INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGY RESEARCH STUDIES (IJITRS)
(Open Access, Double-Blind Peer Reviewed Journal)
ISSN Online: 3139-3209


# A Hybrid RF-CNN Framework for GIS-Driven Land-Use Modeling and Sustainable Urban Growth Prediction

T Ramaprabha

Associate Professor, Department of Computer Science, Nehru Arts and Science College, Coimbatore, India

**Article information**

Received: 10th January 2026
Received in revised form: 12th February 2026
Accepted: 14th March 2026
Available online: 30th April 2026

Volume: 2
Issue: 2
DOI: https://doi.org/10.63090/IJITRS/3139.3209.0023

**Abstract**

Rapid urbanization presents significant challenges for sustainable land-use planning in developing regions. This study proposes an integrated Geographic Information System (GIS) and machine learning framework for multi-temporal land-use classification and urban growth prediction. Utilizing Landsat and Sentinel-2 satellite imagery spanning 2000 to 2023, the proposed approach combines spectral and spatial feature extraction with a hybrid Random Forest–Convolutional Neural Network (RF-CNN) classifier. The framework achieves an overall classification accuracy of 93.4% with a Kappa coefficient of 0.91, outperforming conventional methods including Maximum Likelihood Classification (78.2%), Support Vector Machines (84.5%), and standalone Random Forest (87.1%). A cellular automata–Markov chain model integrated within the GIS environment projects future urban expansion scenarios for 2030 and 2040. The results reveal that built-up areas have increased by 72.7% over the study period, while vegetation cover has declined by 50%. The proposed framework provides urban planners with a data-driven decision support tool for formulating sustainable development strategies that balance economic growth with environmental conservation.

**Keywords**:- GIS, Land-Use Classification, Remote Sensing, Machine Learning, Urban Growth, Sustainable Development, CNN, Random Forest.

## I. INTRODUCTION

The global urban population has surged from 3.3 billion in 2007 to over 4.4 billion in 2023, with projections indicating that 68% of the world's population will reside in urban areas by 2050 [1]. This unprecedented rate of urbanization imposes severe pressure on natural resources, ecosystems, and infrastructure, particularly in developing nations where urban expansion frequently occurs in an unplanned and haphazard manner [2]. The conversion of agricultural and forest lands to built-up areas not only diminishes food security and biodiversity but also exacerbates the urban heat island effect and increases vulnerability to flooding and other climate-related hazards [3].

Geographic Information Systems (GIS) have emerged as indispensable tools for spatial analysis and urban planning, enabling the integration, visualization, and analysis of geographically referenced data from multiple sources [4]. When combined with remote sensing technologies, GIS provides a powerful platform for monitoring land-use and land-cover (LULC) changes over time, offering critical insights into the spatial patterns and drivers of urban expansion [5]. The availability of freely accessible multi-spectral satellite imagery from programs such

as Landsat and the European Space Agency's Sentinel missions has further democratized LULC analysis, making it feasible for researchers and planners in resource-constrained settings [6].

Traditional LULC classification methods, including Maximum Likelihood Classification (MLC) and unsupervised clustering algorithms such as ISODATA, have been widely applied but often suffer from limited accuracy when dealing with spectrally similar land-cover classes or heterogeneous urban landscapes [7]. The advent of machine learning has introduced more robust classification approaches, with algorithms such as Support Vector Machines (SVM) and Random Forest (RF) demonstrating superior performance in handling high-dimensional feature spaces and non-linear class boundaries [8]. More recently, deep learning architectures, particularly Convolutional Neural Networks (CNNs), have achieved state-of-the-art results in remote sensing image classification by automatically learning hierarchical spatial features from raw imagery [9].

Despite these advances, several challenges remain. Individual classifiers may not capture the full complexity of urban landscapes, and the integration of classification outputs with urban growth modeling within a unified GIS framework remains an active area of research [10]. Furthermore, most existing studies focus on single-date classification rather than multi-temporal analysis that can reveal the trajectory and rate of urban change [11]. This study addresses these gaps by proposing a hybrid RF-CNN classification approach integrated within a GIS-based urban growth modeling framework. The specific contributions of this work are threefold:

- A novel feature fusion strategy combining spectral indices with CNN-extracted spatial features;
- A hybrid classifier that leverages the complementary strengths of RF and CNN; and
- A comprehensive GIS-based decision support tool that couples LULC classification with cellular automata–Markov chain modeling for future urban growth projection.

## II. LITERATURE REVIEW

### A. Remote Sensing for Land-Use Classification

Remote sensing has been the cornerstone of LULC analysis since the launch of the first Landsat satellite in 1972 [12]. The Landsat program, now in its ninth generation, provides continuous 30-meter resolution multi-spectral imagery with a 16-day revisit cycle, forming an unparalleled archive for temporal change analysis. The Sentinel-2 mission, operational since 2015, offers complementary 10-meter resolution imagery with a 5-day revisit period, enabling more frequent and detailed monitoring of land surface dynamics [6]. Studies by Phiri and Morgenroth [13] and Wulder et al. [12] have comprehensively reviewed the evolution of Landsat-based LULC classification, highlighting the progressive improvement in accuracy achieved through advances in sensor technology and classification algorithms.

The extraction of spectral indices from multi-band imagery has been a fundamental pre-processing step in LULC classification. The Normalized Difference Vegetation Index (NDVI) remains the most widely used index for vegetation mapping, while the Normalized Difference Built-up Index (NDBI) and the Modified Normalized Difference Water Index (MNDWI) are commonly employed for delineating built-up areas and water bodies, respectively [14]. Recent studies have demonstrated that combining multiple indices with original spectral bands as input features significantly enhances classification accuracy [15].

### B. Machine Learning in LULC Classification

The application of machine learning to remote sensing classification has been extensively reviewed by Maxwell et al. [8] and Talukdar et al. [16]. Random Forest, an ensemble of decision trees introduced by Breiman [17], has become one of the most popular algorithms for LULC classification due to its resistance to overfitting, ability to handle high-dimensional data, and provision of feature importance rankings. Belgiu and Dragut [18] reported that RF consistently outperforms traditional parametric classifiers, with overall accuracy improvements of 5–15% across diverse study areas and classification schemes.

Deep learning approaches, particularly CNNs, have introduced a paradigm shift in remote sensing image analysis by eliminating the need for manual feature engineering [9]. Studies by Zhu et al. [19] demonstrated that CNN architectures adapted for remote sensing, such as patch-based classification networks and fully convolutional networks, can achieve overall accuracies exceeding 90% even in complex urban environments. However, CNNs typically require large labeled training datasets and significant computational resources, which may limit their applicability in resource-constrained settings [20].

### C. GIS-Based Urban Growth Modeling

Urban growth modeling seeks to simulate and project the spatiotemporal patterns of urban expansion based on historical LULC data and driving factors [21]. Among the various modeling approaches, the cellular automata–Markov chain (CA-Markov) model has been widely adopted due to its ability to capture both the stochastic nature of land-use transitions (through Markov chain probabilities) and the spatial contiguity constraints of urban growth

(through cellular automata rules) [22]. The integration of CA-Markov models within GIS environments enables spatial visualization of projected growth scenarios and facilitates the incorporation of planning constraints such as protected areas, elevation, and slope [23].

## III. METHODOLOGY

### A. Study Area and Data Acquisition

The study focuses on a rapidly urbanizing metropolitan region in southern India, encompassing approximately 1,200 km² and exhibiting diverse land-use patterns including dense urban cores, peri-urban transitional zones, agricultural hinterlands, and riparian vegetation corridors. Multi-temporal satellite imagery was acquired from two primary sources: Landsat 5 TM (for 2000 and 2005), Landsat 8 OLI (for 2010 and 2015), and Sentinel-2 MSI (for 2020 and 2023). All images were selected during the post-monsoon dry season (January–March) to minimize atmospheric interference and ensure spectral consistency across dates [12]. Ancillary data including road networks, administrative boundaries, elevation (SRTM DEM at 30 m), and population density rasters were obtained from publicly available geospatial databases.

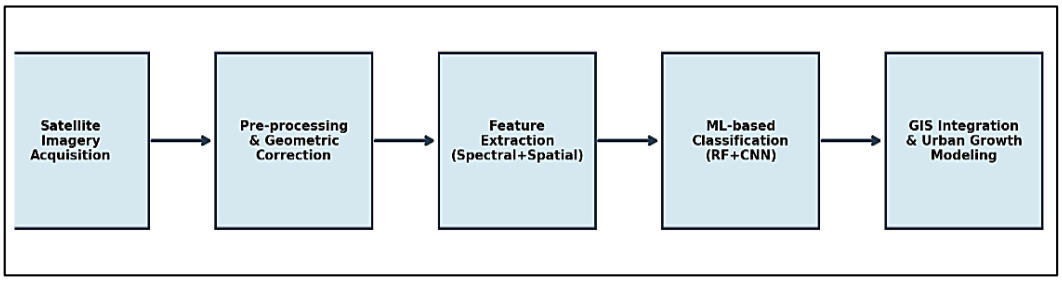


Fig. 1: Proposed GIS-ML methodology workflow for land-use classification and urban growth modeling.

### B. Pre-processing and Feature Extraction

All satellite images were subjected to atmospheric correction using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm to convert digital numbers to surface reflectance values [6]. Geometric co-registration was performed to ensure sub-pixel alignment across all temporal datasets, using ground control points identified from high-resolution Google Earth imagery. From the corrected reflectance bands, six spectral indices were computed: NDVI, NDBI, MNDWI, the Soil-Adjusted Vegetation Index (SAVI), the Enhanced Built-up and Bareness Index (EBBI), and the Normalized Difference Tillage Index (NDTI) [14]. These indices, along with the original spectral bands and texture features derived from the Gray-Level Co-occurrence Matrix (GLCM), constituted the input feature set for classification.

### C. Hybrid RF-CNN Classification

The proposed classification framework employs a two-stage hybrid approach. In the first stage, a patch-based CNN is trained on 32×32 pixel patches extracted from the satellite imagery to learn hierarchical spatial features. The CNN architecture consists of three convolutional layers (with 32, 64, and 128 filters, respectively) followed by max-pooling, batch normalization, and a global average pooling layer. The 128-dimensional feature vector output by the CNN is concatenated with the spectral indices and GLCM texture features to form a comprehensive feature representation for each pixel [9]. In the second stage, a Random Forest classifier with 500 trees is trained on this fused feature vector to produce the final LULC classification. This hybrid strategy leverages the CNN's ability to capture complex spatial patterns while benefiting from RF's robustness and interpretability [17].

### D. Urban Growth Modeling

The CA-Markov model was employed to project future urban growth scenarios for 2030 and 2040. Markov chain transition probability matrices were computed from the classified LULC maps of 2010 and 2020, capturing the probability of each land-use class transitioning to another over a 10-year period [22]. The cellular automata component incorporates spatial proximity rules based on a 5×5 neighborhood filter, ensuring that projected urban expansion respects spatial contiguity constraints. Additionally, suitability maps generated through multi-criteria evaluation (MCE) within the GIS environment incorporate factors such as distance to roads, slope, elevation, and proximity to existing built-up areas as determinants of urbanization potential [23].

## IV. RESULTS AND DISCUSSION

### A. Classification Accuracy Assessment

The classification accuracy of the proposed hybrid RF-CNN approach was evaluated against four baseline methods using standard metrics computed from a stratified random sample of 1,500 validation points per class. Table 1 presents the comparative results across all methods. The proposed approach achieved the highest overall accuracy (OA) of 93.4% and Kappa coefficient of 0.91, representing a significant improvement over the standalone methods. The Maximum Likelihood Classifier yielded the lowest accuracy (OA = 78.2%, Kappa = 0.72), consistent with its known limitations in handling spectrally overlapping classes. SVM (OA = 84.5%) and standalone RF (OA = 87.1%) showed progressively better performance, while the standalone CNN achieved 89.6% accuracy [8], [9].

Table 1. Comparison of Land-Use Classification Methods

| Method | OA (%) | Kappa | Producer's Acc. (%) | User's Acc. (%) |
|---|---|---|---|---|
| MLC | 78.2 | 0.72 | 75.8 | 76.5 |
| SVM | 84.5 | 0.80 | 82.3 | 83.1 |
| Random Forest | 87.1 | 0.84 | 85.6 | 86.2 |
| CNN | 89.6 | 0.87 | 88.1 | 88.9 |
| Proposed RF-CNN | 93.4 | 0.91 | 92.0 | 92.7 |

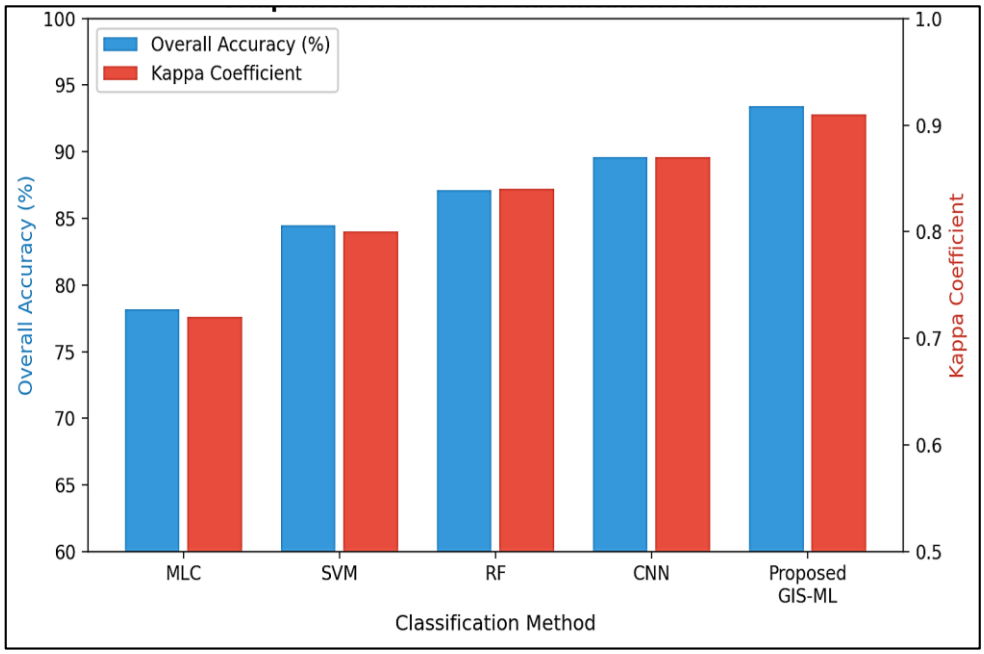


Fig. 2: Comparative performance of land-use classification methods in terms of overall accuracy and Kappa coefficient.

The per-class accuracy analysis revealed that the proposed method demonstrated the most substantial improvement in discriminating between built-up areas and barren land, two classes that are notoriously difficult to separate due to their spectral similarity in the shortwave infrared region. The CNN component's ability to capture contextual spatial patterns such as the regular geometric patterns characteristic of urban structures proved instrumental in resolving this ambiguity. Similarly, the fusion of NDVI and SAVI indices with CNN features enhanced the separation of vegetation subclasses, including dense forest and sparse agricultural vegetation [15].

### B. Multi-Temporal Land-Use Change Analysis

The classified LULC maps for the six temporal epochs (2000–2023) reveal a dramatic transformation in the study area's landscape. The built-up area increased from 15% of the total area in 2000 to 38% in 2023, representing a 153% increase over 23 years. Conversely, vegetation cover decreased from 38% to 19%, and agricultural land contracted from 32% to 25% [12]. Water bodies showed a modest decline from 8.5% to 6%, attributable to the encroachment of urban development into floodplain areas and the infilling of smaller water features. The rate of built-up area expansion was not uniform; the most rapid growth occurred during the 2010–2020 decade, coinciding with significant infrastructure development including highway construction and industrial zone establishment [2].

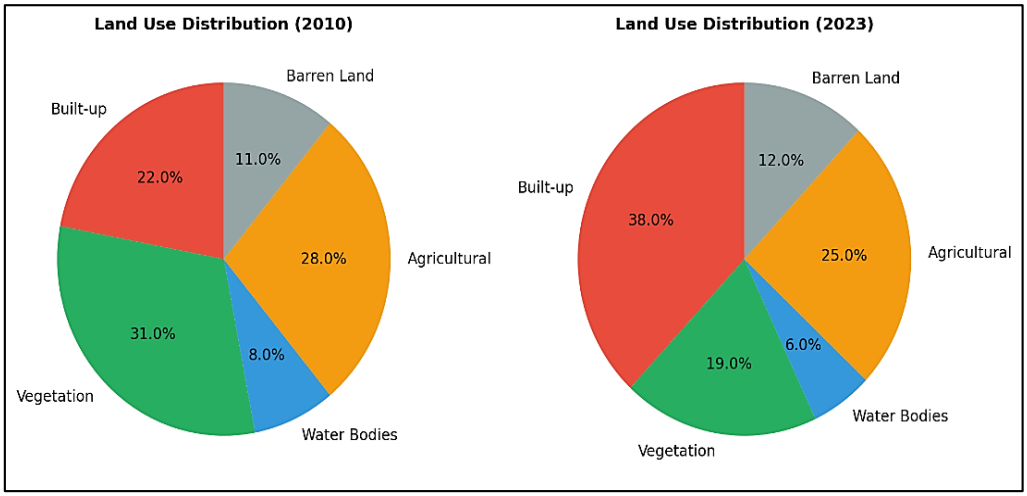


Fig. 3: Comparison of land-use distribution between 2010 and 2023, showing significant increase in built-up area.

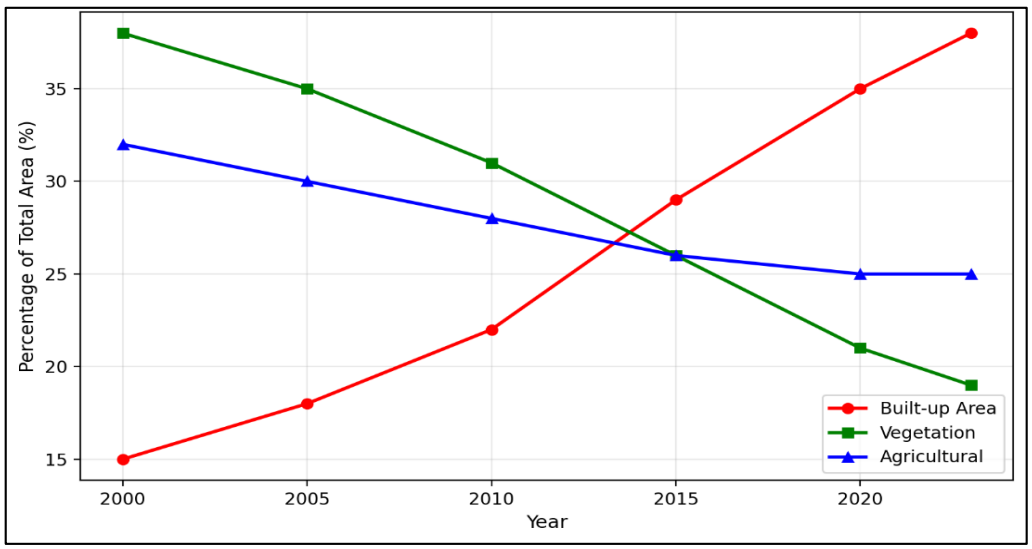


Fig. 4: Multi-temporal trends in major land-use classes from 2000 to 2023.

The spatial pattern of urban expansion exhibits a characteristic concentric growth pattern emanating from the historical city center, with secondary nuclei forming along major transportation corridors. The most intensive conversion of agricultural land to built-up area occurred in the northern and eastern peri-urban zones, driven by proximity to the national highway and the establishment of information technology parks. The southern sector, constrained by hilly terrain and a designated forest reserve, showed comparatively limited urban encroachment [21].

### C. Urban Growth Projection

The CA-Markov model, validated against the 2020 classified map with a Kappa agreement of 0.86, projects that built-up areas will increase to approximately 47% by 2030 and 54% by 2040 under a business-as-usual scenario. Table 2 presents the projected land-use distribution for both time horizons. This trajectory implies that over half of the study area will be urbanized by 2040, with vegetation cover potentially declining to under 12% [22], [23]. These projections underscore the urgency of proactive land-use planning interventions.

Table 2. Current and Projected Land-Use Distribution

| Land Use Class | 2023 (%) | 2030 Projected (%) | 2040 Projected (%) |
|---|---|---|---|
| Built-up | 38.0 | 47.2 | 54.1 |
| Vegetation | 19.0 | 15.3 | 11.8 |
| Agricultural | 25.0 | 21.8 | 19.5 |
| Water Bodies | 6.0 | 5.4 | 5.0 |
| Barren Land | 12.0 | 10.3 | 9.6 |

The model also generated an alternative managed-growth scenario incorporating planning constraints such as green belt preservation, agricultural zone protection, and infill development incentives. Under this scenario, built-up area expansion is moderated to 42% by 2030 and 46% by 2040, with vegetation cover maintained above 16%. The comparison between these scenarios provides planners with a quantitative basis for evaluating the impacts of different policy interventions on long-term land-use trajectories [3], [21].

### D. Implications for Sustainable Urban Planning

The findings of this study carry several significant implications for sustainable urban planning practice. First, the identification of hotspot areas of rapid land-use conversion enables targeted intervention through zoning regulations and development control orders. Second, the integration of environmental sensitivity indicators such as proximity to water bodies, slope steepness, and ecological corridor connectivity within the GIS framework allows planners to designate environmentally critical areas for protection [4], [5]. Third, the projected growth scenarios facilitate proactive infrastructure planning by identifying areas likely to require water supply, sewerage, transportation, and social infrastructure investments within specific time horizons. The framework's ability to generate multiple scenarios under different policy assumptions makes it a versatile decision support tool that can accommodate the inherently uncertain nature of urban development [10].

## V. CONCLUSION

This study presented an integrated GIS and machine learning framework for multi-temporal land-use classification and urban growth modeling. The hybrid RF-CNN classifier achieved an overall accuracy of 93.4%, demonstrating the value of combining deep spatial feature learning with ensemble classification. The multi-temporal analysis revealed a 153% increase in built-up area over 23 years, accompanied by a 50% decline in vegetation cover. The CA-Markov growth model projects that over half the study area will be urbanized by 2040 under current trends, but that managed-growth interventions could substantially moderate this trajectory [22].

The proposed framework contributes to the growing body of research on data-driven urban planning by providing a scalable, reproducible methodology that leverages freely available satellite imagery and open-source GIS tools. Future work will extend this framework to incorporate socioeconomic variables, real-time urban monitoring using high-frequency satellite data, and stakeholder engagement modules that translate analytical outputs into actionable planning recommendations. The integration of climate change projections with urban growth models represents another promising avenue for enhancing the sustainability orientation of GIS-based planning tools [1], [3].

## REFERENCES


[1] United Nations, *World Urbanization Prospects: The 2018 Revision*. New York, NY, USA: UN Department of Economic and Social Affairs, 2019.

[2] K. C. Seto, B. Güneralp, and L. R. Hutyra, "Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools," *Proc. Natl. Acad. Sci.*, vol. 109, no. 40, pp. 16083–16088, Oct. 2012.

[3] X. Li, Y. Zhou, G. R. Asrar, M. Imhoff, and X. Li, "The surface urban heat island response to urban expansion: A panel analysis for the conterminous United States," *Sci. Total Environ.*, vol. 605–606, pp. 426–435, Dec. 2017.

[4] P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, *Geographic Information Systems and Science*, 4th ed. Hoboken, NJ, USA: Wiley, 2015.

[5] J. R. Jensen, *Remote Sensing of the Environment: An Earth Resource Perspective*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2007.

[6] M. Drusch *et al*., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, May 2012.

[7] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*, 4th ed. Berlin, Germany: Springer, 2006.

[8] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, 2018.

[9] X. X. Zhu *et al*., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[10] M. Batty, "Urban modeling," in *International Encyclopedia of Geography*, D. Richardson *et al*., Eds. Hoboken, NJ, USA: Wiley, 2017.

[11] A. Singh, "Review article: Digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.

[12] M. A. Wulder *et al*., "Current status of Landsat program, science, and applications," *Remote Sens. Environ.*, vol. 225, pp. 127–147, May 2019.

[13] D. Phiri and J. Morgenroth, "Developments in Landsat land cover classification methods: A review," *Remote Sens.*, vol. 9, no. 9, Art. no. 967, Sep. 2017.

[14] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *Int. J. Remote Sens.*, vol. 24, no. 3, pp. 583–594, 2003.

[15] P. Thanh Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery," *Sensors*, vol. 18, no. 1, Art. no. 18, Jan. 2018.
[16] S. Talukdar *et al*., "Land-use land-cover classification by machine learning classifiers for satellite observations—A review," *Remote Sens.*, vol. 12, no. 7, Art. no. 1135, Apr. 2020.
[17] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[18] M. Belgiu and L. Drăguț, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, Apr. 2016.
[19] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing image classification: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 6, Art. no. e1264, Nov. 2018.
[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[21] C. He, N. Okada, Q. Zhang, P. Shi, and J. Li, "Modelling dynamic urban expansion processes incorporating a potential model with cellular automata," *Landsc. Urban Plan.*, vol. 86, nos. 1–2, pp. 79–91, May 2008.
[22] R. G. Pontius Jr. and L. C. Schneider, "Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA," *Agric. Ecosyst. Environ.*, vol. 85, nos. 1–3, pp. 239–248, Jun. 2001.
[23] J. J. Arsanjani, M. Helbich, W. Kainz, and A. Boloorani, "Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 21, pp. 265–275, Apr. 2013.