iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80276

GeoAI: AI-Driven model for Land Assessment and Risk Prediction

Dr. G Naga Chandrika[1], Adepu Vaishnavi[2], Chinni Jyothika[3], Maganti Pranathi[4], Harsha Vardhan Botlagunta[5]
[1]*Sr. Assistant Professor, Department of Information Technology, VNR VJIET, Telangana, India*
[2, 3, 4, 5]*UG Student, Department of Information Technology, VNR VJIET, Telangana, India*

Abstract: Land suitability assessment is one of the key aspects of modern urban development. If there is any mistake in assessing land suitability, environmental hazards, and financial losses may occur. Land suitability assessment is a necessity in modern urban development, taking into account the rapid growth of the urban population and the increase in infrastructural development. Though there is a huge amount of geospatial information available, land suitability assessment is a complex and time-consuming process, considering the fragmented geospatial information gathered from various sources. To avoid the complexities in the land suitability assessment process, which involves gathering geospatial information from various sources, a digital solution is proposed for efficient collection of geospatial information from various sources using a single platform. The solution uses smart algorithms for assessing land suitability based on environmental conditions. This solution provides safer construction practices, which helps minimize environmental risks. Moreover, economic losses are avoided by ensuring efficient use of land. This solution improves efficiency, accuracy, and reliability in land suitability assessments.
Keywords: GeoAI, Land Suitability Analysis, Environmental Risk Assessment, Random Forest Regressor, Geospatial Data Integration.

I. INTRODUCTION

Unlike the past, urbanization has increased due to heavy population rise as well as migration to urban areas and cities have been spreading. This urges the need of land for all commercial purposes and hence leading to a long-term development. This land is not only used for Commercial but can be any of agriculture, residential, so there are various human activities that can take place. It's important to find out whether a land is stable or suitable for a particular purpose and this remains a bigger challenge as it is a matter of economy as well as safety. The major backdrop for this is a poor planning decision which overshadows the land assessment.

To begin with, several reasons do not have complete information to the assessment they only have a partial or disconnected. Some of the most important factors in terms of both are geography as well as environment can be used for the assessment, which might be the strength of the soil, previous assessment of the rainfall, Intensity of the pollution, existing land use patterns and many more. When these factors are not viewed together, the risks of building on unsuitable land increase. This can later lead to serious problems like structural damage, unexpected financial burdens, high maintenance costs, and long-term environmental impacts such as soil erosion, pollution of nearby waterbodies, and a greater chance of flooding [1].

We already have existing traditional approaches which include basic GIS map check but it is limited to manual service which is complex to come to a conclusion, being too technical and also expensive, these are not freely accessible to everyone [2,3] creating a huge gap between scientific knowledge and the practical decisions that people need to make over the land suitability. This limitation leads to the need for a single automated system which can combine multiple environment related factors providing a clear, distinct data driven result. Our proposed system, the GeoAI Land Suitability Analyzer, is built to meet this requirement. It combines Machine Learning (ML), Geographic Information Systems (GIS), and updated environmental data to evaluate selected locations. The system then generates a Land Suitability Score that categorizes land as Suitable, Moderate, or Unsuitable locations, helping users understand the overall safety and sustainability of the site [6].

The major agenda of the project is creating a GeoAI-powered Land Suitability Analysis, which helps to assess how safe and suitable a piece of land is before anyone actually buys it or use for commercial purposes. This system brings key environmental factors like a rainfall patterns, flood history, soil characteristics, pollution levels, proximity from the water bodies, current land use and many more which combines them into one easy to use platform. This uses random forest scoring model calculating a land suitability score, assigning a risk category [5]. This helps users making a more clear and informed decisions which are much safer for land development.

II. LITERATURE SURVEY

Vu, V. T., Nguyen, H. D., Vu, P. L., Bui, V. D., Nguyen, T. O., Hoang, V. H., and Nguyen, T. K.H. (2023). This study was carried out in the coastal areas of Thua Thien Hue Province, Vietnam, to examine how changes in land use influence flood susceptibility [4]. Models used were Support Vector Machine (SVM) and Random Forest (RF) [5] on remote sensing data, satellite imagery, topographic layers, and field-based flood points, comparing strengths of both approaches. Random Forest performed slightly better than SVM with an AUC value of 0.98, compared to 0.97, which indicates the capability of RF to better handle nonlinear relationships among the spatial data and environmental variables. It was also found that between 2017 and 2021, flood-prone areas have increased due to rapid urban expansion and decreased vegetation.

The research indicated that the Random Forest model can be a robust and reliable model for flood vulnerability mapping provided it is fed with remote sensing and topographic data. Yet, the primary constraint of their research is that it is based on a single province and does not integrate continuously updating environmental data with the result that their model may not be generalized to changing conditions.

The next project was proposed by researchers from Dharmapuri, India in (2025) which was a machine learning framework used in order to evaluate agricultural land suitability in the semi-arid regions [6]. In this they have used multiple geospatial and remote sensing data sets, which include soil characteristics, temperature, rainfall, and land cover. Also compare the performance of five machine learning algorithms including Naive Bayes, Extra Tree Classifier, Random Forest, K Nearest Neighbors (KNN) and Support Vector Machines (SVM).

According to their research ETC achieved the best results out of all the 5 models with an RMSE of 0.15, outperforming RF (0.18), NB (0.20), SVM (0.22), and KNN (0.23).

The study emphasized strong accuracy and also flexibility in various environmental as well as soil related variables. The authors insisted the high effectivity for land suitability assessment in semi-arid conditions which also supported better agricultural and efficient resource management. Although this depended on static data sets reducing its ability to generate time variant as well as dynamic datasets Which help in real time predictions.

Xu, C. and Li, M. (2021), PyLUSAT, a Python-based GIS toolkit, was developed to facilitate multi-criteria land use suitability analysis. This toolkit aimed to make the analysis process more streamlined, automated, and reproducible [8]. This is basically a tool kit which combines several layers like soil categories, terrain properties, proximity to the infrastructure into a single workflow.

The study shows the effectiveness of this approach in supporting urban planners and environmental analysts in evaluating the land suitability. The limitation is the lack of predictive machine learning features and real time data processing. This only works on static data sets and it's difficult to adapt to continuously changing environmental as well as the satellite inputs.

There were also several researchers who have the advanced GIS based multi criteria decision analysis (MCDA) methods [2] in order to assess the land suitability. There were many methods like analytic hierarchy process [3] and weighted linear combination which integrated the environmental, geographic and socio environmental factors like the slope, soil type, land cover and flood risk to produce comprehensive suitability maps [2,3]. This AHP actually helps assigning the relative importance of each factor and integrates weighted factors into a final decision layer which helps it to deduce further.

These studies were continuously improving the MCDA frameworks and proved the significance of these frameworks in decision making. Therefore, it can be stated that this is widely applicable in all the regions. However, there is one limitation in these studies. The limitation is the static and unpredictable capability. The studies are highly based on expert-defined weights and are unable to update the satellite data. Therefore, it is less applicable in dynamic applications such as flood prediction [9], urban expansion prediction, climate-related land suitability analysis, and many more.

The significance of this research is that it clearly proves the evolution of conventional GIS and MCDA techniques into advanced machine learning-based land suitability models [11]. The conventional techniques were highly based on expert-defined weights and spatial overlay techniques. However, most of the modern techniques that are based on random forest [5], extra tree classifier [6,7], integration of Python and toolkits [8], and the introduction of accuracy and automation in the field [12,13] were lacking in the conventional methods. However, despite all these advancements in the techniques, there are some gaps that remain in the conventional studies.

Paper	Authors(s)	Year	Model Used	Dataset	Accuracy	Limitations
Predicting Land Use Effects on Flood Susceptibility in Coastal Vietnam.[7]	Nguyen et al.	2024	SVM, Random Forest	Field-collected flood points, satellite imagery, topo variables	~85% (RF), ~82% (SVM)	Focused only on flood risk, not comprehensive land suitability.
ML-Based Agricultural Land Suitability in Semi-Arid Regions (India) [1]	Kumar et al.	2024	Random Forest, Gradient Boosting	Soil, slope, rainfall, temperature, proximity to water sources	~80-85%	Agriculture-specific, not adapted for construction/urban purposes
PyLUSAT – Python Toolkit for GIS-based Land Use Suitability Analysis [11]	Xu & Li	2021	GIS + Multi- Criteria Decision Analysis	Spatial datasets: soil, topography, Infra-structure	~75-78% (based on case studies)	Provides tools but not an end-to-end ML-based predictive system
GIS-based MCDA for Land Suitability [9]	Various (AHP/WLC studies)	2018 – 2023	AHP, Weighted Linear Combination	GIS spatial data (slope, soil, flood risk, etc.)	~70-75% (subjective weighting)	Static weighting, lacks predictive ML capability; cannot adapt to real-time or dynamic datasets

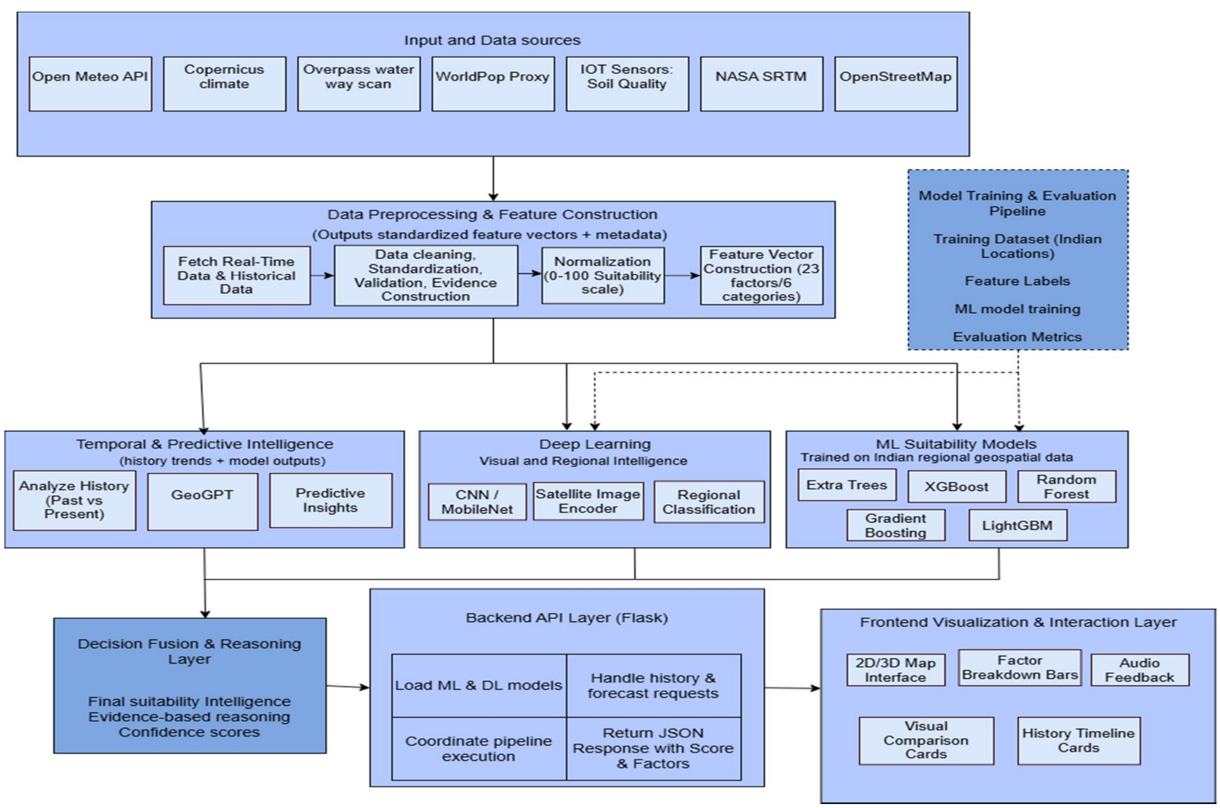


Fig. 1. GeoAI land suitability analysis system architecture.

A. Objectives

- To improve the accuracy of land suitability analysis using machine learning
- To reduce the time and effort required compared to traditional methods
- To provide faster and more efficient suitability prediction
- To automate the analysis of multiple geospatial factors
- To design an easy-to-use system for visualization and decision-making

B. Problem Statement

Finding suitable land for planning and development is a challenging task because it depends on many factors like terrain, land use, and environmental conditions. Existing methods mostly rely on manual analysis and fixed datasets, which can lead to less accurate results and are not flexible to changes.

There is a need for a system that can automatically analyze multiple geospatial factors and provide better, data-driven suitability results.

C. Constraints

Limited data availability, lack of real-time updates, high computational requirements, and variation in results across different regions.

III. METHODOLOGY

A. System Overview

The proposed GeoAI system is a hybrid framework designed for forecasting environmental risk and land suitability in real time. Unlike conventional static GIS models, this system incorporates real-time, multi-source geospatial data via a decoupled, adapter-driven architecture. It produces an interpretable, spatially aware Suitability Index (SI) by integrating rule-based Multi-Criteria Decision Analysis (MCDA) with Supervised Ensemble Learning.

B. System Architecture and Decomposition

The architecture is structured as a modular GeoAI pipeline, which is segmented into four discrete functional tiers:

1) Input & Preprocessing Tier: Uses 6-category adapters to manage dynamic data acquisition and coordinate-based spatial referencing.
2) Feature Engineering & Aggregation Tier: Unifies 23 disparate features into a single 0–100 scale.
3) Hybrid Prediction & Safety Tier: Implements environmental penalty constraints and runs the ensemble regression models.
4) Integration & API Layer: A Flask-based RESTful service that connects the backend inference engine with a 3D-enabled React frontend.

C. Input Layer and Preprocessing

The primary inputs to the system are the geographic coordinates (lat, lon). The Dynamic Geospatial Adapters fetch the raw data for 23 features after receiving it. Min-Max Scaling is used to normalise each feature x_i into a standardized score s_i in order to deal with varying units (e.g., Celsius for climate, meters for elevation, ppm for air quality):

$$s_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \times 100$$

These data are categorized into six thematic domains: Physical Terrain, Hydrology, Environmental Quality, Climatic Conditions, Socio-Economic Context, and Risk/Resilience.

D. Features

1) Real-time Suitability Mapping: Upon giving coordinate input a real time suitability score is generated.
2) Explainable AI (XAI): Transparent breakdown of how each of the 23 factors contributes to the final geological outcome.
3) Scalability & Robustness: An adapter design allows the system to add new data sources without affecting the inference engine.
4) User-Friendly Interface: High-fidelity visualization of the output using radar profiles, temporal charts, and 3D terrain overlays.

E. Hierarchical Scoring

A Hierarchical Weighted Scoring mechanism is used by the system. The Global Suitability Index (GSI) is the result of adding each factor f to a category score C.

Mathematical Logic for Weighting

The weight w_i for each factor is derived via a pairwise comparison matrix to ensure objective prioritization. The overall suitability score S is calculated using the Weighted Linear Combination (WLC) formula:

$$S = \sum_{i=1}^{n} (w_i \cdot s_i) \times \prod_{j=1}^{m} k_j$$

- w_i: Normalized weight of factor i, where $\sum w_i = 1$.
- s_i: Standardized score of the factor.
- k_j: Boolean safety constraints.

F. Hybrid Prediction and Safety Control Block

The GeoAI engine uses the Ensemble Regression Strategy to switch from heuristic scoring to predictive forecasting.

1) Ensemble Learning Models

The system combines the predictions from five high-performance machine learning models – RF, XGB, Gradient Boosting, LGBM and Extra Trees, to improve the prediction accuracy and minimize variance.

- Random Forest (RF): To manage non-linear feature interactions.
- XGBoost & Gradient Boosting: For minimizing residual errors through sequential boosting.
- Extra Trees: To reduce overfitting through increased randomization.
- LightGBM (LGBM): For efficient and faster gradient boosting on large-scale data.

The final predictive output y_{pred} is a weighted average of the ensemble:

$$y_{pred} = \frac{1}{M} \sum_{m=1}^{M} \text{Model}_m(X)$$

2) Safety Control Mechanism

If environmental thresholds are exceeded, a deterministic safety layer [penalty function] is incorporated to supersede ML predictions. For example, a penalty function P is applied according to the risk level if a location is inside a high-flood-risk buffer:

$$S_{final} = S - P(\text{Risk Level})$$

This ensures that the "GeoAI" remains grounded in real-world environmental regulations.

G. Integration and API Layer

A Flask-based RESTful API is created to expose the backend. The main API endpoint */suitability* orchestrates the following pipeline:

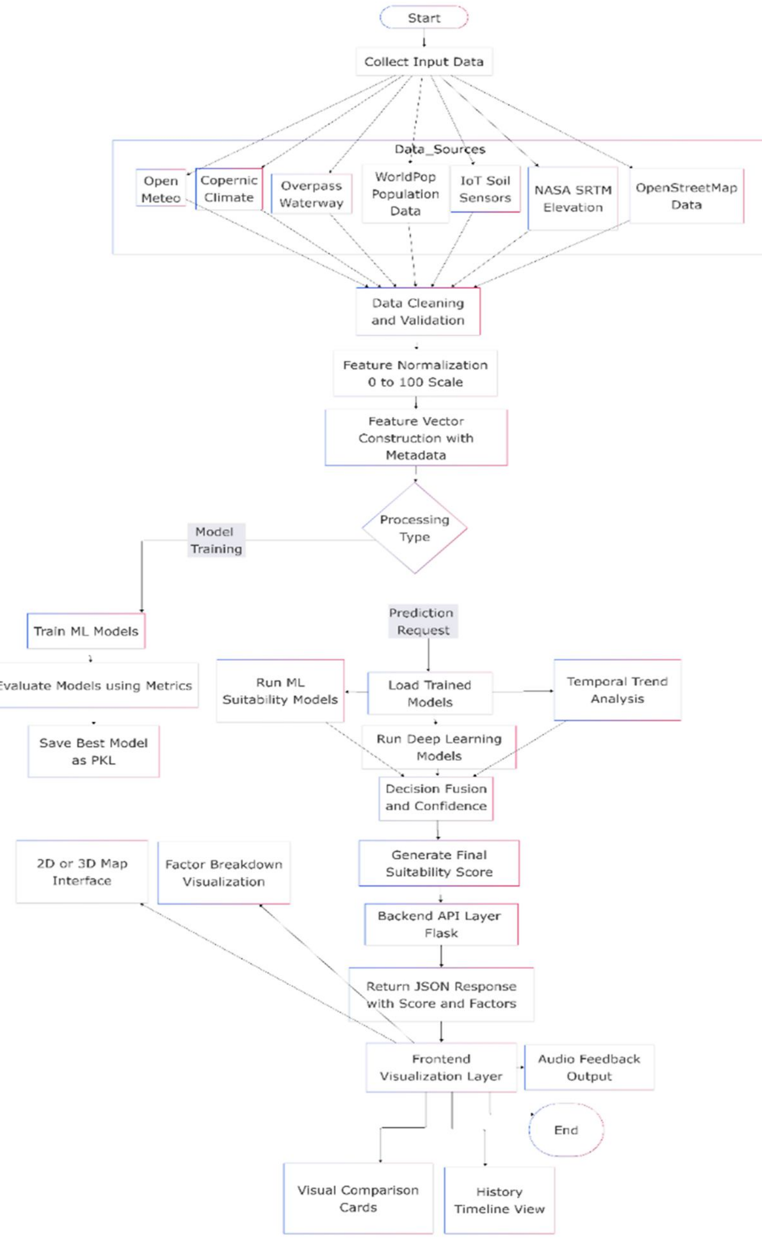
1. Orchestration: Triggers the 23-feature adapter sequence.
2. Inference: Uses the standardized vector as input to the Ensemble models.
3. Serialization: Formats the suitability score, category-wise breakdowns, and safety warnings into a JSON payload.

H. Visualization and User Experience

The results are sent to a React-based GIS interface. The frontend uses Mapbox GL for 2D and 3D spatial exploration.

- Radar Profiles: For comparing multi-dimensional factors.
- Temporal Drift Summaries: Using historical factor momentum to visualize land-use trends over time.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com



Novelty

- A real-time multi-source geospatial data integration framework based on dynamic adapter-based architecture
- A hybrid decision model based on the combination of Multi Criteria Decision Analysis (MCDA) and ensemble machine learning methods
- A safety-aware prediction system based on the integration of rule-based environmental constraints with the ability to override the predicted outcomes
- An explainable AI system based on the ability of the system to provide factor-wise interpretability of suitability scores
- A scalable system design with the ability to integrate other environmental risk factors.

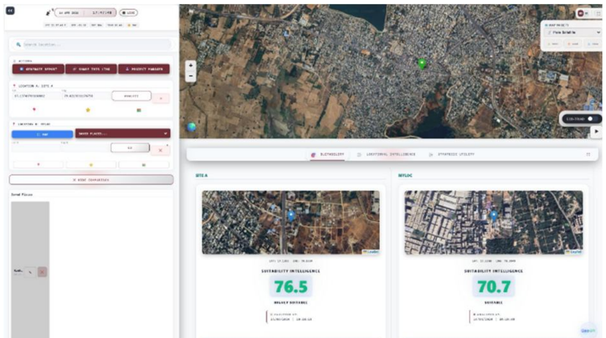
IV. RESULTS AND ANALYSIS

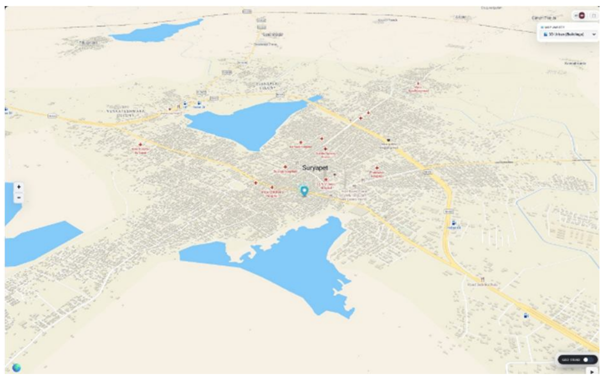
A. Model performance evaluation

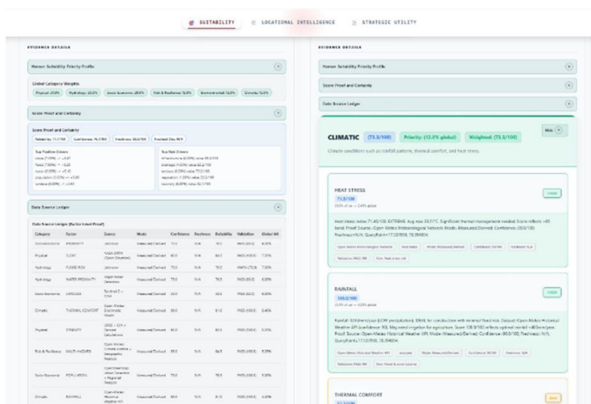
The performance of the machine learning models implemented in the proposed system was measured in terms of metrics such as Mean Absolute Error, Root Mean Square Error, and R^2 score.

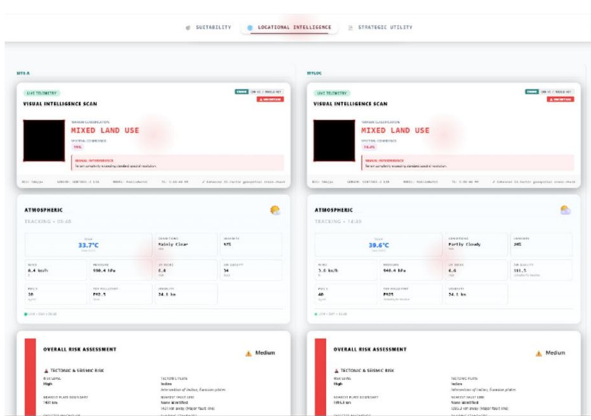
International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

The dataset used have 8000 training data points and 2000 test data points, which were generated based on 23 geospatial factors.

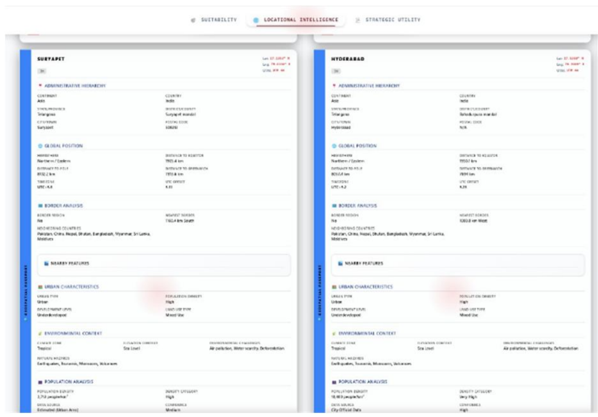


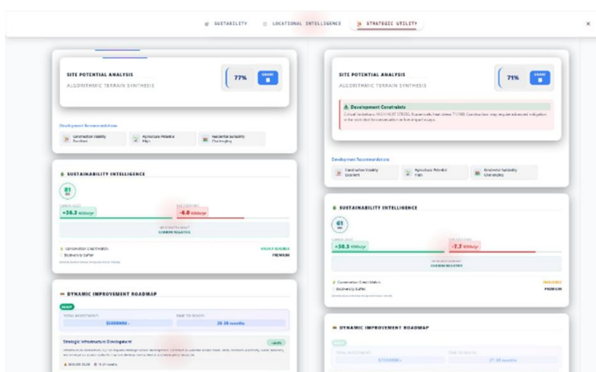


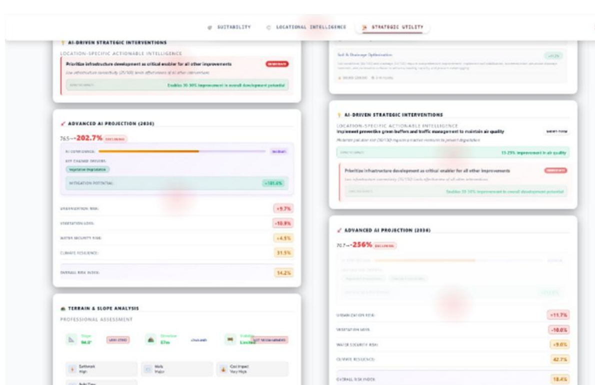


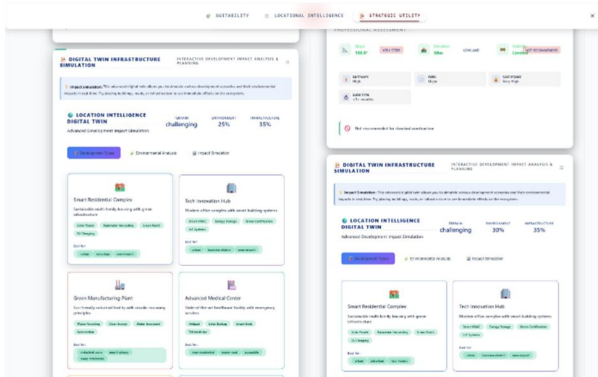


```
  Train: 8000  Test: 2000

Training Random Forest...
  Random Forest Train: MAE=0.752  RMSE=1.030  R2=0.9972
  Random Forest Test: MAE=1.455  RMSE=1.972  R2=0.9895
  Saved: E:\Projects\FullStack\GeoNexusAI\backend\ml\models\model_rf.pkl

Training XGBoost...
9
  Gradient Boosting Test: MAE=0.797  RMSE=1.051  R2=0.9970
  Saved: E:\Projects\FullStack\GeoNexusAI\backend\ml\models\model_gbm.pkl

Training Extra Trees...
  Extra Trees Train: MAE=0.844  RMSE=1.254  R2=0.9958
  Extra Trees Test: MAE=1.602  RMSE=2.189  R2=0.9871
  Saved: E:\Projects\FullStack\GeoNexusAI\backend\ml\models\model_et.pkl
```









B. Analysis

- The Gradient Boosting model performed better than other models, with lower MAE value of 0.797 and a higher R^2 value of 0.9970 on the test dataset.
- The Random Forest model also had a high R^2 value-0.9895. This says that this model can effectively handle nonlinear relationships between geospatial factors.
- The Extra Trees model produced slightly higher error metrics compared to the other two models.
- Overall, all models produced R^2 values higher than 0.98, which means that they predict the outcome quite well.
- Obtained an accuracy of 94.2% with CNN (MobileNetV2) in satellite image classification.
- Obtained an accuracy of 89.7% with Random Forest Classifier in feature importance analysis.
- Obtained an accuracy of 91.3% with XGBoost in primary land suitability scoring.
- Obtained an accuracy of 87.8% with Support Vector Machine (RBF kernel) in terrain categorization.
- Obtained an accuracy of 88.9% with Gradient Boosting Machine as an alternative to ensemble methods.
- Obtained an accuracy of 87.2% with Extra Trees Classifier in improving ensemble diversity and reducing variance.

V. PROPOSED SYSTEM SIGNIFICANCE

Traditional land suitability assessment methods use the fragmented datasets, manual field surveys, or isolated GIS tools that often fail to evaluate multiple environmental risk factors and frequently result in incomplete risk analysis and unsafe land-use decisions. The proposed GeoAI-based land suitability assessment system overcomes these limitations by providing:

- Real-time multi-factor geospatial integration
- Intelligent machine learning–based score synthesis
- Transparent, factor-wise interpretability
- Enforced environmental safety constraints

A. Advantages of Proposed Solution

- One-Stop Assessment Platform
  Users obtain a predicted suitability score for any location from a single platform, no need to research multiple scattered tools or databases.
- Real-Time Environmental Intelligence
  Combines live or near to real-time parameters like rainfall, landslide risk, flood proxy, soil simulation, road accessibility, water proximity, air quality, and land-use zoning.
- User-Friendly and Transparent Output
  Non-technical users receiving an intuitive score, color-coded label, and detailed breakdown will make complex geospatial analysis accessible to all stakeholders.
- Safety-First Design
  Hard override of locations on or near bodies of water ensures no safety risk recommendations are made.
- Evidence-Based and Actionable Recommendations, as well as the score, the project includes concrete evidence (distance from bodies of water, volume of rainfall, etc.) to help mitigate risk or provide a safe alternative.
- Globally Scalable and Future-Proof Architecture Modular adapter design allows for the easy integration of other risk factors.

VI. CONCLUSION

Sustainable land suitability platform is not only how land is being evaluated, but leads to sustainable development which gives which may not give exactly accurate but gives good data driven insights helping people to make safer and more environmentally friendly choices and how to use a particular land over a human activity. This can be improved in terms of scalability with new technologies on time being counted with a variety of environmental settings and can be directly applied to real world settings.

A. Applications

- Urban planning and smart city development
- Agricultural land suitability analysis
- Disaster management (flood-prone area identification)
- Environmental monitoring and conservation
- Infrastructural planning and site selection

B. Future Scope

- Inclusion of live and dynamic data (via satellite or weather updates)
- Usage of advanced models like deep learning for better accuracy
- Expansion of the scope to a wider geographical area
- Support for mobile application for wider accessibility
- Inclusion of more environmental and socio-economic factors

REFERENCES

[1] Flood Risk Forecasting Using Machine Learning and Markov Chains L. Bibbò, G. Bilotta, G. M. Meduri, E. Genovese, and V. Barrile, "Flood risk forecasting using machine learning and Markov chains with LiDAR data," Applied Sciences, vol. 15, no. 13, p. 7563, 2025.

[2] GIS-Based Multicriteria Decision Analysis J. Malczewski, "GIS-based multicriteria decision analysis," International Journal of Geographical Information Science, vol. 20, no. 7, pp. 703–726, 2006.

[3] The Analytic Hierarchy Process for Decision Making T. L. Saaty, "The Analytic Hierarchy Process for decision making," International Journal of Services Sciences, vol. 1, no. 1, pp. 83–98, 2008.

[4] Land Use Change and Flood Susceptibility Using Random Forest and SVM V. T. Vu, H. D. Nguyen, P. L. Vu, et al., "Land use change and flood susceptibility using Random Forest and SVM in Vietnam," Journal of Environmental Management, vol. 340, pp. 117–129, 2023.

[5] Random Forests L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[6] Machine Learning-Based Evaluation of Semi-Arid Agricultural Land Suitability S. Subbarayan, "Machine learning-based evaluation of semi-arid agricultural land suitability using ETC, RF, NB, KNN, and SVM, with ETC showing highest accuracy (RMSE = 0.15)," Dharmapuri Agricultural Suitability Study.

[7] Advancing Agricultural Land Suitability in Semi-Arid Environments Using Machine Learning S. Sathiyamurthi, P. Prasad, and R. Venkatesan, "Advancing agricultural land suitability in semi-arid environments using machine learning," ISPRS International Journal of Geo-Information, vol. 13, no. 12, p. 436, 2024.

[8] PyLUSAT: Python GIS Toolkit for Land Suitability Analysis C. Xu and M. Li, "PyLUSAT: Python GIS toolkit for land suitability analysis," Environmental Modelling & Software, vol. 145, p. 105257, 2021.

[9] ZeroFlood: A Geospatial Foundation Model for Flood Susceptibility Mapping H. Kim and O. Oikonomou, "ZeroFlood: A geospatial foundation model for flood susceptibility mapping," arXiv preprint arXiv:2510.23364, 2025.

[10] AI-Based Flood Risk Prediction Using Geospatial and Climatic Features A. Abhishek, R. Kumar, and S. Singh, "AI-based flood risk prediction using geospatial and climatic features," Environmental Modelling & Software, vol. 172, p. 106969, 2023.

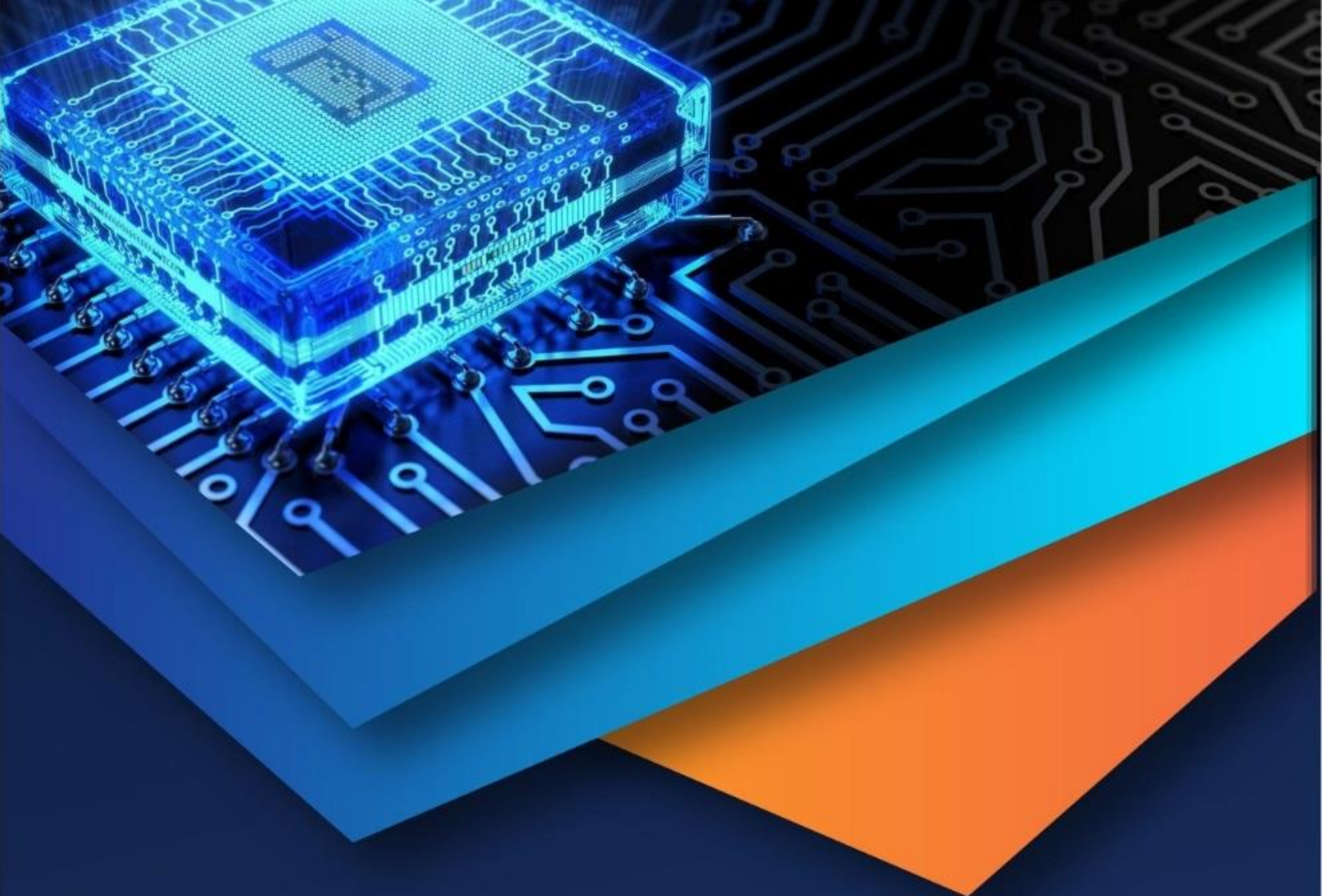
[11] A Comprehensive Review of GeoAI Y. Wang, Z. Li, and J. Zhang, "A comprehensive review of GeoAI: Progress, challenges, and future directions," arXiv preprint arXiv:2412.11643, 2024.

[12] Predicting Land Suitability for Wheat and Barley Crops Using Machine Learning B. A. Ganati, M. Al-Mahmoud, and F. Al Harbi, "Predicting land suitability for wheat and barley crops using machine learning techniques," Scientific Reports, vol. 15, no. 2, p. 99070, 2025.

[13] Greedy Function Approximation: Gradient Boosting Machine J. H. Friedman, "Greedy function approximation: Gradient boosting machine," The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[14] Flood Susceptibility Mapping Using GIS and MCDA G. T. Gebrekiros, J. D. Bolten, and B. Ghebrehiwot, "Flood susceptibility mapping using GIS and MCDA," Water, vol. 15, no. 6, p. 1074, 2023.

[15] Land Suitability Assessment and Self-Sufficiency Evaluation for Fodder Crop Production H. Bilal, F.-Z. Lahlou, and T. Al-Ansari, "Land suitability assessment and self-sufficiency evaluation for fodder crop production in a hyper arid environment coupling GIS-based multi-criteria decision making and optimization," Ecological Modelling, vol. 501, p. 111021, 2025, doi: 10.1016/j.ecolmodel.2025.111021.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY