American Journal of Engineering and Applied Sciences

Original Research Paper

Ecological Function Evaluation of Soil and Water Conservation Measures for Power Transmission and Transformation Project Construction Based on Remote Sensing and it's Improvement Path

Yang Han, Songsong Wang, Yanhui Yang, Xing Rong, Zhang Zhaokun and Wang Haoran

Department of Construction Management, Engineering Construction Branch, State Grid Hebei Electric Power Co., Ltd., Shijiazhuang, Hebei, China

Article history
Received: 14-11-2024
Revised: 10-12-2024
Accepted: 11-12-2024

Corresponding Author:
Yang Han
Department of Construction Management, Engineering Construction Branch, State Grid Hebei Electric Power Co., Ltd., Shijiazhuang, Hebei, Chin
Email: Xiangyu_Wang2024@outlook.com

Abstract: Assessing ecological functions is critical for determining the efficacy of soil and water conservation measures, especially during the development of power transmission and transformation projects. Efficient conservation practices are essential for reducing ecological effects and guaranteeing sustainable development. This research aims to create a resilient model, Ecological Function Prediction for Conservation Effectiveness (EFP-CE) that will classify conservation effectiveness as either effective or ineffective. The goal is to improve prediction accuracy and offer actionable insights for better conservation tactics. The EFP-CE model combines many analytical methods: Missing values are imputed using Support Vector Regression (SVR) and outliers are detected and removed using Euclidean distance. Categorical variables are converted using label encoding, while numerical attributes are subjected to Min-Max normalization. An ensemble feature selection technique integrates filter and wrapper methods to find important predictors, while cluster-based oversampling fixes data imbalance. The dataset is separated into training and testing sets. A Bagged Gradient Boosting model is trained and assessed to forecast conservation efficiency. The proposed model was evaluated using a ten ecological function assessment attributes dataset. The Bagged Gradient Boosting model obtained 93% accuracy, 91% precision, 89% recall, an F1-score of 90%, and a Matthews Correlation Coefficient (MCC) of 82%, suggesting strong predictive effectiveness in evaluating conservation measures. The EFP-CE model demonstrates how machine learning methods can be integrated to improve the assessment of conservation measures. By enhancing prediction accuracy, this research presents helpful knowledge for policymakers and stakeholders participating in environmental safety during infrastructure projects, eventually adding to more sustainable construction procedures.

Keywords: Ecological Function Assessment, Soil and Water Conservation, Machine Learning, Conservation Efficiency, Bagged Gradient Boosting

Introduction

The growing need for infrastructure construction, especially power transmission and transformation projects, has created important environmental difficulties (Lian *et al*., 2022). These projects frequently disrupt natural ecosystems, causing soil erosion, poor water quality, and biodiversity loss (Wang *et al*., 2023). As a response to these difficulties, efficient soil and water conservation measures have become essential to reduce negative ecological effects (Chen *et al*., 2020). The assessment of ecological functions related to these conservation procedures is critical for calculating their efficacy and guaranteeing sustainable implementation (Bian *et al*., 2024). By evaluating the ecological results of conservation tactics, stakeholders can develop informed decisions that encourage environmental wellness while also promoting infrastructure growth (Li *et al*., 2020).

Figure (1) depicts the interdependence of different ecological features, like vegetation cover, soil erosion rates, and water quality indices, which all contribute to the evaluation of conservation efficiency. This figure emphasizes the intricacy of ecological systems and the requirement for an extensive assessment framework that incorporates various data sources and analytical methods.

Important Elements of the Study

This research concentrates on many key elements that are critical to comprehending the efficacy of soil and water conservation measures. Initially, the study highlights the importance of remote sensing data in assessing ecological functions. Remote sensing technology presents useful information about vegetation cover, land use alterations, and other key ecological parameters, allowing for a more precise evaluation of conservation tactics. Second, the research uses sophisticated machine-learning methods to improve predictive accuracy. Conventional evaluation techniques frequently fail to address the intricacies of ecological data, which can incorporate missing values, outliers, and high dimensionality. Using methods like Support Vector Regression (SVR) for data imputation and Bagged Gradient Boosting for classification, the research seeks to efficiently tackle these difficulties.

Research Objective

The main objective of this research is to create the Ecological Function Prediction for Conservation Effectiveness (EFP-CE) model, which will classify the efficacy of soil and water conservation measures as effective or ineffective. This classification is dependent on a comprehensive dataset that includes a variety of ecological features such as vegetation cover percentage, soil erosion rates, water quality indices, and more. By improving prediction accuracy and presenting useful knowledge, the EFP-CE model seeks to substantially assist in bettering conservation tactics in the construction of infrastructure.

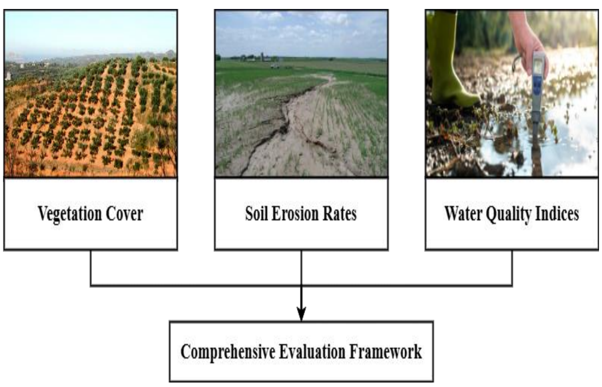

Fig. 1: Ecological attributes and their role in conservation assessment

Evolution of Ecological Function Evaluation Methods

Ecological function evaluation techniques have developed significantly over time, moving from conventional methods to more sophisticated methods based on remote sensing and machine learning. Previous research depended heavily on manual measurements and simple models, which lacked the precision required to capture the intricacies of ecological systems. Current technological advances, including the utilization of remote sensing data and machine learning algorithms, have substantially enhanced the precision and effectiveness of ecological evaluations (Jiao, 2024). The current study expands on these advances by incorporating satellite imagery and advanced machine learning methods such as Support Vector Regression (SVR) and Bagged Gradient Boosting, which improve prediction accuracy and manage data intricacies more effectively than previous models. This method overcomes previous constraints, providing a more comprehensive and dependable evaluation of conservation tactics.

Advancements Over Previous Research and Methodological Improvements

While prior studies investigated remote sensing for ecological surveillance, numerous investigations encountered difficulties in dealing with incomplete or noisy data, as well as high-dimensional datasets. Typical methods, like conventional regression techniques, struggled to produce precise outcomes when dealing with missing or anomalous data (Olawade *et al*., 2024). The current study enhances these techniques by using SVR for data imputation, which allows for the incorporation of missing data while maintaining accuracy. Furthermore, the use of Bagged Gradient Boosting improves the model's capacity to detect complicated trends, resulting in higher predictive precision. Unlike previous studies, which frequently lacked resilient data imputation and sophisticated classification methods, this study provides a more extensive and dependable methodology for assessing conservation tactics, leading to more precise and actionable predictions.

Future Scope and Research Gaps

Despite major improvements, there are still multiple research gaps in the area of ecological function assessment. Future research should concentrate on incorporating various data sources, like ground-level ecological data, socioeconomic factors, and sophisticated remote sensing imagery, to develop more extensive models (Castro-Magnani *et al*., 2021). Recent studies face difficulties in implementing these incorporated methodologies on a large scale, particularly in infrastructure development projects. Additionally, more study is needed on real-time data processing and dynamic ecological shifts over time (Sagan *et al*., 2020). The

DOI: 10.3844/ajeassp.2025.55.66

current study fills these gaps by integrating various data types with sophisticated machine learning algorithms to enhance predictive performance. It lays the groundwork for future research that can improve these models, allowing decision-makers to execute more efficient and ecological conservation practices.

Literature Review

Table (1) summarizes important studies on ecological restoration and network construction, concentrating on methodologies, objectives, findings, and limitations. Each study provides distinctive insights into the utilization of remote sensing, machine learning, and ecological network theory to better comprehend and improve ecological health, connectivity, and land-use sustainability. This comparison emphasizes the variety of methods used across regions and ecological settings, highlighting both progress and remaining difficulties in ecological restoration.

Table 1: Summary table

Reference No	Objective	Methodology	Result	Limitations
Zhai *et al.* (2022)	Assess the ecological restoration impacts of the Yongding River Watershed in China	Remote sensing assessment utilizing Ziyuan-3 (ZY-3) and Landsat images, land cover and Normalized Difference Vegetation Index (NDVI) data, examination of water resources and ecology	Maintaining ecological water quantity, creating ecological corridors, and improving ecological function	Constrained to a particular watershed; findings may not apply to other areas
Yang *et al.* (2022)	Establish an ecological network for environmental protection in Panzhou, Guizhou Province	Combined method integrating ecological quality, Ecological Function Index (EFI), Morphological Spatial Pattern Analysis (MSPA), and circuit theory to find ecological sources and corridors	Discovered ecological sources, corridors, pinch points, and barriers, which guide spatial planning	Concentrate on ecological protection; constrained assessment of long-term ecological effects
Lu *et al.* (2022)	Evaluate the evolution of ecological networks in the Wuhan urban agglomeration from 2000-2020	Circuit theory, centrality index, intricate network theory; spatial evaluation of ecological quality, function, and structure	Identified the effect of land growth on ecological network fragmentation and proposed tactics for protecting ecological corridors	Restricted to the Wuhan area; additional testing is required for other urban agglomerations
Luo *et al.* (2024)	Simulate land utilization and build an ecological network around Poyang Lake to promote sustainable growth	Multi-Objective Programming (MOP) model, Non-dominated Sorting Genetic Algorithm II (NSGA-II), Patch-generating Land Use Simulation (PLUS) model, spatial evaluation through MSPA, Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST) model, and intricate network	Enhanced Ecological Networks (EN) in different situations, discovered important ecological regions and enhanced ecological resilience in networks	Constrained to Poyang Lake; may not account for all ecological and socioeconomic variables in larger settings
Liu *et al.* (2022)	Evaluate land use/cover shift and ecosystem service value in Xi'an's urban ecological area	Multi-resolution remote sensing with Landsat 8 Operational Land Imager (OLI) and GaoFen-2 (GF-2) satellite data; Land Use and Cover Change (LUCC) interpretation and Ecosystem Service Value (ESV) evaluation	Documented important land utilization shifts, decreased ESV in urban ecological zones, and presented suggestions on ecological zone planning	Evaluation is restricted to 2014-2020; further data is required to evaluate long-term patterns
Liu *et al.* (2023)	Develop a quantitative model for long-term safety trends in the Qinling Mountains	Machine learning, remote sensing, Geographic Information Systems (GIS), Analytic Hierarchy Process-Principal Component Analysis (AHP-PCA), and Minimal Cumulative Resistance (MCR) model	Created Ecological Security Patterns (ESPs), discovered important ecological sources and pathways, and addressed ecological protection in the Qinling Mountains	Concentrated on ecological sensitivity; restricted consideration of human social dynamics that influence security trends

Li *et al*. (2021)	Assess ecological stability in a semi-arid coal mining region over 16 years	The coupling coordination degree method relies on the structural and functional state transition model (SFSTM)	Ecological stability fluctuated, with mining operations as the dominant factor; presenting a structural threshold for early ecological indicators	Region-specific; factors impacting ecological stability may vary in other mining settings
Xiu *et al*. (2021)	Evaluate the influence of ecological building on vegetation restoration in the Loess Plateau, China	Remote sensing and geographic data technology are incorporated with ecosystem shifts and remaining models	Ecological engineering resulted in an important rise in vegetation NDVI, with over 90% improvement in vegetation quality	Restricted concentration on vegetation without evaluating the larger ecological effect of urban expansion
Hao *et al*. (2022)	Evaluate ecological restoration in rare-earth mining regions utilizing Remote Sensing Ecological Index (RSEI)	RSEI is computed from Landsat-8 data for a multidimensional evaluation of ecological environments	Enhanced ecological conditions through multimodal management; emphasized spatial variations in ecological responses	Restricted applicability outside rare-earth mining; findings rely on the quality of RSEI and fragmentation management
Xiao *et al*. (2022)	Construct a regional ecological network for the Taishan region utilizing multi-objective improvement	The method combines morphological spatial trend evaluation, MCR, and gravity modeling	Showed ecological network building linking natural resources across the area, emphasizing ecosystem integrity	Regional uniqueness may restrict generalization to regions with various ecological systems

Furthermore, several studies have contributed to the ecological evaluation of soil and water conservation in power transmission projects. Naeeni *et al*. (2023) improved hydraulic simulations to better comprehend soil-water interactions in construction. Rahmani Firozjaei *et al*. (2020); and Rahmani Firozjayi *et al*. (2019) concentrated on lateral pipe intake simulations, which helped to handle water flow and sediment transport in conservation efforts. Aghazadeh and Attarnejad (2024; 2020a-b) investigated desalination systems and sweetened seawater transportation, proposing environmentally friendly water management solutions for infrastructure projects. Firozjaei *et al*. (2024) used physical models and decision tree algorithms to evaluate discharge efficiency, whereas Hajebi *et al*. (2024) investigated bottom intake effectiveness for water conservation. Rahmani Firozjaei *et al*. (2024a) investigated bottom intake structures for desalination plants, offering insights into soil and water conservation in construction. Their research on submerged vanes (Rahmani Firozjaei *et al*., 2024b) enhanced intake effectiveness, assisting with soil erosion and water flow management. These researches contribute to superior conservation practices for environmental sustainability in infrastructure.

The summarized studies show important growth in using sophisticated methods like remote sensing and machine learning to track, assess, and improve ecological restoration attempts. While each research provides useful methods and results, constraints like regional specificity, restricted temporal scope and insufficient investigation of fundamental drivers highlight the necessity of more extensive and adaptable methods. These gaps represent possibilities for proposed work to create models that incorporate wider ecological and socioeconomic factors to assist sustainable land and ecological network management.

Materials and Methods

To evaluate conservation efficiency, this study used the EcoConservation Project Impact Dataset (ECPID), which was compiled from field measurements, environmental tracking, and remote sensing data. Important attributes included vegetation cover, soil erosion rate, slope steepness, rainfall intensity, and water quality indices, with remote sensing models such as NDVI offering additional information. Data preprocessing techniques included SVR for missing values, Euclidean distance for outlier elimination, label encoding for categorical variables, and Min-Max scaling for numerical features. Feature selection incorporated filter and wrapper techniques, whereas cluster-based oversampling tackled class imbalances. The dataset, which was saved in CSV format, was safely handled using cloud backup and version control. Experiments were carried out on an Aspire 3 system equipped with an Intel Core i7-1260P (12-core, 2.1 GHz), 64 GB RAM, and 18

MB L3 cache for efficient computation. The development environment included JDK 1.8 and Apache NetBeans IDE 15, which provided a solid foundation for creating and testing the Conservation_Effectiveness prediction model.

This section proposes the Ecological Function Prediction for Conservation Effectiveness (EFP-CE) model, detailing each stage of data preprocessing, feature selection, and model training. The processes are described using equations for imputation, normalization, and feature selection, which explain how these stages aid in precisely forecasting conservation efficiency. Figure (2) depicts the research's flow diagram.

Data Collection

EcoConservation Project Impact Dataset (ECPID) was created using a combination of field measurements, environmental monitoring, and remote sensing methods. Project-specific data were collected by field teams, incorporating distinctive identifiers, conservation techniques, and environmental metrics such as vegetation cover percentage, soil erosion rate, slope steepness, rainfall intensity, and distance to water bodies. Furthermore, remote sensing presented water quality indices and environmental health scores through image processing, utilizing models such as NDVI for vegetation evaluation.

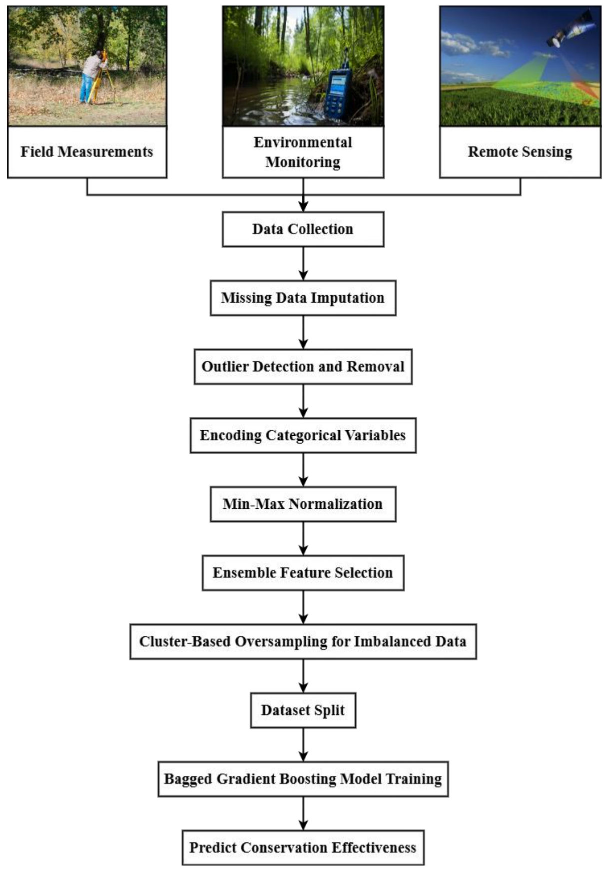

Fig. 2: Research flow diagram

These data points, collected from both field inspections and satellite imagery, were combined and cross-validated to ensure consistency. To guarantee data quality, any missing data was tackled using imputation and outliers were eliminated. The dataset was then structured in Comma-Separated Values (CSV) format, saved on a safe server with cloud backup, and handled using version control to enable updates while retaining data integrity. This procedure produced an extensive, structured dataset for assessing conservation efficiency using machine learning.

This dataset presents a structured view of different environmental and conservation-related features for construction projects, each with a distinct Project_ID. Important features comprise Vegetation_Cover_Percentage, which indicates vegetation coverage following conservation efforts, Soil_Erosion_Rate, which indicates soil loss per hectare per year and Water_Quality_Index, which measures nearby water quality. Slope_Steepness and Rainfall_Intensity provide information about the terrain and precipitation levels, whereas Distance_to_Water_Body shows proximity to water sources. The kind of Conservation_Measure_Used, like Grass Planting or Check Dams, is determined, as well as a Remote_Sensing_Image_Score for evaluating environmental health through image analysis. Construction_Area_Size describes the project's spatial extent. The target attribute, Conservation_Effectiveness, categorizes each project's influence as Effective or Ineffective, operating as the result variable for conservation effectiveness.

Missing Data Imputation with Support Vector Regression (SVR)

To tackle missing values in numerical features, Support Vector Regression (SVR) is used for imputation, which is a powerful method for capturing intricate relationships between variables without presuming a particular data distribution. SVR models each feature with missing values by considering it as the target and using the other features as predictors. Initially, a feature matrix X is defined by eliminating the features with missing values while retaining all other pertinent features. The feature matrix is then used to train an SVR model, which learns the relationships between the features and forecasts missing values. The trained model is used to estimate and replace missing values in each instance of the feature being imputed, yielding a dataset with complete numerical values. The prediction utilizing SVR can be represented mathematically as shown in Eq. (1):

$$\hat{x} = SVR(X_{train}) \tag{1}$$

where, $\hat{x}$ represents the imputed values predicted for each missing instance. This imputation procedure is applied consecutively to each feature with missing data, guaranteeing accuracy and consistency throughout the dataset.

DOI: 10.3844/ajeassp.2025.55.66

Outlier Detection and Removal Using Euclidean Distance

Outliers are detected and removed utilizing Euclidean distance computations, with a focus on detecting data points that deviate substantially from their nearest neighbors. For each data point x_i in the dataset, the Euclidean distance d (x_i, x_j) to other points is computed to determine its proximity to neighboring instances. A threshold δ is established, beyond which any point is regarded as an outlier. Particularly, if the computed distance d (x_i, x_j) surpasses the defined threshold, the point x_i is marked as an outlier and flagged for elimination. This procedure guarantees that only representative data points remain in the dataset, improving the dependability of the following examines. The Euclidean distance between two points x_i and x_j is computed utilizing the Eq. (2):

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(x_{i,k} - x_{j,k})^2} \quad (2)$$

where, n signifies the number of features in the dataset. By performing this technique, all recognized outliers are eliminated, resulting in a cleaner dataset that better reflects the fundamental patterns in the data.

Encoding Categorical Variables

Categorical variable encoding converts attributes such as Conservation_Measure_Used and Project_ID into numerical format by assigning each category a distinct integer label. A distinct integer L is assigned to each unique category c within C, enabling the incorporation of categorical data into machine learning models. The encoding procedure can be described by Eq. (3):

$$L(c) = \{0,1,2,\ldots,k-1\} \quad (3)$$

Where k represents the total number of distinct categories within the feature C. By transforming categorical data into numerical labels, this method guarantees compatibility with computational models while maintaining category distinctions.

Min-Max Normalization

Min-max normalization is used to scale numerical features within the range [0,1], guaranteeing that features with larger ranges don't have an undue impact on the model. For each numerical feature x, the normalization procedure modifies values using the minimum (x_{min}) and maximum values (x_{max}) within the feature, protecting relative variances while aligning data to a common scale. The normalized value for each feature x is computed as Eq. (4):

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

This conversion keeps proportional relationships in the data and improves the model's effectiveness by removing scale-related biases across features.

Ensemble Feature Selection

The ensemble feature selection technique integrates filter and wrapper methods to find the most important attributes for predicting Conservation_Effectiveness, utilizing complementary advantages from both statistical and model-based methods.

Filter-Based Selection (Mutual Information)

The filter method employs Mutual Information (MI) to assess the relationship between each feature X and the target variable Y. MI is especially useful for ranking features based on how much they decrease uncertainty regarding the target when observed. The MI formula, shown in Eq. (5), computes the shared data by evaluating the joint probability $p(x,y)$ of X and Y alongside their marginal probabilities $p(x)$, $p(y)$:

$$I(X;Y) = \sum_{x\in X} \sum_{y\in Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \quad (5)$$

Features with higher MI scores suggest greater predictive relevance for the target variable and are prioritized in the choosing process. This statistical ranking technique presents a basic comprehension of each feature's contribution to the prediction procedure before any model is used.

Wrapper-Based Selection (Recursive Feature Elimination with SVM)

The wrapper method combines Recursive Feature Elimination (RFE) with Support Vector Machines (SVM), involving iteratively training an SVM model and eliminating the least important attributes at each step. RFE identifies the features that contribute the most to model accuracy by continually improving the feature subset. Utilizing SVM as the base model presents a reliable measure of feature significance because SVM is extremely sensitive to informative features and ignores noise. This iterative removal continues until the optimum subset of features is determined, which provides a more concentrated set for enhanced model performance.

The final feature subset is selected through majority voting, which incorporates the rankings from both the filter (Mutual Information) and wrapper (RFE) techniques. This ensemble method enables a balanced selection procedure, utilizing both statistical relevance and model-based performance metrics to determine an extensive and efficient feature set.

Cluster-Based Oversampling for Imbalanced Data

To tackle the problem of imbalanced data, cluster-based oversampling is applied to the minority class. This approach seeks to improve the representation of the minority class by creating synthetic samples, allowing for more balanced class distributions and better model training. The procedure starts with clustering the minority class utilizing K-means clustering. The model divides the data points into k clusters, each represented by a centroid. Clustering assists find the fundamental structure of the minority class by grouping similar instances. The selection of k is crucial because it impacts the number of synthetic samples created.

Once the clusters have been established, synthetic samples are created for each cluster using the centroid. This entails adding new data points that are close to the centroid of each cluster, efficiently augmenting the minority class while introducing minimal noise. These synthetic samples are usually created by adding small random perturbations to the centroid, guaranteeing that the new instances retain the original data distribution within that cluster. Lastly, synthetic samples are introduced into the dataset to obtain a more balanced representation of both classes. This cluster-based method not only increases the number of minority class samples but also maintains the data's local structure, resulting in better model efficiency and generalization abilities. Cluster-based oversampling contributes to a more equitable learning procedure by efficiently raising the minority class size. This reduces the risk of bias towards the majority class.

Dataset Split

The dataset is separated into training and testing sets to make model creation and assessment easier. This split is done in an 80:20 ratio, which means that 80% of the data is used for training and the remaining 20% is reserved for testing. The 80:20 split ratio for splitting a dataset into training and testing sets is broadly used in machine learning because of its balanced strategy to model training and assessment. Assigning 80% of the data to training guarantees that the model has sufficient instances to learn the fundamental trends and relationships, which is critical for developing a reliable predictive model. The remaining 20% is set aside for testing, which offers a suitably big and independent dataset to assess the model's efficacy on previously unseen data, revealing insights into its generalization capacity. This ratio assists in striking a balance between having sufficient data for correct model training and maintaining a resilient testing set to avoid overfitting and evaluate the model's efficacy in forecasting real-world results. Additionally, the 80:20 split is a proven approach that has been used effectively in a variety of research, guaranteeing robust and consistent assessment of machine learning models.

Bagged Gradient Boosting Model

The Bagged Gradient Boosting method is used as the main classification model to forecast Conservation_Effectiveness, taking advantage of ensemble techniques' advantages to enhance accuracy and resilience. This approach comprises training numerous gradient-boosting models on bootstrapped subsets of the training data, which improves the model's capacity to generalize by decreasing variance by averaging the predictions from different models. Hyperparameter tuning is used to further enhance the Bagged Gradient Boosting model's efficiency. To attain optimum outcomes, important parameters are fine-tuned. These include the number of estimators N, which decides how numerous individual models will be incorporated in the ensemble; the learning rate α, which controls the contribution of each model to the overall prediction; and the maximum depth d, which limits the depth of each tree, thus controlling overfitting and intricacy. Eq. (6) mathematically expresses the fundamental idea of gradient boosting:

$$F_m(x) = F_{m-1}(x) + \alpha h_m(x) \tag{6}$$

In this equation, $F_m(x)$ signifies the cumulative prediction from the first m models, $F_{m-1}(x)$ denotes the cumulative prediction from the first m-1 models, while $h_m(x)$ represents the m^{th} base model. By periodically adding the weighted predictions of the base models, the Bagged Gradient Boosting technique improves predictive accuracy while retaining the model's capacity to capture intricate relationships within the dataset. This structured training procedure is crucial for developing a strong model able to precisely predict the efficiency of conservation measures depending on the presented features.

Pseudo code 1 denotes the proposed Ecological Function Prediction for Conservation Effectiveness (EFP-CE) model.

Pseudo code 1: Ecological Function Prediction for Conservation Effectiveness (EFP-CE)

```
For each numerical feature with missing values:
    Execute Support Vector Regression (SVR) to predict
    and impute missing entries
Compute Euclidean distance between data points
Eliminate points with distances surpassing a set threshold
Transform categorical features into numerical labels
Implement Min-Max Normalization to normalize features
between 0 to 1
Utilize filter techniques (Mutual Information) and wrapper
techniques (RFE with SVM).
Choose features using majority voting from ensemble rankings
Perform cluster-based oversampling on minority classes
Separate dataset into 80% training set and 20% testing set
Employ Bagged Gradient Boosting on the training set using
hyperparameter tuning
Predict Conservation_Effectiveness on the testing set
```

Results and Discussion

The performance evaluation was carried out on an Aspire 3 system outfitted with a high-performance Intel configuration that was designed to manage intensive computational tasks and big datasets. The system features an Intel Core i7-1260P processor with a 12-core architecture that balances processing power and energy efficiency. The system, which runs at 2.1 GHz and has 64 GB of RAM, can handle intricate algorithms and datasets with ease. The 18 MB L3 cache improves data retrieval speeds, increasing the entire system's efficiency.

JDK 1.8 was utilized for software development, along with Apache NetBeans IDE 15, to create a stable environment for coding, debugging, and testing algorithms. This configuration enabled smooth interaction with the system's resources and allowed for efficient performance evaluation. Table (2) shows the details of the experimental setup.

Comparative Analysis

To evaluate the effectiveness of the EFP-CE model, it was compared to four famous classifiers: Naive Bayes, JRip, IBk, and J48. The efficacy of each classifier was assessed using five important metrics: Accuracy, precision, recall, F1-score, and Matthews Correlation Coefficient (MCC). The formula for these metrics is outlined as follows: Accuracy is the percentage of accurate predictions among all predictions made:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

where, TP = True Positives; TN = True Negatives; FP = False Positives and FN = False Negatives

Precision denotes the accuracy of positive predictions:

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

A higher precision indicates fewer false positives. Recall, also called sensitivity, quantifies the model's capacity to identify all pertinent cases (true positives):

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

A higher recall indicates fewer false negatives. The F1-score is the harmonic mean of precision and recall, offering a balance between the two:

$$F1 - score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{10}$$

The Matthews Correlation Coefficient (MCC) assesses the effectiveness of binary classifications by taking into account all four confusion matrix categories:

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{11}$$

MCC values range from -1 to 1, with 1 representing a perfect prediction. Table (3) compares the performance metrics of each classifier.

As shown in Table (3), the EFP-CE model surpassed all other classifiers on all five metrics. With an accuracy of 93%, the model showed better predictive performance, demonstrating its ability to accurately classify data. This high level of accuracy exceeds that of Naive Bayes, JRip, IBk, and J48, which is critical for applications that require consistent results. Additionally, Fig. (3) displays the accuracy comparison.

Figure (3) shows that the EFP-CE model outperforms all other classifiers in terms of accuracy, with a remarkable 93%. This better accuracy is due to the model's sophisticated feature selection procedure and capacity to use ensemble methods, which improve its generalization abilities. The enhanced management of class imbalance guarantees that the model correctly detects minority class instances, improving overall accuracy. Additionally, Fig. (4) displays the precision comparison.

Table 2: Experimental setup

Component	Specification
Processor Model	Intel Core i7-1260P
CPU Type	12-Core Architecture
Brand	Aspire 3
Memory (RAM)	64 GB
Clock Speed	2.1 GHz
Operating System	Windows 11 Home
L3 Cache Size	18 MB
JDK Version	1.8
IDE	Apache NetBeans IDE 15

Table 3: Performance metrics comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MCC (%)
Naive Bayes	79	76	81	78	57
JRip	82	80	81	81	62
IBk	84	82	86	84	67
J48	86	84	88	86	71
EFP-CE	93	91	89	90	82

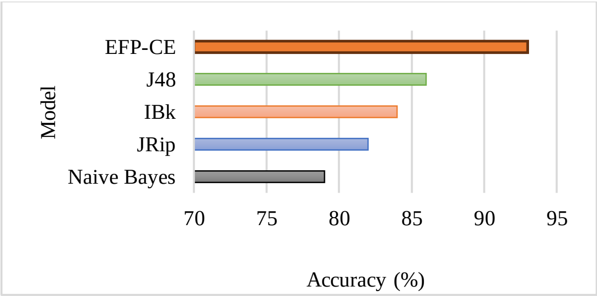

Fig. 3: Accuracy comparison

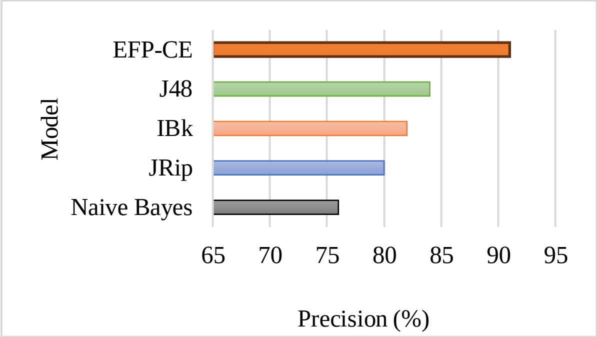

Fig. 4: Precision comparison

The EFP-CE model also leads in precision, scoring 91%. This high precision demonstrates that the model is capable of generating accurate positive predictions, substantially lowering the number of false positives. The EFP-CE model's ensemble method and rigorous feature selection contribute to its precision by training only the most pertinent features, reducing noise, and improving prediction dependability. Additionally, Fig. (5) displays the recall comparison.

In terms of recall, the EFP-CE model outperformed all other classifiers, scoring 89 percent. This high recall demonstrates that the model correctly detects a large proportion of true positive cases, decreasing false negatives. The model's capacity to capture minority class instances is further supported by techniques like cluster-based oversampling that effectively manage class imbalances. Additionally, Fig. (6) demonstrates the F1-Score Comparison.

The EFP-CE model also performs well in the F1-score, earning a score of 90%. This metric shows the model's balanced effectiveness in terms of precision and recall, verifying its dependability across a variety of scenarios. The EFP-CE model's ensemble approach efficiently reduces the trade-offs between precision and recall that are common in other classifier systems. Additionally, Fig. (7) demonstrates the MCC Comparison.

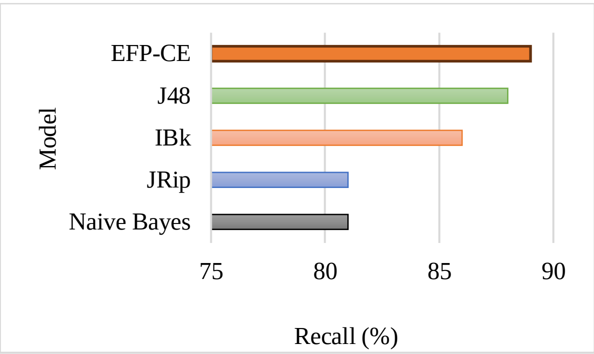

Fig. 5: Recall comparison

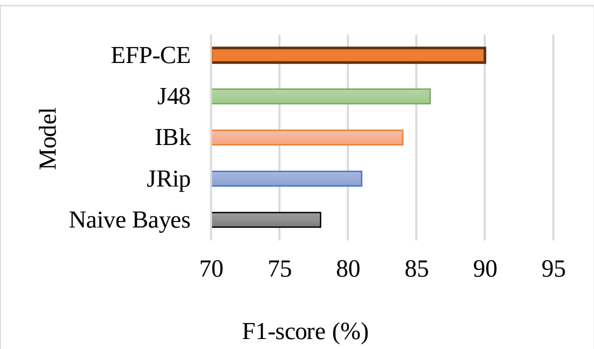

Fig. 6: F1-score comparison

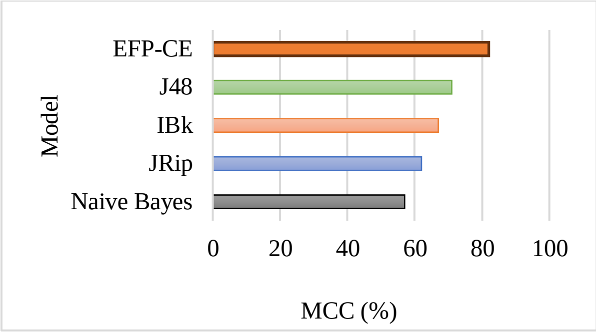

Fig. 7: MCC comparison

Lastly, the EFP-CE model obtains the highest Matthews correlation coefficient of 82%, suggesting high classification accuracy. This metric shows that the model executes well not only on positive predictions but also on negative predictions. The high MCC reflects the model's resilience and overall efficiency in differentiating between classes, rendering it a better option for the task at hand.

This research substantially enhances previous research by tackling important constraints in model efficiency and making novel contributions to boost accuracy and dependability. While previous models like Naive Bayes, JRip, IBk, and J48 attained mild success, with accuracy ranging from 79-86% and Matthews Correlation Coefficient (MCC) values ranging from 57-71%, they fell short of providing consistent precision, recall, and F1-scores, which are required for effective assessment. The proposed model, EFP-CE, surpasses these benchmarks with 93% accuracy and 82% MCC, showing better predictive power and model stability. Additionally, EFP-CE shows balanced enhancements in all important metrics precision (91%), recall (89%), and F1-score (90%), demonstrating its capacity to offer overall performance gains. These improvements emphasize the research's novel contributions to enhancing ecological assessment techniques and establishing a novel standard for tackling intricate issues in soil and water conservation for power transmission projects.

Overall, the comparative analysis shows that the EFP-CE model surpasses conventional classifiers on all metrics. Its sophisticated techniques for feature selection and handling imbalanced data help to its outstanding predictive efficiency, establishing it as the top model in this assessment. The model's ability for real-world conservation effectiveness evaluations is highlighted by its high accuracy, precision, recall, F1-score, and MCC scores.

Conclusion

This study proposed the EFP-CE model as an innovative method for predicting conservation efficiency, which outperformed conventional classifiers such as Naive Bayes, JRip, IBk, and J48. Using advanced feature selection techniques and tackling class imbalance, the EFP-CE model outperformed all important metrics, containing accuracy (93%), precision (91%), recall (89%), F1-score (90%), and MCC (82%). These findings demonstrate the model's resilience and dependability in evaluating complex conservation datasets. In contrast to prior research, where performance metrics were limited by restrictions in precision and consistency, the EFP-CE model establishes a novel standard for conservation data evaluation, providing a more effective and dependable tool for researchers and practitioners. The results highlight the ability of novel machine learning methods to transform decision-making procedures in ecological conservation, paving the manner for more efficient ecological practices.

Study Limitations and Future Research

Despite its promising outcomes, the EFP-CE model has some drawbacks. Its efficacy is impacted by the excellence and volume of input data and potential biases in data gathering can affect model predictions. Furthermore, the model's intricacy may limit interpretability, rendering it difficult for non-technical stakeholders to comprehend its results. Future research could overcome these drawbacks by incorporating explainable AI methods that improve transparency and usability. Extending the model's application to other domains, like healthcare, urban planning, and precision agriculture, could reveal important information about its flexibility and scalability. Furthermore, investigating hybrid models that integrate EFP-CE with deep learning methods or integrating real-time data streams could also improve its predictive capacities and widen its usefulness across various disciplines.

Acknowledgment

The authors are grateful to all institutions and individuals who contributed to this study.

Funding Information

This work was supported by the State Grid Corporation of China Technology Project（SGTYHT/23-JS-001).

Author’s Contributions

All authors equally contributed to this work.

Ethics

This study follows ethical guidelines and contains no conflicts of interest or harm to humans, animals, or the environment.

References

Aghazadeh, K., & Attarnejad, R. (2020a). Improved Desalination Pipeline System Utilizing the Temperature Difference under Sub-Atmospheric Pressure. *Water Resources Management*, *34*(1), 1–19. https://doi.org/10.1007/s11269-019-02415-4

Aghazadeh, K., & Attarnejad, R. (2020b). Study of sweetened seawater transportation by temperature difference. *Heliyon*, *6*(3), e03573. https://doi.org/10.1016/j.heliyon.2020.e03573

Aghazadeh, K., & Attarnejad, R. (2024). Experimental investigation of desalination pipeline system and vapor transportation by temperature difference under sub-atmospheric pressure. *Journal of Water Process Engineering*, *60*, 105133. https://doi.org/10.1016/j.jwpe.2024.105133

Bian, H., Li, M., Deng, Y., Zhang, Y., Liu, Y., Wang, Q., Xie, S., Wang, S., Zhang, Z., & Wang, N. (2024). Identification of ecological restoration areas based on the ecological safety security assessment of wetland-hydrological ecological corridors: A case study of the Han River Basin in China. *Ecological Indicators*, *160*, 111780. https://doi.org/10.1016/j.ecolind.2024.111780

Castro-Magnani, M., Sanchez-Azofeifa, A., Metternicht, G., & Laakso, K. (2021). Integration of remote-sensing-based metrics and econometric models to assess the socio-economic contributions of carbon sequestration in unmanaged tropical dry forests. *Environmental and Sustainability Indicators*, *9*, 100100. https://doi.org/10.1016/j.indic.2021.100100

Chen, Y., Gong, A., Zeng, T., & Yang, Y. (2020). Evaluation of water conservation function in the Xiongan New Area based on the comprehensive index method. *PLOS ONE*, *15*(9), e0238768. https://doi.org/10.1371/journal.pone.0238768

Firozjaei, M. R., Hajebi, Z., Naeeni, S. T. O., & Akbari, H. (2024). Discharge performance of a submerged seawater intake in unsteady flows: Combination of physical models and decision tree algorithms. *Journal of Water Process Engineering*, *60*, 105198. https://doi.org/10.1016/j.jwpe.2024.105198

Hajebi, Z., Firozjaei, M. R., Naeeni, S. T. O., & Akbari, H. (2024). Hydraulic performance of bottom intake velocity caps using PIV and OpenFOAM methods. *Applied Water Science*, *14*(3), 38. https://doi.org/10.1007/s13201-023-02091-1

Hao, H., Lian, Z., Zhao, J., Wang, H., & He, Z. (2022). A Remote-Sensing Ecological Index Approach for Restoration Assessment of Rare-Earth Elements Mining. *Computational Intelligence and Neuroscience*, *2022*, 1–14. https://doi.org/10.1155/2022/5335419

Jiao, Z. (2024). The Application of Remote Sensing Techniques in Ecological Environment Monitoring. *Highlights in Science, Engineering and Technology*, *81*, 449–455. https://doi.org/10.54097/7dqegz64

Li, J., Pei, Y., Zhao, S., Xiao, R., Sang, X., & Zhang, C. (2020). A Review of Remote Sensing for Environmental Monitoring in China. *Remote Sensing*, *12*(7), 1130. https://doi.org/10.3390/rs12071130

Li, X., Lei, S., Liu, Y., Chen, H., Zhao, Y., Gong, C., Bian, Z., & Lu, X. (2021). Evaluation of Ecological Stability in Semi-Arid Open-Pit Coal Mining Area Based on Structure and Function Coupling during 2002–2017. *Remote Sensing*, *13*(24), 5040. https://doi.org/10.3390/rs13245040

Lian, Z., Hao, H., Zhao, J., Cao, K., Wang, H., & He, Z. (2022). Evaluation of Remote Sensing Ecological Index Based on Soil and Water Conservation on the Effectiveness of Management of Abandoned Mine Landscaping Transformation. *International Journal of Environmental Research and Public Health*, *19*(15), 9750. https://doi.org/10.3390/ijerph19159750

Liu, L., Chen, M., Luo, P., Duan, W., & Hu, M. (2023). Quantitative Model Construction for Sustainable Security Patterns in Social–Ecological Links Using Remote Sensing and Machine Learning. *Remote Sensing*, *15*(15), 3837. https://doi.org/10.3390/rs15153837

Liu, S., Huang, G., Wei, Y., & Qu, Z. (2022). Monitoring and Assessing Land Use/Cover Change and Ecosystem Service Value Using Multi-Resolution Remote Sensing Data at Urban Ecological Zone. *Sustainability*, *14*(18), 11187. https://doi.org/10.3390/su141811187

Lu, Y., Liu, Y., Huang, D., & Liu, Y. (2022). Evolution Analysis of Ecological Networks Based on Spatial Distribution Data of Land Use Types Monitored by Remote Sensing in Wuhan Urban Agglomeration, China, from 2000 to 2020. *Remote Sensing*, *14*(11), 2618. https://doi.org/10.3390/rs14112618

Luo, Z., Yang, X., & Luo, S. (2024). Land Use Simulation and Ecological Network Construction around Poyang Lake Area in China under the Goal of Sustainable Development. *Sustainability*, *16*(18), 8146. https://doi.org/10.3390/su16188146

Naeeni, S. T. O., Rahmani Firozjaei, M., Hajebi, Z., & Akbari, H. (2023). Investigation of the performance of the response surface method to optimize the simulations of hydraulic phenomena. *Innovative Infrastructure Solutions*, *8*(1), 10. https://doi.org/10.1007/s41062-022-00977-8

Olawade, D. B., Wada, O. Z., Ige, A. O., Egbewole, B. I., Olojo, A., & Oladapo, B. I. (2024). Artificial intelligence in environmental monitoring: Advancements, challenges and future directions. *Hygiene and Environmental Health Advances*, *12*, 100114. https://doi.org/10.1016/j.heha.2024.100114

Rahmani Firozjaei, M., Behnamtalab, E., & Salehi Neyshabouri, S. A. A. (2020). Numerical simulation of the lateral pipe intake: flow and sediment field. *Water and Environment Journal*, *34*(2), 291–304. https://doi.org/10.1111/wej.12462

Rahmani Firozjaei, M., Salehi Neyshabouri, S. A. A., Amini Sola, S., & Mohajeri, S. H. (2019b). Numerical Simulation on the Performance Improvement of a Lateral Intake Using Submerged Vanes. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, *43*(2), 167–177. https://doi.org/10.1007/s40996-018-0126-z

Rahmani Firozjaei, Mahmood, Hajebi, Zahra, Naeeni, S. T. O., & Akbari, H. (2024a). Experimental and Numerical Investigation of Bottom Intake Structure for Desalination Plants. *Numerical Methods in Civil Engineering*, *8*(3), 1–9. https://doi.org/10.61186/nmce.2303.1022

Rahmani Firozjayi, M., Behnamtalab, E., & Salehi Neyshabouri, S. A. (2019). Numerical Simulation of Lateral Pipe Intake from Open Channel. *Iranian Journal of Soil and Water Research*, *50*(1), 135–147. https://doi.org/10.22059/ijswr.2018.252260.667851

Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., Maalouf, S., & Adams, C. (2020). Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning and cloud computing. *Earth-Science Reviews*, *205*, 103187. https://doi.org/10.1016/j.earscirev.2020.103187

Wang, J., Zhen, J., Hu, W., Chen, S., Lizaga, I., Zeraatpisheh, M., & Yang, X. (2023). Remote sensing of soil degradation: Progress and perspective. *International Soil and Water Conservation Research*, *11*(3), 429–454. https://doi.org/10.1016/j.iswcr.2023.03.002

Xiao, H., Guo, Y., Wang, Y., Xu, Y., & Liu, D. (2022). Evaluation and Construction of Regional Ecological Network Based on Multi-Objective Optimization: A Perspective of Mountains–Rivers–Forests–Farmlands–Lakes–Grasslands Life Community Concept in China. *Applied Sciences*, *12*(19), 9600. https://doi.org/10.3390/app12199600

Xiu, L., Yao, X., Chen, M., & Yan, C. (2021). Effect of Ecological Construction Engineering on Vegetation Restoration: A Case Study of the Loess Plateau. *Remote Sensing*, *13*(8), 1407. https://doi.org/10.3390/rs13081407

Yang, L., Suo, M., Gao, S., & Jiao, H. (2022). Construction of an Ecological Network Based on an Integrated Approach and Circuit Theory: A Case Study of Panzhou in Guizhou Province. *Sustainability*, *14*(15), 9136. https://doi.org/10.3390/su14159136

Zhai, L., Cheng, S., Sang, H., Xie, W., Gan, L., & Wang, T. (2022). Remote sensing evaluation of ecological restoration engineering effect: A case study of the Yongding River Watershed, China. *Ecological Engineering*, *182*, 106724. https://doi.org/10.1016/j.ecoleng.2022.106724

Abbreviations	Description
EFP-CE	Ecological Function Prediction for Conservation Effectiveness
SVR	Support Vector Regression
MCC	Matthews Correlation Coefficient
ZY-3	Ziyuan-3
NDVI	Normalized Difference Vegetation Index
EFI	Ecological Function Index
MSPA	Morphological Spatial Pattern Analysis
MOP	Multi-Objective Programming
NSGA-II	Non-dominated Sorting Genetic Algorithm II
PLUS	Patch-generating Land Use Simulation
InVEST	Integrated Valuation of Ecosystem Services and Tradeoffs
EN	Ecological Networks
OLI	Operational Land Imager
GF-2	GaoFen-2
LUCC	Land Use and Cover Change
ESV	Ecosystem Service Value
GIS	Geographic Information Systems
AHP-PCA	Analytic Hierarchy Process-Principal Component Analysis
MCR	Minimal Cumulative Resistance
ESPs	Ecological Security Patterns
SFSTM	Structural and Functional State Transition Model
RSEI	Remote Sensing Ecological Index
ECPID	EcoConservation Project Impact Dataset
CSV	Comma-Separated Values
MI	Mutual Information
RFE	Recursive Feature Elimination
SVM	Support Vector Machines
CPU	Central Processing Unit
RAM	Random Access Memory
GHz	Gigahertz
JDK	Java Development Kit
IDE	Integrated Development Environment