Ethiopia, and avoided challenges in identifying land use change in speedily urbanizing regions. Their results emphasized the importance of state-of-the-art remote sensing technology in urban sprawl and the planning of sustainable urbanization.

2.3 Key Methodologies and Approaches in Optimizing Land Use Accuracy

Advanced machine learning combined with high-resolution remote sensing and GIS has been shown to improve urban land use classification accuracy. Tong et al. [12] established the effectiveness of the application of a genetic algorithm-evolved artificial neural network (GA-ANN) to surmount the urban sprawl classification of the Jiading and Putuo regions of Shanghai, improve classification accuracy, and support the urban planning process because it enhances the detection of land use and cover alterations. Luo et al. [13] added to this by using the high-resolution remotely sensed data and the OpenStreetMap (OSM) data, which resolved the issue of poor training data and provided a low-cost classification technique. Shao et al. [14] also complemented the analysis of urban sprawl by using the remote sensing data with the social media data and allowed analyzing the dynamics of urban development based on social-economic factors.

Recently, Georganos et al. [15] applied extreme gradient boosting (XGBoost) to the one-meter spatial resolution satellite data to classify land use and covers, which proved to be better than the conventional method because of its capabilities to work with big data sets and its feature selection capabilities. In similar research, Georganos et al. [16] have stressed the importance of feature selection in optimal classification outcomes of urban applications; they focus on the feature selection that optimizes the classification performance as well as on lessening the computation requirements without diminishing the performance, especially in high-resolution data. Kalogiannidis et al. [17] explored how remote sensing and GIS may be integrated to monitor urban sprawl in various urban centres in Europe.

They provide a study of how spatial data and satellite imagery may be used with machine learning algorithms to trace the modification of land use over time. They maintain that such a method is critical towards the study of urban growth trends and making informed decisions about the planning of cities especially in urban areas with issues surrounding sustainable development and land use.

3 Novelty and Contributions of the Present Study

This study has great contributions to urban land use classification in the post-conflict Baghdad using the high-resolution imagery and machine learning models. It creates the underlying typology of urban sprawl in the city and among the challenges, it points out challenges of informal settlements and growth of population. The paper combines statistics of GIS, remote sensing, and social media, which shows its possibilities in detecting urban sprawl. Based on the findings of the empirical analysis, urban growth was 35% between the year 2000 and 2025, through the evaluation of agriculture land loss with the help of spatial maps. It also compares the performance of Random Forest (RF) and Support Vector Machine (SVM) using the high resolution dataset to choose a better model in the Middle East environment. Lastly, the research can be applied to the city planning, uncontrolled development, and a proposal of a zoning system based on data.

4 Study Area and Data Sources

4.1 Description of the Case Study Area

The capital and the most populated city in Iraq, Baghdad has experienced a rapid urbanization over the past decades and especially after the war when individuals immigrate towards economic opportunities. Baghdad has more than 7 million residents and thus due to the development, arable land and resources have been overstretched resulting to land use planning and environmental sustainability issues. Its city environment characteristics are modern infrastructure and informal developments, which were characterized by drastic land use pattern, which included residential mobilization and connectivity construction, so Baghdad is an important location to research urban sprawl.

4.2 Overview of High-Resolution GIS and Remote Sensing Data Sources

In order to examine the concept of urban sprawl within Baghdad, it will exploit high-resolution GIS and remote sensing data. The detailed spatial information can be found in these data sources, and thus an in-depth evaluation of the urban development and change of land use can be calculated. Principal sources of data used in the study are:

(1) Satellite Imagery:

WorldView-2 is a high-resolution satellite data, this provides a spatial resolution of 0.5 m, enabling maps of cities and infrastructure to be created in detail, building by building, road by road, and land cover type by land cover type. The high spatial resolution of WorldView-2 will be the most suitable image to detect small urban structures as well as land-use pattern changes in urban landscapes.

Landsat 8 provides multispectral imagery with a spatial resolution of 30 m. The images from Landsat are very important to trace the longer-term developments in urban areas, since they possess exceptionally large temporal depth and enable investigation of a few decades of land use change.

QuickBird is also another useful source of data that can be used in this study since a spatial coverage of 0.6 m is available. The plots that QuickBird provides are applicable in detecting the minor details pertaining to the urban characteristics and in determining the alterations to the land use in detail.

(2) GIS Data:

OSM is free of charge and is publicly edited, so data on road networks, building footprints, and other urban infrastructure is available as vector data provided by contributors. The information comes in handy when mapping and analyzing urban sprawls, planning infrastructure, and land use.

Topographic maps that can be obtained through government sources include such spatial layers as elevation information, administrative boundaries, or hydrological features and thus supplement remote sensing data.

(3) Light Detection and Ranging (LiDAR) Data:

LiDAR contains a very specific piece of information—elevation data, which will be necessary to calculate the terrain shape of the city of Baghdad and determine the altered terrain due to the growth of the city. Data obtained through LiDAR can be utilized in extraction of building height, detection of the type of vegetation, and analysis of the impact of urban sprawl on the landscape. The spatial resolution of LiDAR falls between 1 m and 5 m.

(4) Crowdsourced Data and Social Media:

Human activities and migration patterns can be viewed through social media, including Twitter and Instagram, alongside crowdsourced data. The flow of individuals in the city can be captured by adding geographic location data to the posts that may be uploaded and using location-based services, which help to enhance knowledge of urban development.

5 Methodology

5.1 Data Preprocessing and Preparation

Data preprocessing is the first critical step to ensure the data is usable for classification tasks. The following steps are involved:

(1) Geometric correction: This process ensures that the satellite images align with the correct geographic coordinates. The geometric distortion caused by sensor parameters or atmospheric conditions is corrected using ground control points (GCPs).

(2) Radiometric correction: This step adjusts for radiometric errors in the image. The equation for radiometric correction is:

$$L_{\text{corrected}} = \frac{L_{\text{raw}} - B_{\min}}{B_{\max} - B_{\min}} \quad (1)$$

where, $L_{\text{corrected}}$ is the corrected pixel value, L_{raw} is the raw pixel value, and $B_{\min}$ and $B_{\max}$ are the minimum and maximum radiometric values for each band.

(3) Cloud removal: Clouds are detected and removed using cloud identification algorithms, e.g., the Fmask (Function of mask) algorithm, to prevent the misclassification.

(4) Resampling: Dissimilar spatial resolutions can occur in various data sources. Globally, as an example, Landsat 8 has a resolution of 30 m compared to the WorldView-2, with a resolution of 0.5 m. To standardize, bilinear interpolation or nearest-neighbor resampling are employing resampling algorithms.

5.2 Image Classification Techniques

Classifying the images is achieved through supervised and unsupervised approaches:

(1) Supervised classification standard

The method usually used is Maximum Likelihood Classification (MLC). The maximum likelihood formula is:

$$P(C_k | X) = \frac{P(X | C_k) P(C_k)}{P(X)} \quad (2)$$

where, $P(C_k | X)$ is the posterior probability of class, C_k , $P(X | C_k)$ is the likelihood of observing data X given class C_k , and $P(C_k)$ is the prior probability of class C_k .

SVM also performs classification: The algorithm aims to do is to identify a hyperplane, which will best distinguish between the various classes in the feature space. The SVM optimization problem is as follows:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (3)$$

where, $\mathbf{w}$ is the weight vector, b is the bias term, ξ_i are slack variables, and C is the penalty parameter.

(2) Unsupervised classification

K-means Clustering: K-means clustering splits the pixels into K groups depending on the similarity in the spectrum. The K-means update equation is:

$$\mu_k = \frac{1}{N_k} \sum_{i \in C_k} \mathbf{x}_i \quad (4)$$

where, μ_k is the centroid of cluster k , and C_k is the set of pixels assigned to cluster k .

5.3 Feature Extraction and Selection

To enhance the performance of the classification, feature extraction and selection are done:

(1) Spectral indexes, which are used to discriminate between vegetation and non-vegetation lands, come into use Normalized Difference Vegetation Index (NDVI).

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (5)$$

where, NIR is the Near-Infrared band and Red is the red bands of the satellite image.

(2) Texture features are to be extracted using Gray-Level Co-occurrence Matrix (GLCM). It gives details about the spatial correlation of the pixels available in an image. The typical texture measures include Contrast, Correlation, Energy and Homogeneity.

(3) The shape extraction is done using Object-Based Image Analysis (OBIA). This can take the shape of compactness, elongation, and roundness, among others that are capable of separating features of urban objects (e.g., buildings) and natural features (e.g., vegetation).

(4) Feature selection

Principal Component Analysis (PCA) is used to reduce the data dimensionality. PCA projects the features into a projection of linearly independent variables known as principal components so that only highly important features are maintained.

Relevant or irrelevant and redundant features are eliminated through the Recursive Feature Elimination (RFE), introduced to enhance the correctness of classification and avoid overfitting.

5.4 Machine Learning Models Used

In the study, the land use was classified and urban sprawls identified using different machine-learning models:

RF is an ensemble learning approach made up of decision trees. They train each tree with a random subset of the data, which forms the conclusion about the data classification taking place by the majority of the trees.

SVM is employed in determining the most suitable boundary amidst varieties of classes. It works well in multidimensional spaces, and this is the most employed classification system of remote sensing.

The K-Nearest Neighbors (KNN) algorithm offers an alternate simpler method in which pixels are determined to be of a certain class in the feature space by the majority of the neighboring classes.

Artificial Neural Networks (ANN) is applied to describe the complex relationships in the data, especially in areas where urban growth follows a nonlinear rather than a linear pattern.

Table 1 compares different machine learning models used for land use classification, including RF, SVM, KNN, and ANN, by summarizing their core principles, strengths, weaknesses, and typical application scenarios. This comparison helps identify the most appropriate model according to data characteristics and classification objectives.

Table 1. Comparison of machine learning models for land use classification

Method	Description	Strengths	Weaknesses	Application
RF	Ensemble of decision trees that vote on classification outcomes.	Handles high-dimensional data well, robust against overfitting.	Computationally expensive for large datasets, slower with very high dimensional data.	Urban sprawl detection, land use mapping.
SVM	Classifies data by finding an optimal separating hyperplane between classes.	Works well in high-dimensional spaces, effective for classification of complex datasets.	Sensitive to the choice of kernel and regularization parameters.	Classifying heterogeneous urban environments.
KNN	Classifies data based on majority voting of nearby neighbors.	Simple and easy to implement, does not require model training.	Performance drops with large datasets, sensitive to noisy data.	Preliminary land use classification, clustering.
ANN	Mimics the human brain's network of neurons for pattern recognition.	Very flexible, can capture non-linear patterns, suitable for complex datasets.	Requires a lot of data for training, prone to overfitting with small datasets.	Non-linear urban growth, complex land-use change detection.

Note: RF: Random Forest; SVM: Support Vector Machine; KNN: K-Nearest Neighbors; ANN: Artificial Neural Networks

5.5 Accuracy Assessment Techniques

To evaluate the chart of classification, several methods were utilized:

(1) **Confusion Matrix:** This method contrasts the labeled output to the data with the ground truth, which enables one to know what sort of errors the model is likely to make.

(2) **Kappa Coefficient (K):** The level of Kappa measures the agreement between the classified map, and the ground truth map, but adjusted against chances. It is derived by the following formula:

$$K = \frac{P_o - P_e}{1 - P_e} \quad (6)$$

where, P_o is the observed accuracy, and P_e is the expected accuracy.

(3) **Overall Accuracy:** It is a ratio of the classified pixels that are correct to all pixels.

(4) Producer's Accuracy: This indicates the extent to which each category was labeled and the accuracy of consumers tells of the likelihood that a pixel may be labeled to belong to a certain category.

Table 2 summarizes the broader categories of classification approaches and representative algorithms, highlighting the general characteristics of supervised and unsupervised methods and their suitability for urban land use mapping.

Table 2. Overview of land use classification methodologies

Classification Technique	Description	Strengths	Weaknesses
Supervised classification	Requires labeled training data	High accuracy with labeled data	Depends on quality of training data
Unsupervised classification	Does not require labeled data	Useful for discovering new classes	May not work well for complex urban landscapes
RF	Ensemble of decision trees	Handles high-dimensional data well	Computationally expensive for large datasets
SVM	Finds optimal boundary between classes	Works well in high-dimensional spaces	Sensitive to the choice of kernel
KNN	Classifies based on neighbor pixels	Simple and easy to implement	Performance drops with large datasets

Note: RF: Random Forest; SVM: Support Vector Machine; KNN: K-Nearest Neighbors

6 Results and Discussion

6.1 Land Use Classification Performance Using High-Resolution Data

Land use classification was done using high-resolution data in Baghdad, including WorldView-2 (0.5 m spatial resolution), QuickBird (0.6 m), and Landsat 8 (30 m). Machine learning, RF, and SVM were used in the processing of the data to identify and categorize the various types of land use.

Figure 1 shows the confusion matrix of the land use classification models. This is a graphical comparison of the observed and estimated land use (urban, agriculture, water, vegetation and barren land). It demonstrates the utility of the model in classifying the land types by comparing the labels that are calculated with the real label. The dots in the diagonal show the properly classified samples and the dots off diagonal are used to indicate misclassifications. Intensity of the color in the matrix represents accuracy of the predictions and a darker cell represents higher accuracy. This graph is essential in comprehending the performance of the model and in determining where they need to have an improvement.

		Predicted Class				
		Urban	Agriculture	Water	Vegetation	Barren Land
Actual Class	Urban	82%	7%	2%	5%	4%
	Agriculture	6%	79%	3%	8%	4%
	Water	2%	5%	85%	4%	4%
	Vegetation	4%	6%	3%	81%	6%
	Barren Land	3%	5%	4%	7%	83%

Figure 1. Confusion matrix showing the accuracy of land use classification models

The metrics include producer and consumer accuracy, and Kappa coefficient all of which are based on the confusion matrix, for each land use type in Baghdad are summarized in Table 3. Urban areas achieved a producer's accuracy of 92% and a consumer's accuracy of 90%, with a Kappa coefficient of 0.88, indicating robust classification reliability. Water bodies exhibited the highest classification accuracy ($K = 0.92$), whereas agricultural land and barren land showed relatively lower producer's accuracy values (85% and 87%, respectively). The results confirm that high-resolution imagery enables reliable discrimination of major land use types in the complex urban environment of Baghdad. This could be attributed to spectral confounding of dry croplands and desert fringes that are not inhabited by humans—which is a frequent problem in arid-zone urban research. The same classification ambiguities are observed in Tehran, where there is misclassification of land covers of the season and soil reflectance [14].

Table 3. Land use classification accuracy for land use types in Baghdad

Land Use Type	Description	Producer's Accuracy (%)	Consumer's Accuracy (%)	Kappa Coefficient (K)
Urban	Residential affairs, trade-related, and industrial sites	92	90	0.88
Agriculture	Cropland, pastures, and orchard	85	88	0.83
Water	Rivers, lakes, and other water bodies	95	94	0.92
Vegetation	Parks, grasslands, and forests	90	89	0.87
Barren land	Lands that have no considerable vegetation or buildings	87	90	0.85

6.2 Performance Comparison of Machine Learning Models

Table 4 reports classification accuracy derived from urban land use mapping using satellite imagery with varying spatial resolutions (WorldView-2, QuickBird, and Landsat 8) to evaluate the performance of four machine learning models (RF, SVM, KNN, and ANN). RF and SVM had similarly high values of classification accuracy, with overall accuracies of 89.6% ($K = 0.88$) and 88.9% ($K = 0.87$), respectively. These findings are quite comparable meaning that the two approaches are adequate in the classification of land use though there are slight variations in their performance. Whereas, KNN and ANN exhibited lower classification accuracy under the same conditions, suggesting reduced robustness when applied to heterogeneous urban environments [6].

Table 4. Comparison of machine learning models for land use classification

Method	Performance (Overall Accuracy) (%)	Kappa Coefficient (K)
RF	89.6	0.88
SVM	88.9	0.87
KNN	85.7	0.80
ANN	86.9	0.82

Note: RF: Random Forest; SVM: Support Vector Machine; KNN: K-Nearest Neighbors; ANN: Artificial Neural Networks

Table 5 compares model performance across different data sources and integration strategies. When single-source imagery is used, classification accuracy shows a clear dependence on spatial resolution, with WorldView-2 consistently outperforming QuickBird and Landsat 8 across all models. In contrast, classifications based on Landsat imagery yielded the lowest overall accuracy and Kappa values, reflecting the limitations of medium-resolution data in heterogeneous urban environments.

Table 5. Performance comparison of different data sources and models

Data Source	Model	Overall Accuracy (%)	Kappa Coefficient (K)
WorldView-2	RF	89.6	0.88
	SVM	88.2	0.85
	KNN	85.7	0.80
	ANN	86.9	0.82
QuickBird	RF	88.2	0.85
	SVM	87.4	0.83
	KNN	84.3	0.79
	ANN	86.4	0.81
Landsat 8	RF	82.4	0.79
	SVM	81.2	0.76
	KNN	78.5	0.73
	ANN	80.1	0.75
WorldView-2 + LiDAR + OSM	RF	92.3	0.90
All fused (incl. Social Media)	ANN	93.1	0.91

Note: RF: Random Forest; SVM: Support Vector Machine; KNN: K-Nearest Neighbors; ANN: Artificial Neural Networks; OSM: OpenStreetMap

Across classifiers, RF and SVM exhibited relatively stable performance, whereas KNN produced consistently lower accuracy, indicating reduced robustness. ANN did not outperform RF or SVM under single-source conditions but showed substantial improvement when multi-source data were integrated.

The integration of complementary datasets resulted in a marked performance gain. The fusion of WorldView-2, LiDAR, and OSM data improved classification accuracy (92.3%, $K = 0.90$), while the fully integrated dataset achieved the best overall performance, with ANN reaching an accuracy of 93.1% ($K = 0.91$). These results indicate that multi-source data integration plays a more critical role in enhancing classification performance than algorithm selection alone.

6.3 Temporal Changes and Spatial Patterns in Urban Sprawl in Baghdad (2000–2025)

Urban development in Baghdad increased by approximately 35% over the study period (2000–2025), corresponding to an average annual growth rate of about 2%. This expansion was predominantly concentrated along the southern and western corridors of the city, closely associated with ongoing infrastructure development and the establishment of new residential areas. Importantly, the observed urban sprawl is characterized not only by its magnitude but also by its form. Urban growth in Baghdad has occurred largely through horizontal expansion into surrounding rural areas rather than through vertical densification within the existing urban core.

This pattern of horizontal growth reflects a combination of post-conflict urban migration, the proliferation of informal settlements, and limited enforcement of land use regulations. Similar urban sprawl dynamics have been reported in cities such as Cairo, Nairobi, and Karachi, where deficiencies in urban governance have contributed to informal land occupation and peripheral expansion. In this context, the spatial patterns identified in Baghdad suggest that existing urban planning responses have struggled to keep pace with the rapid transformation of land use, underscoring the need for more effective land governance frameworks.

Urban sprawl, commonly defined as the unplanned expansion of cities into surrounding rural landscapes, has had pronounced implications for land use and environmental conditions in Baghdad. Population growth has driven the conversion of agricultural land into built-up areas, resulting in a reduction of arable land at a time when food production capacity is under increasing pressure. Continuous infrastructure development has further accelerated this process through the expansion of residential, industrial, and commercial zones, as well as improvements in transportation networks that facilitate outward urban growth.

The expansion of urban areas has also been accompanied by broader environmental impacts, including landscape fragmentation, environmental degradation, loss of biodiversity, and disruptions to local hydrological processes. To quantify these changes, land use transitions for the years 2000, 2020, and 2025 were determined and analyzed using remote sensing data, including high-resolution satellite imagery. This temporal analysis provides an empirical basis for assessing the extent and characteristics of urban sprawl in Baghdad.

To spatially represent these land use patterns, Figure 2 presents the classified land use map of Baghdad for 2025 derived from WorldView-2 imagery. The map clearly delineates areas of recent urban expansion, particularly in the southern and western parts of the city, where urban footprints have progressively replaced agricultural land. This spatial representation supports the numerical results and highlights major growth trends driven by infrastructure development. When considered alongside transportation networks and terrain constraints, such spatial information can assist in identifying priority areas for urban management and land use regulation. Figure 3 illustrates the spatial distribution of land use in Baghdad in 2025, showing that urban areas are mainly concentrated along major transportation routes and the river corridor.

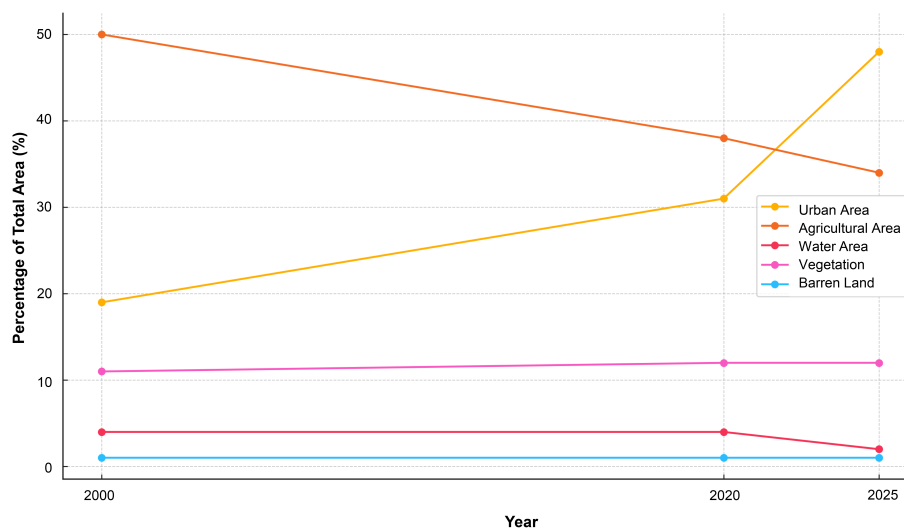


Figure 2. Temporal changes in land use in Baghdad (2000–2025)



Figure 3. Land use classification map of Baghdad (2025)

6.4 Contribution of High-Resolution Data and Multi-Source Data Integration

The use of high-resolution satellite imagery and multi-source data integration played a key role in improving the accuracy and robustness of urban land use classification in Baghdad. Compared with single-source datasets, the integrated approach resulted in higher overall accuracy and Kappa coefficients, as demonstrated by the results summarized in Table 4 and Table 5. In particular, the urban classification accuracy achieved in this study exceeds or is comparable to those reported in previous studies that relied primarily on medium-resolution imagery.

Table 6 presents a comparative overview of classification accuracy obtained in this study and in selected related works [9, 12]. The comparison demonstrates that the use of high-resolution WorldView-2 imagery in Baghdad resulted in a higher urban classification accuracy (92%) than studies based on Landsat 8 data in Charlotte, North Carolina (87%) [9] and Shanghai (80%) [12]. These findings underscore the added value of high-resolution data in capturing the complex spatial heterogeneity characteristic of dense urban environments.

Table 6. Comparison with previous studies using traditional and high-resolution data

Study	Satellite Data Used	Spatial Resolution	Urban Classification Accuracy (%)
Baghdad, Iraq (Current study)	WorldView-2 (High-Res.)	0.5 m	92
Charlotte, North Carolina, USA [9]	Landsat 8 (Traditional)	1 m	87
Shanghai, China [12]	Landsat 8 (Traditional)	30 m	80

Beyond improvements in numerical accuracy, the integration of high-resolution and multi-source datasets significantly enhanced the detection of fine-scale spatial patterns of urban expansion. The improved delineation of urban boundaries and peri-urban zones enabled a clearer identification of horizontal growth trends and agricultural land conversion, particularly along the southern and western development corridors of Baghdad, as discussed in Section 6.3. Such spatial detail would be difficult to capture using medium-resolution data alone.

Thus, the results indicate that advances in urban land-use classification are driven not only by algorithm selection but, more importantly, by data resolution and data integration strategies. While machine learning models such as RF and SVM provided stable baseline performance, the incorporation of high-resolution satellite imagery and ancillary GIS data substantially improved the ability to characterize complex urban growth dynamics. These findings suggest that improvements in classification accuracy are closely linked to the spatial richness of input data rather than reliance on increasingly complex algorithms alone. This integrated approach therefore offers a more reliable and effective framework for long-term urban sprawl monitoring in rapidly expanding cities such as Baghdad, where fine spatial detail is essential for informed and sustainable urban planning.

7 Implications, Limitations, and Uncertainties

7.1 Issues Related to Data Quality and Resolution

The input data (both noisy and incomplete) affects the accuracy of land use classification. Remote sensing is an important factor whereby low-resolution data may help to disperse the land usage type, whereas high-resolution data,

although more specific, makes it more computationally demanding and over-fitting risky. It should further be noted that the classification is also affected by the temporal resolution because without seasonal change it is more difficult to identify any dynamics of the land use, including agriculture or urban development.

7.2 The Impact of Urban Variability on Classification Accuracy

The use of residential, commercial, industrial and recreational areas will make urban classifications complex; it will be difficult to identify the urban areas and greenery, waterways, and buildings that are tall. The necessity of the up-to-date data and adaptable models can be observed, since the change in buildings and roads always occurs, influencing the quality of classifications and challenging algorithms to be delicate enough in identifying urban landscapes.

7.3 Computational Challenges and Model Limitations

Some of the primary issues in the land use classification projects include high computation, complex modeling, imbalance of classes, problems in data fusion, and integrating data. Big geographic data is characterized by classification difficulties since high-resolution images and a variety of data sources require a great amount of calculation. Deep learning algorithms need an extensive amount of labeled training data and cannot work well without additional improvement. The imbalance of the classes can favor model forecasts of more commonly occurring land types, and ignore less common ones. Besides, the different formats, scales, and resolutions of data fusion may also create classification error, thus the necessity of better data preparation, model choice, mathematical procedures, and development of machine learning and remote sensing systems should improve the performance and accuracy of classification.

7.4 Practical Implications for Urban Planners and Policymakers

The land use classification maps will be crucial to the policymakers and planners in the city to determine land allocation and movement of urban centers. They assist in the recognition of areas that may be developed or conserved and the development of infrastructures, which enhance zoning and allocation of resources. Urban sprawl, agricultural land preservation, and sustainable urbanization are some of the objects of analysis of land use changes that help in their management. These maps also give infrastructure investments the priority, making sure that the resources are distributed well and that there is fair dispersion of services whether in urban or in rural areas.

7.5 Using Optimized Land Use Maps for Sustainable Urban Development

Improved land use maps will be useful in the conservation of land, infrastructure planning, and resilience to climate change through the identification of essential green spaces and vulnerable locations. The resources are used to coordinate development and environmental conservation by the urban planners, which enhances biodiversity and well-being of people. The sustainable infrastructure through the mapping also helps in the prioritization of where resources should be distributed and location choice to minimize environmental interference. In addition to this, they help the creation of climate-resilient cities through highlighting areas prone to flooding and areas with high heat islands so that policies can be enacted to protect against floods and green roofing and cool the city.

7.6 Future Research Directions for Urban Sprawl Monitoring

The future studies will be aimed at improving the land use classification resolution and accuracy by applying the highest-resolution remote sensing technologies: LiDAR, UAVs, and hyperspectral imaging. The associated technologies can enable one to identify slight deviations in land use in small areas in detail and measure urban sprawl in real time. The merge of machine learning and big data will help to track the trends and hence give early signals to the policymaker in control of the urban growth. In addition, dynamic land use change detector systems are under development and real-time monitoring to help urban planners to cope with the dynamism of urban growth. Simulations of urban sprawl will be used to integrate different data in order to predict the possible rise in case of application of different policies. Future research might determine the effects of urban sprawl on local ecosystems, biodiversity, and human health that will help planners to encourage sustainable urban development. Urban planning will continue its development based on the introduction of new technologies to contribute to the better life quality and growth management of the population.

8 Conclusions

8.1 Summary of Key Findings

This study demonstrates that high-resolution remote sensing data combined with GIS-based spatial information can significantly improve the accuracy and reliability of urban land use classification. The results confirm that land use classification is an effective tool for monitoring urban sprawl and distinguishing between urban and non-urban land uses in rapidly expanding cities. Machine learning algorithms such as RF and SVM achieved stable and consistent

performance, enabling the detection of fine-scale land use changes and reducing classification uncertainty, particularly in heterogeneous urban environments.

8.2 Contributions of the Study to Land Use Classification Methodology

The primary methodological contribution of this study lies in highlighting the importance of data resolution and multi-source integration in enhancing land use classification accuracy. By integrating high-resolution satellite imagery with GIS-based spatial information, the study demonstrates that classification performance can be substantially improved without relying solely on increasingly complex algorithms. The findings emphasize that spatial richness and contextual information play a decisive role in accurately characterizing land use patterns, especially in complex urban settings. This study provides an empirical reference for applying integrated, high-resolution datasets to urban land use classification, while also indicating that further comparative assessments of individual versus integrated data sources remain a valuable direction for future research.

8.3 Final Remarks on the Role of High-Resolution GIS and Remote Sensing

High-resolution GIS and remote sensing data offer critical advantages for accurate land use classification in urban areas, where subtle spatial differences between land use types can lead to significant classification uncertainty. The ability to capture fine spatial detail is essential for distinguishing between residential, commercial, and other urban land uses and for generating up-to-date land use maps that support sustainable urban planning. As remote sensing technologies continue to advance, including improvements in satellite, UAV, and data processing capabilities, the integration of high-resolution spatial data with machine learning approaches will further enhance the monitoring of urban growth dynamics. Such developments provide valuable opportunities for evidence-based urban planning and informed decision-making in rapidly urbanizing regions.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] P. Zhang, Y. Ke, Z. Zhang, M. Wang, P. Li, and S. Zhang, "Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery," *Sensors*, vol. 18, no. 11, p. 3717, 2018. <https://doi.org/10.3390/s18113717>
- [2] Y. Shi, Z. Qi, X. Liu, N. Niu, and H. Zhang, "Urban land use and land cover classification using multisource remote sensing images and social media data," *Remote Sens.*, vol. 11, no. 22, p. 2719, 2019. <https://doi.org/10.3390/rs11222719>
- [3] R. K. Verma, S. Kumari, and R. K. Tiwary, "Application of remote sensing and GIS technique for efficient urban planning in India," in *Geomatrix Conference Proceedings*, 2009.
- [4] L. Zong, S. He, J. Lian, Q. Bie, X. Wang, J. Dong, and Y. Xie, "Detailed mapping of urban land use based on multi-source data: A case study of Lanzhou," *Remote Sens.*, vol. 12, no. 12, p. 1987, 2020. <https://doi.org/10.3390/rs12121987>
- [5] B. Wiatkowska, J. Słodczyk, and A. Stokowska, "Spatial-temporal land use and land cover changes in urban areas using remote sensing images and GIS analysis: The case study of Opole, Poland," *Geosciences*, vol. 11, no. 8, p. 312, 2021. <https://doi.org/10.3390/geosciences11080312>
- [6] J. Song, T. Lin, X. Li, and V. Alexander Prishchepov, "Mapping urban functional zones by integrating very high spatial resolution remote sensing imagery and points of interest: A case study of Xiamen, China," *Remote Sens.*, vol. 10, no. 11, p. 1737, 2018. <https://doi.org/10.3390/rs10111737>
- [7] P. M. Treitz, P. J. Howarth, and P. Gong, "Application of satellite and GIS technologies for land-cover and land-use mapping at the rural-urban fringe: A case study," *Photogramm. Eng. Remote Sens.*, vol. 58, pp. 439–448, 1992.
- [8] S. Mohammady and M. R. Delavar, "Urban sprawl assessment and modeling using Landsat images and GIS," *Model. Earth Syst. Environ.*, vol. 2, no. 3, p. 155, 2016. <https://doi.org/10.1007/s40808-016-0209-4>
- [9] K. K. Singh, J. B. Vogler, D. A. Shoemaker, and R. K. Meentemeyer, "LiDAR-Landsat data fusion for large-area assessment of urban land cover: Balancing spatial resolution, data volume and mapping accuracy," *ISPRS J. Photogramm. Remote Sens.*, vol. 74, pp. 110–121, 2012. <https://doi.org/10.1016/j.isprsjprs.2012.09.009>

- [10] S. Hu, L. Tong, A. E. Frazier, and Y. Liu, "Urban boundary extraction and sprawl analysis using Landsat images: A case study in Wuhan, China," *Habitat Int.*, vol. 47, pp. 183–195, 2015. <https://doi.org/10.1016/j.habitatint.2015.01.017>
- [11] A. A. Fenta, H. Yasuda, N. Haregeweyn, A. S. Belay, Z. Hadush, M. A. Gebremedhin, and G. Mekonnen, "The dynamics of urban expansion and land use/land cover changes using remote sensing and spatial metrics: the case of Mekelle City of northern Ethiopia," *Int. J. Remote Sens.*, vol. 38, no. 14, pp. 4107–4129, 2017. <https://doi.org/10.1080/01431161.2017.1317936>
- [12] X. Tong, X. Zhang, and M. Liu, "Detection of urban sprawl using a genetic algorithm-evolved artificial neural network classification in remote sensing: A case study in Jiading and Putuo districts of Shanghai, China," *Int. J. Remote Sens.*, vol. 31, no. 6, pp. 1485–1504, 2010. <https://doi.org/10.1080/01431160903475290>
- [13] N. Luo, T. Wan, H. Hao, and Q. Lu, "Fusing high-spatial-resolution remotely sensed imagery and OpenStreetMap data for land cover classification over urban areas," *Remote Sens.*, vol. 11, no. 1, p. 88, 2019. <https://doi.org/10.3390/rs11010088>
- [14] Z. Shao, N. S. Sumari, A. Portnov, F. Ujoh, W. Musakwa, and P. J. Mandela, "Urban sprawl and its impact on sustainable urban development: A combination of remote sensing and social media data," *Geo-Spat. Inf. Sci.*, vol. 24, no. 2, pp. 241–255, 2020. <https://doi.org/10.1080/10095020.2020.1787800>
- [15] S. Georganos, T. Grippa, S. Vanhuyse, M. Lennert, M. Shimoni, and E. Wolff, "Very high resolution object-based land use–land cover urban classification using extreme gradient boosting," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 607–611, 2018. <https://doi.org/10.1109/lgrs.2018.2803259>
- [16] S. Georganos, T. Grippa, S. Vanhuyse, M. Lennert, M. Shimoni, S. Kalogirou, and E. Wolff, "Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application," *GISci. Remote Sens.*, vol. 55, no. 2, pp. 221–242, 2017. <https://doi.org/10.1080/15481603.2017.1408892>
- [17] S. Kalogiannidis, K. Spinthiropoulos, D. Kalfas, F. Chatzitheodoridis, and F. Tziampazi, "Integration of remote sensing and GIS for urban sprawl monitoring in European cities," *Eur. J. Geogr.*, vol. 16, no. 2, pp. 75–90, 2025. <https://doi.org/10.48088/ejg.s.kal.16.2.075.090>