Accuracy Assessment of Land Use Land Cover Classification Using Machine Learning Classifiers in Google Earth Engine; A Case Study of Jammu District

Sheeba Khan[1], Ashutosh Bhardwaj[2], Mathivanan Sakthivel[3]

[1]University of Madras, Department of Geography, Chennai, Tamil Nadu, India - Sheebakhan1802@gmail.com
[2]Indian Institute of Remote Sensing, Dehradun, Uttarakhand, India - ashutosh@iirs.gov.in
[3]Professor and Head, University of Madras, Department of Geography, Chennai, Tamil Nadu, India - mathisakthi22@gmail.com

Keywords: LULC, Random Forest, Support Vector Machine, Gradient Boosted Trees, Classification and Regression Trees, Machine Learning

ABSTRACT

LULC (Land Use and Land Cover) involves classifying and describing different land types and their usage. Using satellite imagery for LULC mapping is increasing in remote sensing. This study focuses on Jammu district in India, situated between mountain ranges from north and south make it eco-sensitive zone. Expanding of human activity and loss of natural resources make it vulnerable if mismanaged. Study of LULC is important because of it and this study deals with efficiency and accuracy of various machine learning classifiers for LULC. This study uses machine learning classifiers - Random Forest (RF), Support Vector Machine (SVM), Gradient Boosted Trees (GTB), and Classification and Regression Trees (CART) - for the task, leveraging Google Earth Engine (GEE). Sentinel-2 satellite data from January 1 to March 31, 2023, with specific spectral bands (B4, B3, B2 & B8, B4, B3), were used for image preprocessing. Six classes: built-up, water, agricultural land, fallow land, forest, and barren land, each with 100 sample points, were used for classification. After training, the classifiers' accuracy was evaluated using Overall Accuracy and Kappa Coefficient. The results showed RF as the top performer with an overall accuracy of 99.36% and a Kappa coefficient of 99.11%, followed by SVM, GTB, and CART. This highlights the effectiveness of machine learning classifiers, especially RF and SVM, in accurately mapping LULC patterns in Jammu district, suggesting RF's potential as a reliable tool for remote sensing-based LULC mapping.

1. INTRODUCTION

1.1 LULC CHANGES CLASSIFICATION

Land Use Land Cover (LULC) classification is a vital process in understanding and monitoring the changes in the Earth's surface, as it involves categorizing various land areas based on their usage and cover type. From the toposheets to remote sensing imagery played a crucial role in LULC classification. In recent times, Google Earth Engine (GEE) is a powerful cloud-based platform that provides extensive geospatial data and computational resources, enabling large-scale environmental data analysis. GEE plays a crucial role in LULC by offering an accessible and efficient way to process satellite imagery and other geospatial data, making it easier to track land use changes over time (Talukdar et al., 2020). In preparation and analysis of LULC classification maps, various methodology is used from various GIS tools like interpolation techniques etc. Analysis of these is conducted through from various statistical techniques in softwares like- SPSS, STATA etc. Emerging of Machine learning, a subset of artificial intelligence, involves training algorithms to recognize patterns in data can be significant in preparation of LULC classification maps and analysis of their accuracy.

In the context of LULC, machine learning classifiers can analyze vast amounts of satellite imagery to accurately categorize land cover types, enhancing the precision and efficiency of environmental monitoring. Traditional methods for LULC classification often rely on manual interpretation and supervised classification techniques, which can be time-consuming, labour intensive, and prone to human error. In contrast, machine learning methods offer enhanced accuracy and efficiency by automating the classification process, handling large datasets with ease, and continuously improving through learning from new data. The integration of machine learning with platforms like Google Earth Engine further amplifies these benefits, enabling scalable and real-time analysis of vast geospatial datasets.

This study focuses on leveraging machine learning classifiers within the Google Earth Engine platform to perform LULC classification in Jammu District. The objective is to demonstrate the effectiveness of integrating advanced machine leaning classifiers in LULC classification. By comparing the performance of different machine learning classifiers—RF, SVM, GTB, and CART—the study aims to identify the most effective algorithm for LULC classification in this region.

1.2 LAND USE LAND COVER (LULC) CHANGES IN JAMMU DISTRICT

Jammu District, situated in the northernmost part of the India, and winter capital of union territory Jammu & Kashmir boasts a variety of physical characteristics influenced by its geographical location, spanning roughly 32.5° to 33.5° North latitude and 74° to 75.5° East longitude. Positioned within the temperate climatic zone, the area experiences distinct seasonal changes, with mean annual rainfall typically falling between 700 to 1100 mm. From south to the north slope is gradually increasing and bounded by pir panjal ranges of the outer Himalaya in the north. Southern part of the Himalayas makes its southern boundary and between these mountain ranges, fertile plain is situated and tavi river and its tributaries flow from here. Most of the human activity is in this plain. Aspect o the district is very diverse. In the southern plains along the

Tawi River, the aspect is predominantly characterized by gentle slopes and relatively low relief, fostering agricultural activities and human settlement. Moving northward towards the foothills of the Shivalik range, the aspect becomes more varied, with slopes ranging from moderate to steep, interspersed with valleys and ridges. Here, the aspect is influenced by the northeast-facing slopes of the Shivalik Hills, which receive less solar radiation and tend to be cooler and more humid compared to their southwest-facing counterparts. As one traverses further north into the Pir Panjal range, the aspect becomes increasingly rugged and mountainous, with steep slopes, deep valleys, and towering peaks dominating the landscape. The aspect of the Pir Panjal range is influenced by its east-west orientation, with aspects ranging from north-facing to south-facing slopes, each with distinct microclimatic conditions and vegetation patterns. Additionally, the aspect of Jammu District plays a crucial role in determining land use patterns, hydrological processes, and ecological habitats across the region. These diverse physical features not only enhance the district's visual appeal but also shape its socio-economic and ecological dynamics. While the fertile plains support agriculture, the mountainous areas offer opportunities for ecotourism. Previous study which are based on the remote sensing techniques shows significant changes in the district. Agricultural lands (6.71%), Barren land (6.45%), and settlement (4.12%) increased while vegetation decreased by 16.57% (Saleem et al., 2023). Previous studies highlight the issue related to this where pir panjal range is prone to landslide specially in monsoon season which can alter the geography of the district. Changes in the land use pattern can make it vulnerable to worse situation because of the its geographical situation.

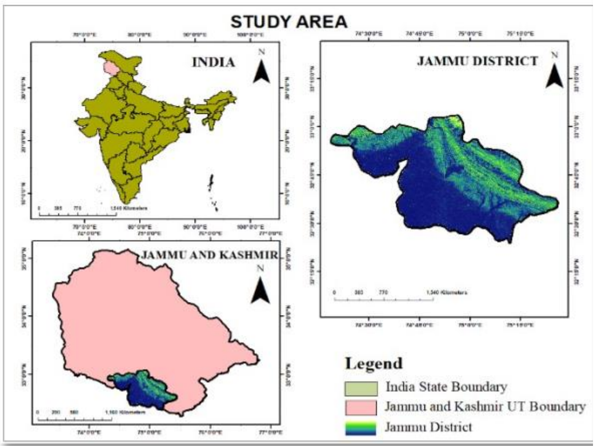

Figure 1: Study Area Map

In this paper we are trying to achieve the LULC pattern of the Jammu district using four machine learning classifier which will contribute for further research. Accuracy of each machine learning is discussed and will help to research community about suitability of each machine learning classifiers for further assessment of LULC pattern not only bounded to the study area but other different study areas also.

2. METHODOLOGY

Jammu district detailed analysis of land use and land cover (LULC) was conducted using advanced machine learning techniques within the Google Earth Engine platform. This process involved the application of four classifiers, including Random Forest, Support Vector Machine (SVM), Gradient Tree Boosting (GTB), and Classification and Regression Trees (CART). These classifiers are powerful algorithms capable of discerning patterns and making accurate predictions based on input data (Loukika et al.,2021).

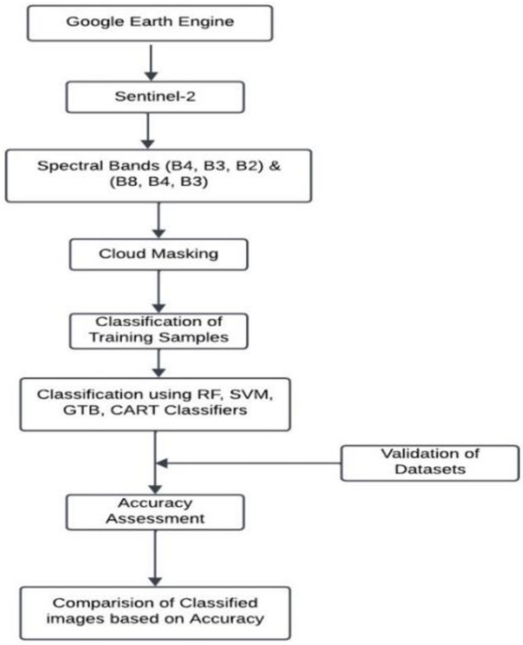

Figure 2: Research Methodology Chart

2.1. DATA PROCESSING AND IMAGE CLASSIFICATION

The primary data source for this analysis is Sentinel-2 multispectral imagery, a critical tool for capturing comprehensive information across various wavelengths of light. Specifically, data acquired between January 1st, 2023, to March 30th, 2023, is chosen due to its suitability in capturing seasonal variations and land cover changes during this period. Sentinel-2 imagery is renowned for its high spatial resolution and spectral detail, making it ideal for precise land cover classification tasks, which are essential for understanding environmental dynamics and planning sustainable land use. Cloud masking step is implemented to remove any cloud cover from the images, ensuring that the analysis is based on clear and unobstructed views of the land surface. To classify the land cover, six distinct classes are identified: built-up areas, water bodies, agricultural land, forested areas, barren land, and grasslands. The classification process involves using four machine learning classifiers, including Random Forest (RF), Support Vector Machine (SVM), Gradient Tree Boosting (GTB), and Classification and Regression Trees (CART). These classifiers analyze the training samples and apply the learned patterns to classify the entire image. This classification is based on the level-1 classification system established by Anderson et al. (1976), which provides a robust framework for categorizing land cover types. Ensuring the accuracy and reliability of the classification results is paramount, hence a balanced dataset comprising 100 samples for each class is employed. This meticulous approach helps to mitigate biases and ensures that each land cover category is adequately represented in the analysis, thereby enhancing the robustness and credibility of the results.

For visual interpretation and validation purposes, the spectral bands B4 (Red), B3 (Green), and B2 (Blue) are utilized to generate RGB representations of the imagery. This allows for intuitive visualization of different land cover types based on their colour characteristics, facilitating easy differentiation and analysis. Additionally, False Colour Composite (FCC) images are generated using bands B8, B4, and B3. The FCC images enhance the discrimination of land cover features by highlighting variations in spectral reflectance that may not be readily apparent in traditional RGB representations. These enhanced visualizations are invaluable for identifying subtle differences in vegetation health, water bodies, and built-up areas, providing deeper insights into the landscape's composition and changes over time. Through the integration of Sentinel-2 imagery and advanced classification techniques, this study provides a detailed and accurate depiction of land cover dynamics in the study area. The methodology employed ensures that the data is not only comprehensive but also reliable, laying a solid foundation for subsequent analyses and decision-making processes.

2.2 ACCURACY ASSESSMENT OF MACHINE LEARNIG CLASSIFIERS

The integration of machine learning classifiers and Sentinel-2 multispectral imagery enables the generation of accurate and detailed maps depicting the distribution of various land cover types across Jammu district. This innovative approach leverages the high spatial resolution and spectral detail provided by Sentinel-2 imagery, combined with the analytical power of machine learning algorithms, to produce comprehensive land use and land cover (LULC) maps. The accuracy of these maps is assessed using two different metrics to ensure their reliability and validity.

(1) The first metric, overall accuracy, measures the likelihood of accurate classification in a test. This is mathematically represented as:

$$Overall\ Accuracy = \frac{TP + TN}{P + N}$$

where TP represents True Positives and TN represents True Negatives. This metric provides a straightforward assessment of how well the classifier performs in correctly identifying land cover types.

(2) The second metric, the Kappa coefficient, evaluates the agreement between the classified data and a reference dataset, taking into account the possibility of agreement occurring by chance. The Kappa coefficient is calculated using the formula:

$$Kappa\ coefficient = \frac{Po - Pe}{(1 - Pe)}$$

Where Po is the observed agreement and Pe is the expected agreement. The Kappa coefficient provides a more robust measure of classification accuracy, particularly in scenarios where the distribution of classes is imbalanced.

In this study, the same set of training samples is used for both classifier training and validation. This approach ensures a consistent evaluation of both overall accuracy and the Kappa coefficient. Post-classification analysis involves generating maps that clearly depict the spatial distribution of land cover types in Jammu District. These maps are essential for visualizing the results and understanding the spatial patterns and changes in land cover over time. The study's findings provide valuable insights into effective LULC mapping methodologies, highlighting the benefits of integrating machine learning techniques with high-resolution satellite imagery. The resulting maps serve as critical tools for land management, offering detailed information that can guide sustainable development practices and inform policy decisions. By advancing the accuracy and reliability of LULC mapping, this research contributes to the broader goal of promoting sustainable land management and environmental conservation in Jammu District and beyond.

Sr. No.	Machine Learning Classifiers	Description
1.	Random Forest (RF)	Random Forest is a versatile machine learning algorithm that constructs multiple decision trees independently, each trained on random subsets of features and data samples. By aggregating predictions through voting or averaging, it mitigates overfitting and enhances model robustness.
2.	Support Vector Machine (SVM)	It is a powerful algorithm for classification and regression tasks, known for its effectiveness in high-dimensional spaces, ability to maximize margin between classes, capability to handle non-linear boundaries through kernel functions, and efficiency in using only a subset of training data points.
3.	Gradient Tree Boosting (GTB)	It is a robust machine learning algorithm known for its efficiency in handling complex data patterns, its ability to sequentially improve model performance by minimizing errors, and its capability to handle various types of data, including numerical and categorical features.
4.	Classification And	It is a simple yet powerful algorithm that recursively partitions the feature space

	Regression Trees (CART)	with binary splits, handling both numerical and categorical data. It provides insights into feature importance and produces understandable decision trees for classification and regression tasks.

Table 1: Detail of used Machine Learning Classifiers

3. RESULTS AND DISCUSSION

In conducting Land Use and Land Cover (LULC) classification over Jammu district, we employ various machine learning classifiers within the Google Earth Engine platform to ensure a comprehensive and accurate analysis. The LULC classification is based on six key parameters: built-up areas, water bodies, agricultural land, fallow land, forested areas, and barren land. These parameters are carefully selected to represent the diverse land cover types present in the region, each contributing uniquely to the landscape's ecological and socioeconomic fabric. The built-up areas parameter encompasses all urban and infrastructural developments, reflecting the extent of human settlement and industrial activity. Water bodies include rivers, lakes, and reservoirs, crucial for understanding the region's hydrological dynamics and water resource management. Agricultural land represents areas under cultivation, essential for assessing food production and rural livelihoods. Fallow land, which includes temporarily unused agricultural fields, helps in understanding crop rotation practices and land management strategies.

3.1 ACCURACY ASSESSMENT OF MACHINE LEARNIG CLASSIFIERS

In the LULC classification of Jammu District using machine learning classifiers in Google Earth Engine, Random Forest classifier achieved an impressive overall accuracy of 99.36%, indicating that it correctly classified the vast majority of land cover types in the study area. Additionally, the Kappa coefficient for RF was 99.11%, reflecting a high level of agreement between the RF classification results and the reference data. This high Kappa coefficient underscores the robustness of the RF classifier in accurately distinguishing between different land cover types, making it the most effective tool in this study. Following the Random Forest classifier, the Support Vector Machine (SVM) also showed strong performance, attaining an overall accuracy of 90.48%. The Kappa coefficient for SVM was 86.70%, which, while lower than that of RF, still indicates a substantial agreement between the classified data and the reference datasets. SVM's performance highlights its capability to effectively handle complex classification tasks, although it falls slightly short compared to RF in this particular study. The Gradient Boosting Trees (GTB) classifier achieved an overall accuracy of 86.27%, with a Kappa coefficient of 80.76%. While GTB performed well, its accuracy and Kappa values were lower than those of RF and SVM, suggesting that GTB, although useful, may not be as effective for this specific LULC classification task. The results indicate that GTB can still provide valuable insights but might require further tuning or additional data for optimal performance. Lastly, the Classification and Regression Trees (CART) classifier yielded an overall accuracy of 86.24% and a Kappa coefficient of 80.87%. These metrics are comparable to those of GTB, indicating that CART is capable of performing reliable LULC classification, but with similar limitations as GTB. The slightly higher Kappa coefficient compared to GTB suggests a marginally better agreement with the reference data, although the difference is not substantial.

Classes	**Classifier**			
	RF	**SVM**	**GTB**	**CART**
Built-up	637.05	611.01	103.505	746.794
Water	27.79	14.93	20.74	20.309
Agriculture	1224.75	1403.91	1195.102	1217.702
Fallow Land	759.47	794.54	797.895	9016.68
Forest	1653.06	1502.46	1257.222	1417.57
Barren land	38.77	14.01	34.84	36.86

Table 2: Results of Various Parameters Using Machine Learning for LULC Classification

Classifier	**Accuracy Assessment**	
	Overall Accuracy	**Kappa Coefficient**
RF	99.35	99.1
SVM	90.48	86.7
GTB	86.26	80.76
CART	86.23	80.86

Table 3: Accuracy Result of Various Machine Learning for LULC Classification

The results highlight the Random Forest classifier stands out as the most effective method for LULC classification in Jammu District, providing the highest accuracy and Kappa coefficient. The Support Vector Machine, Gradient Boosting Trees, and Classification and Regression Trees also contribute valuable results, with varying degrees of effectiveness. the study's results emphasize the importance of choosing the right machine learning classifier for LULC classification tasks. The demonstrated effectiveness of the RF classifier provides a benchmark for future studies and applications, highlighting its potential to deliver accurate and reliable land cover maps. This contributes significantly to the advancement of land management practices and the development of informed, sustainable planning strategies in Jammu District and similar regions.

3.2 LULC MAP OF JAMMU DISTRICT BASED ON MACHINE LEARNING CLASSIFIERS

LULC maps of the Jammu district is prepared based on the google earth engine (GEE) after using 4 machine leaning classifiers. Subsequently four maps are prepared after each machine learning classifiers.

https://doi.org/10.5194/isprs-archives-XLVIII-4-2024-263-2024 | © Author(s) 2024. CC BY 4.0 License.

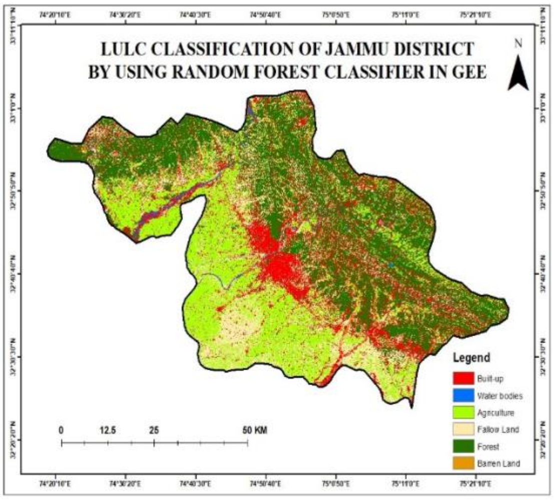

Figure 3: LULC Classification of Jammu District Using Random Forest

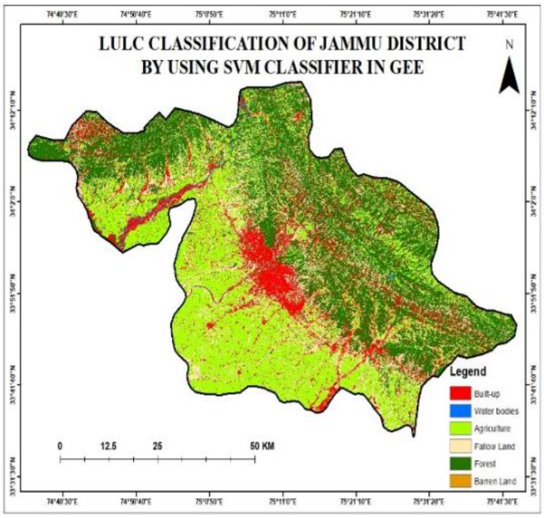

Figure 4: LULC Classification of Jammu District Using SVM

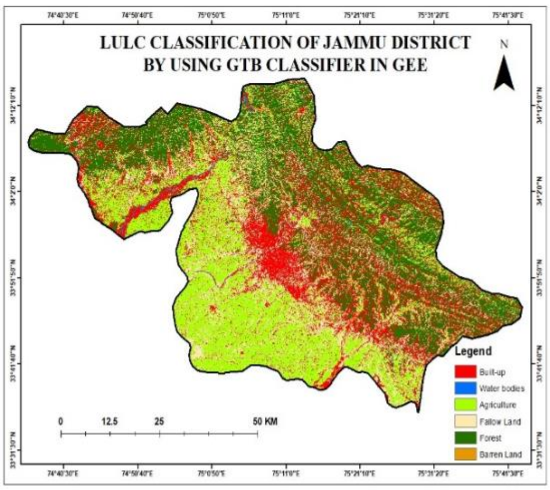

Figure 5: LULC Classification of Jammu District Using GTB

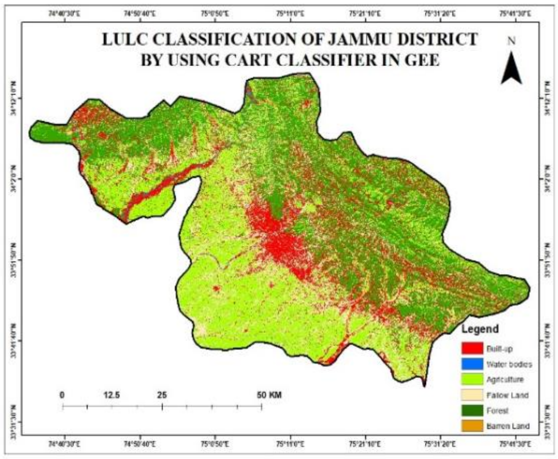

Figure 6: LULC Classification of Jammu District Using CART

Figures 3, 4, 5, and 6 present the maps prepared using the Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Trees (GTB), and Classification and Regression Trees (CART) classifiers, respectively. These maps visually depict the results of the Land Use and Land Cover (LULC) classification over Jammu District, highlighting the distribution of the six major land cover classes: built-up areas, water bodies, agricultural land, fallow land, forested areas, and barren land.

In Figure 3, the map generated using the Random Forest classifier stands out for its remarkable accuracy. Each land cover class is clearly identified and distinctly mapped, reflecting the classifier's superior performance. Figure 4 shows the map produced using the Support Vector Machine classifier. While this map is also accurately represent the different land cover classes, it is slightly less precise than the RF map. The map generated by the Gradient Boosting Trees classifier is presented in Figure 5. GTB performs well, the map reveals areas where the classifier struggles to distinguish between similar land cover types, resulting in less clarity compared to the RF and SVM maps. Finally, Figure 6 displays the map produced using the Classification and Regression Trees classifier. The map is comparable to the one generated by GTB, showing reasonable accuracy but with certain ambiguities in class differentiation. The CART map is useful for general analysis but lacks the detailed precision seen in the RF and, to a lesser extent, SVM maps.

4. CONCLUSION

The study demonstrates the effectiveness of machine learning classifiers in accurately mapping Land Use and Land Cover (LULC) in Jammu district using satellite imagery processed through Google Earth Engine. Among the classifiers tested—Random Forest (RF), Support Vector Machine (SVM), Gradient Boosted Trees (GTB), and Classification and Regression Trees (CART)—RF exhibited superior performance, achieving an overall accuracy of 99.36% and a Kappa coefficient of 99.11%. These results highlight RF's robustness and reliability in LULC classification, making it a potent tool for remote sensing applications. This high accuracy is critical for various practical applications, including urban planning, agricultural management, forest conservation, and disaster management. Accurate LULC maps can assist policymakers and stakeholders in making informed decisions, optimizing land resource use, and implementing sustainable development strategies. Furthermore, the demonstrated efficacy of RF and SVM in this context underscores their potential for broader applications including urban planning, agricultural monitoring, and environmental conservation. Future research could explore the integration of additional data sources and further refinement of classification algorithms to enhance the accuracy and applicability of LULC studies.

REFERENCES

Ahmed, S. A., & N, H., 2023. Land use and land cover classification using machine learning algorithms in google earth engine. *Earth Science Informatics*, 16(4), 3057-3073.

Amini, S., Saber, M., Rabiei Dastjerdi, H., & Homayouni, S., 2022. Urban land use and land cover change analysis using random forest classification of landsat time series. *Remote Sensing*, 14(11), 2654.

Anderson, J. R., 1976. A land use and land cover classification system for use with remote sensor data. *US Government Printing Office.* (Vol. 964)

Aryal, J., Sitaula, C., & Frery, A. C., 2023. Land use and land cover (LULC) performance modeling using machine learning algorithms: a case study of the city of Melbourne, Australia. *Scientific Reports*, 13(1), 13510.

Iqbal, M. F., & Khan, I. A. (2014). Spatiotemporal land use land cover change analysis and erosion risk mapping of Azad Jammu and Kashmir, Pakistan. *The Egyptian journal of remote sensing and space science*, 17(2), 209-229.
Loukika, K. N., Keesara, V. R., & Sridhar, V., 2021. Analysis of land use and land cover using machine learning algorithms on google earth engine for Munneru River Basin, India. *Sustainability*, 13(24), 13758.

Macaspac, K. L., Valenzuela, M. L., & Ramos, R. V., 2024. Land use land cover (LULC) change analysis in Baguio city using GIS-based transition potential modeling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 48, 357-363.

Malik, M. M., Tali, J. A., & Nusrath, A. (2017). Assessment of land use and land cover change in district Baramulla, Jammu and Kashmir. *Research & Reviews: Journal of Space Science & Technologyl*, 6(3), 1-14p.
Meer, M. S., Mishra, A. K., & Rafiq, M. (2021). Spatio-temporal patterns of land use land cover changes over a district in Northern India and their impact on environment and society. *Journal of the Geological Society of India*, 97(6), 656-660.
Nguyen, H. T. T., Doan, T. M., Tomppo, E., & McRoberts, R. E., 2020. Land use/Land cover mapping using multitemporal Sentinel-2 imagery and four classification methods—A case study from Dak Nong, Vietnam. *Remote Sensing*, 12(9), 1367

Saleem, Haadiya, Rayees Ahmed, Shaista Mushtaq, Shahid Saleem, and Mudigandla Rajesh. 2024. Remote Sensing-Based Analysis of Land Use, Land Cover, and Land Surface Temperature Changes in Jammu District, India. *International Journal of River Basin Management,* March, 1–16. doi:10.1080/15715124.2024.2327493.

Shafiullah, G., Al-Ruwaih, F., & Umar, S. M., 2023. Morphometric analysis using multivariate statistical approaches for the prioritization of potential watersheds in Kuwait. *International Journal of Energy and Water Resources,* 1-14.

Shetty, S., 2019. Analysis of machine learning classifiers for LULC classification on Google Earth Engine (Master's thesis, University of Twente).

Singh, R. K., Singh, P., Drews, M., Kumar, P., Singh, H., Gupta, A. K., & Kumar, M., 2021. A machine learning-based classification of LANDSAT images to map land use and land cover of India. *Remote Sensing Applications: Society and Environment,* 24, 100624.

Talukdar, S., Singha, P., Mahato, S., Pal, S., Liou, Y. A., & Rahman, A., 2020. Land-use land-cover classification by machine learning classifiers for satellite observations—A review. *Remote sensing*, 12(7), 1135.

Tassi, A., & Vizzari, M., 2020. Object-oriented LULC classification in google earth engine combining SNIC, GLCM, and Machine learning algorithms. *Remote Sensing,* 12(22), 3776.

Zahoor, S., Wani, A. A., Masoodi, T. H., Islam, M. A., Gatoo, A. A., & Gangoo, S. A. (2020). Land Use Land Cover Classification and Mapping using Geospatial Techniques in Ganderbal District, J & K, India. *International Journal of Current Applied Sc*ience, 9(11), 3838-3847.