Machine Learning Based Decadal Land Use Land Cover Analysis of Karnataka

Pamarthi Chennarao[1*], Pandala Madhavi Latha[2], Kesari Bhavani Prasad Reddy[3] and Veeramallu Satya Sahithi[4]

[1,2,3,4] Department of IT, VR Siddhartha Engineering College, Vijayawada, India
[1*]chennapamarthi93@gmail.com
[2]madhavi.p@vrsiddhartha.ac.in
[3]prasadreddykesari2004@gmail.com
[4]sahithi.geo@gmail.com

Abstract. The use of machine learning algorithms to analyze land use and land cover (LULC) change has become more important for environmental monitoring and urban planning. This paper looks at how well ensemble machine learning models can map and quantify land transformation patterns across Karnataka State over four decades, from 1994 to 2024. The study uses Support Vector Machine (SVM) and Random Forest (RF) algorithms along with Geographic Information System (GIS) techniques to improve the accuracy of spatial quantification and classification. Multi-temporal satellite images were analyzed to observe changes between urban and non-urban areas, as well as shifts in vegetation, forests, and agricultural land. The results show a significant rise in built-up areas from 21.76 sq.km in 1994 to 48.25 sq.km in 2024, along with a slow decline in vegetation and forest cover. The ensemble model proved to be more accurate than individual classifiers, offering a better representation of land cover distribution. These findings showes how effective it is to combine machine learning with geospatial tools for sustainable land management and long-term environmental assessment.

Keywords: LULC, Ensemble Machine Learning, Support Vector Machine, Satellite Imagery, GIS.

1 Introduction

Land Cover and Land Use (LULC) analysis is significant in terms of understanding spatial changes, ecological equilibrium, and the effects of urbanization on natural resources. During the past three decades, Karnataka State [7] has undergone drastic changes in land cover attributed to fast urban expansion, agricultural development, and forest reduction. Such changes, if not observed, may cause environmental degradation, loss of biodiversity, and irrational use of land. Hence, precise detection and evaluation of these variations are crucial in successful regional planning and sustainable growth. Traditional land monitoring techniques—like field surveys and manual interpretation—are usually time-consuming and of limited extent. Large-scale analysis of land cover

B. Singh et al. (eds.), *Proceedings of the International Conference on Advances in Computing Technology and Artificial Intelligence (COMPUTATIA 2026)*, Atlantis Highlights in Intelligent Systems 18,
https://doi.org/10.2991/978-94-6239-713-2_24

has become easier with the introduction of satellite-based remote sensing and Geographic Information Systems (GIS). The complexity and heterogeneity in satellite images, however, need sophisticated computational methods for accurate classification and change detection. Machine learning models have proven to be efficient tools in this field, being able to extract subtle patterns in multi-temporal data. Support Vector Machine (SVM) and Random Forest (RF) classifiers have been especially dominant in classifying various land categories with robust performance [10].

In this research, the two models are combined in an ensemble model to improve classification accuracy and spatial conformity over Karnataka State. The comparison of LULC patterns during 1994-2024 indicated a tremendous shift in land cover composition. Urban-built-up areas rose from 21.76 sq.km to 48.25 sq.km during the period, testifying to the fast-paced urbanization of the state. On the other hand, vegetation cover reduced from 52.14 sq.km to 34.09 sq.km, and forest areas saw marginal decrease, reflecting pressure from spreading settlements and agricultural lands. The above findings point towards the increasing rate of urban expansion and the concomitant reduction in green cover. Through the utilization of an ensemble of SVM and RF classifiers, this study presents a solid and scalable methodology for multi-class LULC mapping with satellite images. The outcomes not only advance knowledge of spatiotemporal land transformation but also aid in evidence-based policy-making for sustainable land and resource management.

1.1 Study area collection:

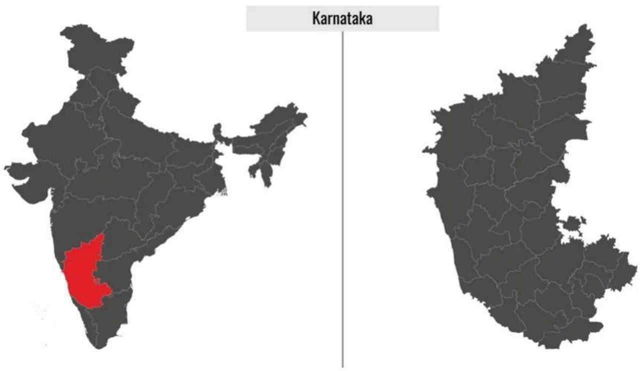

Fig. 1. Study Area

The area considered for the study is Karnataka State, which is situated in the southern part of India and falls between latitudes approximately 11°30′N to 18°30′N and longitudes 74°00′E to 78°30′E. It covers an area of about 191,791 sq.km, presenting a wide range of landforms including coastal plains, forested regions of the Western Ghats, agricultural landscapes, and fast-expanding urban centers. The state has been subjected to significant land use and land cover changes in recent decades due to rapid urbanization, industrial growth, and expansion of agriculture. Therefore, all these characteristics together make Karnataka a suitable region for large-scale LULC monitoring and change detection. Further, multi-temporal satellite imagery from the years 1994, 2004, 2014, and 2024 has been used to analyze the spatial and temporal variations in the state's land cover as shown in Fig.1.

2 Literature Study

Several machine learning and remote sensing developments have enhanced land cover classification. [1] illustrated that Random Forest is effective for high-resolution land-cover mapping, while [2] underscored the crucial role of proper data selection in enhancing classification accuracy. [3] also mentioned how rapid urbanization has shaped land transformation [4] illustrated how GIS and remote sensing could be used to monitor urban growth. The foundational work [5] established RF as a strong classifier, and this was further reaffirmed through reviews such [6], who provided extensive information about the strengths of SVM within remote sensing. Further work by [7] corroborated RF's high performance in vegetation classification, while [8] further strengthened the classifier's reliability for multispectral land mapping. [9] showed the utility of machine learning algorithms across a range of LULC environments. Object-based approaches reviewed [10] contributed to explaining some of the limitations in classifications [11] improved the accuracy through post-classification refinement [12] offered a categorical analysis of a classification approach and its applications. General remote sensing principles were laid out [13] and [14] presented some of the key techniques when it comes to multi-temporal change detection.

3 Methodology:

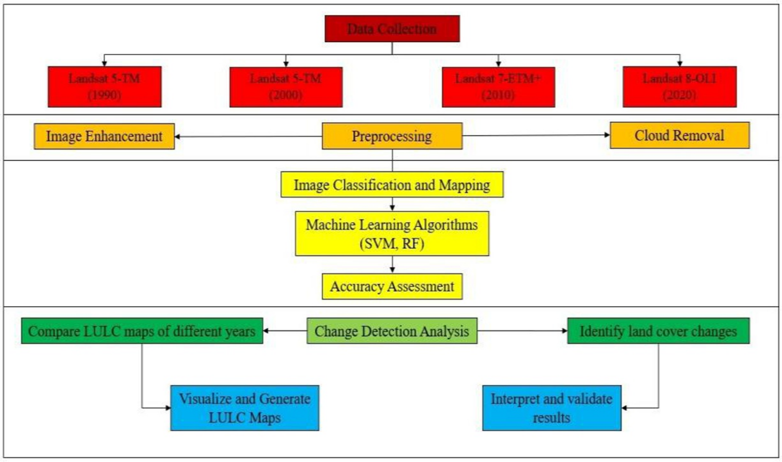

Fig. 2. Proposed Architecture

Four-time period multi-temporal data has been gathered in this study. All three Landsat 5 TM, Landsat 7 ETM+, and Landsat 8 OLI images had very consistent multispectral information. All the images were pre-processed to enhance image quality, atmospheric correction, and cloud removal in order to improve data quality and ensure temporal consistency. The pre-processed images were subjected to supervised machine learning algorithms, namely SVM and RF, in order to classify them into their respective LULC features and to generate a thematic map for each study year. Later, the accuracy of the classification has been quantitatively assessed by standard measures of accuracy assessment. Temporal change detection was done based on the comparison of LULC mapping of different years to identify the spatial and temporal changes in land cover. After visual interpretation, the results were interpreted and validated on the existence of urban growth and environmental metamorphosis over the entire study area as shwn in Fig.2.

3.1 Data Collection

This research makes use of multi-temporal satellite imagery for the analysis of Land Use and Land Cover (LULC) change over Karnataka State for the years 1994, 2004, 2014, and 2024. Satellite imagery for the analysis has been obtained from the USGS and Google Earth Engine platforms, where Landsat 5 TM for 1994 and 2004, Landsat 7 ETM+ for 2014, and Landsat 8 OLI/TIRS for 2024 were used. AOTS samples were collected from NRSC datasets and Google Earth high-resolution imagery for ground truthing, and the dataset has been classified into eight LULC types: built-up, vegetation, forest, agricultural, fallow, bare land, water bodies, and coastal land. Broad pre-processing techniques like cloud masking, atmospheric correction, and geometric correction were carried out.

Karnataka State, located in South India, with a geographical area of around 191,791 sq.km, shows a varied topography, including coastal, forest, agricultural, and rapidly increasing urban areas. The last thirty years have shown substantial changes in land cover due to population and industrial growth, including rising built-up areas and a decrease in vegetation and forests. The impact of these modifications will be analyzed using multi-date Landsat satellite images.

3.2 Classification of SVM and RF:

Support Vector Machine (SVM): SVM is indeed a powerful, widely used supervised learning algorithm for LULC classification because of its capability to handle high-dimensional and nonlinear data. In this study, SVM will be used to classify different land cover types based on the spectral characteristics of satellite imagery. The algorithm works by identifying an optimal hyperplane that separates various classes with the maximum possible margin. The points which lie closest to the hyperplane are called support vectors, which define the decision boundary. When data are not linearly separable, kernel functions such as linear, polynomial, or RBF are applied to project the data into a higher-dimensional space where separation becomes possible.

In the n-dimensional feature space, the separating hyperplane can be mathematically defined as:
where:

- w is the weight vector that is perpendicular to the hyperplane.
- x represents the feature vector, or input data, and
- b is the bias term, which shifts the hyperplane from the origin. SVM was used in this study as a supervised classifier, utilizing labeled training samples to discriminate among multiple LULC categories. Its ability to create clear and well-defined decision boundaries makes it suitable for complex satellite imagery classification.

Random Forest: Random Forest is an ensemble-based machine learning algorithm that works by constructing numerous decision trees in the course of training and then combines outputs for final classification. Each tree is grown on a random subset of training data and features, hence assuring diversity within the model itself and reducing overfitting. In the context of LULC classification, RF skillfully manages large datasets with mixed spectral information along with noise, making it highly robust for applications in remote sensing.

The basic concept of RF is to create N decision trees. Each tree will give one classification result, and the final class will be determined by the majority vote from all the trees. The quality of each tree's split is measured using metrics such as Gini impurity or entropy, defined as:

where p_i is the probability of samples belonging to class i.

Being an ensemble technique, RF improves generalization and accuracy, especially in the discrimination of spectrally similar classes such as vegetation, forest, and agricultural regions. High tolerance to noise and handling complex and nonlinear relationships make the model quite reliable for large-scale LULC mapping and monitoring applications.

4 Results:

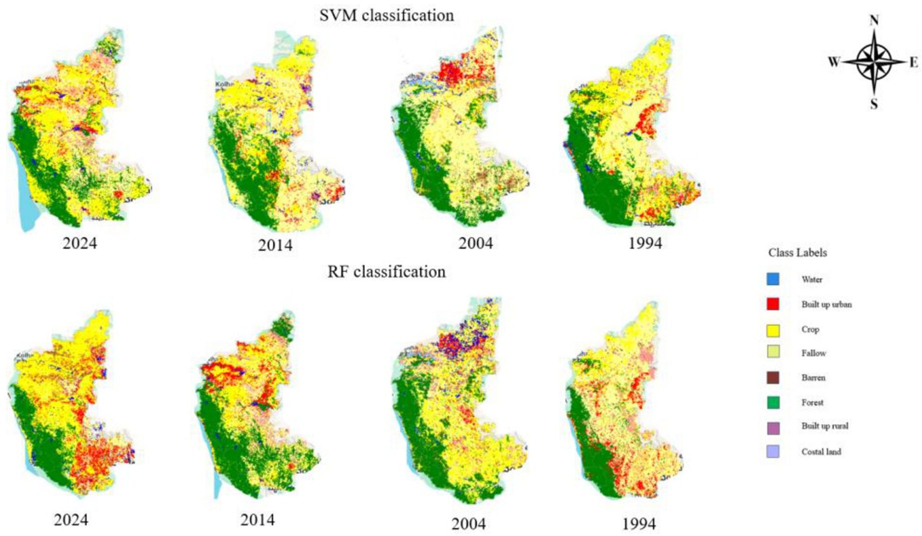

Fig. 3. SVM, RF classification

Figure 3 depicts the classification maps for Land Use and Land Cover (LULC) in the state of Karnataka for the years 1994, 2004, 2014, and 2024, produced using the classification results provided by the Support Vector Machine (SVM) and Random Forest (RF) classifiers. The outcome makes it clear that the level of Urban and Rural areas has gradually increased over the years, with the subsequent reduction in the extent of forest, cropland, and fallow land. The LULC mapping for the year 1994 shows that natural LULC mostly occupies the state, with less built-up land and scattered patterns, while the extent of the forest appears most prominent in the shape of the Western Ghats. By 2004-2014, the result provided by the classification carried out by SVM and RF classifiers shows the extent of Urban areas growing gradually with the emergence of Urban areas radiating from the nearby Urban areas and transportation routes. The mapping for the year 2024 shows the cluster representation for the built-up areas, achieving the demonstration of Urbanization in the Karnataka. By comparison, the features of the Urban areas with stiff boundaries are more prominent in the result provided by the SVM classifier, while the RF classifier results have smooth boundaries with more consideration to the Vegetation, Forest, and Agricultural areas. Table.1 shown below.

Table 1. Year-wise LULC Area Statistics Using SVM (sq. km)

LULC Class	1994	2004	2014	2024
Built-up	21.76	30.24	37.82	48.25
Vegetation	52.14	48.06	41.13	34.09
Forest	26.50	25.40	24.10	22.88
Agricultural	38.20	36.50	34.80	32.10
Fallow land	18.30	17.60	16.20	15.00
Barren land	12.10	13.20	14.50	16.30
Water bodies	14.80	13.90	12.70	11.40
Coastal land	7.90	7.90	7.90	7.90

From the statistics regarding the area change in different classes of LULC for each year, it can be observed that there have been notable transitions in the land use/cover classes between the years 1994 and 2024. While built-up land use/cover classes have been steadily mounting from 21.76 sq.km in 1994 to 48.25 sq.km in 2024, vegetation, forestland, agricultural land, fallow land, and water bodies have been steadily decreasing, which reveals that natural and agricultural lands are being transformed into built-up/urban lands and barren lands. On the other hand, barren lands have been steadily mounting, which indicates that the lands are degrading due to development-related activities. Coastal land use/cover classes have been the same in every year, which reveals that these lands remain at the same geographic location. Table.2 is shown below.

Table. 2. Year-wise LULC Area Statistics Using RF (sq. km)

LULC Class	1994	2004	2014	2024
Built-up	19.50	27.10	34.00	45.20
Vegetation	54.60	50.30	44.80	38.90
Forest	28.40	27.10	25.80	24.60
Agricultural	39.80	38.10	36.20	34.50
Fallow land	17.20	16.50	15.40	14.30
Barren land	10.90	11.80	12.90	14.10
Water bodies	15.90	15.00	13.80	12.60
Coastal land	7.90	7.90	7.90	7.90

Analysis of the LULC area statistics between the years 1994 and 2024 shows that built-up areas have been steadily increasing. This implies that the rates of urbanization have been steadily increasing. On the other hand, vegetation, forestland, agricultural land, fallow land, and water bodies have been steadily declining. This implies that the rates at which natural and agricultural land is being used for urbanization and barren land have been steadily increasing. Barren land has been steadily increasing as well. The coastal land has remained constant. All these imply a steady transformation of the environment from a natural system to a more urbanized environment. Table 3 is shown below.

Table 3. Accuracy assessment

Year	Overall Accuracy (%)		Kappa Coefficient (%)	
	SVM	**RF**	**SVM**	**RF**
1994	89.15	90.48	88.78	89.81
2004	94.62	95.92	93.81	94.84
2014	93.34	92.36	91.83	91.86
2024	94.96	96.12	92.85	94.88

5 Conclusion

Over the 1994-2024 period, there was a great rise in Karnataka's built-up land, from 21.76 sq.km to 48.25 sq.km, indicating continuous urban expansion. Vegetation reduced from 52.14 sq.km to 34.09 sq.km; with increasing development, forest cover also reduced. This is further reiterated by both the classifiers, SVM and RF, with a sharper rise in urban growth for SVM and more maintenance of natural cover for RF. In general, the state transformed from being predominantly green to more urbanized regions. These changes reflect long-term environmental transformation driven by increasing urbanization.

References

1. Hayes, M.M., Miller, S.N., Murphy, M.A.: High-resolution landcover classification using random forest. Remote Sens. Lett. 5, 112–121 (2014)
2. Knight, J.F., Tolcser, B.P., Corcoran, J.M., Rampi, L.P.: The effects of data selection and thematic detail on the accuracy of high spatial resolution wetland classifications. Photogramm. Eng. Remote Sens. 79, 613–623 (2013)
3. Cohen, B.: Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. Technol. Soc. 28(1–2), 63–80 (2006)

4. Hegazy, I.R., Kaloop, M.R.: Monitoring urban growth and land use change detection with GIS and remote sensing techniques in Daqahlia governorate Egypt. Int. J. Sustain. Built Environ. 4(1), 117–124 (2015)
5. Breiman, L.: Random forests. Mach. Learn. 45(1), 5–32 (2001)
6. Mountrakis, G., Im, J., Ogole, C.: Support vector machines in remote sensing: A review. ISPRS J. Photogramm. Remote Sens. 66(3), 247–259 (2011)
7. Hayes, M.M., Miller, S.N., Murphy, M.A.: High-resolution landcover classification using random forest. Remote Sens. Lett. 5(2), 112–121 (2014)
8. Pal, M.: Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26(1), 217–222 (2005)
9. Phalke, H.D., Özdemir, M.S.: Land use/land cover classification with machine learning algorithms—A case study of Konya, Turkey. Remote Sens. 12(3), 493 (2020)
10. Liu, D., Xia, F.: Assessing object-based classification: Advantages and limitations. Remote Sens. Lett. 1(4), 187–194 (2010)
11. Manandhar, R., Odeh, I.O.A., Ancev, T.: Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement. Remote Sens. 1(3), 330–344 (2009)
12. Tewkesbury, A.P., Comber, A.J., Tate, N.J., Lamb, A., Fisher, P.F.: A critical synthesis of remotely sensed land cover classification techniques: Towards a hierarchical framework. Remote Sens. Environ. 156, 426–440 (2015)
13. Giri, C.P.: Remote Sensing of Land Use and Land Cover: Principles and Applications. CRC Press (2012)
14. Singh, A.: Digital change detection techniques using remotely sensed data. Int. J. Remote Sens. 10(6), 989–1003 (1989)