A Practical Framework for Large-scale Land Use and Land Cover (LULC) Classification: Methods, Validation, and Applications

Abstract

Reliable Land Use and Land Cover (LULC) maps are essential for regional planning and environmental monitoring in India, yet existing frameworks often fail to balance high spatial resolution with the thematic depth required for local decision-making. Furthermore, remote sensing classifiers frequently suffer from significant performance decay when transferred across space and time, a limitation that renders many research prototypes unsuitable for operational deployment. Our work provides a comprehensive blueprint for an open-source LULC system designed to address these twin challenges of thematic richness and spatio-temporal robustness. By detailing methodologies for dataset construction, hierarchical classification, and real-world deployment, this work moves beyond one-off snapshots toward a modular geospatial infrastructure adaptable to diverse planning applications.

To address the challenge of data scarcity and fragmented availability, we developed scalable curation strategies to generate large-scale, open datasets. We leveraged crowd-sourced information from OpenStreetMap (OSM) and government MIS portals, ensuring that annotations were sampled from across India’s diverse agro-ecological zones to embed spatial transferability into the models from the outset. For dynamic classes requiring ground-level precision, we utilized a mobile-based field collection infrastructure to facilitate on-ground interventions. This thesis details the data sanitization workflows designed to process these field observations and reflects on the key learnings gained from these on-ground deployment experiences.

Using these datasets, we engineered a hierarchical LULC classification framework that captures both static features, such as built-up areas, and dynamic intra-annual patterns, such as cropping intensity and water seasonality. Recognizing that a single "flat" classifier cannot effectively resolve all classes due to spectral confusion and varying data availability, the framework employs an ensemble of specialized sub-models. This architecture incorporates targeted methods for challenging tasks, such as crop-tree differentiation and unsupervised cropping intensity estimation, and is optimized for scalable deployment on Google Earth Engine. Spatio-temporal robustness is reinforced through spatially diverse sampling and post-classification temporal corrections that preserve potentially true landscape transformations while filtering mis-classifications.

Building upon this hierarchical foundation, we extend our LULC taxonomy to identify crop labels. We propose a trait-based crop classification framework that trains independent classifiers to predict key biophysical traits like crop periodicity, height, structure, canopy density, water consumption, and usage. The definition of these traits profiles (like tall grass crop) remain spatially invariant and we use these traits to deduce final crop labels in any given region. This framework improves classification across spectrally similar crop groups and under-represented crops in training data, delivering a practical solution to crop type mapping.

The utility of the proposed LULC solution is demonstrated through two distinct applications. First, we utilize a multi-temporal built-up series fused with road network data to track the evolution of major Indian cities, illustrating the framework's capacity for high-resolution urbanization monitoring. Second, as a foundational component of the CoRE Stack digital public infrastructure, the framework supports decentralized planning initiatives across 500+ villages and powers the Know Your Landscape(KYL) tool for data-driven landscape analysis. By bridging the gap between methodological innovation and operational utility, our thesis provides a foundational utility for monitoring India's rapidly transforming environment.