A Novel Adaptive Ensemble Learning Approach for Driving Factor Identification in Land Degradation

Gangamma Hediyalad
Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Affiliated to Visvesvaraya Technological University, Belagavi- 590018,India
gangammah_12@rediffmail.com (corresponding author)

K. Ashoka
Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Visvesvaraya Technological University, Belagavi-590018, India
ashokakukkuvada@gmail.com

Received: 11 May 2025 | Revised: 28 June 2025, 24 July 2025, 1 August 2025, and 6 August 2025 | Accepted: 9 August 2025

 DOI: https://doi.org/10.48084/etasr.12057

ABSTRACT

Land degradation is an urgent global concern, driven by complex interactions among environmental and anthropogenic factors. Identifying and understanding the optimal factors driving land degradation are essential for effective prediction and sustainable land management. This study introduces a novel Adaptive Ensemble Driving Factor Learning Model (AEDFLM) designed to identify and validate the most significant factors affecting land degradation. The model incorporates a Random Forest Feature Importance Module (RFFIM) for feature selection, a Driving Factor Bootstrap Sampler (DFBS) to ensure stability and robustness through random sampling, and a Driving Factor Expansion Module (DFEM) for scalable and efficient learning. Besides these core ensemble components, Support Vector Regressor (SVR) and Linear Regressor are employed to assess the influence of specific features, offering insights into both non-linear and linear interactions of less significant driving factors. The model's predictive accuracy was evaluated using Mean Squared Error (MSE) and R-squared (R^2), while feature importance and permutation importance analyses validated the significance of the chosen factors. The model can be further expanded to include predictions for the Desertification Vulnerability Index (DVI), thereby enhancing its broader applicability in assessing land degradation risks across various geographical regions.

Keywords-land degradation; driving factor; AEDFLM; RFFIM; SVR

I. INTRODUCTION

Global climate change is transforming ecosystems and societies, worsening issues like pandemics, food insecurity, and political instability. With alarming rises in land surface temperatures and decreasing humidity, nearly 40% of the world's croplands and pastures are at risk of degradation [1]. Projections indicate that global food demand will increase by 110% by 2050 [2, 3], putting unprecedented pressure on agricultural systems. At the same time, rising temperatures, changing snow-water equivalents, and greenhouse gas emissions are endangering the fragile climate balance needed for crop productivity [4-6].

Land degradation—caused by both environmental and anthropogenic factors—reduces soil fertility, hampers economic growth, and worsens food insecurity. This issue is particularly critical in regions like Ukraine, which is transitioning to an open agricultural land market. In this context, accurate and timely assessment of land degradation and productivity becomes imperative. Remote sensing technologies provide an effective means of deriving essential vegetation indices from multispectral satellite imagery. Indicators such as the Normalized Difference Vegetation Index (NDVI), Leaf Area Index (LAI), and Net Primary Productivity (NPP) are essential for monitoring vegetation health and evaluating agroecological conditions [7]. The integration of NDVI with climatic data further enhances productivity evaluations, contributing to the refinement of global methodologies, such as UN SDG 15.3.1, by incorporating local agroclimatic variations.

Desertification—a severe form of land degradation—is a pressing concern, especially in arid and semi-arid regions, where it results in biodiversity loss, soil erosion, and diminished agricultural productivity. Predicting land degradation requires understanding its multifaceted drivers, including soil moisture, precipitation, wind speed, slope, and human pressures such as population density and land-use

change. Traditional models often fail to capture the nonlinear, dynamic interactions among these variables [8]. Land degradation is also caused by human activities, such as intensive mining and industrial growth [9].

To overcome these limitations, Machine Learning (ML) has become a powerful tool capable of handling large, multidimensional datasets and uncovering complex patterns. While models like Random Forests (RF) and Gradient Boosting Machines (GBM) have demonstrated promise for predicting degradation risk, they often lack interpretability and robustness across diverse landscapes and temporal scales. Furthermore, they may overfit or yield inconsistent results due to variability in environmental datasets. This study presents an innovative hybrid framework to enhance the accuracy and stability of land degradation forecasts. Key contributions encompass:

- Ensemble-Based Feature Selection and Modeling (EFSM). A novel ensemble method is developed, which integrates RF for feature selection, employs bagging to enhance robustness, and utilizes the DFEM to capture complex nonlinear relationships among variables. DFEM allows scalable modeling of high-order interactions between ecological, climatic, and topographic drivers of land degradation.
- DFBS is introduced to enhance factor consistency through Bootstrap Sampling, reducing overfitting and improving model generalizability.
- Critical Factor Identification. The framework provides a methodical approach to identifying key drivers such as wind 10m(u) and soil water L1, enabling precise, context-specific land management interventions.
- Enhanced Prediction Reliability. The model emphasizes interpretability and stability, providing actionable insights for sustainable land-use planning.

This research leverages the DVI, a well-known measure for evaluating land susceptibility to degradation [10]. Precise identification of driving factors enables targeted mitigation strategies, especially in high-risk areas.

A wide range of studies have explored the use of remote sensing, statistical modeling, and ML to monitor and predict desertification [11-13]. In [14], point-to-point and point-to-line models were developed using Landsat 8 OLI and Albedo-Modified Soil Adjusted Vegetation Index (MSAVI) to monitor desertification in Naiman Banner. The point-to-line model, which better captured non-linear albedo–vegetation relationships and soil background effects, outperformed the point-to-point model with 93.8% accuracy, compared with 88.9%. Both methods worked well for moderate desertification, but the point-to-line model was better suited to severe cases. In [15], advanced remote sensing techniques were used to assess land cover changes and desertification in Mongolia's Hogno Khaan protected area, using indicators such as NDVI, Topsoil Grain Size Index (TGSI), and albedo to represent vegetation biomass, landscape pattern, and micrometeorology. A Decision Tree (DT) analysis of data from 1990, 2002, and 2011 showed that desertification increased over time, affecting 15.3% of the area, especially in sand dunes, pastures, and grasslands. In [16], multi-year remote sensing data was used to identify desertification zones in Central Asia, especially in Kazakhstan, by utilizing Integral NDVI data from NOAA AVHRR satellite imagery. This method allowed for detecting desertification trends during the growing season, dating back to 2000. In [17], the study examined desertification dynamics on the Mongolian Plateau using intensity analysis and the geographical detector method to identify root causes. The aim was to pinpoint the primary factors contributing to land degradation in this area. In [18], the study examined desertification trends in Algeria's Nemamcha region by applying a multivariate linear model to link Landsat spectral data with soil samples collected from ground surveys. The study revealed a strong link between soil properties —such as sand, silt, and clay content —and desertification, highlighting vegetation decline and land cover changes. In [19], surface alterations in the Azraq Oasis, Jordan, were tracked using the Tasseled Cap (TC) transform and Change Vector Analysis (CVA) across a 30-year span from 1984 to 2013. The results indicated the capacity of remote sensing techniques to identify desertification in arid regions by assessing the extent and trajectory of landscape changes. In [20], remote sensing methods, including Landsat TM imagery and change vector analysis, were used to track desertification in a semi-arid region of Mexico. The research used a desertification degree index (DDI) to assess the extent of desertification, finding that 49% of the area exhibited some level of desertification, thereby identifying regions at risk.

In [21], the Grain Size Index (GSI) of topsoil was used to identify desertification in Inner Mongolia, China, based on Landsat TM images from 1993 and 2000. The study found an increase in fines and content, indicating topsoil coarsening and desertification driven by human activities. In [22], remote sensing and GIS technology were used to study desertification in northwest China and its effects on sustainable development. The study proposed an evaluation system for desertification sensitivity and developed a model to compute a desertification sensitivity index for monitoring desertification risks. In [23], the MEDALUS model was used to assess desertification risk in northeastern Iran, selecting eight remote sensing indicators that showed a strong correlation with field data for further study. Four machine learning methods—Support Vector Machine (SVM), GBM, Generalized Linear Models (GLM), and RF—were employed to assess desertification risk, with an ensemble model's weighted average predicting the area's susceptibility to desertification [24]. These studies demonstrate the variety of methods used to identify the main causes of desertification and land degradation through remote sensing, machine learning, and multi-temporal analysis. However, none of these approaches offer a comprehensive framework that integrates feature selection, stability improvement, and model scalability.

This study aims to fill these gaps by creating the AEDFLM, designed to identify and confirm the key driving factors behind land degradation [25, 26]. The study uses LU/LC data from Landsat satellite images across four periods—1993, 2003, 2013, and 2024—to effectively capture historical land cover changes. This temporal dataset, enhanced through spectral modeling, offers a solid basis for precise calibration and validation of the simulation model. Furthermore, the research

adopts a hybrid modeling framework that integrates spectral indices, machine learning-generated transition potential maps, and the CA-Markov model. This integrated approach markedly improves the accuracy and dependability of LU/LC change prediction [27].

II. PROPOSED METHODOLOGY

In this study, the RFFIM combined with the DFBS is used to improve the performance of the DFEM within the Enhanced Driving Factor Prediction Model (EDFPM). RFFIM selects key attributes, while DFBS creates multiple sub-training datasets by randomly sampling them. Each subset is used to build individual DFEM models, which are then integrated to generate the final prediction, as shown in Figure 1. To further analyze the impact of less significant factors identified by RFFIM—such as slope and population density—we use SVR and Linear Regression. These models identify both non-linear and linear interactions, providing more detailed insights into how individual factors contribute to land degradation. During the integration phase, each DFEM model functions as a weak learner within the ensemble framework. The diagram given above describes the phases of the model, starting with choosing features, and various factors are considered for the driving factor. Although a few features may have decreased relevance for evaluating the spread and may also recur, this reduces the model's ability to learn efficiently. We implement the RFFIM method described in algorithm 1 for selecting the most essential features.

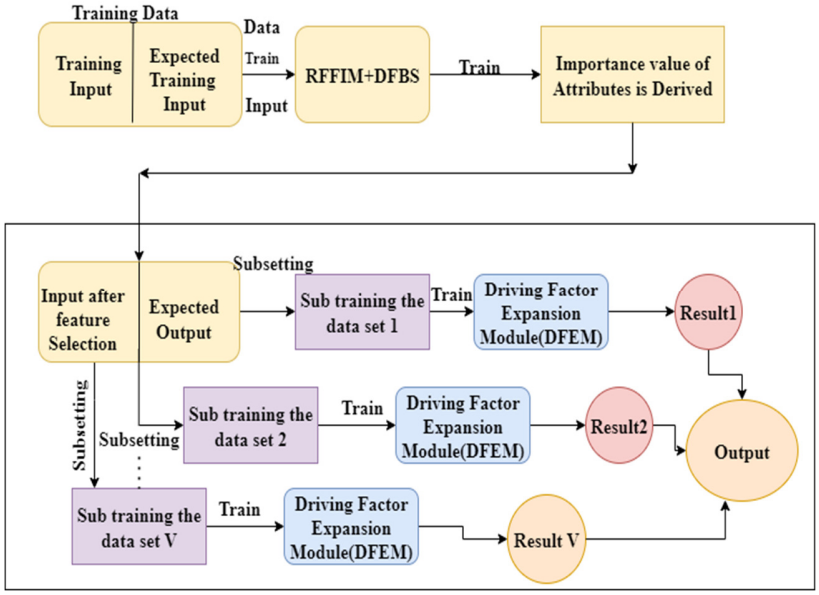

Fig. 1. Proposed Adaptive Ensemble Driving Factor Learning Model (AEDFLM)

In the next phase, the entire dataset is split into training and test sets. Samples are taken from the training data using the DFBS method. This sampling process is repeated to prepare the sub-training data. In the final phase, when the model is integrated, each sub–DFEM model is considered weak for learning. Therefore, multiple weak models are combined to form a strong model. As a result, the output is given as:

$$a_r = \left(a_{r_1} + a_{r_2} + \cdots . + a_{rv}\right)(V)^{-1} \quad (1)$$

where a_{r_k} is the value that is predicted for the $k-th$ learning model, while producing the output a_r as the resulting expected value.

A. Driving Factor Expansion Module (DFEM)

The proposed study implements an enhanced DFEM featuring a flat structural network, eliminating the necessity for a deep neural architecture. This architecture reduces the complexity of the training process, resulting in faster approximations and helping achieve results that compete with those of deep learning networks. These learning methods have a universal approximation ability, and the global loss function can be determined. Additionally, the DFEM can effectively and efficiently update the model with new attributes or information. Considering the DFEM discussed in this paper, the attributes are initially retrieved from the training data utilizing the mapped attribute elements. Furthermore, the adopted attributes are transformed into improved elements. Lastly, the linear fusion of all mapped attributes, along with the enhanced elements, forms the output. It is known through prior studies that the weights used for connection in a DFEM are formulated using ridge regression algorithm.

Let's assume we have a training dataset provided as: $z(k), a(k) \mid z(k)$ belongs to $\mathbb{T}^F$, $a(k)$ belongs to $\mathbb{T}^E$, where k=1,2,…,P, here the sizes of every sample is denoted as F and their respective outputs is given as E. Furthermore, the pattern for the input is given as Z belongs to $\mathbb{T}^{u\times F}$, A belongs to $\mathbb{T}^{u\times E}$, here $Z = [z(1), z(2), \ldots . , z(u)]^V, A = [a(1), a(2), \ldots . , a(u)]$, the count of samples used as input is shown as u. During the phase of learning, the matrix used as input Z is mapped onto attribute elements p by p feature mapping θ_k for generation of

random attributes. We formulate the mapped attribute B_k using feature functions.

$$B_k = \theta Z Y_{gk} + \theta \mu_{gk}, \quad k = 1,2, \dots, p \tag{2}$$

In (2), the weight is represented as Y_{gk} belonging to $\mathbb{T}^{F \times m_k}$ and μ_{gk} is used to denote bias, both these variables are generated at random, having applicable size considering the correct possibility for distribution. Previously utilized activation functions for the mapped attribute elements are given as $\theta(\cdot)$. The feature elements that are grouped is represented as $B^p = [B_1, B_2, \dots., B_p]$.

where B_k belongs to $\mathbb{T}^{u \times m_k}$ and B^p belongs to $\mathbb{T}^{u \times \sum_{k=1}^{p} m_k}$ is used to show the attributes that are mapped from p feature elements. We assume that they are o groups of elements that are enhanced which are mapped to B^p at random.

$$J_l = \beta_l B^p Y_{jl} + \beta_l \mu_{jl}, \quad l = 1,2, \dots, o \tag{3}$$

here, the weights and bias are given as Y_{jl} and μ_{jl} generated at random, the $l-th$ enhanced element uses the activation function that is represented as $\beta_l(\cdot)$.

Therefore, the total output for layer utilized in this phase is given as $J^o = [J_1, J_2, \dots., J_o]$, where J^o belongs to $\mathbb{T}^{u \times o}$. However, the DFEM has its output formulated as given below:

$$A = \left[B_1, B_2, \dots, B_p \middle| J_o\right] Y_p^o = [B^p | J_o] Y_p^o \tag{4}$$

The weight that links the attribute layer as well as the enhanced elements to the output is shown as Y_p^o.

B. Random Forest Feature Importance Module (RFFIM)

The RFFIM is a widely used method that employs multiple decision trees. The recurrence across various decision trees can be eliminated by implementing a specific strategy. A random sample from the initial training data set is used to build each decision tree. Every decision tree produces either classification or regression output, and the method used improves the accuracy of that output. Therefore, this method is used to portray an enhanced performance for the resolution of high-dimensional, regression and classification problems. To configure deep learning models for prediction, start by considering a feature z_l having increased correlation along with the objective score denoted as a that is not necessarily an essential feature that aids in the process of prediction, while some features that have decreased correlation are also more beneficial. The issue lies in manually assessing the importance of each attribute and selecting the most essential one that aids the prediction. Compared with other attribute-selection methods, the RFFIM is the most efficient and effective. The main advantage of this method is that the importance value of each attribute can be determined, which is used to evaluate the importance of individual attributes based on the prediction output. Therefore, the RFFIM method is regarded as the most effective technique for attribute selection. To implement this technique for the proposed study, we derive the value of each attribute to understand its importance. Regression trees are built using the Mean Squared Error method to determine the outputs of each tree. The node x has the mean squared error is formulated as given below:

$$\text{Mean sq error}(x) = \sum_{k=1}^{K} (\hat{a}_k - a_k)^2 \tag{5}$$

The output of the regression is shown as $\hat{a}_k$ from the sample k from the node x. The number of samples that x is split into is given as K. The Mean squared error for attribute z_k for dividing the node x is given as:

$$\text{uncertainty}(z_k, x) = \text{Mean sq error}(z_k, x) - \text{Mean sq error}(z_k, x^N) Y_N - \text{Mean sq error}(z_k, x^T) Y_T \tag{6}$$

In this case, the $\text{uncertainty}(z_k, x)$ is used to denote the impurity of x. The child nodes are given as x^N and x^T on the left and right respectively. Each child node has a fraction of samples that are given as Y_N and Y_T for the left and right, respectively. Hence, the importance value for the feature z_k from $\text{uncertainty}(z_k, x)$ is expressed as:

$$\text{importance}_k^l = \left(\sum_{\text{d belongs to } X_{u_k}^l} \text{uncertainty}(z_k, x)\right)\left(\text{uncertainty}(X^l)\right)^{-1} \tag{7}$$

The set of nodes that are divided here for the tree-l is given as X^l. The feature set for splitting of node is given asd belongs to $X_{u_k}^l$. The summation of impurities in the tree$-$x is given as $\text{uncertainty}(X^l)$. The normalization of this value is define as

$$\text{Normimportance}_k^l = \left(\text{importance}_k^l\right)\left(\text{importance}_{sum}^l\right)^{-1} \tag{8}$$

where 0 is less than equal to Normmprtamce $\frac{l}{k}$ is less than equal to 1.

In conclusion, the final formulation to retrieve the importance value using RFFIM method is given as:

$$\text{importance}_k^{\text{random forest}} = \left(\sum_{l=1}^{p_{tree}} \text{Normimportance}_k^l\right)(p_{tree})^{-1} \tag{9}$$

where p_{tree} is used to express the count of trees as given in algorithm 1 (Table I).

C. Driving Factor Bootstrap Sampler (DFBS)

This technique is mainly used to enhance models in machine learning and has a broad range of applications for both classification and regression tasks in prediction models. Consider a data set with O number of samples, where each sample is chosen randomly and added to the initial set of samples. This process is repeated P times. A subset containing the data with P samples is gathered. The probability of each sample being selected from the initial data is the same. The algorithm used here is for establishing the subset V containing samples p_u. These subsets are used to identify learners who are weak. Consider a dataset that contains related data $F = \{U_1, U_2, \dots, U_P\}$, in this case U_k denotes a sample of the features as well as values that make up the data. These samples are split into training and test data, each with specific dimensions P_1 and P_2, respectively. The number of samples retrieved is determined by a ratio r from the initial dataset for sub-training data. This DFBS technique ensures that the selection of the

training data subset is random. algorithm 2 below provides a detailed explanation of the entire process (Table II).

TABLE I. RANDOM FOREST FEATURE IMPORTANCE MODULE (RFFIM)

Algorithm 1	**RFFIM method to retrieve Feature Value**
Input	Dataset for training: $\text{Data}_{\text{train}} = \{U_1, U_2, \dots, U_{P_1}\}$; Feature count of $\text{Data}_{\text{train}}$ given as p_h; count of chosen feature: p_u(lesser than p_h); attribute set of $\text{Data}_{\text{train}}$: H_g
Output	Chosen attribute set H_g
Step 1	$\text{Data}_{\text{train}}$ is used to construct a RFFIM Model
Step 2	Count of tree $= p_{tree}$
Step 3	For $l = 1$ to p_{tree} do
Step 4	Set $\text{importance}^l_{sum} = 0$
Step 5	For $k = 1$ to p_h do
Step 6	Set $\text{importance}_{initial} = 0$;
Step 7	Choose the set of divided node or elements X^l
Step 8	Choose the set of divided node or elements on the features z_k as $X^l_{z_k}$
Step 9	For d in $X^l_{z_k}$ do
Step 10	$\text{importance}_{initial} = \text{importance}_{initial} + \text{uncertainty}(z_k, x)$
Step 11	End for
Step 12	apply Equation (9)
Step 13	$\text{importance}^l_{sum} = \text{importance}^l_{sum} + \text{importance}^l_k$
Step 14	End for
Step 15	For $k = 1$ to p_h do
Step 16	$\text{Normimportance}^l_k = (\text{importance}^l_k)(\text{importance}^l_{sum})^{-1}$
Step 17	End for
Step 18	End for
Step 19	For $k = 1$ to p_h do
Step 20	Set $\text{Normimportance}_k = 0$
Step 21	For $l = 1$ to p_{tree} do
Step 22	$\text{Normimportance}_k = \text{Normimportance}_k + \text{Normimportance}^l_k$
Step 23	End for
Step 24	Apply equation (10)
Step 25	End for
Step 26	$\text{importance}^{\text{Random Forest}} = \{\text{importance}_1^{\text{random forest}}, \text{importance}_2^{\text{random forest}}, \dots\dots\dots, \text{importance}_{p_h}^{\text{random forest}}\}$
Step 27	$\text{importance}^{\text{Random Forest}}$ is arranged from biggest to smallest and the sorted set is obtained
Step 28	After sorting $\text{importance}^{\text{Random Forest}}$ is used to retrieve attribute set H_g with first values as p_u Return H_g

TABLE II. DRIVING FACTOR BOOTSTRAP SAMPLER (DFBS) ALGORITHM

Algorithm 2	**DFBS Algorithm**
Input	Ratio of the sample denoted as r; V number of iterations; Dataset for training: $\text{Data}_{\text{train}} = \{U_1, U_2, \dots, U_{P_1}\}$
Output	The data after being sampled given as Data_{bs}
Step 1	Subset of samples is set: $P_d = [P_1 . r]$
Step 2	For $k = 1$ to V do
Step 3	For $p = 1$ to P_d do
Step 4	Random samples are used in sampling of $U_p$from $\text{Data}_{\text{train}}$
Step 5	U_p is put back into sample data set $\text{Data}_{\text{train}}$
Step 6	P_d put in Data_{bs_k}
Step 7	End for
Step 8	$\text{Data}_{bs_k} = \{U_1, U_2, \dots, U_{P_h}\}$
Step 9	Put Data_{bs_k}into Data_{bs}
Step 10	End for
Step 11	Return $\text{Data}_{bs} = \{\text{Data}_{bs_1}, \text{Data}_{bs_2}, \text{Data}_{bs_3}, \dots, \text{Data}_{bs_V}\}$

III. PERFORMANCE EVALUATION

The AEDFLM was assessed using key performance metrics like MSE and R^2 to evaluate its ability to predict land degradation and identify important driving factors. The evaluation highlighted the model's capacity to accurately represent the complexity of land degradation patterns while maintaining robustness and stability through its ensemble method.

The methodology combined the RFFIM for selecting key driving factors, the DFBS to enhance stability across datasets, and the DFEM to expand the representation of selected features. The evaluation process confirmed the model's ability to balance accuracy and interpretability, ensuring that the most important factors influencing land degradation were identified and validated. Overall, the AEDFLM showed consistent accuracy in providing precise predictions and meaningful insights into the main factors causing land degradation.

A. Dataset Details

For this study, a comprehensive collection of satellite datasets and geospatial indices was used to evaluate land degradation trends. Vegetation indices like the NDVI and the Enhanced Vegetation Index (EVI) were calculated from MODIS and Landsat 8 images to track vegetation health. Land Surface Temperature (LST) data were collected from MODIS Terra and Aqua satellites, while rainfall and evapotranspiration estimates came from CHIRPS and ERA5 reanalysis datasets. Soil moisture conditions were evaluated using data from the Soil Moisture Active Passive (SMAP) mission. Additionally, topographic variables, including elevation and slope, were extracted from the SRTM Digital Elevation Model (DEM) at 30-meter resolution. These datasets were processed and analyzed with QGIS, ArcGIS, and Google Earth Engine, and their results were verified against ground-truth field observations to confirm spatial accuracy and reliability.

1) Driving Factors

- Wind Speed (Wind 10m(u)): Wind has been consistently identified as one of the most influential factors, especially in contributing to soil erosion and the displacement of topsoil. This makes it a key factor in land degradation, particularly in arid and semi-arid areas.
- Soil Moisture (Soil Water L1): Soil moisture is crucial for supporting vegetation and soil health. When soil moisture drops, it heightens the risk of degradation and desertification.
- Slope: Slope influences water runoff and soil retention. Steeper slopes are more susceptible to erosion, leading to increased land degradation.

- Total Evaporation: High evaporation rates can cause substantial soil drying, which diminishes the soil's capacity to sustain vegetation and hold nutrients, thereby speeding up land degradation..
- Population Density: Human activity, influenced by population density, directly impacts land use, causing greater pressure on land resources and leading to degradation through deforestation, overgrazing, and urban expansion.

B. Results

The MSE and R^2 metrics are essential for evaluating the predictive accuracy of the AEDFLM. A lower MSE indicates a more precise model prediction, demonstrating its capacity to capture the relationships between the identified driving factors and land degradation outcomes. Similarly, the R^2 value indicates how well the model explains variation in land degradation, helping verify that the most important factors are included in the prediction process. In addition to these metrics, the correlation matrix is essential for understanding how various driving factors interact. It provides insights into how factors such as soil moisture, wind speed, and slope interact, confirming that these features not only affect land degradation individually but also have a combined influence on the model's results. Additionally, feature importance from the RFFIM helps rank the key factors contributing to land degradation. This enables the model to concentrate on the most significant features, leading to more accurate predictions and a better understanding of the degradation process. Finally, permutation importance provides an additional layer of validation by measuring how model performance changes when individual feature values are shuffled. This method helps verify the importance of each factor, ensuring that the identified driving factors—whether primary or secondary—play a significant role in predicting land degradation. By combining these evaluation metrics, the AEDFLM not only improves prediction accuracy but also provides a comprehensive analysis of the factors driving land degradation. This ensures that the key features are identified and confirmed, providing valuable insights for land management and mitigation efforts.

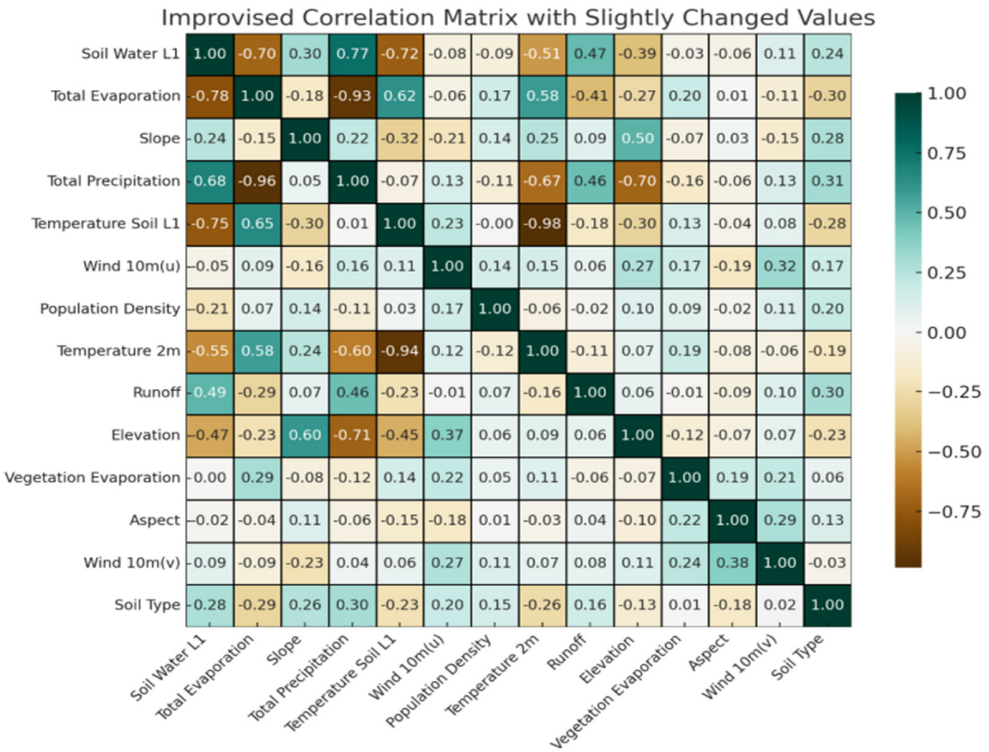

Fig. 2. Correlation matrix

1) Correlation Matrix

The correlation analysis shown in Figure 2 significantly enhances our understanding of how key environmental and demographic factors relate to land degradation, supporting the identification of the best datasets for predicting it. The correlation matrix showed several meaningful interactions among factors such as Soil Water L1, Total Precipitation, Soil Temperature L1, and others, providing deeper insights beyond their individual importance, as shown by the feature importance in the ensemble model. Key findings from the matrix reveal that Soil Water L1 and Total Precipitation have a strong positive correlation (0.68), suggesting that areas with more rainfall tend to have higher soil moisture, which directly affects land degradation risk. Similarly, Temperature at 2m and Total Evaporation show a significant positive correlation (0.58), confirming that higher surface temperatures lead to

increased evaporation rates, a key driver of land drying and desertification. Additionally, the correlation of 0.60 between Slope and Elevation highlights how these geographic features are essential for understanding water runoff and erosion risk, both of which contribute to land degradation.

Conversely, several factors show negative correlations, such as Soil Water L1 and Temperature 2m (-0.55), indicating that higher temperatures reduce soil moisture content and increase land degradation risk. Another notable negative correlation exists between Soil Temperature L1 and Total Precipitation (-0.67), indicating that areas with higher rainfall tend to have cooler soil temperatures, which can impact vegetation growth and land stability. These correlations are highly relevant for identifying optimal datasets for predicting land degradation because they not only confirm the importance of key features such as Soil Water L1, Total Precipitation, and Soil Temperature L1, but also provide additional context by showing how these factors interact to influence land degradation. This analysis complements the feature importance results derived from the ensemble model by providing a more comprehensive perspective of the data. The correlation matrix helps select the best datasets and ensure the key factors are included in the model. This enhances its predictive accuracy and offers better insight into the complex processes causing land degradation.

2) Feature Importance

The feature importance results shown in Figure 3 from the ensemble learning model offer essential insights into the main factors affecting land degradation prediction. The highest-ranked factors, Wind 10m(u) and Soil Water L1, with importance scores of 0.1448 and 0.1399, respectively, highlight the crucial roles of wind speed and soil moisture in influencing erosion and soil health. These are followed by Slope and Total Evaporation, each with a score of 0.0986, underscoring the importance of topographical features and water loss in driving degradation processes. Additional factors such as Wind at 10 meters (v), Vegetation Evaporation, and Population Density also play a moderate role, indicating the impact of both natural and human activities on land use and ecosystem health.

3) Permutations' Importance

This analysis helps identify the optimal driving factors by ranking their impact on the model's predictive performance. Factors such as wind, soil moisture, and slope are shown to be the primary forces behind degradation, enabling more targeted and accurate land degradation prediction models. Focusing on these top factors guides more effective land management strategies, ensuring that the most influential drivers of land degradation are addressed. This ultimately supports the goal of improving predictions and mitigation efforts for vulnerable regions.

4) MSE and R²

The model achieves an MSE of 0.113567, indicating reasonable accuracy in predicting land degradation using the identified driving factors. Although the MSE is higher than in earlier versions, it still shows that the model effectively captures the relationships between key environmental variables and land degradation processes. This MSE value indicates that the model's predictions are, on average, slightly off, but key factors—such as Wind 10m(u), Total Evaporation, and Soil Water L1—remain significant contributors to the model's performance. These factors still significantly influence the degradation process, allowing the model to make fairly accurate predictions about land vulnerability.

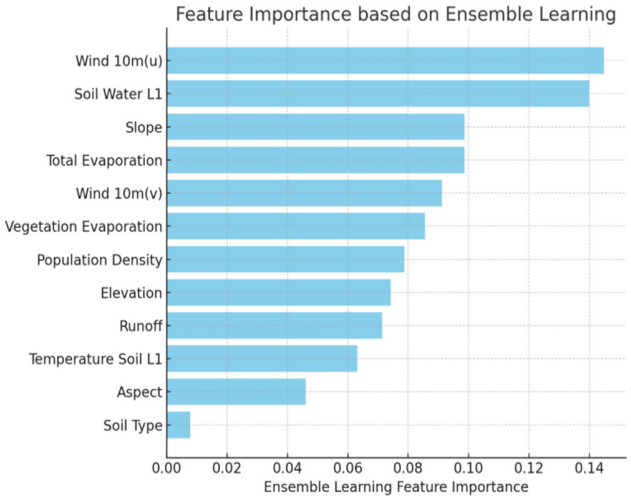

Fig. 3. Feature importance analysis

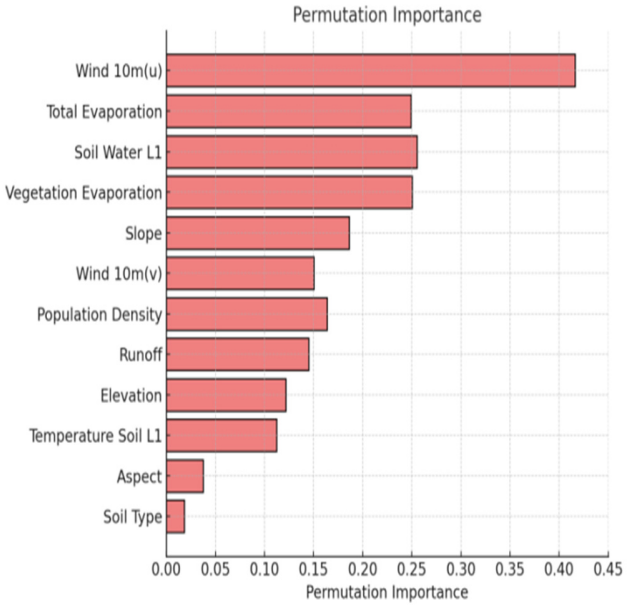

Fig. 4. Permutation importance

The model has an R-squared (R^2) of 0.512, indicating it explains about 51.2% of the variation in land degradation and provides a slight improvement in understanding the variance. This suggests that the identified driving factors effectively influence the predictive model. Principal factors, such as Wind 10m(u), Total Evaporation, and Soil Water L1, stand out as key contributors to land degradation, thus enhancing the model's predictive accuracy. The high R^2 value shows that more than

half of the variation in land degradation is now explained by the chosen driving factors, strongly supporting the goal of identifying the most influential drivers.

5) Result Analysis

The AEDFLM efficiently identifies key drivers of land degradation, offering valuable insights into how environmental and human factors influence degradation patterns. The model's performance was assessed using various metrics, including MSE, R^2, feature importance, and permutation importance, all of which showed the model's robustness and accuracy. The MSE value indicates a low average error between the predicted and actual land degradation values, demonstrating the model's strong predictive ability. The R^2 value shows that a significant portion of the variation in land degradation can be explained by the identified driving factors, reinforcing the effectiveness of the ensemble learning approach. Key factors such as Wind 10m(u), Total Evaporation, and Soil Water L1 were consistently identified as the most important contributors through feature importance analysis. These factors directly influence soil erosion, moisture retention, and vegetation health, making them essential for predicting land degradation. The permutation importance analysis further confirmed these results by quantifying each factor's effect on the model's performance, thereby demonstrating their essential roles in shaping land degradation patterns.

IV. CONCLUSION

This study introduced an Adaptive Ensemble Driving Factor Learning Model (AEDFLM) to tackle the challenge of identifying and validating the key driving factors behind land degradation. The model incorporates the Random Forest Feature Importance Module (RFFIM) for feature selection, the Driving Factor Bootstrap Sampler (DFBS) for robustness, and the Driving Factor Expansion Module (DFEM) for efficient learning. Additionally, Support Vector Regressor (SVR) and Linear Regressor were used to analyze both non-linear and linear interactions among specific driving factors, providing deeper insights. The results show that the AEDFLM significantly improves land degradation predictions, with a Mean Squared Error (MSE) of 0.113567 and an R-squared (R^2) of 0.512, indicating that the model explains over 51% of the variance in land degradation. Key factors such as Wind 10m(u), Total Evaporation, and Soil Water L1 were the most influential contributors to the predictions. Combining SVR with a Linear Regressor reinforced the significance of features like slope and population density, showing how these factors influence land degradation in both non-linear and linear ways. The AEDFLM not only improves prediction accuracy but also offers a comprehensive framework for understanding the complex relationships among driving factors, enabling more accurate land management and conservation strategies. By pinpointing the most influential factors, this model helps decision-making processes focused on reducing land degradation. Future research will expand the use of the AEDFLM model to forecast the Desertification Vulnerability Index (DVI). This will involve applying the model to assess the vulnerability of different regions to desertification, while accounting for additional environmental and socioeconomic factors. This extension will enable broader use across various geographical areas, offering further insights into reducing desertification risks.

ACKNOWLEDGEMENT

I want to express our sincere gratitude to all those who have supported and contributed to this research project. Primarily, I extend our heartfelt thanks to our guide for his unwavering guidance, invaluable insights, and encouragement throughout the research process. No funding is raised for this research.

REFERENCES

[1] S. Dharumarajan *et al.*, "Remote Sensing Sensors and Recent Techniques in Desertification and Land Degradation Mapping—A Review," in *Advances in Understanding Soil Degradation*, E. Saljnikov, L. Mueller, A. Lavrishchev, and F. Eulenstein, Eds. Cham: Springer International Publishing, 2022, pp. 701–716.

[2] D. Tilman, C. Balzer, J. Hill, and B. L. Befort, "Global food demand and the sustainable intensification of agriculture," *Proceedings of the National Academy of Sciences*, vol. 108, no. 50, pp. 20260–20264, Dec. 2011, https://doi.org/10.1073/pnas.1116437108.

[3] N. Alexandratos and J. Bruinsma, "World Agriculture Towards 2030/2050: The 2012 Revision," *ESA Working Paper No. 12-03*, Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, 2012.

[4] Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press, 2023.

[5] D. B. Lobell and C. B. Field, "Global scale climate–crop yield relationships and the impacts of recent warming," *Environmental Research Letters*, vol. 2, no. 1, Nov. 2007, Art. no. 014002, https://doi.org/10.1088/1748-9326/2/1/014002.

[6] T. Wheeler and J. von Braun, "Climate Change Impacts on Global Food Security," *Science*, vol. 341, no. 6145, pp. 508–513, Aug. 2013, https://doi.org/10.1126/science.1239402.

[7] H. Li, X. Yang, and K. Zhang, "Understanding global land degradation processes interacted with complex biophysics and socioeconomics from the perspective of the Normalized Difference Vegetation Index (1982–2015)," *Global and Planetary Change*, vol. 198, p. 103431, Mar. 2021, https://doi.org/10.1016/j.gloplacha.2021.103431.

[8] A. Abolhasani, H. Khosravi, G. Zehtabian, O. Rahmati, E. Heydari Alamdarloo, and P. D'Odorico, "Contribution of predictive factors of land degradation occurrence applying maximum entropy model," *Arid Land Research and Management*, vol. 38, no. 3, pp. 299–317, July 2024, https://doi.org/10.1080/15324982.2023.2298996.

[9] T. Qian, H. Bagan, T. Kinoshita, and Y. Yamagata, "Spatial–Temporal Analyses of Surface Coal Mining Dominated Land Degradation in Holingol, Inner Mongolia," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 5, pp. 1675–1687, May 2014, https://doi.org/10.1109/JSTARS.2014.2301152.

[10] S. Kalyan, D. Sharma, and A. Sharma, "Spatio-temporal variation in desert vulnerability using desertification index over the Banas River Basin in Rajasthan, India," *Arabian Journal of Geosciences*, vol. 14, no. 1, Jan. 2021, Art. no. 54, https://doi.org/10.1007/s12517-020-06417-0.

[11] J. Reiche, C. M. Souza, D. H. Hoekman, J. Verbesselt, H. Persaud, and M. Herold, "Feature Level Fusion of Multi-Temporal ALOS PALSAR and Landsat Data for Mapping and Monitoring of Tropical Deforestation and Forest Degradation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 5, pp. 2159–2173, Oct. 2013, https://doi.org/10.1109/JSTARS.2013.2245101.

[12] S. Prakash, M. C. Sharma, R. Kumar, P. S. Dhinwa, K. L. N. Sastry, and A. S. Rajawat, "Mapping and assessing land degradation vulnerability in Kangra district using physical and socio-economic indicators," *Spatial Information Research*, vol. 24, no. 6, pp. 733–744, Dec. 2016, https://doi.org/10.1007/s41324-016-0071-5.

[13] S. Kaliraj, M. Parmar, I. M. Bahuguna, and A. S. Rajawat, "Assessment of desertification and land degradation vulnerability in humid tropics

and sub-tropical regions of India using remote sensing and GIS techniques," in *Challenges of Disasters in Asia: Vulnerability, Adaptation and Resilience*, H. Sajjad, L. Siddiqui, A. Rahman, and M. Tahir, Eds., Singapore: Springer Nature Singapore, 2022, pp. 15–25, https://doi.org/10.1007/978-981-19-3567-1_2.

[14] Y. Wen, B. Guo, W. Zang, D. Ge, W. Luo, and H. Zhao, "Desertification detection model in Naiman Banner based on the albedo-modified soil adjusted vegetation index feature space using the Landsat8 OLI images," *Geomatics, Natural Hazards and Risk*, vol. 11, no. 1, pp. 544–558, Jan. 2020, https://doi.org/10.1080/19475705.2020.1734100.

[15] M. Lamchin *et al.*, "Assessment of land cover change and desertification using remote sensing technology in a local region of Mongolia," *Advances in Space Research*, vol. 57, no. 1, pp. 64–77, Jan. 2016, https://doi.org/10.1016/j.asr.2015.10.006. (8).

[16] L. Spivak, I. Vitkovskaya, M. Batyrbayeva, and A. Terekhov, "Detection of Desertification Zones Using Multi-year Remote Sensing Data," in *Use of Satellite and In-Situ Data to Improve Sustainability*, Dordrecht, 2011, pp. 235–239.

[17] Y. Wang, E. Guo, Y. Kang, and H. Ma, "Assessment of Land Desertification and Its Drivers on the Mongolian Plateau Using Intensity Analysis and the Geographical Detector Technique," *Remote Sensing*, vol. 14, no. 24, Dec. 2022, Art. no. 6365, https://doi.org/10.3390/rs14246365.

[18] A. Bouzekri *et al.*, "Assessment of the spatial dynamics of sandy desertification using remote sensing in Nemamcha region (Algeria)," *The Egyptian Journal of Remote Sensing and Space Sciences*, vol. 26, no. 3, pp. 642–653, Dec. 2023, https://doi.org/10.1016/j.ejrs.2023.07.006.

[19] A. Zanchetta, G. Bitelli, and A. Karnieli, "Monitoring desertification by remote sensing using the Tasselled Cap transform for long-term change detection," *Natural Hazards*, vol. 83, no. 1, pp. 223–237, Oct. 2016, https://doi.org/10.1007/s11069-016-2342-9.

[20] R. Becerril-Piña, C. Díaz-Delgado, C. A. Mastachi-Loza, and E. González-Sosa, "Integration of remote sensing techniques for monitoring desertification in Mexico," *Human and Ecological Risk Assessment: An International Journal*, vol. 22, no. 6, pp. 1323–1340, Aug. 2016, https://doi.org/10.1080/10807039.2016.1169914.

[21] J. Xiao, Y. Shen, R. Tateishi, and W. Bayaer, "Development of topsoil grain size index for monitoring desertification in arid land using remote sensing," *International Journal of Remote Sensing*, vol. 27, no. 12, pp. 2411–2422, June 2006, https://doi.org/10.1080/01431160600554363.

[22] W. Wei *et al.*, "Spatiotemporal changes of land desertification sensitivity in northwest China from 2000 to 2017," *Journal of Geographical Sciences*, vol. 31, no. 1, pp. 46–68, Jan. 2021, https://doi.org/10.1007/s11442-021-1832-1.

[23] V. A. Dave, M. Pandya, and R. Ghosh, "An Assessment of the Desertification Vulnerability based on MEDALUS model," in *2019 International Conference on Intelligent Computing and Remote Sensing (ICICRS)*, July 2019, pp. 1–6, https://doi.org/10.1109/ICICRS46726.2019.9555853.

[24] A. Boali, H. R. Asgari, A. Mohammadian Behbahani, A. Salmanmahiny, and B. Naimi, "Remotely sensed desertification modeling using ensemble of machine learning algorithms," *Remote Sensing Applications: Society and Environment*, vol. 34, Apr. 2024, Art. no. 101149, https://doi.org/10.1016/j.rsase.2024.101149.

[25] M. Kaddour, L. De Oto, A. F. Tandjaoui, H. Guerid, and M. Khodadadzadeh, "High Resolution Remote Sensing Images from the Algerian steppic zone." DANS Data Station Physical and Technical Sciences, June 20, 2024, https://doi.org/10.17026/PT/POJGN2.

[26] M. Kaddour, L. De Oto, A. F. Tandjaoui, H. Guerid, and M. Khodadadzadeh, "Remote sensing-based driving factors of desertification in the Algerian steppic zone." *DANS Data Station Physical and Technical Sciences*, May 16, 2024, https://doi.org/10.17026/PT/0DNFS0.

[27] R. A. Abbood, A. A. Jafarzadeh, S. Oustan, and A. I. Hamad, "Regional Change Prediction of Land Degradation Risks using Cellular Automata-Markov Modeling in Dhi Qar Province, Iraq," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 22908–22914, June 2025, https://doi.org/10.48084/etasr.10483.