

# AI-Based Urban Fabric Classification using Satellite Imagery

Ondřej Maxian, Kamil Kovář, Martin Metzler, Martin Fleischmann, Dani Arribas-Bel

## Table of contents

<b>1</b>	<b>AI-Based Urban Fabric Classification using Satellite Imagery</b>	<b>2</b>
1.1	Background . . . . .	2
1.2	Data and Methods . . . . .	2
1.2.1	Satellite Imagery . . . . .	2
1.2.2	Urban Fabric Classes . . . . .	3
1.3	Model Design . . . . .	3
1.3.1	Classification vs. Segmentation . . . . .	3
1.3.2	Scale . . . . .	4
1.3.3	Sampling . . . . .	4
1.4	Experimental Approaches . . . . .	4
1.5	Model Evaluation . . . . .	4
1.6	AI Model Results . . . . .	5
1.7	Final Model . . . . .	6
1.7.1	Addressing Class Imbalance . . . . .	6
1.7.2	Classification Schemes . . . . .	8
1.7.3	Model Performance . . . . .	9
1.8	Temporal Analysis . . . . .	9
1.8.1	Diversity Analysis (Shannon Index) . . . . .	10
1.8.2	Spatial Patterns of Change . . . . .	10
<b>2</b>	<b>Discussion</b>	<b>14</b>
2.1	Key Findings . . . . .	14
2.2	Lessons Learned . . . . .	14
2.3	Potential Research Directions . . . . .	15
<b>3</b>	<b>Software and Example Datasets</b>	<b>15</b>
3.1	Software: AI Method for Urban Fabric Classification and Morphometric Characterization . . . . .	16

3.2	Example Datasets Generated During Verification Exercises . . . . .	16
3.3	London . . . . .	17
3.4	Liverpool . . . . .	19
<b>References</b>		<b>20</b>

# 1 AI-Based Urban Fabric Classification using Satellite Imagery

## 1.1 Background

Urban areas represent complex and dynamic environments, making traditional analysis challenging due to their diverse characteristics. Remote sensing technologies, particularly satellite imagery, provide powerful tools for understanding and characterizing urban areas at scale. As part of the EuroFab project, we employed AI modeling techniques using Sentinel-2 satellite imagery to classify and analyze urban fabric across Great Britain. The aim was to generate detailed insights into urban spatial structures and track temporal changes spanning from 2016 to 2021.

## 1.2 Data and Methods

### 1.2.1 Satellite Imagery

We utilized two different sets of satellite images derived from Sentinel-2 imagery:

1. **GHS-composite-S2 R2020A dataset** (Corbane et al. 2020): Covering January 2017 to December 2018, this dataset offers cloud-free, consistent, high-quality RGB (red, green, blue band) composite imagery at a resolution of 10 meters per pixel. We used these images of Great Britain (GB) to train our prediction model.
2. **Annual Sentinel-2 composites**: Created for each year from 2016 to 2021, these composites include only cloud-free RGB images from their specific year. We acquired these using the Google Earth Engine API through automated Python scripts (as described in [GEE pipeline](#)).

Sentinel-2 data is particularly suitable for our application due to several advantages:

- Consistent and openly available global coverage, ideal for cross-regional comparisons
- Temporal flexibility through composites from multiple time points, allowing analysis of urban patterns over time
- Resolution of 10 meters per pixel, providing sufficient detail to distinguish urban features at the neighborhood scale
- Compatibility with many pre-trained geospatial AI models, facilitating transfer learning

### **1.2.2 Urban Fabric Classes**

The urban fabric predictions from our AI model are based on labels from the Spatial Signatures Framework (Arribas-Bel and Fleischmann 2022; Fleischmann and Arribas-Bel 2022), which provides a typology of British urban environments characterized by both form (physical structure) and function (usage). This framework captures the complexity of urban areas, offering insights into how different spaces look and operate.

While our primary focus is on urban fabric classification centered on form—an approach that may be simpler since form is visible in imagery—we currently use the Spatial Signatures Framework as a proxy due to its conceptual alignment with our goals for urban characterization. This decision was influenced by the limited project duration and the fact that the morphotope-based classification only became available in Q4 2024.

To ensure progress in developing the AI model, we use separate input classifications for the morphological and AI components of the project. The expectation is that the two classifications are conceptually similar, allowing the model architecture developed for Spatial Signatures to later support building an equally or more precise predictive model for the morphotope-based classification.

## **1.3 Model Design**

### **1.3.1 Classification vs. Segmentation**

In satellite image analysis, classification and segmentation address spatial labeling at different levels of granularity:

- **Classification:** Assigns a single urban fabric type to an image tile, offering generalized insights into dominant urban characteristics without the need for pixel-perfect alignment.
- **Segmentation:** Assigns urban fabric types at a pixel level, providing detailed boundary-specific mapping but facing challenges due to ambiguous class boundaries in urban environments.

In our study, the label dataset does not always correspond directly to identifiable features in the imagery, making classification a potentially more suitable approach as it generalizes each tile's dominant land cover type without requiring exact pixel alignment. However, we explored both approaches to evaluate how each method performs given the scale and nature of the dataset.

### 1.3.2 Scale

We tested multiple scales for our analysis: - Segmentation at the pixel level (preprocessed into  $224 \times 224$  pixel tiles for compatibility with downstream models) - Classification at scales of  $56 \times 56$  pixels ( $560 \times 560$  meters on the ground) and  $25 \times 25$  pixels ( $250 \times 250$  meters on the ground)

### 1.3.3 Sampling

We evaluated two sampling approaches to assess their impact on spatial generalization and F1-score performance:

1. **Random sampling:** Ensures diversity and captures localized patterns, but risks spatial leakage, potentially inflating performance metrics.
2. **H3 resolution 3 regional sampling:** Reduces spatial leakage and offers a more realistic evaluation of generalization but can suffer from unfair penalization due to the heterogeneity of regions.

## 1.4 Experimental Approaches

We structured experiments around three distinct modeling approaches:

- **Approach A:** Baseline model employing SatlasPretrain embeddings coupled with an XGBoost classifier.
- **Approach B:** Fine-tuned three geospatial foundation models (Satlas, Clay, Prithvi) specifically for segmentation tasks.
- **Approach C:** Fine-tuned the Clay model for classification tasks based on its superior segmentation performance.

## 1.5 Model Evaluation

Multiple metrics were utilized to comprehensively evaluate model performance:

- **Weighted Accuracy (Macro):** Reflects overall predictive accuracy adjusted for class frequency.
- **Weighted F1 Score:** Balances precision and recall, particularly crucial for handling imbalanced datasets.

## 1.6 AI Model Results

Our experimental results demonstrated that classification approaches outperformed segmentation methods. This was not only in terms of accuracy but also in the shapes that were predicted. While the  $56 \times 56$  pixel scale ( $560 \times 560$ m) yielded marginally better statistical performance, we determined that the  $25 \times 25$  pixel scale ( $250 \times 250$ m) offered a more appropriate resolution for policy applications and urban planning contexts. A key finding was that incorporating regional context through H3 spatial indexing significantly enhanced predictive performance across all configurations. Models combining embeddings with H3 contextual information provided the most balanced performance, substantially outperforming other approaches on all metrics. Notably, when comparing sampling strategies, we observed that H3 regional sampling showed slightly lower performance metrics than random sampling, suggesting potential spatial leakage in the random sampling approach that may artificially inflate performance metrics.

The results in the table below represent either pixel-level values or tile level values in brackets.

<b>Approach</b>	<b>Scale</b>	<b>Sampling</b>	<b>Clas/seg</b>	<b>Regional info</b>	<b>Global Acc</b>	<b>Macro Acc</b>	<b>F1 (macro)</b>
C (Clay)	1x1	random	segmentation		<b>0.73</b>	<b>0.31</b>	<b>0.30</b> <b>(0.58)</b>
C (Satlas)	1x1	random	segmentation		0.57		(0.41)
C (Prithvi)	1x1	random	segmentation		0.62		(0.42)
A (embed- dings)	25x25	random	classification		(0.73)	(0.31)	(0.30)
A (embed- dings)	25x25	random	classification		<b>(0.81)</b>	<b>(0.46)</b>	<b>(0.53)</b>
A (embed- dings)	25x25	random	classification	lat/lon	(0.89)	(0.71)	(0.78)
A (embed- dings)	25x25	random		lat/lon	(0.91)	(0.78)	(0.83)
A (embed- dings)	25x25	H3 res 3 (55,743)	classification	H3 res 5 (lat/lon)	(0.58)	(0.15)	(0.15)
A (embed- dings)	25x25	H3 res 5 (2,125)	classification	H3 res 5 (lat/lon)	(0.65)	(0.20)	(0.21)

<b>Approach</b>	<b>Scale</b>	<b>Sampling</b>	<b>Clas/seg</b>	<b>Regional info</b>	<b>Global Acc</b>	<b>Macro Acc</b>	<b>F1 (macro)</b>
A (embed- dings)	25x25	H3 res 6 (335)	classification	H3 res 5 (lat/lon)	<b>(0.72)</b>	<b>(0.29)</b>	<b>(0.32)</b>
A (embed- dings)	56x56	random	classification		0.76 (0.66)	0.22 (0.13)	0.23
A (embed- dings)	56x56	random	classification	H3 res 5 (cat)	<b>0.87</b> <b>(0.82)</b>	<b>0.42</b> <b>(0.35)</b>	<b>0.45</b>
A (embed- dings)	56x56	random	classification	H3 res 5 (lat/lon)	<b>0.87</b> <b>(0.81)</b>	0.39 (0.31)	0.42
A (embed- dings)	56x56	random	regression	H3 res 5 (ordinal)	0.80 (0.80)	0.26 (0.26)	0.26
B (Clay)	56x56	random	classification		0.59 (0.68)	0.09	0.12

Based on these insights, our final model implements a Satlas embedding architecture operating at the  $25 \times 25$  pixel scale (250×250m resolution) with integrated H3 resolution 5 regional context, optimizing the balance between spatial detail and classification accuracy while ensuring robust generalization to diverse urban environments.

## 1.7 Final Model

### 1.7.1 Addressing Class Imbalance

Significant class imbalance, particularly for urban classes, posed a considerable challenge in our modeling. To mitigate this, we implemented a sliding-window augmentation strategy for classes comprising less than 10% of the dataset (all classes except *Countryside agriculture* and *Wild countryside*).

This sliding-window approach systematically shifted the sampling window horizontally and vertically by increments of 50, 100, 150, and 200 meters, significantly increasing the volume of available training data for underrepresented classes.

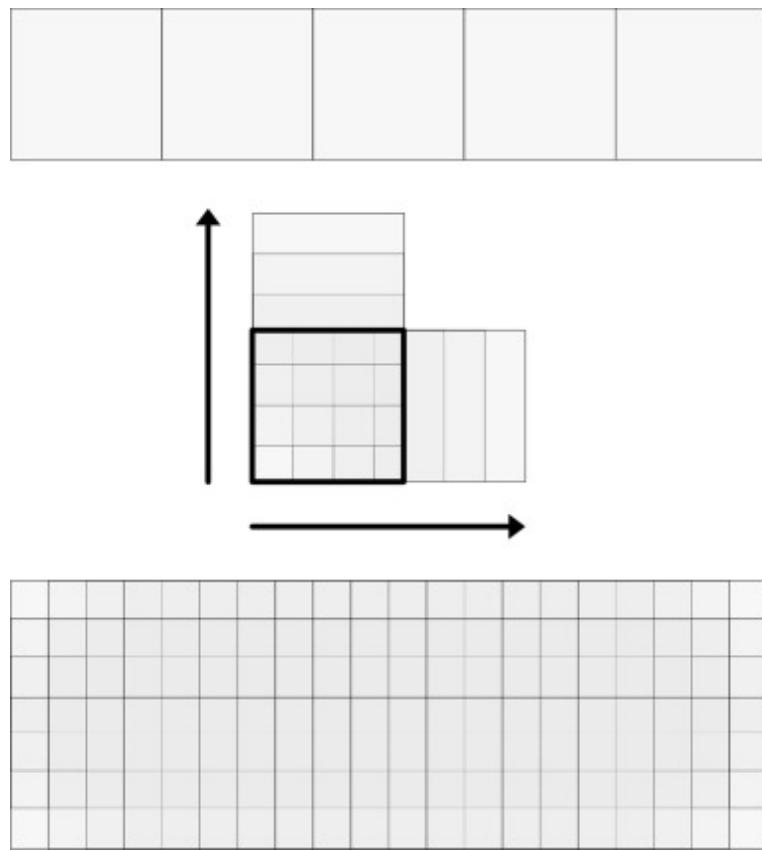
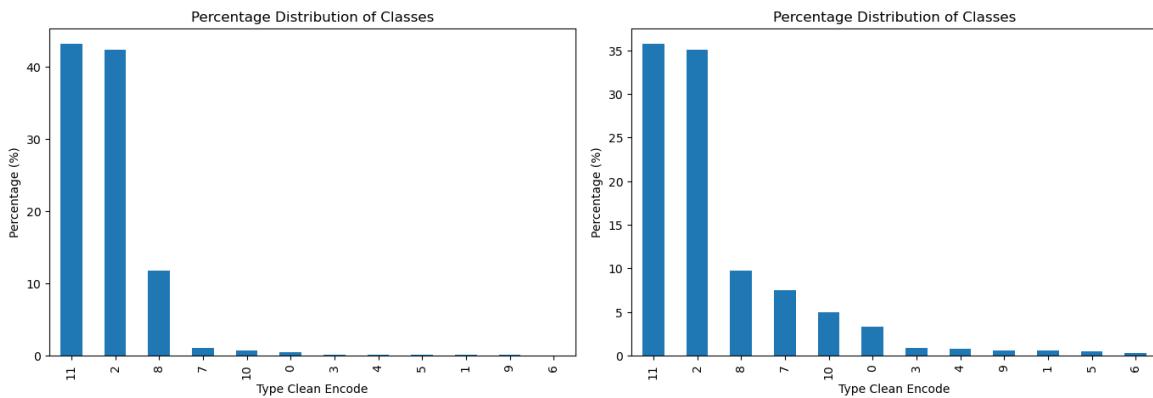


Figure 1: Sliding Window Augmentation

The images below show class distributions before (left) and after (right) augmentation:



The following table summarizes augmentation results, highlighting the substantial increase in training samples for previously underrepresented classes:

Class	Before Augmentation	After Augmentation
Accessible suburbia	15,054	129,620
Connected residential neighbourhoods	2,567	21,021
Countryside agriculture	1,367,999	1,367,999
Dense residential neighbourhoods	4,299	34,507
Dense urban neighbourhoods	3,636	31,657
Disconnected suburbia	2,644	20,113
Gridded residential quarters	1,518	12,849
Open sprawl	33,910	292,884
Urban buffer	381,283	381,283
Urbanity	2,495	21,929
Warehouse/Park land	21,282	195,105
Wild countryside	1,395,048	1,395,048

### 1.7.2 Classification Schemes

We used two versions of our dataset:

1. **12-class scheme:** Maintained all original Spatial Signatures classes:

```
class_labels = {
    'Accessible suburbia': 0,
    'Connected residential neighbourhoods': 1,
    'Countryside agriculture': 2,
    'Dense residential neighbourhoods': 3,
    'Dense urban neighbourhoods': 4,
    'Disconnected suburbia': 5,
    'Gridded residential quarters': 6,
    'Open sprawl': 7,
    'Urban buffer': 8,
    'Urbanity': 9,
    'Warehouse/Park land': 10,
    'Wild countryside': 11
}
```

2. **Simplified 7-class scheme:** Created by reclustering underlying data from the Spatial Signatures Framework using K-means clustering (K=7):

```

class_labels_k7 = {
    'Countryside agriculture': 0,
    'Open sprawl': 1,
    'Compact suburbia': 2,
    'Urban': 3,
    'Urban buffer': 4,
    'Warehouse/Park land': 5,
    'Wild countryside': 6
}

```

### 1.7.3 Model Performance

The final XGBoost classifier was trained using the augmented dataset and evaluated using three metrics: micro accuracy, macro accuracy (where every class has the same weighting), and macro F1 score. We validated the model using 5-fold spatial cross-validation at H3 resolution 6, ensuring an 80/20 training-testing split.

The table below summarizes classifier performance for both classification schemes (7 and 12 classes) and two spatial context scenarios (with and without H3 resolution):

Classes (K)	Spatial Context	Accuracy	Macro Accuracy	Macro F1 Score
7	None	0.4924	0.3856	0.3389
7	H3 (res 5)	0.6959	0.5713	<b>0.5221</b>
12	None	0.4617	0.2666	0.2127
12	H3 (res 5)	0.6654	0.4328	0.3654

Including spatial context (H3 resolution) notably improved classification accuracy and F1 scores, demonstrating the importance of spatial context in predicting urban fabric classes. As anticipated, the model with fewer classes (7) performed better compared to the one with 12 classes.

## 1.8 Temporal Analysis

We used the trained XGBoost classifier to make predictions across the years 2016 to 2021. The overall overlap between the initial year (2016) and final year (2021) remained high at 88%, confirming that Spatial Signatures classes remained relatively stable across the study period. However, minor variations may indicate either genuine change or model uncertainty.

Year-to-year class stability rates:

- 2016 → 2017: 88%

- 2017 → 2018: 88%
- 2018 → 2019: 86%
- 2019 → 2020: 86%
- 2020 → 2021: 88%

### 1.8.1 Diversity Analysis (Shannon Index)

We assessed changes in urban fabric diversity using the Shannon Index across the studied years. The results are summarized in the table below, indicating slight fluctuations, with the most notable increase in diversity occurring in 2019:

Year	Shannon Index
2016	1.868
2017	1.863
2018	1.868
2019	2.007
2020	1.873
2021	1.872

The marked increase in the Shannon Index in 2019 suggests an increase in class diversity during that year, followed by a subsequent return to previous levels. This could indicate some differences in the imagery caused by weather conditions or sensor characteristics specific to 2019.

### 1.8.2 Spatial Patterns of Change

To make predictions across years, we used the trained model and fed in all GB tiles to get predictions for the 12 classes.

Spatial analysis identified areas across England with frequent class transitions (map below), particularly around major urban centers and suburban zones. These frequent transitions may either represent genuine urban transformations or result from classifier uncertainty, especially in ambiguous zones between visually similar Spatial Signatures classes.

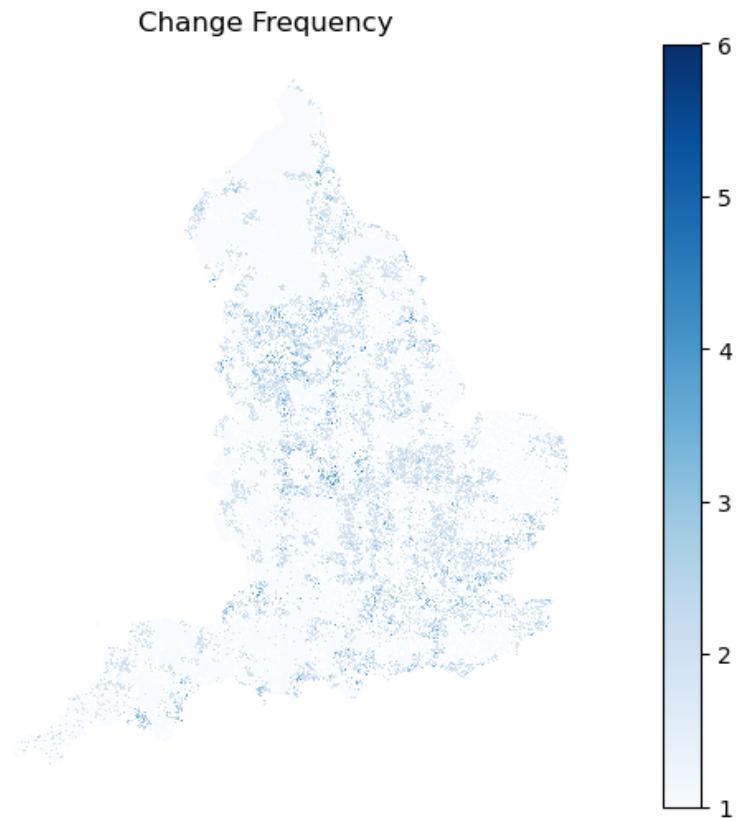


Figure 2: Frequency Map

Class-specific analysis further highlighted particular Spatial Signatures types prone to transitions. The figure below shows which classes experienced frequent transitions:

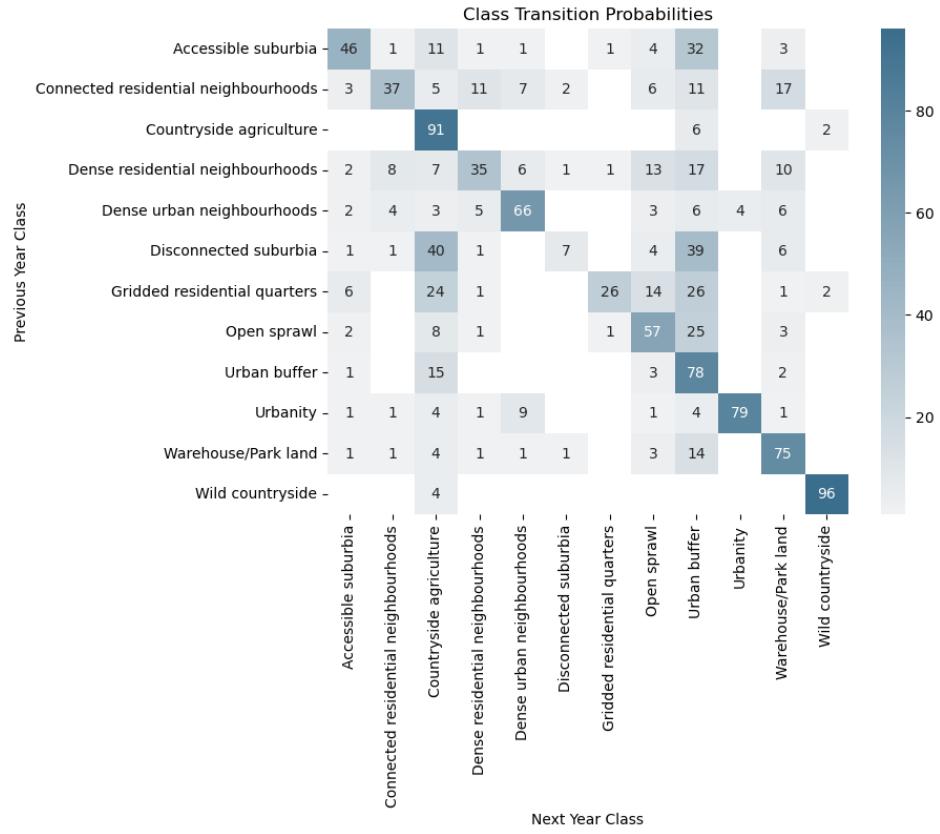


Figure 3: Class Transitions

To better interpret these transitions, we calculated transition probabilities and organized them into a structured confusion matrix. This matrix clearly shows the urban fabric classes most likely to interchange over the studied period:

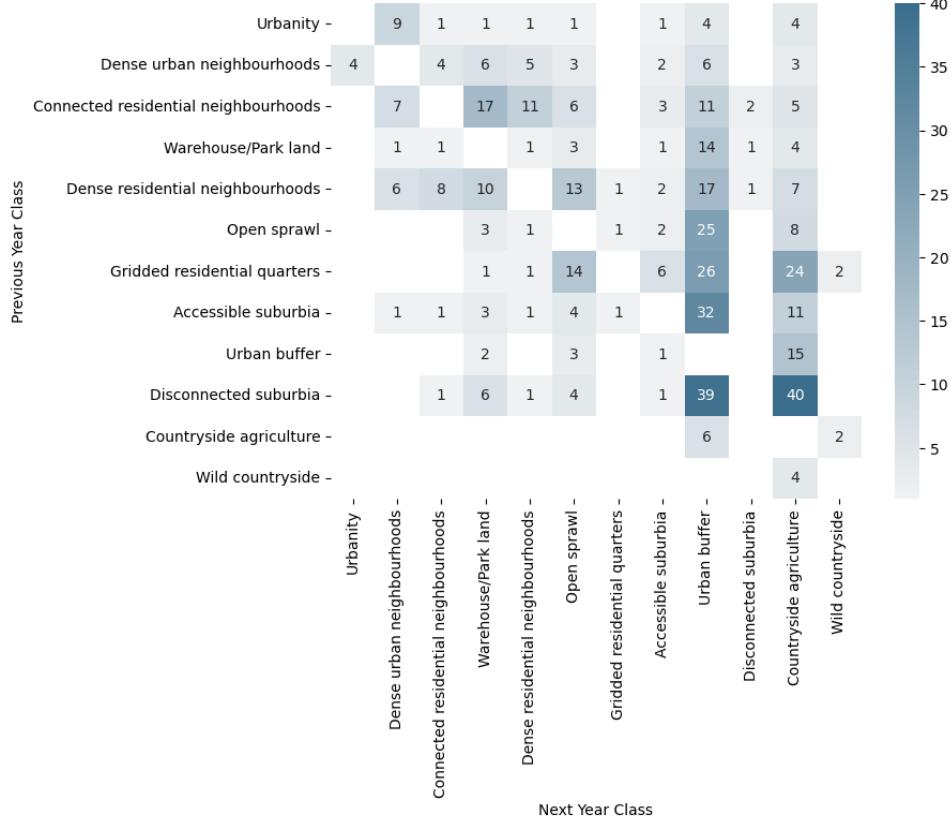


Figure 4: ‘Reasonable’ confusion matrix

Lastly, we analyzed the inverse-probability-based distances of urban fabric classes over time. Shorter distances represent a higher likelihood of transitioning or changing classes from year to year, whereas longer distances indicate greater stability:

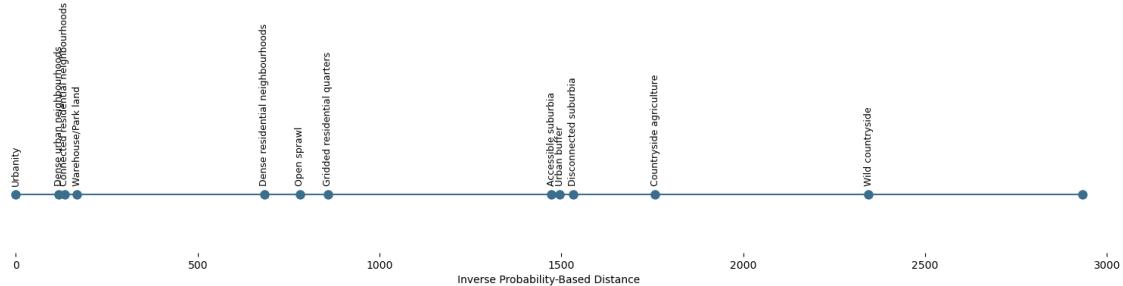


Figure 5: Probability of change across classes

This analysis shows urban fabric classes such as *Urbanity*, *Warehouse/Park land*, *Dense urban*

*neighbourhoods*, and *Connected residential neighbourhoods* have shorter inverse distances, indicating higher levels of dynamic change or redevelopment. Conversely, classes such as *Wild countryside* and *Countryside agriculture* have longer distances, suggesting greater temporal stability. Although these results align with expectations about urban and rural dynamics, they may also reflect the classifier's varying uncertainty across these visually distinct environments.

## 2 Discussion

### 2.1 Key Findings

The analysis demonstrates that urban fabric classifications exhibit distinct temporal and spatial dynamics, reflecting varying levels of stability and diversity over time. Notably, urban classes such as *Dense urban neighbourhoods* and *Connected residential neighbourhoods* displayed higher probabilities of transition, indicating active urban transformation. Similarly, suburban classes, such as *Accessible suburbia*, *Disconnected suburbia*, and *Urban buffer*, also show a higher probability of change between these classes. Conversely, rural classes showed significant stability. It's important to note that these observed changes may be more related to classifier uncertainty than actual environmental changes.

### 2.2 Lessons Learned

- **Scale:** Urban fabric classes, like spatial signatures, have fuzzy boundaries. Pixel-level classifications provide the highest possible spatial resolution, which is beneficial for detailed analysis. However, pixels often lack clear visual cues indicating exact boundaries between classes, making pixel-level predictions challenging for the model. Patch-level classification, although lower in spatial resolution depending on patch size, provided clearer visual context and resulted in better overall performance.
- **Embeddings vs. Fine-Tuned Foundation Model:** Fine-tuning foundation models involves significant complexity and requires careful design decisions. In our case, the limited number of training examples was insufficient to achieve noticeable improvements through fine-tuning. The effort required for fine-tuning did not outweigh the simpler alternative of using off-the-shelf embeddings.
- **Regional Trends:** Including regional contextual information substantially improved the classifier's accuracy. Nonetheless, we found it essential that embeddings themselves already capture enough visual detail for accurate classification, ensuring that predictions remain robust even without regional context (and simply do not just rely on the spatial information to make predictions).

- **Data Augmentation:** The sliding window augmentation approach effectively addressed class imbalances, significantly improving model performance by increasing representation of previously underrepresented urban fabric classes. This approach could boost classifier performance by an additional 10-20% in terms of accuracy.

### 2.3 Potential Research Directions

There are several promising directions for future research:

- **Handling Misclassifications:** Misclassifications typically occur between visually similar urban fabric classes, indicating inherent uncertainty in predictions. Incorporating prediction probabilities into a secondary model could help address this issue. By explicitly using probability scores from the initial classification as input for a refinement model — as previously shown by Fleischmann and Arribas-Bel<sup>1</sup> — we could better distinguish between ambiguous cases. This approach may “smooth” predictions, reducing noise and improving overall classification accuracy. Future work should explore how prediction confidence scores can be systematically utilized, either by employing spatial smoothing algorithms or by applying secondary machine learning models trained specifically to correct uncertain predictions.
- **Generalizability Testing:** Evaluating the generalizability of this methodological framework is crucial for its wider applicability. Future research should test this modeling approach in different European regions, assessing whether the chosen methods, including data preprocessing, augmentation strategies, spatial embeddings, and classifier architectures, perform consistently outside Great Britain. This would involve exploring variations in urban form and regional urban planning contexts across Europe. Understanding these factors will help identify potential adjustments needed to ensure reliable predictions when extending the model beyond the original study area.

## 3 Software and Example Datasets

All analyses presented here are supported by openly accessible software hosted on [GitHub](#). The AI prediction pipeline, including preprocessing, embedding generation, and prediction of spatial signatures, is fully documented and accessible in the EO repository.

---

<sup>1</sup>Fleischmann and Arribas-Bel, 2024. Decoding (urban) form and function using spatially explicit deep learning. *Computers, Environment and Urban Systems*, 31, p.105737.

### **3.1 Software: AI Method for Urban Fabric Classification and Morphometric Characterization**

All the work supporting this analysis can be found on GitHub. The main prediction pipeline, which includes data preprocessing, embedding creation, and spatial signature prediction, can be used as follows:

```
# Run the pipeline
pipeline.spatial_sig_prediction(
  geo_path= "../spatial_signatures/eo/data/example/london_25_25_grid_clipped.geojson", ## V
  vrt_file= "../satellite_demoland/data/mosaic_cube/vrt_allbands/2017_combined.vrt", ## Vr
  xgb_weights = "../spatial_signatures/classifier/k12_h5_slided_gb_xgb_model.bin", ## Model
  model_weights = "../satellite_demoland/models/satlas/weights/satlas-model-v1-lowres.pth"
  output_path= "../vjgo8416-demoland/spatial_signatures/eo/data/predictions/test_london_h6
  h3_resolution=5 ## h3 resolution to be added to analysis (spatial context)
)
```

More details and documentation on how to run the pipeline can be found in the example on the [EO repository](#).

### **3.2 Example Datasets Generated During Verification Exercises**

The final datacube including predictions for the years 2016 to 2021 for 7 and 12 classes can be found in the repository.

Here are some example visualizations showing London and Liverpool from the dataset:

### 3.3 London

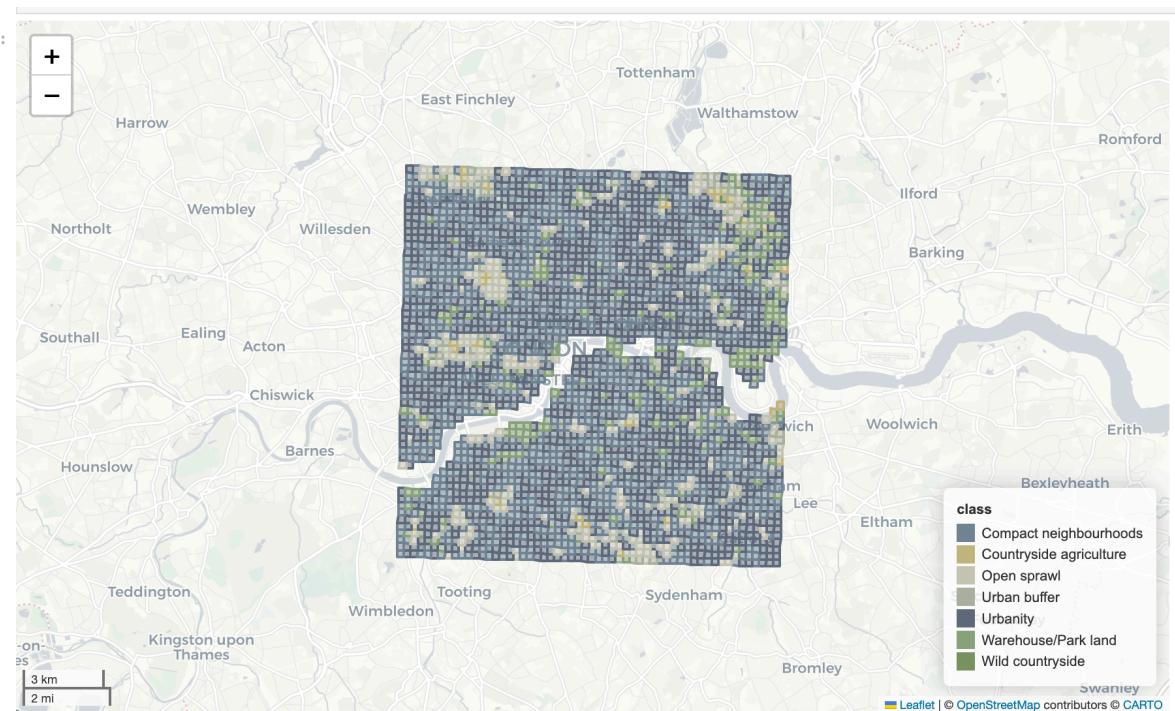


Figure 6: London, 7 classes

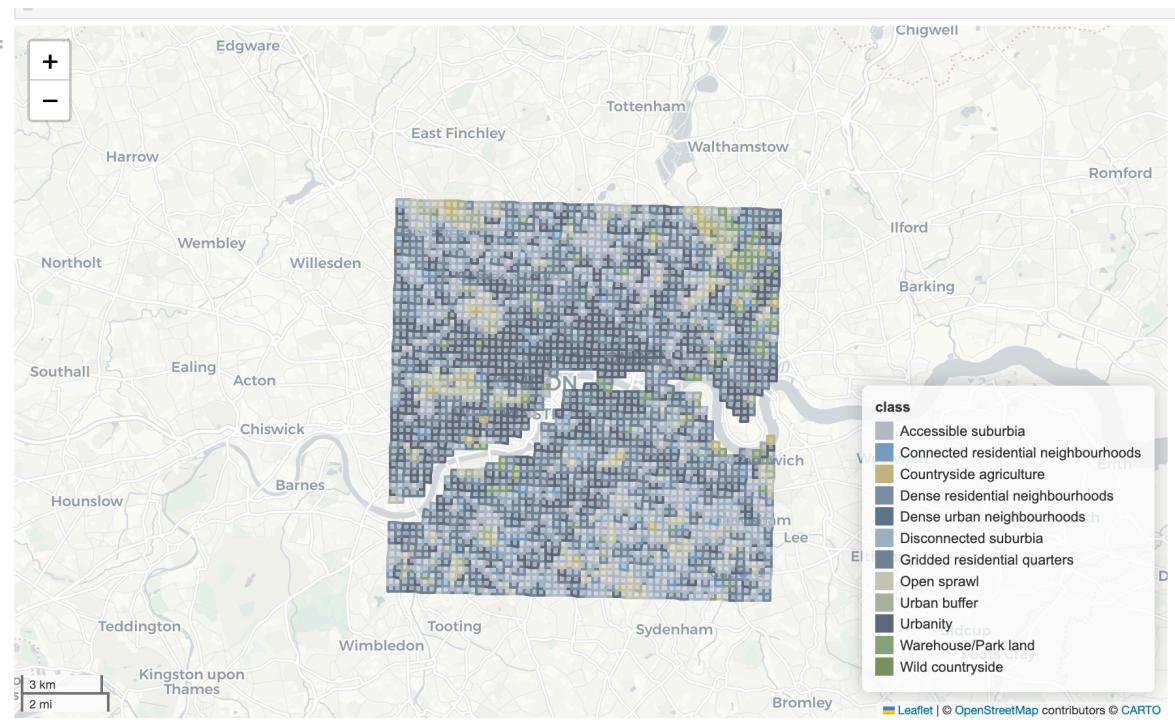


Figure 7: London, 12 classes

### 3.4 Liverpool

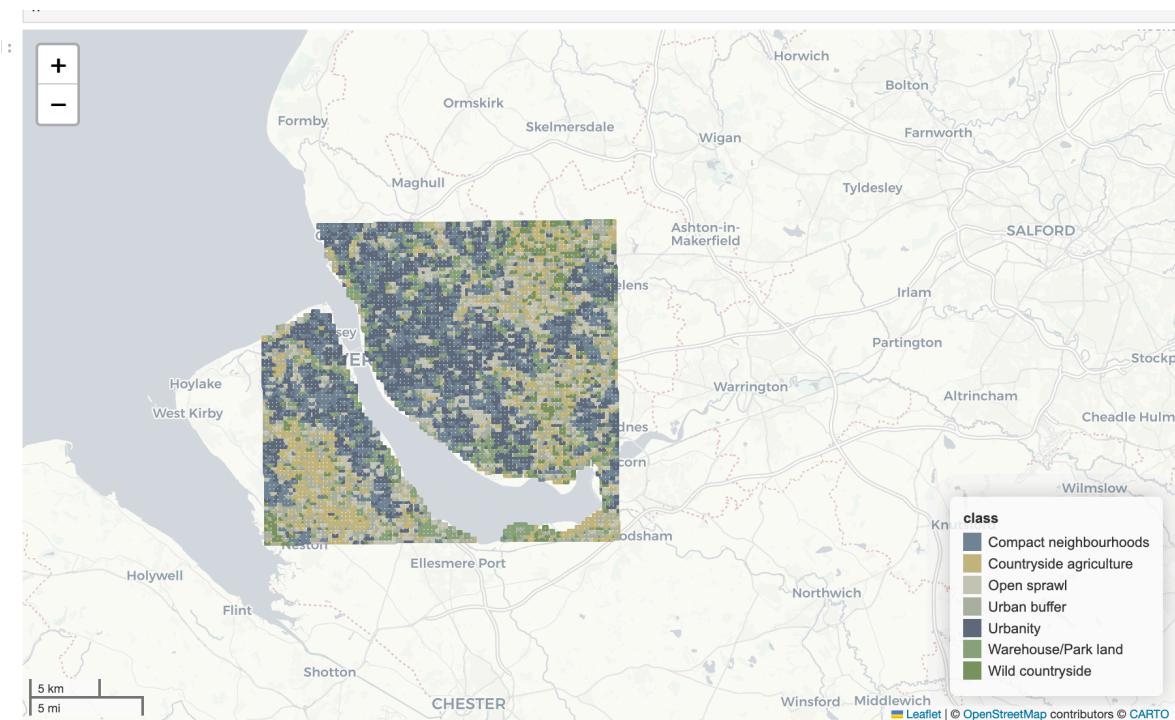


Figure 8: Liverpool, 7 classes

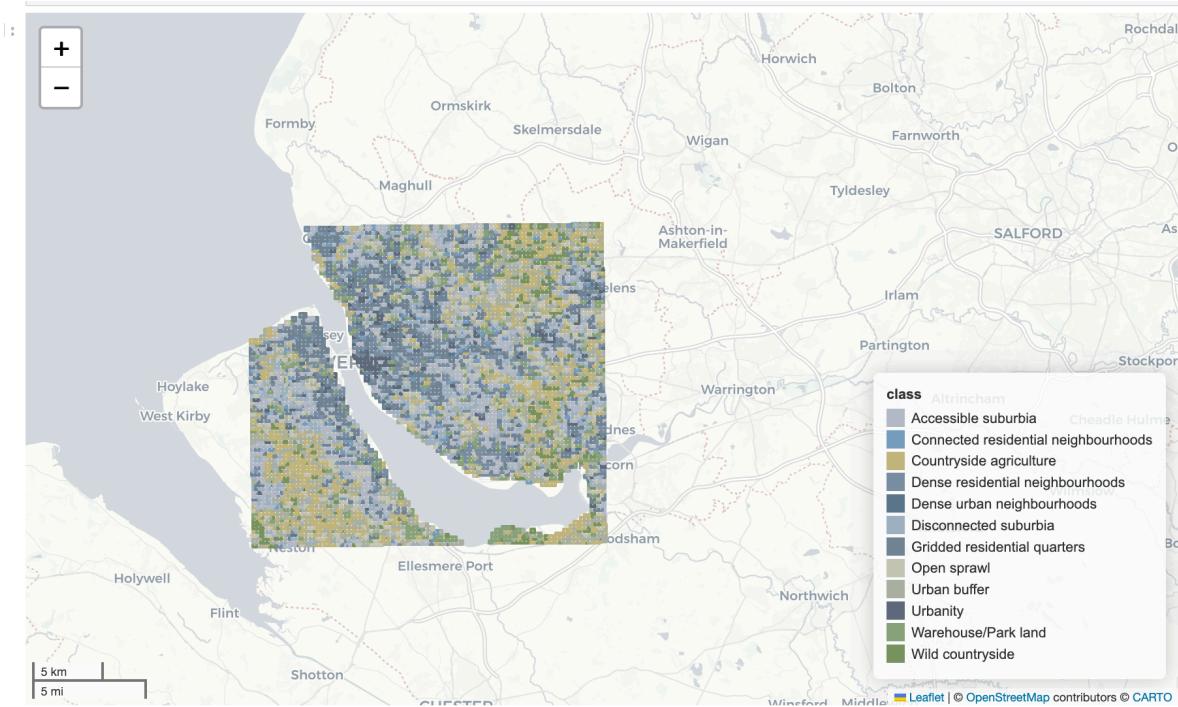


Figure 9: Liverpool, 12 classes

## References

- Arribas-Bel, Daniel, and Martin Fleischmann. 2022. "Spatial Signatures-Understanding (Urban) Spaces Through Form and Function." *Habitat International* 128: 102641.
- Corbane, C., P. Politis, P. Kempeneers, D. Simonetti, P. Soille, A. Burger, M. Pesaresi, F. Sabo, V. Syrris, and T. Kemper. 2020. "A Global Cloud Free Pixel- Based Image Composite from Sentinel-2 Data." *Data in Brief* 31 (August): 105737. <https://doi.org/10.1016/j.dib.2020.105737>.
- Fleischmann, Martin, and Daniel Arribas-Bel. 2022. "Geographical Characterisation of British Urban Form and Function Using the Spatial Signatures Framework." *Scientific Data* 9 (1): 546.