earth

MDPI

Article

Machine Learning and SHAP-Based Analysis of Deforestation and Forest Degradation Dynamics Along the Iraq–Turkey Border

Milat Hasan Abdullah [1,2,3] and Yaseen T. Mustafa [4,5,6,*]

[1] Department of Forestry, College of Agricultural Engineering Sciences, University of Duhok, Duhok 42001, Kurdistan Region, Iraq; milat.abdullah@visitor.uoz.edu.krd
[2] Department of Social Science, College of Basic Education, University of Zakho, Zakho 42002, Kurdistan Region, Iraq
[3] Department of Petroleum Geology, Technical College of Zakho, Duhok Polytechnic University, Zakho 42002, Kurdistan Region, Iraq
[4] Department of Environmental Science, College of Science, University of Zakho, Zakho 42002, Kurdistan Region, Iraq
[5] Department of Environmental Engineering, College of Engineering, Knowledge University, Erbil 44001, Kurdistan Region, Iraq
[6] Applied Remote Sensing & GIS Center, University of Zakho, Zakho 42002, Kurdistan Region, Iraq
* Correspondence: yaseen.mustafa@uoz.edu.krd

Abstract: This study explores the spatiotemporal patterns and drivers of deforestation and forest degradation along the politically sensitive Iraq–Turkey border within the Duhok Governorate between 2015 and 2024. Utilizing paired remote sensing (RS) and high-end machine learning (ML) methods, forest dynamics were simulated from Sentinel-2 imagery, climate datasets, and topographic variables. Seven ML models were evaluated, and XGBoost consistently outperformed the others, yielding predictive accuracies (R^2) of 0.903 (2015), 0.910 (2019), and 0.950 (2024), and a low RMSE ($\leq$0.035). Model interpretability was further improved through the application of SHapley Additive exPlanations (SHAP) to estimate variable contributions and a Generalized Additive Model (GAM) to elucidate complex nonlinear interactions. The results showed distinct temporal shifts; climatic factors (rainfall and temperature) primarily influenced vegetation cover in 2015, whereas anthropogenic drivers such as forest fires (NBR), road construction (RI), and soil exposure (BSI) intensified by 2024, accounting for up to 12% of the observed forest loss. Forest canopy cover decreased significantly, from approximately 630 km^2 in 2015 to 577 km^2 in 2024, mainly due to illegal deforestation, road network expansion, and conflict-induced fires. This study highlights the effectiveness of an ML-driven RS analysis for geoinformation needs in geopolitically complex and data-scarce regions. These findings underscore the urgent need for robust, evidence-based conservation policies and demonstrate the utility of interpretable ML techniques for forest management policy optimization, providing a reproducible methodological blueprint for future ecological assessment.

Keywords: fractional vegetation cover; climate and topography factors; machine learning models; spatiotemporal analysis; change detection

check for updates

Academic Editor: Charles Jones

Received: 28 April 2025
Revised: 21 May 2025
Accepted: 23 May 2025
Published: 1 June 2025

Citation: Abdullah, M.H.; Mustafa, Y.T. Machine Learning and SHAP-Based Analysis of Deforestation and Forest Degradation Dynamics Along the Iraq–Turkey Border. *Earth* **2025**, *6*, 49. https://doi.org/10.3390/earth6020049

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Forests are vital ecosystems, covering approximately 31% of the Earth's surface, and they play a crucial role in maintaining ecological balance [1]. However, they are becoming increasingly threatened by deforestation and degradation, leading to a loss of biodiversity, climate change, and disrupted livelihoods. Deforestation involves the complete removal

of trees for non-forest use, such as agriculture or urban development, whereas forest degradation refers to the decline in forest quality due to factors such as logging, pollution, or climate change, even though the forest remains standing [1].

The drivers of deforestation and forest degradation are multifaceted and large-scale. These generally involve two broad categories: human activity and natural environmental factors. Human activities include agriculture, logging, urban expansion, road construction, and the use of firewood for livelihoods, which are occasionally worsened by conflict and poor management. Natural environmental factors encompass climate change, particularly changes in temperature and rainfall, earth topography, diseases, and their impact on domestic animals.

Understanding the causes and patterns of forest degradation requires both environmental and socio-political perspectives. Traditional remote sensing techniques have been widely applied to detect land cover change and forest dynamics across different regions [2,3]. However, these approaches often lack the predictive capacity and interpretability needed for complex, data-scarce areas. The advancement of machine learning (ML) has enhanced our ability to process and analyze large-scale spatial data with improved accuracy [4–6]. ML methods also offer flexibility in handling nonlinear relationships and mixed variable types.

Recently, studies have shown that integrating ML with explainable models such as SHAP and GAM improves insights into the spatial and temporal behavior of forest change [7,8]. These models help to quantify the relative influence of ecological and anthropogenic factors, such as rainfall, temperature, fire, and human expansion. In regions like northern Iraq, where conflict and terrain complexity intersect, such approaches are particularly valuable [9,10]. Therefore, expanding the theoretical framework of forest degradation using data-driven, interpretable techniques is essential for informed policy and conservation planning.

ML techniques are increasingly being applied in forestry and related fields. For instance, researchers have used Support Vector Machines (SVMs), Classification Regression Trees (CRTs), and Random Forest (RF) to predict land degradation in Iran and found RF to be the most effective predictive model [4]. Researchers have employed ML to identify the drivers of forest degradation and deforestation in the Congo Basin of Cameroon, categorizing the data into broad, fine-scale, and disaggregated time series [5]. The authors classified satellite images of Abbottabad, Pakistan, using an Artificial Neural Network (ANN) and RF to analyze forest cover and support conservation efforts [6]. ML algorithms are integrated with the SHapley Additive exPlanations (SHAP) method to assess urban flood vulnerability, emphasizing the importance of identifying optimal feature sets [7]. Furthermore, they focused on climate factors, vegetation indices, and Landsat imagery to monitor forest degradation in India through machine learning and utilized the Google Earth Engine (GEE) platform to capture and organize data [8].

Forest management information systems are essential tools for planning, implementing, and monitoring activities. Regions such as the Duhok Governorate have experienced a decline in overall vegetation cover, especially over the past ten years [9]. The economic limitations and conflicts in Duhok have increased the complexity of forest conservation and management. Increasing pressure on the environment highlights the need for greater forest maintenance and sustainable management, particularly in the Iraqi Kurdistan Region (IKR), where they are superimposed on domestic challenges. The border between Iraq and Turkey, the study area for this research, faces high levels of forest degradation and deforestation due to political, economic, and environmental issues, including unregulated logging, land use changes, and conflicts.

The Iraq–Turkey border area is characterized by factors leading to forest degradation and deforestation, including increased conflicts and wars resulting in burned areas, tree cutting, and road construction, as well as climatic and topographic influences. The study area is exposed to climatic challenges and human conflicts, making field data col-lection risky. Traditional forest management methods are inadequate, and data scarcity hinders effective monitoring. Although helpful, RS techniques have not been fully integrated with modern ML techniques for real-time data assessments. The combination of ML and remote RS is still in its early stages, especially in areas such as the Duhok Governorate. In light of these and other diverse challenges and drivers, it is important to focus on a specific region that reflects these complexities. Accordingly, this study explores the sensitive geopolitical area of the Iraq–Turkey border based on the Duhok Governorate. Although socio-economic variables such as population density or economic activities could strongly influence deforestation, such data were excluded due to limited availability and reliability within this geopolitically sensitive region.

This study aims to monitor and analyze forest degradation and deforestation along the Iraq–Turkey border between 2015 and 2024, a region characterized by ecological sensitivity and political instability. The research integrates explainable ML techniques with high-resolution Sentinel-2 RS data to model and interpret the drivers of forest change. Specifically, this study seeks to quantify the extent of forest loss, compare the predictive performance of different ML models, and utilize SHAP and GAM to improve interpretability. It also aims to evaluate the relative influence of climatic, topographic, and anthropogenic factors over time. The research is guided by two key questions: (1) What are the prevailing drivers of deforestation and forest degradation in the study area? (2) How have these drivers evolved temporally in relation to ecological and human-induced pressures? By addressing these objectives, this study provides a data-driven, reproducible framework for understanding landscape dynamics in geopolitically complex and data-limited regions.

2. Study Area

The study area is located north of the Duhok Governorate, IKR, on its border with Turkey. Geographically, it is located at latitudes 37°0′0″ N to 37°20′0″ N and longitudes 43°0′0″ E to 44°0′0″ E with an area coverage of approximately 1645 km^2. The study area features varying elevations, ranging from 583 m to 2597 m above sea level, with the steepest slope reaching 85% (Figure 1). It is worth noting that the study area covers the northern parts of the two districts of the Duhok Governorate. These are Zakho and Amadi, which are border regions with Turkey. The study area has the highest vegetation cover in the Duhok Governorate [11]. Additionally, the Amedi district has the highest greenspace coverage in the IKR, accounting for 59.8% [12]. The study area is cold and semi-arid, with a Mediterranean rainfall pattern. The region experiences a long winter, an extended spring, a semi-moderate summer, and a short autumn. It has a complex topography, and diverse ecosystems create unique terrains and rich vegetation, including mountain forests, riverine forests, and steppe pastures [13]. These natural features make this area of great value for agriculture and tourism, making the conservation of the area's forests of utmost importance, potentially facilitating the work of other researchers in the region.

However, in recent decades, deforestation and land degradation have threatened dense forests. Human activities, including logging for fuel, infrastructure development, and political conflicts, have led to severe forest loss. This region is particularly affected by forest fires and illegal logging. Additionally, ongoing conflicts between Turkey and the Kurdistan Workers' Party (PKK) have turned parts of the area into restricted zones, further accelerating deforestation and environmental damage [10]. To effectively quantify and analyze the described environmental pressures and spatiotemporal dynamics, a

methodological framework was adopted. This methodology, which combines RS data, climatic and topographic information, and advanced ML approaches, is detailed in the following section.

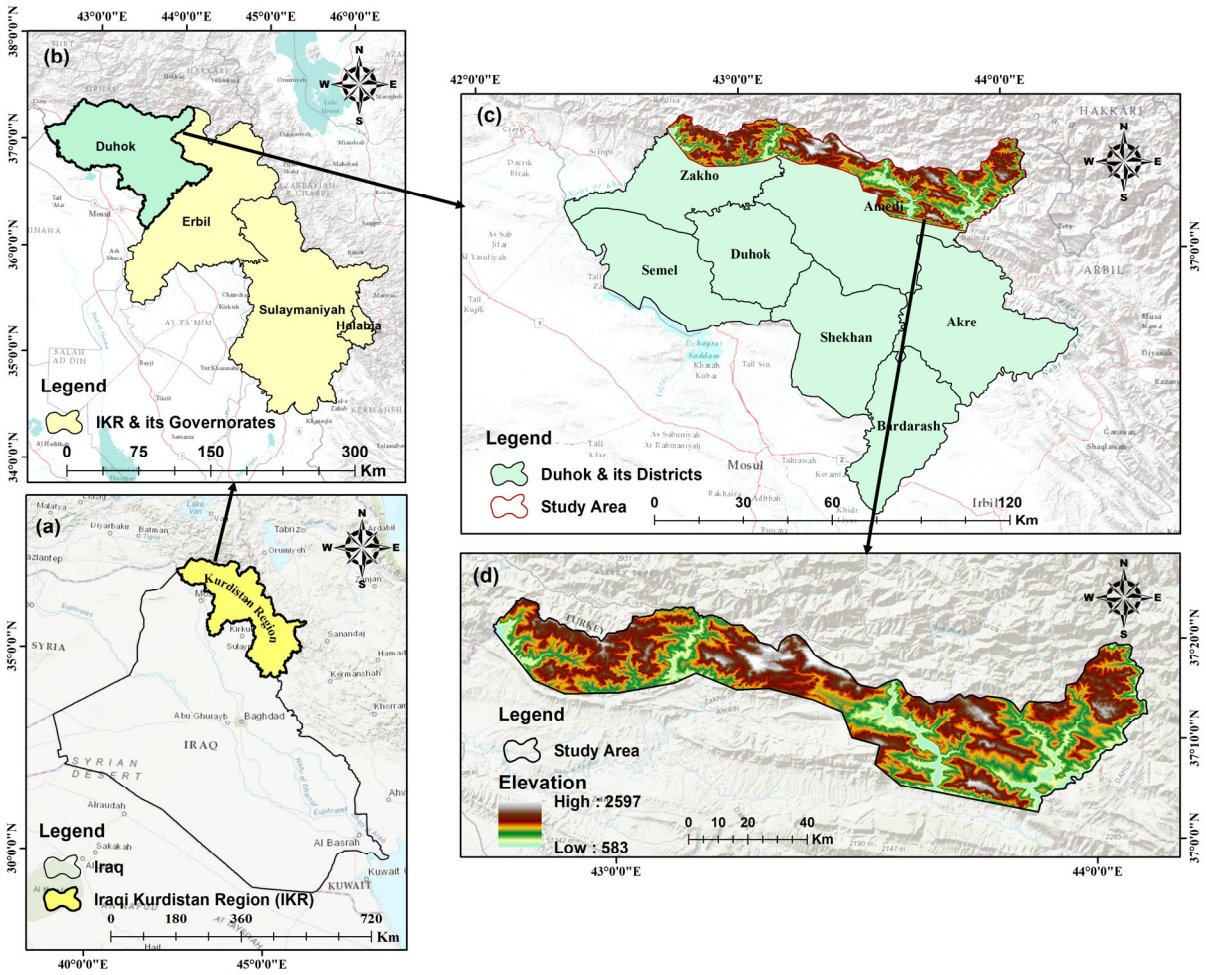


Figure 1. Map of (**a**) Iraq and the IKR with the Turkey border. (**b**) The IKR. (**c**) Duhok Governorate with its districts. (**d**) Study area with elevation features.

3. Materials and Methods

This study used an ML framework to investigate spatiotemporal forest conditions by combining data-driven methods with computational analysis. The entire process, as shown in Figure 2, was organized into three major parts.

The initial phase was data acquisition, whereby relevant features were obtained from various sources and converted into thematic maps with a 10 m spatial resolution. In the second part, a strict process of feature selection with multicollinearity and correlation analysis was used to ensure that only the most relevant variables were utilized in the ML models. The final part involved model development, testing, spatial mapping, and the interpretation of results.

A set of proprietary software packages were used throughout the study. ArcGIS (v. 10.8) was used for GIS mapping and ENVI (v. 5.6) was used to enable image processing and analysis. The R Project for Statistical Computing (v. 4.3) was used for statistical modeling and model building, and the GEE was used to enable satellite data processing and feature extraction.

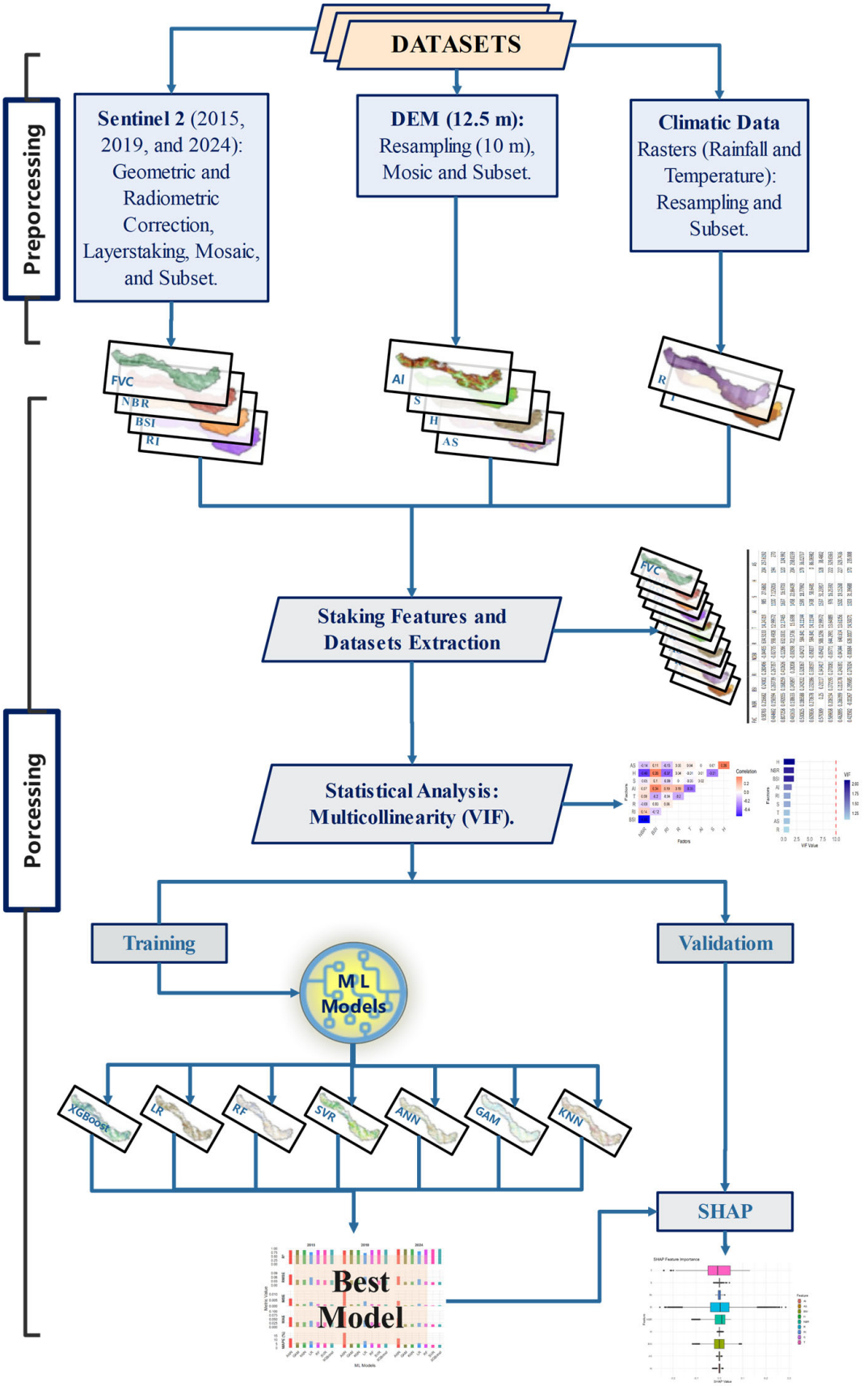


Figure 2. Data and method workflow.

3.1. Data Sources and Preprocessing

Given the significant role of climatic and topographic variables identified in prior research [14,15], their incorporation into our analysis is pivotal for a comprehensive understanding of forest dynamics and deforestation patterns.

The datasets used in this study included the following primary types: satellite images, topographic data, and climate data. These datasets were systematically processed to guarantee spatial resolution uniformity and facilitate accurate environmental assessments. Sentinel-2 images were obtained as the main dataset for land cover change studies, whereas Digital Elevation Model (DEM) data and climate conditions (temperature and precipitation) were useful environmental data. A summary of the datasets, descriptions, derived factors, and sources is presented in Table 1.

Table 1. Dataset resources.

Dataset Type	Extracted Factors	Data Description	Source
Satellite Imagery	Land Cover Indices	Sentinel-2 images (10 m resolution), acquired for August 2015, 2019, and 2024.	https://earthengine.google.com/ (accessed on 25 August 2024)
Topographic Data	Elevation, Slope, Aspect	DEM dataset (12.5 m resolution, resampled to 10 m).	https://search.asf.alaska.edu/ (accessed on 25 August 2024)
Climate Data	Rainfall, Temperature	Annual precipitation from CHIRPS (4800 m, resampled to 10 m), and temperature from ERA5 (9600 m, resampled to 10 m).	https://app.climateengine.org/ (accessed on 12 October 2024)

Sentinel-2 data, acquired by the Google Earth Engine (GEE) platform, were chosen for three cloud-free dates in August 2015, 2019, and 2024. The images were pre-processed with regular radiometric, atmospheric, and geometric corrections. Additional preprocessing procedures, such as layer stacking and masking of the study area, were performed to improve analytical precision. Due to the multi-year intervals of available Sentinel-2 imagery, shorter-term deforestation events might have been missed. However, the intervals chosen were representative of the most significant changes documented by secondary reports.

The climate data included the annual average temperature (T) from ERA5 and the yearly precipitation (R) from the CHIRPS dataset. These data were originally available at coarse spatial resolutions of 9600 m and 4800 m, respectively (Figure 3m–r). The Climate Engine platform from which these data were obtained has been thoroughly tested and used in other research investigations, such as urban heat island research [16] and environmental monitoring [17].

A digital elevation model (DEM) of 12.5 m resolution with radiometrically corrected data was downloaded from the Earth Data website of the National Aeronautics and Space Administration (NASA) (Figure 3s). The data were used to perform a spatial analysis of the study area and to measure its effect on forest dynamics. The DEM scenes were initially mosaicked, followed by extraction of the study area through a masking process. To ensure compatibility with Sentinel-2 images, all the datasets were resampled to a 10 m resolution using the nearest neighbor resampling method in ArcGIS 10.8. This step was crucial for maintaining consistency in the spatial analysis.

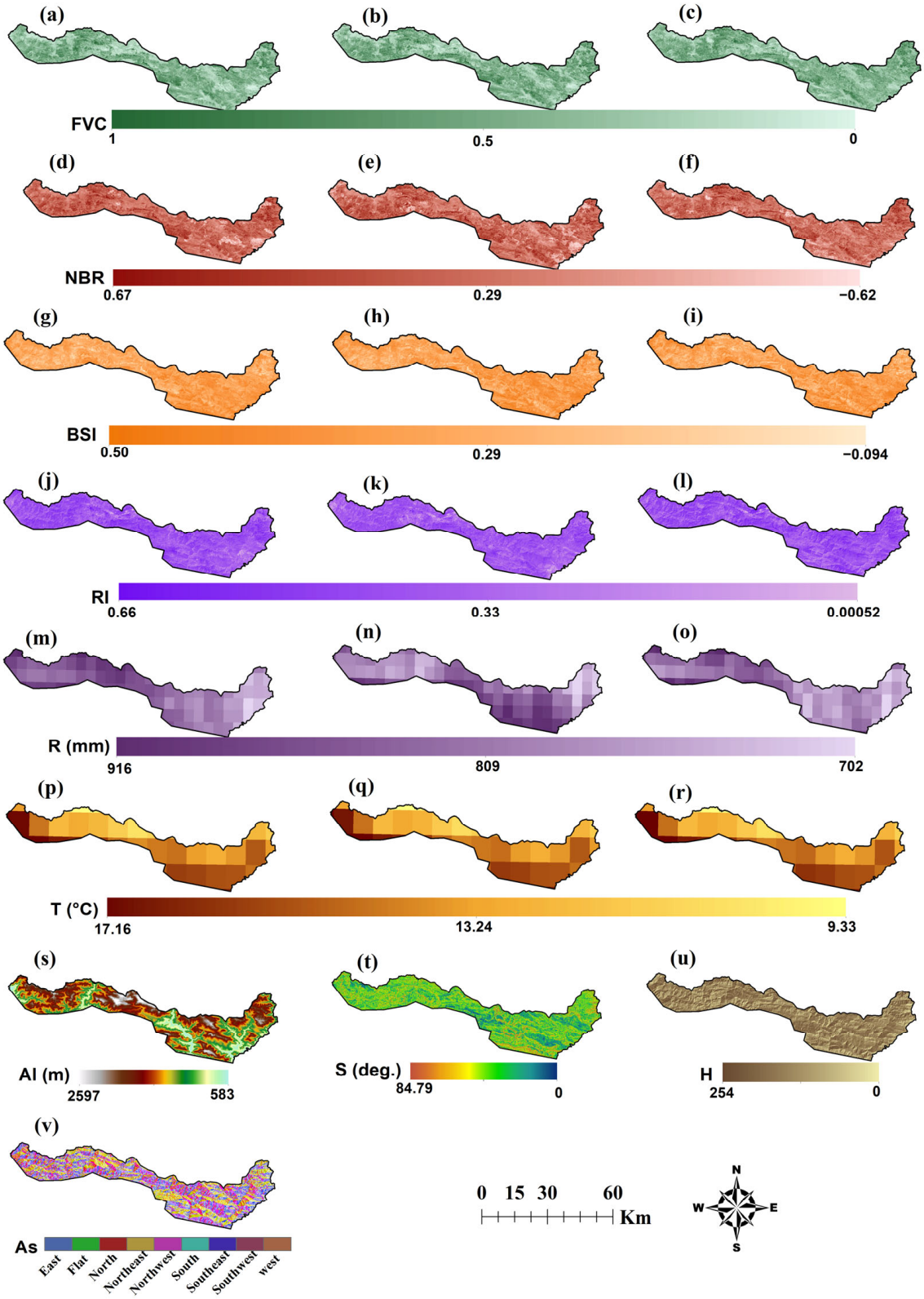


Figure 3. Analysis of key drivers of forest cover in the study area. FVC ((**a**) 2015; (**b**) 2019; (**c**) 2024), NBR ((**d**) 2015; (**e**) 2019; (**f**) 2024), BSI ((**g**) 2015; (**h**) 2019; (**i**) 2024), RI ((**j**) 2015; (**k**) 2019; (**l**) 2024), R ((**m**) 2015; (**n**) 2019; (**o**) 2024), T ((**p**) 2015; (**q**) 2019 (**r**) 2024), Al (**s**), S (**t**), H (**u**), and As (**v**).

3.1.1. Land Cover Indices

Numerous studies have employed various indices extracted from satellite images to examine land use and land cover in specific areas. In this study, four indices were used to

identify the factors affecting the forest vegetation cover, namely the fractional vegetation cover (FVC), normalized burn ratio (NBR), Bare Soil Index (BSI), and Road Index (RI). All indices were derived from Sentinel-2 images and were calculated using the GEE platform. Each of these indices is calculated using a unique formula and is specified in terms of certain value ranges based on its recommended use (Table 2, Figure 3a–l). The values of all indices fall within the range of −1 to +1.

FVC is a valuable indicator for tracking environmental changes, assessing vegetation conditions, and estimating deforestation and land degradation [18]. It has been identified as an elementary phenotypic trait in agriculture, forestry, and ecological science, providing a measurable indicator of vegetation distribution and coverage pattern [19]. In this study, FVC was used as a critical parameter in the estimation of the forest vegetation area and as a dependent variable in the analysis. It was obtained based on NDVI in combination with its minimum and maximum values to accurately estimate vegetation.

NBR is a widely used metric in the study of wildfires to detect areas burned over large fire-affected regions [20]. It is a standard metric for assessing fire damage; a high NBR value generally indicates healthy vegetation, while a low value indicates barren land and severely burned areas [21]. The NBR was calculated based on normalizing Shortwave Infrared 2 (SWIR 2) and red bands (Table 2). The forests in the study area experience annual burning, primarily driven by human-related conflicts. This study applied the NBR index to monitor fire-damaged areas and evaluate their role in forest degradation and the acceleration of deforestation.

The BSI serves as a quantitative index for assessing bare soil extent, forest cover, and soil erosion rates [22]. It is derived using the square root of the ratio between the difference and sum of the SWIR2 and Green (G) spectral bands, multiplied by 100 [23] (Table 2). The BSI was used in this study to map soil exposure and examine temporal land-cover changes during the last decade. It is particularly relevant for this region due to widespread land clearing and increasing bare soil exposure caused by illegal logging, conflict-driven fires, and overgrazing.

RI is a key parameter used to detect road networks. Road feature extraction from satellite images is a challenging task, but it is extremely important for a wide range of applications, such as urban planning, cartography, and traffic management [24]. RI was computed based on the saturation ratio method from particular Sentinel-2 bands (11, 8, and 2) [25] (Table 2). In this study, RI was used to assess the impact of road construction on forested areas in the study area. It is especially important here, as road expansion is a primary driver of forest fragmentation and accessibility-related degradation in the Iraq–Turkey border zone.

Table 2. Land cover indices derived from Sentinel-2 bands.

Index	Formula	Reference
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR\ (B8)-Red(B4)}{NIR(B8)+Red\ (B4)}$	[26]
Fractional Vegetation Cover (FVC)	$\frac{NDVI-NDVImin}{NDVImax-NDVImin}$	[27]
Burn Ratio (NBR)	$\frac{SWIR2\ (B12)-Vegtation\ red\ edge(B8A)}{SWIR2\ (B12)+Vegtation\ red\ edge(B8A)}$	[28]
Bare Soil Index (BSI)	$\sqrt{\frac{SWIR2(B12)-Green(B3)}{SWIR2(B12)+Green(B3)}} \times 100$	[23]
Road Index (RI)	$1-\frac{3\times Min[SWIR1(B11),NIR(B8),Blue(B2)]}{SWIR1(B11)+NIR(B8)+Blue(B2)}$	[24]

3.1.2. Topographic Variables

Topography is a key factor influencing the spatial distribution of forest cover across the study region. This region is characterized by rugged mountain terrain and has the highest elevation in the IKR. Topographic variables have previously been proven to be closely linked with forest degradation, and their controlling influence on the distribution of vegetation and ecosystem processes has been explored [14,29]. In this study, four significant topographic factors, including altitude (AL), slope (S), aspect (AS), and hillshade (H), were investigated to determine their roles in forest dynamics (Figure 3s–v). The variables were extracted using ArcGIS 10.8, where all topographic data were resampled to a spatial resolution of 10 m to match the dataset consistency.

Altitude (AI) serves as a primary determinant of forest distribution by influencing temperature, precipitation, and ecological zones. Higher altitudes are normally associated with cold climates and increased moisture availability, supporting vegetation growth [30]. In contrast, lower altitudes are more susceptible to human activities such as deforestation and agricultural land use.

Slope (S) is another critical factor for assessing forest stability and regeneration. It is challenging to preserve forests on steep slopes, especially those with high inclinations, because they prone to erosion, landslide, and poor soil retention [14,31]. High-incline, sharp-slope areas can also pose effective fire management challenges, thereby exposing forests to loss.

Aspects (AS), or slope orientation, also affect vegetation distribution by controlling sunlight exposure and microclimatic conditions. Northern-facing slopes tend to retain more moisture and support denser vegetation because of reduced solar exposure, whereas southern slopes receive more direct sunlight, leading to higher evaporation rates and drier conditions [32]. Variations in vegetation density and soil properties can influence the resilience of forests to environmental stressors and degradation.

Additionally, the degree of shading on the ground, known as hillshade (H), is a key factor for forest health. Areas with lower shade exposure receive increased solar radiation, resulting in elevated soil temperatures and drier conditions, which can cause physiological stress on vegetation and accelerate forest decline [15]. In contrast, shaded slopes can support greater moisture content and therefore greater forest cover, whereas slopes with higher evaporation and drought stress potential occur on exposed slopes [33]. By examining these topographic variables, this study sought to determine the combined effect of forest cover and environmental resilience in the study area.

3.2. Machine Learning Models

In this study, ML techniques were employed to identify the drivers of forest cover dynamics and to quantify their influence. Seven ML models, including XGBoost, Linear Regression (LR), Random Forest (RF), Support Vector Regression (SVR), an Artificial Neural Network (ANN), a Generalized Additive Model (GAM), and K-Nearest Neighbors (KNN), were assessed to measure correlations between dependent and independent variables. All the algorithms were run using the R programming package. The independent variable was FVC, while the independent variables included spectral indicators (NBR, BSI, RI, and MBI), climatic predictors (R and T), and topographic predictors (Al, S, H, and As).

XGBoost is an ensemble learning algorithm based on the gradient-boosting framework [7]. It enhances classical gradient boosting by using second-order gradient statistics to optimize loss functions, achieving faster convergence and higher accuracy than models such as Gradient Boosting Decision Trees (GBDT) [34,35]. XGBoost builds sequential decision trees, each of which corrects errors made in previous iterations and uses regularization methods to avoid overfitting [36,37].

Linear Regression (LR) defines linear relationships between one dependent variable and a set of multiple independent variables [38]. LR requires minimal multicollinearity among predictors and works best when the relationships are approximated to be linear. However, LR performs poorly when modeling nonlinear interactions or working with high-dimensional data.

Random Forest (RF), a bagging ensemble algorithm, creates an array of decision trees using bootstrap aggregating and feature selection by chance [39,40]. Learning was carried out in the trees using a portion of the data (two-thirds through bootstrapping), and the OOB samples cross-validated the model performance [41]. RF is particularly useful for handling nonlinear data without overfitting and is therefore applicable to both regression and classification issues.

Support Vector Regression (SVR) is the generalization of Support Vector Machines (SVMs) to regression problems, projecting nonlinear data onto higher-dimensional spaces using kernel functions [42,43]. SVR regulates model complexity and errors in prediction through tunable parameters (e.g., ε-insensitive loss function and penalty factor C), offering flexibility in handling various datasets [44,45].

Artificial Neural Networks (ANNs) are computing systems inspired by biology composed of groups of interconnected nodes (neurons) in layers [46]. ANNs learn complex nonlinear patterns by repeatedly adjusting weights and are therefore used for regression, classification, and pattern recognition tasks [36].

Generalized Additive Models (GAMs) are a broadening of generalized linear models (GLMs) that allow smooth, nonlinear functions to capture predictor–response relationships [14]. GAMs were applied in this case to illustrate and clarify the correlation effects of the independent variables on FVC.

K-Nearest Neighbors (KNN) is an instance-based, nonparametric algorithm that makes predictions based on similarity measures (e.g., Euclidean distance) between input attributes and training instances [47,48]. This approach is based on assessing the similarity between variables, which is a nonparametric case-based learning approach.

3.3. SHAP-Based Machine Learning Interpretability

ML models are often perceived as "black boxes" due to their complexity and difficulty in interpreting how individual features influence predictions. Thus, their uninterpretability makes it difficult to ascertain the positivity or negativity of a feature regarding model output. SHapley Additive exPlanations (SHAP) is widely acknowledged to be one of the best methods for enhancing the interpretability of ML and provides an explicit methodology for approximating feature contributions. It interprets each feature as a player in a cooperative game where the model's output represents the final score of the game.

SHAP values are based on the contribution of each feature to the prediction as compared to other features and thus help in ranking the feature importance in the model [49]. These values were calculated using built-in functions to provide an equal and objective evaluation of various datasets. Equation (1) outlines the SHAP value calculation [50].

$$f(X') = J_0 + \sum_{i=1}^{K} J_i X'_i \quad (1)$$

where f represents the interpretation model, X' denotes the basic features, K is the maximum feature subset size, and J represents feature attributions.

The novelty of the SHAP is its consistency in measuring feature importance in a way that is theoretically justified, making it suitable for a wide range of ML tasks [49]. This approach has been extensively employed in areas such as pedestrian safety evaluations [51] and predictive analytics [52]. SHAP achieves this by comparing model predictions both

with and without a specific feature, allowing it to assess the strength of the influence of that feature. SHAP analysis was performed in this study using the "SHAP" package in R, providing deeper insights into the role of each variable in the ML models.

3.4. Sampling Strategy and Dataset Preparation

Random sampling was conducted to ensure proper assessment of forest cover within the study area. For the entire study area of 1645 km^2, 15,650 sample points were randomly allocated, representing approximately 10.5% of the study area (Figure 4). The sampling design achieved a minimum interval of 1 km between consecutive points, with maximum spatial coverage across a range of topographic features. The sampling was carried out in ArcGIS 10.8 and the sample points were further separated into training and validation sets for ML model development.

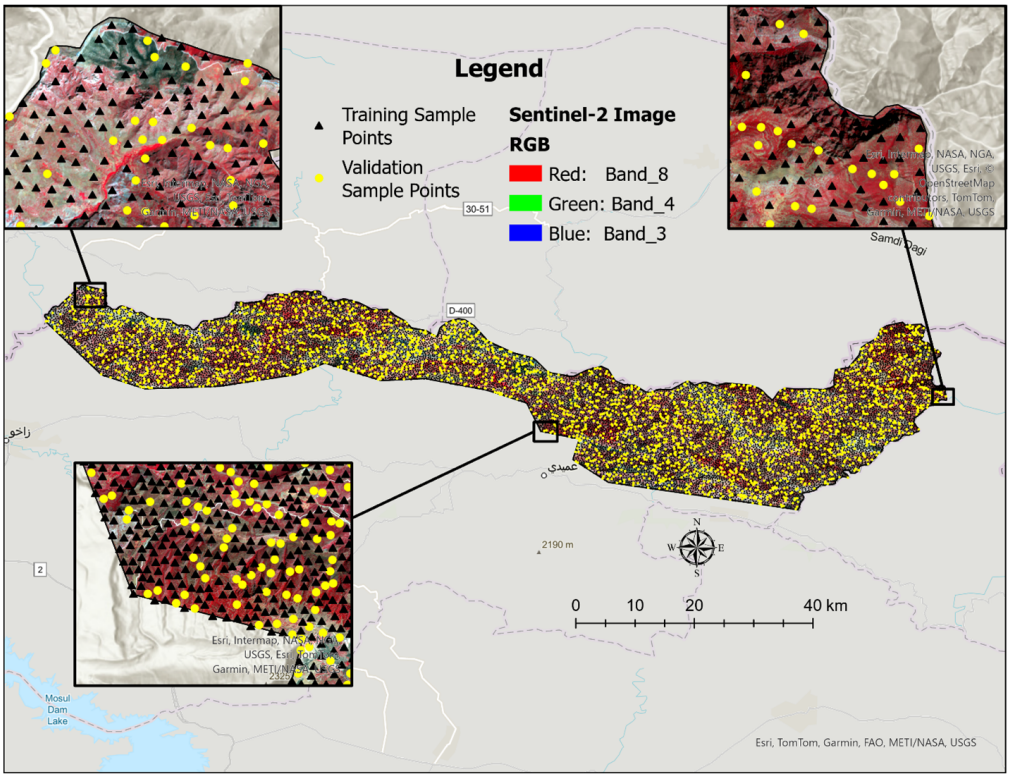


Figure 4. Spatial distribution of sample points across the study area. Black triangles indicate training sites, while yellow circles mark validation points, ensuring broad and representative coverage for model development.

To create a comprehensive dataset, thematic maps of all the factors of interest (FVC, NBR, BSI, RI, R, T, AL, S, H, and AS) were stacked and merged using ENVI software (v. 5.6). The values corresponding to 15,650 sampling points were then extracted using ArcGIS, ensuring that each point contained information from all layers. These extracted values formed the basis for model calibration and validation. Field validation was not conducted due to ongoing security risks in the study region. Instead, robust random sampling and validation sets were employed to ensure the reliability and representativeness of the results.

The random sampling method was applied to balance the datasets and improve ML models' performance. The dataset was divided into two categories: 80% for training, enabling the model to learn subtle patterns for estimating forest vegetation cover, and 20% as a hold-out to check the accuracy of the model. This approach enhanced the models' ability to generalize new data and refine predictive performance. Figure 4 demonstrates the geographic location of sample points, where training samples (black triangles) were

used for model construction and validation samples (yellow circles) were withheld from the training process to allow for objective assessment.

3.5. Feature Selection

To improve model accuracy, a rigorous feature selection process was applied to eliminate redundant and irrelevant attributes. The Variance Inflation Factor (VIF) was employed as the primary statistical metric to identify multicollinearity between independent variables [53]. By measuring how much the variance of a regression coefficient is inflated because of the intercorrelation between predictors, VIF identifies troublesome dependencies.

VIF analysis was performed using R (v. 4.3) packages Caret and Car, which also supported the VIF and Pearson correlation evaluation calculations. The relationships between variables were visualized using heatmaps to identify and remove less important features. Independent variables that are highly correlated can form spurious parameter estimates, which provide unstable standard errors. By deleting the highly correlated predictors, the VIF analysis enhanced the model interpretability and stability of the predictions.

The VIF was used for the independent variable x_i and was calculated as follows:

$$\mathrm{VIF}(x_i) = \frac{1}{1 - R_i^2} \tag{2}$$

where R_i^2 represents the coefficient of determination when X_i is regressed against other predictors. The VIF interpretation rules were as follows: VIF = 1, no multicollinearity; VIF = 1–5, moderate but acceptable relationship; VIF > 5, high correlation that must be carefully monitored; and VIF > 10, extreme multicollinearity that requires dropping the variable.

3.6. Model Performance Evaluation

Five statistical metrics were used to assess the accuracy and reliability of the ML models. These measures were the mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) (Equations (3)–(7)). All measures were used for all three years of the study (2015, 2019, and 2024) to ensure a complete analysis.

The SHAPE package and the GAM in R Studio (v. 2023.12.1+402) were used to test the interaction between FVC and its influencing factors. The MAE, MSE, RMSE, and MAPE measure the prediction accuracy with a gap between the estimated and observed values. These metrics typically range from 0 to 1, with lower values indicating a better fit, and ideal model performance is denoted by lower values [54]. Concurrently, R^2 and adjusted R^2 measure the extent of correlation between independent and dependent variables, with values closer to 1 indicating a stronger predictive relationship and higher model accuracy [55]. Each metric was applied in the R. All the statistical tests and model fits were conducted within R, ensuring a robust and data-driven evaluation process. Having established robust evaluation metrics, we applied these criteria comprehensively to assess multiple ML algorithms for their predictive performance, as detailed in the Results section below.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \quad (7)$$

4. Results

4.1. Predictor Variable Evaluation and Model Selection

A comprehensive dataset comprising 15,650 randomly distributed sample points was successfully constructed using the carefully designed sampling strategy detailed earlier (Section 3.4). This robust sampling approach secured representative spatial coverage across the diverse climatic and topographic conditions of the study area, enhancing the reliability and generalizability of subsequent correlation analyses and model performance evaluations.

Correlation and multicollinearity between predictor variables were assessed using Pearson's correlation coefficients and Variance Inflation Factor (VIF) analysis for each study year (2015, 2019, and 2024). The correlation matrix in 2015 (Figure 5a) showed relationships, including a moderate negative correlation between the Bare Soil Index (BSI) and normalized burn ratio (NBR) (−0.56), indicating that increased bare soil is associated with reduced forest fire activity. Temperature (T) was negatively correlated with altitude (Al) (−0.35) and slope (S) (−0.03). Figure 5b illustrates the VIF values for all variables, indicating minimal multicollinearity (VIF < 2) for the subsequent regression analysis, which is quite good, as it is less than 5.

Correlation analysis for 2019 (Figure 6a) revealed moderate to weak positive relationships between normalized burn ratio (NBR) and BSI (0.45), suggesting an association between burned areas and heightened exposure of the soil. Temperature (T) was negatively correlated with altitude (Al) (−0.35) and positively correlated with aspect (AS) (0.04). The VIF analysis for 2019 (Figure 6b) again showed low multicollinearity among all variables (VIF < 2).

By 2024, Figure 7a indicates stronger correlations between anthropogenic variables; for instance, a moderate negative correlation is observed between the Road Index (RI) and the BSI (−0.48), along with a moderate positive relationship be-tween RI and NBR (0.45). The VIF analysis (Figure 7b) further demonstrated long-standing low multicollinearity, upholding the appropriateness of such variables for follow-up modeling. Collectively, these analyses support the robustness of the predictor variables for use in regression modeling throughout the study period.

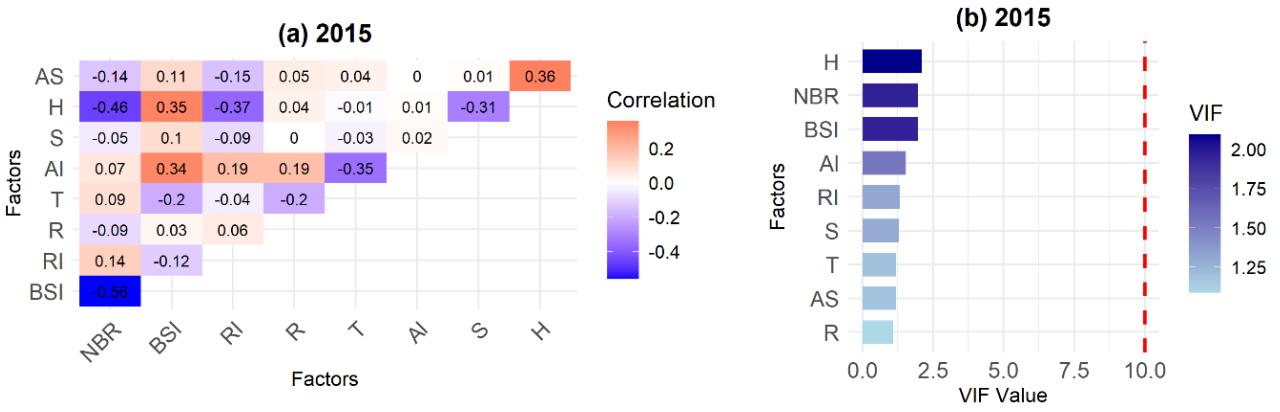


Figure 5. (**a**) The correlation between factors for the year 2015. (**b**) VIF values between independent factors for the year 2015.

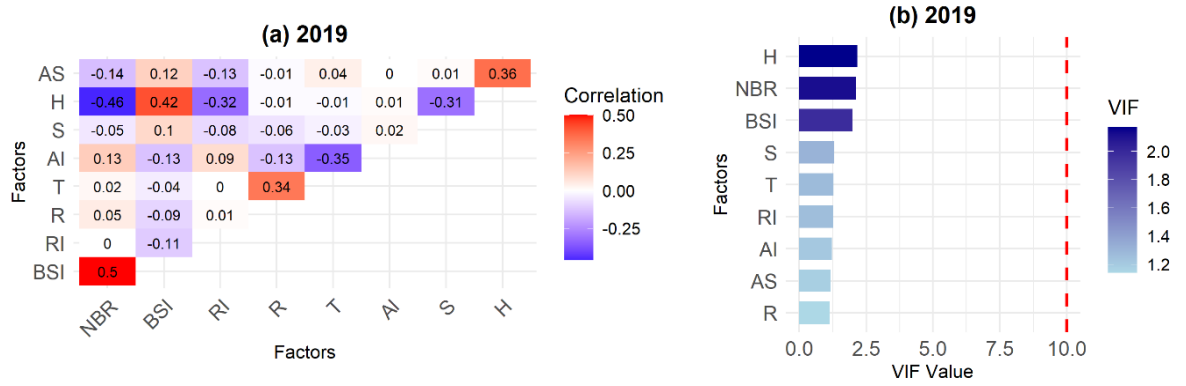


Figure 6. (**a**) The correlation between factors for the year 2019. (**b**) VIF values between independent factors for the year 2019.

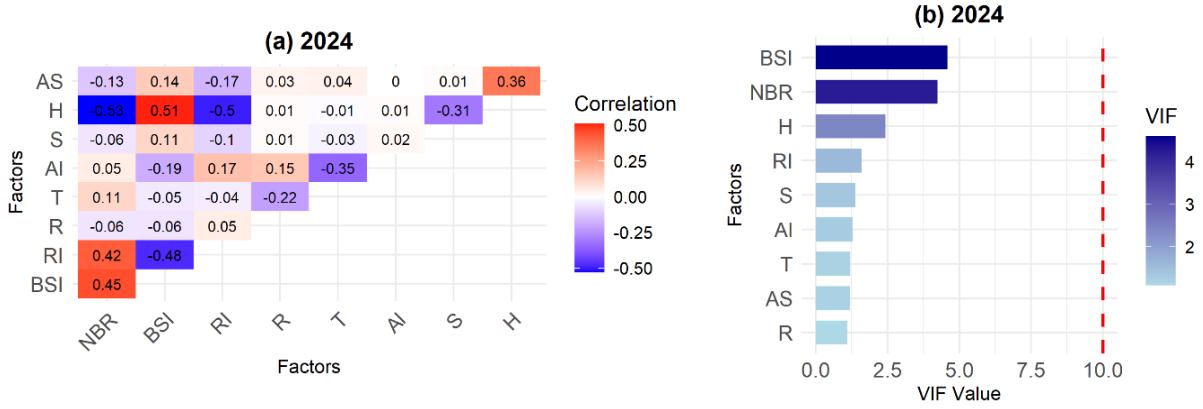


Figure 7. (**a**) The correlation between factors for the year 2024. (**b**) VIF values between independent factors for the year 2024.

Having confirmed the suitability and minimal multicollinearity of the predictor variables, we next evaluated multiple ML algorithms to identify the most effective predictive model.

Seven ML models were rigorously tested to predict FVC, with detailed performance outcomes presented in Table 3 and Figure 8 for all the study years (2015, 2019, and 2024). Of all the tested models, XGBoost repeatedly showed a higher predictive accuracy across all study years. This was observed through the highest R^2 values ($\geq$0.90) and lowest prediction errors, including RMSE values of 0.035 (2015), 0.034 (2019), and 0.022 (2024).

SVR closely followed as the second-best performing model, presenting marginally higher errors and slightly lower R^2 values than XGBoost. Linear Regression (LR), on the other hand, had uniformly poor predictability, with significantly lower R^2 values (e.g., 0.698 in 2019) and larger error readings during all the study intervals. The ANN recorded the largest prediction errors, indicating its unsuitability for this dataset.

Owing to its strong and improved performance, XGBoost was identified as an ideal model for further analysis and interpretation. The reliable performance of XGBoost highlights the ability of this algorithm to predict sophisticated forest-cover patterns driven by heterogeneous environmental and anthropogenic factors. Therefore, subsequent analyses, including SHAP and GAM, were based on XGBoost predictions to accurately interpret variable impacts. This further validates the real-world applicability of the model for forest conservation and planning endeavors in this geopolitically sensitive region.

Table 3. Comparative performance metrics of ML models for predicting FVC in 2015, 2019, and 2024.

Year	Metric	XGBoost	LR	RF	SVR	ANN	GAM	KNN
2015	MAE	0.0273	0.0387	0.0308	0.0275	0.0605	0.0297	0.0298
	MSE	0.0012	0.0032	0.0016	0.0013	0.0062	0.0015	0.0015
	RMSE	0.0352	0.0563	0.0394	0.0355	0.0784	0.0381	0.0384
	MAPE (%)	5.61	8.11	6.43	5.66	11.58	6.19	6.22
	R^2	0.903	0.751	0.887	0.900	0.873	0.886	0.885
2019	MAE	0.0258	0.0417	0.0295	0.0254	0.0961	0.0269	0.0281
	MSE	0.0012	0.0038	0.0015	0.0011	0.0127	0.0012	0.0014
	RMSE	0.0344	0.0613	0.0391	0.0334	0.1129	0.0349	0.0378
	MAPE (%)	4.96	8.33	5.76	4.85	18.56	5.04	5.51
	R^2	0.910	0.699	0.886	0.905	0.861	0.902	0.887
2024	MAE	0.0161	0.0249	0.0195	0.0154	0.0564	0.0168	0.0182
	MSE	0.0005	0.0018	0.0007	0.0004	0.0042	0.0005	0.0006
	RMSE	0.0219	0.0426	0.0270	0.0211	0.0650	0.0231	0.0244
	MAPE (%)	3.69	5.85	4.55	3.53	11.52	3.88	4.15
	R^2	0.950	0.808	0.930	0.935	0.940	0.944	0.938

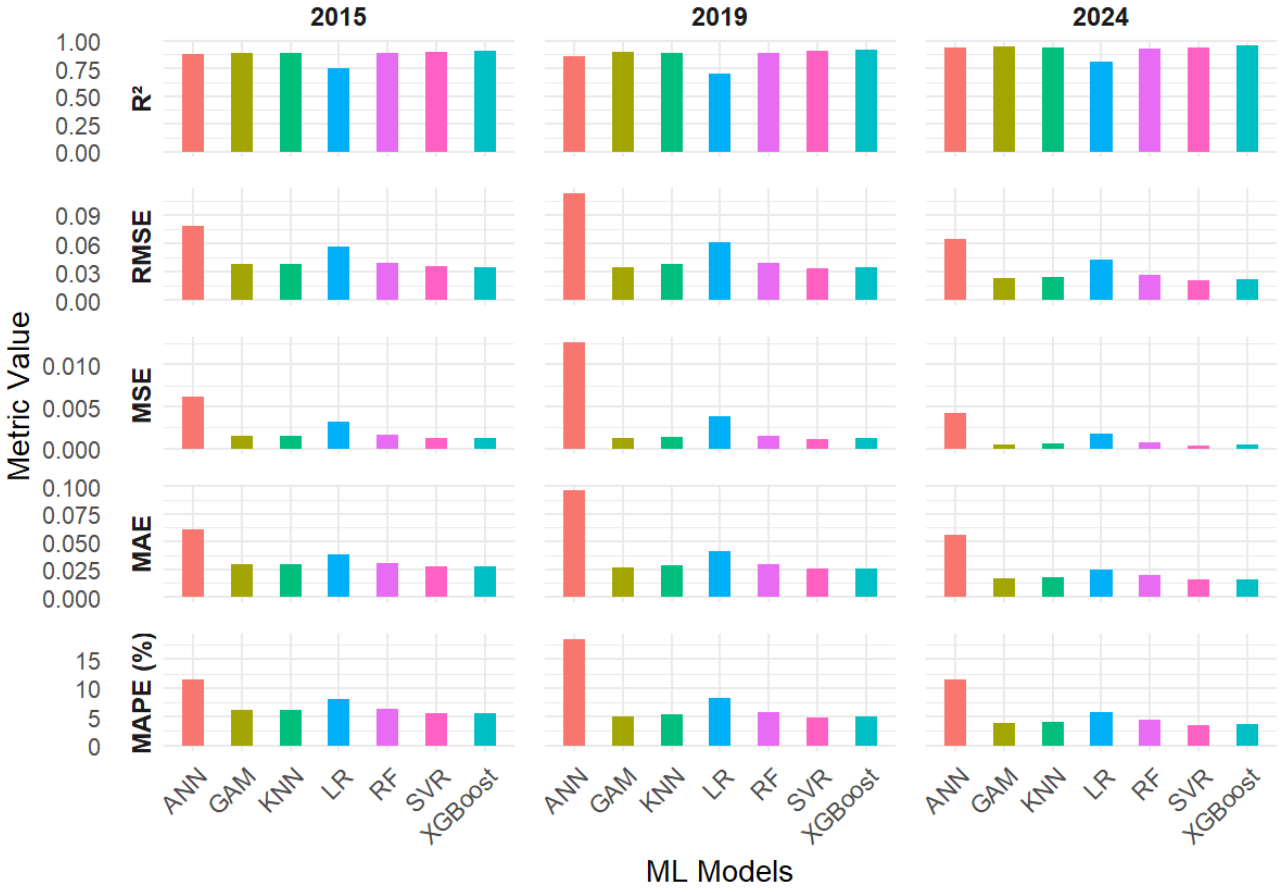


Figure 8. Comparative performance of ML models for FVC Prediction over three years (2015, 2019, 2024).

4.2. Spatiotemporal Dynamics and Drivers of Forest Change

A comparative analysis of the influencing factors across the three years revealed a significant transition from climate-dominated vegetation dynamics to anthropogenically driven deforestation and degradation. In 2015, climatic variables played a dominant role in influencing forest cover. The SHAP values indicated that rainfall (+0.32) contributed positively, whereas temperature contributed negatively (−0.24) to forest cover (Table 4).

This suggests that ecological conditions primarily governed forest distribution and density in 2015, thereby supporting natural growth cycles and seasonal regeneration.

Table 4. Temporal evolution of influencing factors based on SHAP values (2015–2024).

Year	Primary Positive Factors	Primary Negative Factors	Dominant Factor Type	XGBoost R^2	RMSE
2015	Rainfall (+0.32),	Temperature (−0.24), BSI (−0.22), NBR (−0.17)	Ecological Dominance	0.903	0.035
2019	Rainfall (+0.22)	NBR (−0.34), BSI (−0.18)	Anthropogenic Emergence	0.910	0.034
2024	—	RI (−0.35), BSI (−0.25), NBR (−0.18)	Anthropogenic Dominance	0.950	0.022

By 2019, however, climatic drivers had diminished dominance as drivers of a more anthropogenic nature emerged. SHAP values demonstrated that the Normalized Burn Ratio (NBR) (−0.34) and Bare Soil Index (BSI) (−0.18) became increasingly dominant (Table 4). This emphasizes the occurrence of fire events and exposed soil, which are frequently associated with illegal deforestation and overgrazing. Although rainfall retained a modest positive effect (+0.22), its relative importance decreased dramatically in 2015.

This trend persisted in 2024, where the Road Index (RI) was the highest contributor, with a negative correlation with forest cover (SHAP value of −0.35), followed by NBR (−0.25) and BSI (−0.18) as the next two contributors (Table 4). These shifts suggest that infrastructure expansion, particularly roads, had become a major driver of forest fragmentation and loss by 2024. Roads serve as the entry point for previously inaccessible peat-swamp forests for human-based activities, including illegal logging, expansion of settlements, and construction of secondary roads.

Spatial analysis confirmed this trend, with hotspots of degradation shifting considerably across the study period (Figures 10, 12 and 14). In 2015, the affected regions were largely isolated and attributed to natural stressors such as diminished rainfall or small fire outbreaks. In 2019, hotspots were spread out, particularly in the mid-elevation zones and areas surrounding agricultural land. In 2024, degradation patterns were concentrated along new roads and in the vicinity of urbanized areas, thereby confirming the growing role of anthropogenic activities in driving forest disturbance. The observed hotspots of degradation notably coincide with regions known from local reports to experience intensified human activities such as illegal logging, infrastructure development, and military operations.

To examine the nature, direction, and magnitude of the relationships between forest cover and its drivers, SHAP and GAM were applied to the outputs of the XGBoost model outputs, and only the top five most effective factors were presented in the GAM plot. Interpretability tools were applied to all three temporal snapshots to provide nuanced insights related to variable effects and interactions.

In 2015, GAM analysis indicated that rainfall exhibited a distinct positive linear correlation with FVC. Temperature showed a threshold response in which vegetation health improved when raising temperatures up to 20 °C, after which it declined (Figure 9). These findings were corroborated by the SHAP, with rainfall and temperature being the most significant contributing variables (Figure 10). In contrast, BSI and NBR demonstrated moderately negative effects, suggesting that the early emergence of anthropogenic stress began to manifest.

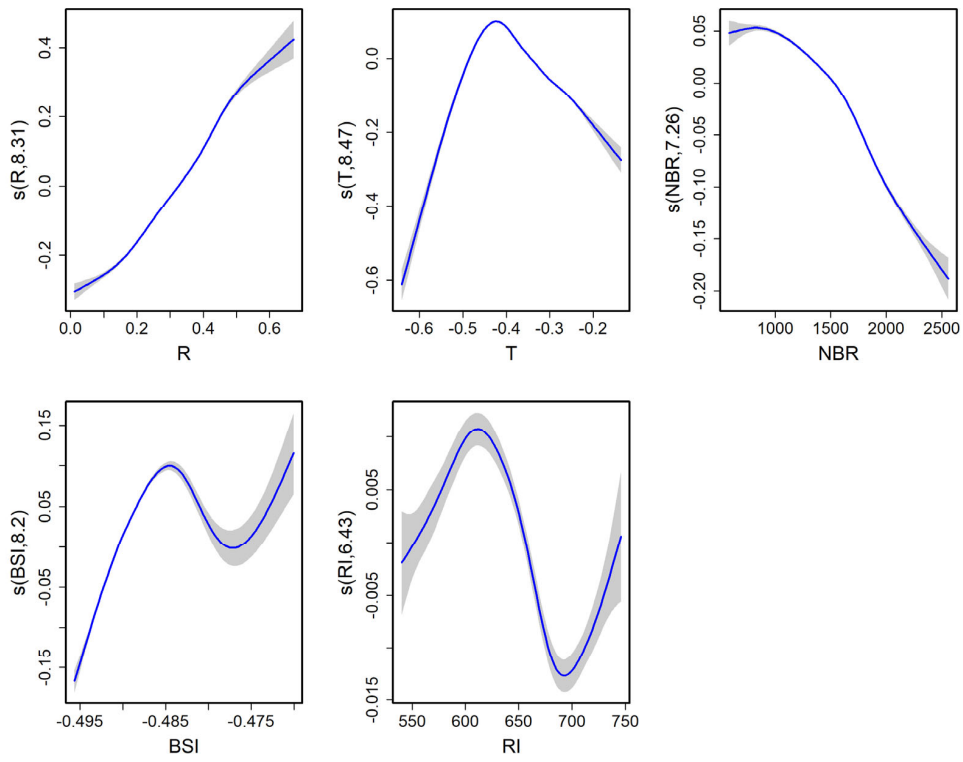


Figure 9. GAM plot showing the correlation pattern between features in 2015.

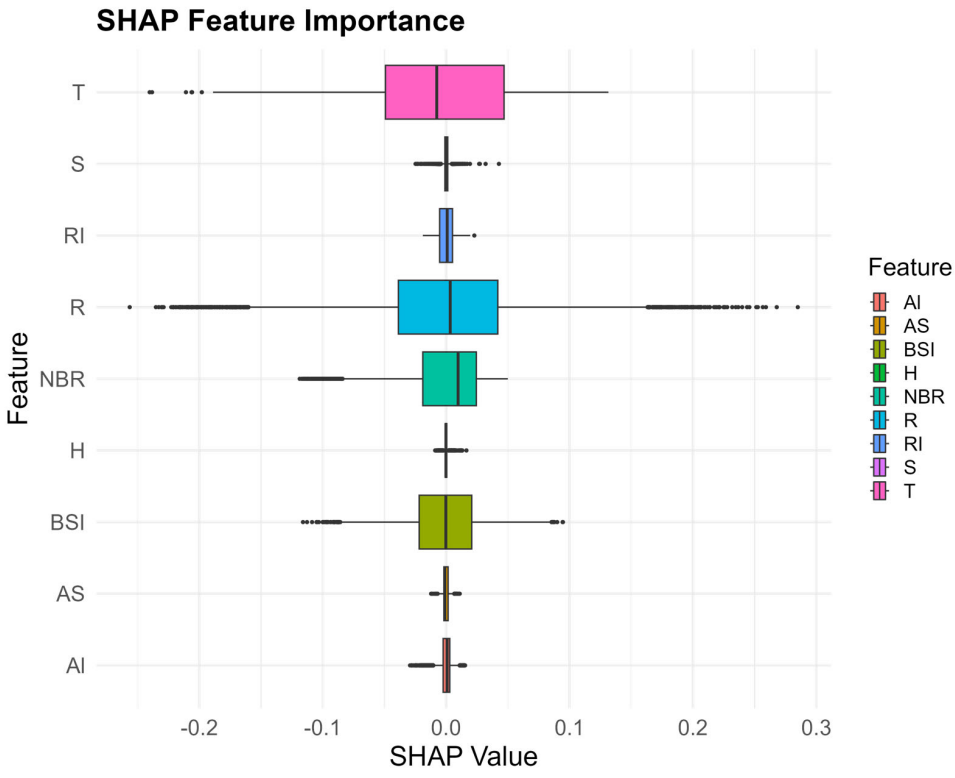


Figure 10. SHAP chart showing the values of important features in 2015.

By 2019, the GAM plots showed a stronger nonlinear influence from BSI and NBR (Figure 11). FVC declined markedly as bare soil exposure and burn severity increased, and SHAP analysis confirmed these variables as the dominant negative contributors (Figure 12). Notably, the influence of rainfall was minimal, which paralleled the transition from natural

to human pressure on the forests. However, aspect and hillshade were not significant contributors during this year.

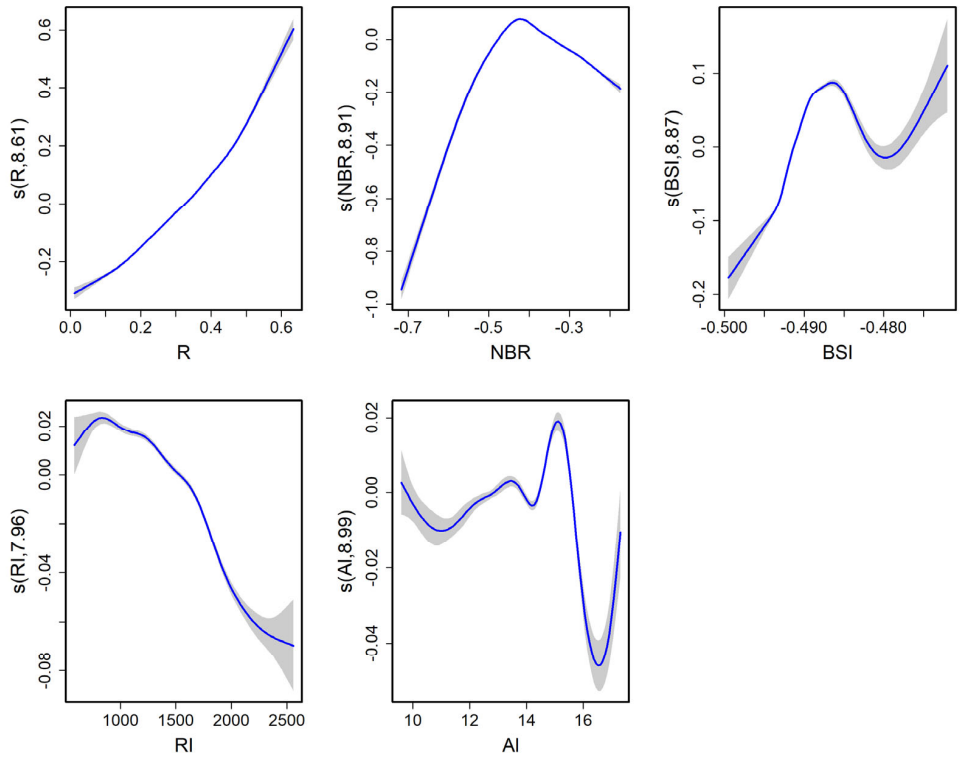


Figure 11. GAM plot showing the correlation pattern between features in 2019.

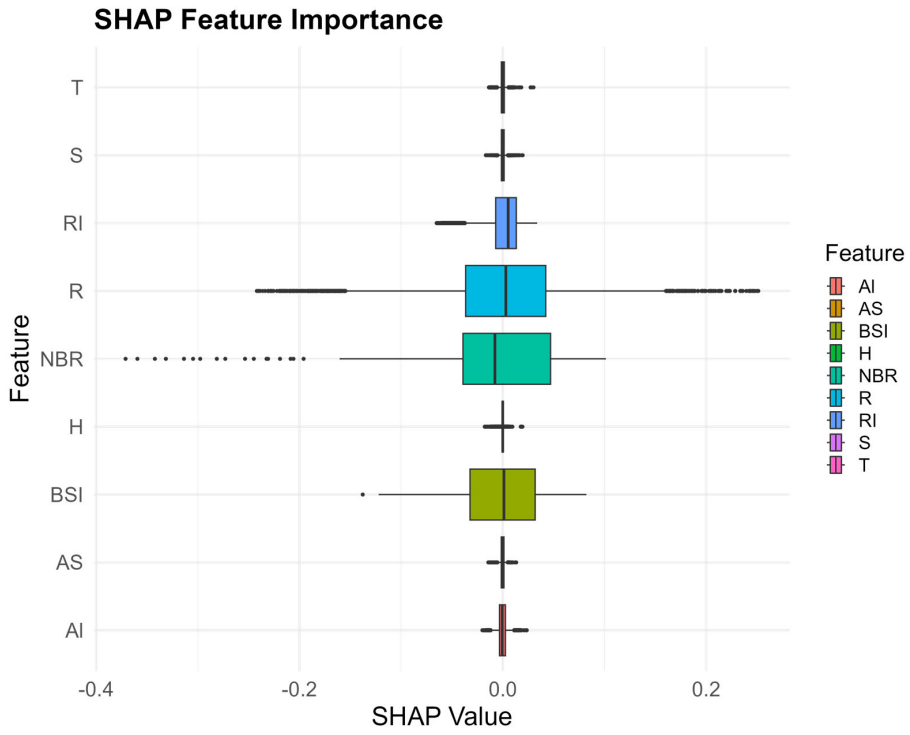


Figure 12. SHAP chart showing the values of important features in 2019.

The interpretability plots for 2024 revealed a decisive shift. In the best-performing linear model (GAM), FVC showed the highest linear decline with respect to road presence (Figure 13), and RI was ranked as the highest negative contributor (−0.36) by SHAP

(Figure 14). The effect of the BSI and NBR remained strongly negative, but their impacts were amplified by ongoing infrastructure expansion. This indicates an important phase in which human activity thwarted natural control over forest dynamics. The minimal influence of topographic variables, such as altitude and hillshade, indicates that anthropogenic pressures became spatially indiscriminate, impacting a wide array of terrains.

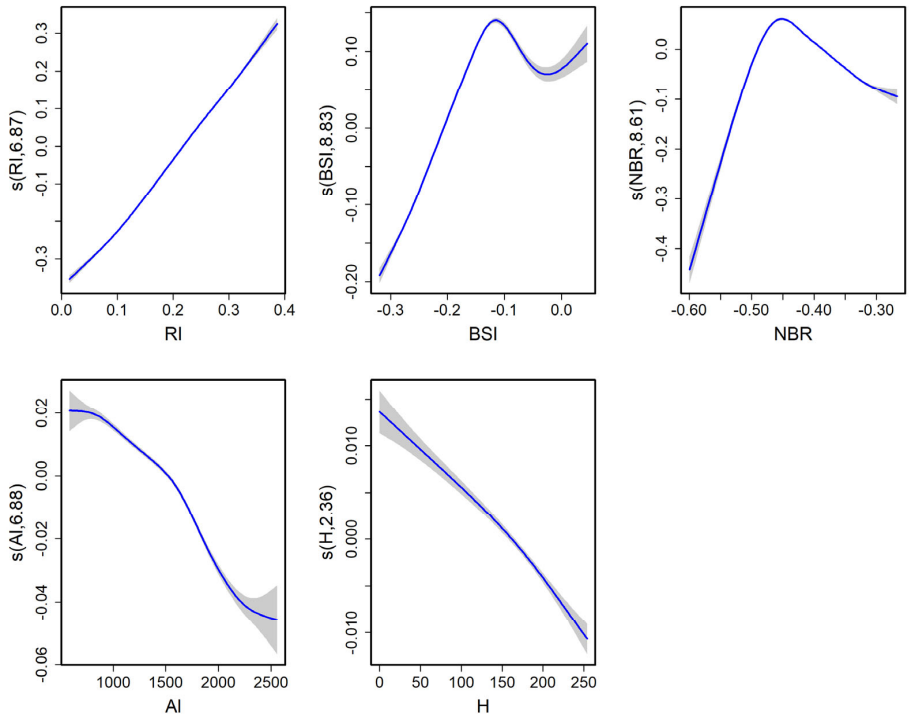


Figure 13. GAM plot showing the correlation pattern between features in 2024.

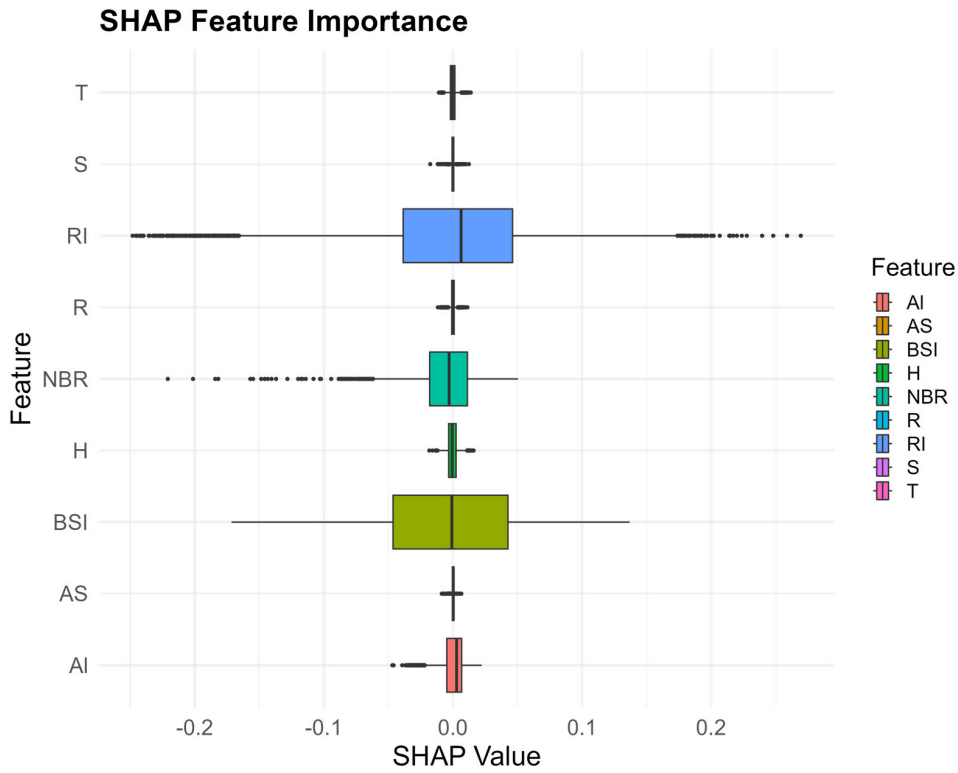


Figure 14. SHAP chart showing the values of important features in 2024.

A comparative visualization in Figures 10 and 14 summarizes the critical transition from climate-driven ecosystem-regulating drivers in 2015 to increasingly complex, fragmented, and human-altered ones in 2024.

4.3. Spatial Predictions, Accuracy Assessment, and Forest Cover Trends

The forest cover maps are shown, with observed and predicted FVC values for each of the three years (Figure 15), along with the respective error maps. The observed FVC maps derived from Sentinel-2 data illustrate the spatial heterogeneity in vegetation density, with greener shades representing denser cover. Moreover, the FVC maps predicted by the XGBoost model showed very similar trends, confirming the predictive accuracy of the model.

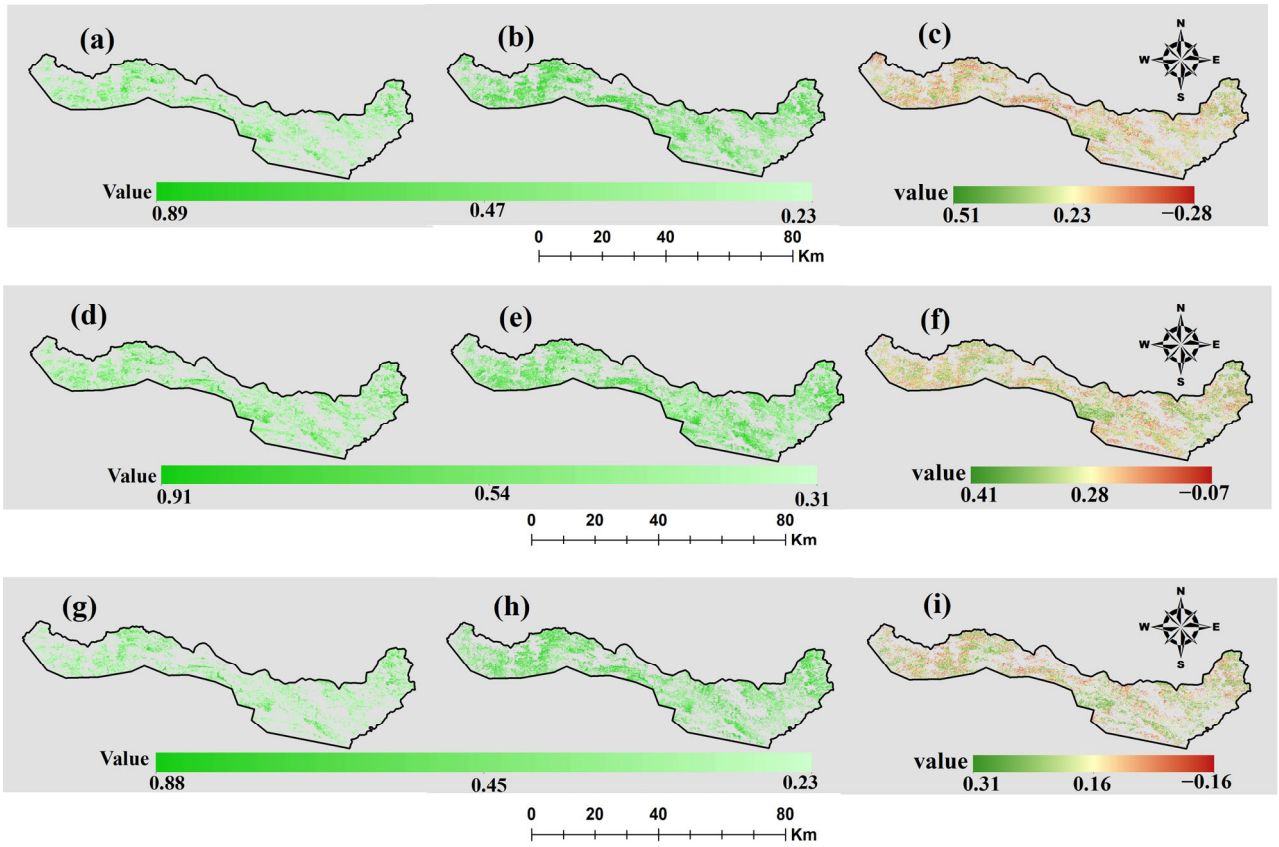


Figure 15. Mapping of forest cover in 2015, 2019, and 2024: satellite-derived data ((**a**) 2015; (**d**) 2019; (**g**) 2024); XGBoost Model predictions ((**b**) 2015; (**e**) 2019; (**h**) 2024); and error maps ((**c**) 2015; (**f**) 2019; (**i**) 2024).

The error maps, which are calculated as the difference between the observed and predicted FVC, are interpreted as patterns of spatial overestimation and underestimation. Positive error values indicated over-prediction, whereas negative values indicated underestimation. The distribution of errors decreased dramatically over time, particularly in 2024, indicating improved accuracy and spatial generalization. These further underscores the reliability of the ML-based approach for mapping dynamic forest cover when data are scarce. The high accuracy of the spatial predictions provided confidence for further quantitative analysis, allowing us to reliably evaluate trends in forest areas and quantify the extent of deforestation.

Forest area defined based on a calibrated FVC threshold showed a constant decrease throughout the period analyzed. As of 2015, the forest area was approximately 630 km^2. This dropped slightly to 625 km^2 in 2019 and then collapsed to 577 km^2 by 2024 (Figure 16).

While the initial decline was subtle, the sharp reduction post-2019 was consistent with the rise in anthropogenic drivers, such as road construction and burn incidents.

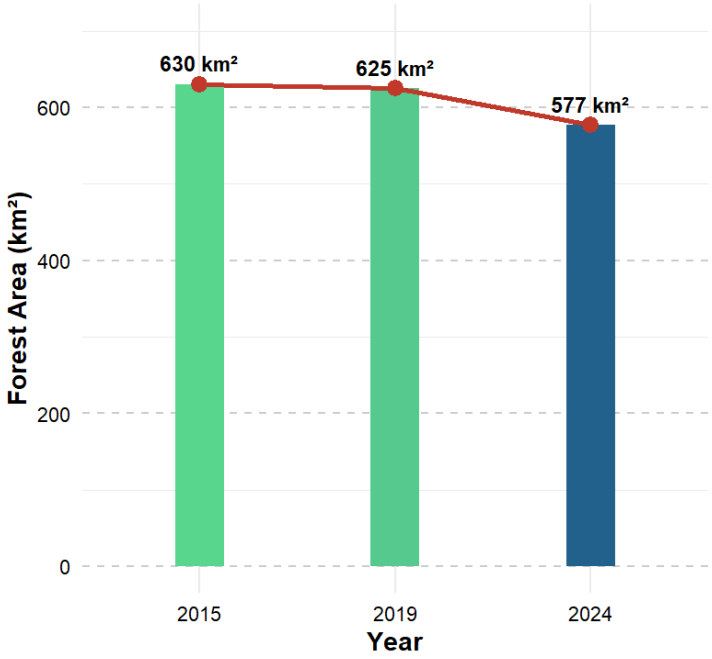


Figure 16. FVC trend over time (2015, 2019, and 2024).

This trend confirms the spatial and interpretive results derived from the model predictions and driver analysis. A forest cover loss of more than 50 km^2 within the last decade indicates a more than 12% decline in forest cover within the study area. This alarming rate of decline is largely due to the ecological sensitivity and geopolitical significance of the study area. This highlights the urgency of implementing evidence-based conservation policies and spatial planning measures to mitigate forest degradation in this area.

Therefore, our findings synthesize rigorous model evaluations and interpretable ML insights using spatial analysis and quantitative forest change assessments. This ultimately provides a comprehensive understanding of the deforestation dynamics in the Iraq–Turkey border region between 2015 and 2024. These results demonstrate the ability to combine remotely accessed data with explainable ML for the monitoring of ecologies within complex geopolitical landscapes. These empirical findings and identified driver shifts clearly necessitate an interpretative discussion to contextualize the results against broader ecological and socio-political frameworks, as addressed in the following section.

5. Discussion

This study aimed to investigate the complex dynamics of deforestation and forest degradation along the Iraq–Turkey border within the Duhok Governorate over ten years (2015–2024). This was achieved by applying interpretable ML models in combination with high-resolution RS datasets, incorporating both climatic and topographic variables. These results provide compelling evidence that deforestation in the study area has transitioned from being predominantly climate-driven in 2015—when rainfall was the most influential variable—to being increasingly governed by anthropogenic factors such as road expansion (RI), burn severity (NBR), and bare soil exposure (BSI) by 2024. This temporal shift was robustly captured using SHAP and GAM and is consistent with independent reports of accelerating infrastructure development and conflict-induced degradation in northern Iraq [56,57].

Predictive modeling showed that the XGBoost model consistently outperformed other ML models with accuracy metrics of R^2 values exceeding 0.90 and low RMSE across all prediction years. This aligns with recent studies highlighting that XGBoost can shape nonlinear ecological phenomena and performs better than a simple regression approach in environmental modeling [7,35,36]. The use of SHAP values and GAM plots further facilitates intuitive visualization of each variable's direction and the magnitude of influence exerted by each variable, thereby aligning the gap between predictive performance and ecological interpretability [49,50].

Alongside the clear gain in the predictive performance of XGBoost, the interpretive insights developed through SHAP and GAM are also noteworthy, clarifying the nuanced drivers of deforestation and forest degradation.

In 2015, climate was the main driver of vegetation dynamics. The strongest relationships with positive impacts on FVC were observed for rainfall and temperature. This is aligned with previous studies analyzing forest health in the Kurdistan Region, where the main drivers of forest vitality were linked to precipitation amount and average temperature [30,58]. This timeframe corresponded to a time of little human disturbance when forest cover was generally stable.

However, by 2019, the ecological balance had begun to tip. Fire- and soil-exposure-related anthropogenic disturbances were among the strongest drivers of forest degradation. These results are consistent with those published by [9,59], who recorded increased fire incidences and land use pressures throughout northern Iraq. These disturbances increasingly occur through the intensification of illegal logging, land clearing, and localized fire outbreaks, particularly in mid-elevation zones close to encroaching settlements.

The situation had intensified by 2024, as this study revealed road infrastructure expansion to be the number one key driver of forest cover loss. The Road Index (RI) was the top negative contributor to FVC, as confirmed by the SHAP values, which further emphasizes the strong correlation between accessibility and forest exploitation results. This trend was corroborated by field observations and reports of widespread road construction by military and civilian actors from 2020 onwards [56,57]. The establishment of over 40 military bases and connecting roads in the Amedi and Zakho districts also enabled deep penetration into wooded areas, which hastened forest fragmentation and area clearing.

The impact of conflict, as a hidden driver of deforestation, is enormous. The study of Eklund and Dinc [10] and a report in IraqClimateChange [56] have highlighted the ecological toll exerted on militarized landscapes, especially in the borderlands of contention. Our findings align with Sur et al. [8], who also observed that machine learning combined with remote sensing effectively identifies vegetation degradation trends. However, unlike their study in India, where climatic factors remained dominant, our results indicate a stronger influence of political and socio-economic disturbances. This discrepancy may be attributed to the militarized and geopolitically sensitive nature of our study area, which accelerates forest loss beyond climatic thresholds. Frequent air strikes, wars, and tourism in some areas cause immediate tree loss, fires, and long-term environmental disturbances. Our findings support these statements, demonstrating an unprecedented increase in NBR and BSI in 2019 and 2024, consistent with reported fire events and burned vegetation.

Livelihood pressure also contributed to the deforestation trends. Well-established studies, such as those by Hassan and Taha [60] and Karim [61], show that local populations are dependent on forests for fuel, timber, and the expansion of agricultural land. These results reflect the consistent BSI into the non-favorable range of the study period, indicating incremental degradation of landscape vegetation through strip clearing of trees and land clearing practices.

Notably, ecological variables (altitude, aspect, and hillshade) had little influence in the later years. This indicates that human-driven degradation has reduced topographic resilience, enabling exploitation in previously untouched higher mountain ranges. Steep slopes and high elevations are traditionally considered natural buffers to disturbances [14,30], and our findings show that roads and military activities have increasingly penetrated these zones, diminishing their protective function. This shift marks a significant departure from conventional forest degradation models, where topographic resilience (e.g., slope, elevation) is typically associated with lower deforestation rates. In our study, however, conflict dynamics and road presence instead seem to overwhelm these natural defenses. This spatial indiscriminateness suggests that military activity and infrastructure expansion are not merely co-occurring with degradation—they are actively enabling forest loss in previously inaccessible highland areas.

Understanding these drivers and how they are changing is critical; however, predicting how the changing drivers manifest spatially is also critical for implementing effective conservation measures. Our spatial modeling framework successfully closes this gap, yielding accurate predictions.

Despite these trends, the modeling approach has demonstrated considerable promise. The excellent congruency of the observed and predicted FVC maps, with low residual errors, emphasized the usefulness of XGBoost in mapping forest cover dynamics. Most importantly, tens of SHAP and GAM were combined to ensure not only accuracy in the results but also ecological interpretability, an aspect often missing from standard black-box ML applications [49,52].

The prediction errors were clustered spatially over time. This further confirms that this modeling framework is well-suited to long-term forest monitoring in data-scarce or geopolitically complex regions.

This study affirms that deforestation along the Iraq–Turkey border stems from a combination of climatic variability, conflict, infrastructure development, and subsistence activities. The paradox revealed in the reduction in the forest area, decreasing from 630 km^2 in 2015 to 577 km^2 in 2024, indicates more than 12% forest loss over less than a decade and underscores the need for land use planning and conflict-sensitive conservation to be undertaken immediately. This methodological framework provides a replicable basis for similar assessments in other geopolitically contested and data-scarce regions.

Future research should address several methodological limitations observed in this study. The absence of direct socio-economic variables—such as population density, fuel consumption, and land tenure systems—was due to limited availability and data reliability in the study area. Incorporating such information, where accessible, or using spatial proxies like settlement proximity or nighttime lights could strengthen future analyses. Another limitation is the reliance on multi-year Sentinel-2 imagery, which may overlook short-term or seasonal forest disturbances. Access to higher-temporal-resolution datasets would enable more detailed trend detection. Furthermore, the lack of direct ground-truthing, constrained by regional insecurity, remains a limitation. While a robust sampling design was applied, future studies could benefit from collaborative field validation involving local communities or remote-sensing-based verification protocols. Overall, future efforts should aim to integrate socio-economic and conflict-related data into predictive models and work closely with local stakeholders to produce actionable, context-sensitive conservation strategies. The rapid pace of forest loss emphasizes the need for targeted policy and management responses, supported by interpretable model outputs and a comprehensive understanding of local drivers.

6. Conclusions

This study presents a comprehensive, interpretable analysis of deforestation and forest degradation along the Iraq–Turkey border within the Duhok Governorate over a decade (2015–2024). Using high-resolution Sentinel-2 imagery, climatic and topographic data, and advanced machine learning, it demonstrates the effectiveness of explainable artificial intelligence in complex, data-limited regions.

Among the tested algorithms, XGBoost achieved the highest performance ($R^2 > 0.90$), with minimal prediction errors. SHAP and GAM improved model interpretability and revealed a clear temporal shift in forest drivers—from climate-related factors (rainfall, temperature) in 2015 to anthropogenic pressures (fires, road construction, land clearing) by 2024. This shift was strongly linked to conflict-related development and infrastructure encroachment. Spatial modeling showed alignment between degradation hotspots and road expansion, with forest area declining by 12% (from 630 km^2 to 577 km^2). These patterns confirm the reported trends of militarization, illegal logging, and fire-induced degradation.

This study highlights the compounded effect of ecological vulnerability and political instability on forest loss. It also shows that forest degradation in geopolitically sensitive regions requires more than ecological monitoring. It requires integrating socio-political insights into predictive analysis.

Regulating road construction, strengthening legal enforcement against deforestation, and involving local communities in forest monitoring and restoration are critical. Public policies should combine environmental conservation with conflict-aware land management and access to higher temporal and socio-economic data.

The methodological framework developed here offers a reproducible, robust approach for forest monitoring. It integrates predictive analysis with ecological interpretation and holds significant potential to inform evidence-based policies in fragile landscapes.

Author Contributions: Conceptualization, M.H.A. and Y.T.M.; methodology, M.H.A. and Y.T.M.; software, M.H.A. and Y.T.M.; validation, M.H.A. and Y.T.M.; formal analysis, M.H.A.; investigation, M.H.A.; resources, M.H.A. and Y.T.M.; data curation, M.H.A.; writing—original draft preparation, M.H.A.; writing—review and editing, Y.T.M. visualization, M.H.A.; supervision, Y.T.M.; project administration, Y.T.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data were not made publicly available owing to privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. FAO. *Global Forest Resources Assessment 2020*; FAO: Rome, Italy, 2020. [CrossRef]
2. Shimabukuro, Y.E.; Arai, E.; Duarte, V.; Jorge, A.; dos Santos, E.G.; Gasparini, K.A.C.; Dutra, A.C. Monitoring deforestation and forest degradation using multi-temporal fraction images derived from Landsat sensor data in the Brazilian Amazon. *Int. J. Remote Sens.* **2019**, *40*, 5475–5496. [CrossRef]
3. Ketema, T.; Lemi, G.; Liban, J. GIS and Remote Sensing Based Spatiotemporal Analysis of Deforestation in Dati Wolel National Park, Western Ethiopia. *Am. J. Agric. For.* **2022**, *10*, 1–8. [CrossRef]
4. Yousefi, S.; Pourghasemi, H.R.; Avand, M.; Janizadeh, S.; Tavangar, S.; Santosh, M. Assessment of land degradation using machine-learning techniques: A case of declining rangelands. *Land Degrad. Dev.* **2021**, *32*, 1452–1466. [CrossRef]
5. Debus, A.; Beauchamp, E.; Acworth, J.; Ewolo, A.; Kamga, J.; Verhegghen, A.; Zébazé, C.; Lines, E.R. A labelled dataset to classify direct deforestation drivers from Earth Observation imagery in Cameroon. *Sci. Data* **2024**, *11*, 564. [CrossRef]
6. Aziz, G.; Minallah, N.; Saeed, A.; Frnda, J.; Khan, W. Remote sensing based forest cover classification using machine learning. *Sci. Rep.* **2024**, *14*, 69. [CrossRef] [PubMed]
7. Zhao, J.; Zhang, C.; Wang, J.; Abbas, Z.; Zhao, Y. Machine learning and SHAP-based susceptibility assessment of storm flood in rapidly urbanizing areas: A case study of Shenzhen, China. *Geomat. Nat. Hazards Risk* **2024**, *15*, 2311889. [CrossRef]

8. Sur, K.; Verma, V.K.; Panwar, P.; Shukla, G.; Chakravarty, S.; Nath, A.J. Monitoring vegetation degradation using remote sensing and machine learning over India–a multi-sensor, multi-temporal and multi-scale approach. *Front. For. Glob. Change* **2024**, *7*, 1382557. [CrossRef]
9. Habeeb, H.N.; Mustafa, Y.T. Deep Learning-Based Prediction of Forest Cover Change in Duhok, Iraq: Past and Future. *Forestist* **2025**, *75*, 1–13. [CrossRef]
10. Eklund, L.; Dinc, P. Fires as collateral or means of war: Challenges of environmental peacebuilding in the Kurdistan Region of Iraq. *Ecol. Soc.* **2024**, *29*, 25. [CrossRef]
11. Al-Khudhairy, A.A.; Al-Timimi, Y.K. Analysis of the LST and Vegetation Indices relationship using Landsat-8 data in Duhok Governorate, Iraq. *Al-Mustansiriyah J. Sci.* **2021**, *32*, 6–12. [CrossRef]
12. Mustafa, Y.; Salar, S.G.; Khatab, A.; Ahmed, F.; Habeeb, H. *Mapping and Estimating Vegetation Coverage in Iraqi Kurdistan Region Using Remote Sensing and GIS*; Iraq-Final Report; Horticulture, Forestry, and Rangeland, Ministry of Agriculture and Water Resources: Erbil, Kurdistan Region, Iraq, 2015; pp. 1–136.
13. Tahir, S.A.; Uzun, A.; Galalaey, A.M.K. Plant Biodiversity in the Old World (A Case Study for Amedi District in the Northern Iraq). In Proceedings of the 3rd International Symposium on EuroAsian Biodiversity, Minsk, Belarus, 5–8 July 2017.
14. Vaca, R.A.; Golicher, D.J.; Rodiles-Hernández, R.; Castillo-Santiago, M.Á.; Bejarano, M.; Navarrete-Gutiérrez, D.A. Drivers of deforestation in the basin of the Usumacinta River: Inference on process from pattern analysis using generalised additive models. *PLoS ONE* **2019**, *14*, e0222908. [CrossRef]
15. Najafifar, A.; Hosseinzadeh, J.; Karamshahi, A. The role of hillshade, aspect, and toposhape in the woodland dieback of arid and semi-arid ecosystems: A case study in Zagros woodlands of Ilam province, Iran. *J. Landsc. Ecol.* **2019**, *12*, 79–91. [CrossRef]
16. Gkatzioura, P.E.; Perakis, K. Analysis of Urban Heat Island (UHI) through Climate Engine and Arcgis Pro in Different Cities of Bulgaria. In Proceedings of the Eighteenth International Scientific Conference, Sofia, Bulgaria, 24–27 June 2022; pp. 19–21.
17. Huntington, J.L.; Hegewisch, K.C.; Daudert, B.; Morton, C.G.; Abatzoglou, J.T.; McEvoy, D.J.; Erickson, T. Climate engine: Cloud computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 2397–2410. [CrossRef]
18. Lin, X.; Chen, J.; Lou, P.; Yi, S.; Qin, Y.; You, H.; Han, X. Improving the estimation of alpine grassland fractional vegetation cover using optimized algorithms and multi-dimensional features. *Plant Methods* **2021**, *17*, 96. [CrossRef]
19. Gao, L.; Wang, X.; Johnson, B.A.; Tian, Q.; Wang, Y.; Verrelst, J.; Mu, X.; Gu, X. Remote sensing algorithms for estimation of fractional vegetation cover using pure vegetation index values: A review. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 364–377. [CrossRef]
20. Roy, D.P.; Zhang, H.; Ju, J.; Gomez-Dans, J.L.; Lewis, P.E.; Schaaf, C.; Sun, Q.; Li, J.; Huang, H.; Kovalskyy, V. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sens. Environ.* **2016**, *176*, 255–271. [CrossRef]
21. Alcaras, E.; Costantino, D.; Guastaferro, F.; Parente, C.; Pepe, M. Normalized Burn Ratio Plus (NBR+): A new index for Sentinel-2 imagery. *Remote Sens.* **2022**, *14*, 1727. [CrossRef]
22. Nguyen, C.T.; Chidthaisong, A.; Kieu Diem, P.; Huo, L.-Z. A modified bare soil index to identify bare land features during agricultural fallow-period in southeast Asia using Landsat 8. *Land* **2021**, *10*, 231. [CrossRef]
23. Wentzel, K. Determination of the overall soil erosion potential in the Nsikazi District (Mpumalanga Province, South Africa) using remote sensing and GIS. *Can. J. Remote Sens.* **2002**, *28*, 322–327. [CrossRef]
24. Reddy, S.L.K.; Rao, C.; Kumar, P.R.; Anjaneyulu, R.; Krishna, B.G. An index based road feature extraction from LANDSAT-8 OLI images. *Int. J. Electr. Comput. Eng.* **2021**, *11*, 1319.
25. Ahmed, M.W.; Saadi, S.; Ahmed, M. Automated road extraction using reinforced road indices for Sentinel-2 data. *Array* **2022**, *16*, 100257. [CrossRef]
26. Rouse, J.W.; Haas, R.H.; Deering, D.; Schell, J.; Harlan, J.C. *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*; Texas A&M University: College Station, TX, USA, 1974.
27. Zeng, X.; Dickinson, R.E.; Walker, A.; Shaikh, M.; DeFries, R.S.; Qi, J. Derivation and evaluation of global 1-km fractional vegetation cover data for land modeling. *J. Appl. Meteorol.* **2000**, *39*, 826–839. [CrossRef]
28. Ch, K. Landscape assessment: Ground measure of severity, the Composite Burn Index; and remote sensing of severity, the Normalized Burn Ratio. In *Firemon: Fire Effects Monitoring and Inventory System*; CreateSpace: Scotts Valley, CA, USA, 2004.
29. Habeeb, H.N.; Mustafa, Y.T. Analyzing the Impact of Geoenvironmental Factors on the Spatiotemporal Dynamics of Forest Cover via Random Forest. *Earth* **2025**, *6*, 3. [CrossRef]
30. Mzuri, R.T.; Omar, A.A.; Mustafa, Y.T. Spatiotemporal analysis of vegetation cover and its response to terrain and climate factors in Duhok Governorate, Kurdistan Region, Iraq. *Iraqi Geol. J.* **2021**, *54*, 110–126. [CrossRef]
31. Sissakian, V.K.; Kadhim, T.H.; Jab'bar, M.F.A. Geomorphology of the high folded zone. *Iraqi Bull. Geol. Min.* **2014**, *6*, 7–51.
32. Singh, M.; Shahina, N.; Das, S.; Arshad, A.; Siril, S.; Barman, D.; Mog, U.; Panwar, P.; Shukla, G.; Chakravarty, S. Forest resources of the world: Present status and future prospects. In *Land Degradation Neutrality: Achieving SDG 15 by Forest Management*; Springer: Singapore, 2022; pp. 1–23.

33. Xian, Y.; Wang, T.; Yu, P.; Du, Y. To Ignore or Not: Understanding the Influence of Hillshade. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 9009–9017. [CrossRef]
34. Jang, E.; Kim, Y.J.; Im, J.; Park, Y.-G.; Sung, T. Global sea surface salinity via the synergistic use of SMAP satellite and HYCOM data based on machine learning. *Remote Sens. Environ.* **2022**, *273*, 112980. [CrossRef]
35. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
36. Rash, A.; Mustafa, Y.; Hamad, R. Quantitative assessment of Land use/land cover changes in a developing region using machine learning algorithms: A case study in the Kurdistan Region, Iraq. *Heliyon* **2023**, *9*, e21253. [CrossRef]
37. Bui, Q.-T.; Chou, T.-Y.; Hoang, T.-V.; Fang, Y.-M.; Mu, C.-Y.; Huang, P.-H.; Pham, V.-D.; Nguyen, Q.-H.; Anh, D.T.N.; Pham, V.-M.; et al. Gradient boosting machine and object-based CNN for land cover classification. *Remote Sens.* **2021**, *13*, 2709. [CrossRef]
38. Kim, S.-J.; Bae, S.-J.; Jang, M.-W. Linear regression machine learning algorithms for estimating reference evapotranspiration using limited climate data. *Sustainability* **2022**, *14*, 11674. [CrossRef]
39. Zhang, G.; Wang, M.; Liu, K. Forest fire susceptibility modeling using a convolutional neural network for Yunnan province of China. *Int. J. Disaster Risk Sci.* **2019**, *10*, 386–403. [CrossRef]
40. Liu, Z.; Peng, C.; Work, T.; Candau, J.-N.; DesRochers, A.; Kneeshaw, D. Application of machine-learning methods in forest ecology: Recent progress and future challenges. *Environ. Rev.* **2018**, *26*, 339–350. [CrossRef]
41. Woznicki, S.A.; Baynes, J.; Panlasigui, S.; Mehaffey, M.; Neale, A. Development of a spatially complete floodplain map of the conterminous United States using random forest. *Sci. Total Environ.* **2019**, *647*, 942–953. [CrossRef] [PubMed]
42. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [CrossRef]
43. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.
44. Dewi, C.; Chen, R.-C. Random forest and support vector machine on features selection for regression analysis. *Int. J. Innov. Comput. Inf. Control* **2019**, *15*, 2027–2037.
45. Lee, H.; Wang, J.; Leblon, B. Using linear regression, random forests, and support vector machine with unmanned aerial vehicle multispectral images to predict canopy nitrogen weight in corn. *Remote Sens.* **2020**, *12*, 2071. [CrossRef]
46. Mas, J.F.; Flores, J.J. The application of artificial neural networks to the analysis of remotely sensed data. *Int. J. Remote Sens.* **2008**, *29*, 617–663. [CrossRef]
47. Maurya, P.; Kumar, A. Performance assessment of k-nearest neighbor algorithm for classification of forest cover type. In Proceedings of the Third International Conference on Sustainable Computing: SUSCOM 2021, Jaipur, India, 19–20 March 2021; Springer: Singapore, 2022; pp. 43–51.
48. Guo, G.; Wang, H.; Bell, D.; Bi, Y.; Greer, K. KNN model-based approach in classification. In *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE, Proceedings of the OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, 3–7 November 2003*; Proceedings, 2003; Springer: Berlin/Heidelberg, Germany, 2003; pp. 986–996.
49. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017.
50. Guo, R.; Yang, B.; Guo, Y.; Li, H.; Li, Z.; Zhou, B.; Hong, B.; Wang, F. Machine learning-based prediction of outdoor thermal comfort: Combining Bayesian optimization and the SHAP model. *Build. Environ.* **2024**, *254*, 111301. [CrossRef]
51. Chang, I.; Park, H.; Hong, E.; Lee, J.; Kwon, N. Predicting effects of built environment on fatal pedestrian accidents at location-specific level: Application of XGBoost and SHAP. *Accid. Anal. Prev.* **2022**, *166*, 106545. [CrossRef]
52. García, M.V.; Aznarte, J.L. Shapley additive explanations for NO2 forecasting. *Ecol. Inform.* **2020**, *56*, 101039. [CrossRef]
53. O'brien, R.M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **2007**, *41*, 673–690. [CrossRef]
54. Wei, L.; Yuan, Z.; Wang, Z.; Zhao, L.; Zhang, Y.; Lu, X.; Cao, L. Hyperspectral inversion of soil organic matter content based on a combined spectral index model. *Sensors* **2020**, *20*, 2777. [CrossRef] [PubMed]
55. Taghizadeh-Mehrjardi, R.; Schmidt, K.; Amirian-Chakan, A.; Rentschler, T.; Zeraatpisheh, M.; Sarmadian, F.; Valavi, R.; Davatgar, N.; Behrens, T.; Scholten, T. Improving the spatial prediction of soil organic carbon content in two contrasting climatic regions by stacking machine learning models and rescanning covariate space. *Remote Sens.* **2020**, *12*, 1095. [CrossRef]
56. IraqClimateChange. The Politics of Environment in the Kurdistan Region: Unveiling the Crisis of Deforestation and Forest Fires. Available online: https://iraqclimatechange.org/the-politics-of-environment-in-the-kurdistan-region-unveiling-the-crisis-of-deforestation-and-forest-fires/?utm (accessed on 3 February 2025).
57. Rudaw. Turkish Bombardment Sparks Fire in Duhok's Amedi District. Available online: https://www.rudaw.net/english/kurdistan/270720218 (accessed on 25 October 2024).
58. Al-Hedny, S.M.; Muhaimeed, A.S. Drought monitoring for Northern Part of Iraq using temporal NDVI and rainfall indices. In *Environmental Remote Sensing and GIS in Iraq*; Springer: Cham, Switzerland, 2019; pp. 301–331.
59. Habeeb, H.N.; Mustafa, Y.T. Deep learning approaches for estimating forest vegetation cover and exploring influential ecosystem factors. *Earth Sci. Inform.* **2024**, *17*, 3379–3396. [CrossRef]

60. Hassan, R.A.; Taha, M.A. Determination of Heating Value of Quercus Aegilops and Populus Nigra Growing in Kurdistan-Iraq for an Effective Woody Biomass Utilization. *J. Duhok Univ.* **2019**, *22*, 61–70.
61. Karim, S.I. Iraqi Kurdistan People Cutting Trees to Secure Their Home Heating Fuel. Available online: https://theinsightinternational.com/kurdistan-trees-heating-2016-10-12 (accessed on 12 December 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.