International Journal on Emerging Research Areas (IJERA) **ISSN:2230-9993**

A Review Based on Satellite-Based Land Cover Classification System

Hitha P S
Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirappally, India
pshitha@amaljyothi.ac.in

Ezra Tom George
Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirappally, India
ezratomgeorge2026@cs.ajce.in

Fathima N
Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirappally, India
fathiman2026@cs.ajce.in

Izabel Joseph
Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirappally, India
izabeljoseph2026@cs.ajce.in

Karun Jidhish
Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirappally, India
karunjidhish2026@cs.ajce.in

Kausalya Sumesh
Computer Science and Engineering
Amal Jyothi College of Engineering
Kanjirappally, India
kausalyasumesh2026@cs.ajce.in

***Abstract*—Accurate land-cover information is essential for understanding environmental change, supporting sustainable development, and assisting planning in rapidly evolving urban and agricultural landscapes. Although satellite imagery is widely accessible, transforming raw multispectral data into reliable large-scale land-cover maps remains a non-trivial task due to data complexity and the expertise typically required for interpretation. This study presents an automated land use and land cover (LULC) classification framework that generates pixel-level thematic maps directly from Sentinel-2 Level-2A GeoTIFF imagery. The proposed system integrates selected spectral bands with reference annotations derived from the ESA WorldCover dataset and enhances the feature representation through domain-driven spectral indices such as NDVI, NDBI, and MNDWI. Training samples are obtained from geographically diverse regions and balanced across categories to improve robustness and generalization. Several supervised machine learning algorithms—including Support Vector Machine, Random Forest, and XGBoost—are evaluated, with LightGBM selected as the final classifier due to its computational efficiency and scalable leaf-wise boosting mechanism. Experimental results demonstrate that the framework delivers spatially consistent and accurate land-cover maps while maintaining lower computational complexity compared to deep learning-based alternatives. The modular design further enables seamless extension toward time-series land analysis and automated environmental monitoring workflows.**

***Index Terms*—Land Cover Classification, Sentinel-2, LightGBM, Remote Sensing, GeoTIFF.**

I. Introduction

Accurate understanding of the changing patterns of the Earth's surface is very important for environmental sustainability and urban planning. In recent decades, rapid urbanization, deforestation, and environmental changes due to climate change have altered the natural landscape of the Earth's surface to a great extent. Hence, the accurate understanding of land use and land cover patterns has become an important area of research in the field of remote sensing.

Satellite remote sensing has been recognized as an effective tool to monitor environmental changes on the Earth's surface. The recent satellites launched by the European Space Agency, like Sentinel-2, can provide high-resolution multispectral imagery to monitor environmental changes on the Earth's surface. However, manual analysis of satellite imagery is very difficult because of the huge amount of data and the complexity of information contained within the imagery.

To address this challenge, automated classification techniques using machine learning and deep learning approaches have gained popularity in remote sensing research. Machine learning techniques can effectively recognize complex relationships between spectral features and land cover classes, enabling the transformation of images obtained from satellites into thematic maps showing different classes of land cover. Among the different techniques proposed in the literature, Random Forest, Support Vector Machine, XGBoost, and LightGBM techniques have been thoroughly investigated in terms of their reliability, efficiency, and accuracy.

Recent advances in the field have proposed different deep learning techniques using convolutional neural networks and U-Net models to enhance spatial feature extraction and segmentation at the pixel level using images obtained from satellites. Even with these advances, choosing appropriate classification techniques and understanding their relative strengths is considered a challenge since their performance is often dependent on datasets, feature engineering techniques, and geographical conditions.

The aim and objective of this review paper is to present a comprehensive study on the application of different machine learning and deep learning techniques in the classification of land cover using satellites. This paper presents a summary of existing research works and their analysis with respect to the models used in the classification process. In addition, different advances in time series analysis and change detection techniques are also discussed with respect to the application of the classification framework.

II. Satellite Data for Land Cover Classification

Satellite imagery acts as the basis for modern land cover classification systems. Among the popular satellite imagery used for land cover classification, Sentinel-2 imagery has attracted considerable attention because of its high spatial resolution and rich spectral information. The Sentinel-2 mission offers multispectral imagery that carries thirteen spectral bands ranging from visible to near-infrared and shortwave infrared parts of the spectrum [2]. These spectral bands can be used to identify vegetation, water bodies, soil, and infrastructure.

Apart from Sentinel-2 imagery, other satellites like Landsat and PlanetScope play important roles in land cover classification studies. Landsat imagery has a long history of providing Earth observation data and is considered useful for long-term land cover classification studies [1]. PlanetScope imagery offers very high spatial resolution imagery that can be used to classify vegetation and agriculture.

Before any classification, a series of preprocessing steps are required to be carried out on the satellite images. Some of these preprocessing steps include atmospheric correction, cloud detection, geometric correction, and normalization [4], [12]. All these preprocessing steps are carried out to ensure consistency in data over time and space.

III. Machine Learning Models for Land Cover Classification

Machine learning algorithms are now considered essential tools for satellite image classification due to their potential for processing large-scale multispectral images and recognizing the intricate relationships between spectral features and land cover classes. A variety of supervised learning algorithms has been extensively investigated in the literature for pixel-level land cover classification. Among the algorithms, Support Vector Machine (SVM), Random Forest (RF), and gradient boosting algorithms such as XGBoost and LightGBM are commonly reported as successful tools for the analysis of multispectral images for remote sensing purposes [7], [15].

In general, the majority of the satellite-based land cover classification research works follow a similar framework, where the pre-processing of satellite images, feature extraction, training, and pixel-level classification are the essential components. Both machine learning and deep learning algorithms are commonly adopted for the purpose of reliable land cover mapping. Figure 1 shows the general workflow that has been observed across several research works, which explains how different models contribute to the overall classification process. If we conduct a comparative analysis of these models based on several research works and experimental studies, we can see that each algorithm has its own advantages when it comes to classification accuracy, computational efficiency, scalability, and robustness to noise. Insights obtained from the literature and comparative studies reveal the benefits and drawbacks of these models when used with high-resolution satellite imagery like Sentinel-2.

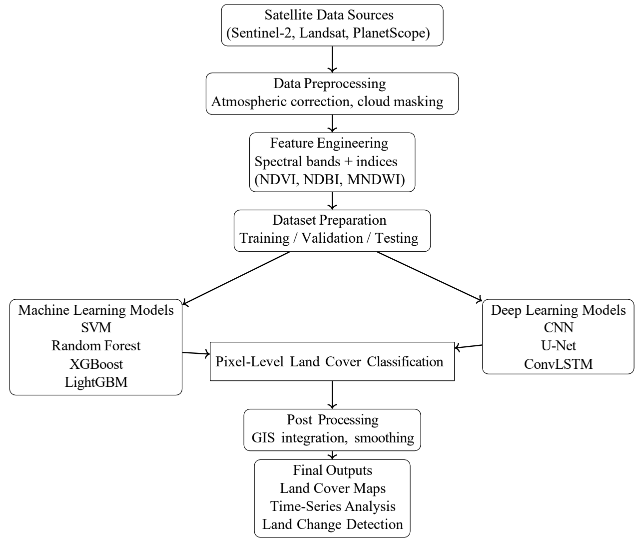

Fig. 1: General workflow of machine learning and deep learning approaches used for satellite-based land cover classification.

A. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning technique for creating optimal decision boundaries in the form of hyperplanes in high-dimensional space [6], [16]. SVM is widely used in remote sensing due to the effectiveness of the technique in classifying spectrally similar land cover types. Furthermore, the use of the kernel method in the technique is advantageous in the sense that it is able to handle nonlinear relationships between the variables.

Previous studies have shown the effectiveness of the SVM technique in remote sensing. For example, Singh et al. showed the effectiveness of the technique in the classification of remote sensing images. Specifically, the authors used the technique to demonstrate the effectiveness of the kernel-optimized SVM technique in the classification of remote sensing images from the Sentinel-2 satellite for the purpose of land cover mapping [10]. Previous studies have shown the effectiveness of the technique in the classification of hyperspectral images. For example, in the study by Xie et al. on the use of the technique in the classification of hyperspectral images for the purpose of distinguishing different types of vegetation as well as agricultural lands. However, despite the high performance of the algorithm in classification tasks, there are a few limitations to the application of the SVM algorithm. These limitations include the need to tune the parameters of the algorithm for the choice of kernel functions. Moreover, the training of the algorithm is computationally intensive for large satellite images with millions of pixels. Based on the observations from the reviewed literature, although the algorithm performs well for the accuracy of the classified images, the limitations of the algorithm to deal with large geospatial images make the ensemble algorithm a better choice.

B. Random Forest

Random Forest (RF) is another popular ensemble learning algorithm that uses multiple decision trees to classify data and improve the stability of the classification model by reducing overfitting problems [5]. RF uses randomness during the bootstrap sampling of the data and the selection of the features to improve the model's ability to generalize and classify complex data sets. Random Forest is one of the most commonly used classifiers in the field of remote sensing due to the strong performance of the algorithm in the presence of noisy data. RF is also efficient in handling high-dimensional data. Previous studies have shown the strong performance of the RF classifier in the field of remote sensing for specific land cover classification problems involving multispectral images. For example, the study by Gislason *et al.* showed the strong performance of the RF classifier in classifying different types of land cover in remote sensing images [9]. Similarly, the comparative study of the RF classifier for the Sentinel-2 image showed the strong performance of the classifier in remote sensing images [7]. Another advantage of RF is that it can estimate the importance of the features. This will allow researchers to identify the best bands of the spectrum used in the classification. This is particularly useful in remote sensing studies where various indices like NDVI, NDBI, and MNDWI are incorporated for better separability of the classes.

From the above studies, it is evident that the performance of the RF algorithm is satisfactory in terms of accuracy as well as robustness for the problem of land cover mapping. Although RF is providing good results for the problem of land cover mapping, recent studies have shown that slightly better accuracy can be achieved through gradient boosting models.

C. Gradient Boosting Models: XGBoost and LightGBM

Gradient boosting algorithms form a sophisticated set of ensemble learning techniques that involve the construction of decision trees sequentially to reduce the prediction error. These techniques aim to learn from the errors of the previous trees to effectively deal with the complex non-linear relationships that exist in multispectral satellite images. Recently, gradient boosting techniques like XGBoost and LightGBM have received considerable attention for application to RS images due to their prediction accuracy and computational speed [15]. XGBoost is a highly optimized implementation of the gradient boosting algorithm that also includes regularization techniques to enhance the stability of the model. XGBoost has been reported to achieve good results for land cover classification tasks due to its ability to effectively deal with the non-linear relationships between the spectral bands or vegetation indices and the land cover classes. Some research carried out on Sentinel-2 images reported that the boosting algorithm performs better for land cover classification by achieving high accuracy compared to other machine learning algorithms [7], [15]. Moreover, the algorithm can deal with missing values and structured data, making it appropriate for land cover classification tasks.

LightGBM is a more recent gradient boosting system developed to enhance the scalability of the computation process. Unlike other conventional boosting algorithms, which use level-wise tree growth, LightGBM uses leaf-wise tree growth along with the histogram-based learning approach to reduce the memory size and the computation time, making it more appropriate for pixel classification problems with millions of data points.

Comparative analysis results reported in the literature also indicate that gradient boosting models often yield superior performance compared to conventional machine learning classifiers for handling large multispectral datasets [15]. Some observations reported in the literature also indicate that the proposed LightGBM model can yield marginally improved accuracy with reduced training time due to the efficient handling of complex interactions between features.

Overall, the literature reviewed above indicates that the gradient boosting model offers an optimal balance between accuracy and computational efficiency. Therefore, the gradient boosting algorithms are becoming popular for implementing land cover mapping systems for handling large areas.

IV. Deep Learning Approaches in Land Cover Mapping

Significant improvements have been seen in the ability of satellite image analysis systems due to the advancements in the capabilities of deep learning techniques, which have enabled the automatic extraction of spatial and spectral features from images. Unlike other machine learning approaches, which rely on the use of handcrafted features, deep learning models are able to learn the features through the application of multiple non-linear transformations. Convolutional Neural Networks (CNNs) have proven to be useful for the analysis of satellite images because of the ability to extract features through the application of convolutional operations.

Among the various deep learning architectures employed for various remote sensing tasks, the most commonly used architecture for semantic segmentation tasks is the U-Net model. Though the model was originally developed for the task of biomedical image segmentation, the architecture of the U-Net model follows an encoder-decoder pattern that allows the model to learn both contextual information and spatial information simultaneously [21]. With the help of the encoder part of the model, the model learns the various features of the images, while the decoder part of the model allows the model to segment the images accurately by performing the task of pixel classification for the various land cover classes.

Some researchers have successfully employed U-Net-based models for land cover mapping using satellite images. To exemplify, Ball *et al.* have demonstrated the effectiveness of U-Net models for classifying satellite images and generating high-quality land cover maps with improved spatial consistency [17]. Further enhancements have also been proposed through architectural extensions to the U-Net model, including the application of attention mechanisms and ensemble U-Net models, which facilitate the improvement of the overall robustness of the classification model for diverse land cover scenarios [19].

Recent research has also proposed the application of temporal modeling approaches with the overall goal of improving the overall effectiveness of the proposed models for land cover classification using satellite images. Convolutional models with the addition of recurrent layers, known as ConvLSTM models, facilitate the analysis of multi-temporal satellite images and the exploration of the overall dependencies between the land cover features [18].

Although the models have good predictive power and the ability to create detailed segmentation maps, they require large datasets and considerable computational power to train the models. Therefore, although the models have higher accuracy than the other models, the lightweight models are good options to consider for large-scale land cover classification problems due to the computational power required by the models.

V. Comparative Insights from Machine Learning Studies

Comparative analysis of various studies reveals that each classification model possesses unique advantages. For example, SVM and Random Forest, which are conventional machine learning techniques, guarantee good performance and are easy to implement. Gradient boosting techniques, such as XGBoost and LightGBM, have shown high precision in their performance compared to others because they are capable of handling complex interactions in features.

An interesting observation from various studies is related to the significance of spectral feature engineering in enhancing classification performance. For example, indices such as NDVI, NDBI, and MNDWI are useful in highlighting vegetation, built-up regions, and water bodies, respectively.

Gradient boosting techniques have shown promising performance while ensuring efficiency in computation, thus making these techniques viable alternatives in designing efficient environmental monitoring systems.

VI. Time-Series Land Cover Analysis and Change Detection

Apart from single-date land cover classification, recent research has also focused on the analysis of land cover change through satellite images. Time series land cover analysis helps to investigate the change in the land cover pattern across various years.

Research carried out on the analysis of land cover change through satellite images revealed that satellite images help to analyze gradual changes in the land cover pattern. Analyzing the land cover images for various years helps to identify the changes occurring in the spatial extent of various land cover classes.

Land change detection techniques involve comparing the land cover images for various years and detecting the changes occurring in the land cover pattern. These changes include the loss of forest cover, urbanization, or changes in agricultural land cover.

VII. Challenges and Future Research Directions

Despite the success that has been achieved with satellite image-based land cover classification, there are a number of challenges that need to be addressed. One of the challenges with satellite image-based land cover classification is the occurrence of cloud cover or any other type of disturbance in the satellite images.

Another challenge with the land cover classification techniques that have been discussed in the paper is the availability of good quality ground truth information. Most land cover datasets do not have balanced classes, meaning that some classes have a large amount of training samples while other classes have a few training samples.

In the future, research is expected to be carried out to develop hybrid machine learning and deep learning techniques for land cover classification. Moreover, the use of multiple data sources for land cover classification could also be explored. Further research could also be carried out to develop techniques for the interpretation of land cover classification techniques.

VIII. Conclusion

This review paper provided a detailed analysis of the machine learning and deep learning techniques that can be used for land cover classification using satellites. The commonly used techniques that were covered in the paper include Support Vector Machine, Random Forest, XGBoost, and LightGBM.

According to the analysis provided in the paper, gradient boosting techniques provide high accuracy with the advantage of high computing efficiency, which can be used to solve environmental problems on a larger scale. Deep learning techniques like U-Net can be used to segment the satellite images at the pixel level.

References

[1] J. Cihlar, "Land cover mapping of large areas from satellites: Status and research priorities," *International Journal of Remote Sensing*, vol. 21, no. 6–7, pp. 1093–1114, 2000.

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

[3] A. Gascon *et al.*, "Copernicus Sentinel-2A calibration and products validation status," *Remote Sensing*, vol. 9, no. 6, p. 584, 2017.

[4] L. Louis *et al.*, "Sentinel-2 Sen2Cor: L2A processor for users," in *Proc. Living Planet Symposium*, 2016.

[5] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[6] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.

[7] S. Ghosh, R. Sharma, and P. Joshi, "Random forest, support vector machine and K-means based land use land cover classification using Sentinel-2 imagery: A comparative study," *Sustainability*, vol. 13, no. 13, p. 758, 2021.

[8] X. Xie, X. Li, Y. Chen, and T. Chen, "Support vector machines for classification of hyperspectral remote sensing imagery," *Remote Sensing Letters*, vol. 2, no. 4, pp. 315–324, 2011.

[9] P. Gislason, J. Benediktsson, and J. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.

[10] R. Singh, S. Sharma, and M. K. Jha, "Comparison of land cover classifiers using different satellite imagery and machine learning techniques," *Remote Sensing Applications: Society and Environment*, vol. 24, p. 100620, 2021.

[11] D. Rodrigues *et al.*, "Preprocessing and classification of Sentinel-2 imagery for vegetation analysis," *Remote Sensing*, vol. 17, no. 480, pp. 1–20, 2022.
[12] J. Li, X. Chen, and Y. Zhang, "Atmospheric correction and preprocessing strategies for Sentinel-2 data," *Remote Sensing*, vol. 17, no. 2138, pp. 1–16, 2022.
[13] A. Alciaturi *et al.*, "Crop monitoring with Sentinel-1 and Sentinel-2 data fusion," *Remote Sensing*, vol. 14, no. 209, pp. 1–17, 2022.
[14] P. D'Odorico *et al.*, "Integrating digital elevation models with Sentinel-2 for land classification," *Sensors*, vol. 25, no. 228, pp. 1–14, 2023.
[15] A. Li *et al.*, "Machine learning approaches for land cover classification using Sentinel-2," *Sustainability*, vol. 13, no. 13758, pp. 1–18, 2021.
[16] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
[17] J. Ball *et al.*, "Application of U-Net for land cover classification using Sentinel-2," *Remote Sensing*, vol. 11, no. 495, pp. 1–14, 2019.
[18] Y. Zhang, M. Zhang, and J. Xu, "ConvLSTM-based U-Net for temporal land cover analysis," *Remote Sensing*, vol. 12, no. 2495, pp. 1–18, 2020.
[19] I. Dimitrovski, D. Kocev, and S. Dzˇeroski, "Ensemble of U-Nets for Sentinel-based mapping," *Applied Sciences*, vol. 11, no. 543, pp. 1–20, 2021.
[20] A. Irfan *et al.*, "Fusion of Sentinel-1 and Sentinel-2 data for deep learning classification," *Sensors*, vol. 23, no. 8966, pp. 1–14, 2023.
[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
[22] M. Abid *et al.*, "Modeling and forecasting land use/land cover changes in Punjab using CA–Markov and ANN," *Earth*, vol. 4, no. 39, pp. 1–20, 2023.
[23] J. Verbesselt, R. Hyndman, A. Zeileis, and D. Culvenor, "Detecting trend and seasonal changes in satellite image time series," *Remote Sensing of Environment*, vol. 114, no. 12, pp. 2970–2980, 2010.
[24] I. Ullah *et al.*, "Integration of Sentinel-2 data into GIS for biomass and land cover mapping," *Forests*, vol. 13, no. 1077, pp. 1–20, 2022.