JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# A Conceptual Framework for AI-Assisted Preliminary Geotechnical Site Investigation Using Remote Sensing, GIS and Spatial Data Analytics

**[1]Chayan Biswas, [2]Sushmita Ghosh, [3]Anirban Maity**

[1]Assistant Professor, [2]Assistant Professor, [3]UG Student
[1]Civil Engineering Department,
[1]Swami Vivekananda University, Barrackpore, North 24 Pargana, West Bengal.

***Abstract:*** Preliminary geotechnical site investigation is a fundamental component of infrastructure planning because decisions concerning foundations, earthworks, slopes, and excavation, drainage and construction methods depend on an adequate understanding of subsurface conditions. Conventional investigation methods, including boreholes, trial pits, standard penetration tests, cone penetration tests, geophysical surveys and laboratory testing, provide essential ground-truth information but are spatially discrete, relatively expensive and often insufficient to fully represent heterogeneous geological environments. Recent advances in remote sensing, geographic information systems (GIS), spatial data analytics and artificial intelligence (AI) create an opportunity to complement conventional investigation through spatially continuous preliminary assessment and uncertainty-informed investigation planning.

This paper proposes a conceptual framework for AI-assisted preliminary geotechnical site investigation that integrates multisource remote sensing, digital elevation models, geological and geomorphological information, hydrological and land-use data, legacy geotechnical records, GIS-based spatial databases, machine-learning algorithms, spatial statistics and uncertainty quantification. The framework comprises seven principal stages: (1) multisource data acquisition, (2) data quality control and spatial harmonization, (3) geotechnical feature engineering, (4) AI-assisted spatial prediction, (5) hazard and uncertainty assessment, (6) investigation-priority optimization, and (7) field validation and iterative model updating. Particular emphasis is placed on the distinction between remotely observable proxies and actual engineering parameters. Remote sensing and AI are therefore positioned as preliminary screening and decision-support technologies rather than substitutes for direct subsurface investigation.

***Index Terms* - Artificial intelligence; Geo AI; geotechnical investigation; remote sensing; GIS; machine learning; spatial analytics; uncertainty quantification; digital ground model.**

## *1 Background*

Geotechnical engineering is fundamentally concerned with the interaction between civil infrastructure and the ground. The design and construction of foundations, retaining structures, slopes, embankments, tunnels, roads, railways, dams and underground infrastructure require adequate knowledge of geological conditions and the engineering behaviour of soils and rocks. Unlike manufactured construction materials, however, natural ground is spatially variable, incompletely observed and affected by geological history, weathering, groundwater, geomorphological processes and human modification.

The purpose of geotechnical site investigation is consequently not merely to obtain a collection of soil or rock samples. Its broader purpose is to develop a sufficiently reliable conceptual model of the ground to support engineering decisions. This conceptual model normally evolves through desk study, geological interpretation, field reconnaissance, boreholes, trial pits, in-situ testing, geophysical investigation and laboratory testing.

A fundamental limitation of conventional site investigation is that subsurface information is obtained at discrete locations. A borehole provides highly valuable information at a particular point and depth, but conditions between boreholes must be inferred. The resulting interpretation is therefore an inherently spatial problem.

This problem becomes particularly significant where ground conditions vary rapidly over short distances. Examples include alluvial deposits, coastal and estuarine environments, residual soils, weathered rock profiles, reclaimed land, urban fill, landslide terrain and areas affected by groundwater fluctuation. The problem is further complicated where historical investigation data are incomplete, inconsistently recorded or unavailable in digital form.

Recent developments in artificial intelligence (AI), remote sensing and geographic information systems (GIS) provide an opportunity to address part of this problem. AI and machine-learning techniques can identify nonlinear relationships between observed geotechnical conditions and environmental predictors. Remote sensing provides spatially extensive information on terrain, land cover, surface moisture, geological expression and ground deformation. GIS provides the spatial environment in which these datasets can be integrated.

AI applications in geotechnical engineering have expanded considerably in recent years. A broad review of AI in geotechnics identifies applications across constitutive modelling, site characterization, slope stability, foundations, tunnelling, liquefaction and other areas, while also highlighting continuing challenges associated with data quality, interpretability, generalisation and engineering adoption. Recent work specifically examining the maturity of AI and machine learning in geotechnical engineering suggests that technical development is progressing faster than practical adoption, reinforcing the need for transparent and engineering-oriented frameworks. Traditional preliminary geotechnical investigation frequently relies on a combination of available geological maps, topographic information, historical reports and engineering judgement. Although these sources remain indispensable, their integration is often qualitative and discontinuous.

At the same time, large volumes of geospatial information are increasingly available. Satellite imagery, digital elevation models, SAR data, LiDAR, UAV imagery, geological datasets and historical land-use information can potentially provide useful indicators of terrain and environmental conditions. Yet these data are rarely incorporated into a unified preliminary geotechnical decision-support system. There is consequently a methodological gap between: remote sensing and spatial observation and geotechnical investigation planning.

Existing AI studies often focus on predicting an individual engineering variable—for example, a soil property, bearing capacity, settlement or slope-safety indicator. Such predictions can be valuable, but they do not necessarily answer the practical question of how AI should influence the design of the investigation program itself.

An engineering-oriented framework should therefore progress beyond prediction toward decision support under uncertainty.

## *2 Aim & Objectives:*

The primary aim of this research is to develop a conceptual framework for AI-assisted preliminary geotechnical site investigation using remote sensing, GIS and spatial data analytics.

1. develop a multisource spatial data architecture for preliminary geotechnical investigation;
2. identify remote-sensing and GIS-derived variables that can serve as geotechnical proxies;
3. establish an AI-based spatial prediction methodology;
4. incorporate uncertainty quantification into preliminary ground assessment;
5. develop a conceptual Investigation Priority Index;
6. link spatial predictions to field-investigation planning;
7. establish an iterative field-validation and model-updating mechanism;

### *2.1 Artificial intelligence in geotechnical engineering*

AI has become increasingly important in geotechnical engineering because many engineering relationships are nonlinear and difficult to represent using simple analytical equations.

Common machine-learning methods include:

- artificial neural networks;
- support vector machines;
- decision trees;
- random forests;
- gradient boosting;
- extreme gradient boosting;
- Gaussian-process models;
- convolutional neural networks; and
- resemble learning.

These methods can learn relationships between predictor variables and observed engineering responses.

The recent literature indicates a rapid expansion of AI applications in geotechnics. Sheil et al. describe AI as a potentially transformative technology for geotechnical engineering while identifying important challenges related to data, physics, uncertainty and practical implementation.

However, predictive accuracy alone is insufficient for engineering adoption. A model may produce a statistically accurate prediction while remaining difficult to interpret, physically implausible outside its training domain, or unreliable where training data are sparse.

The proposed framework therefore adopts five principles:

1. **data quality before model complexity;**
2. **spatial validation rather than random validation alone;**
3. **uncertainty reporting alongside prediction;**
4. **engineering interpretation alongside statistical performance; and**
5. **field validation before engineering application.**

The conceptual ground model can be expressed as a spatially varying function:

[ G(x,y,z)=f(X_1,X_2,,X_n,) ]

where:

- (G) represents the ground condition;
- (x,y,z) represent spatial position;
- (X_i) represents an environmental or geotechnical predictor;

- () represents unexplained variability.

*Remote sensing offers several advantages for preliminary site investigation:*

- **broad spatial coverage;**
- **repeated observations;**
- **historical data availability;**
- **relatively rapid acquisition;**
- **integration with GIS;**
- **detection of geomorphological patterns; and**
- **potential monitoring of deformation.**

### *2.2 Optical imagery*

Multispectral imagery can provide information on vegetation, surface moisture, land cover and exposed soil or rock. Spectral indices can be used to identify differences in surface conditions.

Potential variables include:

[ NDVI= ]

and moisture-related indices derived from near-infrared and short-wave infrared bands.

These variables should be interpreted as proxies rather than direct measurements of soil engineering properties.

---

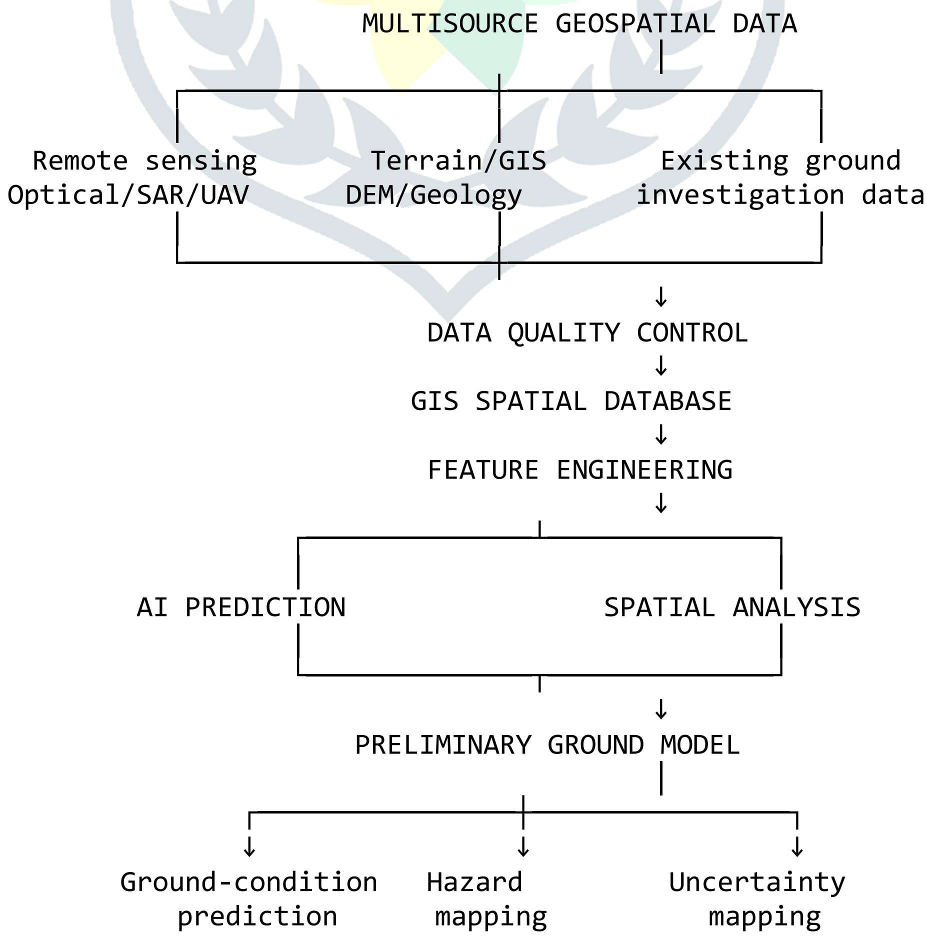
### *2.3 Overall architecture*

The proposed architecture comprises seven interconnected stages:

1. multisource data acquisition;
2. data quality control and harmonisation;
3. spatial feature engineering;
4. AI-based prediction;
5. hazard and uncertainty assessment;
6. investigation optimisation; and
7. field validation and model updating.

The framework is illustrated conceptually below.

*Figure 1. Proposed AI-assisted preliminary geotechnical investigation framework*



```
└───────────────────┼───────────────────┘
                              ↓
          INVESTIGATION PRIORITY
                              ↓
     TARGETED FIELD INVESTIGATION
                              ↓
Boreholes / CPT / SPT / Geophysics
                              ↓
                    VALIDATION
                              ↓
            MODEL CALIBRATION
                              │
                    └─────────► FEEDBACK
```

***3. Target Variables***

The framework distinguishes three categories of outputs.

*3.1 Geological outputs*

Examples:

- geological unit;
- geomorphological unit;
- lithological class;
- weathering class.

*3.2 Preliminary ground-condition outputs*

Examples:

- soft-ground probability;
- shallow-rock probability;
- fill/reclamation probability;
- groundwater potential;
- relative ground-condition class.

*3.3 Hazard outputs*

Examples:

A five-level conceptual classification is proposed.

*G1 — Favourable*

Low predicted hazard and relatively low uncertainty.

*G2 — Generally favourable*

Generally favourable conditions with moderate spatial variability.

*G3 — Intermediate*

Moderate uncertainty or potentially variable conditions.

*G4 — Potentially problematic*

Evidence suggests conditions requiring focused investigation.

*G5 — High priority*

High hazard, high uncertainty, or both.

The G1–G5 system is not intended to replace formal geotechnical classification standards. It is an investigation-planning classification.

### *4. Hazard Modelling*

*4.1 Landslide susceptibility*

A generic susceptibility model can be written as:

[ LS=f(S,C,G,R,D,L,V) ]
where:

- (S) = slope;
- (C) = curvature;
- (G) = geology;
- (R) = rainfall;
- (D) = drainage;
- (L) = land use;
- (V) = vegetation.

Machine learning can be trained using historical landslide inventories where available.

*4.2 Why uncertainty is essential*

A prediction map without uncertainty can create false confidence.

Two areas may both have a predicted G4 classification, but one may have:
[ P(G4)=0.90 ]
while another has:
[ P(G4)=0.55. ]
The engineering response should not necessarily be identical.
The proposed system therefore generates:
[ P(Y|X) ]
rather than only:
[ . ]

*4.3 Rationale*

The key purpose of the proposed framework is to support investigation planning.

A conventional strategy might locate boreholes:

- on a regular grid;
- along a building footprint;
- based solely on geological mapping; or
- according to professional judgement.

The proposed system supplements these approaches with an uncertainty-aware spatial priority map.

*4.4 Mathematical formulation*

The Investigation Priority Index is proposed as:

[ IPI_i= w_HH_i+ w_UU_i+ w_VV_i+ w_CC_i+ w_II_i ]
where:

- (H_i) = hazard score;

The proposed framework introduces information gain as a decision criterion.

If (U_{before}) represents uncertainty before investigation and (U_{after}) represents uncertainty after investigation, then:
[ IG=U_{before}-U_{after}. ]
A candidate investigation location with high expected (IG) may be more valuable than a location with simply high predicted hazard.
This principle transforms investigation from a passive sampling process into an adaptive learning process.

## *5. Spatial Cross-Validation*

*5.1 Importance*

Explain ability is essential for engineering acceptance.

For a prediction at location (x), the system should provide not only:

Predicted condition = G4

but also:

Principal contributing variables = geology, slope, moisture, deformation trend, geomorphology.

SHAP or related methods can be used to estimate contributions:

[ f(x)=*0*+{j=1}^{p}_j ]
where (_j) represents the contribution of feature (j).
Explainability serves three functions:

1. engineering interpretation;
2. model auditing;
3. identification of physically implausible predictions.

## *6. Physics-Informed AI*

Purely data-driven models can generate statistically plausible but physically unreasonable predictions.

The proposed framework explicitly requires physical validation.

*6.1 Boreholes*

Boreholes provide:

- stratigraphy;
- depth to rock;
- groundwater observations;
- sample recovery;
- SPT data.

*6.2 CPT/CPTu*

CPT can provide continuous profiles of:

- cone resistance;
- sleeve friction;
- pore pressure.

*6.3 Geophysical investigation*

Potential techniques include:

- electrical resistivity tomography;
- seismic refraction;
- MASW;
- downhole methods;
- GPR where suitable.

*6.4 Laboratory testing*

Tests can include:

- grain-size distribution;
- Atterberg limits;
- density;
- moisture content;
- shear strength;
- consolidation;
- permeability;
- chemical characteristics.

The field observations then become new training and validation data.

### *7. Proposed Workflow for Practical Implementation*

*Phase 1: Desk study*

Collect:

- geological maps;
- topographic information;
- remote sensing;
- historical records;
- previous geotechnical reports.

*Phase 2: GIS database*

Develop a geospatial database containing all relevant information.

*Phase 3: Feature generation*

Generate terrain, spectral, hydrological and deformation variables.

*Phase 4: AI modelling*

Train and compare candidate algorithms.

*Phase 5: Spatial prediction*

Generate ground-condition and hazard maps.

*Phase 6: Uncertainty*

Generate confidence and uncertainty surfaces.

*Phase 7: Investigation optimisation*

Calculate IPI and expected information gain.

*Phase 8: Field investigation*

Conduct targeted boreholes and tests.

*Phase 9: Validation*

Compare predictions with observations.

*Phase 10: Model updating*

Retrain and recalibrate.

### *8. Limitations of the Framework*

*8.1 Surface–subsurface disconnect*

Remote sensing primarily observes the surface. Subsurface conditions may be completely different from what is inferred from surface characteristics.

*8.2 Training-data dependence*

AI models cannot compensate for fundamentally inadequate training data.

*8.3 Geological transferability*

A model trained in one geological environment may fail in another.

*8.4 Scale dependence*

A satellite pixel, borehole and foundation operate at different spatial scales.

*8.5 Temporal variability*

This paper has proposed a conceptual framework for AI-assisted preliminary geotechnical site investigation using remote sensing, GIS and spatial data analytics.

The framework responds to a fundamental challenge in geotechnical engineering: the need to infer spatially variable ground conditions from incomplete and spatially discrete observations.

Seven interconnected stages have been proposed:

1. **multisource geospatial data acquisition;**
2. **data quality control and harmonisation;**
3. **geotechnical feature engineering;**
4. **AI-assisted spatial prediction;**
5. **hazard and uncertainty assessment;**
6. **investigation-priority optimisation; and**
7. **field validation and iterative model updating.**

The central methodological contribution is the introduction of an **uncertainty-aware Investigation Priority Index** that integrates predicted hazard, uncertainty, spatial variability, project criticality and information gain.

***9. References***


1. Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748.
2. Zhao, C., Gong, W., Juang, C. H., Tang, H., & Wang, L. (2023). Optimization of site exploration program based on coupled characterization of stratigraphic and geo-properties uncertainties. *Engineering Geology*, *317*, 107081. https://doi.org/10.1016/j.enggeo.2023.107081.
3. Sheil, B., Anagnostopoulos, C., Buckley, R., Ciantia, M. O., Febrianto, E., Fu, J., et al. (2026). Artificial intelligence transformations in geotechnics: Progress, challenges and future enablers. *Computers and Geotechnics*, *189*, 107604.
4. Otake, Y. (2026). Perceived maturity of AI and machine learning in geotechnical engineering.