DEVELOPMENT OF A MACHINE LEARNING & ARTIFICIAL INTELLIGENCE MODEL FOR TIME-SERIES ANALYSIS TO FIND OUT ENCROACHMENT AND CHANGE DETECTION

Punjab Urban Land Systems Enhancement (PULSE) Project, Board of Revenue Punjab

April 2025

Terms of Reference (ToR) for Development AI/ML Models to Find out Encroachment and Change Detection

1. Introduction

The objective of this project is to develop a cloud-based AI/ML application capable of processing free-source datasets of building footprints. Starting with the Normalized Difference Built-up Index (NDBI) to calculate land use changes over time, the application will use deep learning models to calculate specific indices and implement a model to classify built-up areas based on Sentinel satellite imagery. This process will incorporate multi-modal data sources, such as Sentinel-1, Sentinel-2, and Digital Elevation Models (DEMs), to enhance classification accuracy and detect building footprints more effectively. The application should support both online and local dataset inputs and generate classified building footprint data, height information, and shapefiles from classified raster data. The system should support a variety of satellite imagery formats, including but not limited to Sentinel-1, Sentinel-2, and their variations for future scalability.

2. Scope of Work The selected vendor will be responsible for

2.1 Data Ingestion & Processing

- Develop an interface to accept and process free-source datasets from online repositories (e.g., Sentinel Hub, Google Earth Engine) and local machine-based inputs.
- Develop ETL (Extract, Transform, Load) pipelines to process datasets from diverse repositories.

2.2 NDBI Analysis

Implement algorithms for the analysis of building footprints using historical and current datasets and calculate their Normalized Difference Built-up Index (NDBI) on large datasets likewise all related indices related to built-up area, agriculture land and water etc.

Building footprint generation is not restricted to freely available online datasets; it can also be carried out using locally acquired datasets obtained through drone or satellite imagery.

2.3 Deep Learning Model for Higher Dimension Analysis on Larger Data Sets

Train and deploy a deep learning-based classification model for automatic detection of building footprints on online and local datasets, both. The model will incorporate the following techniques:

- Multi-temporal satellite imagery (e.g., Sentinel-2 for optical imagery).
- Digital Elevation Models (DEMs) for building height estimation.

- Change detection analysis using multi-temporal data —whether raster or vector—can be performed using freely available online sources or locally archived datasets.
- Accuracy of all AI based analysis & ML model must have minimum 80% precision.

2.4 Application Development & Deployment

- Develop a cloud-based platform with an intuitive user interface.
- Ensure compatibility with multiple data formats (e.g., **GeoTIFF**, **SHP**, **CSV**).
- Provide visualization tools for interactive exploration of classified data.
- Capable to deploy the application on a scalable cloud environment (AWS, Google Cloud, Microsoft Azure) as well on the local server.

2.5 Output Generation & Data Export

- Generate classified raster and vector-based shapefiles of output datasets, including building footprints and height estimation.
- Provide downloadable reports and geospatial analysis outputs, such as maps, statistics of building areas, and height metrics.
- Implement API endpoints for integration with GIS platforms for further analysis and decision-making.

3. Deliverables The contractor is expected to deliver:

- A fully functional cloud-based/local machine learning application.
- A time-series analysis module for building footprint change detection.
- A statistical module for building footprint changes detection.
- A deep learning classification model for building footprint and height estimation.
- A dashboard for interactive visualization and data download.
- Integration with valuation app to scan a change in a parcel to check the if valuation needs to be changed at the time of transfer or challan generation.
- Integration with state land module / maps to detect encroachment using AI.
- Integration with base revenue maps/layers.
- Module for inspecting a file (CAD/SHP or any other valid format) to auto check and generate report as per the criteria of the regulator. Parameters may be defined in the configuration.

- Technical document for deployment steps, third party dependencies and preparation of PROD environment, complete guide.
- API documentation for external system integration.
- User and technical manuals.
- Training sessions for relevant stakeholders.

4. Preferred Technical Requirements

- Programming Languages: Python, JavaScript (for UI/UX development).
- Machine Learning Frameworks: TensorFlow, PyTorch.
- Geospatial Libraries: GDAL, Rasterio, Shapely, OpenCV.
- Cloud Platforms: AWS, Google Cloud, or Microsoft Azure. Ideal deployment would follow a Kubernetes-based infrastructure for scalability.
- Database: PostgreSQL with PostGIS for spatial data storage. Integration with cloud services like AWS Lambda and API Gateway for scalability.
- The application will use Role Base Access Control (RBAC) for user and role implementation and will be configurable.
- Data Augmentation Techniques: The model will employ data augmentation strategies such as rotation, flipping, and scaling to improve generalization on large and diverse datasets.
- Application Code will be entirely ownership of the employer, firm cannot further sell this code or part of the code to any other private or govt. organization.

5. Timeline

The project is expected to be completed within 6 months, with the following milestones:

- Month 1-2: Data pipeline setup and preprocessing module development, including preprocessing of Sentinel-1 and Sentinel-2 data and DEMs.
- Month 3: Implementation of statistical analysis, deep learning model training, and evaluation using accuracy metrics such as F1 score and IoU for building footprint detection.
- Month 4-5: UI/UX development, integration of visualization tools, and model optimization based on performance.
- Month 6: Deployment on cloud, real-time testing, final documentation, and user manual delivery.

6. Qualification & Experience Requirements

- Registration with Income Tax and Sales Tax Department.
- Valid/active NTN and GST Number with Income Tax returns of the last three years.
- At least 3 years of experience in similar nature geospatial machine learning applications projects in Pakistan or other countries.
- Strong Deep Learning and Statistical Techniques Experience, including but not limited to Neural Networks, Transfer Learning Techniques, Image Segmentation and Spatial Imagery Knowledge
- Proven experience in cloud-based application development and deployment projects.
- Strong background in handling satellite imagery, geospatial data processing, and multi-source data integration.