NDVI Time Series Analysis and Temporal ForestCut Detection for Land Cover Change Using Google Earth Engine and LSTM Models

Abstract

This report details the development of an application to analyze the NDVI (Normalized Difference Vegetation Index) time series and detect temporal land cover changes, including ForestCut detection, using Google Earth Engine (GEE). The approach is scalable to any location globally, as demonstrated for Indore, India, in this case study. The methodology involves utilizing Sentinel-2 Surface Reflectance data, cloud masking, seasonal analysis, and ForestCut detection based on NDVI change. Additionally, Long Short-Term Memory (LSTM) models are used to predict future NDVI trends based on the CSV data exported from the analysis.

1 Introduction

Deforestation, land cover change, and ForestCut detection are critical challenges that require efficient and scalable monitoring systems. Satellite data, such as from the Sentinel-2 mission, provides valuable insights into land cover dynamics over time. NDVI is a commonly used metric to track vegetation health, which is influenced by various environmental factors, including seasonal changes and anthropogenic activities [?, ?]. This report demonstrates the use of Google Earth Engine (GEE) for spatial analysis of NDVI over any location globally, with a focus on temporal ForestCut detection and land cover classification [?]. Additionally, the report explores the use of Long Short-Term Memory (LSTM) models to predict future NDVI trends based on the CSV data generated from GEE [?, ?].

2 Methodology

2.1 Study Area and Data Acquisition

The region of interest for this case study is Indore, India (latitude 22.7196, longitude 75.8788). However, the methodology is scalable and can be applied to any location globally by adjusting the coordinates. The study period spans from January 1, 2017, to December 31, 2024. Data is sourced from the *Sentinel-2 Surface Reflectance dataset* (COPERNICUS/S2_SR_HARMONIZED) available on Google Earth Engine [?]. This dataset provides cloud-free, atmospherically corrected images, suitable for vegetation monitoring [?].

2.2 Preprocessing and NDVI Calculation

- **Cloud Masking:** Sentinel-2 data includes a cloud mask layer (SCL) used to filter out cloudy pixels. The cloud mask removes clouds, cloud shadows, and cirrus clouds [?].
- **NDVI Calculation:** NDVI is calculated using the formula:

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{RED})}{(\mathrm{NIR} + \mathrm{RED})}$$

  where NIR (Near Infrared) is band B8, and RED is band B4 from the Sentinel-2 imagery [?].

2.3 Seasonal Analysis

The study area undergoes different seasonal cycles, so NDVI values are analyzed by seasons:

- **Winter:** December to February
- **Summer:** March to May
- **Monsoon:** June to September
- **Autumn:** October to November

This seasonal breakdown follows the climate patterns for India, which are well documented in [?].

2.4 Time Series and Seasonal Analysis

A time series of NDVI values from 2017 to 2024 was generated. The data was analyzed by creating both a **time series chart** for NDVI changes and a **seasonal chart** to compare NDVI across different seasons [?].

2.5 ForestCut Detection and Land Cover Change Classification

- **NDVI Change Detection:** By comparing the first and last NDVI values in the time series, areas of deforestation, afforestation, or no change were identified based on predefined NDVI thresholds [?]:
  - Deforestation: NDVI decrease of more than -0.20
  - Afforestation: NDVI increase of more than 0.20
  - No Change: Minimal NDVI variation within ± 0.20
- **ForestCut Detection:** The approach includes detecting significant forest loss (ForestCut) based on NDVI decrease, thus identifying regions affected by deforestation and land use changes [?].
- **Classification:** A land cover change classification image was generated, identifying deforestation, afforestation, and areas with no significant change. The land cover types were visualized with a color palette representing these classes [?].

2.6 Exporting Data

The processed NDVI time series and classified land cover data were exported as CSV files to Google Drive for further analysis, including training a machine learning model [?].

3 Results

3.1 NDVI Time Series

The NDVI time series chart shows the variation in vegetation cover from 2017 to 2024, highlighting seasonal patterns and long-term trends. Peaks typically correspond to the monsoon season when vegetation is most abundant [?].

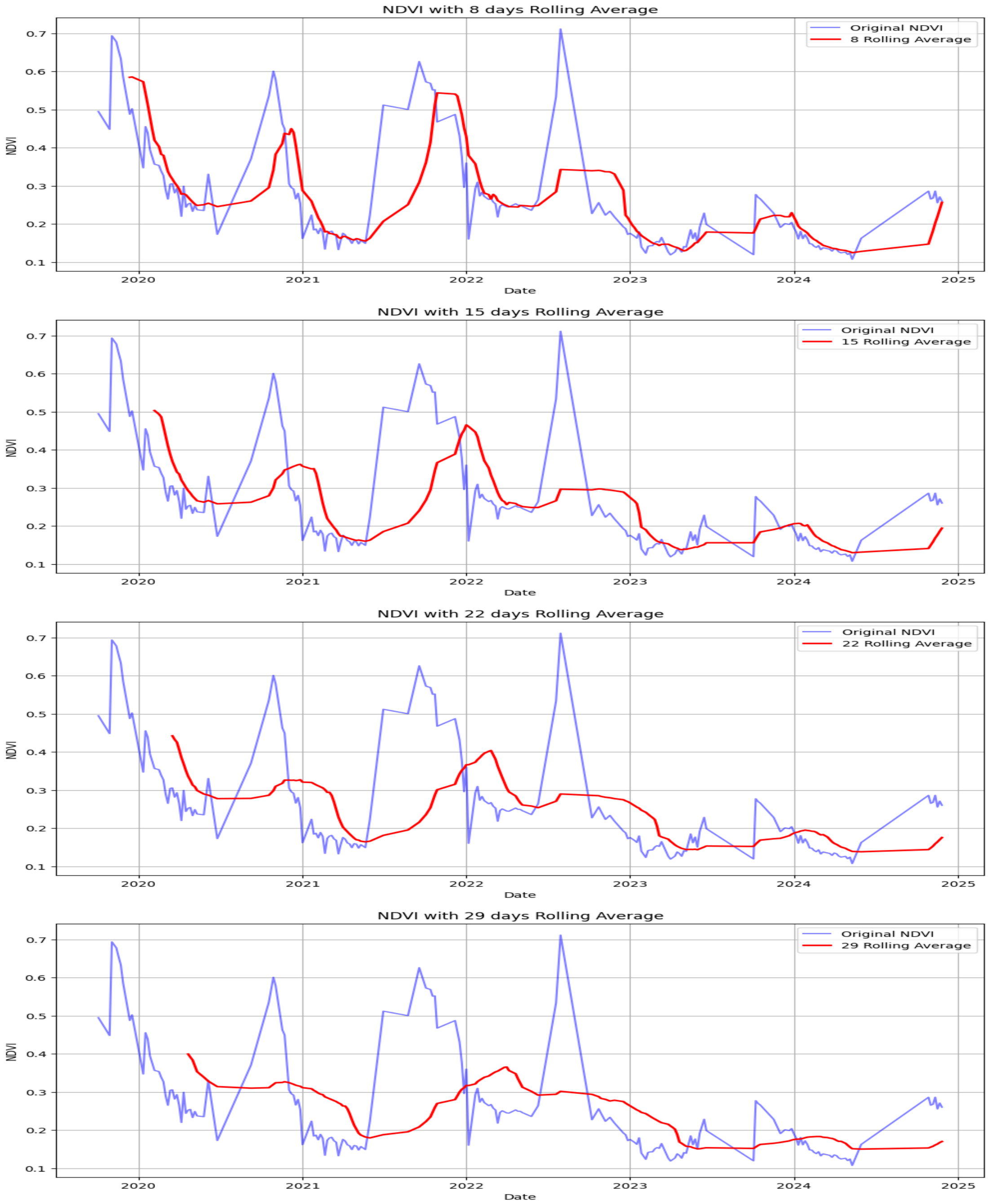


Figure 1: Comparison of rolling averages with window sizes of 8, 16, 22, and 29 days. Rolling averages help smooth noisy data, revealing clearer trends. Among these, the 8-day and 14-day rolling averages perform best, striking a balance between smoothness and fidelity to the data. Larger window sizes tend to over-smooth and obscure finer details in the dataset.

3.2 Seasonal NDVI Distribution

The seasonal analysis chart shows the variation of NDVI across the four seasons, revealing that the highest NDVI values occur during the monsoon season, while the lowest values are observed during winter and summer months [**?**].

3.3 ForestCut and Land Cover Change Classification

The land cover classification based on NDVI change and ForestCut detection identified significant areas of deforestation and afforestation within the study region. Deforestation regions were marked in **red**, afforestation areas in **green**, and stable areas in **gray** [**?**].

3.4 NDVI Change Analysis

The analysis of NDVI changes revealed a decrease in vegetation health (deforestation) in certain regions and an increase (afforestation) in others. The detailed land cover classification provides insights into the dynamics of vegetation cover over time [**?**].

4 Working And Demonstration

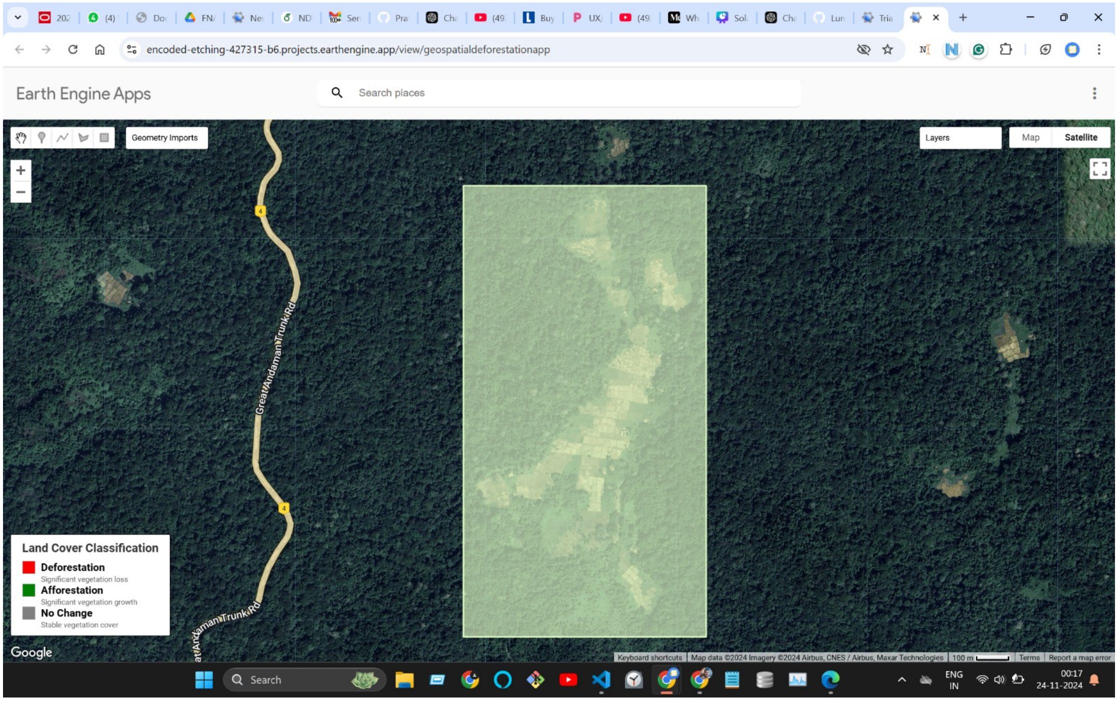


Figure 2: Defining the study area: This step involves selecting the geographic region of interest for analyzing deforestation, afforestation, and stable vegetation zones.

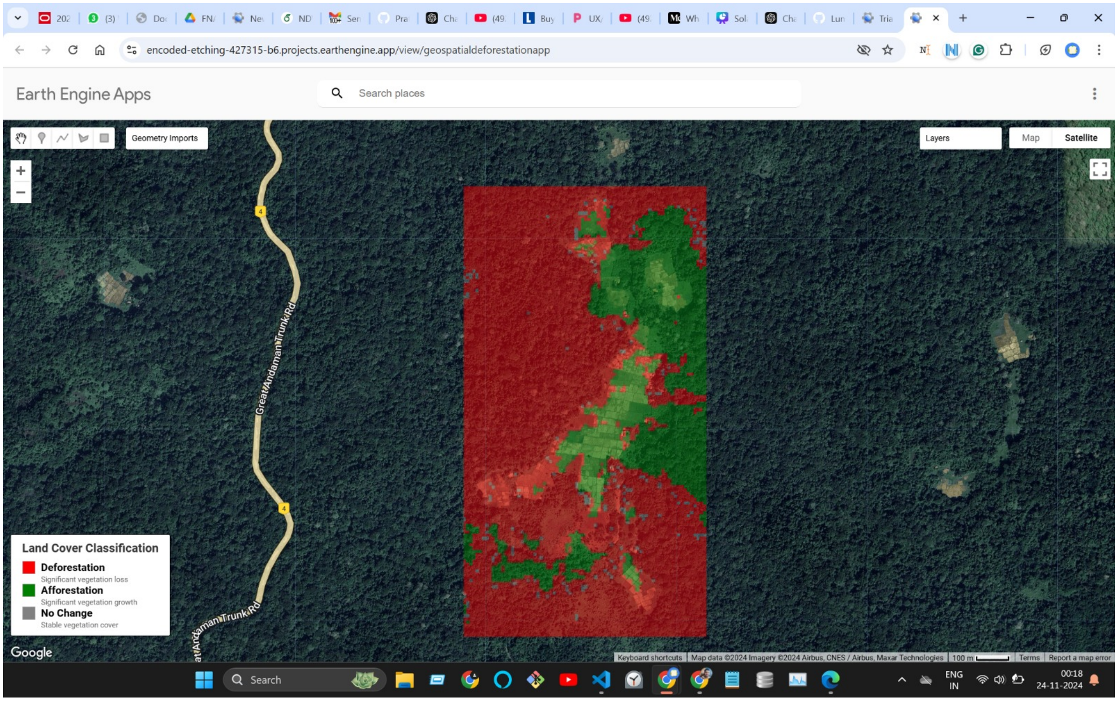


Figure 3: Classification results: A visualization showing areas categorized into deforestation, afforestation, and stable zones based on NDVI change analysis.

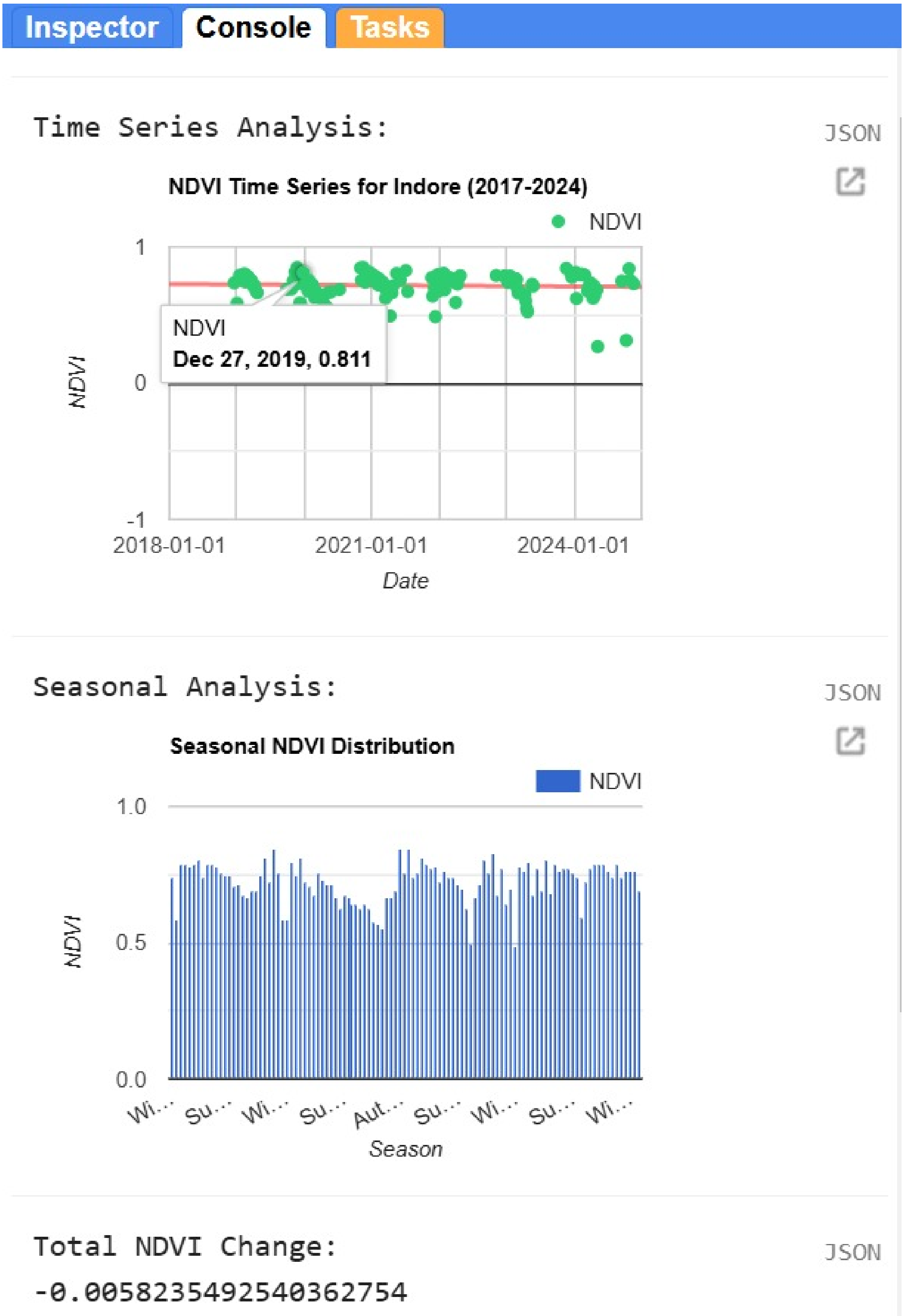


Figure 4: Detailed analysis using graphs and histograms: These visualizations provide insights into NDVI distribution and trends, enhancing the understanding of vegetation health changes.

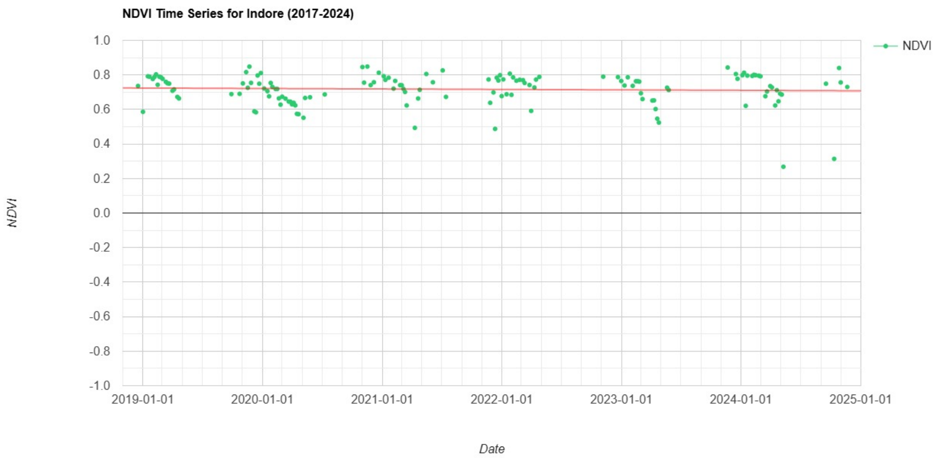


Figure 5: NDVI time series plot: This chart depicts vegetation health trends over the study period, highlighting seasonal variations and long-term changes.

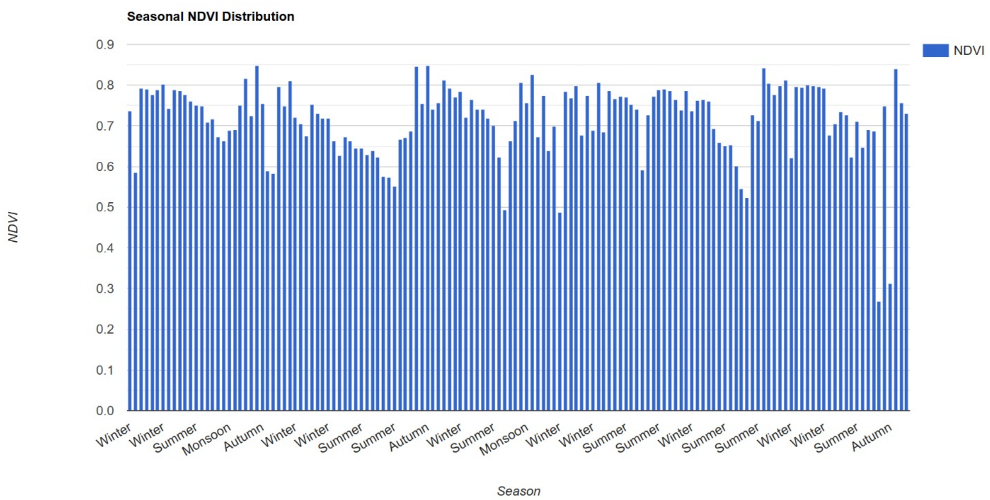


Figure 6: Seasonal NDVI distribution: A comparative analysis of NDVI values across different seasons, illustrating the impact of climatic conditions on vegetation.

5 LSTM Model for Predicting NDVI Trends

5.1 Data Preparation

The CSV data containing NDVI values over time was exported from Google Earth Engine. This data forms the basis for training the LSTM model. LSTM networks are ideal for time series forecasting due to their ability to capture long-term dependencies in sequential data [**?**].

in our case we wanted to analyze how the construction has caused long-term change in vegetation trends. although we have multiple constructions, we focus on the Bheema Hostel Block.

Construction for Bheema hostel began in January 2023. and we can see a clear change in historical NDVI temporal data after that event

The construction of the Bheema Hostel led to significant vegetation clearance, resulting in a notable decrease in NDVI values over time. The deforestation began in November 2019 with the establishment of the college and has continued year to date.

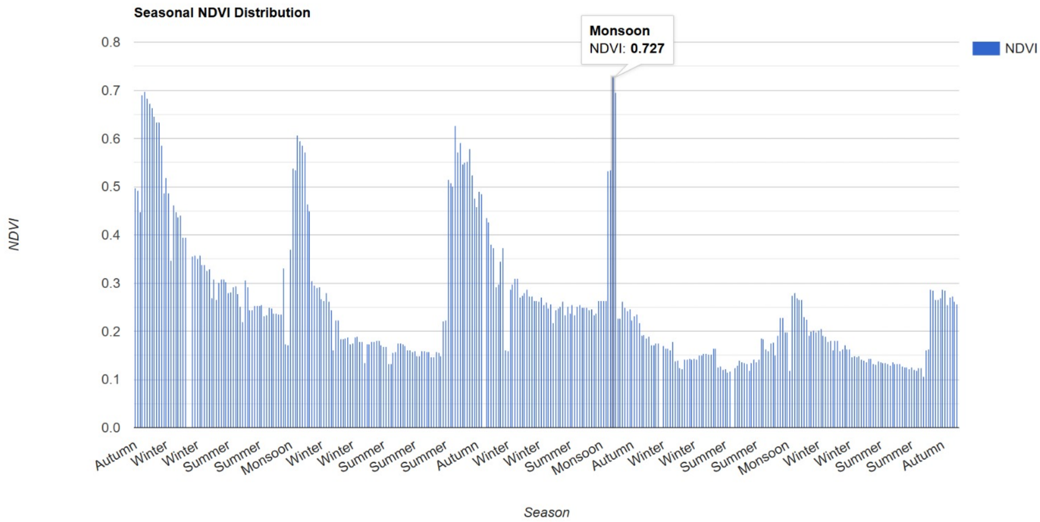
- Peak NDVI before construction: **0.726** (recorded on July 30, 2022, during the monsoon season).
- Construction start date: January 2023.
- Peak NDVI after construction: **0.288** (recorded on July 15, 2023).

This sharp decline in NDVI highlights the substantial impact of vegetation loss caused by the development of the hostel.

We further export this data to drive for analysis as shown in Figure 6.

5.2 LSTM Model Setup

- **Input Features:** NDVI values along with temporal features (e.g., month, season).
- **Target Variable:** NDVI value for the next time step.
- **Model Architecture:** The LSTM network was trained using exported CSV data with a focus on predicting future trends in NDVI [**?**].



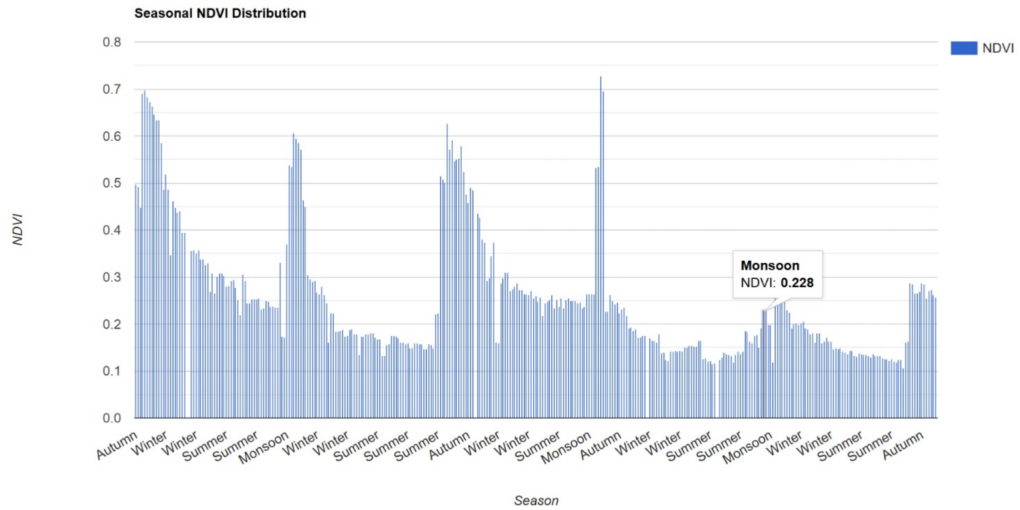


Figure 7: Comparison of NDVI values before and after the construction of Bheema Hostel. The peak NDVI before construction (July 30, 2022) was 0.726, indicating dense vegetation, while after construction (July 15, 2023), it dropped significantly to 0.288 due to vegetation clearance for development.

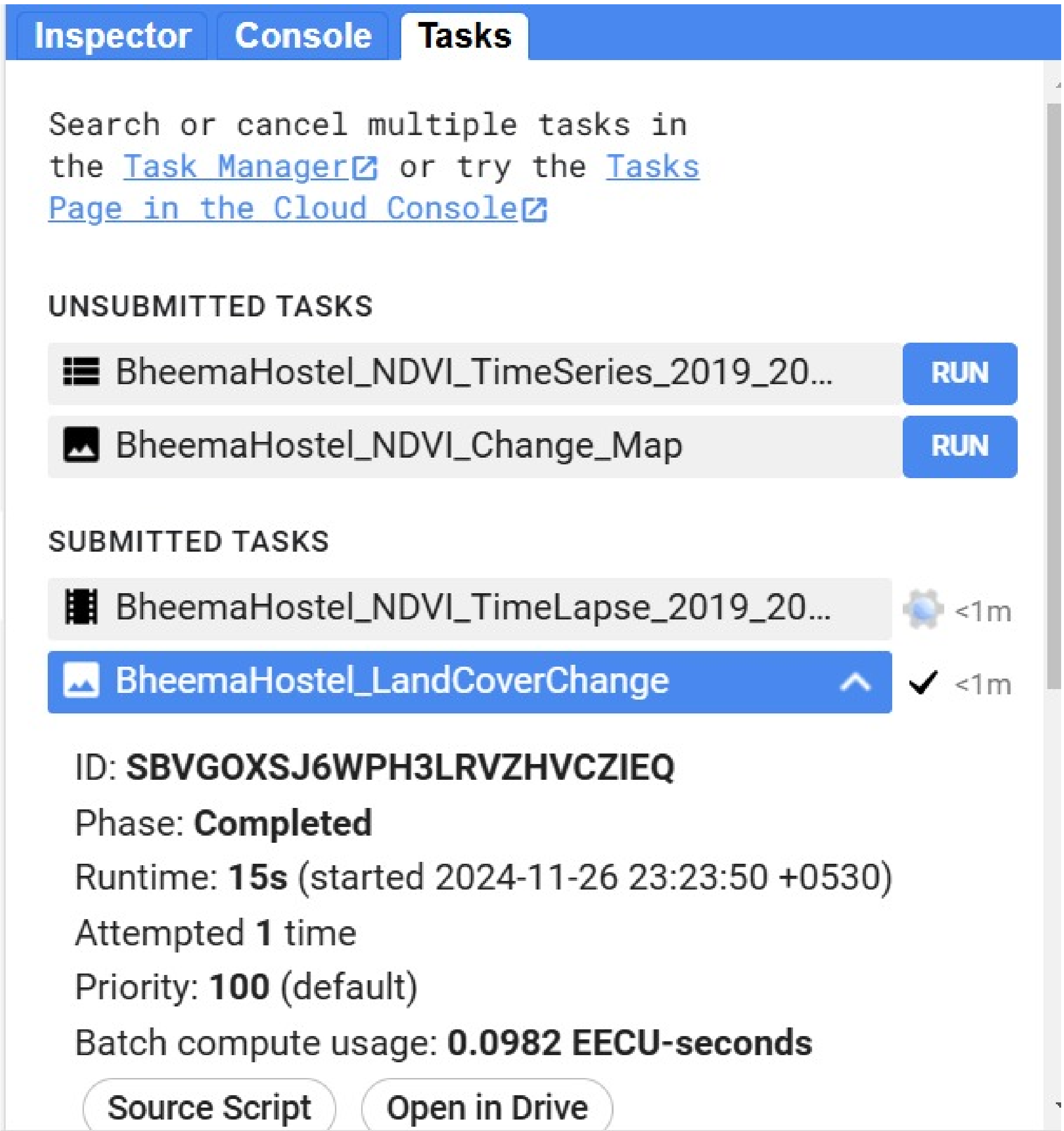


Figure 8: Exporting data using Google Earth Engine: Demonstration of the Export.toDrive function, enabling the export of shapefiles, images, videos, and tabular data for further analysis.

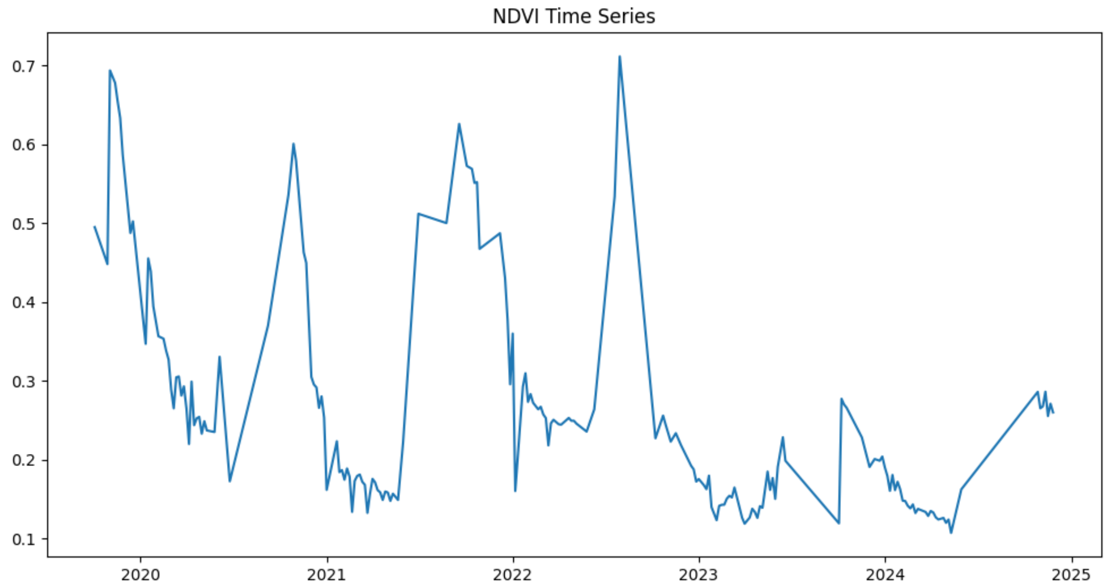


Figure 9: NDVI Values after linear interpolation using inbuilt pandas functions

5.3 Model Evaluation

The LSTM model was evaluated using metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE). The model's accuracy in predicting future NDVI values was assessed by comparing the predicted values with actual NDVI observations [?].

6 Discussion

6.1 Application for Monitoring Deforestation and Forest-Cut

The use of NDVI time series analysis is an effective method for detecting deforestation, afforestation, and ForestCut, and monitoring land cover changes [?]. By leveraging Google Earth Engine's cloud-based platform, large-scale analysis can be performed without the need for local computational resources [?].

6.2 LSTM Model for Prediction

Long Short-Term Memory (LSTM) networks were employed to forecast future vegetation trends based on NDVI time series data. Their ability to retain long-term dependencies and handle sequential data made them an ideal choice for this

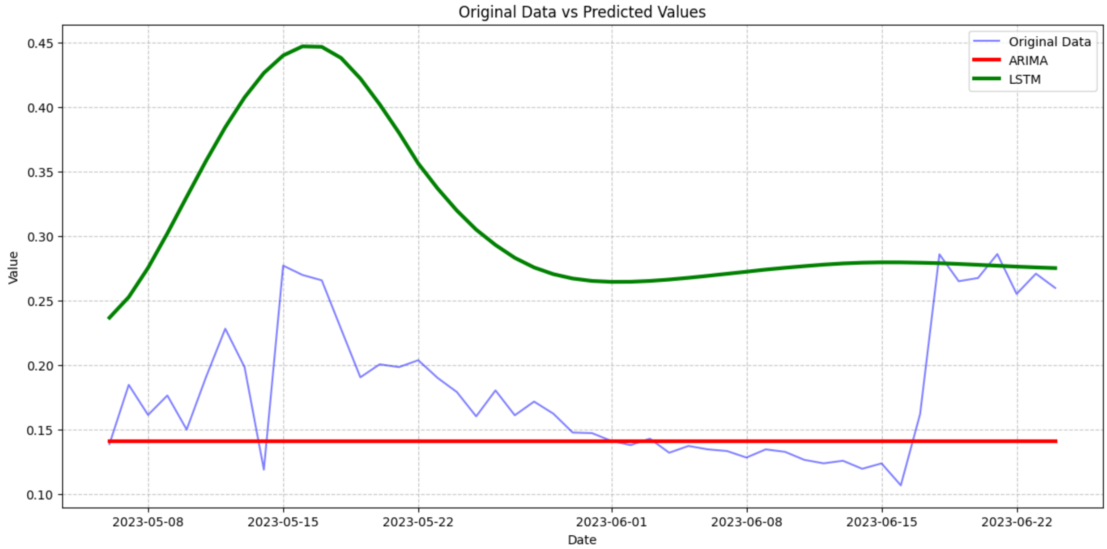


Figure 10: LSTM fits the data and its shape better tham ARIMA across same strech of testing data

application, as NDVI data often exhibit seasonal and temporal variations. This approach supports early detection of changes in vegetation health, enabling proactive interventions for issues such as deforestation and land degradation.

6.3 Design Decisions

Key design choices were made to ensure the methodology balances accuracy, scalability, and practicality:

- **Choice of LSTM over Alternatives:** Unlike simpler models like ARIMA, which are limited to linear trends, or more complex methods like Transformer models, which require significantly larger datasets and computational resources, LSTMs strike a balance between computational efficiency and their ability to capture non-linear, time-dependent patterns in NDVI data.

- **Google Earth Engine for Data Acquisition:** Google Earth Engine was selected for its comprehensive satellite data catalog and powerful computational capabilities, which significantly reduce the overhead associated with traditional methods of downloading, preprocessing, and analyzing imagery. Other tools like QGIS or local GIS platforms, while useful, lack the scalability and ease of cloud-based processing for global applications.

- **Automated Temporal Aggregation and Preprocessing:** To streamline the workflow, automated techniques were used to prepare and clean

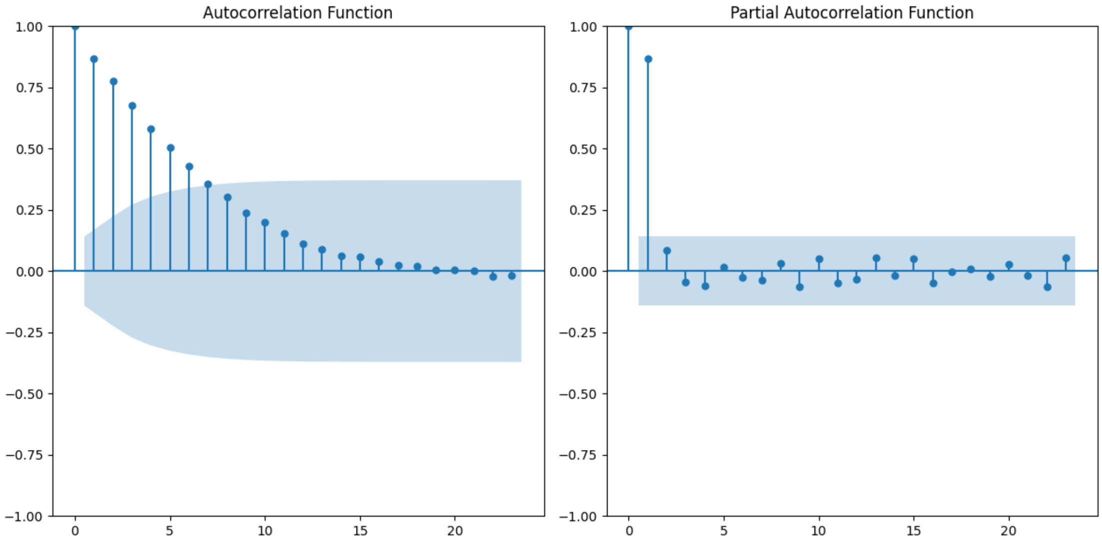


Figure 11: Autocorrelation helps identify the seasonality hyperparameter m for optimal results. A greater distance from the zero line indicates stronger correlation. Partial autocorrelation typically exhibits an exponential decay, with the highest correlations observed at lags 2 and 3 (it is always 1 at lag 0 by default).

time series data. This approach reduces the risk of manual errors compared to fully manual preprocessing, ensuring consistent and reproducible results, particularly for large-scale datasets.

- These choices were made to optimize the process for real-world usability, minimizing computational demands while maintaining robustness across diverse ecological regions and temporal scales.

6.4 Limitations

- **Data Quality and Availability:** Reliance on satellite imagery introduces challenges like missing data due to cloud cover or sensor errors, impacting temporal consistency.

- **Model Generalization:** LSTM models, while powerful, may require fine-tuning to adapt to diverse vegetation types and climatic conditions, potentially limiting their immediate applicability across all ecosystems.

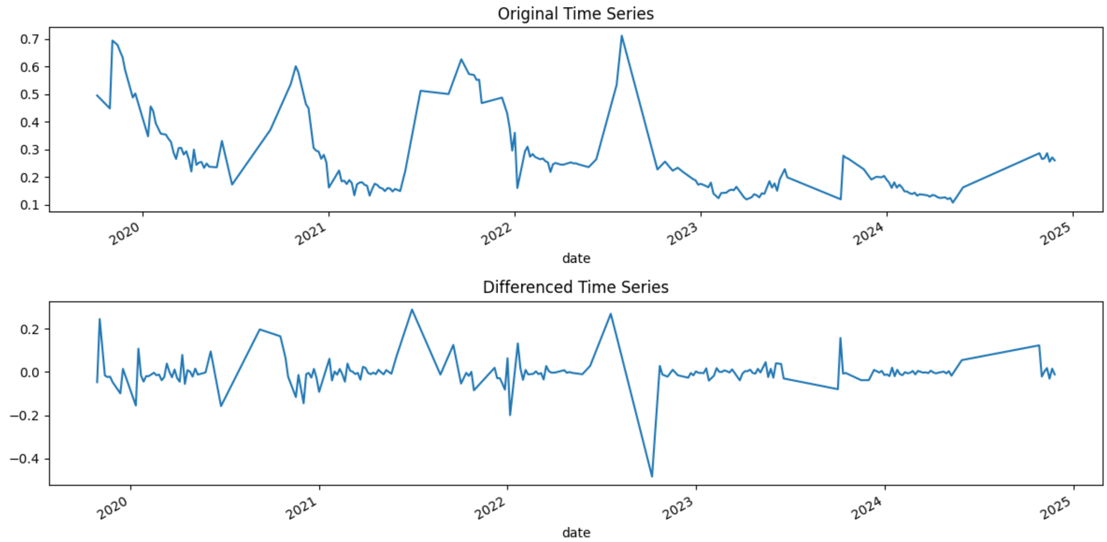


Figure 12: Transformation of the dataset into a stationary time series through differencing: Stationarity ensures that statistical properties like mean, variance, and autocorrelation remain consistent over time, which is crucial for reliable modeling and forecasting in time series analysis

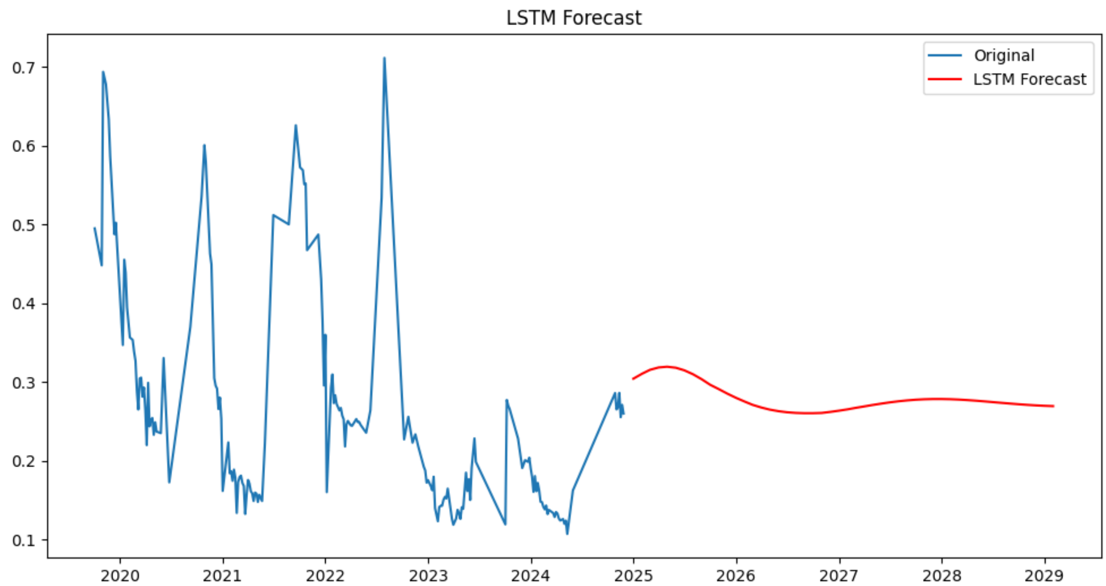


Figure 13: Degradation of forecast accuracy over time: As the prediction horizon extends further from the last observed data point, the uncertainty increases due to accumulating errors and reduced influence of recent trends, leading to diminished reliability in long-term forecasts.

7 Conclusion

This study outlines an efficient, scalable framework for NDVI time series analysis, combining forest loss detection, land cover change classification, and predictive modeling. The strategic use of LSTM networks and Google Earth Engine enhances the system's capacity to monitor vegetation dynamics and predict trends in real-time, providing a robust tool for addressing global environmental challenges.

8 References

- Vermote, E. F., et al. (2016). *MODIS Surface Reflectance (MOD09).* NASA Earth Science Data. https://modis.gsfc.nasa.gov/data/
- Huete, A. R., et al. (1994). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 48(3), 302-313.
- Gorelick, N., et al. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27.
- Drusch, M., et al. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for Land and Water Monitoring. *Remote Sensing of Environment*, 120, 25-36.
- Bastin, J.-F., et al. (2019). The extent of forest in the global land system. *Science*, 366(6463), 904-908.
- Puyravaud, J. P. (2003). Standardizing the calculation of the annual rate of deforestation. *Forest Ecology and Management*, 177(1-3), 307-313.
- Sutskever, I., et al. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27, 3104-3112.
- Zhou, H., et al. (2020). A review on deep learning-based forecasting methods. *Computers, Materials & Continua*, 64(1), 53-66.
- Sezer, M. I., et al. (2020). Predicting stock returns using deep learning models: A systematic review. *Expert Systems with Applications*, 146, 113197.
- Hansen, M. C., et al. (2008). Humid tropical forest disturbance: Drivers, trends, and policy implications. *Ambio*, 37(6), 385-396.
- Lambin, E. F., et al. (2001). The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change*, 11(4), 261-269.