

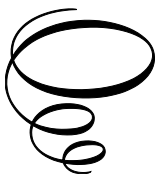
Geospatial Data Engineering and Applications

Geospatial Data Engineering and Applications

Edited by

Abhijit Bora, Kaustubh Bhattacharyya
and Manoranjan Kalita

**Cambridge
Scholars
Publishing**



Geospatial Data Engineering and Applications

Edited by Abhijit Bora, Kaustubh Bhattacharyya and Manoranjan Kalita

This book first published 2026

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2026 by Abhijit Bora, Kaustubh Bhattacharyya,
Manoranjan Kalita and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 978-1-0364-6454-7

ISBN (Ebook): 978-1-0364-6455-4

TABLE OF CONTENTS

Foreword	viii
Preface	ix
List of Illustration	x
List of Tables	xiii
Chapter 1	1
GeoAI Essentials and Applications <i>Abhijit Bora and Kaustubh Bhattacharyya</i>	
Chapter 2	11
Leveraging Machine Learning for Geospatial Data Integration <i>Jonti Deuri, Naveen Kumar and Otabek Puladjonov</i>	
Chapter 3	31
Machine Learning Integration with Geospatial Data <i>Imteyaz Hussain khan and Amar Singh</i>	
Chapter 4	62
Challenges Of Route Alignment in Difficult Terrain and Its Solution Using Geospatial Technologies <i>Victor Saikhom and Manoranjan Kalita</i>	
Chapter 5	76
GIS Data Acquisition and Management: Fundamentals and Applications <i>Abhishek Mandal and Abhijit Bora</i>	
Chapter 6	95
Artificial Neural Network and Machine Learning Approaches for Predicting Morbidity of Malaria <i>Pallavi Mohapatra, Rishi Jain and Nitin Kumar Tripathi</i>	

Chapter 7	120
Spatial Analysis and Visualization	
<i>Ch. Sravani, Vaibhav Bhagwan Pandit, Ch. Ravali, Aqil Tariq and Sumanta Chatterjee</i>	
Chapter 8	156
From Low to High: Transforming Satellite Imagery Through Super-Resolution Techniques	
<i>Wazir Muhammad, Ayaz Hussain and Sohrab Khan</i>	
Chapter 9	181
Exploring The Role of Geospatial AI in The Management and Analysis of IoT	
<i>Moushumi Barman</i>	
Chapter 10	205
Hyperspectral Image Processing: Review On Methods and Applications	
<i>Jonti Deuri, Sagar Saikia and Robiyabonu Akhunova</i>	
Chapter 11	230
GIS Spatial Intelligence for a Sustainable Future and SDG Achievement	
<i>Bamdeb dey, Tusar Kanti Das, Abhijit Bora and Jintu Mani Nath</i>	
Chapter 12	242
Applications Of GIS In Disaster Management	
<i>Vaibhav Bhagwan Pandit, Shende Amit Basveshwar, Dwaram Mrudula, Ch Ravali, Sumanta Chatterjee and Aqil Tariq</i>	
Chapter 13	270
Suitable Site Selection for Solid Waste Dumping Zone Management: A Case Study of Medinipur Municipality	
<i>Sanjib Mahata and Sanjit Patra</i>	
Chapter 14	289
Case Study Investigation of GIS-Based Landslide Susceptibility Analysis	
<i>Suman Hazari and Pori Das</i>	
Chapter 15	313
Mapping Life: GIS For Sustainable Ecosystems Development: A Focus on Human-Ecosystem Connection	
<i>Pawan Muddu and Lima Hazarika</i>	

Chapter 16	344
Late And Early Blight Diseases Detection of Potato Leaves using Resnet 50 And Inceptionv3 <i>Siddharth Sahu, Chiragdeep Kalita, Anjan Kumar Talukdar and Ranjit Sarma</i>	
Chapter 17	369
Heuristic Approaches of GIS For Sustainable Development <i>Moushumi Barman</i>	

FOREWORD

The amount of spatial information is expanding at a rate never seen before in the twenty-first century. The availability and richness of geospatial data, from satellite constellations to sensor networks built into commonplace products, have revolutionized how we see, navigate, and influence our environment. In order to create, manage, and use these enormous datasets in a meaningful way, advanced techniques are desperately needed.

The first edition of “Geospatial Data Engineering and Applications” comes at an extremely important time. Research, techniques, and practical experiences are carefully compiled in this volume to show the depth and scope of contemporary geospatial practices. It covers both basic ideas and new issues, such as incorporating cloud computing, big data analytics, and artificial intelligence into geospatial processes.

Together with a remarkable collection of authors, the Editor has created a useful resource that will be useful to students, academics, engineers, and policymakers. This book promotes better understanding and ongoing innovation in the ever-evolving field of geospatial science and technology by bridging the gap between academic rigor and practical applicability.

Recommending the book entitled “Geospatial Data Engineering and Applications” to readers around the world is a pleasure and an honor. May it act as a lighthouse for more research, cooperation, and learning.

Foreword by
Dr. Mrutyunjay Maharana.
Chief Engineer & Expert,
Hardware & High Voltage System Development,
BYD Auto Co Ltd, Shenzhen China
April 2025

PREFACE

The satellite remote sensing, location-based services and Internet of Things have all contributed to the ever-increasing amount of geospatial data, which has opened up previously unheard-of possibilities for revolutionizing how we view and comprehend our globe. The capacity to derive significant insights from intricate geographical datasets is more important than ever as urbanization picks up speed, climatic trends change, and social demands increase. At the nexus of data engineering, artificial intelligence, and geoinformatics, this book provides a relevant and thorough examination of how these fields interact to solve urgent real-world problems.

With a focus on tackling issues in urban planning, agriculture, forest management, disaster response, and environmental monitoring, this volume focuses on the use of big data analytics, machine learning, and artificial intelligence to satellite image processing and earth observation. It acts as a reference for professionals and students wishing to comprehend and use geospatial intelligence in a variety of fields by fusing theoretical viewpoints with real-world case studies.

This book “Geospatial Data Engineering and Applications” is meant to serve as a link between innovative research and real-world application. We would like to express our profound appreciation to all of the organizations and contributors whose assistance and wisdom have influenced this book. We hope it stimulates more research and creativity in the intriguing field of geospatial analytics.

With Best Wishes.

Dr. Abhijit Bora

Dr. Kaustubh Bhattacharyya

Prof. Manoranjan Kalita

LIST OF ILLUSTRATIONS

- Fig. 2-1. Characteristics of Geospatial Data
- Fig. 3-1. A Venn diagram showcasing the spatial, temporal, and multidimensional aspects of geospatial data
- Fig. 3-2. Machine Learning in Geospatial Analysis
- Fig. 3-3. Classify Land Cover from Satellite Imagery with Convolutional Neural Networks (CNNs) (North American Geoscientists.,2025)
- Fig. 3-4. Graphical illustration of this iterative Geospatial MLOps process
- Fig. 6-1. Geographical Distributions of Malaria-Attributed Mortality and Slide Plasmodium falciparum Rate
- Fig. 6- 2. Life Cycle of Malaria Parasite, Klein, E. Y. (2013)
- Fig. 6-3. Month-wise Comparison of the Accuracy of the J48 (Cross-validation & Percent Split model) to the Multi-Layer Perceptron Model for Accuracy for Sundargarh District
- Fig. 6-4. Month-wise Comparison of KAPPA of the J48 (Cross-validation & Percent Split model) to the Multi-Layer Perceptron Model for Accuracy for Sundargarh District
- Fig. 6-5. Month-Wise Comparison of RMSE of the J48 (Cross-Validation & Percent Split Model) to the Multi-Layer Perceptron Model for Accuracy for Sundargarh District
- Fig. 7-1. Choropleth Map showing showing state-wise distribution of heavy metals and total heavy metal concentration
- Fig. 7-2. Heatmap showing the exposure to climate change in the studied region across various districts.
- Fig. 7-3. Hexagonal map
- Fig.7-4. Dot mapping for spatial distribution visualization
- Fig. 7-5. Cluster map of wheat yield
- Fig. 7-6. Bubble map
- Fig. 7-7. Carograph map of world area in ha under organic agriculture
- Fig. 7-8. Hotspot map
- Fig. 7-9. Flow map
- Fig. 7-10. Contour map
- Fig. 7-11. 3D Surface Map
- Fig. 7-12. Animated time series map: NDVI value between 01 January and 31 December, 2024
- Fig. 8-1. Basic architecture of Satellite Image Super-resolution

- Fig. 8-2. Distribution of Remote Sensing application in different sectors.
- Fig. 8-3. Computational complexity cost in terms of number of parameters and number of layers versus state-of-the-art satellite image SR.
- Fig. 8-4. Name of different type of loss functions used in satellite image SR tasks.
- Fig. 8-5. Name of different optimizers to perform the different satellite image tasks.
- Fig. 9-1. Three-Layer Architecture for GIS integrates with IoT.
- Fig. 9-2. Integration of GIS with Cloud, Edge, and Fog Computing
- Fig. 10-1. Representation of a hyperspectral data cube, showing spatial dimensions (x, y) and spectral dimension (λ).
- Fig. 10-2. Challenges in High-Dimensional Data
- Fig. 10-3. PCA Process for Hyperspectral Data
- Fig. 10-4. Spectral Band Selection Process
- Fig. 10-5. Diagrammatic representation of Step by Step Sequence of Genetic Algorithm
- Fig. 11-1. Block diagram framework
- Fig. 12-1. GIS for Disaster Preparedness
- Fig. 12-2. Enhancing Post-Disaster Recovery Through GIS Technology
- Fig. 13-1 Location map of Medinipur municipality
- Fig. 13-2. Land use/land cover map of Medinipur municipality
- Fig. 13-3. Proposed model
- Fig. 13-4. Reclassified from Medinipur Municipality's main road
- Fig. 13-5. Reclassified slope map of Medinipur municipality area
- Fig. 13-6. Reclassified vegetation map of Medinipur municipality area.
- Fig. 13-7. Dumping zone with a 750-meter buffer zone
- Fig. 13-8. Reclassified map of Geomorphology of Medinipur municipality
- Fig. 13-9. Suitability map of waste dumping zone in Midnapur municipality area
- Fig. 14-1. Topographical features of the study area
- Fig. 14-2. Slope Profile of vulnerable Slope
- Fig. 14-3 Schematic representation of the vulnerable Slope
- Fig. 14-4 Particle Size Distribution of the Soil
- Fig. 14-5. Level of water level in the well
- Fig. 14-6. Geometry of the homogeneous model slope with meshes and boundary conditions
- Fig. 14-7 Sensitivity analysis results to decide the size of finite element mesh
- Fig. 14-8 Deformed mesh after the numerical analysis
- Fig. 14-9. Variation of FOS with surcharge loadings
- Fig. 14-10. Plot of horizontal displacements

- Fig. 14-11. Plot of vertical displacements
- Fig. 14-12 Depth of the phreatic line
- Fig. 14-13. Variation of FOS with depth of water table
- Fig. 14-14 Plot of total stresses
- Fig. 15-1. Timeline provides a comprehensive overview of key milestones in GIS development
- Fig. 15-2. A project on ArcGIS Pro Desktop Version, Image Source: Introduction to ArcGIS Pro
- Fig. 15-3. Application developed using the ArcGIS Online interface, illustrating potential evacuation routes and population risk associated with a hurricane.
- Fig. 15-4. Both Figures Fig.15-4 A & 15-4 B depict the QGIS Interface which consists of a layer panel, browser panel, toolbars, map canvas, Status Bar, Side toolbar, locator Bar, and a processing toolbox for Geoprocessing Analysis
- Fig. 16-1. Early and Late blight potato disease
- Fig. 16-2. ResNet50 Model Architecture
- Fig. 16-3. Inception V3 Model Architecture
- Fig. 16-4. Dataset collection of all images
- Fig. 16-5. Preprocessing of images
- Fig. 16-6. Variation of accuracy with the number of epochs
- Fig. 16-7. Confusion Matrix
- Fig. 16-8. Classification
- Fig. 16-9. Detection model with Confidence level
- Fig. 16-10. ResNet50 results (a) Training and validation accuracy graphs and (b) Training and validation loss graphs
- Fig. 16-11. Inception V3 results (a) Training and validation accuracy graphs and (b) Training and validation loss graphs
- Fig. 16-12. Inception V3 Confusion Matrix
- Fig. 16-13. ResNet50 Confusion Matrix
- Fig. 17-1. Components of GIS

LIST OF TABLES

- Table 2-1. Challenges in Integrating Geospatial Data with Machine Learning
- Table 2-2. Examples of Feature Engineering Techniques
- Table 5-1. Summary of key GIS data sources
- Table 6-1. Month-wise Performance metrics for RMSE, KAPPA, and Accuracy for the Cross-Validation Classifier
- Table 7-1. Spatial Data Types and Their Sources
- Table 7-2. overview of spatial analysis techniques and their visualization methods
- Table 7-3. Common Visualization Techniques for Spatial Data
- Table 8-1. Present additional details on CNN architectures used for satellite image super-resolution. Single scale identifies as “S” and multi-scale identify as “M”.
- Table 8-2. Names of different libraries used in satellite image super-resolution.
- Table 8-3. Overview of satellite images.
- Table 8-4. Overview information about satellite platforms.
- Table 9-1. Summary of Literature Review
- Table 9-2. Comparative analysis on the basis of Scalability, Strengths and Weakness.
- Table 9-3. A comparative analysis of GeoAI techniques highlighting their strengths, weaknesses, and optimal IoT applications.
- Table 10-1. Comparison of hyperspectral and multispectral imaging
- Table 10-2. Comparison of Classification Techniques
- Table 11-1. SDG Key Objective
- Table 13-1. Area of Land Use Land Cover
- Table 13-2. Criteria that were applied during development.
- Table 13-3. Geomorphology suitability scores are provided.
- Table 13-4. Pair-wise comparison ratio
- Table 13-5. Aggregating criterion weights and standardized criterion maps [Suitability Map]
- Table 13-6. Criteria wise different vector
- Table 13-7. Pair-wise matrix
- Table 13-8. Area of final suitable zone
- Table 14-1. Elevations of the Landslide Vulnerable Locations

Table 14-2. Soil properties obtained from laboratory investigation

Table 14-3. Input soil parameters for Mohr-Coulomb model

Table 14-4. Comparison of results between the proposed method and experimental study at $\phi = 37^\circ$, $q = 0$

Table 16-1. Summary of the literature review

Table 16-2. Summary of evaluation matrices

Table 16-3. Classification report of ResNet50

Table 16-4. Classification Report of Inception V3

Table 17-1. Summary of literature review

CHAPTER 1

GEOAI ESSENTIALS & APPLICATIONS

ABHIJIT BORA¹ AND
KAUSTUBH BHATTACHARYYA²

¹DEPARTMENT OF COMPUTER APPLICATIONS, SCHOOL OF
TECHNOLOGY, ASSAM DON BOSCO UNIVERSITY, INDIA

²DEPARTMENT OF ELECTRONICS, ASSAM SKILL
UNIVERSITY, INDIA

Abstract

This chapter highlights the fundamental ideas and wide range of uses of Geospatial Artificial Intelligence, a topic that combines artificial intelligence methods with geographic information systems to improve spatial data analysis and decision-making. With the growing availability of remote sensing imagery, GPS data, and geographic information, the combination of Artificial Intelligence and Geoinformatics allows one for the optimization of urban infrastructure, prediction of environmental changes, and the extraction of intricate patterns. To increase the precision and effectiveness of geographical studies, core technologies, including computer vision, machine learning, deep learning, and natural language processing, are used to automate and improve geospatial operations. Supported by case studies and research on urban land use and environmental forecasts, the chapter discusses current advancements in GeoAI, including its use in ecosystem monitoring, catastrophe risk assessment, and smart city management. In both academic and practical contexts, open-source tools like QGIS and specific plugins demonstrate how accessible platforms are being used to operationalize GeoAI. Real-time interpretation and forecasting capabilities are further enhanced by combining generative AI and big language models with IoT-enabled geographical data. GeoAI offers a scalable and intelligent approach to geographic decision support systems, and despite obstacles like data quality and technical complexity,

it has the potential to revolutionize a variety of industries, including public safety, agriculture, and climate change monitoring.

Keywords: Geospatial Artificial Intelligence, Remote Sensing, Data Engineering, Machine Learning, Raster Data, Vector Data

Introduction

Remote sensing (RS), the Global Positioning System (GPS), and geographic information systems (GIS) are crucial instruments for site-specific management or earth observations. GPS and GIS are technologies that are intrinsically linked. When combined, they provide a powerful tool for measuring, mapping, tracking, and modelling environmental data and resources for use in both commercial and scientific settings.

The ability of GIS to effectively process, compute, and analyse a variety of data and geographic information on computers is crucial for a variety of tasks involving spatial information or geographic location. In the 1970s, GIS technology was first developed. As technology has advanced, the system has progressively been established in sectors like urban construction, logistics transportation, public security patrols, and archaeological exploration (Zhuang et al., 2022). This article attempted to highlight the potent computing power of artificial intelligence (AI) to support GIS systems by first combining AI with conventional GIS.

The term "GeoAI" describes the fusion of geospatial data and analysis with artificial intelligence methods and tools. In order to derive significant insights from spatial data, it integrates the powers of AI with GIS.

By applying sophisticated algorithms and techniques to gain useful valuable insights from spatial data, GeoAI plays a vital role in the geographic information systems sector. Numerous advantages result from its integration with GIS systems, which enhance resource management and decision-making. Fundamentally, GeoAI improves the analysis of spatial data by using machine learning methods to identify complex relationships, patterns, and irregularities in geographic data sets. This analytical expertise fosters a thorough comprehension of spatial dynamics, empowering decision-makers across a range of industries, including urban planning, environmental monitoring, disaster relief, agriculture, and transportation.

Through automation, geospatial AI simplifies GIS data processing and analysis tasks. Productivity is increased by automating labour-intensive processes because it saves time and resources needed for manual analysis. This efficiency, in turn, increases productivity by enabling GIS professionals to concentrate on analysing data and formulating plans based on

the insights gathered from it. By employing historical geospatial data to more precisely identify trends and patterns, GeoAI improves predictive modelling and forecasting capabilities. By continuously analysing geographic data, it also makes real-time monitoring and response feasible. These systems are especially useful in many fields, such as smart cities, because they can swiftly identify environmental changes or anomalies and set off automated reactions or alerts. In order to maximize resource allocation and improve public safety, GeoAI can also keep a close eye on traffic patterns, air quality, and infrastructure health.

Related work

GeoAI has played a crucial role in urban land use research by utilizing sophisticated learning methodologies and incorporating a wide range of geospatial and Earth-related data.

Yao et al. (2022b) utilize a neural network known as TR-CNN to analyze urban land use patterns. This approach integrates remote sensing data with high-spatial and temporal resolution time-series electricity data, which reflect various socioeconomic characteristics. In a separate study, Yang et al. (2022) employ an ensemble classification technique to categorize urban functional regions. This method merges vector-based building data and points of interest (POIs) with an enhanced graph convolutional neural network (GCNN) and a Word2Vec model, resulting in superior classification accuracy compared to applications relying on single-source data and existing methods for multisource data integration. Yao et al. (2022a) stated that an enhanced predictive accuracy and robustness in geographic flood risk modeling with remote sensing data can help in evaluating flash flood situation. The stacking and blending ensemble learning techniques can be used to assess the flash flood situation.

Gevaert and Belgiu (2022) conduct land use classification and identify buildings across three African cities by utilizing a supervised classification model known as the Fully Convolutional Network (FCN). This model incorporates landscape metrics to evaluate the similarity between training and testing images. In a separate study, Pan et al. (2022) examine urban expansion through a combination of a convolutional neural network (U-Net) and a recurrent neural network (LSTM). This integration yields highly accurate land use classification by leveraging the multiscale neighborhood information from U-Net alongside the temporal data on historical urban expansion provided by LSTM.

GeoAI is increasingly being utilized in ecosystem management, facilitating more informed decision-making and promoting sustainable

development. Wang et al. (2021) assess the value of regional ecosystem services within an archipelago by employing satellite imagery, while also identifying the relationships between ecosystem services and socio-economic factors. Tan et al. (2022) create a communication-navigation-remote sensing-integrated ecological environment emergency monitoring chain (CN-RIEEEMC) specifically for tailings areas, establishing a framework for automated and intelligent monitoring of ecological emergencies. Ye et al. (2022) presents high-resolution ecological footprint mapping through GeoUNet, an AI-driven end-to-end multi-scale prediction method that utilizes multi-source data, enabling the generation of more precise downscaling maps for detailed spatio-temporal analysis.

Essential technologies that empower GeoAI

One essential pillar is machine learning, which analyzes and interprets geospatial data using algorithms. Predictive modeling, classification, and task grouping are made possible by its ability to teach GeoAI systems to recognize patterns and relationships in geospatial data. Numerous geospatial data types, including GPS data, geotagged social media data, and multi and hyperspectral satellite imagery, can be processed using these algorithms (Infosys BPM. 2025).

Complex and unstructured geospatial data, including multitemporal data cubes and cloud points from satellite and aerial imagery, are especially suitable for deep learning analysis. For particular tasks, these models are able to automatically extract patterns and features from large geospatial datasets (Wang et al., 2010). The interpretation and analysis of visual geospatial data, including satellite and aerial imagery, are made possible by computer vision. Applications like land use mapping, disaster response, and urban planning are made possible by sophisticated computer vision algorithms that can recognize and categorize features, land cover types, infrastructure, and changes over time from images.

Whereas, the techniques in natural language processing enable the examination of unstructured geospatial textual data, along with the extraction of geospatial details, sentiment insights, and object identification from text data. This capability allows for applications like geospatial trend assessment, disaster surveillance, and opinion-driven urban development. Extensive language models (LLM) and generative artificial intelligence (GenAI) can significantly simplify the process of interpreting maps, drawing insights, and suggesting actions. LLMs possess the capability to thoroughly analyze data represented on maps, discerning patterns, relationships, and trends within the information. Consequently,

GenAI can produce significant insights and conclusions that integrate pertinent geospatial mapping findings, thereby ensuring a clear comprehension of the implications and meanings associated with the mapped data. By combining LLMs with vector storage solutions and third-party APIs, GeoAI applications can effectively access, extract, and synthesize relevant content from extensive datasets. This integration facilitates the dynamic interpretation of geospatial data, allowing the models to generate insights derived from both real-time and historical data. For example, when a user inquiry about environmental changes in a specific region, LLMs can swiftly process the query, identify the intent, and retrieve relevant geospatial information from third-party APIs and vector storage to obtain the most recent environmental reports. The model can then summarize this information, delivering a comprehensive response to the user. Through the utilization of LLMs, GeoAI is poised to revolutionize our interaction with and understanding of geospatial information, rendering complex geographical analysis more accessible than ever before. Enterprises can leverage LLM and GenAI to derive significant insights from geospatial data by uncovering concealed information, recognizing spatial connections, and forecasting future scenarios. Such insights can guide a variety of decisions, including the optimization of transportation routes, urban planning, environmental risk mitigation, and disaster preparedness, among others.

Internet of Things (IoT) oriented devices that come with sensors, GPS units, and communication components produce vast quantities of geospatial data pertaining to environmental factors, infrastructure efficiency, and human actions. GeoAI merges these IoT data streams for various applications, facilitating data-informed decisions and timely interventions.

Applications of GeoAI

AI can be applied in various manners within the GIS industry. Fig. 1-1 shows the some of the applications of GeoAI that are used most commonly in different area.

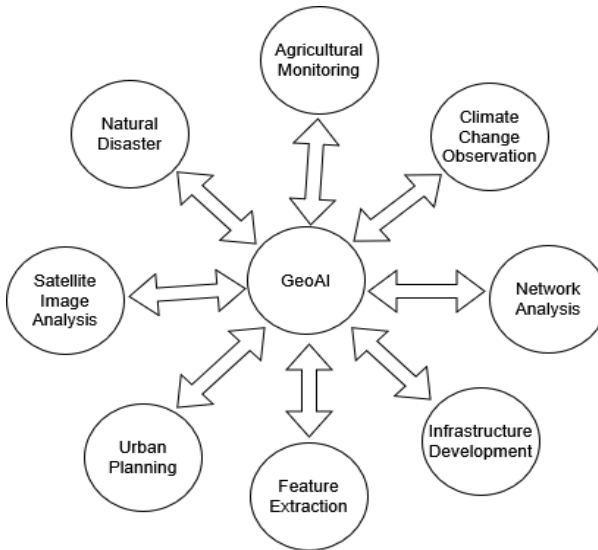


Fig. 1-1 Applications of GeoAI in different area

GeoAI can be used for extraction of features from the geospatial data sources. Different analysis can be carried out for the satellite imagery to identify patterns and changes in images that cannot be seen easily by normal human eye. Natural disasters can be predicted by observing historical data and pattern changes in over time. Urban planning and traffic management as well as network analysis can be carried out for efficient management of the urban resources. Irrigation and harvesting process can be observed and predicted for the better growth in the agriculture sector. GeoAI can also be applied for observation of climate change in different region of the globe. As early prediction of risk and climate change can be made.

Geospatial data service can also play a vital role for GeoAI based assessment (Spyrosoft. 2024). The Fig. 1-2 depicts the geospatial data services that can be used for creating data set which in turn can be used to train the model so that the model can respond based on its intelligence.

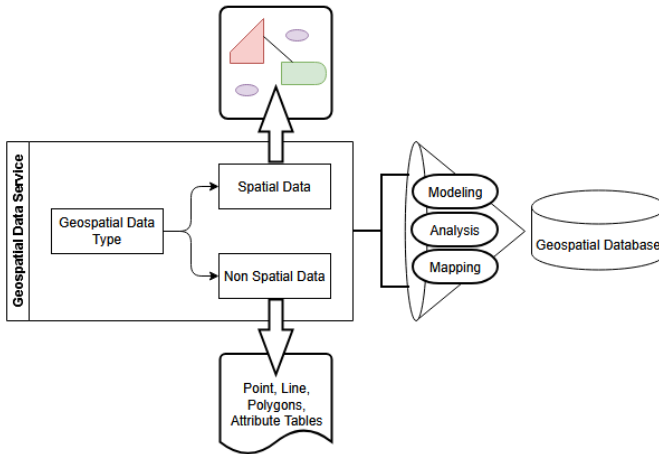


Fig. 1-2 Geospatial data services

While using the open-source tool, the most prominent tool used worldwide is QGIS. The manual process as well as automation process of training and testing activities for different machine learning model such as Support Vector Machine, Random Forest, Gradient Boosting etc can be executed to study different historical data of earth observation. Different pretrained model can be used to detect objects in the specific satellite image. QGIS contains plugin services for implementing deep learning algorithm.

The Deepness plugin makes deep learning accessible to non-expert users by making it simple to conduct segmentation, detection, and regression on raster orthophotos using custom ONNX Neural Network models. Highlights of the features: Any raster layer can be processed, including custom orthophotos from files or layers from internet sources like Google Satellite. Processing range is restricted to a predetermined region (visible portion or region denoted by vector layer polygons). Segmentation, regression, detection, and layer integration (for input data and model output layers) are among the frequently used model types that are supported. An output layer can be manually saved as a file after it has been created.

A free open-source QGIS plugin called Artificial Intelligence Forecasting Remote Sensing (AIRS) enables time series forecasting with deep learning models. AIRS's primary objective is to make data preparation, model training, and prediction less complicated so that one can use data to produce precise and perceptive time series forecasts.

Preprocessing data, creating and training deep learning models, forecasting, and visualizing outcomes are all made possible by the plugin. Tensorflow must be installed in order to use this plugin.

Stream network analysis, remote sensing, GIS, LiDAR data processing, DEM analysis, and spatial hydrology are all capabilities of WhiteboxTools (WBT), a sophisticated geospatial software program and data analysis platform.

The dzetsaka plugin uses SVM, Random Forest, and Gaussian Mixture Model classifiers for semi-automated classification.

The Semi-Automatic categorization Plugin (SCP) offers tools for downloading, preprocessing, and postprocessing photos and enables the supervised categorization of remote sensing images. Images from Sentinel-2 and Landsat can be searched and downloaded. There are numerous algorithms available for classifying land cover. Installing Remotior Sensus, GDAL, OGR, Numpy, SciPy, and Matplotlib is necessary for this plugin.

Conclusion

GeoAI possesses the capacity to automate and optimize the processing of large and intricate geographical datasets. It is one of its main advantages. In addition to speeding up data processing, this automation lowers the possibility of human error, producing results that are more trustworthy. GeoAI improves the capabilities of conventional GIS by utilizing cutting-edge AI techniques like machine learning, deep learning, computer vision, and natural language processing, allowing for more precise, effective, and perceptive geographical analytics. Such systems can offer current insights through the integration of real-time data from IoT devices and remote sensing technologies. They are essential for applications like disaster management and smart city projects. Public safety and resilience can be improved by these systems' ability to track environmental conditions, anticipate possible threats, and facilitate quick response plans. Although GeoAI has many benefits, there are drawbacks to its deployment. Its adoption may be hampered by problems including data quality, technological complexity, and the requirement for specialist knowledge. It can be concluded that GeoAI is a major development in geospatial analysis with the potential to revolutionize a number of industries.

References

1. Gevaert, C.M., and M. Belgiu. 2022. "Assessing the Generalization Capability of Deep Learning Networks for Aerial Image Classification

- Using Landscape Metrics.” *International Journal of Applied Earth Observation and Geoinformation* 114: 103054. <https://doi.org/10.1016/j.jag.2022.103054>.
2. Infosys BPM. 2025. “Geospatial Data Types.” *Infosys BPM*. Accessed April 25, 2025. <https://www.infosysbpm.com/blogs/geospatial-data-services/geospatial-data-types.html>.
 3. Pan, X., Z. Liu, C. He, and Q. Huang. 2022. “Modeling Urban Expansion by Integrating a Convolutional Neural Network and a Recurrent Neural Network.” *International Journal of Applied Earth Observation and Geoinformation* 112: 102977. <https://doi.org/10.1016/j.jag.2022.102977>.
 4. Spyrosoft. 2024. “GIS and Artificial Intelligence: What Is GeoAI?” Available at <https://spyro-soft.com/blog/geospatial/gis-and-artificial-intelligence-what-is-geoai>, Accessed April 20, 2025.
 5. Tan, X., J. Jiao, Y. Zhong, A. Ma, Y. Xu, Z. Sha, F. Huang, Y. Wan, and W. Hu. 2022. “The CNRIEEEMC: A Communication-Navigation-Remote Sensing-Integrated Ecological Environment Emergency Monitoring Chain for Tailings Areas.” *International Journal of Applied Earth Observation and Geoinformation* 108: 102710. <https://doi.org/10.1016/j.jag.2021.102710>.
 6. Wang, L., C. Chen, F. Xie, Z. Hu, Z. Zhang, H. Chen, X. He, and Y. Chu. 2021. “Estimation of the Value of Regional Ecosystem Services of an Archipelago Using Satellite Remote Sensing Technology: A Case Study of Zhoushan Archipelago, China.” *International Journal of Applied Earth Observation and Geoinformation* 105: 102616. <https://doi.org/10.1016/j.jag.2021.102616>.
 7. Wang, J.-F., X.-H. Li, G. Christakos, Y.-L. Liao, T. Zhang, X. Gu, and X.-Y. Zheng. 2010. “Geographical Detectors-Based Health Risk Assessment and Its Application in the Neural Tube Defects Study of the Heshun Region, China.” *International Journal of Geographical Information Science* 24 (1): 107–127. <https://doi.org/10.1080/13658810802443457>.
 8. Yao, Y., X. Yan, P. Luo, Y. Liang, S. Ren, Y. Hu, J. Han, and Q. Guan. 2022b. “Classifying Land-Use Patterns by Integrating Time-Series Electricity Data and High-Spatial Resolution Remote Sensing Imagery.” *International Journal of Applied Earth Observation and Geoinformation* 106: 102664. <https://doi.org/10.1016/j.jag.2021.102664>.
 9. Yao, J., X. Zhang, W. Luo, C. Liu, and L. Ren. 2022a. “Applications of Stacking/Blending Ensemble Learning Approaches for Evaluating Flash Flood Susceptibility.” *International Journal of Applied Earth Observation and Geoinformation* 112: 102932.

- <https://doi.org/10.1016/j.jag.2022.102932>.
10. Ye, R., Z. Huang, L. Li, and X. Shan. 2022. “GeoUNet: A Novel AI Model for High-Resolution Mapping of Ecological Footprint.” *International Journal of Applied Earth Observation and Geoinformation* 112: 102803. <https://doi.org/10.1016/j.jag.2022.102803>.
 11. Zhuang, F., and W. Du. 2022. “Intelligent Decision Support System for Distribution Network Planning Based on Artificial Intelligence and GIS Technology.” In *2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 1934–1938. Chongqing, China. <https://doi.org/10.1109/IMCEC55388.2022.10019346>.

CHAPTER 2

LEVERAGING MACHINE LEARNING FOR GEOSPATIAL DATA INTEGRATION

JONTI DEURI¹, NAVEEN KUMAR²
AND OTABEK PULADJONOV³

¹FACULTY OF ENGINEERING AND TECHNOLOGY,
SHARDA UNIVERSITY, UZBEKISTAN

²AMITY INSTITUTE OF INFORMATION TECHNOLOGY,
AMITY UNIVERSITY PATNA, INDIA

³FACULTY OF ENGINEERING AND TECHNOLOGY,
SHARDA UNIVERSITY, UZBEKISTAN

Abstract

The convergence of machine learning with geospatial data has profoundly reshaped geospatial data engineering, introducing advanced tools for analyzing, modeling, and predicting spatial phenomena. The sources of Geospatial data are satellite imagery, remote sensing, and GIS, which are inherently complex due to their high dimensionality, spatial heterogeneity, and temporal variability. ML effectively solves these challenges by enabling the precision and efficiency of processing and interpreting large, intricate datasets. This chapter examines key methodologies and frameworks for integrating ML with geospatial data, focusing on techniques such as spatial data preprocessing, feature engineering, and the use of supervised, unsupervised, and deep learning models. Real-world applications are explored, including LULC classification, disaster prediction, environmental monitoring, urban planning, and precision agriculture, showcasing the practical utility of this integration. Despite its immense potential, combining ML with geospatial data poses challenges such as managing spatial-temporal variability, ensuring computational scalability, and improving model interpretability. The following issues are

addressed by discussing strategies and frameworks to overcome them, highlighting interdisciplinary collaboration in advancing geospatial data engineering. In *Geospatial Data Engineering and Applications*, this chapter underscores the transformative impact of ML in extracting actionable insights from geospatial datasets and discusses future perspectives for research and innovation in this dynamic field.

Keyword: Machine Learning, GIS, Satellite Image, Geospatial data, Vector, Raster

1 Introduction

Using Machine Learning (ML) for geospatial data has become a game changer in geospatial data engineering, fundamentally altering how we analyse, model, and predict spatial phenomena. Geospatial data, which includes information from satellite imagery, remote sensing, and Geographic Information Systems (GIS), is complex due to its high dimensionality, spatial variability, and changes over time (Zhang & Zhu, 2018). By applying ML techniques to geospatial data, we can create advanced tools that process and interpret these complex datasets with remarkable accuracy and efficiency. This chapter examines the methods and frameworks that support the integration of ML with geospatial data, highlighting both the theoretical foundations and real-world applications of this exciting intersection.

Geospatial data encompasses information linked to specific locations on the Earth's surface. This type of data collection is sourced from satellite imagery, aerial photography, GPS, and remote sensing technologies (Goodchild, 2007). The rise of these technologies has resulted in huge amount of geospatial data available, presenting both opportunities and challenges for data scientists and engineers.

In the past, analyzing geospatial data relied primarily on traditional statistical methods and manual interpretation. However, these techniques often struggled to manage the complexity and scale of contemporary geospatial datasets (Li & Shao, 2014). The emergence of ML has introduced a robust set of tools for automating and improving the analysis of geospatial data. Machine Learning algorithms, especially those used in supervised, unsupervised, and deep learning have demonstrated significant efficacy in activities such as classification, regression, and clustering, and anomaly detection within geospatial contexts (LeCun et al., 2015).

The combination of machine learning and geospatial data represents more than just a technological advancement; it signifies a fundamental

change in how we comprehend and engage with spatial information. By utilizing ML, researchers and practitioners can reveal patterns and insights that were once obscured within extensive and intricate datasets, resulting in improved decision-making and innovative applications in multiple fields (Foody & Mathur, 2004).

1.2 Importance of ML in Geospatial Data

The importance of ML in the field of geospatial data is significant. The distinct features of geospatial data—like spatial autocorrelation, heterogeneity, and scale—pose both challenges and opportunities for analysis. ML techniques are particularly effective in tackling these challenges, as they can manage large datasets, learn from intricate patterns, and deliver predictions with remarkable accuracy (Zhang & Zhu, 2018).

Handling High Dimensionality: Geospatial data frequently consists of multiple layers of information, such as spectral bands in satellite images or various attributes in GIS datasets. ML algorithms, especially those rooted in deep learning, excel at processing and analyzing these high-dimensional datasets, allowing for the extraction of meaningful features and dimensionality reduction when needed (LeCun et al., 2015).

Spatial Heterogeneity: The spatial variability inherent in geospatial data means that patterns and relationships can differ greatly between various areas. ML models, particularly those that consider the spatial context (e.g., convolutional neural networks), can adjust to these differences, leading to predictions that are more precise and tailored to specific locations (Foody & Mathur, 2004).

Temporal Variability: Numerous geospatial datasets are dynamic in nature; they change gradually as a result of natural phenomena and human activities. Models in machine learning, especially those specifically designed for analyzing time-series data, such as recurrent neural networks, excel at identifying these temporal relationships and patterns. This capability allows for predicting future conditions based on past data (Li & Shao, 2014).

Automation and Scalability: Conventional geospatial analysis techniques often demand considerable manual effort and struggle to scale effectively with large datasets. Conversely, machine learning algorithms are capable of automating various aspects of data handling and analysis, enabling efficient examination of extensive geospatial data (Goodchild, 2007).

Enhanced Decision-Making: By delivering more precise and timely insights, The incorporation of machine learning into geospatial analysis enhances decision-making in vital sectors like disaster response, urban development, and ecological monitoring. For example, machine learning algorithms can predict the likelihood of natural disasters, enhance land utilization, or monitor deforestation almost in real-time (Zhang & Zhu, 2018).

1.3 Objectives

The primary aim of this chapter is to provide a thorough overview of how ML integrates with geospatial data, focusing on the methods, challenges, and uses that stem from this integration. Specifically, the chapter aims to:

Explore Methodologies: Investigate the essential methodologies and frameworks for merging ML with geospatial data, which encompasses the preprocessing of spatial data, feature engineering, and the use of various ML models (supervised, unsupervised, and deep learning) (LeCun et al., 2015).

Highlight Applications: Look into real-world uses of ML in geospatial data analysis, such as LULC classification, disaster prediction, environmental monitoring, urban planning, and precision agriculture. These examples will demonstrate the practical benefits and transformative potential of ML in geospatial contexts (Foody & Mathur, 2004).

Address Challenges: Explore the difficulties associated with combining machine learning and geospatial information., such as managing spatial-temporal variability, ensuring computational scalability, and enhancing model interpretability. The chapter will also discuss strategies and frameworks to tackle these challenges (Zhang & Zhu, 2018).

Foster Interdisciplinary Collaboration: Stress the significance of interdisciplinary collaboration in advancing geospatial data engineering. Combining machine learning with geospatial data requires knowledge from multiple disciplines, such as computer science, geography, environmental science, and urban planning. This chapter will underscore the role of collaboration in fostering innovation and addressing complex spatial issues (Goodchild, 2007).

Future Perspectives: Provide perspectives on upcoming research and innovation directions in the realm of ML and geospatial data. As the volume and complexity of geospatial data continue to expand, ongoing research is essential to develop more advanced ML models, enhance

computational efficiency, and improve the interpretability of results (Li & Shao, 2014).

2 Geospatial Data: Characteristics and Challenges

Geospatial data, commonly known as spatial data, refers to information that pinpoints the geographic location and attributes of both natural and man-made features and boundaries on our planet. This kind of data is crucial for a range of applications, such as city planning, environmental observation, emergency response, and navigation. However, the distinct characteristics of geospatial data also pose considerable challenges, particularly regarding its integration with machine learning methods.

2.1 Characteristics of Geospatial Data

Geospatial data is complex due to its multidimensional nature and the various forms it can take. Fig.2-1 shows the characteristics of geospatial data. Below are several essential traits that characterize geospatial data:

Spatial Dependence: Geospatial data exhibits spatial autocorrelation, indicating that nearby data points are likely to be more alike than those that are situated farther away. This characteristic is essential for spatial analysis and modelling.

Scale and Resolution: The scale of geospatial data indicates the level of detail or granularity, while resolution refers to the smallest unit of measurement. High-resolution data provides greater detail, but it also leads to an increase in both the size and complexity of the data.

Temporal Dynamics: Geospatial data often includes a temporal component, capturing changes over time. This factor is crucial for uses such as climate modeling, land-use change detection, and urban growth analysis.

Heterogeneity: Geospatial data can exist in various formats, including vector data (points, lines, polygons), raster data (grids of pixels or cells), and more complex forms like 3D models or LiDAR data. This variety requires flexible processing and analysis techniques.

Volume and Velocity: With the rise of remote sensing technologies, IoT devices, and GPS, the amount of geospatial data being generated is vast. Additionally, the speed at which this data is produced demands efficient storage and processing solutions.

Semantic Richness: Geospatial data often contains rich semantic information, such as land-use categories, population density, or environmental attributes. This depth introduces an additional layer of intricacy to data evaluation.

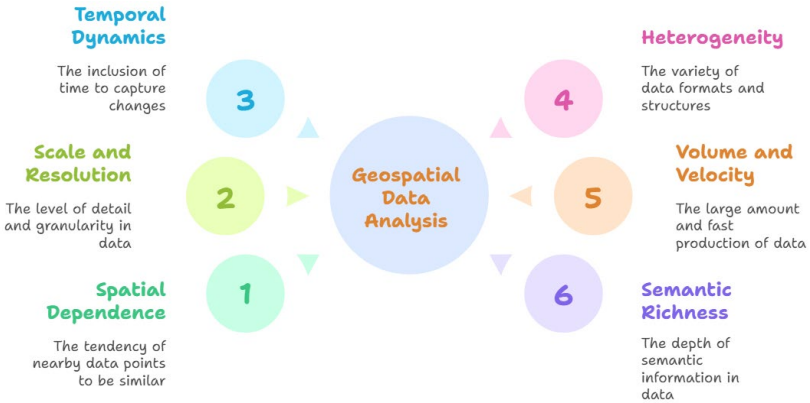


Fig. 2-1 Characteristics of Geospatial Data

2.2 Challenges

Combining geospatial data with machine learning presents numerous challenges, many of which arise from the distinct nature of the data. Table 2-1 shows the challenges in integrating Geospatial Data with Machine Learning. Below are several of the main challenges:

Data Preprocessing and Cleaning: Geospatial data typically requires significant preprocessing, which includes removing noise, filling in missing data, and normalizing values. The different types of data formats contribute an additional level of difficulty to this task.

Scalability: The immense amount of geospatial data can overwhelm conventional machine learning algorithms. It's crucial to have efficient methods for data storage, indexing, and retrieval to manage extensive datasets effectively.

Spatial Autocorrelation: Although spatial dependence is a key characteristic of geospatial data, it can also present difficulties. A number of conventional machine learning algorithms are built on the premise that the data points are independent and identically distributed (i.i.d.), which is not