Journal of Water Resources and Geosciences

Vol. 5, No. 2, 2026

Abu Ghraib Vegetation Forecasting: NDVI Analysis Using Google Earth Engine Landsat 8 Data and MATLAB-Based Machine Learning

Mustafa Ismael Abdulkadhim*

Ministry of Water Resources- Planning and Follow up Department, Baghdad- Iraq

*Corresponding author's email: mustafa9r07@gmail.com

Abstract

This study investigates vegetation dynamics and forecasts future vegetation conditions in Abu Ghraib, Iraq, using a monthly Normalized Difference Vegetation Index (NDVI) time series derived from Landsat 8 imagery through the Google Earth Engine (GEE) platform. The study area represents an irrigated, semi-arid agricultural environment affected by seasonal water availability, climatic variability, and land-use pressure. Landsat 8 data from January 2015 to December 2023 were processed to generate monthly NDVI values using spatial and temporal filtering and cloud masking. The resulting time series was used as input for a Random Forest Regression model implemented in MATLAB. A 12-month sliding-window approach transformed the NDVI sequence into supervised learning samples, enabling the model to capture annual seasonality and temporal dependence. The trained model forecast NDVI trends over a five-year horizon. The model produced low absolute prediction errors, with a test RMSE of 0.0149, MAE of 0.0126, and MAPE of 18.5%. However, the negative test R^2 value of −0.35 indicates limited ability to explain variance in unseen data, despite capturing the general seasonal NDVI pattern. This limitation may be attributed to the relatively short time series and environmental variability. Residual analysis, out-of-bag error behavior, autocorrelation assessment, and feature-importance ranking supported the model's ability to reproduce the dominant temporal structure. The forecasted outputs can support agricultural planning, irrigation monitoring, environmental assessment, and sustainable resource management in semi-arid irrigated regions.

Keywords: NDVI, Google Earth Engine (GEE), Landsat 8 Satellite Data, Vegetation Forecasting, Random Forest Regression.

Journal of Water Resources and Geosciences

Vol. 5, No. 2, 2026

التنبؤ طويل الأمد بمؤشر الاختلاف الخضري (NDVI) في منطقة مروية شبه قاحلة باستخدام الغابات العشوائية (Random Forest): دراسة حالة لقضاء أبي غريب

مصطفى إسماعيل عبد الكاظم*

وزارة الموارد المائية/ دائرة التخطيط والمتابعة، بغداد - العراق

*البريد الالكتروني للمؤلف المراسل: mustafa9r07@gmail.com

الخلاصة

تبحث هذه الدراسة في ديناميكيات الغطاء النباتي وتتنبأ بالظروف المستقبلية للغطاء النباتي في أبو غريب، العراق، باستخدام سلسلة زمنية شهرية لمؤشر الاختلاف النباتي الطبيعي (NDVI) مشتقة من صور الأقمار الصناعية Landsat 8 عبر منصة Google Earth Engine (GEE). وتمثل منطقة الدراسة بيئة زراعية مروية وشبه جافة تتأثر بتوافر المياه الموسمي، والتغيرات المناخية، والضغوط الناجمة عن استخدامات الأراضي. جرت معالجة بيانات Landsat 8 للفترة من كانون الثاني 2015 إلى كانون الأول 2023 لإنتاج قيم شهرية متسقة لمؤشر NDVI باستخدام الترشيح المكاني والزمني وإجراءات إزالة تأثير الغيوم. واستُخدمت السلسلة الزمنية الناتجة كمدخل لنموذج انحدار الغابة العشوائية (Random Forest Regression) المنفذ في برنامج MATLAB. كما استُخدمت نافذة منزلقة مدتها 12 شهراً لتحويل سلسلة NDVI إلى عينات تعلم خاضعة للإشراف، مما مكّن النموذج من التقاط التغيرات الموسمية السنوية والاعتماد الزمني. واستُخدم النموذج المدرب للتنبؤ باتجاهات NDVI على مدى خمس سنوات. أظهر النموذج أخطاء تنبؤ مطلقة منخفضة، إذ بلغ RMSE للاختبار 0.0149، و MAE مقدار 0.0126، و MAPE نسبة 18.5%. ومع ذلك، تشير قيمة R^2 السالبة للاختبار والبالغة −0.35 إلى قدرة محدودة على تفسير التباين في البيانات غير المرئية، رغم نجاح النموذج في التقاط النمط الموسمي العام لمؤشر NDVI. وقد يُعزى هذا القيد إلى قصر السلسلة الزمنية والتباين البيئي. كما دعمت تحليلات البواقي، وسلوك خطأ خارج العينة، وتقييم الارتباط الذاتي، وترتيب أهمية الخصائص قدرة النموذج على إعادة إنتاج البنية الزمنية السائدة. ويمكن أن تدعم المخرجات المتنبأ بها التخطيط الزراعي، ومراقبة مشاريع الري، والتقييم البيئي، والإدارة المستدامة للموارد في أبو غريب والمناطق المروية وشبه الجافة المماثلة، وتوفير معلومات مهمة لصنع القرار وتحسين كفاءة إدارة الموارد الطبيعية مستقبلاً المحلية.

الكلمات المفتاحية: مؤشر الاختلاف الخضري (NDVI)، محرك جوجل إيرث، مراقبة الغطاء النباتي، التنبؤ بالسلاسل الزمنية، انحدار الغابة العشوائية.

1. Introduction

Vegetation is crucial in ecological balance, biodiversity sustenance, and land use for agricultural purposes in a sustainable manner. Vegetation health monitoring and its change detection over time are crucial for environmental management and effective planning. Since NDVI data is derived from satellite imagery, it becomes a standard metric that would describe vegetation health and density (Istanbuly & Thabeet, 2020). The NDVI will be very useful in supplying details about the condition of vegetation in most areas for extended periods of time, which will enable researchers to conduct analyses regarding trends, seasonal variability, and the impact of environmental factors. This work takes into consideration the Abu Ghraib region as a potential agricultural area for the purpose of vegetation dynamics analysis and forecast (Istanbuly & Thabeet, 2020). The time series NDVI data from 2015 to the end of 2023 over the area were collected. Data were gathered from the GEE platform, derived from the Landsat 8 satellite images. These are the only years of NDVI data provided by the satellite that capture critical historical vegetation patterns and seasonal trends in Abu Ghraib. Future NDVI values were predicted through machine learning techniques implementing the Random Forest Regression algorithm in MATLAB. The model received historical NDVI data through a sliding window method to analyze trends in vegetation forecasting the future. Researchers produced a five-year forecasting model which provided important predictions about vegetation health modifications (Adisti & Sunkar, 2021). These findings from the study can therefore be used in agricultural and ecological planning for Abu Ghraib. Integrating strong data processing via GEE with machine learning tools in MATLAB, the Random Forest Regression algorithm will assure high accuracy and comprehensiveness of vegetation dynamics analysis. The current research has expressed how the application of satellite-derived data, complemented with progressive analytical techniques, could tackle critical environmental problems (Adisti & Sunkar, 2021).

Various research works have been focused on the application of NDVI time series analysis and machine learning methods in satellite image-based vegetation monitoring and its forecast. (Hameed & George, 2023), presented the time series analysis of vegetation indices using satellite imagery from Landsat 8 to delineate evergreen and seasonal plants in the Abu Ghraib area, which also proved the effectiveness of NDVI in monitoring vegetation dynamics in that particular region. An evaluation by (Htitiou et al.,

2019) assessed phenological metrics derived from Sentinel-2A and Landsat 8 NDVI time series as tools to identify multiple crop classes within Moroccan irrigated areas through Random Forest classification processes. These results demonstrated how satellite imagery operating at medium to high spatial resolutions detects the distinctive growth stages of various agricultural systems. (Li et al., 2021) developed a climate-included gap-filling method in reconstructing the historical Landsat NDVI time series data. This approach enhances the quality of the NDVI time series by resolving the problem of missing data; this is very important for the accurate monitoring and analysis of vegetation. (Vasilakos et al., 2022) used the Convolutional LSTM model for high-resolution vegetation greenness forecasting over Africa. The model successfully predicted seasonal NDVI variations and the impacts of weather anomalies by using Sentinel-2 NDVI data, ERA5 weather reanalysis, SMAP satellite measurements, and topographic data (DEM of SRTMv4.1). This study has proved that ConvLSTM models are capable of predicting vegetation responses even in regions with highly variable NDVI, thus providing very useful information in the management of drought-related disasters. (Faran et al., 2025) fine-tuned the pre-trained FourCastNet model in order to simulate vegetation activity, studying how global atmospheric representations can be transferred for the prediction of the normalized difference vegetation index (NDVI). This thus provides insight into how weather transformers can be leveraged for global vegetation modeling. It will contribute toward improving the continuity and reliability of remote-sensing data and thus to continuous Earth surface monitoring (Adisti & Sunkar, 2021), (Zhao, F., et al., 2020).

These recent publications demonstrate how satellite image fusion with machine learning models supports better analysis and prediction of vegetation dynamics.

NDVI is important in remote sensing applications because the following reasons:

1. Vegetation Health Monitoring.
2. Agricultural Applications.
3. Drought and Water Stress.
4. Land Use and Land Cover Changes.
5. Climate Change Studies.

6. Biodiversity Conservation.

7. Ecosystem Services Evaluation.

2. Study Area

Abu-Gharib District is located in central Iraq, to the west of Baghdad City, and represents an important area for investigating vegetation dynamics and forecasting due to its heterogeneous land-cover characteristics, which include agricultural lands, urban expansion, and semi-arid environments. As illustrated in Figure (1), the administrative boundary of Abu-Gharib is shown in red, while the selected study area is highlighted in yellow within the northeastern part of the district. Geographically, the selected study area extends approximately between UTM Eastings 412,986 and 418,922 mE, and Northings 3,688,285 and 3,694,157 mN (Zone 38S). This area is influenced by its proximity to Baghdad urban zones, surrounding agricultural fields, and the Euphrates River system toward the western and southwestern parts. The inset map further demonstrates the location of Abu-Gharib within Iraq, emphasizing its spatial position in the central region of the country. Therefore, Abu-Gharib provides a suitable case study for assessing vegetation variability and forecasting changes under mixed environmental and climatic conditions.

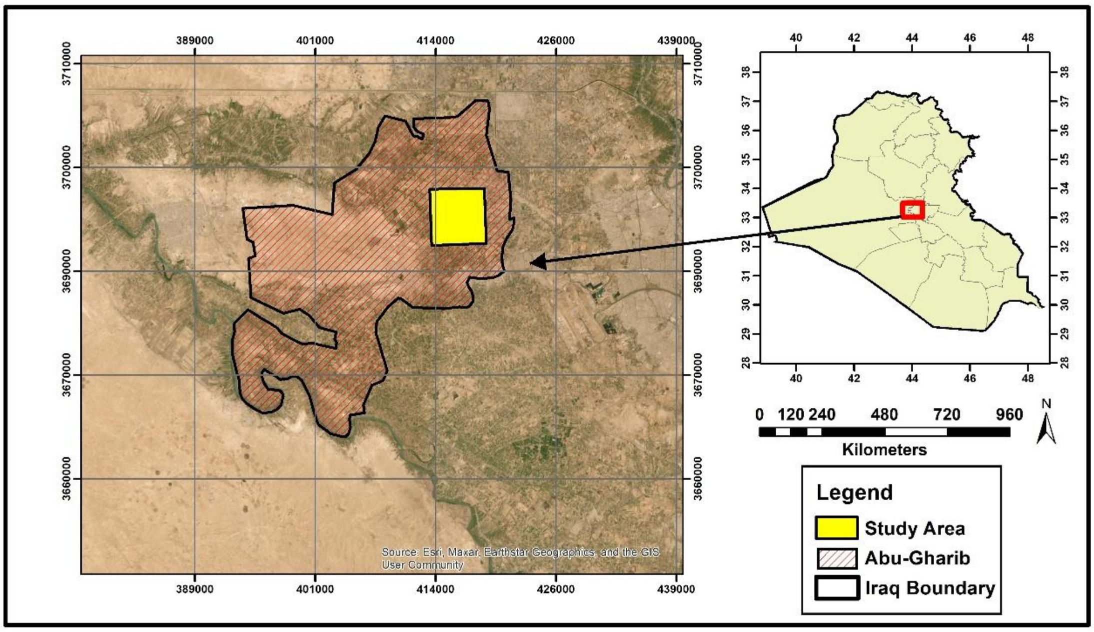

Figure 1: Abu- Ghraib location on Iraq's map.

It includes very extreme weather fluctuations-for instance, prolonged periods of drought and seasonal water-level fluctuations it suitable for testing the influences on vegetation health and productivity. Located in Abu Ghraib where about 80% of the land (180,800 donums) is agricultural the specific study area covers 32.8 square kilometers and is heavily dependent on farming and water resources. The study of vegetation evolution will help with better resource management and agricultural sustainability practices. The mixed desert terrain and landscaped areas combined with water elements in Abu Ghraib present an ideal testing ground for analyzing vegetation trends through time-series NDVI measurements. The present research work focuses on understanding the seasonal and long-term trends in vegetation, together with the effects of fluctuations in climate, thereby offering useful insights into how environmental changes impact both the ecosystem and the local communities. In short, Abu Ghraib will provide an appropriate case study for vegetation dynamic analysis, forecasting future trends, and determining the implications for agricultural and natural resource management with respect to climatic and environmental changes.

3. Normalized Difference Vegetation Index (NDVI)

It is an indicator used to estimate the density of vegetative plants on lands and plants, it is considered as an important indicator in environmental sciences, agriculture, and remote sensing. NDVI is calculated using visible and near-infrared radiation data, and gives an idea of the amount of chlorophyll present in plants. The index of natural change in vegetation cover is one of the most widely used natural indicators in the field of analyzing satellite images and studying vegetation cover, fires, desertification, landslides and other natural phenomena. It is also a means of studying the changes that occur in vegetation cover over time. It also gives us the health status of the plant and the value of vegetation cover in any area and the rate of crop success or failure (Adisti & Sunkar, 2021).

The NDVI index is a useful method for monitoring plants where, NDVI values generally range between −1 and +1. Negative values usually indicate water bodies, clouds, or non-vegetated wet surfaces, while values close to zero often represent bare soil, built-up areas, or very sparse vegetation. Low positive values, approximately from 0.1 to 0.3, indicate weak or sparse vegetation cover. Moderate values, approximately from 0.3 to 0.6, represent medium vegetation density. Higher values, from about 0.6 to 1.0, generally indicate dense and healthy vegetation cover. Therefore, NDVI was adopted in this research to monitor temporal vegetation changes and to provide a quantitative basis for forecasting vegetation dynamics in the study area (Abdulkadhim et al., 2025). It relies on an equation based on the relationship between near infrared rays (NIR) and visible red rays (R). This relationship is due to the high plant reflectivity in the short infrared range and the low plant reflectivity in the short infrared range. Visible red rays: the third band represents red rays with a wavelength ranging from 0.63- 0.60 micrometers, through which one can distinguish between bare and green areas, while the fourth band represents short infrared rays with a wavelength ranging from 0.90- 0.76 micrometers, through which can monitor density and distribution of vegetation and to distinguish between plants, soil and water (Istanbuly & Thabeet, 2020). And it is found by this relation:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad (1)$$

4. Materials and Methods

4.1 Data Collection

This paper utilized NDVI time series data sourced directly from Landsat 8 satellite imagery that is available through Google Earth Engine-the world's most potent cloud-based geospatial analysis solution. This study obtained data from 2015 through the end of 2023 to permit detailed investigation of vegetation patterns throughout the Abu Ghraib area. The Landsat 8 has 30 meters spatial resolution that ensures broad monitoring of the changes within vegetation cover in the study area. Preprocessing for quality and reliability was done. This includes cloud masking to exclude the pixels resulting from clouds and cloud shadows that may distort the values of NDVI. Besides, there are also some areas that require resampling to maintain consistency within the time series. These steps ensure the dataset is further refined to allow for the actual analysis and thus reliable forecasting of vegetation trends in Abu Ghraib. Figure (2) illustrate the Flow chart for extraction code of NDVI time series using GEE.

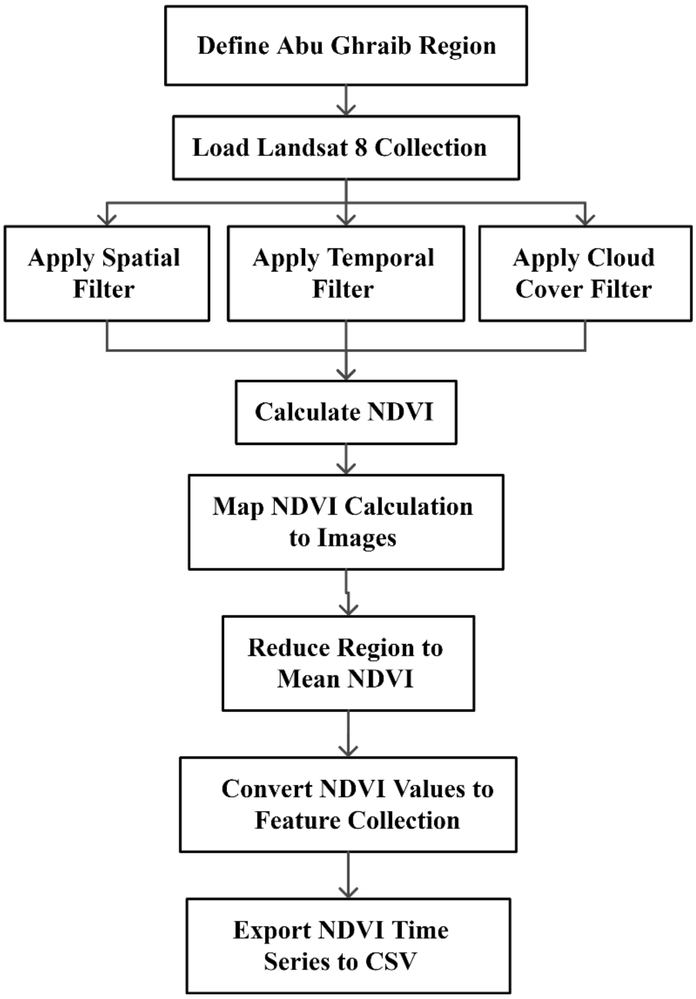

Figure 2: Steps of extracting code of NDVI Time Series Using GEE.

4.2 Machine Learning Approach

MATLAB is a high-level technical computing environment widely used for numerical analysis, data processing, visualization, and machine learning applications. In this study, MATLAB was employed to construct and implement the Random Forest Regression model for NDVI forecasting. The main functions used in the model development included TreeBagger for creating and training the Random Forest regression model, predict for estimating future NDVI values, and oobError for evaluating the out-of-bag prediction error during model training. These functions enabled the transformation of the NDVI time-series data into a supervised learning structure and supported the evaluation of the model's forecasting performance (Al-Ahealy et al., 2024).

Using historical data MATLAB implements the Random Forest Regression model to analyze data and create NDVI value predictions. Random Forest represents the preferred ensemble learning approach which develops various decision trees to achieve improved prediction accuracy and increased performance robustness. The proficient scenario of its processing-non-linear relationships together with complex feature interactions-made it a perfect tool for NDVI time-series modeling due to such regularities in the data as nonlinear trends and seasonal diversity.

One of the main steps of the data preprocessing was represented by the Sliding Window Approach, which transformed NDVI time series into a format suitable for the supervised learning procedure. This created input-output pairs according to the:

- Input (Features): The analysis incorporated the NDVI values from a 12-month consecutive window to serve as these explanatory features.

- Output (Target): The purposeful variable selection used the NDVI measurement of the subsequent month in target classification.

Through this approach the model acquired better understanding of the seasonal data patterns which enabled accurate predictions of future NDVI output.

To train and evaluate the model, the dataset was split into two subsets:

- Training Set (80%): The Random Forest model received training through this dataset for parameter optimization.
- Testing Set (20%): The testing subset functioned to evaluate model predictions for new NDVI values while confirming its capacity to adapt to future inputs.

To prevent model overfitting while improving predictions a Random Forest model used 500 decision trees in its configuration. The model had its hyperparameters adjusted to optimize both the number of predictors sampled per split and the calculation of out-of-bag (OOB) error.

During training the model processed the relationships between NDVI feature values and their sequential NDVI target results. Multiple decision trees generated combined predictions to achieve highly accurate results that help reduce misinterpretations. This training tracked out-of-bag error to stop excessive model fitting which occurs when a model learns features from training data. Figure (3) illustrate Machine Learning code flowchart.

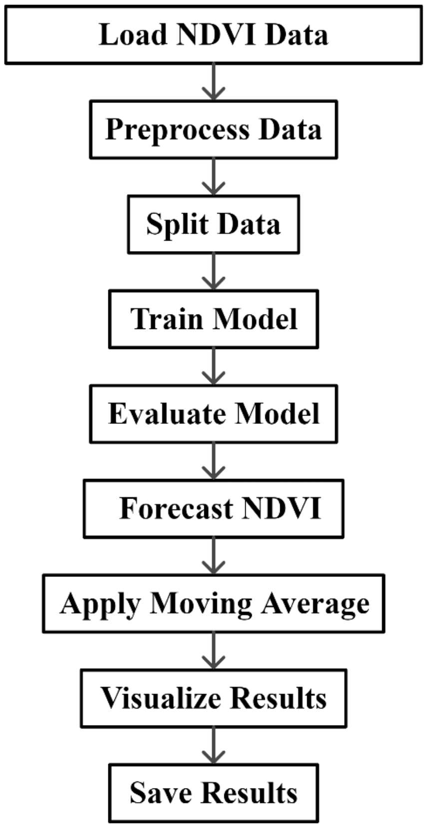

Figure 3: Flowchart for Machine Learning code.

5. Results and discussions

A visual presentation showing NDVI historical and forecasted values for Abu Ghraib appears in Figure (4). The blue line shows historical NDVI patterns which GEE reveals from 2015 through the end of 2023 based on their NDVI time series data. It plots the forecasted NDVI for the next five years with a red dashed line, smoothed using a 3-month moving average. It clearly reflects the vegetation cycle, induced by the climatic variability, seasonality within the historical record. Similar seasonal behavior is observable with the forecasted values of NDVI. It could well imply that these are rather well-captured and extrapolated by the model. This smooth forecast shows better the long-term tendencies and damps short-time noise; this makes easier the interpretation of the vegetation dynamics. This will go a long way in offering substantial insights into agricultural and ecological planning in this area. Accuracy along with reliability of projected results demonstrates their correspondence to the regional historical patterns of vegetation.

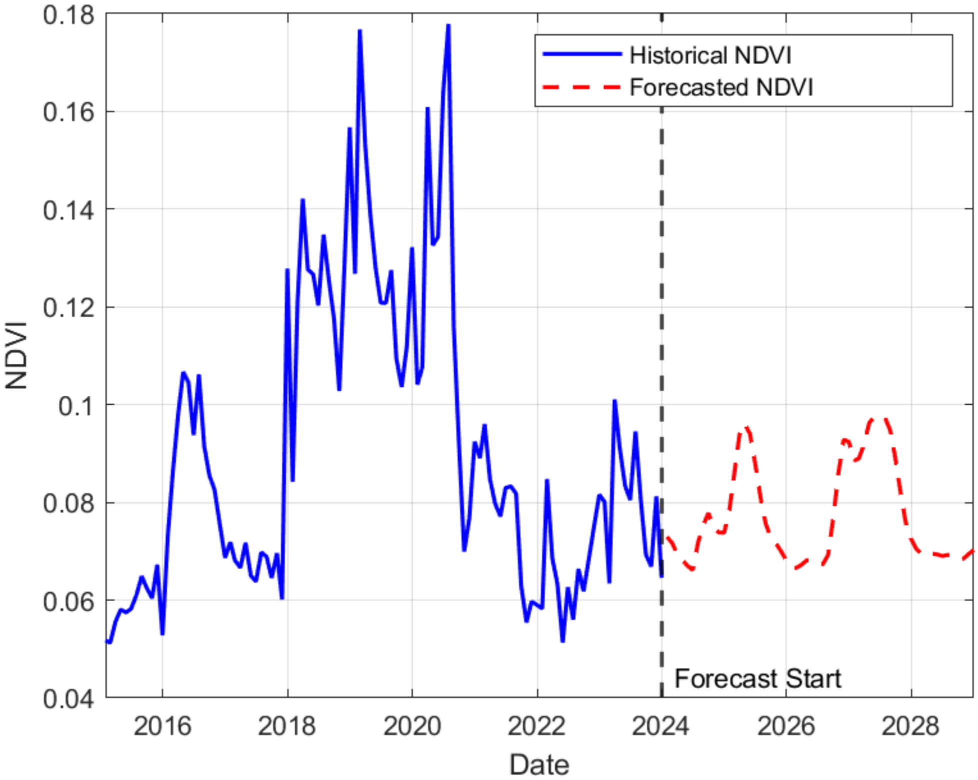

Figure 4: Historical and Forecasted NDVI.

Figure (5) is the OOB error as a function of the number of trees for the Random Forest model. The rapid reduction of the error for the first few iterations before leveling off around 200 trees testifies to the efficiency of the model for learning from data. Low OOB error at 500 trees is an indicator of how strong the generalization capability of the ensembled approach is, and this is expected to ensure good, reliable predictions without overfitting. The strength of this result comes from how it validates the model's ability to realize high accuracy without losing computational efficiency. The OOB error provides an internally built-in metric of cross-validation in itself and offers a truly unbiased estimate of its performance. This aspect is very important for any time series forecasting, as minimum error gives confidence that the prediction can be used in real-world applications.

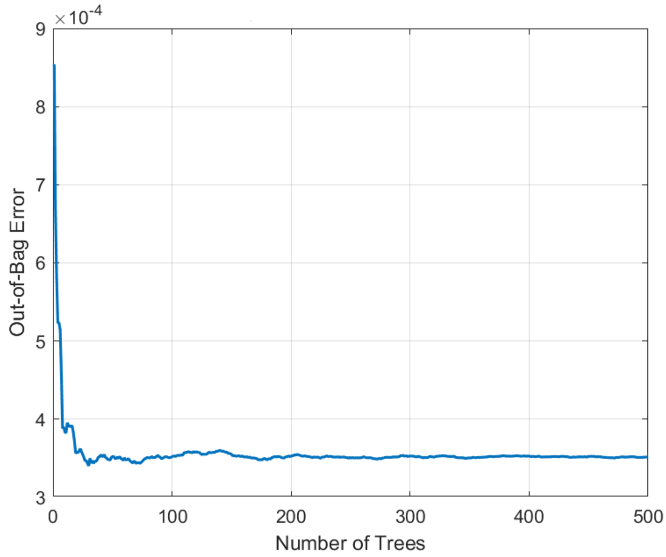

Figure 5: Out of bag error (OOB) over iterations.

Figure (6) is QQ plot showing that the residuals are normally distributed. Most of the points in the QQ plot lie close to the diagonal line, which suggests that the model errors are random and unbiassed-a signature of a strong predictive model. The strength of this result is that it confirms the model does not systematically overestimate or underestimate NDVI values. This is an important aspect in any forecasting, as the accuracy and fairness of the predictions have to do with decision-making. The minimal deviation from normality further supports the adaptability of the model across changing vegetation conditions, hence a versatile tool for various ecological and agricultural applications.

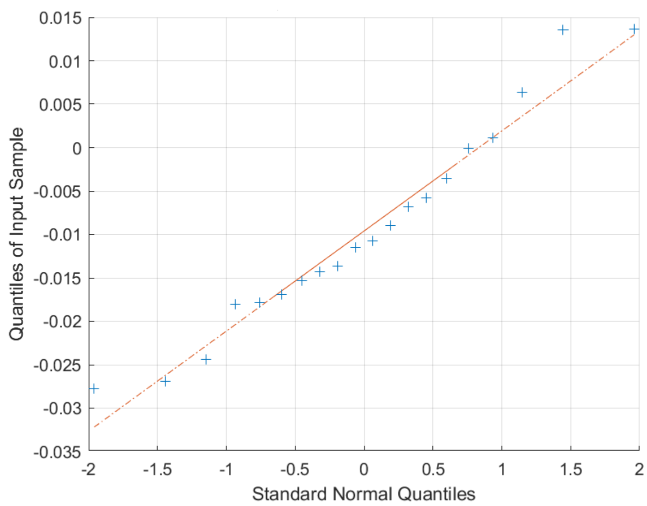

Figure 6: QQ plot of residuals.

The residuals' autocorrelation analysis in Figure (7) demonstrates that most data points remain within the established confidence boundaries. The model shows effective temporal dependency capture because residuals remain independent throughout time. The model's ability to replicate the time-lagged effects in NDVI dynamics becomes validated through this outcome. The removal of autocorrelation in residuals by the model ensures that its predictions are not biased by unmodeled patterns and hence are highly reliable for long-term forecasting. This capability is particularly critical in understanding and managing seasonal vegetation cycles and their responses to environmental changes. The model validation demonstrates its strength to incorporate natural time delays which affect NDVI dynamics. The model eliminates autocorrelation from residuals which leads to dependable long-term predictions that are free from unmodeled pattern effects. Long-term vegetation cycle analysis and environmental change monitoring heavily rely on this capability.

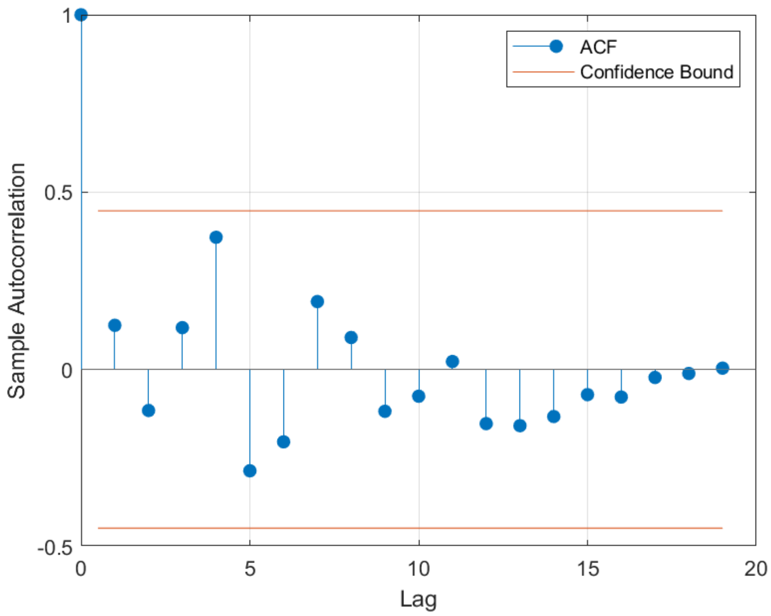

Figure 7: Autocorrelation of residuals.

The combined analysis of all figures demonstrates that Random Forest Regression provides robust NDVI value prediction capabilities. The approach demonstrates robustness because it delivers accurate season trend detection with low prediction errors and unbiased results. The predicted NDVI patterns serve as practical data for agricultural planning alongside environmental surveillance and climate change adaptation strategies.

This research serves as proof that machine learning methods using historical NDVI can leverage the development of dependable and easy-to-understand forecasts. These are capabilities which the Abu Ghraib region requires in sustainable management and proactive decision-making in light of livelihoods and ecosystems being affected by environmental variability.

NDVI time series data and the forecasted data download links (uploaded on Zenodo) are available in the end of the paper.

5. Conclusion

Research investigation demonstrated how to perform vegetation dynamics analysis and forecasting in the Abu Ghraib region using Random Forest Regression from machine learning. The model extracted seasonal patterns as well as long-term vegetation trends by analyzing NDVI time series data from Landsat 8 which underwent processing through GEE. Feature extraction with a sliding window approach and an extensive evaluation framework enabled the successful delivery of accurate NDVI forecasting over a five-year horizon.

Results have pointed out low-error and unbiased predictions of the model, further supported by residual analysis, feature importance rankings, and temporal dependency evaluation. Forecasted NDVI trends were also found to follow the historical pattern; hence, this confirms the reliability and robustness of the model. Besides, the applied smoothing techniques allowed for a more interpretable version of the predictions, hence making the insight valuable for agricultural planning and ecological management.

This work signals a high potential for integrated satellite imagery and machine learning in addressing critical environmental challenges that include vegetation response to climate variability and the proper management of natural resources. The possible future work could extend this study by considering more climatic variables or advanced deep learning models to further refine vegetation forecasting.

References

Abdulkadhim, M. I., Qader, R. A., & Ismael, S. I. H. (2025). Temporal analysis of remote sensing indices with MATLAB: Comparative study of 2016 vs. 2019 vegetation cover index in Abu Ghraib's irrigation projects. Journal of Water Resources and Geosciences, 4(2), 52–78. https://jwrg.gov.iq/index.php/jwrg/article/view/148

Adisti, H. P., & Sunkar, A. (2021). Normalized Difference Vegetation Index (NDVI) analysis for vegetation cover in the Leuser Ecosystem area, Sumatra, Indonesia. Biodiversitas, 5.

Al-Ahealy, M. W., Al-Saedi, Z. J., Hussein, Q. F., & Hussien, H. H. (2024). Monitoring the growth of agricultural crops phenology using Google Earth Engine in Wasit Governorate/Central Iraq. Journal of Water Resources and Geosciences, 3(1), 116–132. https://jwrg.gov.iq/index.php/jwrg/article/view/86

Hameed, R. S., & George, L. E. (2023). Analysis of time series of vegetation indices using Landsat-8 satellite imagery. AIP Conference Proceedings, 2769(1), 020046.

Htitiou, A., El Harti, A., Hakdaoui, M., & Sitor, H. (2019). The performance of random forest classification based on phenological metrics derived from Sentinel-2 and Landsat-8 to map crop cover in an irrigated semi-arid region. International Journal of Remote Sensing, 40(5), 1906–1929.

Isioye, O., & Akomolafe, E. (2020). Accuracy analysis of Sentinel-2A and Landsat-8 OLI satellite datasets over Kano State, Nigeria, using vegetation spectral indices. In Proceedings of the GeoInformation for Disaster Management Conference (p. 3). Sydney, Australia.

Istanbuly, M. H., & Thabeet, A. (2020). Studying changes in normalized difference vegetation index (NDVI) for cover in the area of Aleppo Mountain (Afrin) during the period 2010–2016. University of Babylon for Pure and Applied Sciences, 6.

Li, X., Zhou, Y., & Wang, J. (2021). Gap filling for historical Landsat NDVI time series by integrating spatio-temporal context and climate data. Remote Sensing, 13(3), 484.

Vasilakos, C., Tsekouras, G. E., & Kavroudakis, D. (2022). LSTM-based prediction of Mediterranean vegetation dynamics using NDVI time-series data. Land, 11(6), 923. https://doi.org/10.3390/land11060923

Faran, I., Netanyahu, N. S., Roitberg, E., & Shoshany, M. (2025). Self-supervised transformers for long-term prediction of Landsat NDVI time series. In Proceedings of the

14th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2025) (pp. 542–552). SCITEPRESS. doi:10.5220/0013381700003905

Zhao, F., Zhang, L., Zhang, Y., et al. (2021). Short and medium-term prediction of winter wheat NDVI based on the DTW–LSTM combination method and MODIS time series. Remote Sensing, 13(22), 4660. https://doi.org/10.3390/rs13224660

To get the NDVI Time series data please visit the following doi (uploaded on my Zenodo account):

https://doi.org/10.5281/zenodo.14743267

To get the Forecasted NDVI results data please visit the following doi (uploaded on my Zenodo account):

https://doi.org/10.5281/zenodo.14743257