Spatio-Temporal Analysis of Vegetation Cover Change (LULC) in Diamer using Geospatial Techniques and Supervised Classification.

Silicon Global Tech, Gilgit, Gilgit-Baltistan

Mehtab Ali

03408139030

alimehtab0003@gamil.com

Supervisor: Hafiz u din

Abstract

This research proposes a comprehensive spatio-temporal assessment of vegetation cover change in the Diamer region using advanced geospatial and machine learning techniques. Highland ecosystems such as Diamer are highly vulnerable to climate variability and increasing anthropogenic pressures, making continuous monitoring essential. The study will integrate multi-temporal Landsat datasets with vegetation indices—primarily the Normalized Difference Vegetation Index (NDVI)—to quantify vegetation dynamics across different time periods. These indices will help detect subtle shifts in vegetation health linked to climatic variations, land use changes, and infrastructure development.To ensure accurate land use and land cover (LULC) classification, the study will employ supervised machine learning algorithms including Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT). These models have demonstrated strong performance in handling complex spectral variability and are expected to enhance classification accuracy for mountainous environments. The classification outputs will undergo rigorous accuracy assessment using confusion matrices, producer's/user's accuracy, overall accuracy, and Kappa coefficient metrics. Additionally, temporal change detection techniques will be applied to identify long-term vegetation trends and spatial distribution of degradation or improvement zones. By combining multispectral indices with machine learning, the research aims to provide a robust evaluation of environmental changes and their drivers.The outcomes will offer valuable insights for regional planners, environmental managers, and policymakers working toward sustainable development, climate adaptation, and natural resource conservation in the Diamer district. Ultimately, this study seeks to support evidence-based decision-making aligned with SDG 9 and SDG 13.

1. Introduction

Highland ecologies are the most susceptible to climate change, often experiencing intensified impacts. The mountainous regions of Pakistan, particularly in areas like Diamer, face significant environmental challenges due to both climatic variations and anthropogenic activities. Due to climate change and human activities, there were dramatic changes in the alpine domain of the China-Pakistan Economic Corridor (CPEC), which is a vital project of the Belt and Road Initiative (BRI). These changes necessitate comprehensive monitoring and assessment of land surface dynamics to ensure sustainable development in these fragile ecosystems. The study of vegetation cover change through Land Use and Land Cover (LULC) analysis has become increasingly critical for understanding environmental transformations in mountainous regions., research in this domain provides essential insights for environmental management and policy formulation. Remote sensing techniques and geospatial analysis have emerged as powerful tools for monitoring vegetation dynamics over time. The combination of multispectral indices and the AI provides a comprehensive insight into how various factors affect the mountainous landscape and climatic conditions in the study area. The Normalized Difference Vegetation Index (NDVI) serves as a fundamental parameter for assessing vegetation health and coverage, as demonstrated in recent studies where NDVI distribution shows a decreasing trend (-0.00469/year, $p > 0.05$) in similar mountainous environments. This research on "Spatio-Temporal Analysis of Vegetation Cover Change (LULC) in Diamer using Geospatial Techniques and Supervised Classification" employs advanced machine learning algorithms including Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT) models to analyze vegetation change patterns. From background knowledge, these supervised classification techniques are particularly effective for LULC mapping as they can handle complex spectral signatures and provide high accuracy in distinguishing different land cover classes. The integration of NDVI parameters with these machine learning approaches enables precise detection and quantification of vegetation changes over temporal scales. The significance of this research extends beyond academic inquiry, as this study has practical and highly relevant implications for policymakers and researchers interested in research related to land use and land cover change, environmental and infrastructure development in alpine regions. Understanding vegetation dynamics in Diamer is crucial for sustainable resource management, climate change adaptation strategies, and informed decision-making for regional development initiatives. To assess change detection and find climatic conditions using multispectral indices along the mountainous area, this research aims to achieve several key objectives that align with sustainable development practices in highland environments. (1) **Temporal Vegetation Analysis** Analyze vegetation cover changes over multiple time periods using NDVI parameters to identify trends and patterns in vegetation dynamics. Highland ecologies are the most susceptible to climate change, often experiencing intensified impacts, making temporal analysis crucial for understanding long-term environmental changes in Diamer. (2) **Machine Learning Model Implementation** Implement and compare the performance of three supervised classification algorithms - Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT) - for accurate LULC classification and vegetation mapping. From background knowledge, these models offer different strengths: SVM excels in handling high-dimensional data, Random Forest provides robust ensemble predictions, and Decision Trees offer interpretable classification rules.(3) **Multispectral Index Integration** Utilize comprehensive multispectral indices beyond NDVI to enhance vegetation monitoring capabilities. The combination of multispectral indices and the AI provides a comprehensive insight into how various factors affect the mountainous landscape and climatic conditions in the study area.

2. Literature Review

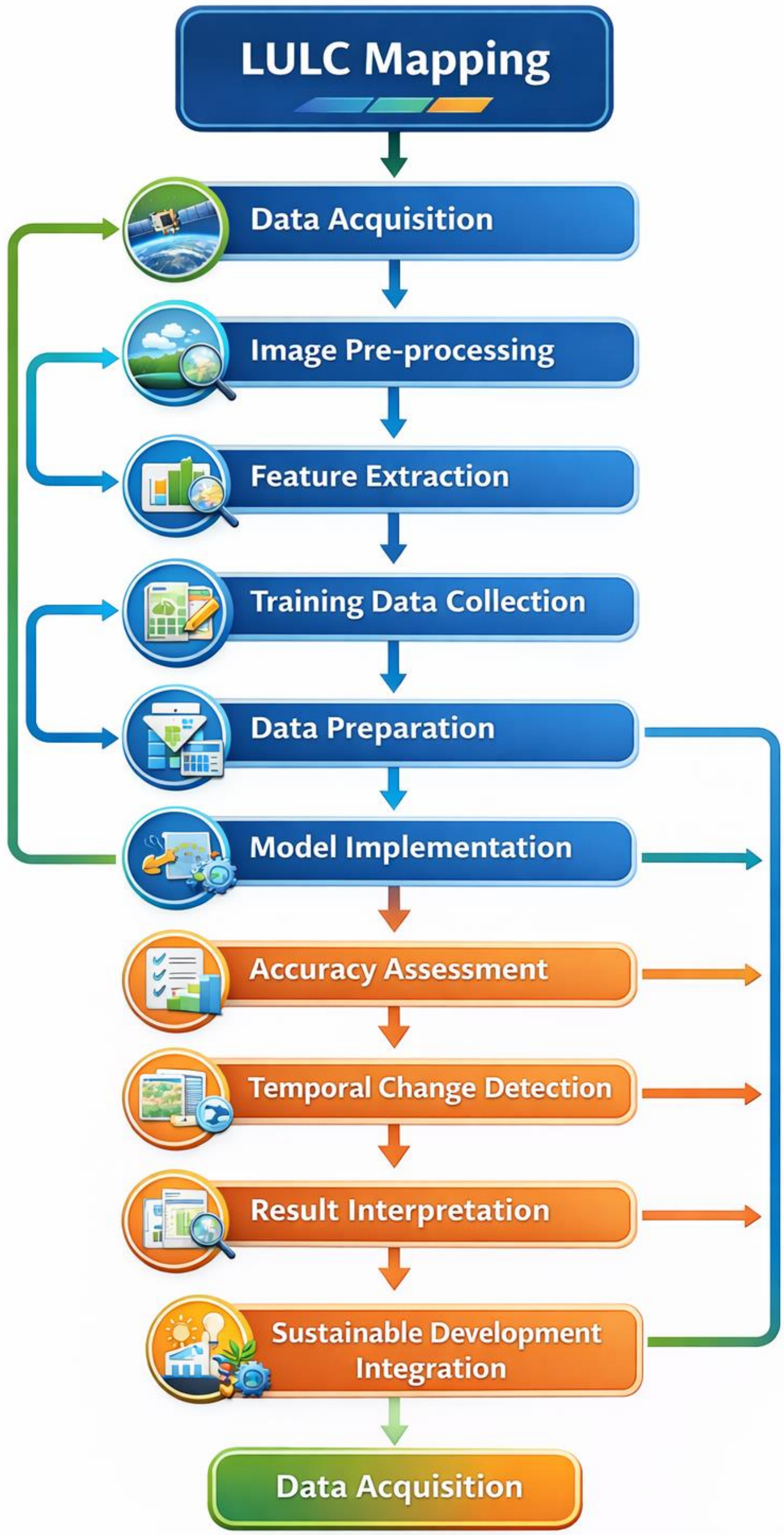
Highland ecologies are among the most vulnerable to climate change, often experiencing intensified impacts; therefore, studying vegetation dynamics in mountainous regions is increasingly critical for environmental management and sustainable development. Gönençgil et al. [1] demonstrated that due to climate change and human activities, dramatic changes have occurred in the alpine domain of the China–Pakistan Economic Corridor (CPEC), a vital project of the Belt and Road Initiative (BRI). Their findings highlight the complex interactions between natural climate variability and anthropogenic pressures in highland environments. The study further emphasized that rapid infrastructure expansion within the CPEC corridor may increase land surface susceptibility, underscoring the urgent need for comprehensive environmental monitoring in these fragile ecosystems.

In their analysis, 30×30 m Landsat 5 (TM), Landsat 7 (ETM+), and Landsat-8/9 (OLI and TIRS) data, along with meteorological records, were used to calculate the Aridity Index (AI), providing a robust multi-temporal framework for environmental assessment [2], [3]. The results revealed concerning patterns: NDVI exhibited a decreasing trend (-0.00469/year, $p > 0.05$), while NDBI showed substantial increases of 79.89%, 87.69%, and 83.85% between 2008 and 2023, indicating vegetation decline concurrent with urban expansion [2]. Additionally, AI results showed a non-significant decreasing trend (-0.0021/year, $p > 0.05$) in Gilgit and a significant decreasing trend (-0.0262/year, $p < 0.05$) in Hunza–Nagar, suggesting a drying tendency and increasing drought risk in an already arid to semi-arid environment [1].

The authors concluded that integrating multispectral indices with AI provides comprehensive insights into how various factors influence mountainous landscapes and climatic conditions, aligning with SDG 9 (industry, innovation, and infrastructure) and SDG 13 (climate action) to support sustainable and regenerative development in highland regions. Furthermore, machine learning methods such as Support Vector Machines (SVM), Random Forest (RF), and Decision Trees (DT) have emerged as powerful tools for land-use and land-cover (LULC) classification. These algorithms outperform traditional pixel-based approaches due to their ability to manage complex spectral data and deliver high classification accuracy [4]–[8]. When combined with vegetation indices like NDVI [2], these classifiers enable precise detection and quantification of vegetation changes across temporal scales, making them especially valuable in complex mountainous environments. This integrated research framework offers strong practical implications for policymakers and researchers focused on LULC change, environmental management, and infrastructure development in alpine regions.

3. Methodology

Methodology Framework


LULC Mapping
Data Acquisition
Image Pre-processing
Feature Extraction
Training Data Collection
Data Preparation
Model Implementation
Accuracy Assessment
Temporal Change Detection
Result Interpretation
Sustainable Development Integration
Data Acquisition

Study Area and Data Acquisition

The methodology follows established approaches for highland environmental monitoring, as demonstrated by Gönençgil et al. (2024) who utilized 30 × 30 m Landsat 5 (TM), Landsat 7 (ETM+), and Landsat-8/9 (OLI and TIRS), and meteorological data were employed to calculate the aridity index (AI) . For this study of Diamer, multi-temporal Landsat imagery will be acquired from the USGS Earth Explorer platform covering the study period. From background knowledge, Landsat data provides consistent temporal coverage with appropriate spatial resolution for regional-scale LULC analysis, making it ideal for vegetation monitoring studies.

Image Pre-processing and Enhancement

The satellite imagery will undergo comprehensive pre-processing including atmospheric correction, geometric correction, and radiometric calibration. From background knowledge, these steps are essential to ensure data quality and comparability across different acquisition dates. Cloud masking and gap-filling techniques will be applied to remove atmospheric interference and ensure data continuity.

Vegetation Index Calculation

The Normalized Difference Vegetation Index (NDVI) will be calculated using the standard formula incorporating near-infrared and red spectral bands. This approach aligns with established methodologies where NDVI distribution shows a decreasing trend (-0.00469/year, $p > 0.05$) has been effectively used to monitor vegetation dynamics in similar mountainous environments. From background knowledge, NDVI values range from -1 to +1, where higher positive values indicate healthier and denser vegetation cover.

Machine Learning Model Implementation

Three supervised classification algorithms will be implemented: Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT). From background knowledge, SVM excels in handling high-dimensional spectral data through kernel functions, Random Forest provides robust ensemble predictions by combining multiple decision trees, and Decision Trees offer interpretable classification rules. Training samples will be collected through stratified random sampling across different land cover classes.

Accuracy Assessment and Validation

Model performance will be evaluated using confusion matrices, overall accuracy, producer's accuracy, user's accuracy, and kappa coefficients. From background knowledge, these statistical measures provide comprehensive assessment of classification reliability and enable comparison between different machine learning approaches.

Temporal Change Analysis

Multi-temporal analysis will be conducted to detect vegetation cover changes over the study period. The combination of multispectral indices and the AI provides a comprehensive insight into how various factors affect the mountainous landscape and climatic conditions in the study area . Change detection techniques including post-classification comparison and direct multi-date classification will be employed.

Sustainable Development Integration

The methodology incorporates sustainable development goals, mainly SDG 9 (industry, innovation, and infrastructure) and SDG 13 (climate action), to evaluate the conservation and management practices for the sustainable and regenerative development of the mountainous region, ensuring that the research outcomes support evidence-based environmental management.

The methodology ensures that this study has practical and highly relevant implications for policymakers and researchers interested in research related to land use and land cover change, environmental and infrastructure development in alpine regions, providing a comprehensive framework for vegetation monitoring in highland environments.

4. Evaluation Metrics

NDVI Formula and Application

The Normalized Difference Vegetation Index (NDVI) will be calculated using the standard formula: NDVI = (NIR - Red) / (NIR + Red), where NIR represents the near-infrared band and Red represents the red band reflectance values. From background knowledge, NDVI values range from -1 to +1, where values closer to +1 indicate dense, healthy vegetation, values near 0 represent bare soil or rock, and negative values typically indicate water bodies or snow. The research follows established methodologies where NDVI distribution shows a decreasing trend (-0.00469/year, $p > 0.05$) , demonstrating the effectiveness of this index for monitoring vegetation changes in mountainous environments.

Kappa Coefficient Assessment

From background knowledge, the Kappa coefficient (κ) will be calculated using the formula: $\kappa = (Po - Pe) / (1 - Pe)$, where Po is the observed accuracy and Pe is the expected accuracy due to chance. The Kappa coefficient ranges from -1 to +1, with interpretation guidelines: $\kappa > 0.8$ indicates strong agreement, $0.6 \leq \kappa \leq 0.8$ represents moderate agreement, $0.4 \leq \kappa \leq 0.6$ shows fair agreement, and $\kappa < 0.4$ indicates poor agreement. This metric provides a robust measure of classification accuracy that accounts for chance agreement, making it particularly valuable for comparing the performance of different machine learning algorithms (SVM, RF, DT).

Overall Accuracy Assessment

From background knowledge, Overall Accuracy will be calculated as the ratio of correctly classified pixels to the total number of reference pixels: Overall Accuracy = (Sum of diagonal elements in confusion matrix) / (Total number of reference pixels) × 100. This metric provides a general measure of classification performance across all land cover classes and serves as the primary indicator of model effectiveness.

Producer's and User's Accuracy

From background knowledge, Producer's Accuracy measures the probability that a reference pixel is correctly classified: Producer's Accuracy = (Correctly classified pixels in a category) / (Total reference pixels in that category) × 100. User's Accuracy represents the probability that a classified pixel actually represents that category on the ground: User's Accuracy = (Correctly classified pixels in a category) / (Total classified pixels in that category) × 100. These metrics are essential for understanding classification errors and model reliability for specific vegetation classes.

F1-Score and Precision-Recall Metrics

From background knowledge, the F1-Score will be calculated as the harmonic mean of precision and recall: F1-Score = 2 × (Precision × Recall) / (Precision + Recall). Precision measures the accuracy of positive predictions: Precision = True Positives / (True Positives + False Positives), while Recall measures the model's ability to identify all positive instances: Recall = True Positives / (True Positives + False Negatives). These metrics are particularly important for evaluating model performance on imbalanced vegetation datasets.

Confusion Matrix Analysis

From background knowledge, the confusion matrix will serve as the foundation for all accuracy assessments, providing a cross-tabulation of predicted versus actual land cover classes. The matrix enables calculation of commission errors (false positives) and omission errors (false negatives) for each vegetation category, which are essential for understanding model performance across different NDVI-derived vegetation classes.

5. Tools and Technologies

Tools / Library	**Category**	Task
ArcMap	Application	Used for traditional GIS operations, spatial analysis, and map production,
Google Earth Pro	Application	To provide high-resolution imagery for visual interpretation and ground truth validation
Python	Programing Language	For Code Implementation
Pandas / Numpy	Frameworks	For data manipulation/ numerical computations
Matplotlib	Frameworks	For visualization
Scikit-learn	ML Framework	For implementing machine learning algorithms
Jupiter Notebook	Development Environment	Will provide an interactive development environment for code execution, data visualization, and documentation
Microsoft Word	Documentations	Will be used for research documentation, report writing, and manuscript preparation.
Microsoft Excel	Tabulations	Will handle tabular data management, statistical calculations, and accuracy assessment tabulations.

6. References

[1] B. Gönençgil, M. U. Khan, and N. G. Barlas, “Assessment of climate and environmental changes in the alpine domain of the China–Pakistan Economic Corridor (CPEC),” 2024.

[2] C. J. Tucker, “Red and photographic infrared linear combinations for monitoring vegetation,” Remote Sensing of Environment, vol. 8, no. 2, pp. 127–150, 1979.

[3] D. P. Roy et al., “Landsat-8: Science and product vision for terrestrial global change research,” Remote Sensing of Environment, vol. 145, pp. 154–172, 2014.

[4] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[5] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, pp. 273–297, 1995.

[6] J. R. Quinlan, “Induction of decision trees,” Machine Learning, vol. 1, no. 1, pp. 81–106, 1986.

[7] M. Pal and P. Mather, “Support vector machines for classification in remote sensing,” International Journal of Remote Sensing, vol. 26, no. 5, pp. 1007–1011, 2005.

[8] M. Belgiu and L. Drăguț, “Random forest in remote sensing: A review of applications and future directions,” ISPRS J. Photogramm. Remote Sens., vol. 114, pp. 24–31, 2016.

[9] V. F. Rodriguez-Galiano et al., “An assessment of the effectiveness of a Random Forest classifier for land-cover classification,” ISPRS J. Photogramm. Remote Sens., vol. 67, pp. 93–104, 2012.

[10] UNEP, World Atlas of Desertification, 2nd ed. London: Arnold, 1997.