



Predictive Analytics for Urban Sprawl Using Machine Learning in Land Cover Mapping

Mohsen Niroomand¹ , Parham Pahlavani^{2✉}

1. Department of GIS, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. E-mail: k.niroomand@ut.ac.ir
2. Corresponding author, Department of GIS, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. E-mail: Pahlavani@ut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 2025-04-13
Received in revised form 2025-08-05
Accepted 2025-08-05
Published online 2025-10-19

Keywords:
Urban Sprawl,
Land Cover Change Modelling,
Machine Learning Algorithms,
dimensionality reduction,
Markov Chains.

ABSTRACT

Urban sprawl, characterized by low-density, uncoordinated, and outward urban expansion, presents critical challenges to sustainable development, particularly in rapidly growing metropolitan regions such as Tehran. This study aims to employ an integrated framework combining remote sensing, spatial urban sprawl indices, and advanced machine learning techniques to analyse and project land cover changes between 2011 and 2026.

Initial land cover maps for the years 2011, 2016, and 2021 were generated using the Random Forest (RF) algorithm applied to Landsat 7 and 8 imagery, achieving overall classification accuracies of 92.53%, 93.27%, and 93.88%, respectively. Subsequently, a comprehensive set of urban sprawl indices—derived from census data, transportation networks, and land use parcel information—was utilized alongside land cover transition maps to train Multi-Layer Perceptron (MLP), Decision Forest (DF), and Support Vector Machine (SVM) models within a Markov chain framework. Dimensionality reduction techniques, including Principal Component Analysis (PCA) and Independent Component Analysis (ICA), were applied to enhance model efficiency.

Among the evaluated models, the MLP trained with the complete feature set demonstrated superior performance, attaining an F1-score of 83.95%. The projections suggest a 6% increase in built-up areas by 2026, predominantly at the expense of barren lands and green spaces.

The results underscore the potential of integrating geospatial technologies with machine learning methodologies to support data-driven urban planning and the formulation of sustainable land management policies in rapidly urbanizing contexts.

Cite this article: Niroomand, M & Pahlavani, P. (2025). Predictive Analytics for Urban Sprawl Using Machine Learning in Land Cover Mapping, *Earth Observation and Geomatics Engineering*, Volume 8, Issue 2, Pages 51-66. <http://doi.org/10.22059/eoge.2025.393376.1173>



© The Author(s).

DOI: <http://doi.org/10.22059/eoge.2025.393376.1173>

Publisher: University of Tehran.

1. Introduction

Urban sprawl emerged as a prominent phenomenon in the mid-1960s, driven by rapid population growth and industrialization. Ewing's widely cited definition characterizes urban sprawl as a pattern of spatial development predominantly occurring in open spaces, rural areas, and the peripheries of large cities (Ewing, 1997). This form of growth is propelled by factors such as demographic expansion, economic development, increased automobile dependency, infrastructure investment, and socio-political dynamics, culminating in the extensive outward expansion of urban areas (Banai & DePriest, 2014; Yasin et al., 2021).

The unregulated nature of this growth, particularly in developing nations, presents multifaceted environmental, social, and economic challenges (Rana & Sarkar, 2021). Low-density, dispersed urban development contributes to the loss of agricultural land, ecosystem fragmentation, elevated levels of air and water pollution, and heightened susceptibility to climate change (Das et al., 2023). Such expansion induces significant land use and land cover (LULC) transformations, resulting in the conversion of natural landscapes into impervious built environments. These changes exacerbate surface temperature increases, intensify urban heat island effects, and degrade overall environmental quality (Moniruzzam et al., 2018).

These challenges become particularly pronounced in regions where urbanization outpaces the development of infrastructure and public services, especially in developing countries (Barman et al., 2024). Empirical studies conducted across a range of geographic contexts underscore the adverse impacts of unchecked urban expansion and reinforce the need for integrated planning approaches (Rimal et al., 2018). Sustainable urban management strategies are imperative and should encompass principles of smart growth, including support for infill development, diversification of transportation options, and inclusive housing policies. Furthermore, regional planning and community engagement are essential in shaping balanced urban development and mitigating environmental degradation (Blair & Wellman, 2017; Duany et al., 2000).

Nowadays, advancements in remote sensing and Geographic Information Systems (GIS) have facilitated comprehensive monitoring of urban expansion processes, offering high-resolution, spatiotemporal data to inform decision-making (Barman et al., 2024; Duany et al., 2000; Yin et al., 2011). Analytical techniques such as Shannon's entropy and the CA-Markov modelling framework have proven effective in quantifying the intensity and spatial pattern of urban sprawl (Baqi et al., 2021; Barman et al., 2024). The integration of satellite imagery, demographic datasets, and machine learning algorithms has further enhanced the precision and interpretability of sprawl analyses, enabling the identification of at-risk areas and the development of targeted policy interventions (Gómez et al.,

2019; Kulkarni & Vijaya, 2022). Numerous studies have utilized these tools and methodologies to examine urban expansion dynamics across diverse geographical areas.

(Barman et al., 2024) conducted a study focusing on the Jalpaiguri region in India, utilizing Landsat satellite imagery, urban metrics, and Shannon's entropy analysis to document the substantial expansion of built-up areas over a span of two decades. In a separate study, (Shi et al., 2023) examined land use, land use efficiency, and population density to evaluate the urban sprawl trend in Shanghai, China, from 1990 to 2020. Their findings indicated that urban sprawl peaks at a distance of 20 to 30 kilometers from the city center, with housing provision methods significantly influencing its extent and intensity.

(Dhanaraj & Angadi, 2022) employed Landsat satellite imagery to analyze the urban expansion of Mangaluru in India, revealing a growth pattern aligned with transportation networks, as indicated by calculated Shannon entropy values. (Rana & Sarkar, 2021) investigated urban growth in Pabna, Bangladesh, utilizing Landsat images and the Maximum Likelihood algorithm. Their study included future development projections using the CA-Markov model and fuzzy set theory, emphasizing the role of satellite imagery and predictive modeling in comprehending urban expansion dynamics.

(Biney & Boakye, 2021) applied the Shannon entropy method to study urban sprawl dynamics in the Sekondi-Takoradi region of Ghana, highlighting accelerated settlement growth post-oil discovery and the economic drivers influencing urban development. (Dadashpoor & Salarian, 2020) utilized the SLEUTH model and land conversion models to analyze urban sprawl in the Mazandaran province of Iran, predicting a surge in built-up areas by 2040 and stressing the importance of effective policy implementation to address urban expansion consequences.

(Dinda et al., 2019) employed Shannon's entropy and Markov chains to map and forecast urban growth in Midnapore, India, emphasizing socio-economic factors' role in unchecked city expansion. (Rimal et al., 2018) utilized the Support Vector Machine (SVM) algorithm and the CA-Markov model to investigate urban expansion in central Nepal and predict future land use changes. (Moniruzzam et al., 2018) analysed land use changes in Khulna, Bangladesh, using Landsat 8, Landsat 7, and Landsat 5 satellite images alongside supervised classification algorithms, noting a significant increase in built-up areas over two decades.

The reviewed literature highlights the essential role of advanced geospatial technologies and modelling approaches in analyzing and predicting urban land use transformations. Despite the diversity in regional contexts, a consistent finding across studies is the profound impact of urban expansion on natural ecosystems and agricultural lands. This convergence underscores the urgent need for informed urban planning and policy measures to ensure

sustainable development. Satellite imagery, GIS, and spatial modelling techniques have proven indispensable in generating actionable insights for managing the complexities of urban sprawl. Given the accelerating pace of urbanization in developing countries and its associated economic, social, and environmental consequences, the present study focuses on assessing urban sprawl in Tehran. It aims to generate predictive land cover maps grounded in established sprawl metrics, guided by a comprehensive review of spatial indicators employed in prior research. The study pursues the following objectives.

- Identifying effective spatial indicators for measuring urban sprawl
- Land Cover Change Modelling Based on Urban Sprawl Measurement Indices
- Evaluating the efficiency and performance of machine learning models through the implementation of dimensionality reduction techniques
- Developing predictive land cover maps

In this context, the present study offers a novel methodological contribution by integrating remote sensing data with a diverse set of spatial and socioeconomic indices of urban sprawl, employing dimensionality reduction techniques, and implementing multiple machine learning algorithms within a Markov chain modeling framework. Through a comparative analysis of the predictive performance of Multi-Layer Perceptron (MLP), Decision Forest (DF), and SVM models using both original and reduced feature sets, the study identifies the most effective approach for land cover change prediction. Moreover, it provides valuable insights into the spatial dynamics of urban expansion in Tehran. The proposed framework aims to support urban planning by offering a data-driven tool capable of forecasting future land transformations and informing more sustainable development strategies in rapidly urbanizing regions.

The structure of this study is organized as follows. Section 2 introduces the methodological framework, detailing the processes of land cover mapping, the selection of urban sprawl indices, the implementation of machine learning models, and the application of the Markov chain approach. Section 3 provides a description of the study area and the datasets employed. Section 4 presents the experimental setup, results, and model evaluation. Section 5 offers a discussion of the findings in the context of urban planning. Finally, Section 6 concludes the study and outlines potential directions for future research.

2. Methodology

To generate land cover prediction maps for the study area, this research employed initial land cover maps in conjunction with urban sprawl measurement criteria. The methodology, outlined in Figure 1, integrated remote sensing data, urban sprawl indices, machine learning algorithms, and the Markov chain modelling approach to analyse land cover changes over the specified period.

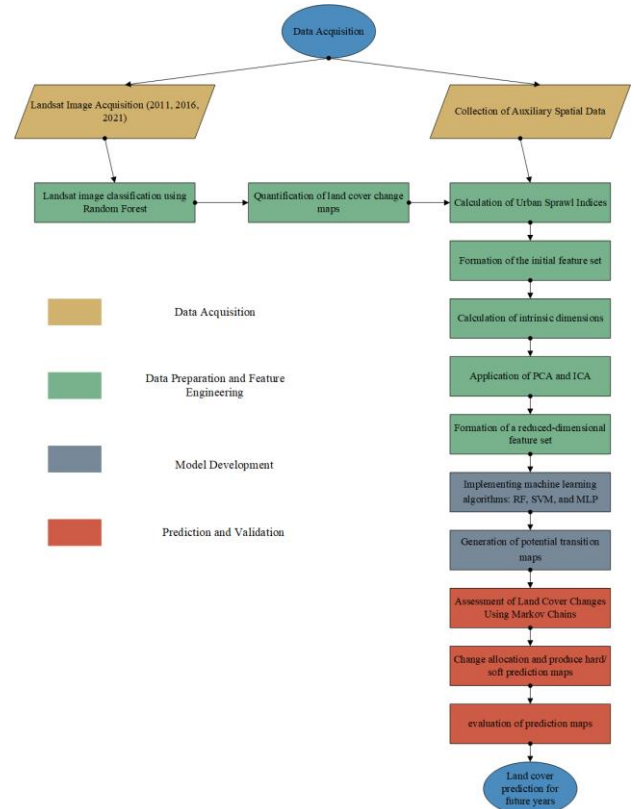


Figure 1. Flowchart of the study process

2.1. Producing initial land cover maps

This research used Landsat 7 and Landsat 8 satellite imagery to produce initial land cover maps of the study area. To enhance the quality of the imagery, atmospheric correction techniques were applied to reduce the effects of aerosols, and cloud masking algorithms were used to eliminate cloud-contaminated pixels. Subsequently, a series of spectral indices were derived to support the classification

process, including EVI¹ (Tatsumi et al., 2015), NBR² (Escuin et al., 2008), NDMI³ (Herbei & Sala, 2016), NDWI⁴ (Kshetri, 2018), NDBI⁵ (Estoque & Murayama, 2015), NDBaI⁶ (Trinh, 2020), NDVI⁷ (Taufik et al., 2016), SAVI⁸ (Oon et al., 2019), GNDVI⁹ (Shaver et al., 2006), as well as Wetness and Greenness components (Hislop et al., 2018). These spectral indices were selected due to their effectiveness in highlighting specific land surface characteristics relevant to land cover classification. For instance, NDVI, EVI, SAVI, and GNDVI are widely used to assess vegetation health and density, while NDBI and NDBaI help in distinguishing built-up and barren areas. NDMI and NDWI are sensitive to soil and vegetation moisture, supporting the detection of water bodies and wet soils, and NBR is particularly useful in identifying disturbance areas such as burned land. By incorporating this diverse set of indices, the classification process gained robustness and improved discriminatory power across land cover types. Combined with the spectral bands and elevation data extracted from the Shuttle Radar Topography Mission (SRTM), this suite of features formed the comprehensive feature set for analysis.

Pixel-based samples representing each land cover class were collected and partitioned into training and testing subsets to facilitate supervised classification. The Random Forest (RF) algorithm was then applied to classify the imagery for each target year (Biau & Scornet, 2016; Breiman, 2001). The RF algorithm was chosen for the initial land cover classification owing to its demonstrated robustness in processing noisy and high-dimensional satellite data. Its capacity to capture complex, non-linear relationships, combined with a low risk of overfitting, made it particularly suitable for this task. Moreover, the ensemble nature of RF contributed to enhanced classification accuracy and stability across heterogeneous land cover classes. The land cover classes examined in this study are presented in Table 1.

Table 1. Classification of Land Cover Types in Landsat Imagery Analysis.

Class	Description	Contains
Built-up	Areas dominated by human-made structures, including residential, commercial, and industrial zones.	Houses, buildings, roads, pavements, urban infrastructure
Barren	Land with very sparse to no	Bare soil, rocks, sand, desert

	vegetation, often found in natural deserts or areas of extreme land use.	landscapes, quarries, exposed ground with minimal to no vegetation
Water	Bodies of water, including both natural and artificial waterbodies.	Lakes, rivers, reservoirs, ponds, swimming pools
Greenery	Areas with significant vegetation cover, not cultivated for agriculture.	Forests, parks, gardens, grasslands, natural vegetation areas
Cropland	Areas used for the cultivation of crops and agricultural production.	Fields of corn, wheat, rice paddies, vegetable plots, orchards, vineyards

2.2. Modelling and Predicting Land Cover Maps Based on Machine Learning and Markov Chain

To model and predict land cover dynamics, this study employed machine learning algorithms alongside a feature set comprising urban sprawl metrics and quantified land cover change maps to develop transition potential surfaces. These surfaces served as inputs to a Markov chain model, which estimated the probabilities of transitions between land cover classes over time. This section presents the methodological framework for generating predictive land use and land cover maps using this integrated approach.

2.2.1. Feature Sets Employed for Machine Learning Model Training

This study used two primary feature sets for training machine learning models. The first set comprised demographic, geographic, and geometric attributes, selected as indices of urban sprawl to evaluate their effectiveness in predicting land cover changes. The selection of these indicators was informed by a comprehensive review of existing literature and empirical findings from previous studies. They collectively represent key dimensions of urban sprawl, including density, spatial configuration, accessibility, and socioeconomic conditions, and have been shown to be effective in capturing patterns of urban growth in various contexts.

¹ Enhanced Vegetation Index

² Normalized Burn Ratio

³ Normalized Difference Moisture Index

⁴ Normalized Difference Water Index

⁵ Normalized Difference Building Index

⁶ Normalized Difference Bareness Index

⁷ Normalized Difference Vegetation Index

⁸ Soil Adjusted Vegetation Index

⁹ Green Normalized Difference Vegetation Index

The second set included quantified transition maps of land cover classes, representing spatial dynamics and transformation patterns over time. The following sections present a detailed analysis of each feature group.

- urban sprawl measurement indices

Urban sprawl indices were analysed by examining the underlying drivers, structural characteristics, and associated consequences of this phenomenon. While initial studies primarily focused on variables such as population growth and land use changes, subsequent research adopted a more comprehensive approach by incorporating a broader range of criteria. In 1998, Club identified four principal indicators: migration to suburban areas, the ratio of population growth to land consumption, increased traffic congestion, and the reduction of open and undeveloped land (Club, 1998).

Recent studies expanded upon these foundations, employing a variety of metrics—often supported by GIS and statistical techniques—to evaluate urban sprawl. These metrics encompassed demographic trends, land development patterns, transportation accessibility, resource consumption, and urban spatial structure, including zoning and service distribution. Commonly utilized indices include population growth rate, extent of spatial expansion, population and residential density, employment patterns, and accessibility indicators. The specific indices adopted in this study are summarized in Table 2.

Table 2. Summary of Features Used for Model Training

Category	Index	Calculation Method	Parameter Definition
Topography	Elevation (Kumar, 2017)	Digital Elevation Model (30 m resolution)	-
	Slope (Herold et al., 2003)	Slope derived from Digital Elevation Model (30 m resolution)	-
Density	Vertical Density (Jiang et al., 2007)	$VD_i = \frac{P_i}{n \times P_a}$	P_i : population in settlement i n : average household size P_a : number of building parcels
	Gross Population Density (Jiang et al., 2007)	$DG_i = \frac{P_i}{UA_i}$	P_i : population in settlement i UA_i : built-up area

	Net Population Density (Frenkel & Ashkenazi, 2008)	$DN_i = \frac{P_i}{RA_i}$	P_i : population in settlement i RA_i : residential built-up area
Urban Geometry	Fractal Dimension (Frenkel & Ashkenazi, 2008)	$F_i = \frac{2 \times \log L_i}{\log A_i}$	L_i : perimeter A_i : area of built-up region i
	Shape Index (Frenkel & Ashkenazi, 2008)	$SH_i = \frac{L_i}{2 \times \sqrt{\pi \times A_i}}$	L_i : perimeter A_i : area of settlement i
Segmentation	Linear Development Index (LDI) (Jiang et al., 2007)	Proximity of new developments to highways	-
	Discontinuous Development Index (Jiang et al., 2007)	Distance between new and existing built-up areas	-
Social and Economic	Immigration Rate (Zhang et al., 2022)	Immigrant ratio indicating urban growth and diversity	-
	Labour Force Participation Rate (Zhang et al., 2022)	Proportion of employed individuals	-
	Housing Value (Hatab et al., 2019)	Market price analysis	-
Composition	Land Use Mix (Frenkel & Ashkenazi, 2008)	$LU_{ij} = \frac{a_{ij}}{A_i}$	a_{ij} : area of land use type j in settlement i A_i : total area of settlement i

- Quantification of land cover changes through the ELT Method

This study employed the Evidence-Based Land Transformation (ELT) method, which is grounded in Bayesian statistical theory, to model and quantify changes in land cover between two temporal land cover maps (Royall, 2017). The procedure involves the following steps.

1. Categorical Change Detection: A comparative analysis of the two land cover maps was conducted to identify and map transitions between different land cover classes.

2. **Boolean Reclassification:** The resulting change map was reclassified into binary form, where each pixel is designated as either changed or unchanged.
3. **Application of ELT:** These binary maps were then used as evidence layers in the ELT framework to estimate the probability of change, thereby converting qualitative spatial observations into quantitative metrics.

2.2.2. Dimension reduction

Dimensionality reduction plays a vital role in facilitating the analysis of high-dimensional data, particularly as datasets increase in size and complexity. This process is typically approached through two main strategies: feature selection and feature extraction (Ma & Zhu, 2013).

Feature extraction is based on the assumption that the response variable $\mathbf{Y}$ is associated with several linear combinations of the predictor variables $\mathbf{x}$. The objective is to identify these combinations and project the original feature space into a lower-dimensional subspace, as represented by Equation (1).

$$pr(\mathbf{Y} \leq \mathbf{y} | \mathbf{x}) = pr(\mathbf{Y} \leq \mathbf{y} | \beta^T \mathbf{x}) \quad (1)$$

In this equation, β denotes a $p \times d$ matrix that transforms the original p -dimensional feature space into a d -dimensional subspace ($d \ll p$), retaining the essential information relevant to $\mathbf{Y}$. The minimal such subspace, known as the central subspace, is identified through the estimation of the intrinsic dimension using Maximum Likelihood Estimation (MLE) (Karbauskaitė & Dzemyda, 2013).

In this study, dimensionality reduction techniques are employed to optimize the urban sprawl indicators, thereby enhancing computational efficiency and enabling more effective training of machine learning models.

- **Principal Component Analysis**

Principal Component Analysis (PCA) is a statistical technique used to reduce the dimensionality of high-dimensional data while preserving as much variance as possible. It transforms the original correlated variables into a new set of uncorrelated variables, known as principal components, which are linear combinations of the original features.

Given a dataset with p -dimensional vectors, PCA projects the data into a d -dimensional subspace as defined in Equation (2) (Nabi & Zhou, 2024).

$$\mathbf{x} = \mathbf{W}(\mathbf{y} - \boldsymbol{\mu}) \quad (2)$$

where $\mathbf{x}$ denotes the transformed data, $\mathbf{y}$ is the original feature vector, $\boldsymbol{\mu}$ is the mean vector, and $\mathbf{W}$ is a $p \times d$ matrix consisting of the eigenvectors corresponding to the largest eigenvalues of the sample covariance matrix $\mathbf{S}$, expressed as Equation (3) (Nabi & Zhou, 2024).

$$\mathbf{S} = \frac{1}{N} \sum_{i=1}^N (\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^T \quad (3)$$

These eigenvectors satisfy $\mathbf{S}\mathbf{v} = \lambda\mathbf{v}$, where λ are the eigenvalues. In the reduced space, the components are uncorrelated, and the covariance matrix becomes diagonal, with the eigenvalues indicating the variance explained by each principal component (Nabi & Zhou, 2024).

- **Independent Component Analysis**

Independent Component Analysis (ICA) is a computational approach designed to decompose a multivariate signal into a set of statistically independent components. It is widely applied in areas such as signal processing and feature extraction due to its effectiveness in uncovering latent structures within complex datasets (Pokorny et al., 2023). The transformation is represented by Equation (4), where the observed signals $\mathbf{X}$ are mapped to independent components $\mathbf{Y}$ through the de-mixing matrix $\mathbf{W}$ (Zhang & Chan, 2005).

$$\mathbf{Y} = \mathbf{W}\mathbf{X} \quad (4)$$

In contrast to techniques that only ensure uncorrelated outputs, ICA emphasizes statistical independence. The estimation of $\mathbf{W}$ is based on maximizing the non-Gaussianity of the components, as informed by the central limit theorem, which posits that the sum of independent non-Gaussian variables tends to approximate a Gaussian distribution. ICA therefore seeks to identify components that exhibit maximal non-Gaussianity, typically using negentropy as a quantitative measure of independence (Cao et al., 2003).

2.2.3. Modelling Potential Transitions between Land Cover Classes

To assess the dynamics of urban expansion in the study area, potential transitions between land cover classes were modelled using three machine learning algorithms: MLP (Chan et al., 2001; Mather & Tso, 2016), DF (Biau & Scornet, 2016), and SVM (Awad & Khanna, 2015). These algorithms were selected based on their proven effectiveness in land cover classification and change modeling tasks. The MLP is capable of capturing complex nonlinear relationships within spatial data, the DF model provides robustness and interpretability, and the SVM performs well in high-dimensional spaces, particularly when training data are limited. Their application enables a comparative evaluation of model performance in capturing the spatial and temporal patterns of urban expansion. The input feature

set includes quantified land cover change maps and urban sprawl measurement indices. Furthermore, to enhance model efficiency and performance, additional features were derived by applying PCA and ICA to the initial feature set. These transformed features contributed to improved classification accuracy by reducing dimensionality while preserving essential information.

Specifically, the machine learning models were trained using spatial indices of urban sprawl and quantified transition maps as input variables. The output consisted of pixel-wise probability surfaces that represent the likelihood of transition to each land cover class. These probability surfaces form the basis for generating predictive land cover maps in the subsequent stages of the analysis.

2.2.4. Markov Chain Modelling of Land Cover Transitions

Markov chains offer a robust framework for modelling stochastic processes in which the future state of a system depends solely on its current state, a property known as the Markov assumption (Tolver, 2016). This characteristic makes them particularly suitable for analysing land cover change, as it simplifies the estimation of transition probabilities between land cover classes. The possible states of the system are defined as Equation (5) (Tolver, 2016).

$$S = \{S_1, S_2, \dots, S_n\} \quad (5)$$

The transition probabilities between states are represented in a matrix form as Equation (6) (Liping et al., 2018).

$$P_{ij} = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix} \quad (6)$$

Where p_{ij} denotes the probability of transition from state S_i to S_j , and n is the number of land cover classes. The land cover state at time $t+1$ is then computed by multiplying the transition matrix with the state vector at time t as Equation (7) (Liping et al., 2018).

$$S_{t+1} = P_{ij} \times S_t \quad (7)$$

Transition matrices are derived from successive land cover maps and are normalized to annual probabilities to ensure temporal consistency and comparability across different time intervals (Liping et al., 2018).

In this framework, the input to the Markov chain model comprises two successive land cover maps, from which the algorithm calculates a transition probability matrix. The output is a projected distribution of land cover classes for a future year, assuming that future changes depend solely on the current state configuration.

2.2.5. Change allocation

The final predictive map was constructed by integrating the potential transition maps with the land cover class transition probability matrices, which are derived through the application of the Markov chain model. The allocation of land cover classes to each pixel was carried out using two distinct approaches: Hard prediction and Soft prediction.

The hard prediction approach assigns each pixel to the land cover class with the highest predicted probability, generating a categorical map that represents the most likely land cover outcome for each location. This method employs the Multi-Objective Land Allocation (MOLA) algorithm, wherein each land cover transition is conceptualized as a spatial shift from a host class (experiencing area loss) to a claimant class (gaining area), as defined by the transition matrix. Land is subsequently reallocated to satisfy projected demands in accordance with these transitions.

In contrast, the soft prediction method yields a continuous probability surface, where each pixel is associated with a full set of likelihood values corresponding to all possible land cover classes. This probabilistic representation captures the inherent uncertainty and complexity of urban transformation by accommodating multiple potential transition pathways. Pixels influenced by multiple driving factors may demonstrate elevated cumulative probabilities of change, rendering this method particularly valuable in exploratory modeling and policy-sensitive planning scenarios.

2.2.6. Validation

The validation of the methods utilized in this study was conducted in two phases. The first phase involved assessing the ability of the RF algorithm to generate accurate land cover maps. This is achieved by employing various performance metrics and evaluating the significance of each feature used in the model. The second phase involved a comparison of the maps generated for 2021 using the Hard Prediction process and different machine learning methods with the reference map for the same year. The comparison was carried out using multiple metrics to evaluate the effectiveness of the feature set and the performance of the different algorithms. The validation metrics applied in this study include Precision (Borenstein, 2001), Accuracy (Sokolova et al., 2006), F1-score (Sokolova et al., 2006), and Mean Squared Error (MSE) (Marmolin, 1986).

3. Study area and Dataset

3.1. Study area

Tehran, the capital of Iran and its primary political, economic, and cultural center, is located in Tehran Province and covers a geographic extent approximately between 35.5°N to 35.9°N latitude and 51.2°E to 51.6°E longitude.

In recent decades, the city has experienced considerable demographic changes, consistent with national trends in urbanization and economic growth. According to data from the Tehran Municipality, the population of Tehran increased from 8,154,051 in 2011 to 8,693,706 in 2016, with the annual growth rate rising from 0.9% to 1.3%. This upward trend is attributed to both natural population growth and significant in-migration, as 996,404 individuals moved to the city during this five-year period. These dynamics render Tehran particularly susceptible to the phenomenon of urban sprawl. The spatial extent of the study area is presented in Figure 2.

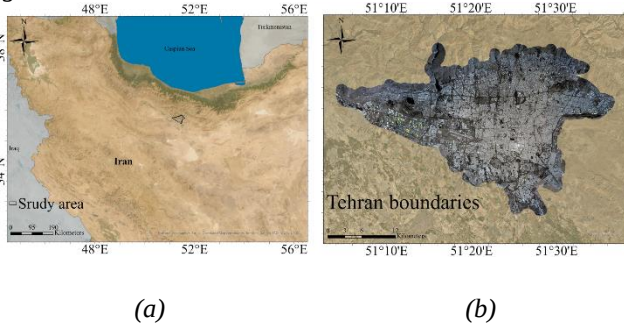


Figure 2. Study area location: (a)Tehran within Iran and (b)its administrative boundaries

3.2. Dataset

This study utilizes a comprehensive dataset to examine rapid urbanization, demographic shifts, and urban sprawl in Tehran, including satellite imagery, census data, land-use parcels, and road maps. All spatial datasets and land cover maps utilized in this study were projected using the Universal Transverse Mercator (UTM) coordinate system, Zone 39N, referenced to the WGS84 datum. This coordinate reference system, which employs the Transverse Mercator projection, was chosen to maintain spatial consistency and to ensure accurate area calculations throughout the study area.

3.2.1. Satellite Imagery

Landsat 8 imagery from 2016 and Landsat 7 imagery from 2011, both with a spatial resolution of 30 meters across multispectral bands, are employed. This resolution enables detailed analysis of urban structures, green cover, croplands, barren lands, and water bodies. The satellites provide images every 16 days, ensuring a consistent dataset for monitoring temporal changes. Although Landsat 7 data may exhibit gaps due to sensor malfunctions, the uniform resolution between the two satellites permits effective comparison of urban development and land-use changes in Tehran over the five-year period.

3.2.2. Census Data

Data from the Population and Housing Censuses of Tehran for 2011 and 2016 are utilized, structured in shape file statistical blocks as defined by the Statistical Centre of

Iran. These blocks, delineated by public access ways or natural barriers, contain vital information on geographic, household, economic, population, and housing statistics.

3.2.3. Land Use Data

The land use map of Tehran, provided by the municipality in shape file format, delineates the boundaries and land-use types for each registered plot across the city. These parcels are fundamental for defining criteria to assess urban sprawl. To ensure compatibility with other datasets, this map is utilized at a spatial resolution of 30 meters.

3.2.4. Road Maps

Road maps of Tehran, extracted from Open Street Map (OSM), are used to analyse traffic patterns, access to infrastructure, and connectivity between urban areas.

4. Experiments

4.1. Land Cover Classification Using Random Forest

In this study, land cover changes were analysed using RF classification applied to Landsat 7 and Landsat 8 imagery for the years 2011 and 2016, with the 2021 map serving as a reference. The RF model was configured with 300 decision trees to optimize the trade-off between computational efficiency and classification accuracy. At each node, the number of features considered for splitting was set to the square root of the total number of input features to reduce bias. Furthermore, the Bootstrapping technique was employed, whereby 50% of the training data was randomly sampled for each tree to ensure diversity and improve generalizability.

The model's performance was evaluated based on overall accuracy, the Kappa coefficient, and the out-of-bag (OOB) error. To provide a clearer understanding of the classification context, the number of training and test samples utilized for each year is also reported. These results are presented in Table 3.

Table 3. Performance metrics of the Random Forest classification for each image.

Imagery	Train Sample Size	Test Sample Size	OOB	Accuracy	kappa
Landsat7 (2011)	32586	8424	0.027	92.53%	0.897
Landsat8 (2016)	35019	8957	0.033	93.27%	0.905
Landsat8 (2021)	32602	7690	0.028	93.88%	0.917

The F1-scores for each land cover class are presented in Table 4, where the model shows excellent classification accuracy for built-up areas and water bodies. However, the F1-scores for barren lands, green spaces, and croplands demonstrate some variability, which could be attributed to

the challenges posed by mixed pixels or spectral similarities between these classes.

Table 4. F1-scores for each land cover class in the Random Forest classification.

Imagery	F1-Score				
	Cropland	Greenery	Water	Barren	Built-up
Landsat 7 (2011)	94.44%	91.03%	97.55 %	91.84%	93.61 %
Landsat 8 (2016)	89.64%	91.41%	98.81 %	94.39%	94.22 %
Landsat 8 (2021)	90.51%	94.07%	98.56 %	94.72%	93.71 %

The feature importance values for the years 2011, 2016, and 2021 provide valuable insights into the landscape changes and sensor-specific differences. Elevation was identified as a significant feature in 2011 and maintained its importance in subsequent years. In the Landsat 8 images from 2016 and 2021, indices such as the NDBaI and Greenness gained prominence, reflecting a greater focus on urban expansion and vegetation health. Traditional spectral bands, including Blue, Green, Red, Near-Infrared (NIR), Short-Wave Infrared 1 (SWIR1), and Short-Wave Infrared 2 (SWIR2), consistently remained essential for land cover analysis. Figure 3 illustrates the normalized importance values of the features used in each of the satellite images.

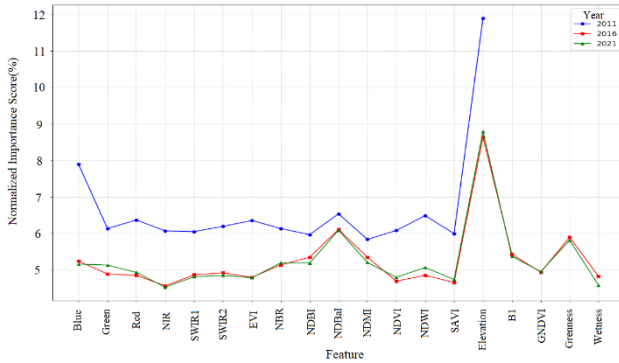


Figure 3. Feature Importance Scores across Years

Land cover maps for 2011 and 2016, depicted in Figure 4, highlight the Tehran area and its surrounding buffer zone segmented into the five classes.

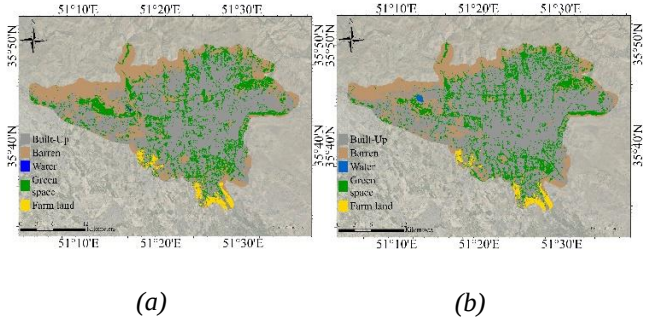


Figure 4. Land Covers Map of Tehran for (a) 2011 and (b) 2016

Over the five-year period, as presented in Table 5, built-up areas exhibited a considerable expansion, and the creation of an artificial lake in western Tehran contributed to an increase in water bodies. In contrast, green spaces and agricultural lands remained relatively unchanged, while barren lands experienced a significant decline.

Table 5. Gains and losses between 2011 and 2016.

Class	Losses(Hectares)	Gains(Hectares)
Built-up	3309	7170
Barren	5943	2034
Water	132	2
Green space	3936	3938
Farm land	243	323

4.2. Modelling and Predicting Land Cover Maps

4.2.1. Generation of Feature Sets for Machine Learning Training

Urban sprawl measurement indices were employed as key input features for training machine learning models and generating potential land cover transition maps. These indices were derived for 110 regions within the study area using census data, land use parcel maps, road networks, and baseline land cover maps. While some of the urban sprawl indices were initially computed for 110 administrative units, conducting the analysis strictly at this coarse resolution would result in mixing heterogeneous land cover patterns within each unit. To overcome this limitation, the indices were rasterized and assigned uniformly to 30-meter pixels within each unit, enabling pixel-wise modeling and supporting land cover unmixing. This approach allows the model to capture intra-unit spatial variations in land transitions, while maintaining the original structure of the input data. Figure 5 presents the spatial distribution of urban sprawl indices in 2011.

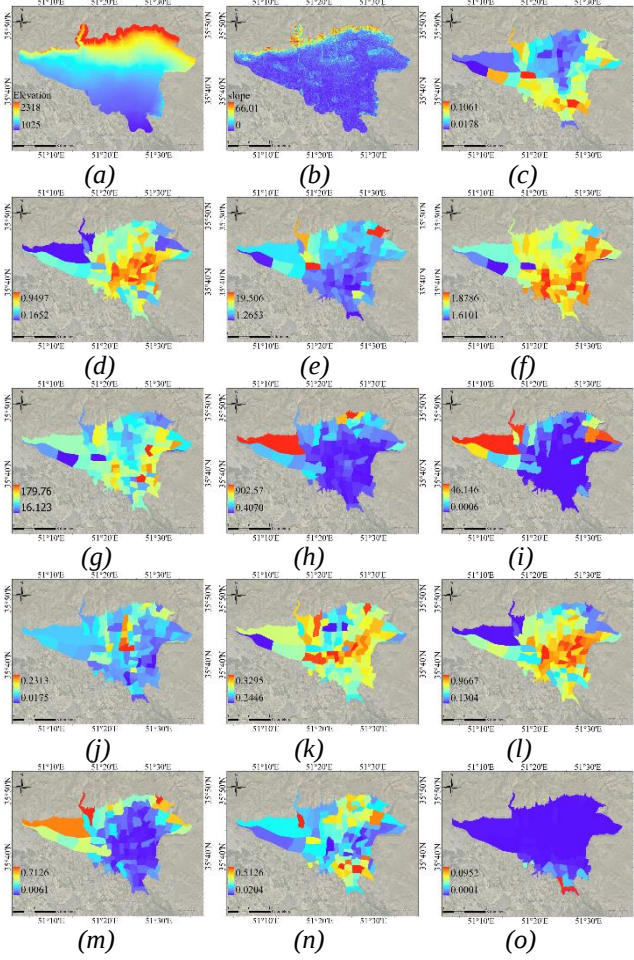


Figure 5. Urban sprawl measurement indices in 2011: (a) Elevation, (b) Slope, (c) Net Population Density, (d) Gross Population Density, (e) Vertical Density, (f) Fractal Dimension, (g) Shape Index, (h) Strip Development, (i) Discontinuous Development, (j) Migration Rate, (k) Employment Rate, (l) Built Land Cover, (m) Barren Land Cover, (n) Greenery Land Cover, (o) Farms Land Cover

Also, Figure 6 shows the set of urban sprawl measurement indices of the study area in 2016.

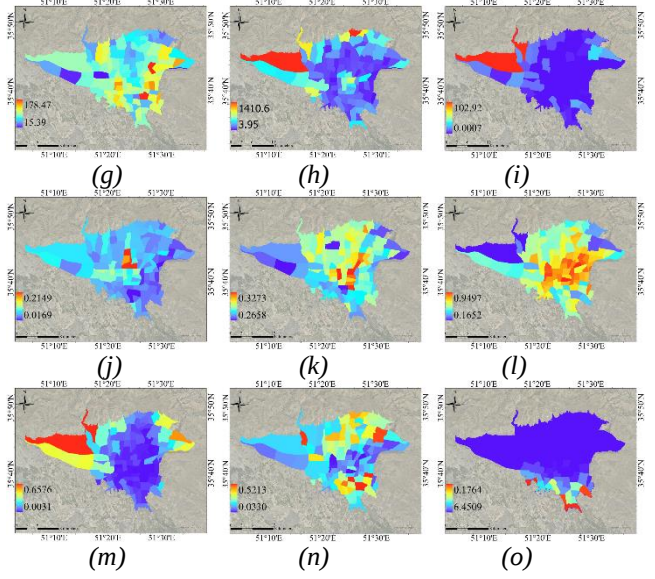
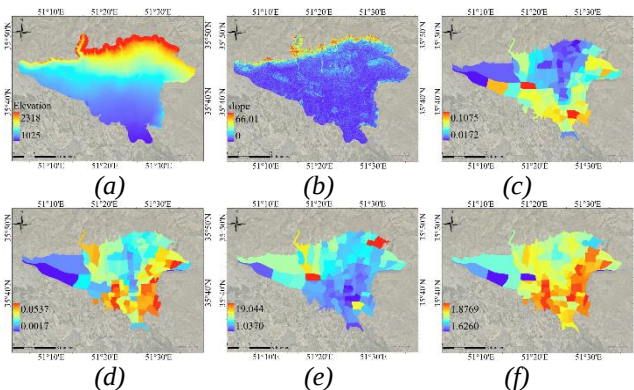


Figure 6. Urban sprawl measurement indices in 2016: (a) Elevation, (b) Slope, (c) Net Population Density, (d) Gross Population Density, (e) Vertical Density, (f) Fractal Dimension, (g) Shape Index, (h) Strip Development, (i) Discontinuous Development, (j) Migration Rate, (k) Employment Rate, (l) Built Land Cover, (m) Barren Land Cover, (n) Greenery Land Cover, (o) Farms Land Cover

The probabilistic maps depicting potential land cover class transitions, in conjunction with the computed urban sprawl indices, constitute the principal feature set employed for training the machine learning models. These maps are illustrated in Figure 7.

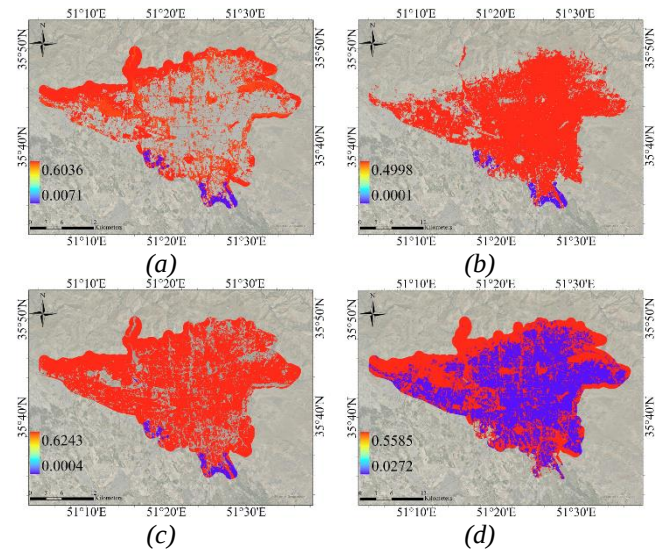


Figure 7. Probability map for the quantization of land cover class transitions into the categories (a) Built, (b) Barren, (c) Green space, (d) Farms using the ELT approach.

To construct a concise yet representative feature set, PCA and ICA were applied. The intrinsic dimensionality of the original feature space was estimated using the MLE method, which indicated that the underlying data structure could be effectively captured by three components, as illustrated in Figure 8. This estimation reflects the complexity and interdependence of the initial features while preserving the most informative variance. To ensure adequate local representation and reduce potential bias from selecting overly high or low values for the number of nearest neighbours, the neighbourhood size in the MLE procedure was restricted to a range of 6 to 10. This range was chosen to balance local sensitivity with the stability of the dimensionality estimation.

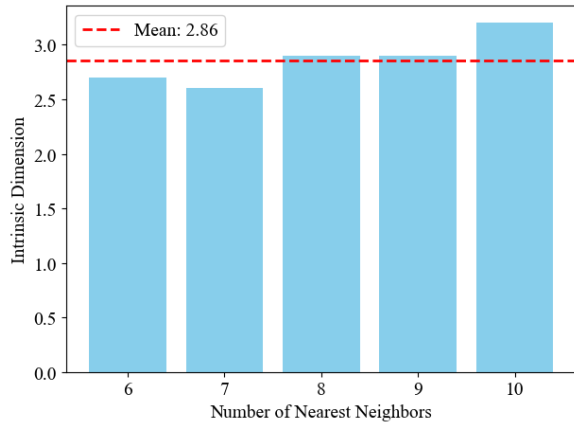


Figure 8. Intrinsic dimensionality estimation using the MLE method

Accordingly, two distinct feature sets were constructed for training the machine learning models based on the components extracted through dimensionality reduction techniques. The resulting components derived from PCA and ICA are presented in Figure 9.

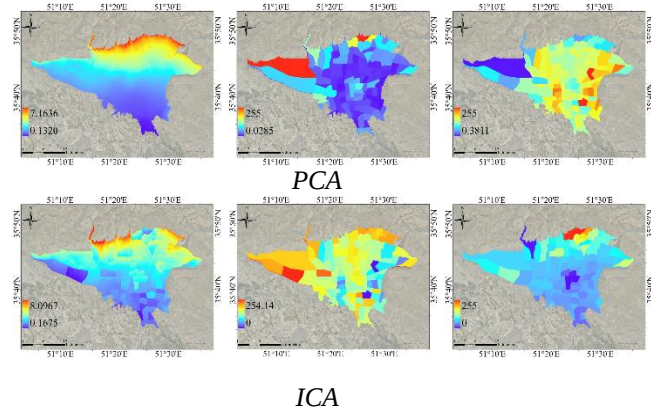


Figure 9. The feature set generated from PCA and ICA

4.2.2. Land Cover Transition Prediction via Markov Chain Analysis

The Markov chain model is utilized to quantify the extent of land cover transitions within the study area. By incorporating land cover maps from 2011 and 2016, the expected land cover changes for the year 2021 are computed, as presented in Table 6. This table outlines the transition probabilities of land cover classes over a five-year period, based on the Markov chain analysis.

Table 6. Probability of Land Cover Class Transitions by 2021.

From/To	Built-up	Barren	Water	Greenery	Farm land
Built-up	0.9453	0.0137	0	0.0409	0
Barren	0.1357	0.8134	0.0032	0.0434	0.0043
Water	0	0.0059	0.9563	0.0378	0
Greenery	0.1169	0.0443	0.0012	0.8332	0.0043
Farm land	0.0160	0.0694	0	0.0385	0.8761

The potential transition maps, derived from these transition probabilities, highlight the expected land cover changes over the five-year period. These maps are created by utilizing the Markov chain projections. In addition, various feature sets are generated for training machine learning models. Subsequently, the study investigates the creation of both hard and soft prediction maps using different machine learning methodologies.

- Land Cover mapping using the MLP model

The MLP algorithm was employed to generate potential land cover transition maps based on different feature sets. The model was trained over 10000 iterations, with the learning rate dynamically adjusted to enhance convergence and model performance. Table 7 summarizes the training and testing results, including the root mean square (RMS)

errors and classification accuracy for each feature configuration.

Table 7. Training the MLP Model with Various Feature Sets.

Feature set	Learning rate	Training RMS	Testing RMS	Accuracy
Initial	0.0001	0.2692	0.2772	60.92%
PCA-Derived	0.0005	0.3578	0.3583	31.09%
ICA-Derived	0.0003	0.3513	0.3520	39.60%

Among the tested configurations, the initial feature set yielded the highest accuracy and was therefore selected for generating the final potential transition maps. The soft and hard land cover maps for the year 2021 for the study area are generated based on the superior model MLP, as illustrated in Figure 10.

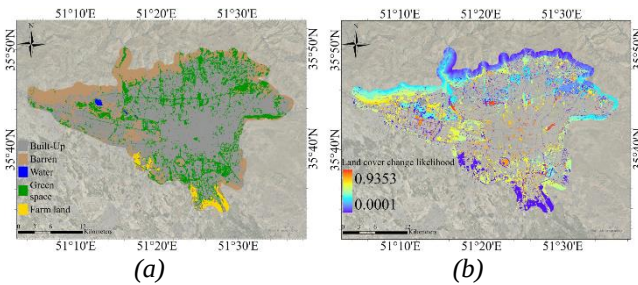


Figure 10. Land cover map (a) hard (b) soft for the year 2021 based on the superior MLP model.

• Land Cover mapping using the DF model

The DF algorithm was applied to various feature sets to model land cover transitions. The best performance was obtained using the initial dataset, with an OOB accuracy of 82.88% achieved using 150 decision trees and five variables per node. In comparison, the PCA- and ICA-derived feature sets yielded lower OOB accuracies of 62.71% and 63.32%, respectively, when configured with 250 trees and two variables per node. The soft and hard land cover maps for the study area in 2021 were generated using the advanced RF model, as shown in Figure 11.

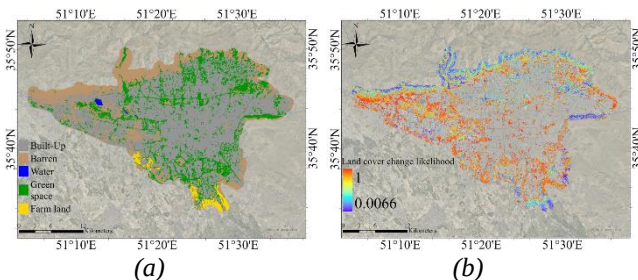


Figure 11. Land cover map (a) hard (b) soft for the year 2021 based on the superior DF model

• Land Cover mapping using the SVM model

The SVM model, employing a radial basis function (RBF) kernel, was initially trained on the complete feature set, resulting in an accuracy of 67.89%. Subsequently, the model was trained on reduced feature sets derived via ICA and PCA, each comprising three components. These reduced datasets yielded improved accuracies of 68.89% and 70.02%, respectively. Consequently, the SVM model trained on the ICA-Derived features is deemed superior. Soft and hard land cover maps for the study area were generated for the year 2021 using the optimized SVM model, as illustrated in Figure 12.

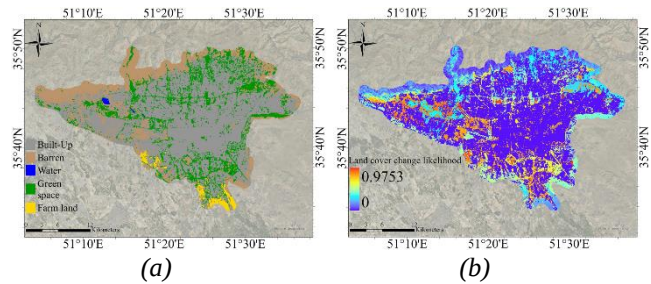


Figure 12. Land cover map (a) hard (b) soft for the year 2021 based on the superior SVM model

4.2.3. Validation of Results

To assess the model performance, the 2021 land cover maps produced by the best-performing models were compared with the reference map for the same year. Validation was conducted using metrics including MSE, Accuracy, Precision, F1-Score, and the Kappa coefficient. A summary of these validation results is provided in Table 8. Among the models, the MLP model trained with the initial feature set demonstrated the highest concordance with the reference map. Nonetheless, the outputs from the other models also yielded comparable results, reflecting the general reliability of the applied approaches.

Table 8. Validation of results.

Model	MSE	Accuracy	Precision	F1-Score	Spatial-k
MLP	0.9825	84.46%	84.04%	83.95%	76.09%
DF	0.9823	84.43%	84.01%	83.92%	76.05%
SVM	0.9900	84.40%	83.99%	83.89%	76.00%

4.2.4. Future Land Cover Projection

The MLP model trained on the initial feature set demonstrated the highest predictive accuracy against the 2021 reference map. Accordingly, this model was employed to project the land cover map for the study area in 2026. The resulting map is presented in Figure 13.

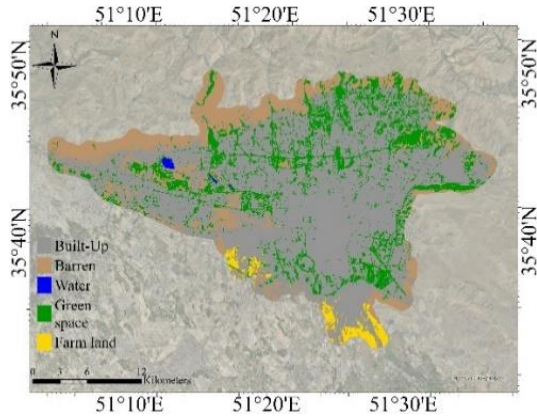


Figure 13. The hard prediction map for the year 2026

5. Discussion

The comparative analysis of the predicted and reference land cover maps for 2021 highlights the promising potential of machine learning-based models in forecasting urban expansion trends in Tehran. Among the various approaches tested, the MLP model, trained with the initial feature set, demonstrated superior performance in both accuracy and F1-score, thereby showcasing its effectiveness in capturing the intricate spatial dynamics of land transformation. Although the DF and SVM models also yielded commendable results, the slight performance advantage of the MLP model emphasizes the importance of leveraging comprehensive, unreduced feature sets in specific urban contexts.

The spatial patterns observed in the prediction maps indicate a continuing trend of urban expansion towards the western, north-western, and southern peripheries of Tehran. Notably, barren lands in these peripheral areas appear to be particularly vulnerable to conversion into built-up zones, which may be influenced by both market-driven development pressures and potential infrastructure-led planning initiatives. This pattern aligns with global trends, where urban growth often encroaches upon ecologically or agriculturally marginal lands, driven by lower economic costs and fewer regulatory constraints.

Projected land cover changes from 2016 to 2026 emphasize the critical need for sustainable land management practices. The anticipated 6% increase in built-up areas, coupled with a 10% reduction in barren land and notable declines in green spaces (8%) and farmlands (1%), reflects a shift towards a higher proportion of impervious surfaces. These transformations are expected to exacerbate several environmental challenges, including increased urban heat island intensity, loss of biodiversity,

decreased agricultural productivity, and elevated flood risks due to the reduced infiltration capacity of the soil.

From a methodological perspective, this study underscores the strategic importance of dimensionality reduction techniques, such as PCA and ICA, in enhancing both model performance and computational efficiency. While the highest predictive accuracy was achieved using the full feature set, the components derived through ICA exhibited superior performance in SVM modelling. This finding suggests that the effectiveness of dimensionality reduction methods may vary depending on the classification algorithm employed and the inherent complexity of the dataset. In large-scale applications, such techniques can significantly reduce computational time all while maintaining robust predictive capabilities.

The integration of satellite remote sensing, spatial indices of urban sprawl, and advanced machine learning techniques offers significant potential for the development of replicable, data-driven tools for urban planning. By combining sprawl indicators—such as density, shape metrics, accessibility indices, and socioeconomic factors—with land cover transition probabilities, this research enhances the understanding of urban dynamics at a more granular level. These tools are not only valuable for predictive purposes but also for the formulation of proactive policies. For example, transition probability maps can support zoning regulations, infrastructure planning, and environmental mitigation strategies by identifying high-risk areas prone to unplanned development.

Nevertheless, the study is subject to certain limitations, primarily stemming from data availability. The lack of detailed traffic and transportation data, which play a crucial role in shaping urban expansion, constrained the inclusion of mobility-related indicators in the analysis. Future research should aim to incorporate more comprehensive datasets, such as real-time traffic flows, public transport accessibility, land value trends, and climate variables, in order to enhance the accuracy of predictions and expand the policy relevance of the model's outputs.

In conclusion, this study emphasizes the pivotal role of integrating spatial intelligence and machine learning techniques in the monitoring and management of urban growth. The findings underscore both the risks associated with unregulated urban sprawl in Tehran and the potential for implementing informed and sustainable planning practices. As urbanization accelerates, particularly in developing countries, the adoption of such analytical frameworks will be crucial for achieving a balance between development, ecological preservation, and urban resilience.

6. Conclusion

This study adopted a comprehensive methodological framework that integrates remote sensing data, spatial indicators of urban sprawl, dimensionality reduction techniques, and machine learning algorithms to model and predict land cover changes in Tehran. By leveraging

Landsat satellite imagery, demographic datasets, and sprawl-related metrics, the research effectively captured the spatial-temporal dynamics of urban expansion over a ten-year period and projected potential land cover transformations through the year 2026.

The findings indicate that the Multi-Layer Perceptron (MLP) model trained with the initial, unreduced feature set yielded the highest predictive accuracy, underscoring the importance of incorporating a comprehensive range of demographic, topographic, and geometric variables in land cover modelling. Although dimensionality reduction techniques such as PCA and ICA enhanced computational efficiency, their impact on model performance varied depending on the algorithm, with notable improvements observed in the SVM model.

Spatial analysis indicates that urban expansion in Tehran is primarily concentrated in the western, north-western, and southern peripheral zones, frequently encroaching upon green spaces and barren lands. These patterns highlight escalating environmental pressures associated with urban sprawl, including landscape fragmentation, vegetation loss, and the contraction of agricultural areas.

The predictive capability of the proposed framework presents substantial value for urban planners, policymakers, and environmental managers by enabling evidence-based decision-making. It supports the identification of areas with a high likelihood of future development and contributes to the formulation of sustainable urban growth strategies. However, the study recognizes limitations related to data availability, particularly regarding transportation and infrastructure datasets. Addressing these gaps in future research may improve model accuracy and expand the framework's applicability across broader planning contexts.

Ultimately, the integration of geospatial technologies with machine learning offers a robust and scalable approach for analysing and directing urban development. As urban expansion accelerates, particularly in rapidly growing regions, the application of such advanced analytical tools will be essential for promoting a sustainable balance between developmental needs and environmental preservation.

References

- Awad, M., & Khanna, R. (2015). *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Springer nature.
- Banai, R., & DePriest, T. (2014). Urban sprawl: Definitions, data, methods of measurement, and environmental consequences. *Journal of Sustainability Education*, 7(2), 1-15.
- Baqa, M. F., Chen, F., Lu, L., Qureshi, S., Tariq, A., Wang, S., Jing, L., Hamza, S., & Li, Q. (2021). Monitoring and modeling the patterns and trends of urban growth using urban sprawl matrix and CA-Markov model: A case study of Karachi, Pakistan. *Land*, 10(7), 700. <https://doi.org/10.3390/land10070700>
- Barman, S., Roy, D., Chandra Sarkar, B., Almohamad, H., & Abdo, H. G. (2024). Assessment of urban growth in relation to urban sprawl using landscape metrics and Shannon's entropy model in Jalpaiguri urban agglomeration, West Bengal, India. *Geocarto International*, 39(1), 2306258. <https://doi.org/10.1080/10106049.2024.2306258>
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25, 197-227. <https://doi.org/10.1007/s11749-016-0481-7>
- Biney, E., & Boakye, E. (2021). Urban sprawl and its impact on land use land cover dynamics of Sekondi-Takoradi metropolitan assembly, Ghana. *Environmental Challenges*, 4, 100168. <https://doi.org/10.1016/j.envc.2021.100168>
- Blair, R., & Wellman, G. (2017). Smart growth principles and the management of urban sprawl. In *Regional Equity* (pp. 73-89). Routledge.
- Borenstein, M. (2001). *Power and precision* (Vol. 1). Taylor & Francis.
- Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3), 199-231. <https://doi.org/10.1214/ss/1009213726>
- Cao, L., Chua, K. S., Chong, W. K., Lee, H. P., & Gu, Q. (2003). A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing*, 55(1-2), 321-336. [https://doi.org/10.1016/S0925-2312\(03\)00433-8](https://doi.org/10.1016/S0925-2312(03)00433-8)
- Chan, J. C.-W., Chan, K.-P., & Yeh, A. G.-O. (2001). Detecting the nature of change in an urban environment: A comparison of machine learning algorithms. *Photogrammetric Engineering and Remote Sensing*, 67(2), 213-226.
- Club, S. (1998). *The dark side of the American dream*. retrieved March, 5, 2003.
- Dadashpoor, H., & Salarian, F. (2020). Urban sprawl on natural lands: Analyzing and predicting the trend of land use changes and sprawl in Mazandaran city region, Iran. *Environment, Development and Sustainability*, 22(2), 593-614. <https://doi.org/10.1007/s10668-018-0211-2>
- Das, B., Khan, F., & Mohammad, P. (2023). Impact of urban sprawl on change of environment and consequences. *Environmental Science and Pollution Research*, 30(49), 106894-106897. <https://doi.org/10.1007/s11356-023-29192-3>
- Dhanaraj, K., & Angadi, D. P. (2022). Land use land cover mapping and monitoring urban growth using remote sensing and GIS techniques in Mangaluru, India. *GeoJournal*, 87(2), 1133-1159. <https://doi.org/10.1007/s10708-020-10302-4>
- Dinda, S., Das, K., Chatterjee, N. D., & Ghosh, S. (2019).

- Integration of GIS and statistical approach in mapping of urban sprawl and predicting future growth in Midnapore town, India. *Modeling Earth Systems and Environment*, 5(1), 331-352. <https://doi.org/10.1007/s40808-018-0536-8>
- Duany, A., Plater-Zyberk, E., & Speck, J. (2000). *Suburban nation: The rise of sprawl and the decline of the American dream*. Macmillan.
- Escuin, S., Navarro, R., & Fernández, P. (2008). Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images. *International Journal of Remote Sensing*, 29(4), 1053-1073. <https://doi.org/10.1080/01431160701281072>
- Estoque, R. C., & Murayama, Y. (2015). 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 indicators*, 56, 205-217. <https://doi.org/10.1016/j.ecolind.2015.03.037>
- Ewing, R. (1997). Is Los Angeles-style sprawl desirable? *Journal of the American planning association*, 63(1), 107-126. <https://doi.org/10.1080/01944369708975728>
- Frenkel, A., & Ashkenazi, M. (2008). Measuring urban sprawl: how can we deal with it? *Environment and Planning B: Planning and design*, 35(1), 56-79. <https://doi.org/10.1068/b32155>
- Gómez, J. A., Patiño, J. E., Duque, J. C., & Passos, S. (2019). Spatiotemporal modeling of urban growth using machine learning. *Remote Sensing*, 12(1), 109. <https://doi.org/10.3390/rs12010109>
- Hatab, A. A., Cavinato, M. E. R., Lindemer, A., & Lagerkvist, C.-J. (2019). Urban sprawl, food security and agricultural systems in developing countries: A systematic review of the literature. *Cities*, 94, 129-142. <https://doi.org/10.1016/j.cities.2019.06.001>
- Herbei, M., & Sala, F. (2016). Classification of land and crops based on satellite images Landsat 8: case study SD Timisoara.
- Herold, M., Goldstein, N. C., & Clarke, K. C. (2003). The spatiotemporal form of urban growth: measurement, analysis and modeling. *Remote Sensing of Environment*, 86(3), 286-302. [https://doi.org/10.1016/S0034-4257\(03\)00075-0](https://doi.org/10.1016/S0034-4257(03)00075-0)
- Hislop, S., Jones, S., Soto-Berelov, M., Skidmore, A., Haywood, A., & Nguyen, T. H. (2018). Using landsat spectral indices in time-series to assess wildfire disturbance and recovery. *Remote Sensing*, 10(3), 460. <https://doi.org/10.3390/rs10030460>
- Jiang, F., Liu, S., Yuan, H., & Zhang, Q. (2007). Measuring urban sprawl in Beijing with geo-spatial indices. *Journal of Geographical Sciences*, 17(4), 469-478. <https://doi.org/10.1007/s11442-007-0469-z>
- Karbauskait'e, R., & Dzemyda, G. (2013). Investigation of the maximum likelihood estimator of intrinsic dimensionality.
- Kshetri, T. (2018). Ndvi, ndbi & ndwi calculation using landsat 7, 8. *GeoWorld*, 2, 32-34.
- Kulkarni, K., & Vijaya, P. (2022). Measuring urban sprawl using machine learning. *Fundamentals and methods of machine and deep learning: algorithms, tools and applications*, 327-340. <https://doi.org/10.1002/9781119821908.ch14>
- Kumar, A. (2017). Analysing urban sprawl and land consumption patterns in major capital cities in the Himalayan region using geoinformatics. *Applied geography*, 89, 112-123. <https://doi.org/10.1016/j.apgeog.2017.10.010>
- Liping, C., Yujun, S., & Saeed, S. (2018). Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques—A case study of a hilly area, Jiangle, China. *PloS one*, 13(7), e0200493. <https://doi.org/10.1371/journal.pone.0200493>
- Ma, Y., & Zhu, L. (2013). A review on dimension reduction. *International Statistical Review*, 81(1), 134-150. <https://doi.org/10.1111/j.1751-5823.2012.00182.x>
- Marmolin, H. (1986). Subjective MSE measures. *IEEE transactions on systems, man, and cybernetics*, 16(3), 486-489. <https://doi.org/10.1109/TSMC.1986.4308985>
- Mather, P., & Tso, B. (2016). *Classification methods for remotely sensed data*. CRC press. <https://doi.org/10.1201/9781420090741>
- Moniruzzam, M., Roy, A., Bhatt, C., Gupta, A., An, N., & Hassan, M. (2018). Impact analysis of urbanization on land use land cover change for Khulna City, Bangladesh using temporal landsat imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 757-760. <https://doi.org/10.5194/isprs-archives-XLII-5-757-2018>
- Nabi, F., & Zhou, X. (2024). Enhancing Intrusion Detection Systems Through Dimensionality Reduction: A Comparative Study of Machine Learning Techniques for Cyber Security. *Cyber Security and Applications*, 100033. <https://doi.org/10.1016/j.csa.2023.100033>
- Oon, A., Mohd Shafri, H. Z., Lechner, A. M., & Azhar, B. (2019). Discriminating between large-scale oil palm plantations and smallholdings on tropical peatlands using vegetation indices and supervised classification of LANDSAT-8. *International Journal of Remote Sensing*, 40(19), 7312-7328. <https://doi.org/10.1080/01431161.2019.1579944>

- Pokorny, V. J., Sponheim, S. R., & Rawls, E. (2023). Impact of reduced- dimensionality independent components analysis on event- related potential measurements. *Psychophysiology*, 60(5), e14223. <https://doi.org/10.1111/psyp.14223>
- Rana, M. S., & Sarkar, S. (2021). Prediction of urban expansion by using land cover change detection approach. *Heliyon*, 7(11). <https://doi.org/10.1016/j.heliyon.2021.e08437>
- Rimal, B., Zhang, L., Keshtkar, H., Haack, B. N., Rijal, S., & Zhang, P. (2018). Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and markov chain. *ISPRS International Journal of Geo-Information*, 7(4), 154. <https://doi.org/10.3390/ijgi7040154>
- Royall, R. (2017). Statistical evidence: a likelihood paradigm. Routledge. <https://doi.org/10.1201/9780203738665>
- Shaver, T., Khosla, R., & Westfall, D. (2006). Utilizing green normalized difference vegetation indices (GNDVI) for production level management zone delineation in irrigated corn. *The 18th World Congress of Soil Science*,
- Shi, Y., Zhou, L., Guo, X., & Li, J. (2023). The multidimensional measurement method of urban sprawl and its empirical analysis in Shanghai metropolitan area. *Sustainability*, 15(2), 1020. <https://doi.org/10.3390/su15021020>
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. *Australasian joint conference on artificial intelligence*, https://doi.org/10.1007/11941439_114
- Tatsumi, K., Yamashiki, Y., Torres, M. A. C., & Taïpe, C. L. R. (2015). Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture*, 115, 171-179. <https://doi.org/10.1016/j.compag.2015.05.001>
- Taufik, A., Ahmad, S. S. S., & Ahmad, A. (2016). Classification of landsat 8 satellite data using NDVI thresholds. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 8(4), 37-40.
- Tolver, A. (2016). An introduction to Markov chains. Department of Mathematical Sciences, University of Copenhagen.
- Trinh, L. H. (2020). Urban Bare Land Classification Using NDBaI Index Based on Combination of Sentinel 2 MSI and Landsat 8 Multiresolution Images. *VNU Journal of Science: Earth and Environmental Sciences*, 36(2). <https://doi.org/10.25073/2588-1094/vnuees.4537>
- Yasin, M. Y., Yusoff, M. M., Abdullah, J., Noor, N. M., & Noor, N. M. (2021). Urban sprawl literature review: Definition and driving force. *Geografia*, 17(2). <https://doi.org/10.17576/geo-2021-1702-10>
- Yin, J., Yin, Z., Zhong, H., Xu, S., Hu, X., Wang, J., & Wu, J. (2011). Monitoring urban expansion and land use/land cover changes of Shanghai metropolitan area during the transitional economy (1979–2009) in China. *Environmental monitoring and assessment*, 177, 609-621. <https://doi.org/10.1007/s10661-010-1660-8>
- Zhang, K., & Chan, L.-W. (2005). Dimension reduction as a deflation method in ICA. *IEEE Signal Processing Letters*, 13(1), 45-48. <https://doi.org/10.1109/LSP.2005.860541>
- Zhang, M., Li, Y., Guo, R., & Yan, Y. (2022). Heterogeneous effects of urban sprawl on economic development: empirical evidence from China. *Sustainability*, 14(3), 1582. <https://doi.org/10.3390/su14031582>