Harnessing AI for Earth Observation Applications

August 27, 2024

SPEAKER: DR. JAEDEO DHARPURE
Post Doctoral Scholar | Ohio State University

SUBMITTED BY: SHUCHIKA SHARMA
Roll No: 24901325

GEOSPATIAL DATA

Information about objects, phenomena, and environments on Earth, collected using technologies like satellites is known as Geospatial data. It is used in monitoring environmental changes, urban planning, and disaster management.

REMOTE SENSING:

We gather information/data without touching the data through sensors.
Two types -

- **Passive Remote Sensing**: Relies on natural sources of energy, mainly sunlight, to capture data. Example: Optical sensors on satellites.
- **Active Remote Sensing**: Uses artificial sources of energy, like radar, to collect data. Example: Synthetic Aperture Radar (SAR).

TECHNIQUES FOR ACQUISITION OF GEOSPATIAL DATA

Geospatial data acquisition involves sensors on satellites and other devices to remotely collect information about the Earth. Techniques include -

- Imaging (such as optical and radar images).
- Laser Scanning (LiDAR) and Radar for mapping landscapes.

Sources of Geospatial data -

- Maps
- Stallite imagery
- GPS data

Softwares -

- ArcGIS Pro
- QGIS
- ERDAS Imagine
- ENVI

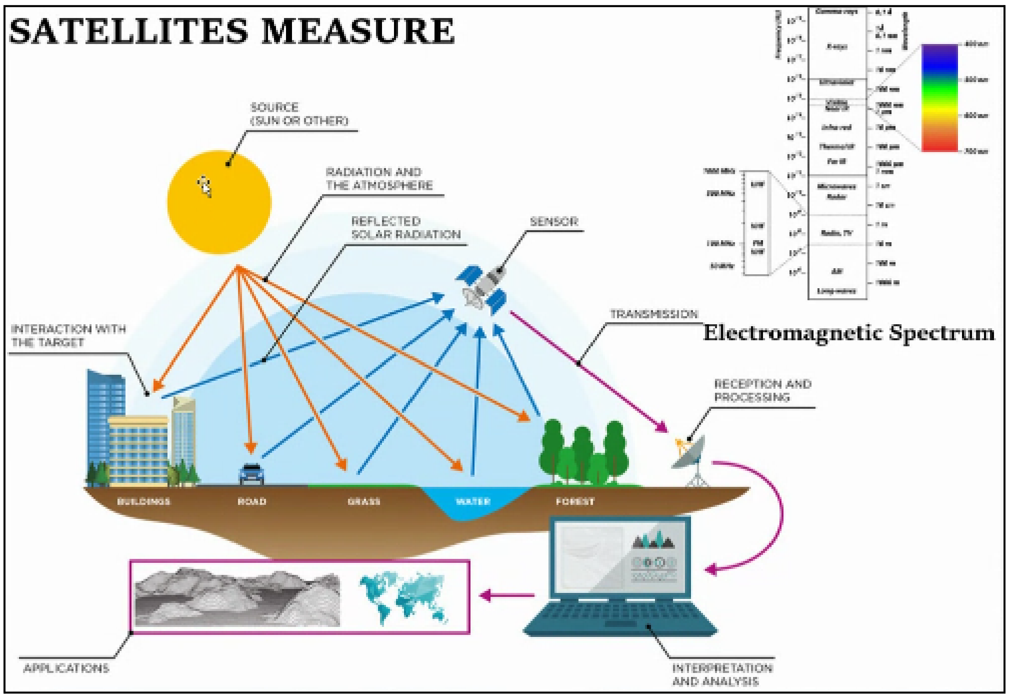
SATELLITES MEASURES

Satellites gather measurements such as:

1. Spectral Recognition - Identifies objects based on the specific light spectrum they emit or reflect. It describes a sensor's ability to define fine wavelength intervals. The finer the spectral resolution, the narrower the wavelength range for a particular channel or band.
2. Spatial Recognition - Detects shapes and physical features by analyzing space in images. Spatial recognition is defined by the size of the pixel.

3. Radiometric resolution - Imagery data are represented by positive digital numbers that vary from 0 to a power of 2. The maximum number of brightness levels available depends on the number of bits (represents radiometric resolution) used in representing the energy recorded.
4. Temporal resolution - Represents how frequently a satellite can provide observation of the same data on the same earth.

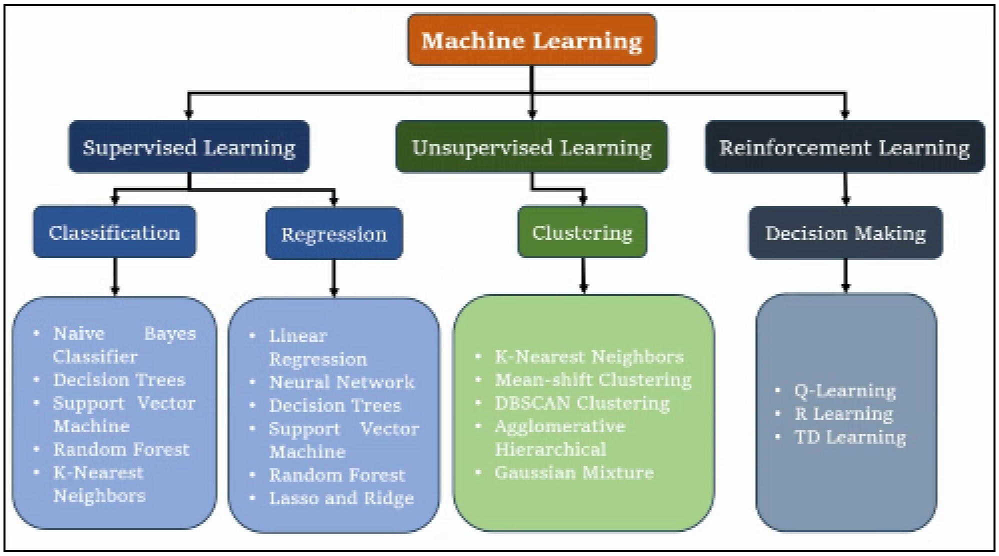
A large portion of solar energy is absorbed by the Earth's surface, affecting temperature and energy distribution.



MACHINE LEARNING APPROACHES

Various ML approaches were discussed, which include -

1. Supervised Learning Methods
2. Unsupervised Learning
3. Reinforcement Learning



CASE STUDY 1 - Snow Cover Variability Detection

Remote sensing images can help provide us data of regions that are not accessible like snow etc. So, we can easily find out where snow cover with be increasing and where decreasing

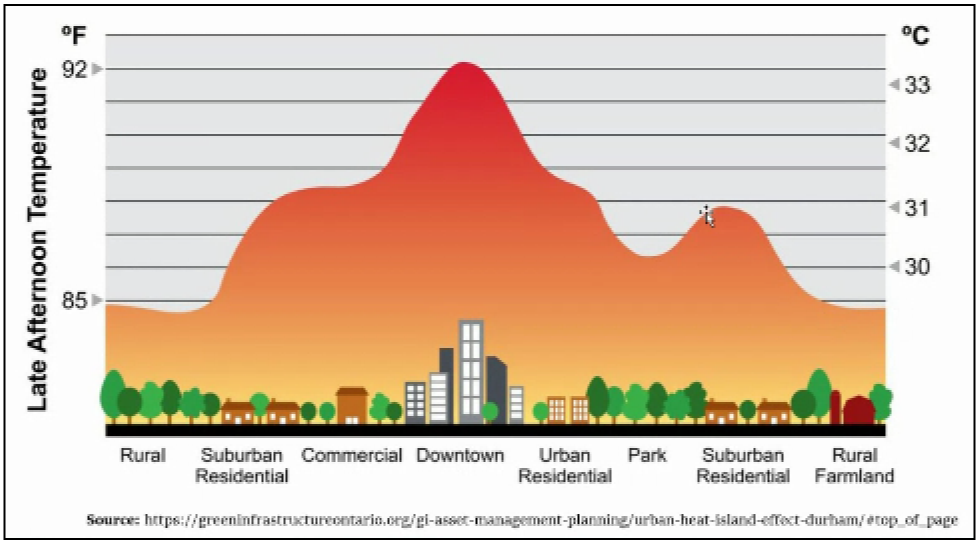
APPLICATION:
Monitoring snow accumulation and melt patterns using satellite data. This is important for -

- water resource management,
- climate studies, and
- forecasting flood risks.

CASE STUDY 2 - Urban Heat Island Analysis

Temperature of urban areas is more than that of rural areas. This is due to factors like less trees, more vehicles, factories etc. We can predict urban heat island using remote sensing data.

UHI Analysis can be done using AI to analyze satellite data and predict areas prone to extreme heat.

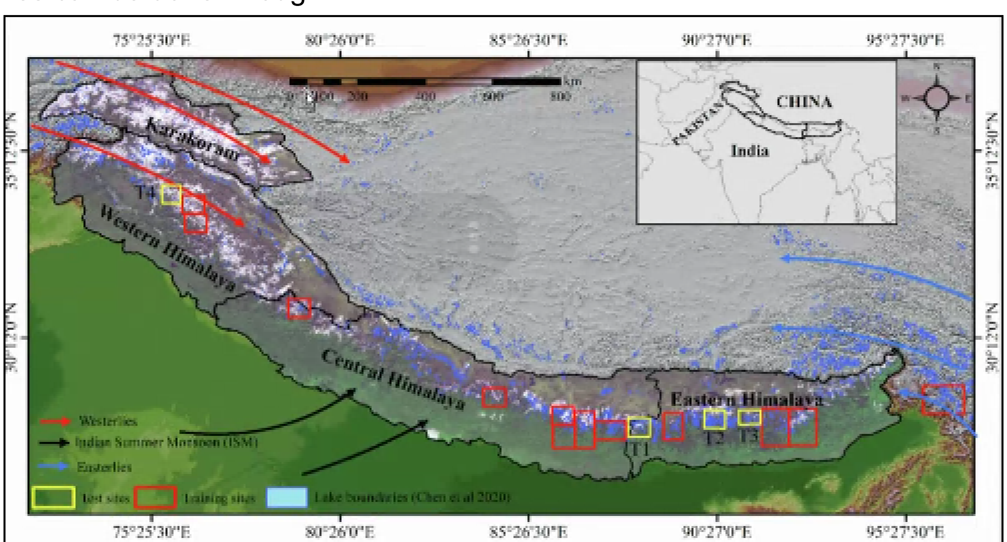


Indexes used by Researchers:

- Normalized Difference Vegetation Index (NDVI): Measures vegetation cover and its health.
- Land Surface Temperature (LST): Measures the temperature of the Earth's surface.
- Normalized Difference Built-up Index (NDBI): Assesses the extent of urbanization by measuring built-up areas.

CASE STUDY 3 - Automated Mapping of Glacial Lakes

In inaccessible areas like the Karakoram range of Himalayas, extraction of the number of Glacial lakes can be done through AI.



(Kaushik et al. 2022)

Method:

- Use of 10 spectral bands from satellite images.

- Manual Labeling: Experts manually label known glacial lakes.
- Machine Learning (ML) Model: Trains on the labeled data using the 10 spectral bands to predict and detect the formation of new glacial lakes.

CASE STUDY 4 - Bridging Gap in Time Series Data

Time series data often has missing values or periods where data wasn't collected. These gaps can hinder the accuracy of predictions or analyses.

Challenges:

- Data Gaps: Missing data points in time series datasets, especially in environmental or geospatial data, can result from satellite malfunctions, cloud cover (in optical sensors), or other technical issues.
- Impact on Analysis: Missing data affects the continuity of environmental monitoring, which can distort the results of models predicting future trends (e.g., climate change, temperature variations).

Methodology:

- Use of AI Models: To address the issue of missing data, advanced AI models like LSTM (Long Short-Term Memory) and Transformers were employed.
- LSTM (Long Short-Term Memory): This is a specialized neural network architecture designed for time-series data. It excels at capturing temporal dependencies and making predictions based on sequences of data.
- Transformers: These are powerful machine learning models typically used for natural language processing but have been adapted for time-series data. They can handle long-range dependencies and provide better context for predicting missing data points.

By applying these AI models, the researchers were able to interpolate or predict the missing data points in the time series, creating a more complete dataset. This approach is crucial in Earth observation, where uninterrupted data is essential for monitoring climate changes, environmental patterns, or urban development.

CASE STUDY 5 - Flood Prediction using Deep Learning Models

Objective: Predict river water discharge in the next 3-6 hours at the end of the river.

Techniques: Use of advanced machine learning models such as:

- LSTM (Long Short-Term Memory): A type of recurrent neural network (RNN) well-suited for time-series predictions like river flow.
- Transformers: Another ML model that can handle sequential data and make predictions based on historical patterns of river discharge.

CHALLENGES

1. DATA QUALITY AND AVAILABILITY
    - Data Scarcity - Limited data in remote regions

- Heterogeneity - Diverse data formats and resolutions

2. Computational Complexity
    - Large-Scale Processing - Handling massive datasets
    - Real-time analysis - Processing data in real-time scenario
    - Model Training - High computational demands for training AI models
3. Interdisciplinary integration
    - Domain knowledge - Need for integrating specialized knowledge
4. Model Generalization
    - Transferability - Difficulty applying models across regions
    - Overfitting - Risk of models performing well only on training data
5. Ethical and Societal Concerns
    - Bias and fairness
    - Privacy
    - Transparency
6. Technical Challenges
    - Multi-Source Data Integration - Aligning and synchronizing varied data
    - Scalability - Maintaining performance with increasing data volumes

FUTURE DIRECTIONS

1. Advanced AI Techniques
    a. Deep Learning Integration - Enhanced accuracy with CNNs and RNNs
    b. Explainable AI - Making models more transparent and understandable
2. Big Data and Real Time Analytics for handling geospatial datasets
3. Interdisciplinary collaboration
    a. Cross domain integration to combine knowledge from multiple fields
    b. Human-in-the-loop to add human expertise
4. Ethics and Privacy
    a. Mitigating bias
    b. Protecting sensitive geospatial data
5. Automation and anonymous systems
    a. Automated mapping
    b. Advanced navigation and environmental mapping using autonomous vehicles/drones
6. Scalability
7. Using crowdsourced data for broader impact