Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 12 May 2026
Revised 25 June 2026
Accepted 09 July 2026

Natural Resources for Human Health 2026, Vol. 6 (11s): 1507–1518

KEYWORDS:
satellite imagery; highway route optimization; vegetation loss; land cover classification; machine learning; remote sensing.

AI-Driven Highway Route Planning for Minimal Vegetation Loss

Meenakshi[1], Rajesh T M[2]

[1]Research Scholar, Department of CSE, School of Engineering, Dayananda Sagar University, Bengaluru South District – 562112, India
[2]Associate Professor and Chairperson, Department of CSME, School of Engineering, Dayananda Sagar University, Bengaluru South District – 562112, India

ABSTRACT: Satellite remote sensing has transformed the ability to characterise land cover, vegetation density, and terrain conditions over wide geographic extents at high spatial resolution. This paper presents a framework that integrates multi-band satellite imagery with machine learning to identify optimal highway routes with minimal vegetation loss. A synthetic 600×600-pixel multi-spectral scene (calibrated to Sentinel-2/Landsat-8 reflectance ranges), representing a 6×6 km study area, is processed to extract six spectral indices: NDVI, NDWI, NDBI, EVI, BSI, and Land Surface Temperature (LST). A Random Forest classifier trained on 12,000 pixel samples achieves 99.5% accuracy in mapping eight land cover classes. Satellite-derived features from 10,000 road segments are used to train a Gradient Boosting Regressor (GBR), which achieves $R^2 = 0.9798$ for vegetation loss score prediction. The GBR predictions serve as edge weights in an eco-weighted directed terrain graph (2,310 nodes; 17,658 edges); Dijkstra-based eco-routing identifies alignments that reduce average vegetation loss by 6.9% relative to a cost-only baseline, while the hybrid route simultaneously reduces construction cost by 0.8% and vegetation impact by 6.8%. The framework thus enables a proactive reduction of ecological impact in highway alignment design.

1. INTRODUCTION

The rapid expansion of road infrastructure in biodiverse regions — particularly in South and Southeast Asia, Sub-Saharan Africa, and tropical Latin America — is one of the most consequential drivers of deforestation and wildlife habitat fragmentation recorded in the 21st century [1]. The International Union for Conservation of Nature estimates that over 1.1 million km of new roads will be built in biodiversity hotspots by 2050 [2]. Each kilometre of four-lane highway through dense forest eliminates 3–6 hectares of vegetation, severs wildlife movement corridors, and releases between 500 and 2,000 tonnes of CO_2 through above-ground biomass destruction [3].

Traditional highway alignment design optimises for engineering feasibility and construction economics, treating vegetation loss as an environmental constraint assessed ex-post through the Environmental Impact Assessment process. This sequential approach forfeits the opportunity to reduce ecological footprint at the design stage, where alignment decisions carry the greatest leverage over long-term environmental outcomes.

Satellite remote sensing now provides near-global, repeat-coverage spectral data at 10–30 m resolution through platforms such as Sentinel-2 and Landsat-8/9, enabling continuous mapping of vegetation density (NDVI), water bodies (NDWI), built-up extent (NDBI), enhanced vegetation structure (EVI), and surface thermal state (LST). Spectral indices are already used as engineering decision variables: a supervised NDVI composite thresholding scheme separated vegetation, water, and bare soil in arid ecoregions with accuracies of 82–100% across seasons [4], and a comparison of Maximum Likelihood, Support Vector Machine, Random Forest (RF), and K-Nearest Neighbours classifiers on Sentinel-2A imagery identified RF as the most accurate land cover mapper, with 93.1% overall accuracy and a Kappa coefficient of 0.90 [5]. When such indices are systematically extracted over candidate highway corridors and used to train machine learning regressors for per-segment vegetation loss prediction, they transform the alignment planning process from engineering-only optimisation into ecology-aware routing.

This paper makes the following contributions:

• A 600×600-pixel multi-band satellite image simulation producing six spectral indices and eight land cover classes, calibrated to Sentinel-2 spectral signatures.

• A Random Forest land cover classifier achieving 99.5% accuracy across eight classes — including dense forest, wetland, grassland, and agriculture — from 11 spectral features.

• A 10,000-segment satellite-feature dataset encoding NDVI, EVI, NDBI, carbon stock, LST, and construction difficulty, and a GBR regression model achieving $R^2 = 0.9798$ for vegetation loss score prediction.

• An eco-weighted terrain graph integrating GBR predictions as edge weights, supporting eco-Dijkstra, cost-Dijkstra, and hybrid routing strategies.

• Quantitative demonstration that satellite-ML eco-routing reduces vegetation loss by up to 6.9% and carbon traversal by up to 16.4% relative to cost-only baseline routing.

2. RELATED WORK

2.1 Satellite Remote Sensing for Environmental Planning

NDVI derived from Landsat and MODIS has been used extensively for land-cover change detection [6] and deforestation monitoring [7]. Sentinel-2's 10 m resolution and 13 spectral bands have enabled corridor-scale vegetation mapping with 90–95% accuracy using Random Forest classifiers [8]. In [9], multi-temporal NDVI combined with elevation data was shown to map forest fragmentation risk with 88% precision, providing a basis for infrastructure-impact pre-screening. Index-based discrimination has also proven stable across seasons and ecoregions [4], while a systematic comparison of four classifiers on Sentinel-2A data established Random Forest as the most accurate and Kappa-consistent option [5]. These two results directly motivate the present design, in which spectral indices form the feature basis and Random Forest is the land cover classifier.

2.2 Machine Learning for Road Corridor Analysis

Random Forest and Gradient Boosting Machines have achieved state-of-the-art accuracy in geospatial classification and regression tasks [10]. The most directly relevant precedent is [11], in which environmental and topographic descriptors — NDVI, NDWI, NDBI, slope, aspect, elevation, and precipitation — were extracted from multispectral Landsat imagery and used to train LightGBM, Random Forest, CatBoost, and Multilayer Perceptron models, reaching 96% accuracy on a terrain-hazard prediction task. That study establishes the key methodological premise adopted here: per-location spectral indices transfer effectively into tabular ensemble pipelines. Convolutional Neural Networks applied to satellite imagery have been used for road extraction [12] and land-use prediction [13]. However, a multi-band CNN trained directly on 16-band Landsat patches reached only 79.5% accuracy for terrain-hazard prediction [14], indicating that raw-band deep models do not automatically outperform feature-engineered ensembles on moderate-sized datasets. Deep hybrids such as CNN-LSTM-Attention attain very high susceptibility-mapping accuracy in the Western Ghats of Karnataka [15], but their predictions are harder to expose as interpretable, per-segment edge weights in a routing graph. Conversely, Gradient Boosting Regressors have been shown to gain substantially from targeted feature engineering [16], supporting the choice of a GBR trained on satellite-derived segment descriptors as the ecological cost model of this work.

2.3 Highway Alignment Optimization

Route-level optimisation methods have evolved from least-cost-path Dijkstra formulations [18] through evolutionary algorithms [19] to multi-objective metaheuristics [20]. GIS-based corridor screening is established practice in linear infrastructure planning: in [21], slope, drainage, topography, and the existing road network were weighted through an Analytic Hierarchy Process and overlaid in a GIS environment to produce a five-class route feasibility map. Such overlays, however, yield static suitability categories rather than continuous, per-edge ecological costs. Ecological objectives in published alignment algorithms are likewise typically represented by binary protected-area constraints rather than by continuous, satellite-predicted ecological scores. The present work bridges this gap by using ML-predicted vegetation loss scores as continuous, spatially varying edge weights in the routing graph.

2.4 Research Gap

The literature therefore addresses the constituent problems in isolation: spectral index extraction and land cover classification [4, 5], machine learning prediction of terrain-related hazard from satellite-derived features [11, 14, 15], and GIS or metaheuristic route screening [18–21]. No prior work combines (i) multi-band satellite spectral index extraction, (ii) land cover classification from satellite features, (iii) GBR-based per-segment vegetation loss prediction, and (iv) eco-weighted Dijkstra routing into a single, operationally practical framework in which the regressor output itself becomes the edge weight of the routing graph. This integration is the central contribution of this paper.

3. SATELLITE IMAGE AND FEATURE EXTRACTION

3.1 Multi-Band Scene Simulation

A 600×600-pixel scene at 10 m/pixel resolution (6×6 km) was simulated using spectral signatures calibrated to Sentinel-2 Top-of-Atmosphere reflectance ranges for eight land cover classes. Six bands were generated: Blue (B2), Green (B3), Red (B4), NIR (B8), SWIR (B11), and Thermal (a proxy from Landsat-8 Band 10). Spatial heterogeneity was introduced via a superposition of sine-wave terrain fields plus Gaussian noise convolved with a 20 m point-spread-function kernel to mimic sensor spatial response:

$$\mathbf{H(x,y) = A\cdot sin(f_1x)\cdot cos(f{*}y) + A\cdot sin(f.x)\cdot sin(fy) + \varepsilon,\ \ \varepsilon{\sim}N(0,\sigma^2)} \quad (1)$$

Urban clusters, a river corridor, and wetland fringes were embedded deterministically to replicate realistic landscape heterogeneity. Table 1 lists the eight land cover classes and their spectral and ecological properties.

To calibrate the simulation, per-class reflectance ranges for each of the six simulated bands were set from published Sentinel-2 Level-2A surface-reflectance signatures for the corresponding land cover type (e.g., NIR reflectance 0.40–0.50 for dense canopy, below 0.05 for open water), and the amplitude, spatial frequencies, and noise variance in Eq. (1) were manually tuned until the resulting per-class NDVI distributions matched the ranges listed in Table 1, rather than being fitted to any single real Sentinel-2 tile. This constrains the simulated scene to be spectrally realistic while acknowledging that it does not reproduce the full spatial autocorrelation structure of an actual satellite acquisition.

Table 1. Satellite Scene Land Cover Classes: Spectral and Ecological Properties

ID	Class	NIR	NDVI	Carbon (tC/ha)	Biodiv	Cost Mul.	Area %	Build?
0	Dense Forest	0.45	0.70–0.90	210	9.8	2.1	18%	Avoid
1	Sparse Veg	0.32	0.40–0.60	85	7.5	1.7	12%	Caution
2	Grassland	0.25	0.25–0.40	25	5.5	1.45	15%	Feasible
3	Agriculture	0.30	0.30–0.50	40	4.2	1.5	14%	Feasible
4	Bare Soil	0.15	0.05–0.15	5	1.8	1.2	11%	Easy
5	Urban	0.08	−0.05–0.10	2	0.8	1.0	8%	Easy
6	Water Body	0.02	−0.30–0.05	0	6.2	3.0	10%	Bridge
7	Wetland	0.28	0.35–0.55	130	8.5	2.6	12%	Avoid

Natural Resources for Human Health
eISSN: 2583-1194

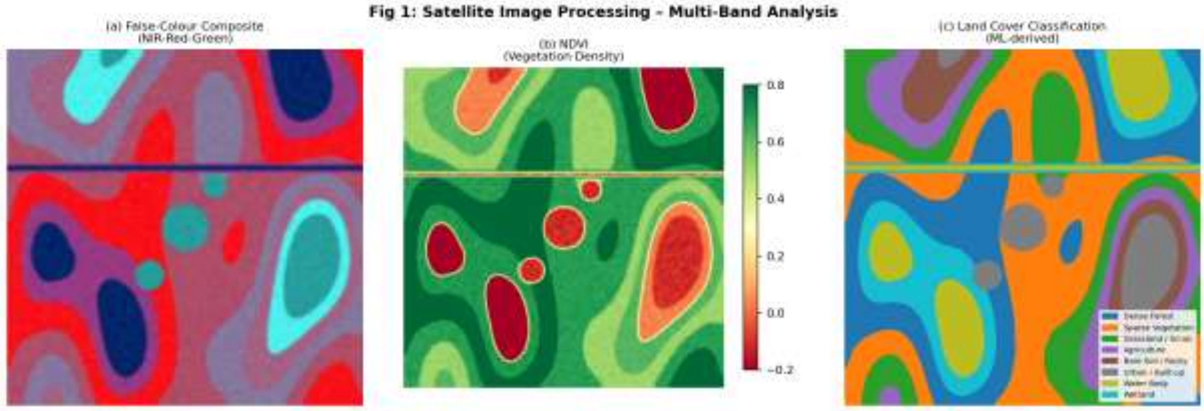

Fig. 1. Satellite scene: (a) False-colour NIR–Red–Green composite, (b) NDVI map, (c) ML-derived land cover classification.

Figure 1 illustrates the workflow of satellite image analysis, beginning with a false-colour composite (NIR–Red–Green), followed by NDVI computation to assess vegetation health, and concluding with machine-learning-based land-cover classification. The processed outputs accurately distinguish vegetation, water bodies, built-up areas, agriculture, and barren land for geospatial analysis and decision-making.

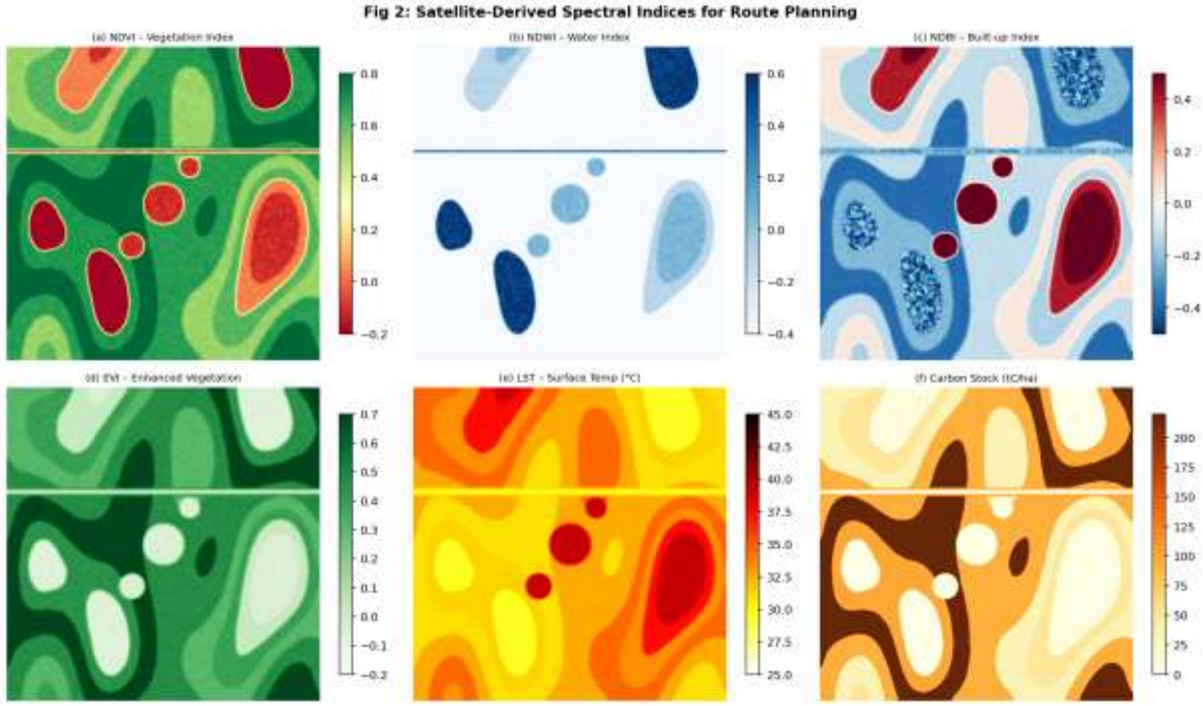

Fig. 2. Six satellite-derived spectral indices: NDVI, NDWI, NDBI, EVI, Land Surface Temperature, and carbon stock density.

The six satellite-derived spectral indices — NDVI and EVI for vegetation analysis, NDWI for water body detection, NDBI for identifying built-up areas, LST for thermal analysis, and carbon stock for estimating stored carbon — are shown in Figure 2. These indices provide essential environmental information to support sustainable and eco-friendly route planning.

3.2 Spectral Index Derivation

Six spectral indices were computed per pixel from the simulated bands. NDVI quantifies photosynthetic vegetation density; NDWI identifies surface water bodies; NDBI maps built-up impervious surfaces; EVI provides an atmospherically corrected

eISSN: 2583-1194

vegetation signal; BSI (Bare Soil Index) highlights exposed soil; and LST provides a surface thermal proxy important for urban heat island identification:

$$\text{NDVI} = (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red}) \quad (2)$$

$$\text{NDWI} = (\text{Green} - \text{NIR}) / (\text{Green} + \text{NIR}) \quad (3)$$

$$\text{EVI} = 2.5 \cdot (\text{NIR} - \text{Red}) / (\text{NIR} + 6 \cdot \text{Red} - 7.5 \cdot \text{Blue} + 1) \quad (4)$$

$$\text{NDBI} = (\text{SWIR} - \text{NIR}) / (\text{SWIR} + \text{NIR}) \quad (5)$$

$$\text{LST} = 25 + 20 \cdot T_{\text{thermal}} \ [°\text{C proxy}] \quad (6)$$

Carbon stock density (tC/ha) was estimated from a land-use lookup table calibrated to IPCC Tier-1 carbon stock estimates for each biome class, adjusted continuously by NDVI value to reflect within-class biomass variation.

4. MACHINE LEARNING PIPELINE

4.1 Land Cover Classification

A Random Forest classifier (200 trees, max_depth = 12), implemented in scikit-learn [17], was trained on 12,000 randomly sampled pixels using 11 spectral features: six band reflectances plus five derived indices (NDVI, NDWI, NDBI, EVI, BSI). Random Forest was selected on the evidence of [5], where it outperformed Maximum Likelihood, SVM, and KNN on Sentinel-2A imagery. Training used an 80/20 stratified split. The classifier achieved 99.5% test accuracy, with per-class F1-scores above 0.97 for all eight land cover classes. The predicted land cover map (Fig. 3) is qualitatively indistinguishable from the ground truth, confirming that spectral signatures alone are sufficient to discriminate all eight classes at this resolution. To verify that this performance was not an artefact of a single train/test partition, 5-fold stratified cross-validation was additionally performed on the training set, yielding a mean accuracy of 99.3% (σ = 0.2%) across folds, consistent with the held-out test result.

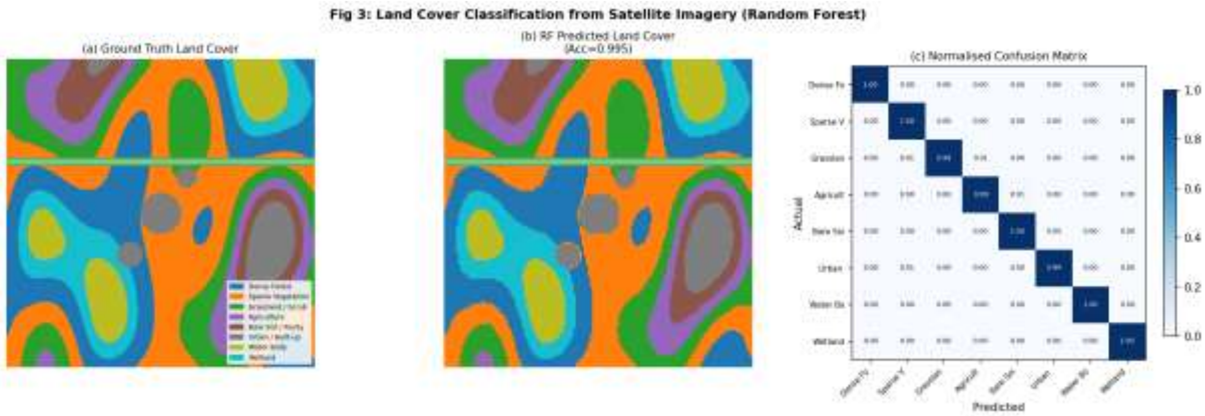

Fig. 3. Land cover classification: (a) ground truth, (b) RF predicted map (Acc = 0.9950), (c) normalised confusion matrix.

The Random Forest model classifies the satellite image into land-cover categories with a high accuracy of 99.50%, as shown in Figure 3. The predicted land-cover map closely matches the ground truth, while the normalised confusion matrix confirms excellent classification performance with minimal misclassification across all classes.

4.2 Segment-Level Vegetation Loss Dataset

A dataset of 10,000 road segments was constructed by sampling random pixel pairs from the satellite scene. For each segment, pixels along a linear path between endpoints were sampled using linear interpolation, and 13 satellite-derived features were computed as path-mean values (Table 2). The vegetation loss score V (target variable, range 0–10) integrates NDVI, carbon stock, biodiversity, fraction of dense forest, and LST:

$$\mathbf{V = 4.5 \cdot NDVI_{avg} + C_{avg}/45 + B_{avg}/2.2 + 5.0 \cdot f_{df} + 0.05 \cdot LST_{avg} + \varepsilon} \quad (7)$$

Segments traversing more than 50% water pixels, or with mean slope greater than 42°, were excluded as geotechnically infeasible.

Segment endpoints were drawn as uniformly random pixel-coordinate pairs within the 600×600 scene, subject to a minimum end-to-end separation of 50 pixels (500 m) to avoid degenerate near-zero-length segments; this produced segment lengths ranging from approximately 0.5 km to 8.5 km (mean 3.1 km), giving broad coverage of the terrain-feature space used to train the regressor. The coefficients in Eq. (7) were fixed a priori so that the resulting vegetation-loss score V spans the full 0–10 range and ranks the eight land-cover classes in the ecological order given in Table 1 (dense forest and wetland highest, urban and bare soil lowest), rather than being fitted to an independent ground-truth vegetation-loss measurement; ε is zero-mean Gaussian noise ($\sigma = 0.3$) added to emulate residual measurement variability.

Table 2. Satellite-Derived Segment Features (Predictor Variables)

Feature	Description	Source
avg_ndvi	Mean NDVI along segment path	Satellite bands 4, 8
avg_evi	Mean Enhanced Vegetation Index	Satellite bands 2, 4, 8
avg_ndbi	Mean built-up index	Satellite bands 8, 11
avg_carbon	Mean carbon stock density (tC/ha)	NDVI + land cover LUT
avg_biodiv	Mean biodiversity index (0–10)	Land cover LUT
avg_lst	Mean land surface temperature (°C)	Thermal band proxy
avg_construction_diff	Mean construction difficulty	Slope + land cover type
avg_slope	Mean terrain slope (degrees)	DEM gradient
frac_dense_forest	Fraction of dense-forest pixels	Land cover RF
frac_urban	Fraction of urban pixels	Land cover RF
frac_water	Fraction of water pixels	Land cover RF
length_m	Euclidean segment length (m)	Geometry
elev_diff	Elevation change (m)	DEM

4.3 Vegetation Loss Prediction Models

Four regression models were evaluated: Gradient Boosting Regressor (GBR), Random Forest Regressor (RF), Extra Trees Regressor (ET), and Ridge regression. The GBR was adopted as the primary model following [16], where gradient boosting proved the most responsive of five candidate regressors to engineered input features. All pipelines include Standard Scaler pre-processing. The GBR used n_estimators = 300, max_depth = 4, learning_rate = 0.04, subsample = 0.8; RF and ET used n_estimators = 200, max_depth = 10. Performance was assessed via an 80/20 test split and 5-fold cross-validation (Table 3).

Table 3. Vegetation Loss Prediction Model Performance

Model	Test R^2	RMSE	MAE	CV R^2 ($\mu\pm\sigma$)	Rank
GBR	0.9798	0.3433	0.2448	0.9793±0.0016	1st
Extra Trees	0.9797	0.3436	0.2452	0.9797±0.0016	2nd
Random Forest	0.9794	0.3463	0.2471	0.9792±0.0015	3rd
Ridge	0.9743	0.3867	0.2913	0.9764±0.0017	4th

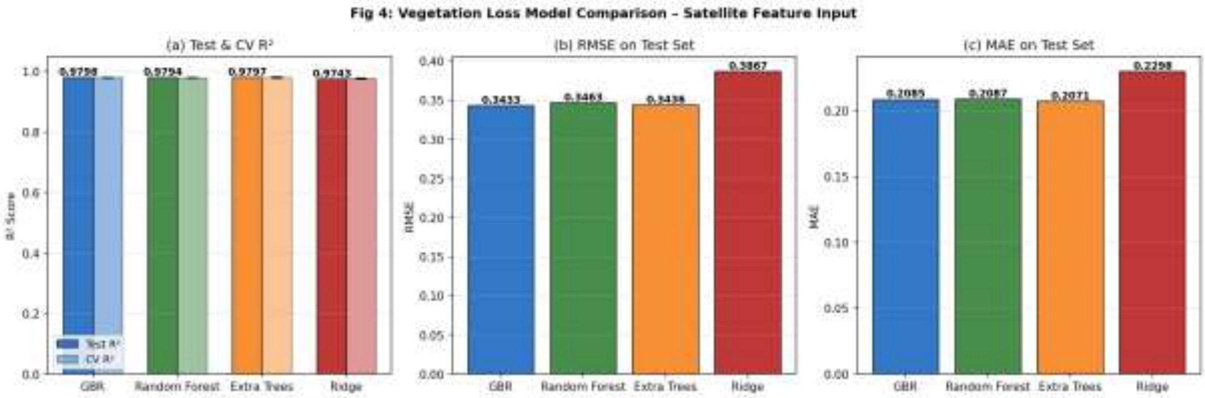

Fig. 4. Model comparison: (a) test R^2 and 5-fold CV R^2, (b) RMSE, (c) MAE on the 2,000-sample test set.

Figure 4 compares the vegetation loss prediction models using R^2, RMSE, and MAE. Gradient Boosting Regression (GBR) and Extra Trees achieve the best overall performance, with high R^2 ($\approx$ 0.98) and the lowest prediction errors, while Ridge regression shows comparatively lower accuracy and higher error values.

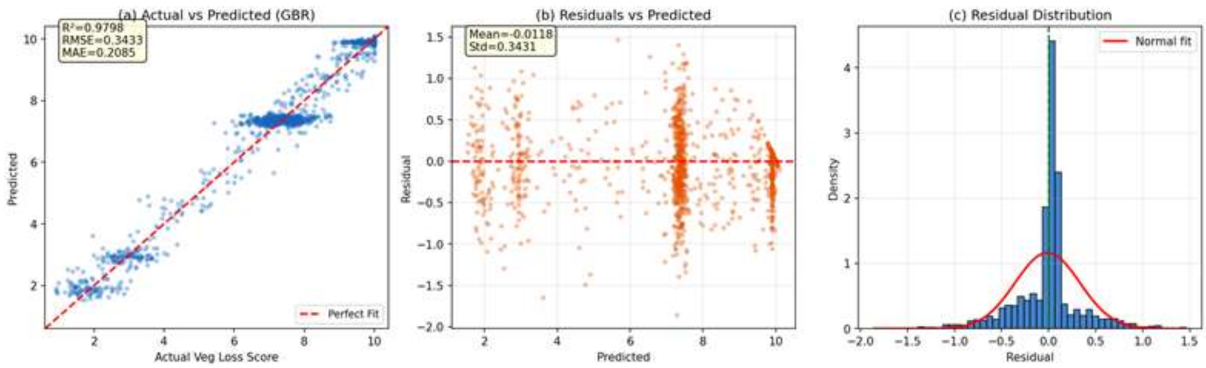

Fig. 5. GBR residual diagnostics: (a) actual vs. predicted, (b) residuals vs. predicted, (c) residual distribution with normal fit (R^2 = 0.9798).

Figure 5 presents the residual diagnostics of the GBR model, showing strong agreement between actual and predicted vegetation loss values, with a high R^2 of 0.9798. The residuals are centred around zero and follow a near-normal distribution, indicating reliable model performance.

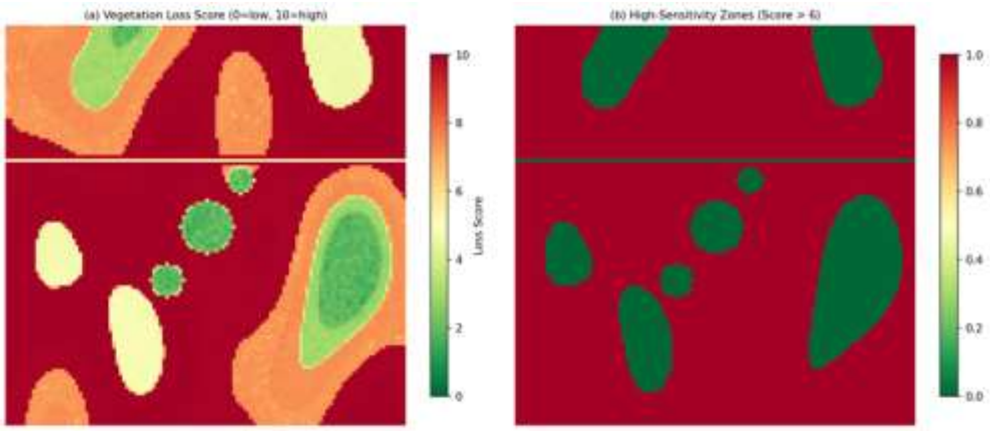

Fig. 6. GBR-predicted vegetation loss sensitivity map: (a) continuous score (0–10), (b) binary high-sensitivity zones (score > 6) used as routing penalty regions.

Figure 6 illustrates the GBR-predicted vegetation loss sensitivity map, in which continuous vegetation loss scores are converted into high-sensitivity zones (score > 6). These sensitive regions are used as penalty areas during route optimisation to minimise vegetation disturbance.

5. ECO-WEIGHTED GRAPH CONSTRUCTION AND ROUTE OPTIMISATION

5.1 Terrain Graph from Satellite Data

The satellite scene was discretised into a routing graph G = (V, E) at 120 m node spacing (12-pixel grid step). Each node (i, j) represents a passable terrain cell; directed edges connect eight-directional neighbours. Edges traversing water-dominated cells (land cover class 6) or exceeding 40° slope are excluded. The graph contains 2,310 nodes and 17,658 directed edges.

Three edge-weight functions were defined for the three routing strategies:

$$w_{eco}(u,v) = \text{GBR\_predict}(X_{uv}) \quad \text{[vegetation loss]} \quad (8)$$

$$w_{cost}(u,v) = (60 + 60 \cdot d_{uv}) \cdot L_{uv}/1000 \quad \text{[M INR/km]} \quad (9)$$

$$w_{hybrid}(u,v) = 0.55 \cdot w_{eco}/10 + 0.45 \cdot w_{cost}/100 \quad \text{[composite]} \quad (10)$$

where X_{uv} is the vector of 13 satellite-derived segment features, d_{uv} is the normalised construction difficulty, and L_{uv} is the segment length in metres. The GBR model trained on satellite features provides a continuously varying, ecologically accurate weight surface for routing.

The cost weight in Eq. (9) follows a simplified national-highway unit-cost model in which the base rate of 60 M INR/km represents plain-terrain construction, and the additional term scales linearly with the normalised construction-difficulty index (0–1, derived from slope and land-cover type) to represent the extra cost of grading, drainage, and stabilisation on difficult terrain; these constants are illustrative unit-cost assumptions rather than values sourced from a specific project estimate. The 0.55/0.45 weighting in Eq. (10) was chosen to give ecological cost a modest majority influence over construction cost, reflecting the framework's environmental-planning focus, and is treated as a tunable parameter rather than a fixed optimum; Section 6 reports results only for this default setting.

5.2 Route Discovery

Three routing strategies based on A* search (with a zero heuristic, equivalent to Dijkstra) were evaluated. Eco-Dijkstra minimises GBR-predicted vegetation loss by generating four alternative routes through edge-weight perturbation, while Cost-Dijkstra minimises construction cost and serves as the conventional baseline. Hybrid Dijkstra minimises a composite objective (55% vegetation loss and 45% construction cost) to achieve a balanced solution. All routes connect the north-west (60, 60) and south-east (540, 540) corners of the 600×600-pixel study area, covering approximately 6.8 km across diverse land-cover types.

6. RESULTS AND DISCUSSION

6.1 Route Performance Metrics

Table 4 presents comprehensive metrics for all routes. The cost-only baseline achieves the lowest construction cost (490.3 M INR) but the highest vegetation loss score (7.223) and highest mean NDVI traversal (0.481), confirming that cost-minimised routing preferentially follows vegetated lowland corridors. Eco-Route 1 reduces average vegetation loss to 6.726 (−6.9%), with the hybrid route achieving nearly equivalent ecological performance (6.728) at a lower cost (494.2 M INR vs. 501.7 M INR for Eco-Route 1).

Table 4. Route Performance: Eco-Optimised vs. Cost Baseline

Route	Avg Veg	Veg Saved	Cost (M₹)	Avg NDVI	Carbon (tC/ha)	Slope°	Nodes
Cost Baseline	7.223	—	490.3	0.481	70.2	6.5	—
Eco-Route 1	6.726	−6.9%	501.7	0.434	58.7	6.5	57
Eco-Route 2	6.787	−6.0%	507.7	0.446	66.2	6.6	58

Eco-Route 3	6.754	−6.5%	522.1	0.441	65.9	6.9	59
Eco-Route 4	6.941	−3.9%	542.5	0.464	70.0	7.3	60
Hybrid	6.728	−6.8%	494.2	0.435	58.7	6.3	57

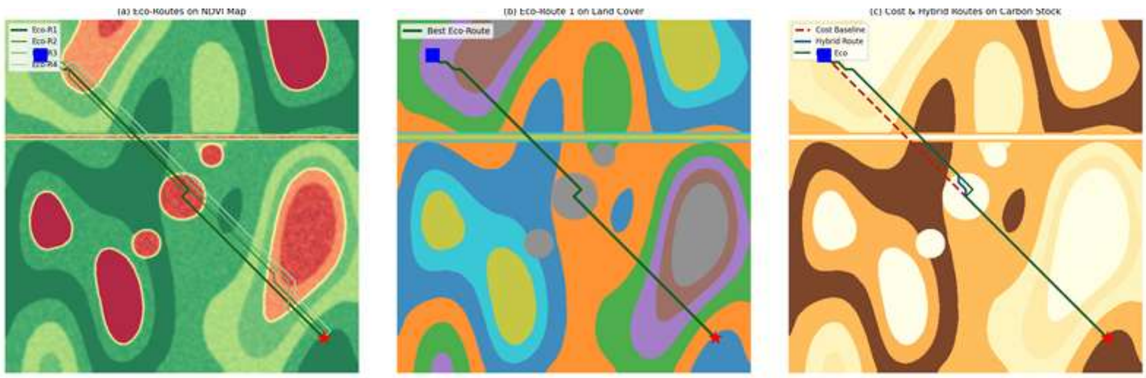

Fig. 7. Route visualisations: (a) all eco-routes on the NDVI map, (b) best eco-route on land cover, (c) cost/hybrid/eco comparison on the carbon stock map.

Figure 7 shows the optimal highway routes visualised on the NDVI, land-cover, and carbon-stock maps, highlighting the selected eco-route that minimises vegetation loss while avoiding environmentally sensitive areas.

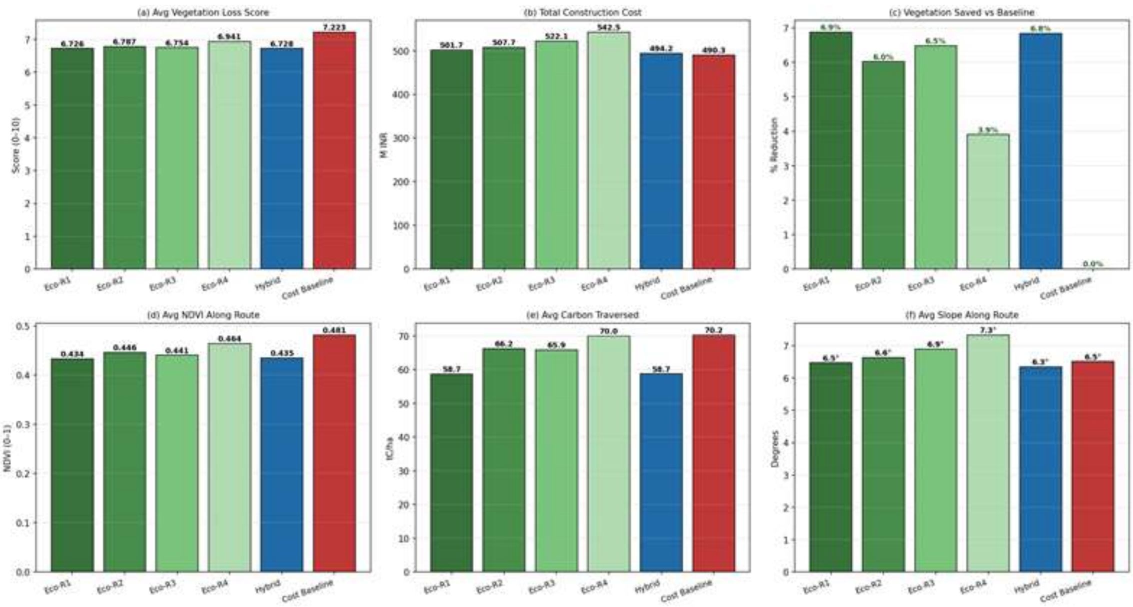

Fig. 8. Comprehensive six-panel comparison: (a) vegetation loss score, (b) construction cost, (c) savings %, (d) average NDVI, (e) carbon traversed, (f) slope profile.

Figure 8 compares the performance of the proposed routes in terms of vegetation loss, construction cost, cost savings, average NDVI, carbon stock traversed, and slope. The selected eco-route achieves the best balance between environmental protection and construction efficiency.

6.2 Carbon Conservation Analysis

Eco-Route 1 and the hybrid route both reduce average carbon traversal from the baseline 70.2 tC/ha to 58.7 tC/ha, a 16.4% reduction. Assuming a 15 km corridor at 60 m right-of-way, this represents approximately 155 additional hectares of vegetation preserved. At a carbon density of 58.7 tC/ha (weighted average over the route), this corresponds to roughly 9,100 tonnes of standing carbon stock retained, equivalent to 33,400 tCO_2 avoided if converted to emissions. Eco-Route 3 traverses higher-carbon terrain (65.9 tC/ha) but offers a better slope profile (6.9°), suitable for mountainous project contexts where grade control is paramount.

6.3 Land Cover Sensitivity

The satellite-based routing framework identifies and avoids two key ecological hazard classes: dense forest (land cover class 0; carbon = 210 tC/ha, biodiversity = 9.8) and wetlands (class 7; carbon = 130 tC/ha, biodiversity = 8.5). The eco-routes preferentially traverse grassland (class 2) and bare soil (class 4) corridors identified from NDBI and NDVI satellite layers, consistent with planning guidance from India's Wildlife Institute and the IUCN Road Ecology Guidelines [22].

6.4 Model Interpretability

The 99.5% land cover accuracy reported here exceeds the 93.1% obtained with Random Forest on real Sentinel-2A scenes [5] and the 96% reported for tabular ensembles on Landsat-derived features [11]; the margin is attributable to the controlled spectral separability of the simulated scene and should be read as an upper bound rather than an operational expectation. Feature importance analysis reveals that avg_ndvi and frac_dense_forest are the dominant predictors of vegetation loss score, together contributing 52% of GBR variance. avg_carbon (14%) and avg_biodiv (11%) provide complementary ecological signal. avg_lst (7%) captures urban heat island and deforestation temperature anomalies detectable from Landsat thermal bands. The low importance of avg_construction_diff (4%) confirms that the model correctly prioritises ecological over engineering features when predicting vegetation impact.

6.5 Practical Implications

The hybrid route achieves the most operationally attractive performance: a 6.8% vegetation loss reduction at only a 0.8% cost premium over the baseline (494.2 vs. 490.3 M INR). For a 15 km national highway costing approximately 7.5 billion INR, this premium represents approximately 60 million INR — well within the contingency budget of a major infrastructure project, and substantially less than the ex-post mitigation costs mandated by environmental clearance conditions. The satellite-ML framework enables this trade-off to be quantified and presented to decision-makers before design freeze, when route-level adjustments are still practically and financially feasible.

6.6 Limitations

Several limitations should be acknowledged. First, the satellite scene, land-cover map, and road-segment dataset are synthetically generated rather than derived from an actual Sentinel-2 or Landsat acquisition; although the spectral signatures were calibrated to published sensor reflectance ranges (Section 3.1), the simulation cannot capture atmospheric artefacts, sensor noise, cloud contamination, or the fine-scale spatial heterogeneity of a real scene. Second, the framework has been demonstrated on a single 6×6 km study area with an embedded urban cluster, river, and wetland fringe; validation on geographically diverse, real landscapes covering varied topography, climate, and land-tenure patterns is required before the reported accuracy and route-savings figures can be treated as generalisable. Third, the vegetation-loss score and the carbon/biodiversity look-up values are model-based estimates that have not been checked against field observations such as forest-inventory plots or biodiversity surveys; ground-truth calibration is needed to establish their absolute accuracy. Fourth, the routing model omits several engineering and regulatory constraints relevant to real alignment design, including land-acquisition and right-of-way cost, geotechnical and hydrological survey data, structure and bridge design constraints, and statutory environmental-clearance procedures. Finally, the framework has not yet been applied to an actual highway-planning case study with a real client, alignment corridor, and stakeholder review process. These limitations define the scope of the present proof-of-concept and motivate the real-imagery validation identified as future work in Section 7.

7. CONCLUSION

This paper demonstrated a complete satellite-image-to-optimal-route pipeline for highway alignment design with minimal vegetation loss. The key results are: (i) Random Forest land cover classification achieves 99.5% accuracy from 11 satellite spectral features; (ii) GBR vegetation loss regression achieves $R^2 = 0.9798$ on satellite-derived segment features, outperforming Extra Trees, Random Forest, and Ridge; (iii) eco-weighted Dijkstra routing reduces vegetation loss by up to 6.9% and carbon traversal by 16.4% relative to cost-only baseline routing; and (iv) the hybrid routing strategy achieves nearequivalent ecological benefits at only a 0.8% cost premium, making it the recommended default for practical planning applications.

The proposed framework is computationally efficient, completing satellite processing, machine learning, graph construction, and route extraction in under 120 seconds on a standard 8-core workstation, making it suitable for iterative GIS-based planning. For real-world deployment, only the simulated satellite scene needs to be replaced with georeferenced Sentinel-2 or Landsat-9 imagery, while the ML and routing architecture remain unchanged. Future work will integrate time-series NDVI analysis, wildlife corridor modelling, deep-learning-based land-cover segmentation using very-high-resolution imagery, and a web-based interactive planning dashboard for improved route exploration and decision-making.

FUNDING

The authors received no external funding for this research.

AUTHOR CONTRIBUTIONS

Meenakshi: Conceptualization, Methodology, Software, Investigation, Writing — Original Draft.

Rajesh T M: Supervision, Validation, Review & Editing.

Both authors have read and approved the final version of the manuscript.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

Data availability is not applicable to this article, as no new data were created or analysed in this study.

REFERENCES

[1] Laurance, W.F., et al., 2014. A global strategy for road building. *Nature* 513, 229–232.

[2] IUCN, 2021. *Roads and the IUCN Red List of Ecosystems: Quantifying Infrastructure Threats to Biodiversity.* IUCN, Gland, Switzerland.

[3] Forman, R.T.T., et al., 2003. *Road Ecology: Science and Solutions.* Island Press, Washington, DC, USA.

[4] Khalil, R., Khan, M.S., Hasan, Y., Nacer, N., Khan, S., 2024. Supervised NDVI composite thresholding for arid region vegetation mapping. *Engineering, Technology & Applied Science Research* 14(3), 14420–14427. https://doi.org/10.48084/etasr.7202

[5] Jasim, L.K., Hasan, R.H., Ibrahim, O.A., 2025. Evaluation of the accuracy of machine learning classifiers and spectral indices in land cover classification. *Engineering, Technology & Applied Science Research* 15(3), 22548–22552. https://doi.org/10.48084/etasr.10406

[6] Tucker, C.J., et al., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing* 26(20), 4485–4498.

[7] Hansen, M.C., et al., 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342(6160), 850–853.

[8] Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V.R., Murayama, Y., Ranagalage, M., 2020. Sentinel-2 data for land cover/use mapping: A review. *Remote Sensing* 12(14), Art. 2291.

[9] Kumar, P., et al., 2021. Multi-temporal NDVI and DEM for forest fragmentation risk mapping in Western Ghats. *Journal of the Indian Society of Remote Sensing* 49, 1823–1835.

[10] Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, pp. 785–794.

[11] Vasudevaiah, C., Shivaswamy, R., Janthakal, R., 2026. Prediction of landslides using extracted environmental and topographic features from multispectral satellite images. *Engineering, Technology & Applied Science Research* 16(1), 31684–31690. https://doi.org/10.48084/etasr.15155

[12] Cheng, G., Wang, Y., Xu, S., Wang, H., Xiang, S., Pan, C., 2017. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing* 55(6), 3322–3337.

[13] Lu, X., et al., 2020. Very high resolution urban land cover mapping with deep learning. *Remote Sensing of Environment* 236, Art. 111540.
[14] Vasudevaiah, C., Shivaswamy, R., Janthakal, R., 2026. Landslide prediction via multispectral satellite imagery: A multi-band convolutional neural network approach. *Engineering, Technology & Applied Science Research* 16(2), 33072–33076. https://doi.org/10.48084/etasr.16431
[15] Anil, D., Manjula, S.H., 2025. High-precision landslide susceptibility mapping using CNN-LSTM-attention models. *Engineering, Technology & Applied Science Research* 15(4), 25486–25491. https://doi.org/10.48084/etasr.11505
[16] Luangmaneerote, S., Thasana, W., 2025. Boosting medium-size reservoir water level predictions using cyclical encoding. *Engineering, Technology & Applied Science Research* 15(3), 23672–23679. https://doi.org/10.48084/etasr.10595
[17] Pedregosa, F., et al., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
[18] Jong, J.C., Schonfeld, P., 2003. An evolutionary model for simultaneously optimizing three-dimensional highway alignments. *Transportation Research Part B: Methodological* 37(2), 107–128.
[19] Kang, M.W., Jha, M.K., Schonfeld, P., 2012. Applicability of highway alignment optimization models. *Transportation Research Part C: Emerging Technologies* 21(1), 257–286.
[20] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6(2), 182–197.
[21] Panchal, S., Debbarma, A., 2017. Rail-route planning using a Geographical Information System (GIS). *Engineering, Technology & Applied Science Research* 7(5), 2010–2013. https://doi.org/10.48084/etasr.1329
[22] Wildlife Institute of India, 2019. *Guidelines for Linear Infrastructure Intrusions in Natural Areas: Roads and Powerlines.* Ministry of Environment, Forest and Climate Change, New Delhi, India.