HANDLING SPARSE INVENTORY DATA IN DEEP-LEARNING BASED LANDSLIDE SUSCEPTIBILITY MAPPING

Presented by

Golam Murad

RESEARCH CONTEXT

Global Challenge:

Landslides cause more than **4,000** fatalities annually (*Magee et al., 2018*)

Regional Criticality:

The Chittagong Hill Tracts are more prone to landslides due to their rugged terrain and underlying geological conditions.

RAJOSTHOLI BORDER ROAD

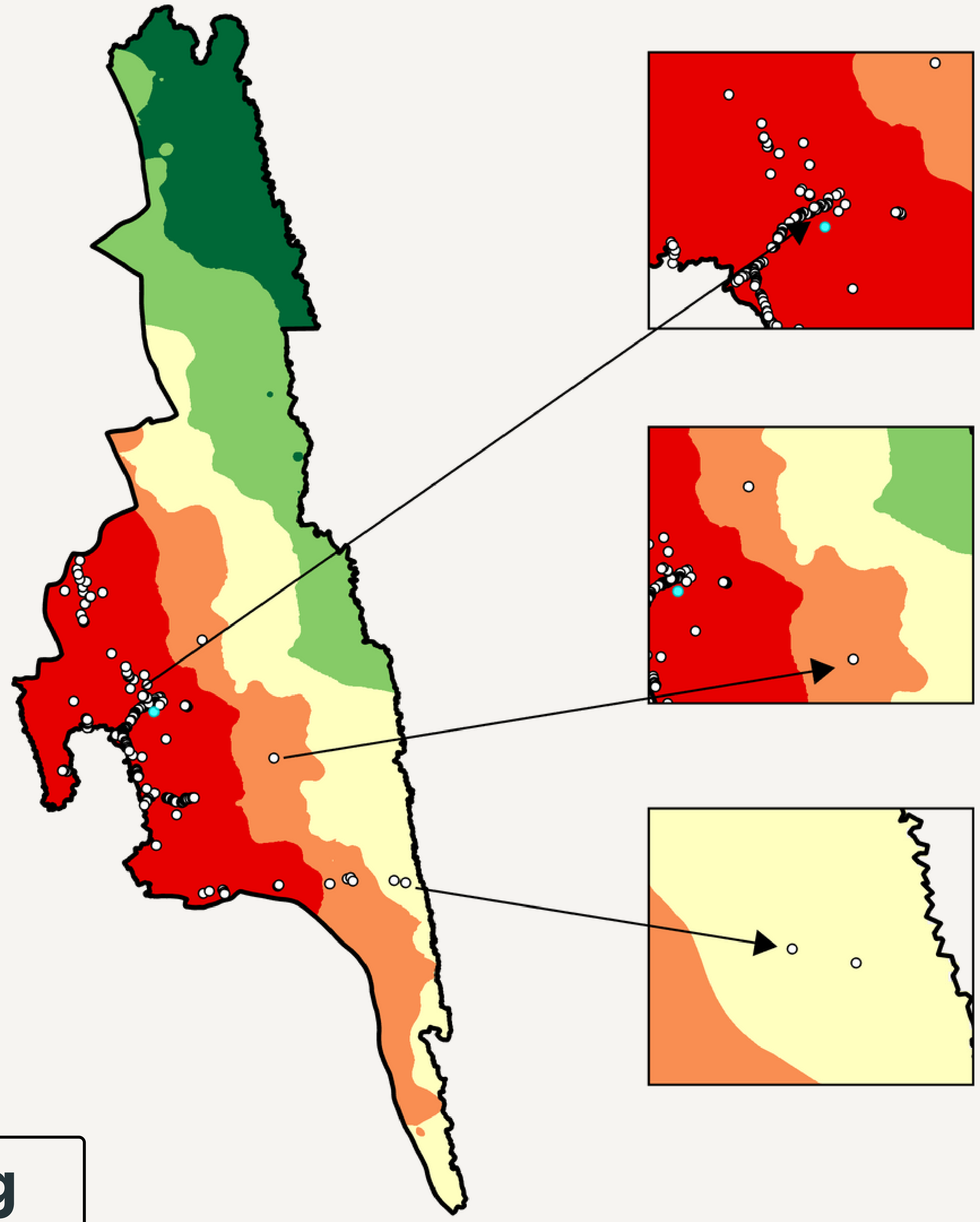
PROBLEM STATEMENT

Core Scientific Gap:

- Spatial sparsity in landslide inventories **(n=198 points)** induces sampling bias, violating stationarity assumptions in geospatial modeling *(Yi et al., 2020)*

Consequence:

- Reduced generalizability of susceptibility models in data-scarce regions



This map was created without addressing Spatial sparsity in the inventory dataset

RESEARCH OBJECTIVES

- To mitigate **Spatial Sparsity**, by density-based clustering (DBSCAN)
- To evaluate **DL architectures** (DNN, 1D-CNN, LSTM) on both data sets
- To generate a Comprehensive **Landslide susceptibility map**

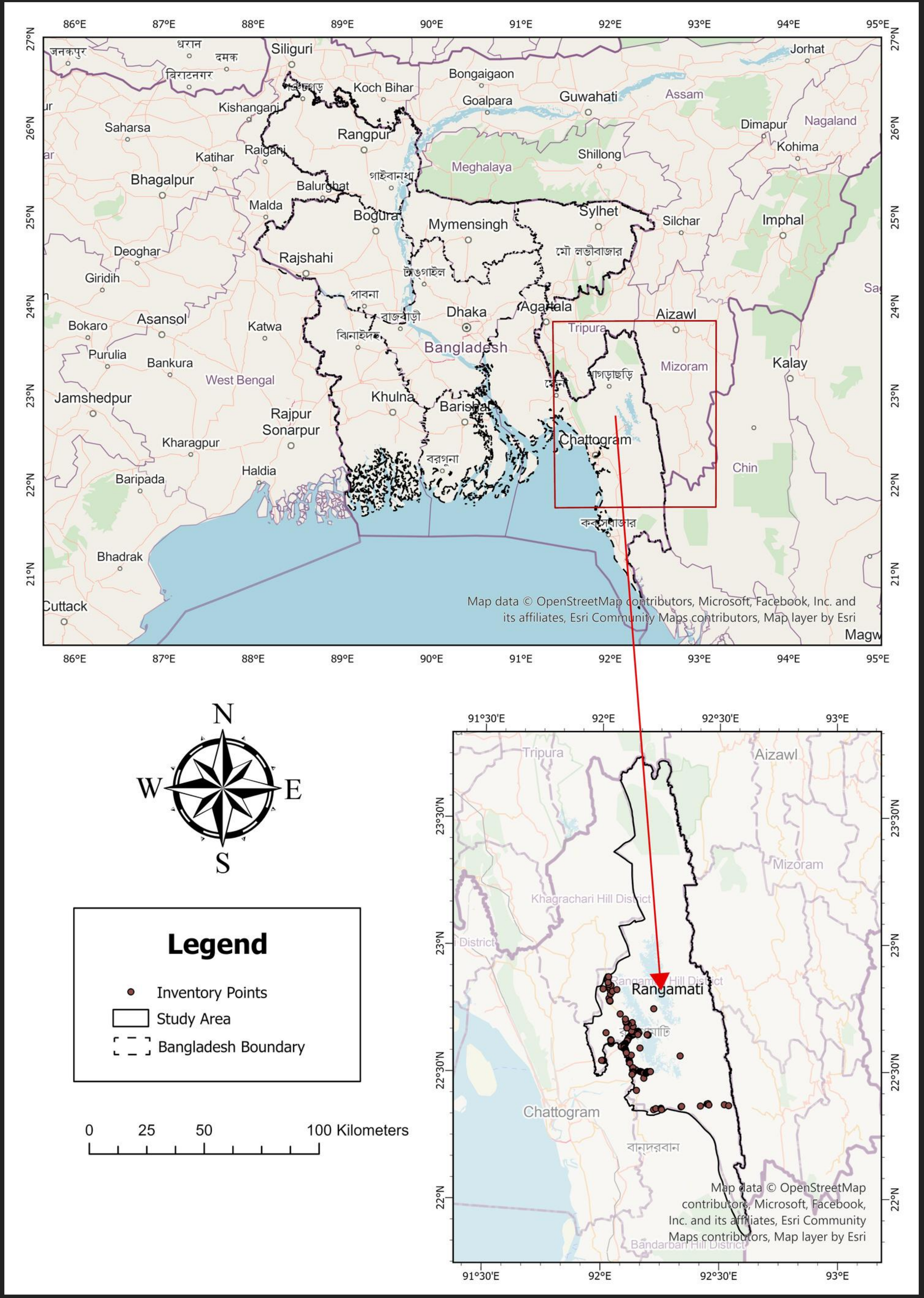
STUDY AREA: RANGAMATI

Geological Setting:

- Steep Terrain
- Bhuban-Bokabil Formation with high weathering potential *(Humayain & Biplob, 2021)*

Climatic Hazard:

- 80% of the annual rainfall is concentrated in the June-September monsoon



1

STUDY AREA

Rangamati

2

DATA COLLECTION

- SRTM DEM
- ESRI Sentinel-2 imagery
- Geofabrik data
- University of East Anglia

3

DATA PREPROCESSING

- Rasterization
- Extract Multipoint Data
- Lebel the dataset

4

SPARSITY CHECK & MITIGATION

- CONFIRMED SPARSIY
- APPLY THE MITIGATION TECHNIQUE

5

TensorFlow Keras

DEEP LEARNING ARCHETECTURE

- DNN
- 1D-CNN
- LSTM

6

ACCURACY

MODEL EVALUATION AND PREPARE LANDSLIDE SUSCEPTIBILITY MAP

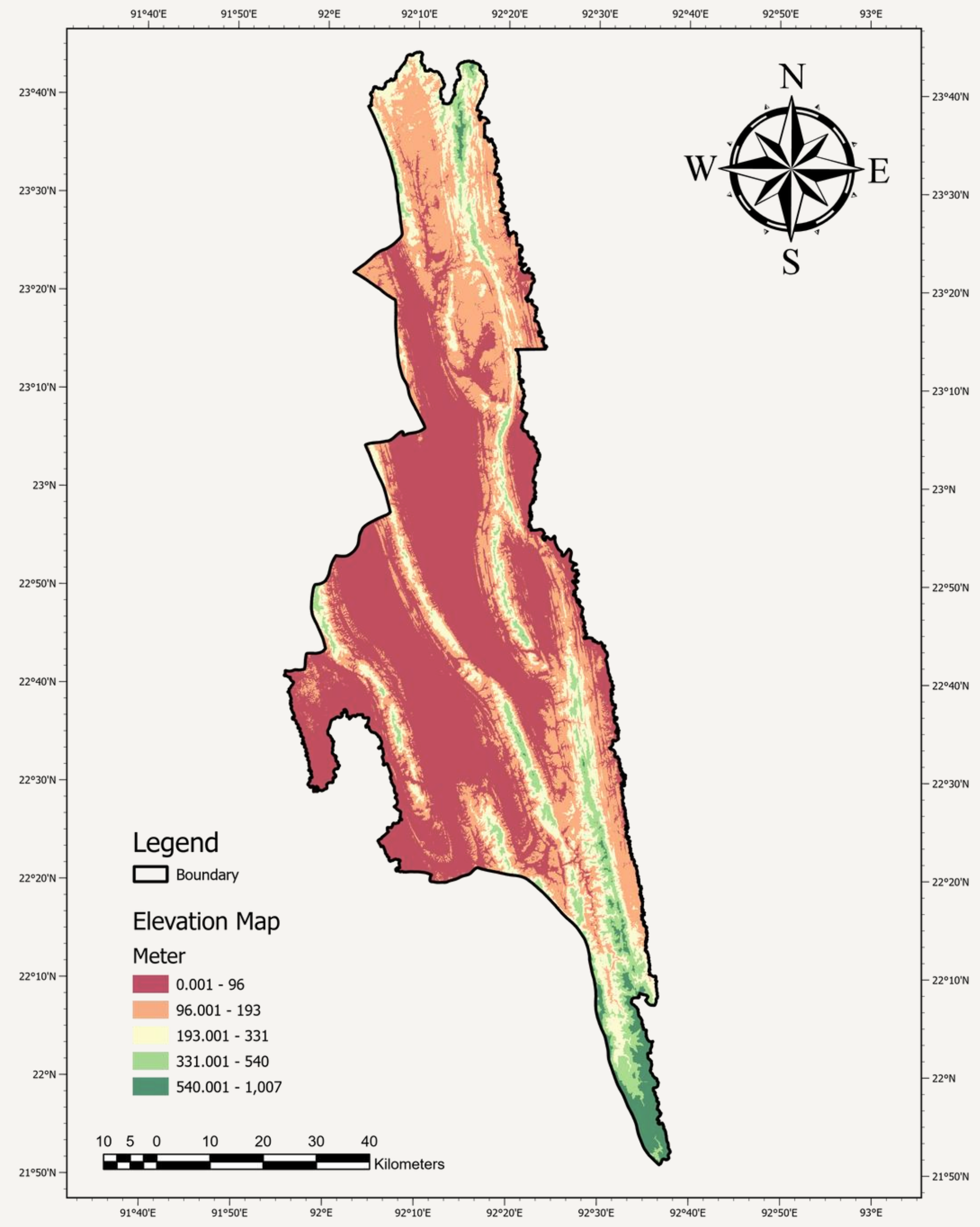
REMOTE SENSING DATA SOURCES

INVINTORY DATA

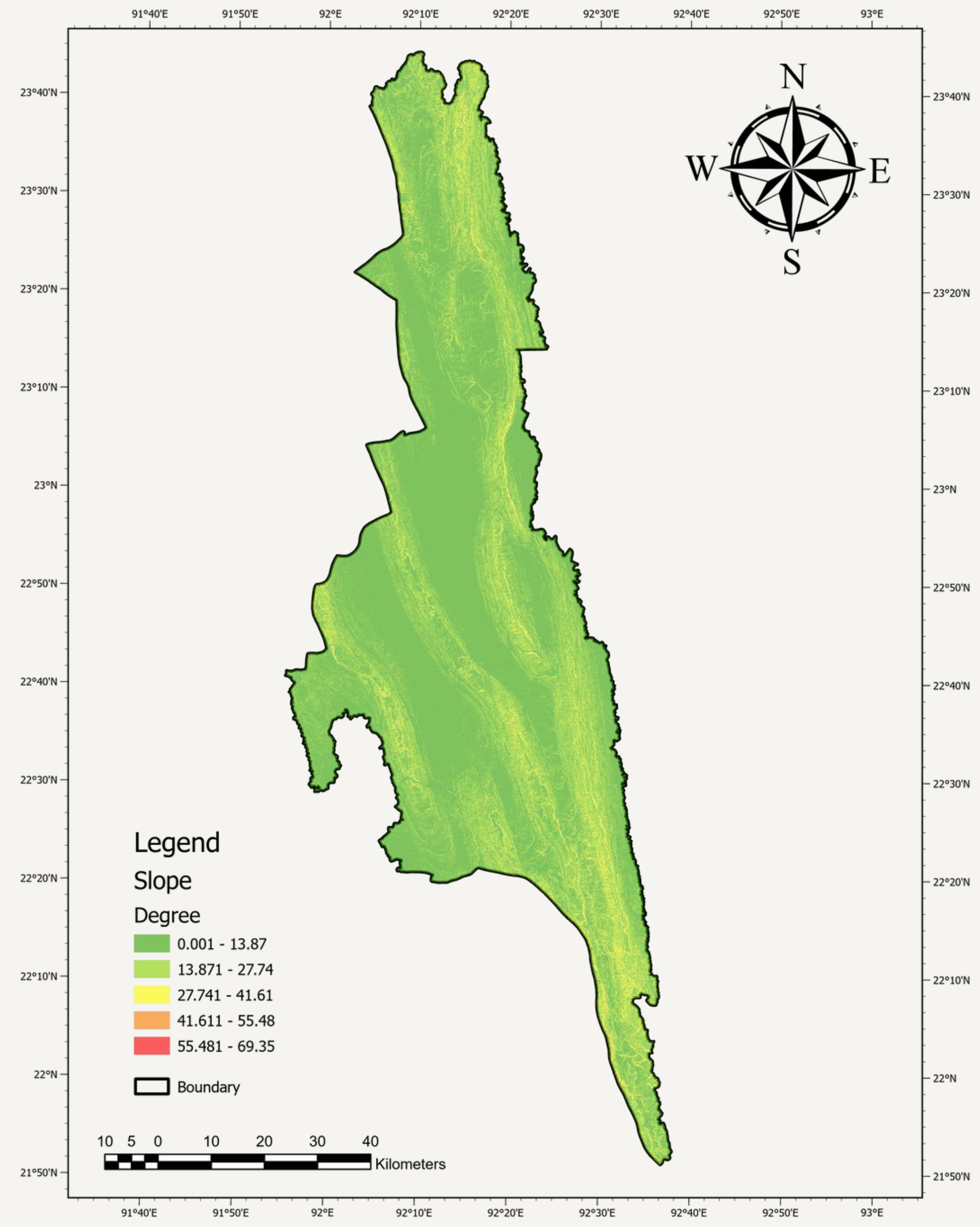
- Global Landslide Catalog BY **NASA**
- The Center for Environmental and Geographic Information Services (**CEGIS**)
- 34 Engineer Construction Bridge, **Bangladesh Army**

LANDSLIDE CONDITONING FACTORS

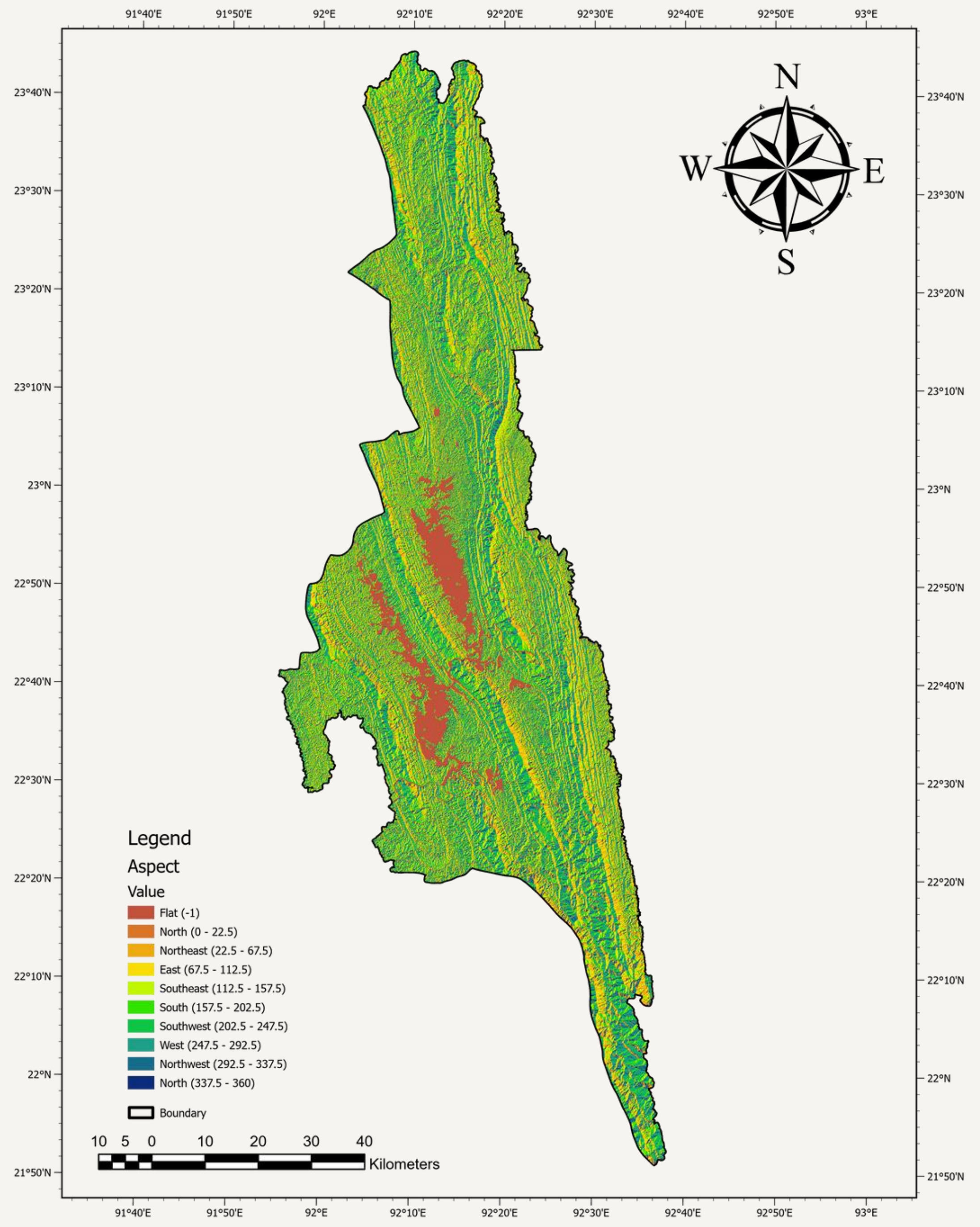
Factor	Source of Data	Spatial Regulation (m)
Elevation	https://earthexplorer.usgs.gov/ SRTM DEM	30m x 30m
Slope	https://earthexplorer.usgs.gov/ SRTM DEM	30m x 30m
Aspect	https://earthexplorer.usgs.gov/ SRTM DEM	30m x 30m
Curvature	https://earthexplorer.usgs.gov/ SRTM DEM	30m x 30m
SPI	https://earthexplorer.usgs.gov/ SRTM DEM	30m x 30m
TWI	https://earthexplorer.usgs.gov/ SRTM DEM	30m x 30m
TRI	https://earthexplorer.usgs.gov/ SRTM DEM	30m x 30m
NDVI	https://earthexplorer.usgs.gov/ Landsat 8 level 2 imagery	10m x 10m
LULC	https://livingatlas.arcgis.com/landcover/ Esri | Sentinel-2 Land Cover	10m x 10m
Rainfall	https://crudata.uea.ac.uk/cru/data/hrg/	
Distance to Drainage	https://earthexplorer.usgs.gov/ SRTM DEM	30m x 30m
Soil Texture	https://soilgrids.org/ ISRIC (International Soil Reference and Information Centre)	250-meter resolution



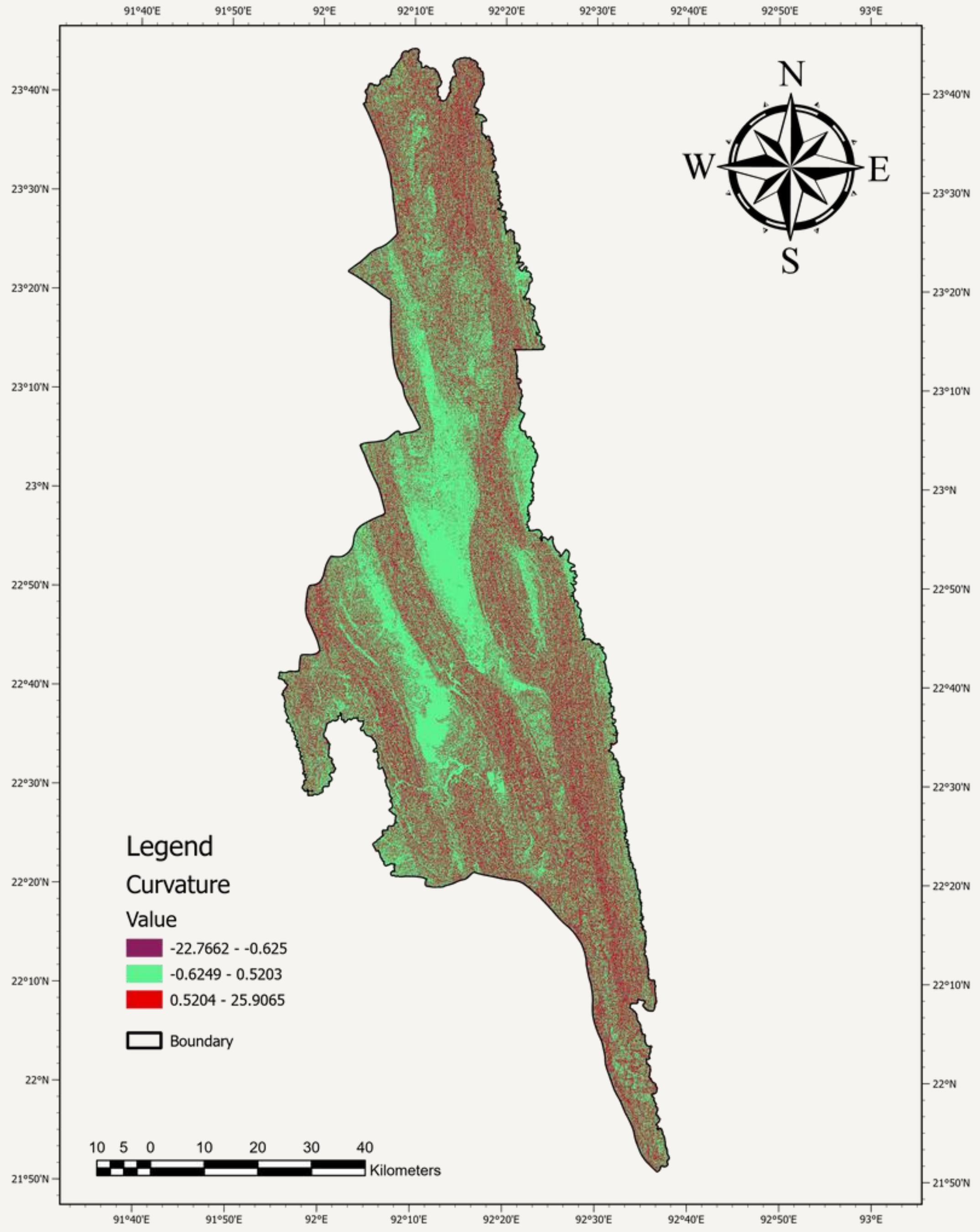
Elevation Map



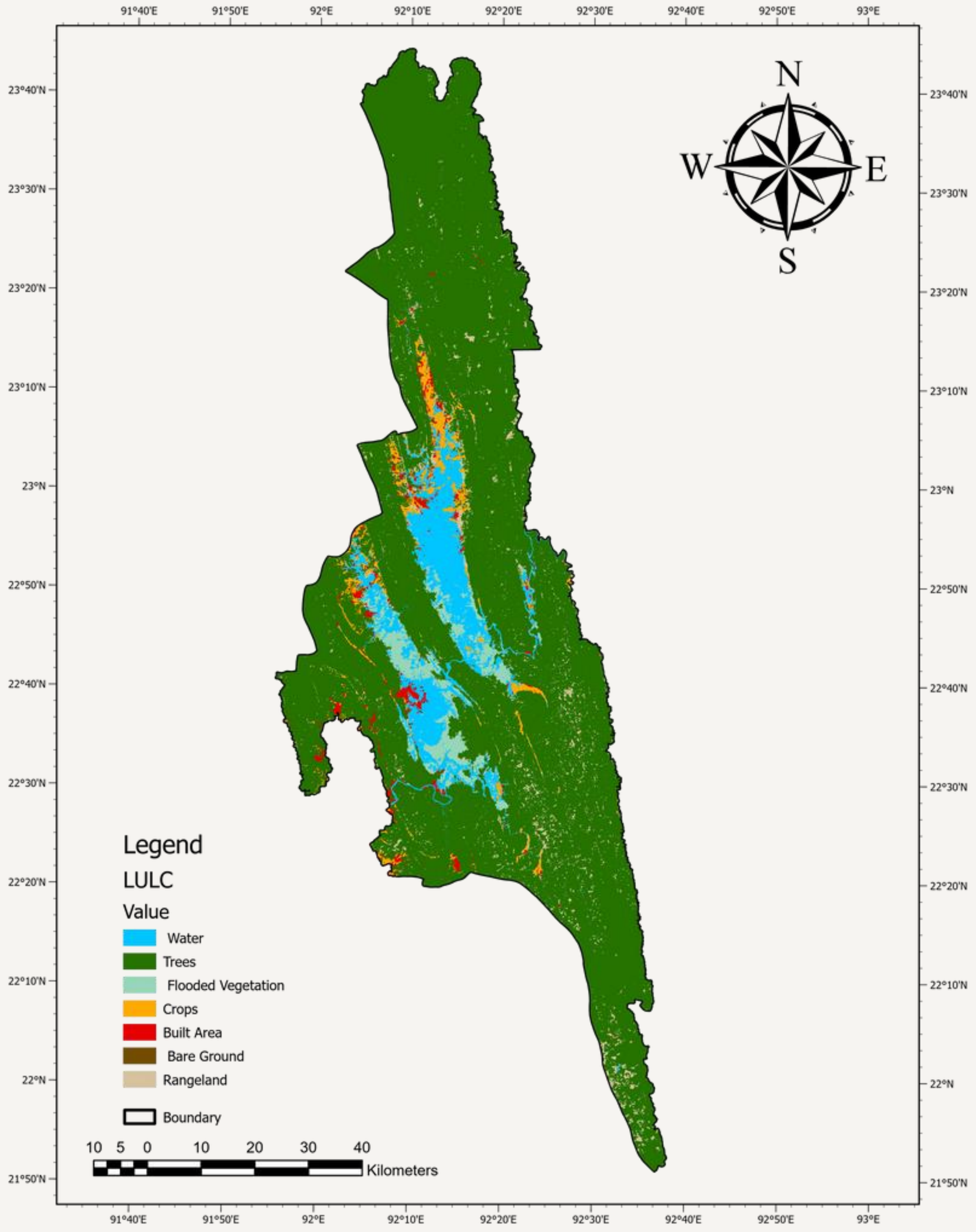
Slope Map



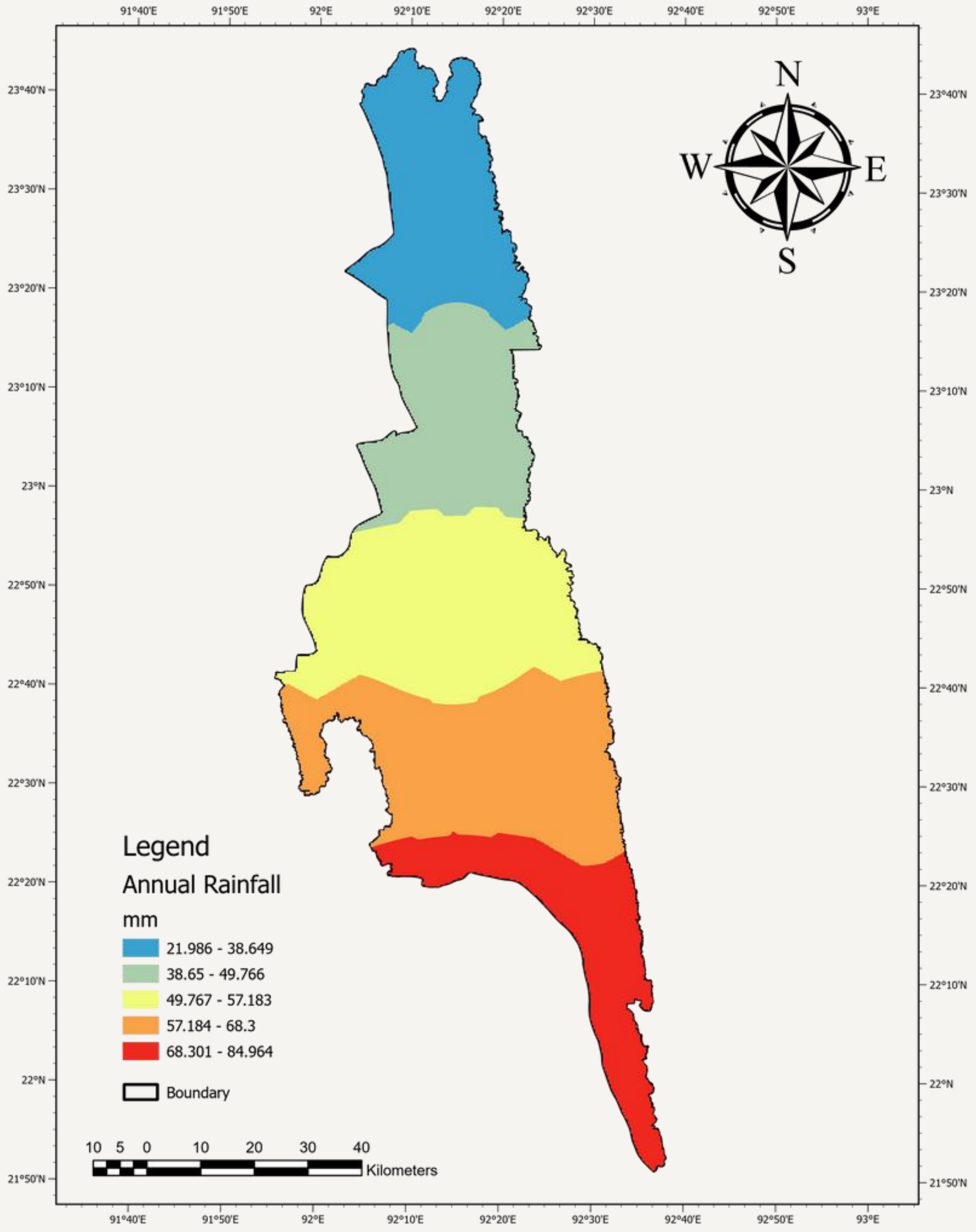
Aspect Map



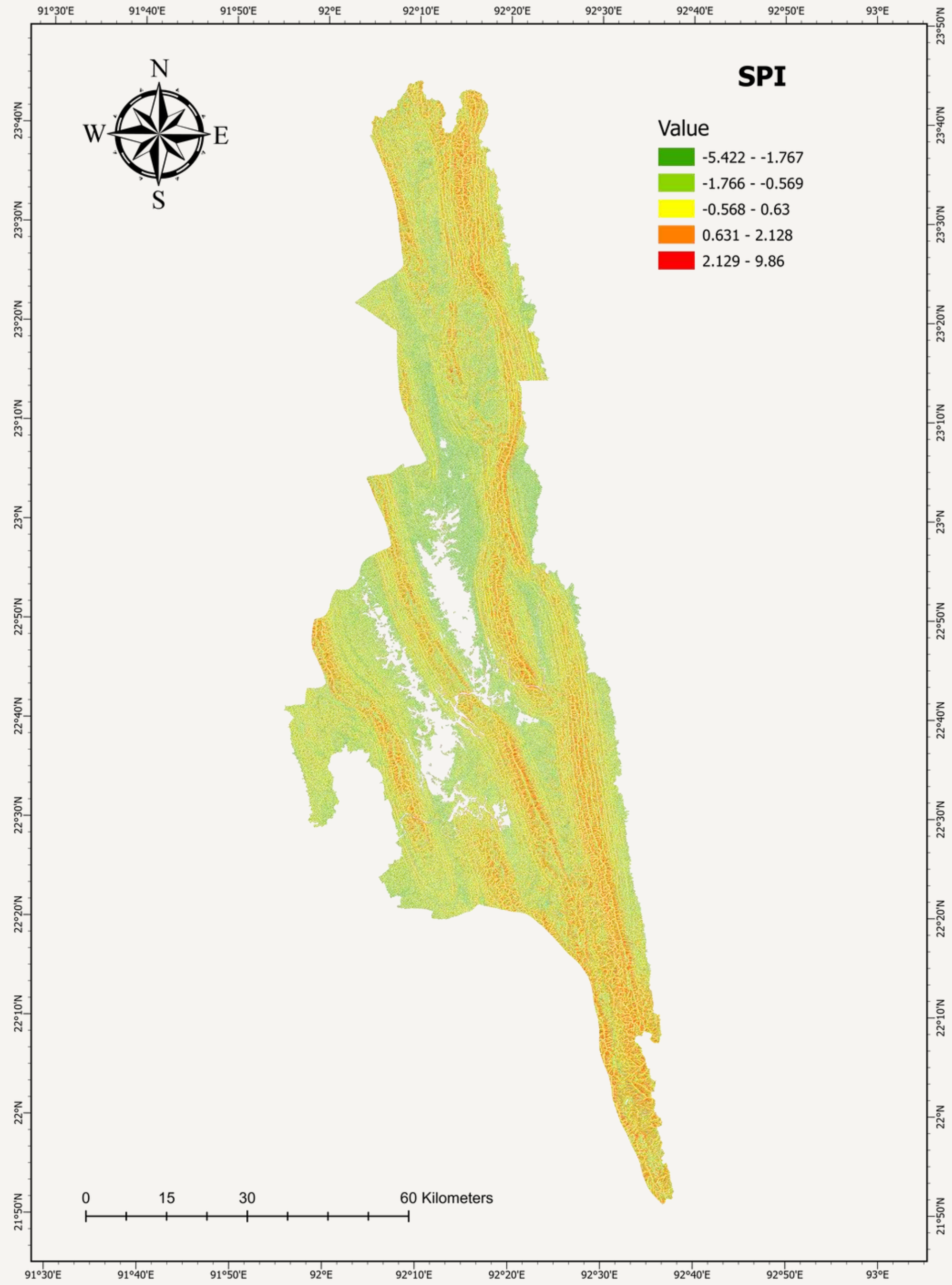
CURVATURE MAP



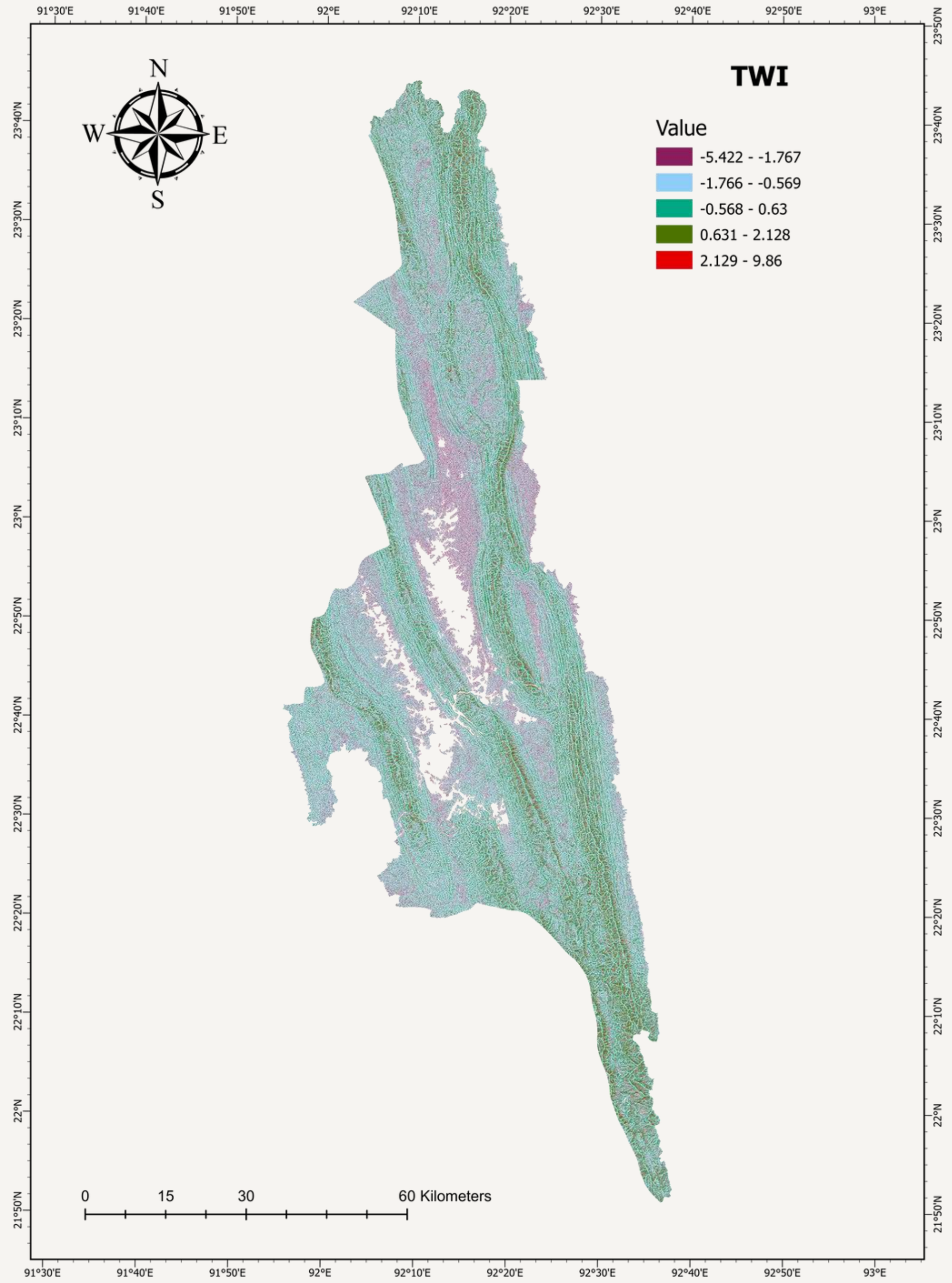
LUSC MAP



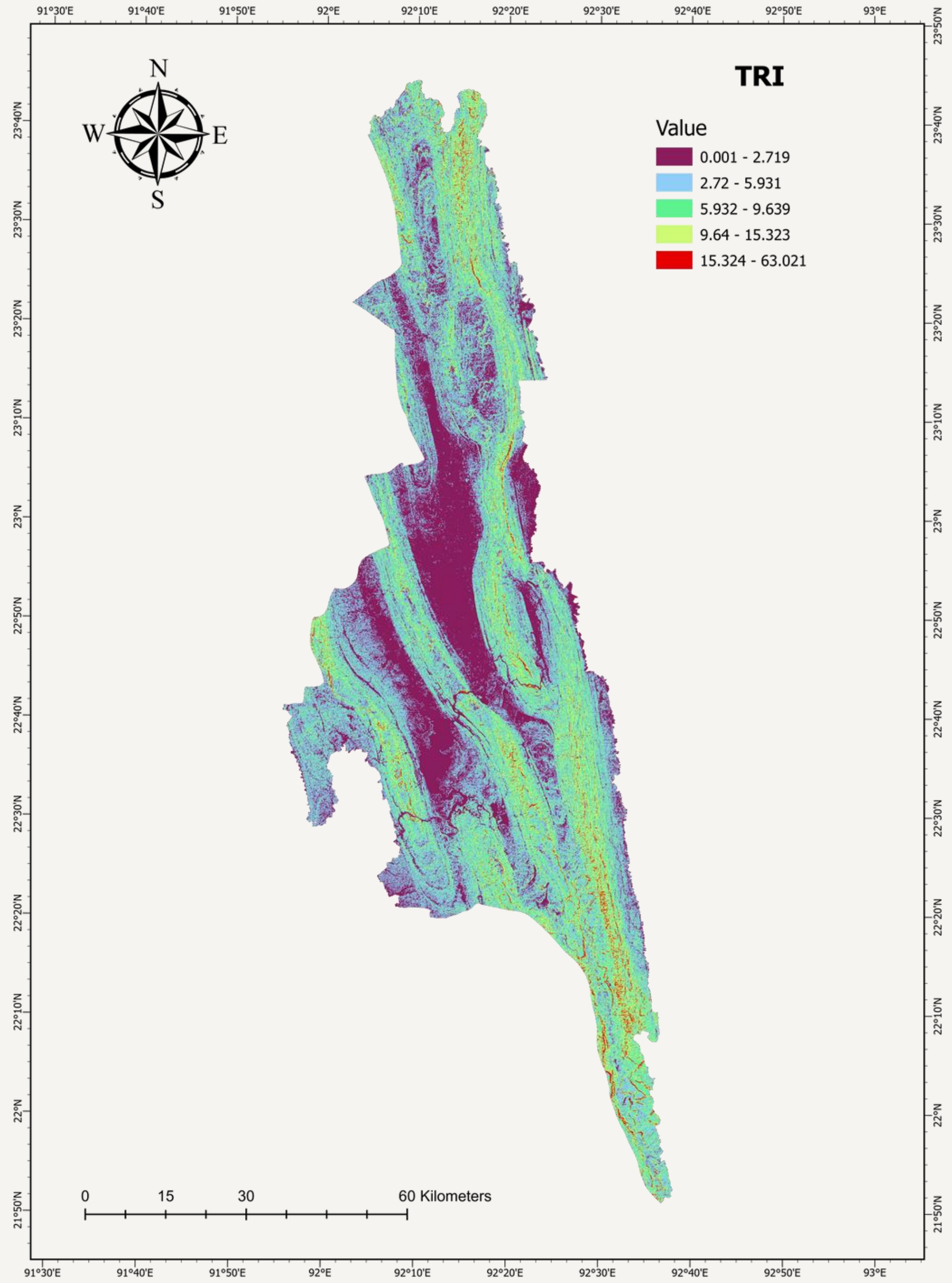
ANNUAL RAINFALL MAP



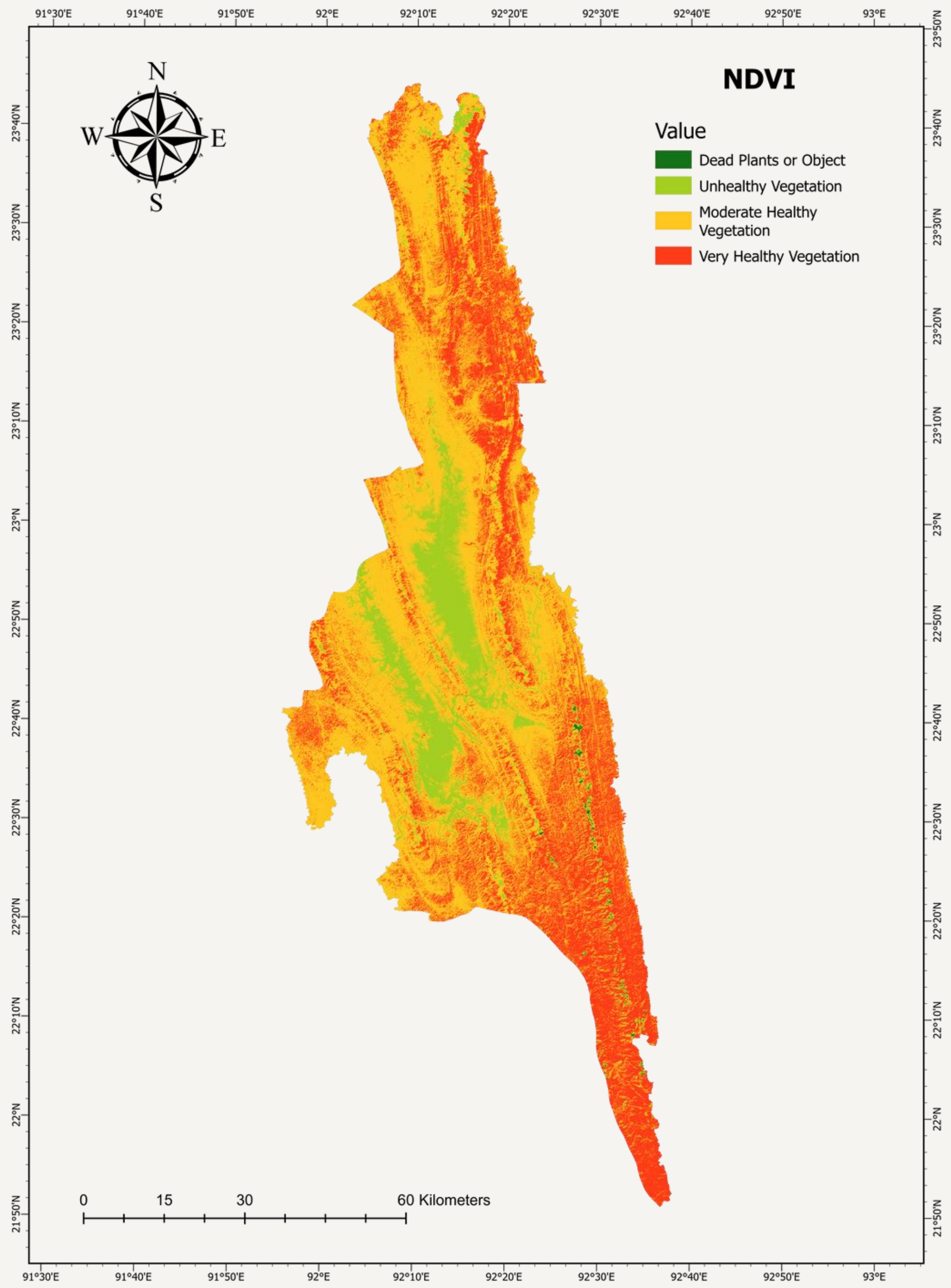
SPI MAP



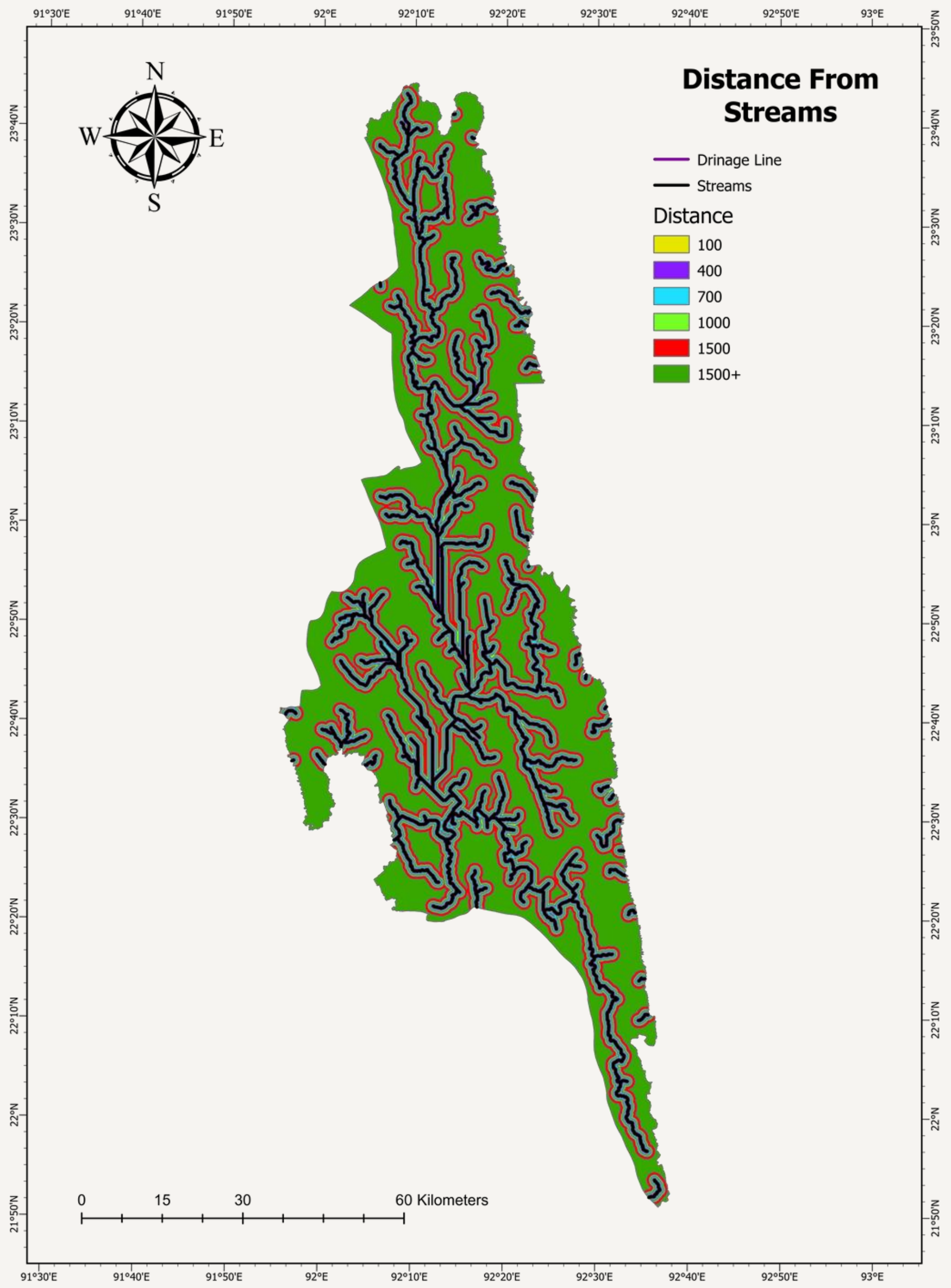
TWI MAP



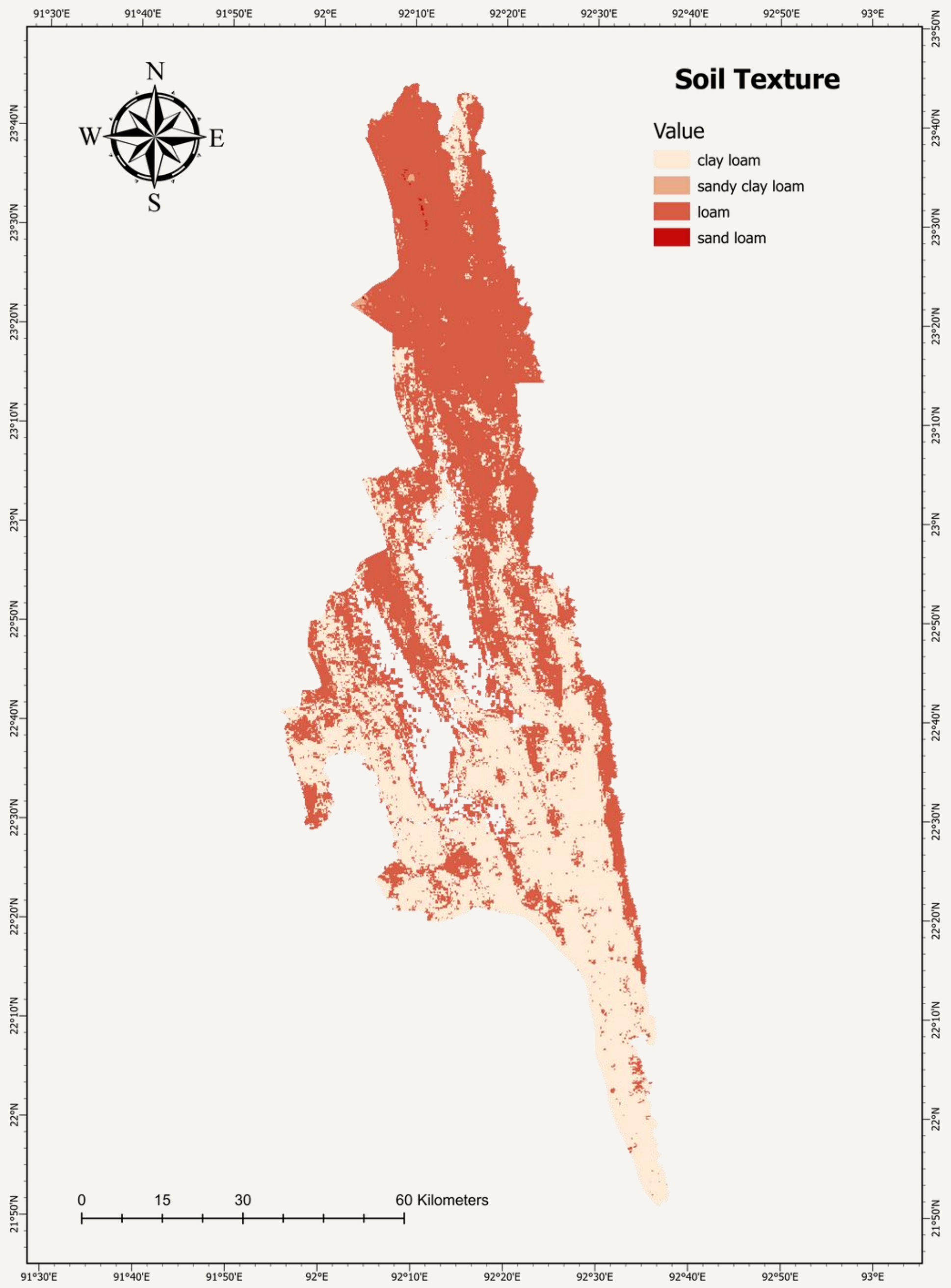
TRI MAP



NDVI MAP

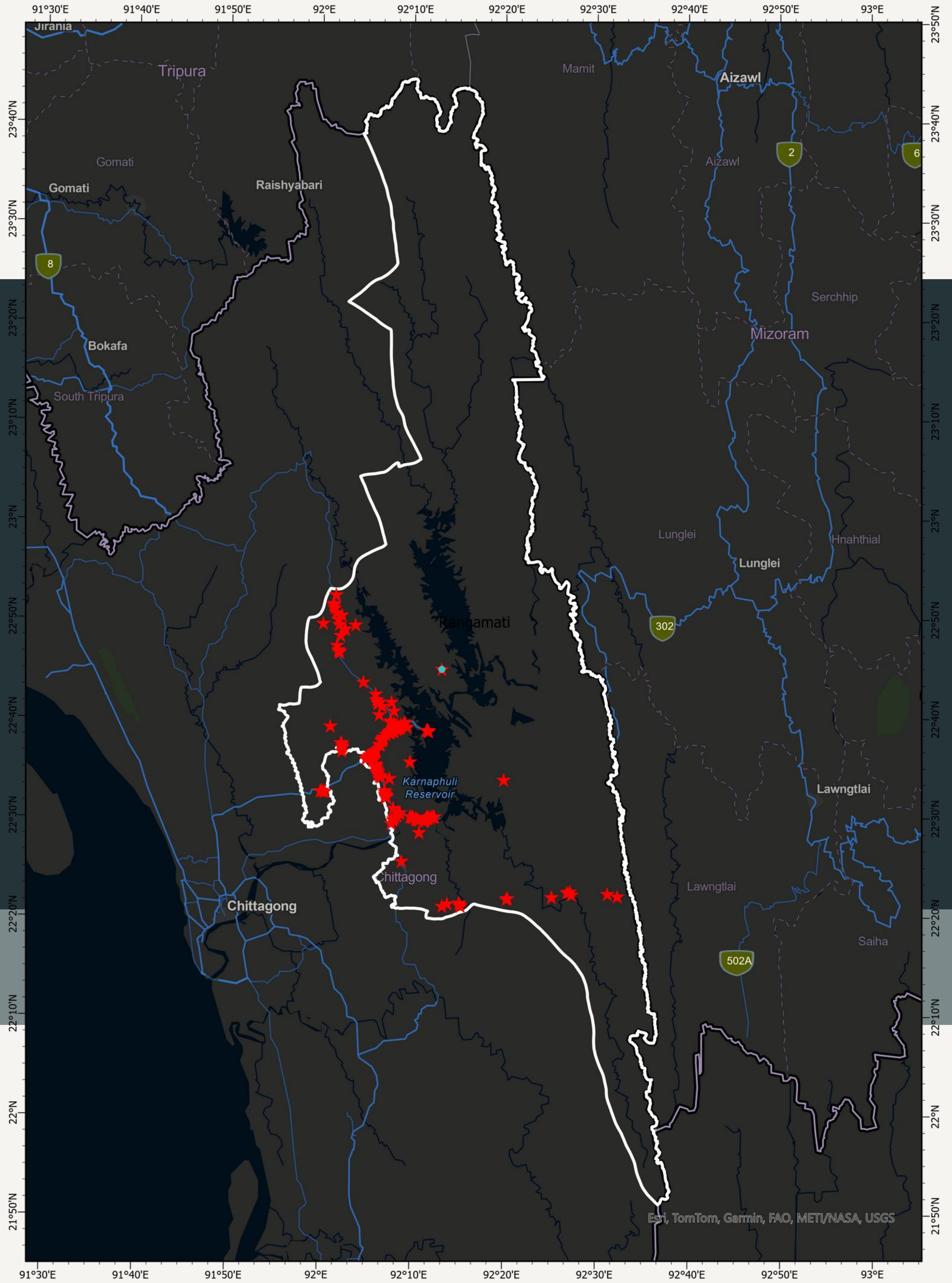


DISTANCE FROM STREAM MAP



SOIL TEXTURE MAP

WHY WOULD ANYONE BELIEVE THIS IS A SPARSE DATASET?

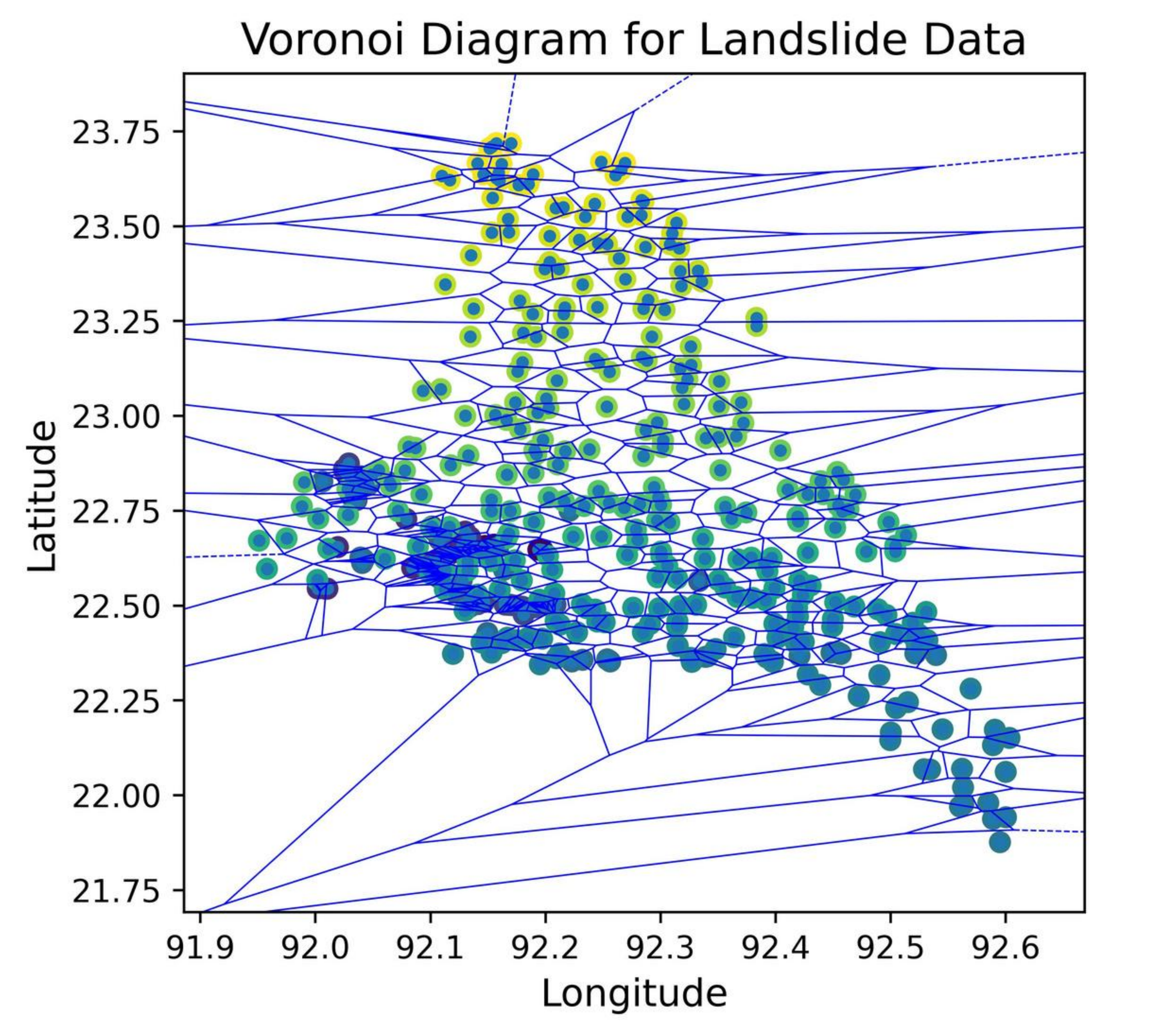


VORONOI-BASED SPARSITY QUANTIFICATION

Mathematically, the Voronoi cell of a point y in a set X is defined as:

$$VorX(y) := u \in R^n : y \in argmin_{(x \in X)} \parallel x - u \parallel 2$$

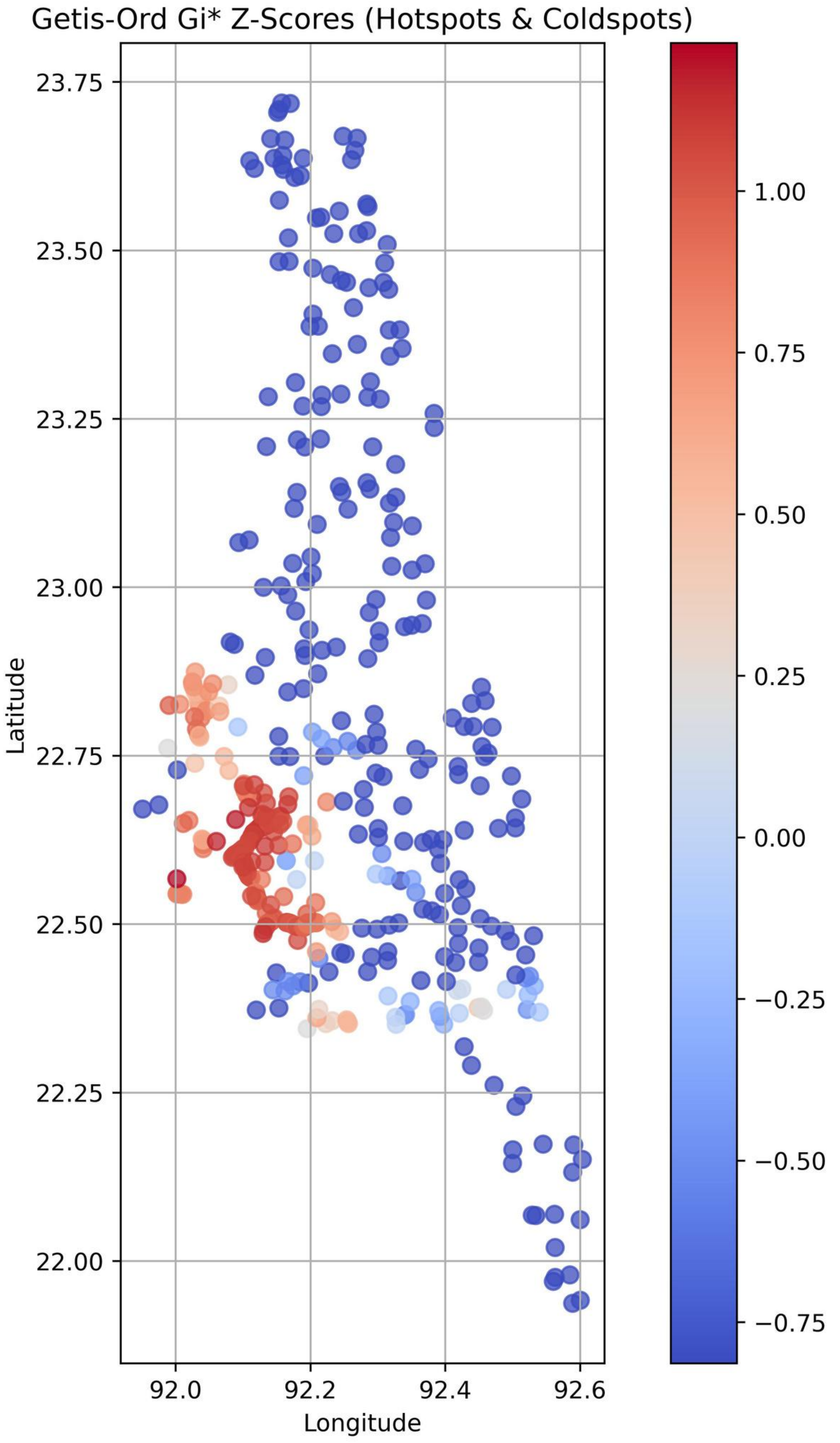
- Large Voronoi polygons indicate low data density and poor spatial coverage.
- Small polygons reflect high landslide concentration.



GETIS-ORD GI* HOTSPOT ANALYSIS

Detects statistically significant spatial clusters (hotspots & cold spots).

Hotspot **Neutral** **Coldspot**



MORAN'S I

Moran's I is one of the most widely used statistical tools that simultaneously measures spatial autocorrelation based on feature locations and values

$$I = \frac{n}{W} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where:

- n: Number of spatial units
- w_{ij}: Spatial weight between units i and j
- x_i, x_j: Attribute values of units i and j
- $\bar{x}$: Mean of all values
- W: Sum of all spatial weights

MORAN'S I SCATTER PLOT

I=0.7636

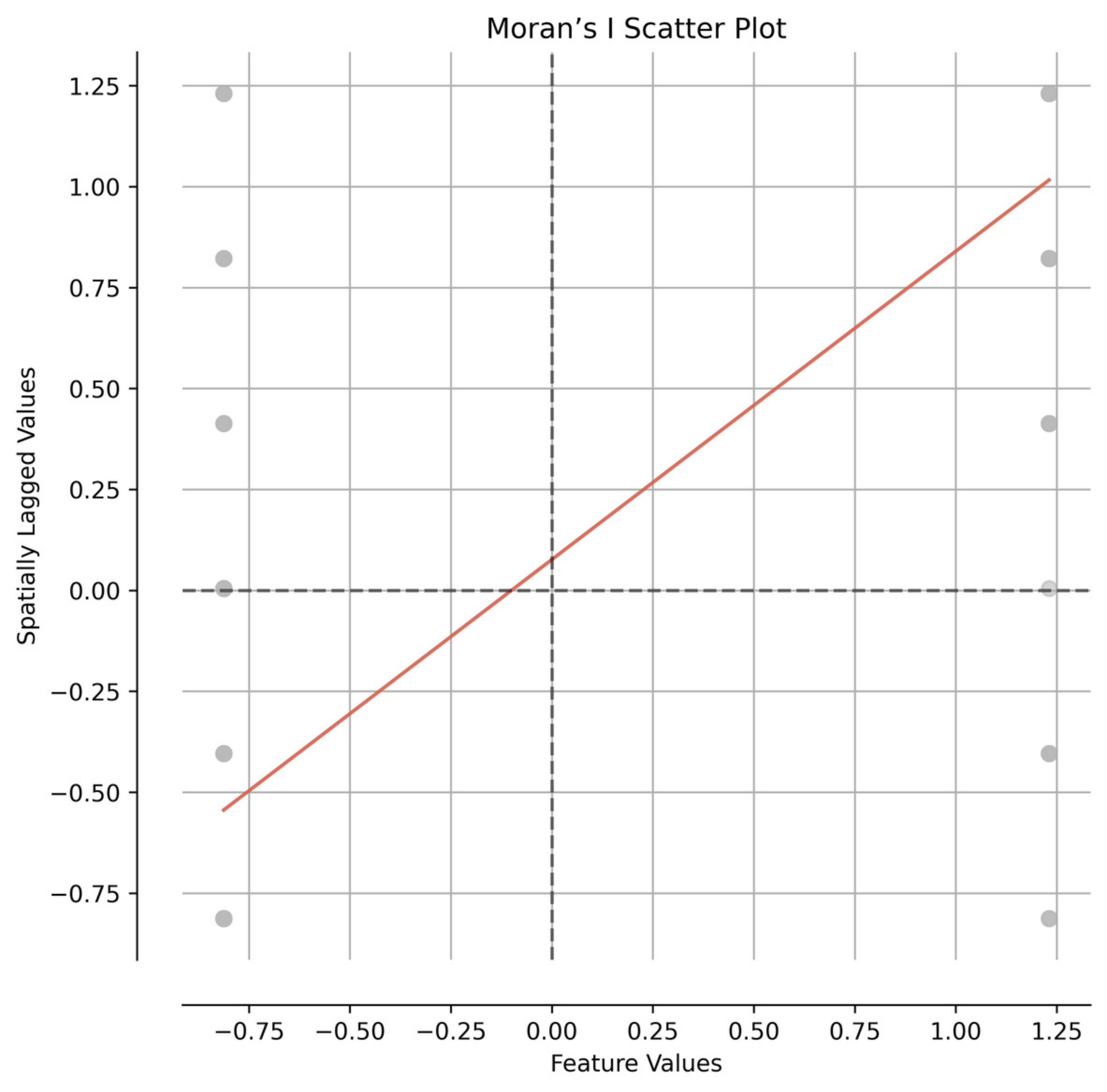
Strong Positive Autocorrelation:

Landslide points exhibit significant clustering (I>0)

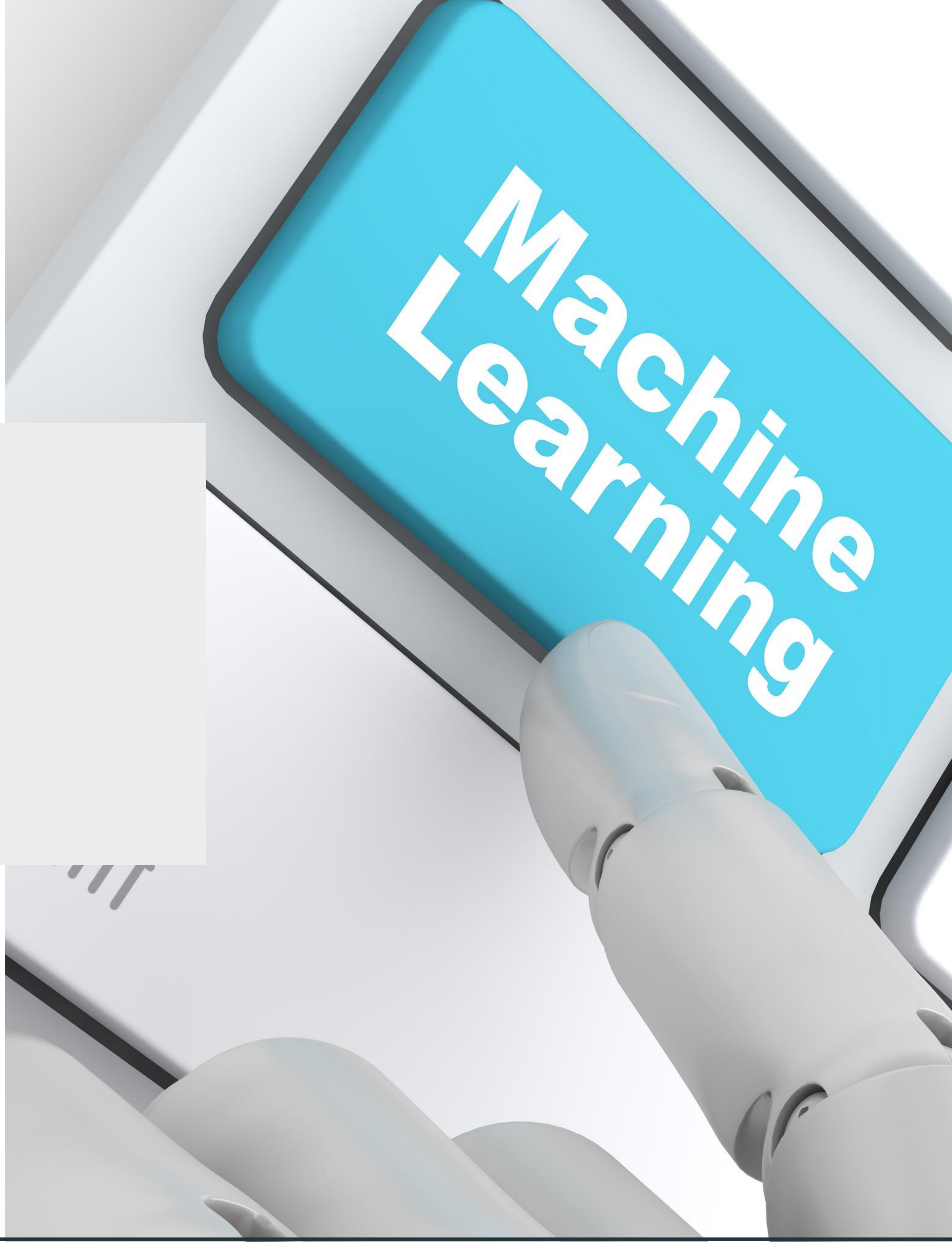
Regression Slope, β=0.82

Confirms strong dependence

Quantifies inherent spatial bias → justifies sparsity mitigation



APPLICATION OF CLUSTERING



DBSCAN

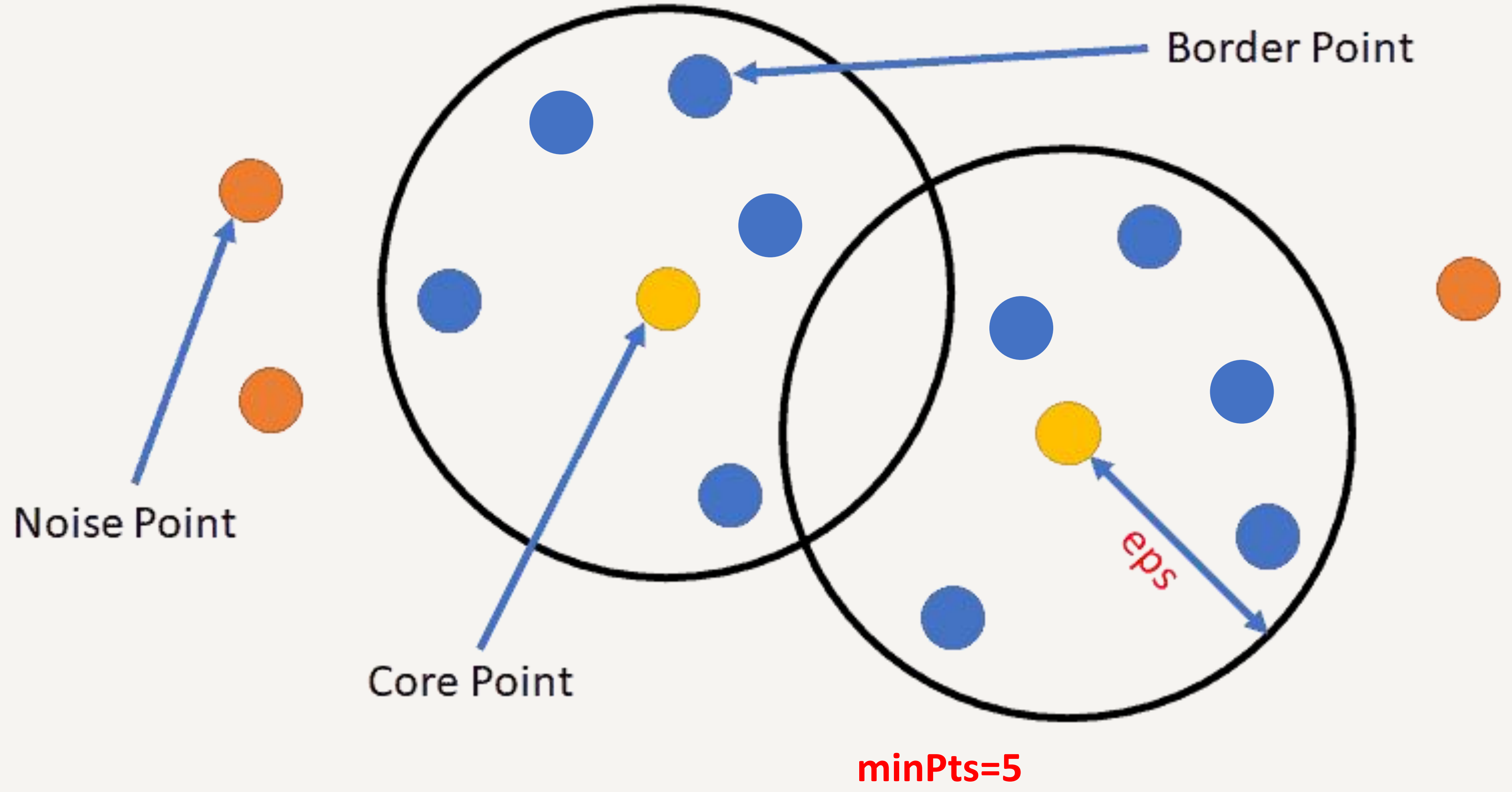
DBSCAN - **Density-Based Spatial Clustering of Applications with Noise**. Finds core samples of high density and expands clusters from them. Good for data that contains clusters of similar density.

DBSCAN Parameters

ε (eps): 0.01º (~1.1 km)

min_point: 5

Metric: Euclidean



SPATIAL DENSITY ANALYSIS

DBSCAN Clustering Results

- 9 clusters (IDs 0–8)
- noise points (labeled -1): Landslide points
- Red: Cluster 0 (non-landslide points)

Threshold:

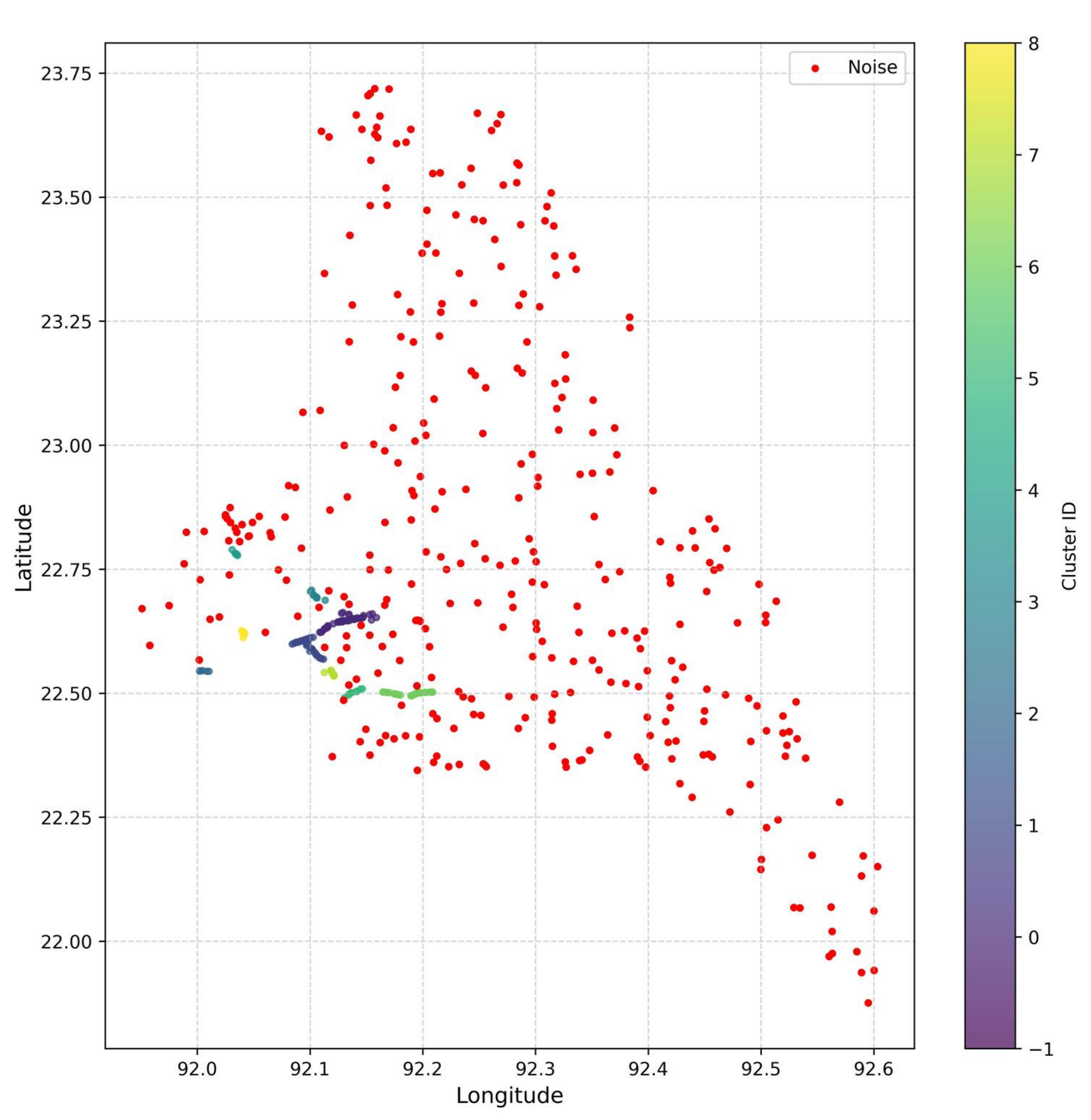
T = 0.2 × N / k = 4.4

→ Sparse if < 5 points

N = 198 landslide points

k = 9 clusters

- Sparse: Clusters 3, 5, 7
- Dense: Clusters 0, 1, 2, 4, 6, 8



OLD VS ENHANCHED DATA SET

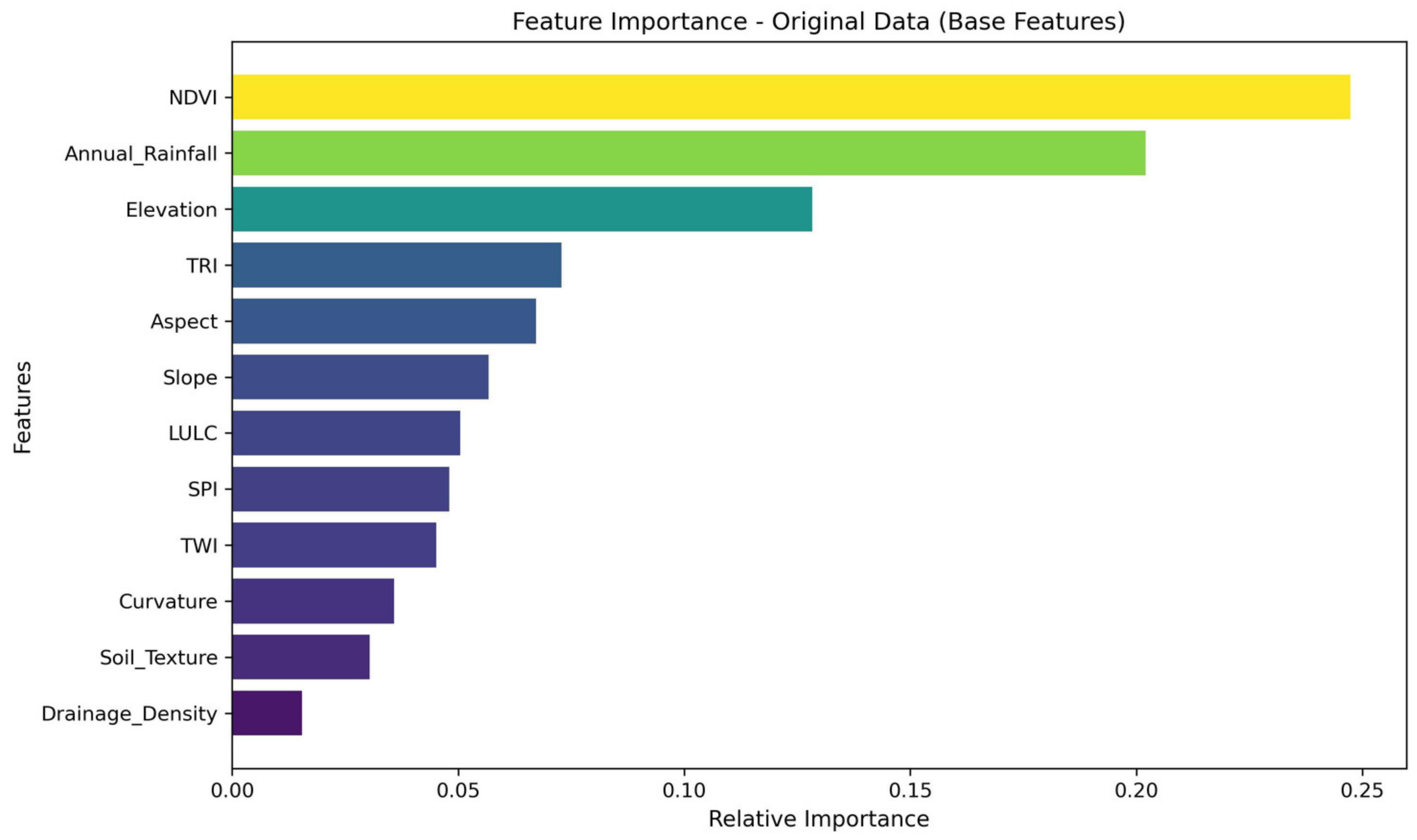
Original Dataset:

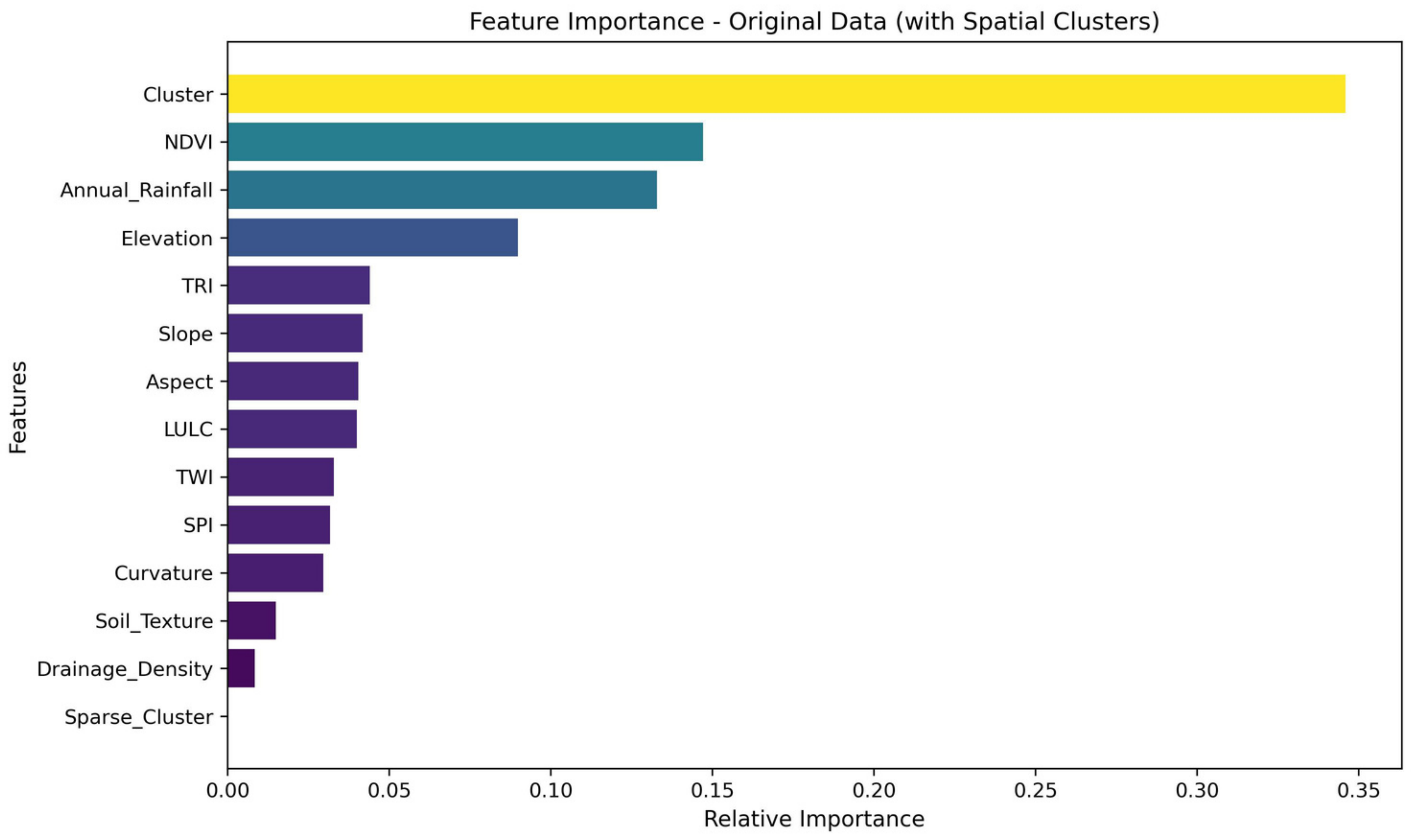
12 landslide conditioning factors (elevation, slope, rainfall, etc.) with no spatial context → models biased toward sampled areas.

Enhanced Dataset:

Original 12 factors + Cluster ID and Sparse_Cluster flag → **captures spatial patterns**

FEATURES IMPORTANCE

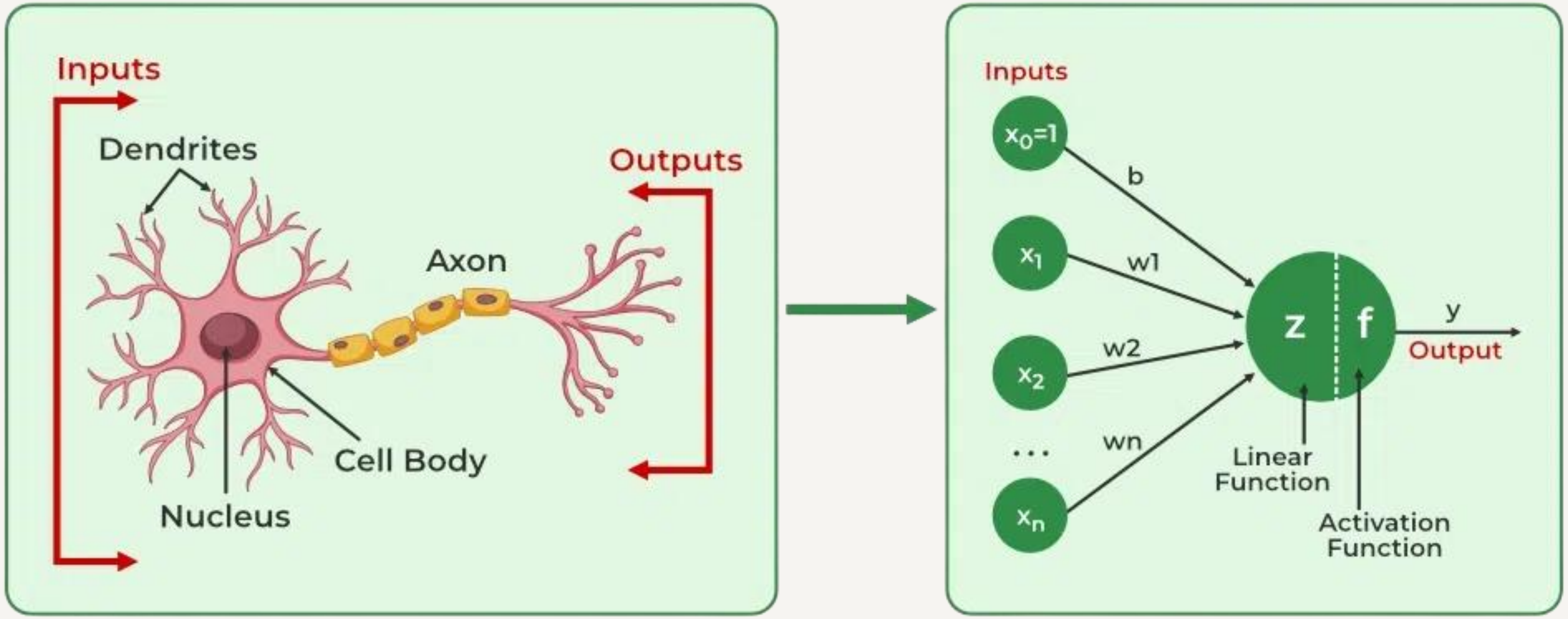




APPLICATION OF DEEP LEARNING

WHAT IS DEEP LEARNING

Deep learning is a subset of machine learning that uses neural networks with multiple layers to automatically learn complex patterns from data.



Source: https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/

GRIDENT DESCENT

$$w_{new} = w_{old} - \eta \times \partial L/\partial w$$

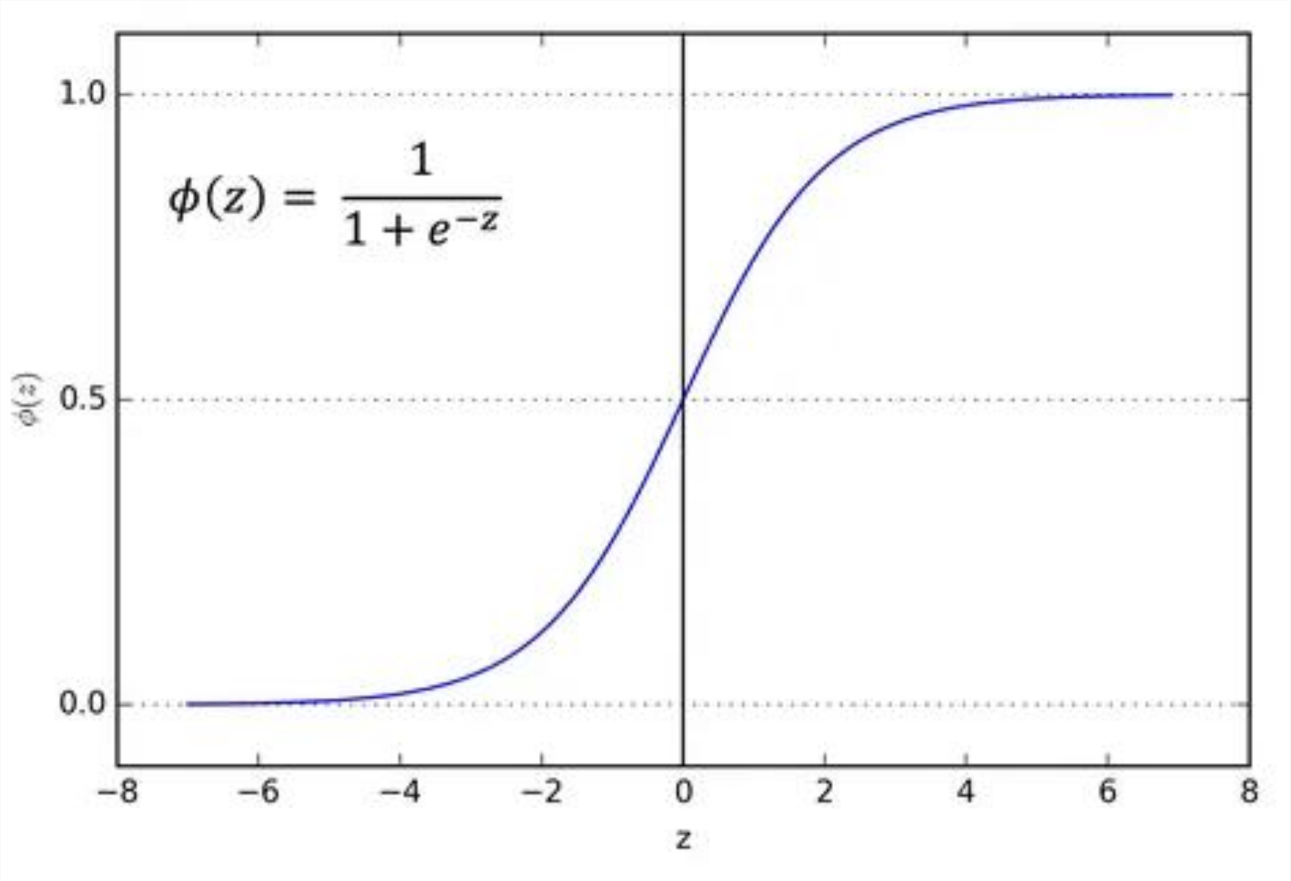
Where:

- w_{new} = Updated weight
- w_{old} = Current weight
- η = Learning rate
- $\partial L/\partial w$ = Gradient of the loss with respect to the weight

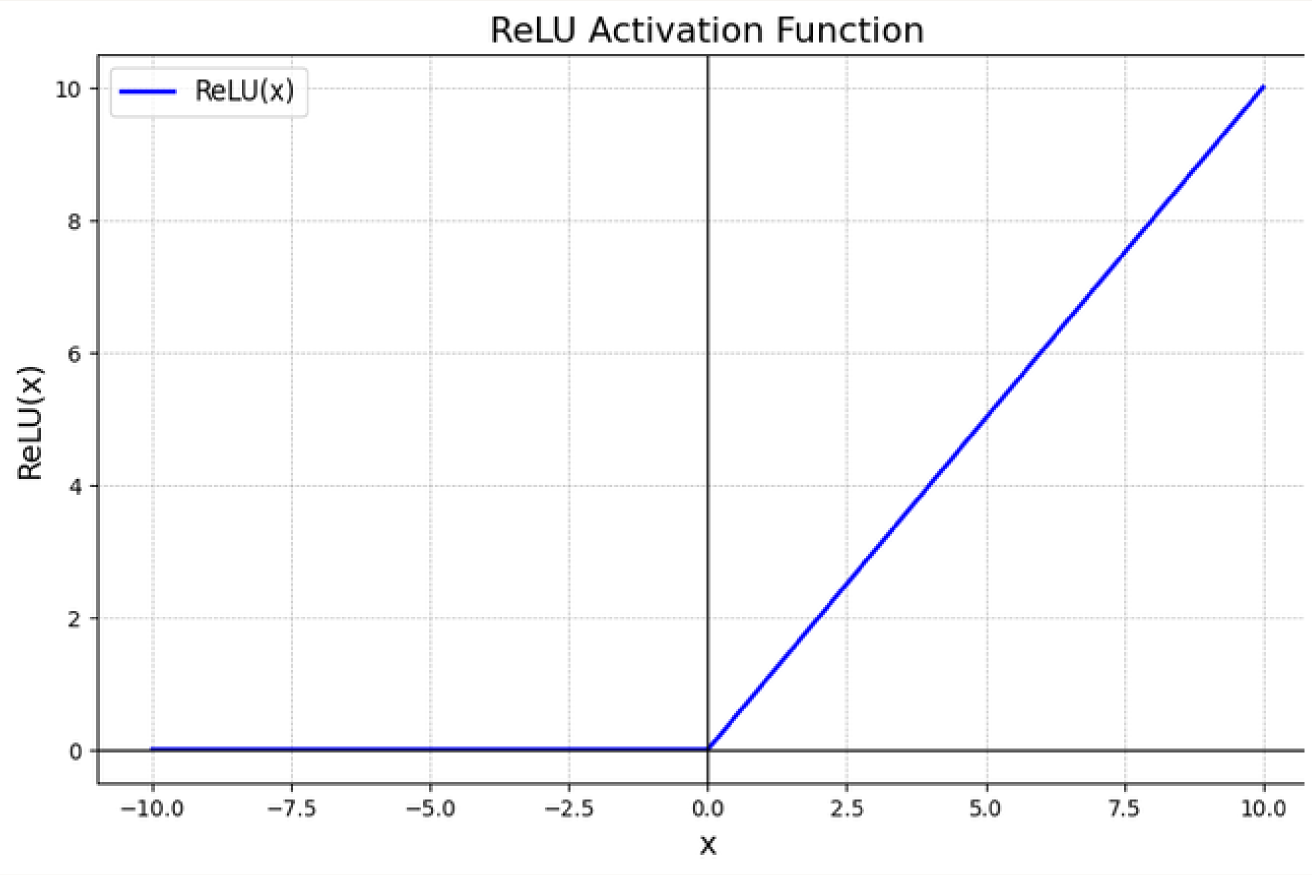
Source: https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/

ACTIVATION FUNCTION

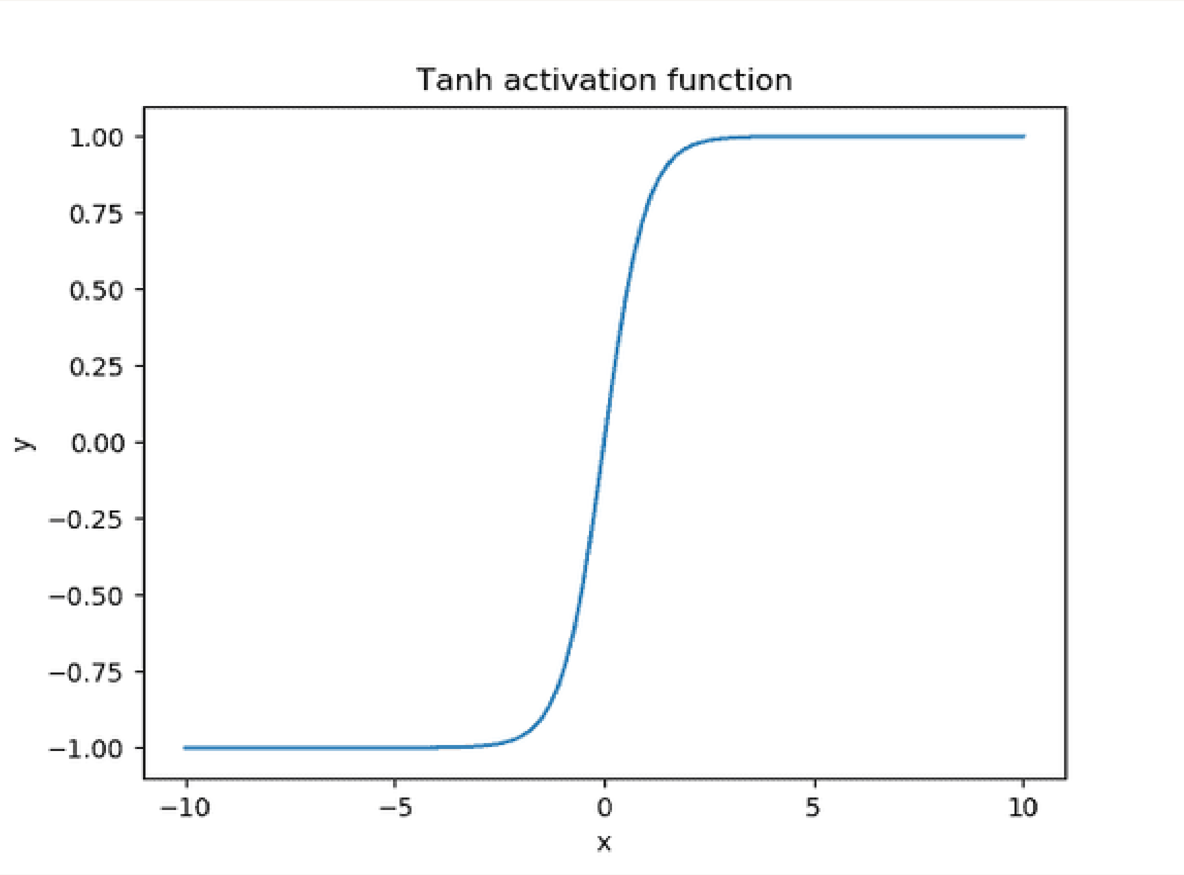
Sigmoid



ReLU



Tanh

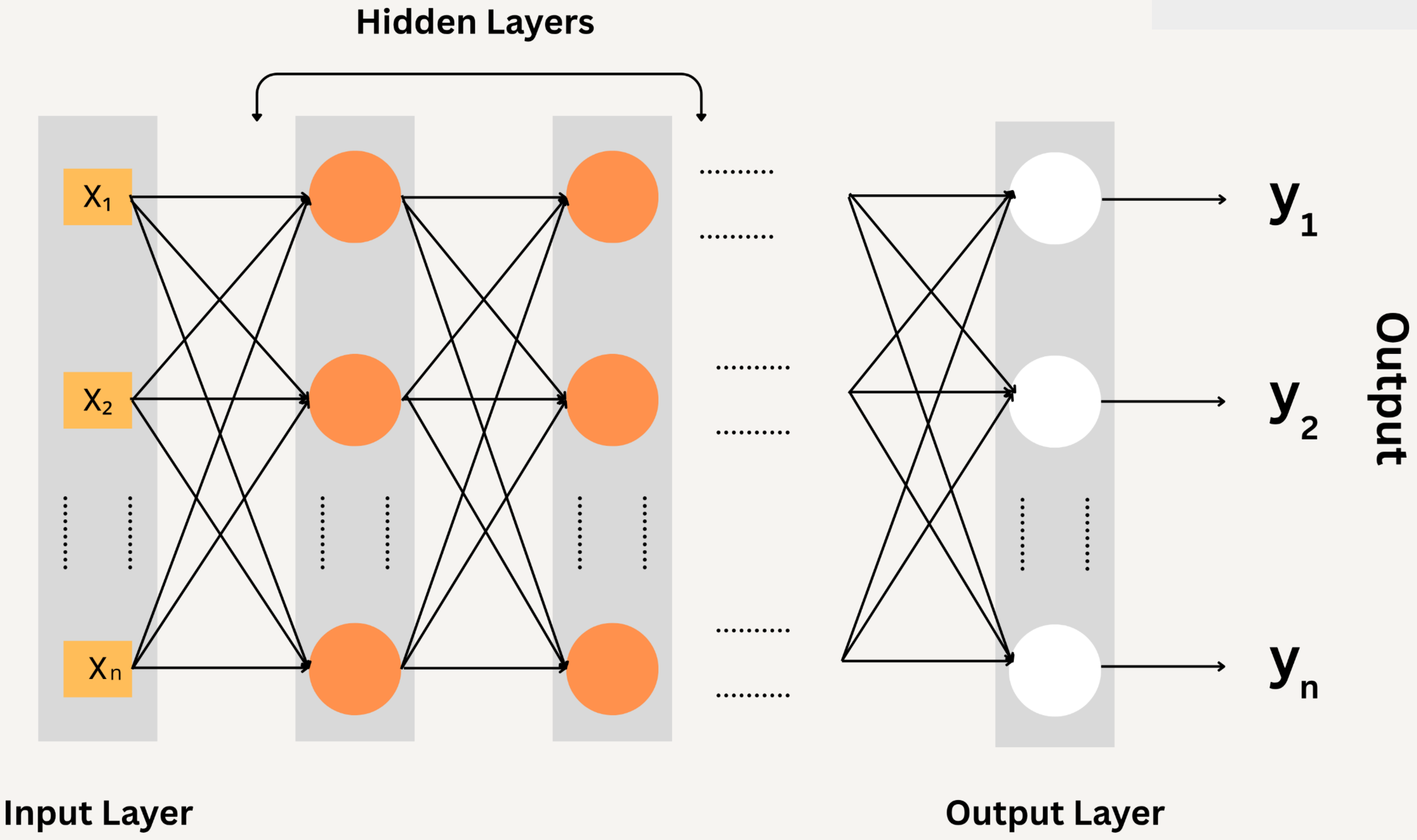


Source: https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/

DNN

DNN Architecture

- Input: 12 or 14 features
- Hidden Layers:
  - Layer 1: 64 neurons, ReLU, L2 + Dropout
  - Layer 2: 32 neurons, ReLU, L2 + Dropout
  - Layer 3: 16 neurons, ReLU, L2 only
- Output Layer: 1 neuron, Sigmoid → Probability (0–1)



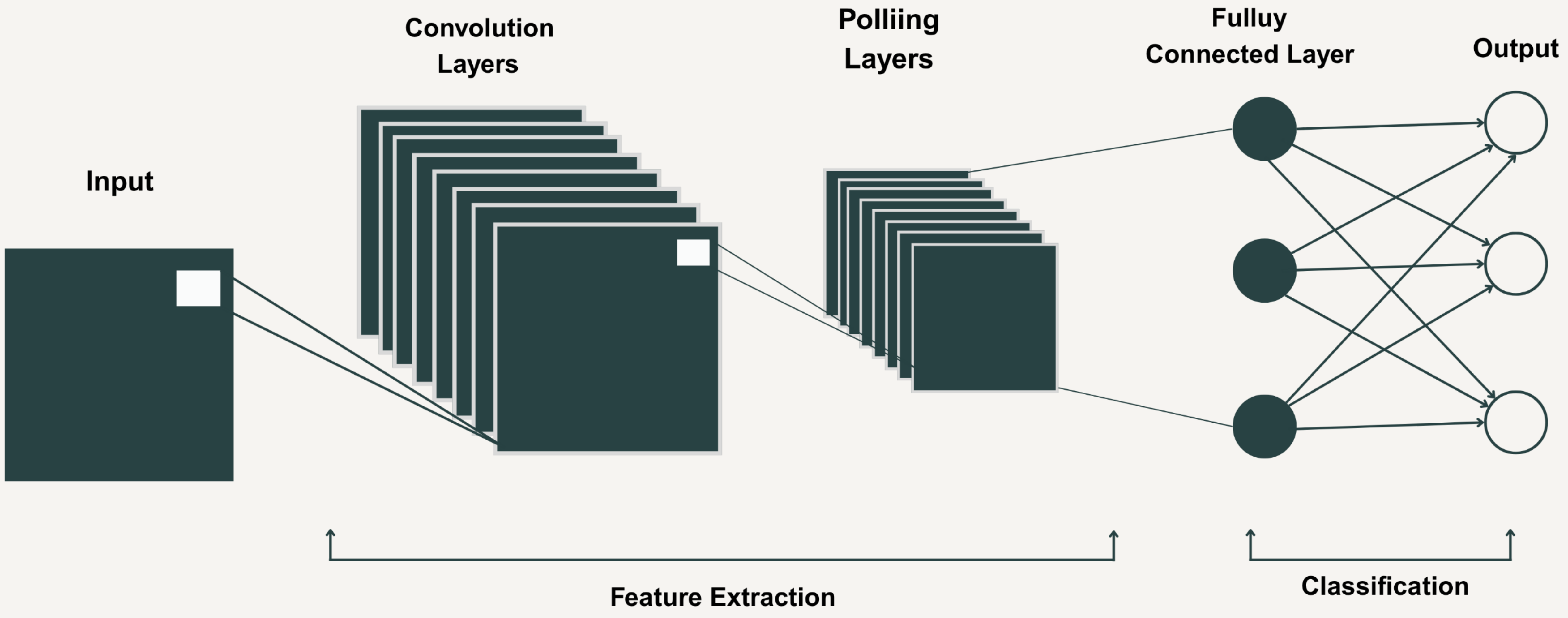
1D-CNN

Conv1D(64, ReLU) → Dropout(20%) → Flatten → Dense(32, ReLU) → Dropout → Dense(16, ReLU) → Dense(1, Sigmoid)

Input Shape: (features, 1)

Regularization: L2 (λ=0.01)

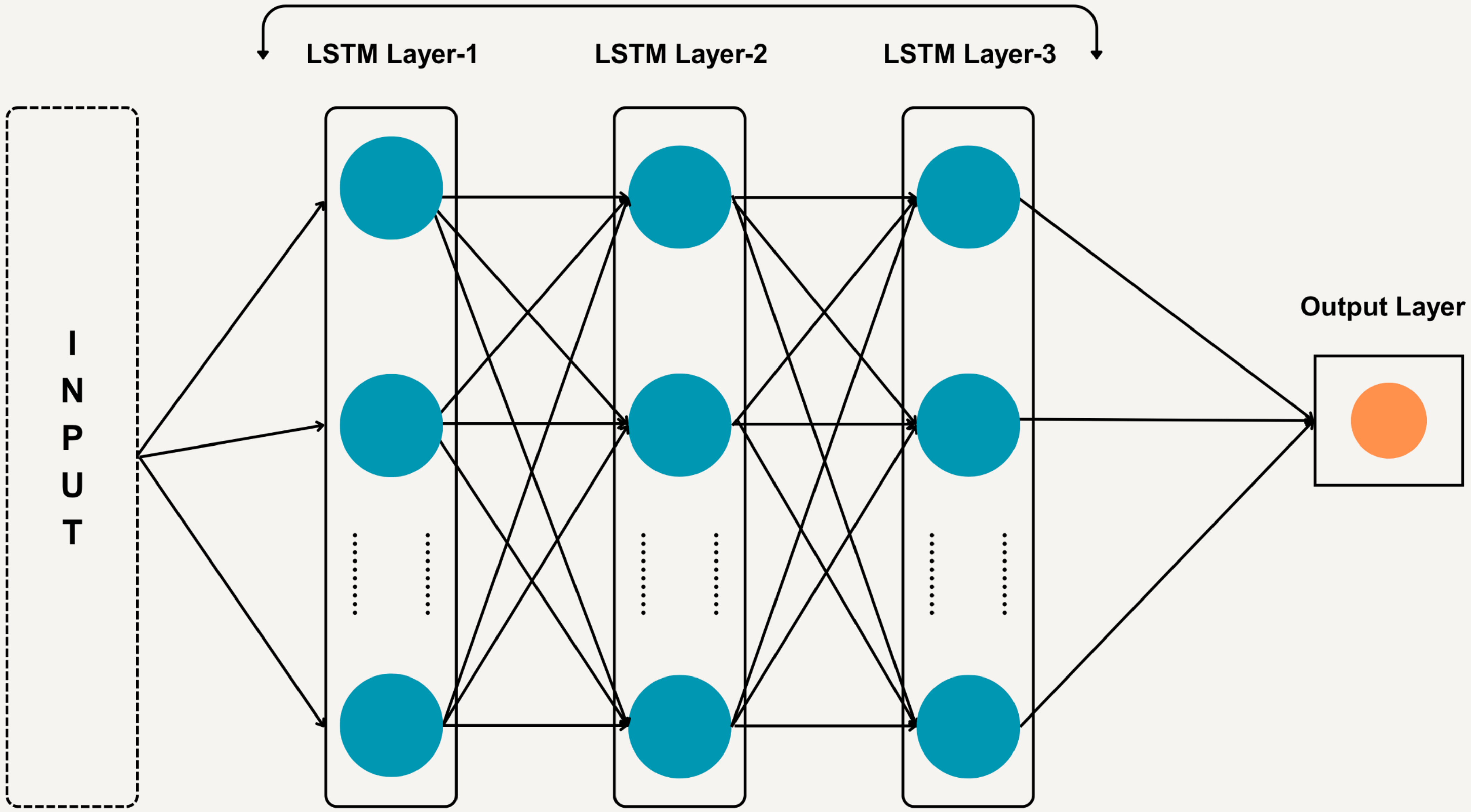
Output: Landslide Probability (0–1)



LSTM

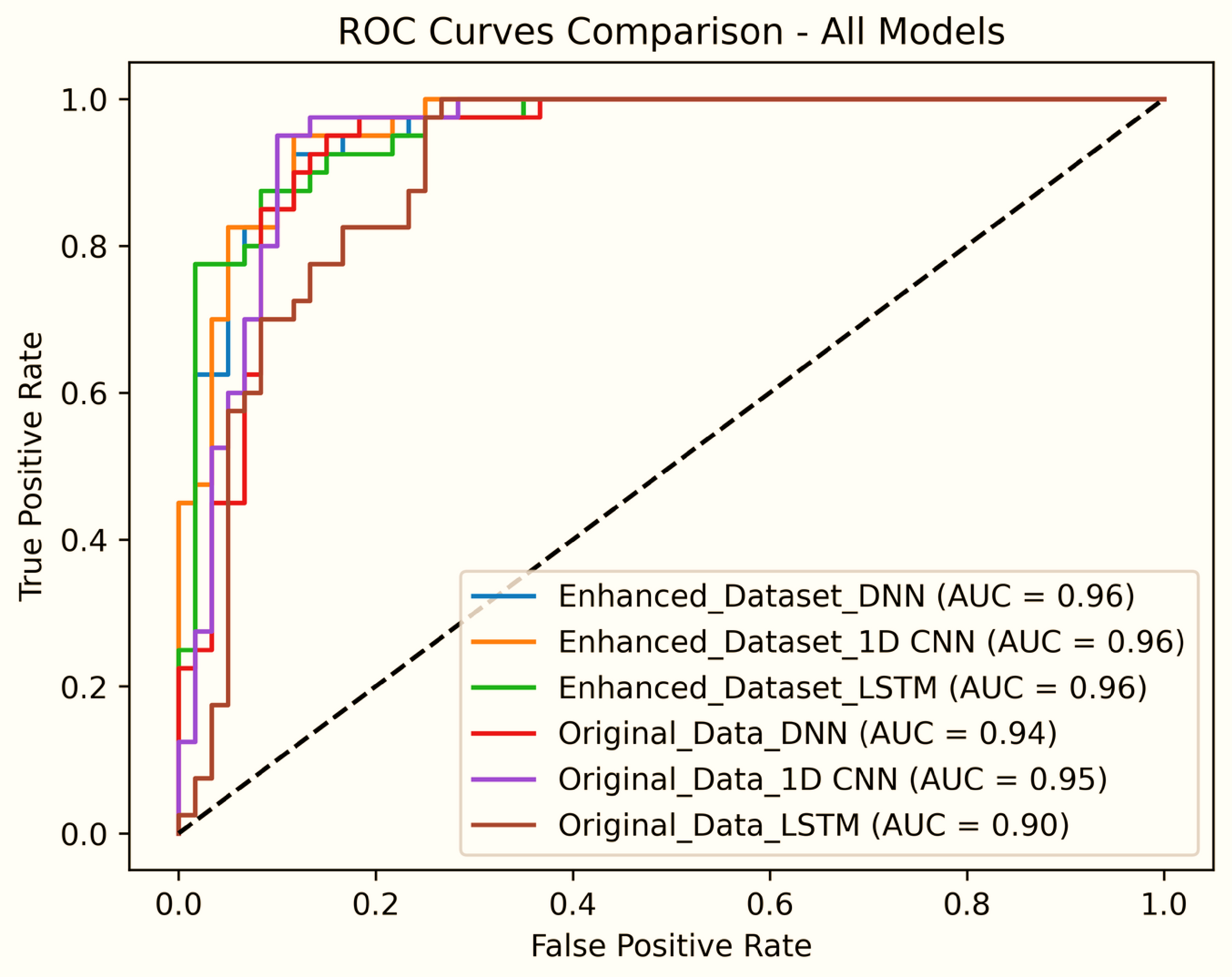
LSTM(128, return_seq=True) → LSTM(64) → Dense(32, ReLU) → Dense(1, Sigmoid)

Input Shape: (features, 1) Gate Activations: Sigmoid (gates), Tanh (state) Output: Landslide Probability (0–1)



ROC CURVE COMPARISION

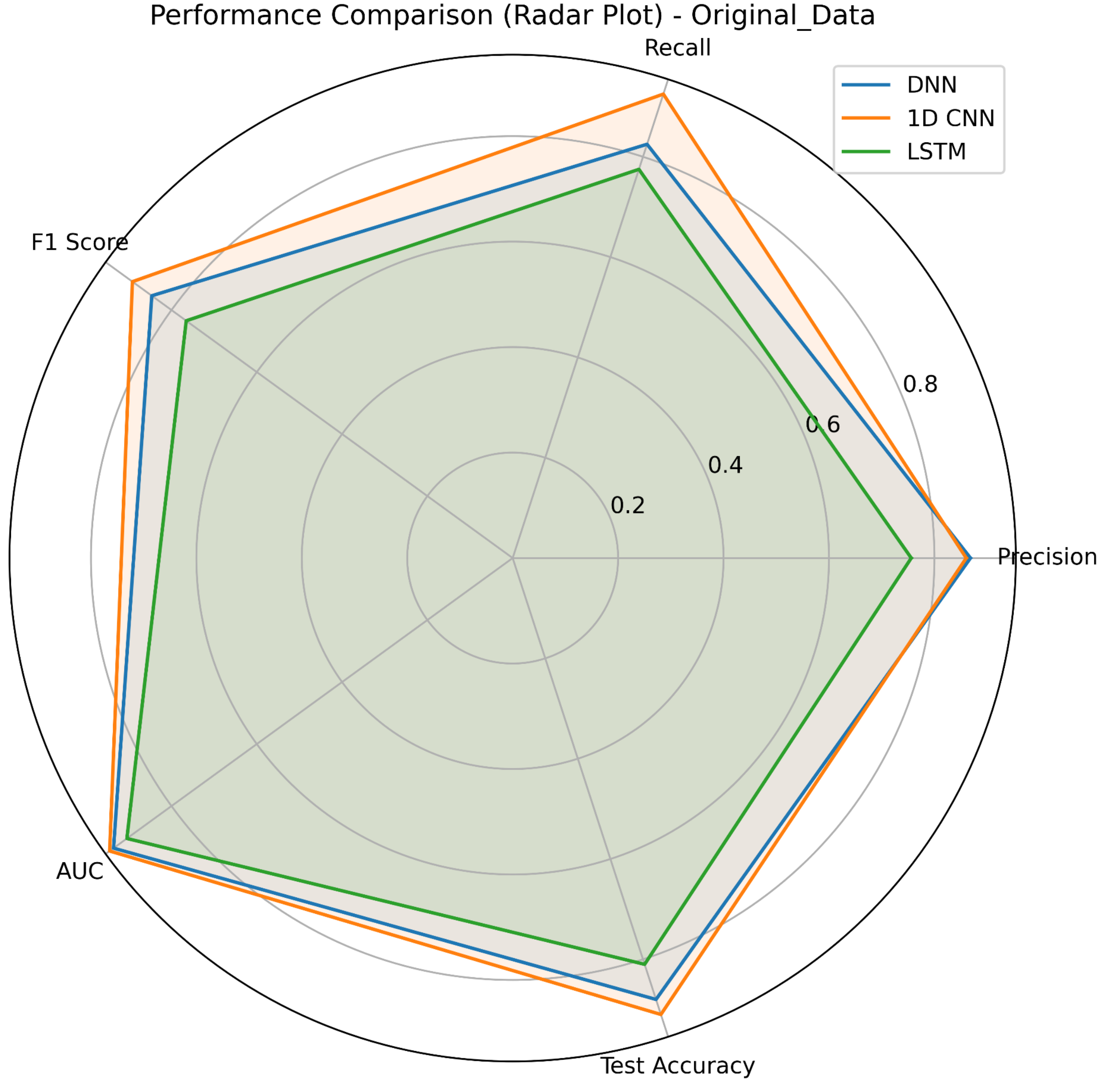
Model	Dataset	AUC
DNN	Enhanced Dataset	0.96
1D CNN	Enhanced Dataset	0.96
LSTM	Enhanced Dataset	0.96
1D CNN	Original Dataset	0.95
DNN	Original Dataset	0.94
LSTM	Original Dataset	0.90



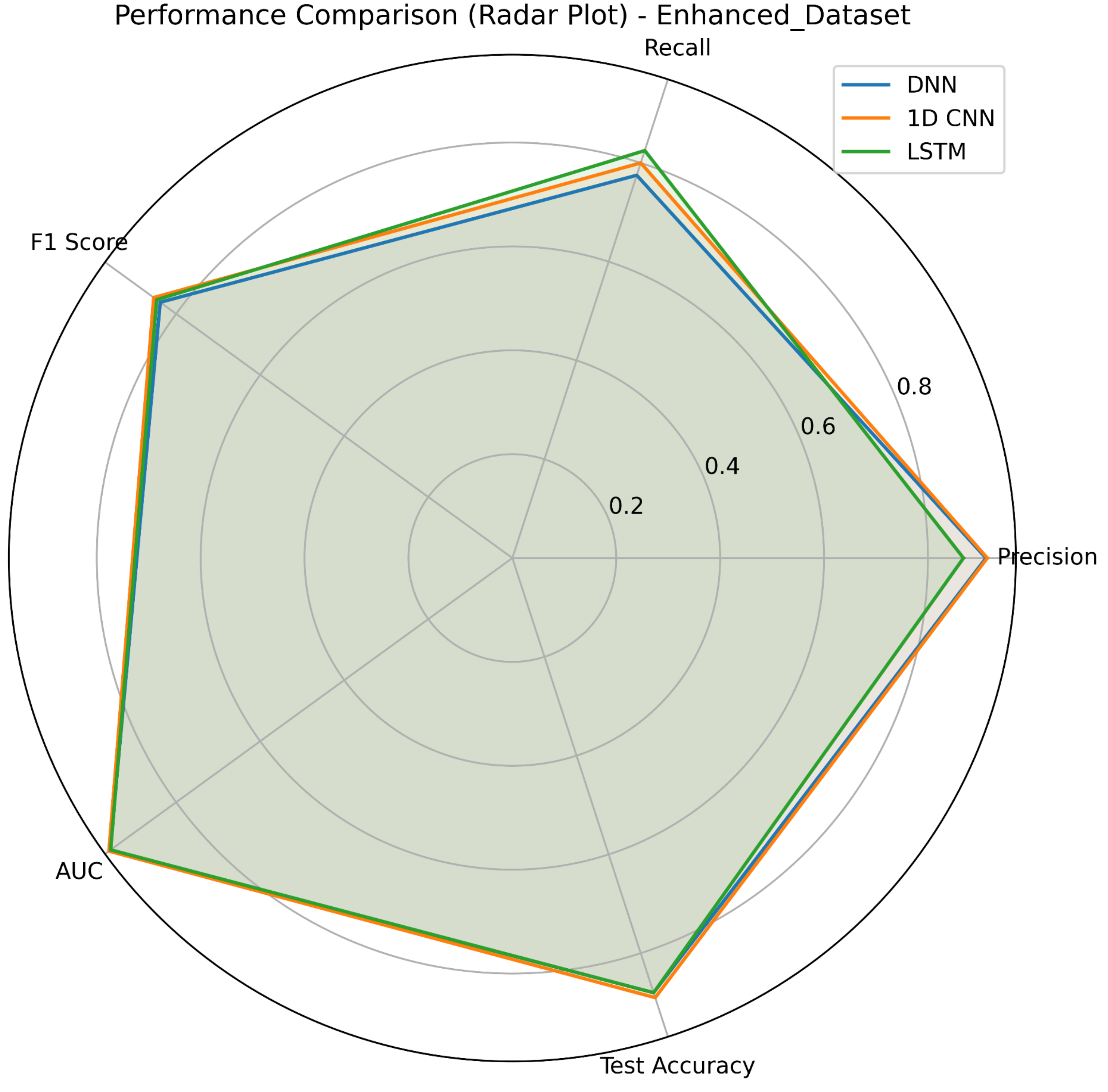
EVALUATION METRIX

Model	AUC	Train Accuracy	Test Accuracy	Precision	Recall	F1 Score
Enhanced_Dataset_DNN	0.96	0.919	0.87	0.909	0.75	0.821
Enhanced_Dataset_1D CNN	**0.962**	**0.896**	**0.89**	**0.914**	**0.8**	**0.853**
Enhanced_Dataset_LSTM	0.955	0.937	0.89	0.939	0.775	0.849
Original_Data_DNN	0.940	0.871	0.9	0.875	0.875	0.875
Original_Data_1D CNN	0.937	0.874	0.9	0.840	0.925	0.880
Original_Data_LSTM	0.927	0.884	0.81	0.8	0.7	0.746

RADAR PLOT



Performance Comparison (Radar Plot) - Original_Data
Recall
F1 Score
AUC
Test Accuracy
Precision
0.2
0.4
0.6
0.8
DNN
1D CNN
LSTM



Performance Comparison (Radar Plot) - Enhanced_Dataset
Recall
F1 Score
AUC
Test Accuracy
Precision
0.2
0.4
0.6
0.8
DNN
1D CNN
LSTM

EVALUATION: 1D-CNN

Metric	Value	Interpretation in Landslide Susceptibility Context
AUC	0.962	The model has a 96.2% probability of correctly ranking landslide zones higher than stable areas → Excellent discriminative power
Test Accuracy	0.890	89% of grid cells correctly classified → Reliable overall predictions
Precision	0.914	91.4% of "high-risk" predictions are correct → Minimal false alarms for efficient resource allocation
Recall	0.800	80% of true landslide zones detected → 20% omission risk
F1-Score	0.853	Optimal balance between precision/recall → Best single metric for policy decisions
Train-Test Gap	0.006	<1% performance drop on unseen data → **High generalizability to new terrain**

SUSCEPTIBILITY MAP

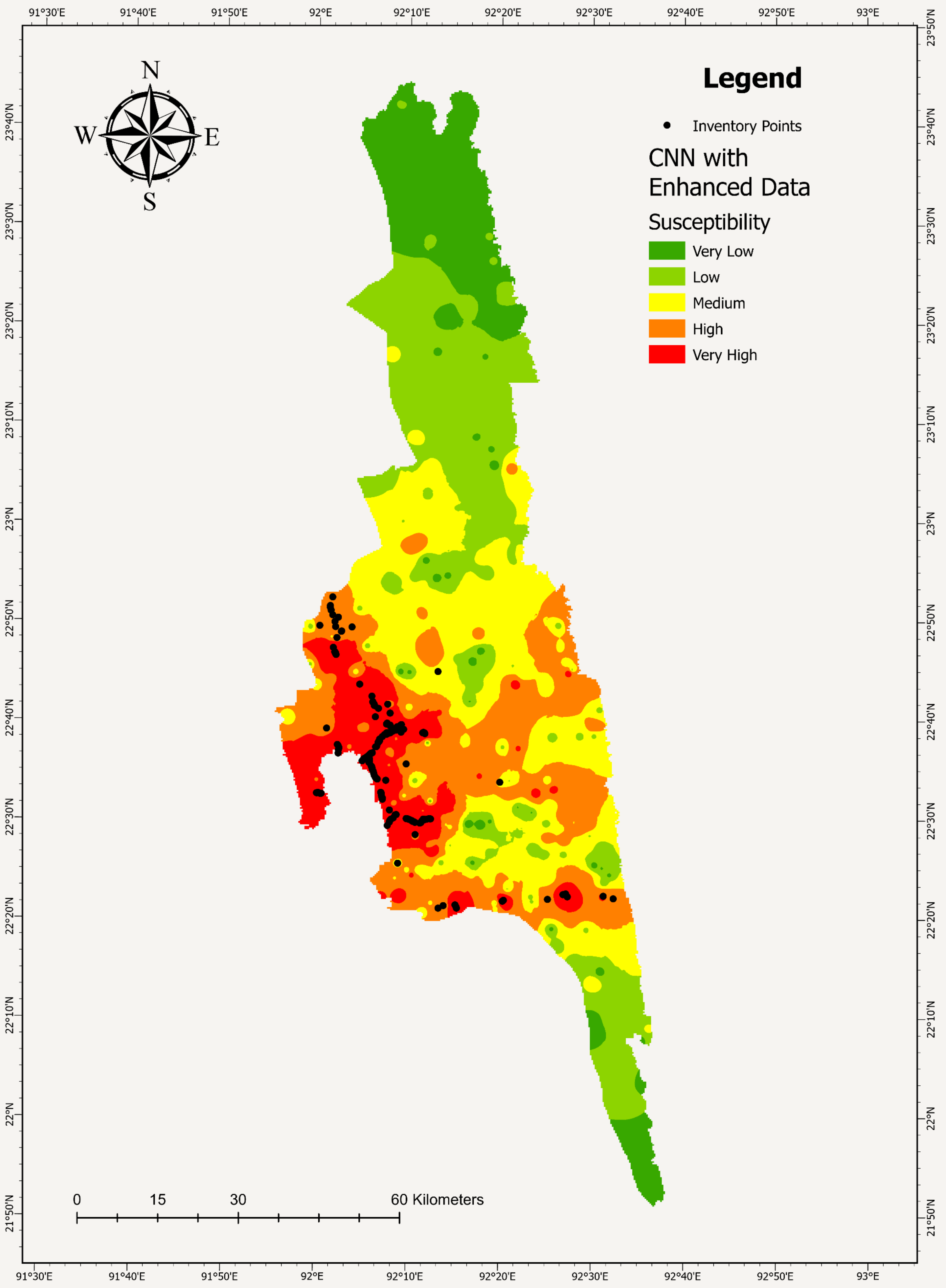
CRITICAL FINDINGS

Spatial Pattern:
Clear **N → S risk gradient** (Low → Very High)
78% of *Very High* zones lie **south of 22.5°N**

Validation Accuracy:
89% match with **historical landslides** (black dots)

High-Risk Cluster Alignment:
Steep slopes (>30°)
Monsoon rainfall hotspots

Anomalies:
Isolated red zones in **northern lowlands**
→ *Field verification recommended*

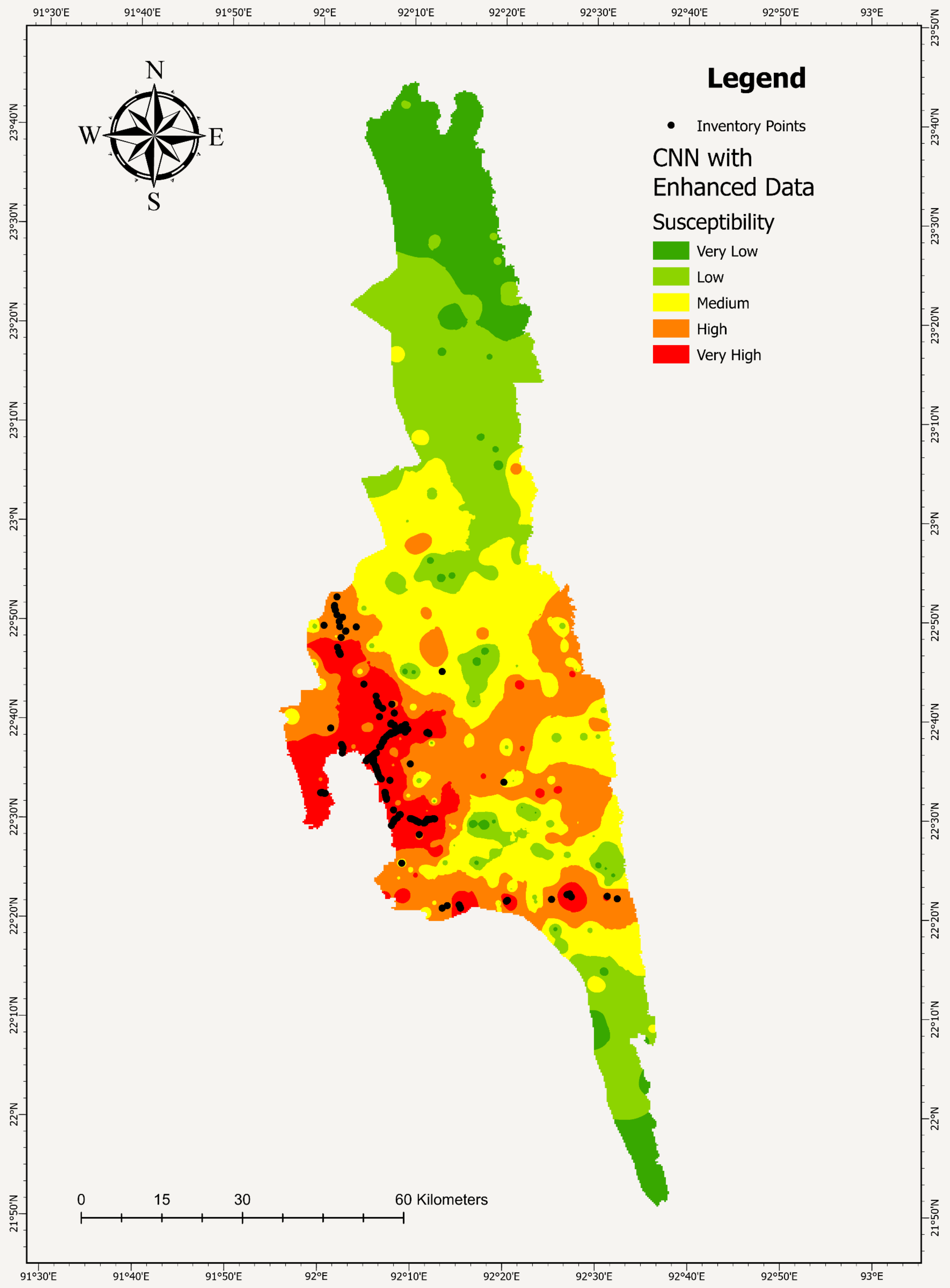


SUSCEPTIBILITY MAP

Disaster Management Implications

Priority Zones (Red/Orange = 40%):
Restrict infrastructure development
Install **real-time slope monitoring**

Evacuation Planning:
Emergency routes mapped to **avoid red zones**



CONCUSION

SPATIAL SPARSITY MITIGATED

- DBSCAN resolved the effect of sparsity
- Enhanced dataset boosted AUC by +2.7%

OPTIMAL MODEL IDENTIFIED

- 1D-CNN (AUC: 0.962) outperformed DNN & LSTM
- Precision: 0.914   Recall: 0.800   F1-Score: 0.853

ACTIONABLE SUSCEPTIBILITY MAP

- 40% of Rangamati is classified as High/Very High Risk
- Southern Zones Prioritised (e.g., Belaichari, Uraichari)

NOVEL CONTRIBUTIONS

- First Framework Combining Spatial Sparsity Metrics With Dl For Landslides
- Reproducible Python Workflow For Global Adaptation

LIMITATIONS

LIMITATION	CONSEQUENCE	SEVERITY
SPARSE SOUTHERN DATA	20% recall gap in critical zones	High
STATIC RAINFALL	Underestimates monsoon-triggered risks	Medium
BLACK-BOX MODEL	Reduced stakeholder trust	Medium
COARSE SOIL DATA	Local-scale inaccuracies	Low

FUTURE RESEARCH RECOMMENDATIONS

Data & Resolution

- Acquire 10m data via USGS/NASA
- Add field sensors in a sparse region
- Integrate 3-hourly rainfall from CHIRPS-GEFS

Deployment Strategy

- GEE Dashboard: Monthly risk updates
- Digital Twin: 3D terrain + real-time sensor feeds

Incorporate the concept of XAI

- SHapley Additive exPlanations (SHAP)
- Local Interpretable Model-agnostic Explanations (LIME)

Advanced Models

Model	Benefit
CNN-LSTM Hybrid	Spatial-temporal forecasting
Graph Neural Net	Slope network relationships
Attention Mechanism	Explain high-risk zones

REFERANCES

Ahmed, B. (2015). Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. Landslides, *12*(6), 1077–1095. https://doi.org/10.1007/s10346-014-0521-x

Anselin, L. (1995). Local indicators of spatial association—LISA. Geographical Analysis, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, 226–231.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. Geographical Analysis, *24*(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Humayain, K., & Biplob, C. (2021). Geological vulnerability assessment of Rangamati District. Unpublished technical report, Chittagong University of Engineering & Technology.

Abadi, M. et al. (2016). TensorFlow: A system for large-scale machine learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, 265–283.

Chollet, F. (2015). Keras. GitHub repository. https://github.com/keras-team/keras

Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, *12*, 2825–2830.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. Proceedings of the IEEE International Conference on Computer Vision, 1026–1034. https://doi.org/10.1109/ICCV.2015.123

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Python Software Foundation. (2023). Python (Version 3.11.4) [Computer software]. https://www.python.org/

ESRI. (2023). ArcGIS Pro (Version 3.1) [Computer software]. https://www.esri.com/

USGS. (2015). SRTM 30m digital elevation model. EarthExplorer. https://earthexplorer.usgs.gov/

SoilGrids. (2023). Global soil property maps [Data set]. International Soil Reference and Information Centre. https://soilgrids.org/

THANK YOU!