

Using Satellite Imagery and Geo-AI for Long-Term Urban Growth Monitoring

A Case Study of Riyadh

Prepared by **Waleed Ahmad AlSharif**

Supervisor **Dr. Khaled Eskaf**

MSc in Artificial Intelligence · 2026

Presentation Outline

01 Motivation & Problem Definition

02 Objectives & Contributions

03 Related Work & Positioning

04 Methodology

05 Results & Discussion

06 Comparison, Limitations & Conclusion

Motivation & Problem Definition

Riyadh is growing fast — but the tools to track it aren't keeping up.

- Rapid, fragmented urban expansion driven by Saudi Vision 2030 outpaces infrastructure and service planning.
- Built-up surfaces and bare desert soil share near-identical spectral signatures — a core arid-region challenge.
- Rocky, mountainous terrain intensifies this confusion (exposed rock vs. built structures).
- Current monitoring is slow, manual, and based on outdated data.

+14.0%

Riyadh built-up growth, 2018–2025 (this study)

~5%

Built-up ↔ bare-ground spectral confusion rate observed

Objectives & Contributions

OBJECTIVES

- Build a radiometrically consistent Sentinel-2 series (2018–2025).
- Classify land cover with a U-Net into 4 urbanization-relevant classes.
- Quantify city-wide built-up growth and its annual rate.
- Diagnose & correct a terrain-related spectral-confusion failure mode.
- Validate predictions independently in the field.
- Produce a transparent, trend-based 2030 growth projection.

KEY CONTRIBUTIONS

- Diagnosed a radiometric-inconsistency artifact from single-scene compositing — fixed via median compositing.
- Otsu/NDVI correction resolving systematic over-labeling of desert as vegetation (43–47% → 3.3–3.5%).
- Rule-based, terrain-aware correction for a previously undocumented built-up/rock confusion mode.
- Independent field validation distinct from the training-label source.
- Direct benchmark against the closest prior Riyadh study.

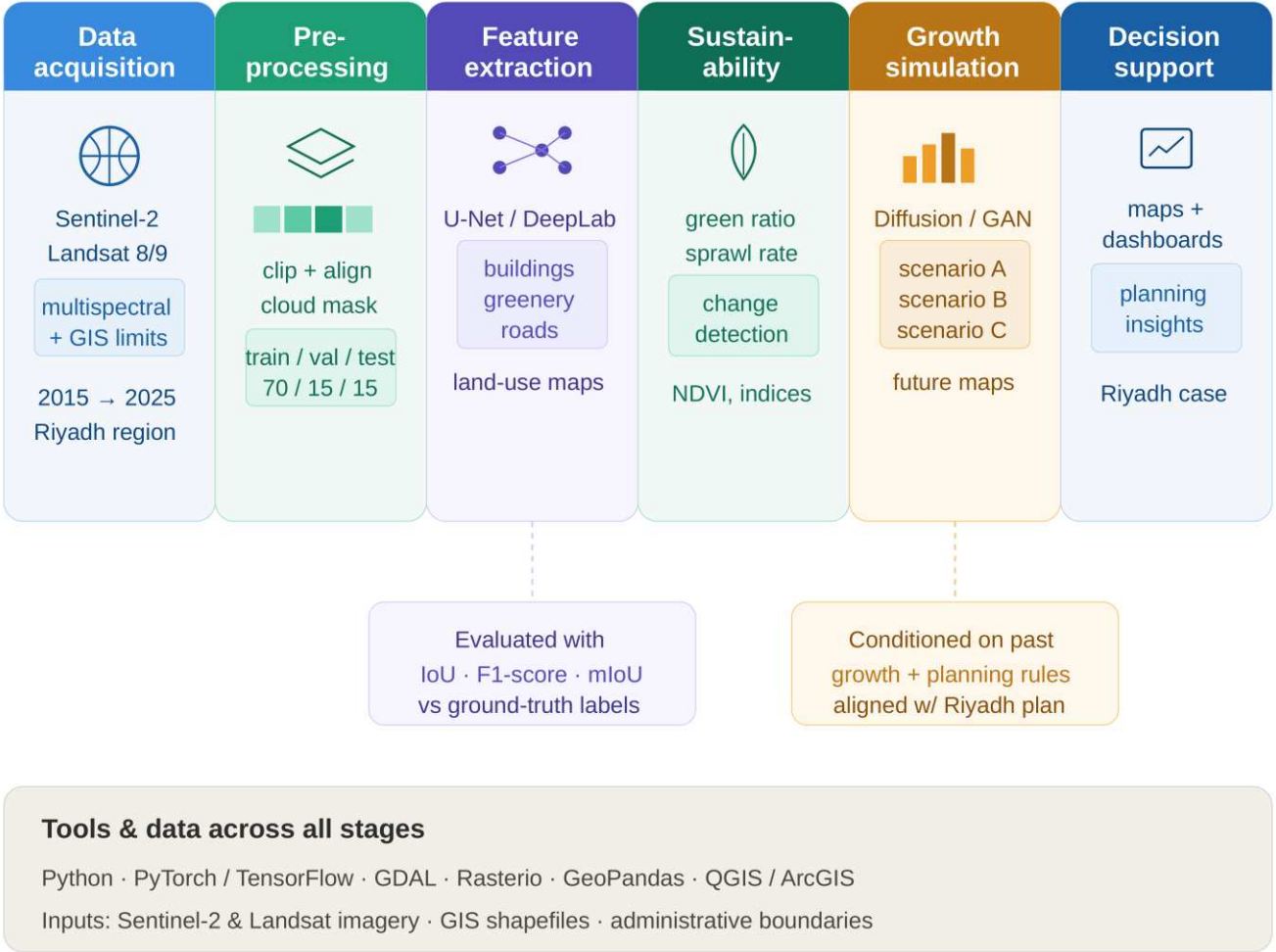
Related Work & Positioning

Closest prior study of Riyadh's urban expansion:

	Abdelsattar (2026)	This Thesis (2026)
Method	Random Forest (classical ML)	U-Net (deep segmentation)
Data / Resolution	Landsat, 30 m, 1990–2025	Sentinel-2, 10 m, 2018–2025
Classes	Binary (urban / non-urban)	4-class (incl. vegetation split)
Validation	600 pts/epoch + WorldCover	Field sample + independent rerun
Accuracy	97.7% OA	97.44% (harder, multi-class task)

Near-parity in accuracy on a substantially harder, higher-resolution, multi-class task suggests this pipeline is a genuine methodological advance — not a simpler restatement of prior work.

End-to-End Geo-AI Pipeline



Data & Feature Engineering

STUDY AREA

- Riyadh + periphery: 46.40–47.00°E, 24.40–25.10°N
- CRS: EPSG:32638 (UTM Zone 38N)
- Sentinel-2 L2A via Google Earth Engine, 2018–2025
- Per-pixel median composite (cloud-masked via SCL)

8-BAND FEATURE STACK

- 5 raw bands: Red, Green, Blue, NIR, SWIR
- NDVI — vegetation (Rouse et al., 1974)
- NDBI — built-up, SWIR-based (Zha et al., 2003)
- BSI — bare soil, 4-band composite (Rikimaru et al., 2002)

8

input channels per pixel

23,424

total 128×128 training patches

2018–25

8-year observation window

Two Key Methodological Fixes

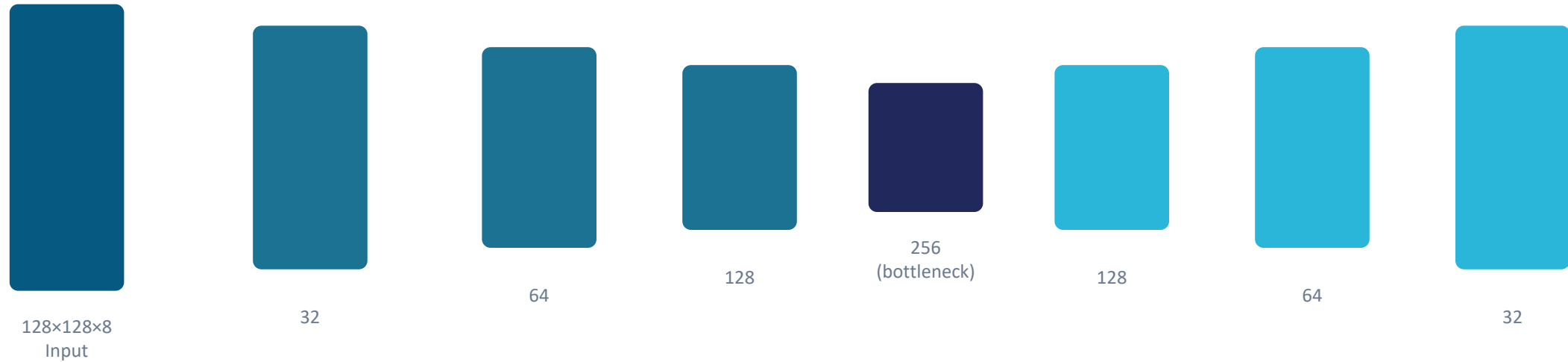
Fix 1 — Compositing Method

- Mosaicking (least-cloudy pixel) caused non-physical year-to-year area swings.
- Root cause: radiometric inconsistency between neighboring pixels from different scenes.
- Fix: per-pixel median composite across all cloud-masked scenes per year.
- Result: smooth, physically plausible growth trend.

Fix 2 — Label Noise Correction

- Raw Esri labels over-labeled bare desert as vegetation (43–47% of study area).
- Fix: Otsu-threshold NDVI correction, computed per year (Otsu, 1979).
- Threshold far more stable on median composites ($\sigma=0.014$) than mosaics (range 0.05–0.37).
- Result: realistic 3.3–3.5% vegetation share.

Model Architecture — U-Net



Output: 128×128×4 class-probability map

1.93M

trainable parameters

3 + 1

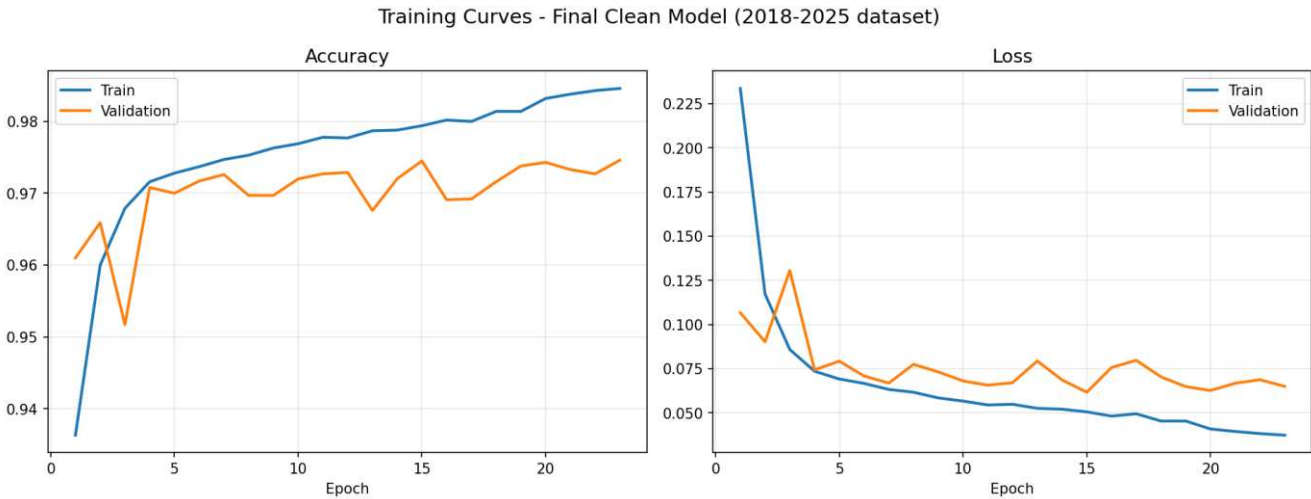
encoder/decoder stages + bottleneck

Skip Links

preserve fine spatial detail

Training Configuration & Convergence

Setting	Value
Optimizer	Adam (lr = 0.001)
Loss	Sparse categorical cross-entropy
Batch size	16
Callbacks	EarlyStopping · Checkpoint · ReduceLROnPlateau
Split	Train 2018–22 · Val 2023 · Test 2024–25
Converged	Epoch 23 of 40 (max)



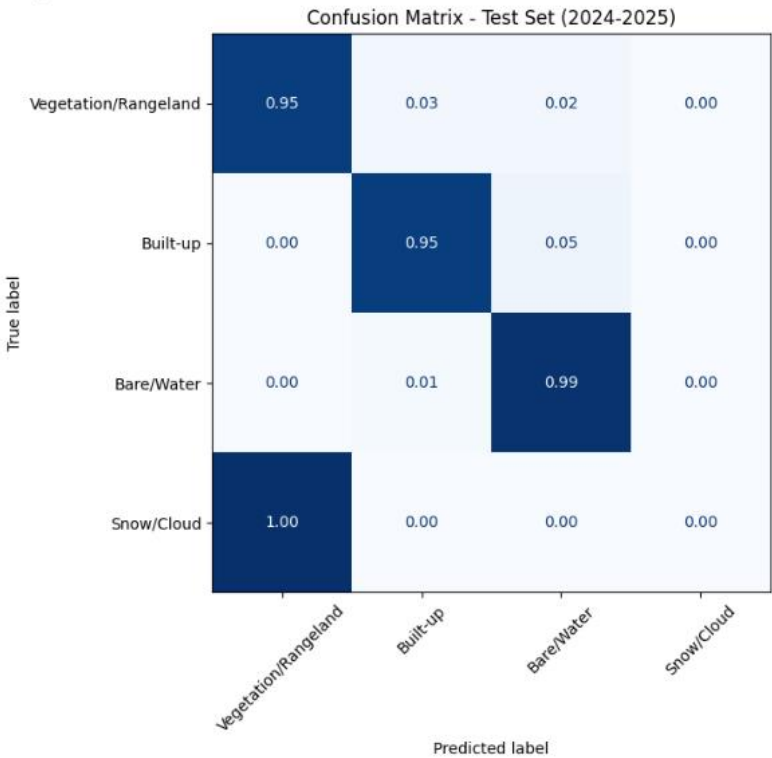
Training & validation accuracy / loss curves

Segmentation Performance

Class	IoU	F1
Vegetation/Rangeland	0.920	0.959
Built-up	0.917	0.957
Bare Ground/Water	0.965	0.982

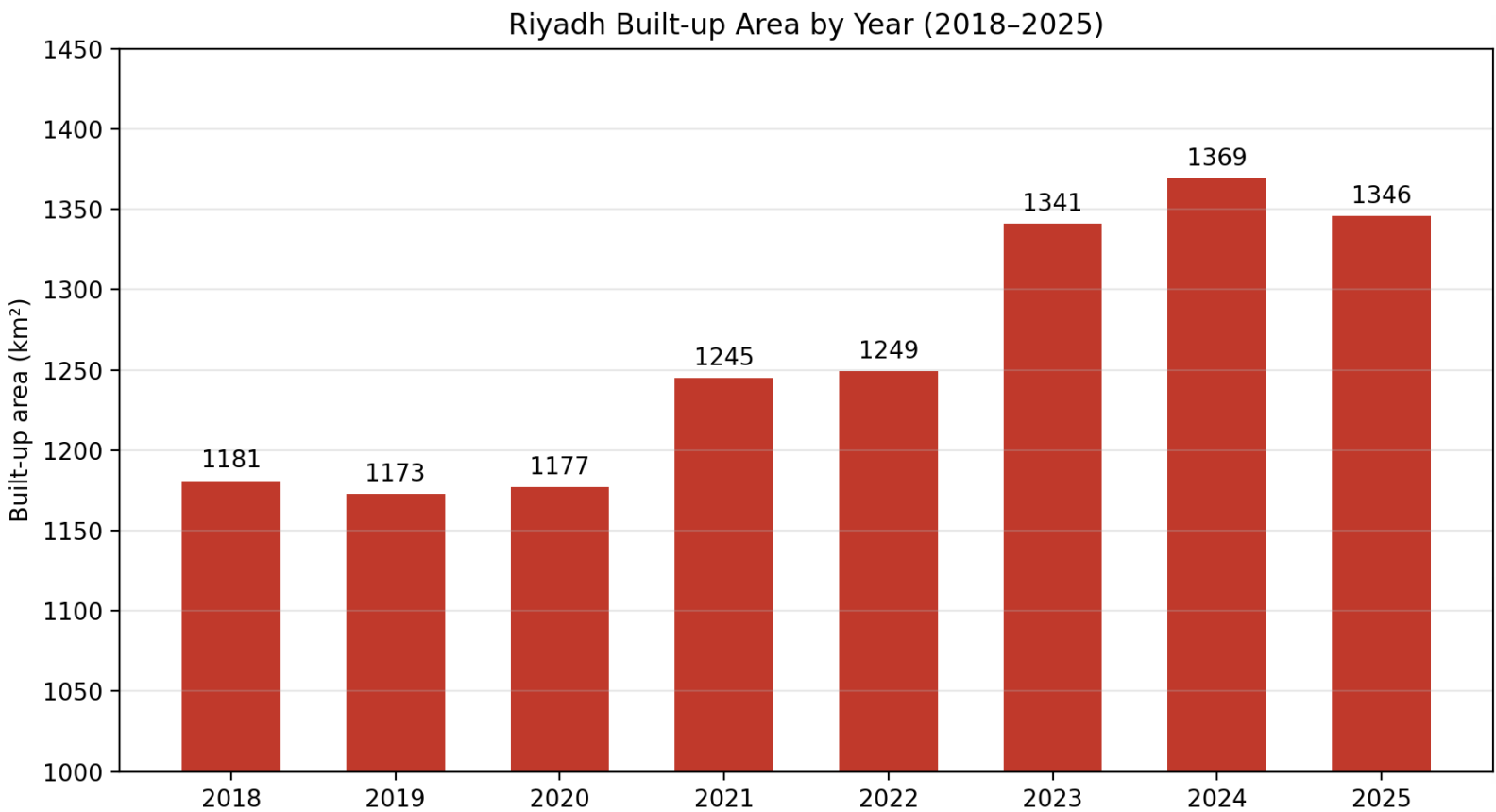
97.44%

overall pixel accuracy on held-out 2024–2025 test set



Normalized confusion matrix (2024–25 test set)

City-Wide Urban Growth, 2018–2025



1,181 → 1,346

built-up area, km² (2018 → 2025)

+14.0%

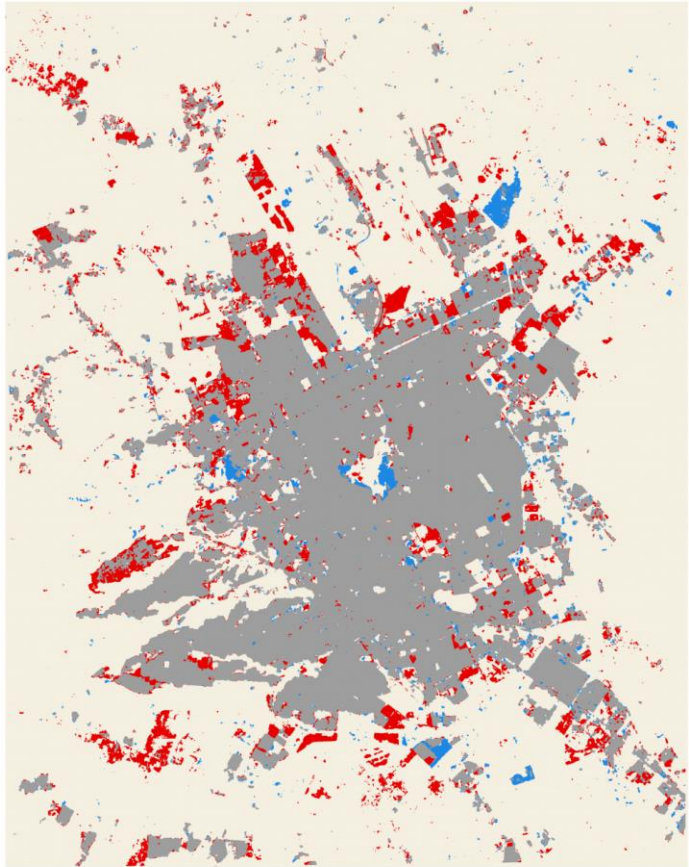
total growth over 8 years

~23.6 km²

average annual growth

Spatial Growth Pattern

Riyadh Urban Growth 2018 -> 2025 (New development highlighted)



Non-built Existing built-up New urban growth Built -> non-built

KEY OBSERVATIONS

- Clear radial, fringe-led expansion pattern from the 2018 urban core.
- New development (red) concentrated at the periphery and in vacant interstitial parcels.
- Sparse built-to-non-built pixels (blue) near the core — consistent with residual classification noise, not large-scale demolition.
- Largest single expansion phase: 2023–2024, concentrated north and east.

Terrain-Aware Correction

A failure mode invisible in the fixed test set — surfaced only when tested on genuinely new imagery in mountainous terrain.



Wadi Laban: live image → raw classification → corrected classification

6.2%

corrected at Wadi Hanifah (primary site)

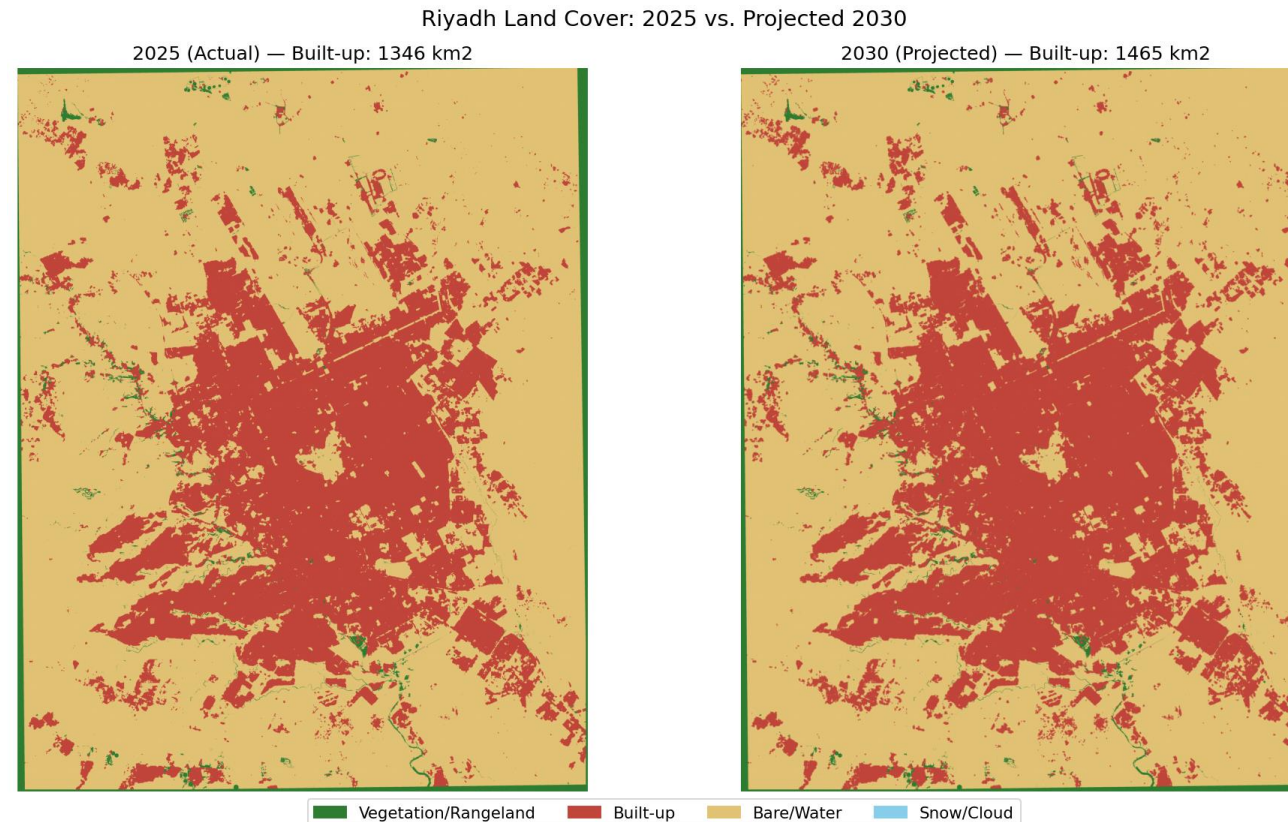
2.4%

corrected at Wadi Laban (secondary,
incomplete)

No Retrain

rule: slope > 5° OR BSI > 0.15

Preliminary Future-Growth Projection



≈1,465 km²

projected built-up area by 2030 (+8.8% vs. 2025)

Trend-based

deterministic simulation — not a validated forecast

Independent Validation & Reproducibility

83.3%

field-validation agreement (5 of 6 stratified sample points, checked against basemap imagery — not the training reference)

97.57%

accuracy of an independently reconstructed re-run of the full pipeline (vs. 97.44% primary)

$\pm 2 \text{ km}^2$

typical year-to-year agreement between the two independent runs' area estimates

Why this matters:

- Field validation gives a reference-independent accuracy signal, since training labels themselves are only 75–85% accurate.
- A second, independently-built pipeline reproducing closely matching results is strong evidence the findings reflect a genuine property of the method — not an artifact of one run.

A Validated, Reproducible Geo-AI Framework for Riyadh

Limitations: reference-label noise (75–85% source accuracy) · small field-validation sample (n=6) · terrain correction proven at 1 of 2 sites · projection is single-scenario, no CI

- 97.44% accuracy · $\text{IoU} \geq 0.91$ on all substantive classes
- +14.0% built-up growth quantified over 8 years, city-wide
- Terrain-aware correction for a newly diagnosed failure mode
- Independent field validation + independently reproduced results
- Benchmarked directly against the closest prior Riyadh study

FUTURE WORK

- Alternative backbones (ResNet-UNet, DeepLabV3+)
- Native 20 m SWIR; systematic terrain-threshold tuning
- Expand field validation to ≥ 50 stratified points
- Generative model for probabilistic 2030+ scenarios
- Add SAR data to reduce built-up/bare-ground confusion



Thank You

Questions & Discussion