Utilizing the Clay Foundation Model and Sentinel-2 Imagery for Urban Growth Monitoring in Johnston County, North Carolina

Benton Tripp

Abstract

This study addressed the challenge of infrequent urban imperviousness data updates by integrating Sentinel-2 multispectral satellite imagery with the Clay Foundation Model, an open-source deep learning framework for Earth observation. Focusing on Johnston County, North Carolina—a region experiencing rapid urban growth—the research aimed to approximate urban density more frequently than the National Land Cover Database (NLCD) updates, which occur approximately every five years. Sentinel-2 imagery, accessed via the Microsoft Planetary Computer and AWS Earth Search APIs, provided high-resolution, multitemporal data suitable for regular monitoring. The Clay Foundation Model generated spatial embeddings from the imagery, capturing detailed spectral information without the need for additional feature engineering or external indices. Deep learning methodologies, including convolutional and recurrent neural networks, were applied to predict urban imperviousness percentages as a proxy for urbanization. This proof-of-concept study demonstrated a scalable, data-efficient framework for monitoring urban growth, offering insights into methodological advancements and highlighting the potential of foundation models to enhance sustainable urban planning by filling gaps left by traditional datasets.

Introduction

Literature Review

Urban growth fundamentally reshapes landscapes, ecosystems, and socioeconomic structures, making it a crucial area of study within environmental and urban planning. The expansion of impervious surfaces, such as roads and buildings, serves as a key indicator of urbanization and impacts ecosystems by increasing surface runoff, reducing groundwater recharge, and altering local climates. Understanding and tracking these changes allows researchers and policymakers to manage the environmental consequences of urban growth and devise strategies to minimize negative impacts on biodiversity, water cycles, and air quality (Goetzke et al., 2008).

Remote sensing has emerged as an essential technology for monitoring urban growth, offering extensive and consistent data across time and space. Multitemporal satellite imagery, particularly from sources like Sentinel-2, enables precise tracking of land cover changes over large geographic areas and prolonged periods, making it ideal for detecting patterns of urban expansion (Ayush et al., 2021; Zhu et al., 2017). This technology facilitates not only the visualization of urban sprawl but also the quantitative analysis of changes in land use and land cover. By providing high-resolution, time-sequenced imagery, remote sensing allows for dynamic monitoring of urbanization processes, yielding insights critical for sustainable urban planning and resource management.

Deep learning architectures have revolutionized remote sensing, with Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers each playing pivotal roles. CNNs excel at extracting spatial features from images and are widely used for analyzing high-resolution satellite imagery, enabling accurate image classification and object detection. RNNs and their variants, such as Long Short-Term Memory networks (LSTMs), are advantageous for capturing temporal patterns within data sequences, facilitating the processing of time-series satellite data. Transformers, originally designed for natural language processing, have been adapted for spatiotemporal analysis, introducing new possibilities in remote sensing applications (He et al., 2022; Jean et al., 2019).

In supervised learning, labeled data are used to train deep learning models to recognize specific patterns, making it suitable for tasks like land cover classification, object detection, and semantic segmentation

(Dionelis et al., 2024; Zhu et al., 2017). These tasks require precise pixel-level predictions, which supervised deep learning models, such as CNNs, are adept at performing. However, this approach is often limited by the availability of labeled data, which can be costly and time-consuming to produce. This limitation has led to increased interest in unsupervised learning and representation learning methods, where models like autoencoders and contrastive learning frameworks learn features from unlabeled data. By enabling models to automatically learn useful representations, these methods make it feasible to work with larger datasets without extensive annotation, enhancing the scope and applicability of deep learning in remote sensing (Ayush et al., 2021; Jean et al., 2019).

Foundation models have transformed fields like natural language processing, where models such as BERT and GPT-3 have set new standards in text processing by pretraining on large datasets and then fine-tuning for specific applications. These models provide a flexible framework for handling diverse tasks with minimal additional training, representing a shift from task-specific models to more generalized architectures. By capturing universal patterns within massive datasets, foundation models enable the efficient transfer of knowledge across related tasks, reducing the need for task-specific training data (Mai et al., 2022).

In Earth Observation (EO) and geospatial artificial intelligence (AI), foundation models demonstrate significant potential for applications such as land cover classification, object detection, and change detection. Unlike traditional models that are trained for single tasks, foundation models can generalize across a wide range of EO applications, learning to recognize spatial and temporal features across multiple datasets (Dionelis et al., 2024). This generalizability allows these models to be highly adaptable to new tasks, even with limited labeled data, making them particularly valuable in remote sensing where annotation resources are often scarce. Their adaptability, combined with label efficiency and robust feature extraction capabilities, positions foundation models as powerful tools for advancing EO and geospatial AI (Ayush et al., 2021; He et al., 2022).

Motivation for the Research

Despite advancements in deep learning and the introduction of foundation models, limited research has focused specifically on using these models to map urban imperviousness. A significant challenge in monitoring urban growth is the infrequent updates of authoritative land cover datasets. For instance, the National Land Cover Database (NLCD) updates its urban imperviousness data approximately every five years, with recent updates in 2016 and 2021 (U.S. Geological Survey, 2021). In rapidly growing areas like Johnston County, this temporal resolution may not be sufficient to capture dynamic changes in urban density. There is a pressing need for more frequent assessments to inform timely urban planning and environmental management decisions.

Utilizing satellite imagery with frequent revisit times, such as the Sentinel-2 five-day cycle, presents an opportunity to approximate urban density more regularly, assuming the availability of models capable of extracting relevant information from the imagery. Foundation models like the Clay Foundation Model offer the potential to address this need by generating robust embeddings from multispectral satellite data, which can be used to predict urban imperviousness with higher temporal resolution. This capability can fill the gap left by infrequent updates of traditional land cover datasets, providing a scalable and accurate means to monitor urban growth.

Research Question and Objectives

The primary objective of this study was to evaluate the potential of foundation models in enhancing the temporal resolution of urban imperviousness mapping. Specifically, the research aimed to leverage the Clay Foundation Model, in conjunction with Sentinel-2 imagery, to predict urban imperviousness in Johnston County, North Carolina, with greater frequency than traditional datasets allow. By using the Clay model's pretrained embeddings extracted from Sentinel-2 images, the study sought to develop and assess different approaches for predicting urban imperviousness at high spatial and temporal resolutions, without incorporating additional feature engineering or external indices.

This proof-of-concept study intended to demonstrate the value of foundation models in filling the gap left by infrequent land cover data updates, offering a method to approximate urban density more frequently and accurately. By doing so, the research contributed to the growing body of work employing foundation models for geospatial applications and highlighted their potential for improving sustainable urban planning and resource management in rapidly developing regions (Clay Foundation, 2023; Goetzke et al., 2008).

Study Site

Johnston County is located in the eastern part of North Carolina, United States, covering approximately 2,050 square kilometers. It lies between latitudes 35.3°N and 35.8°N, and longitudes 78.0°W and 78.6°W. The county is part of the rapidly expanding Raleigh-Durham-Chapel Hill metropolitan area, making it a pertinent case study for urban growth analysis.

For spatial analysis, all coordinates were defined in the Universal Transverse Mercator (UTM) coordinate system, specifically UTM Zone 17N (EPSG:32617), with units in meters. This coordinate system provided spatial accuracy and consistency in measurements across the study area and was compatible with the coordinate systems used by datasets such as the Sentinel-2 imagery.

Figure 1: Johnston County, North Carolina (obtained from Johnston County Economic Development)

Data Overview

Several datasets were utilized to define spatial boundaries, analyze land cover, and estimate urban density within Johnston County, North Carolina. The Johnston County GIS data provided the official county boundary, defining the study's spatial extent (Johnston County Department of GIS, 2011). Sentinel-2 satellite imagery from the European Space Agency (ESA) was employed as the exclusive source of multispectral data. Specifically, the MSI Level-1C Top of Atmosphere (TOA) Reflectance Product, Collection 1, offering 10-meter spatial resolution and including 13 spectral bands suitable for land cover analysis, was used. Sentinel-2's revisit interval of approximately five days allowed for frequent monitoring over time (Copernicus Sentinel-2, 2021). Data access was facilitated through the `pystac_client` python library, drawing from the Microsoft Planetary Computer STAC API (Microsoft Open Source et al., 2022) and the Earth Search AWS STAC API, adhering to the SpatioTemporal Asset Catalog (STAC) API specification (STAC Specification, 2023). Urban imperviousness raster data from the National Land Cover Database (NLCD) provided 30-meter resolution raster data on land cover and impervious surfaces, which were used to calculate urban density percentages within spatial patches (U.S. Geological Survey, 2021).

Table 1 summarizes the datasets used in the study along with their key attributes, including coordinate reference systems (CRS), spatial extents, resolutions, formats, and any transformations applied during preprocessing.

Table 1: Summary of datasets and their attributes used in the study.

Dataset	**Sentinel-2 Satellite Imagery**	**Urban Imperviousness**	**Johnston County Boundary**
Original CRS	EPSG:32617 (UTM Zone 17N)	Albers Conical Equal Area (EPSG:4326)	EPSG:32617 (UTM Zone 17N)
Original Spatial Extent	Queried individually within Johnston County Region	minx: -2493045.0 miny: 177285.0 maxx: 2342655.0 maxy: 3310005.0	minx: 707750.13 miny: 3904356.29 maxx: 765959.87 maxy: 3967194.75
Original Resolution / Format	10x10 meters / Multispectral GeoTIFF	30x30 meters / GeoTIFF	Vector
Additional Data Attributes	Bands: o B02 (Blue) o B03 (Green) o B04 (Red) o B08 (NIR) Cloud Cover: < 1% Dates: o Approximately triannual spread (where available) o 2016-2024 o 20 dates total	-	-
Transformation Details	Used as primary spatial reference; no CRS transformation needed. Cropped to the 600x600-meter tiles covering Johnston County boundary.	Reprojected to EPSG:32617 and cropped to Johnston County. Resampled to 200m resolution using bilinear interpolation.	Converted to EPSG:32617 to match Sentinel-2 and NLCD data. Used to define study area bounds, and divided into 600x600-meter tiles for consistent spatial units.

Methods

The data processing workflow began by defining the spatial extent of Johnston County, North Carolina, using official boundary data from the Johnston County GIS. This boundary data were reprojected to UTM Zone 17N (EPSG:32617) to align with the coordinate reference system of the Sentinel-2 imagery. A grid of 600×600-meter tiles was generated to cover the county boundary, providing consistent spatial units for analysis. Each tile was stored as a polygon within a geospatial dataset to facilitate tracking and associating data files with each specific area of interest. These tiles served as the fundamental units for subsequent data extraction, processing, and urban density analysis across the study area.

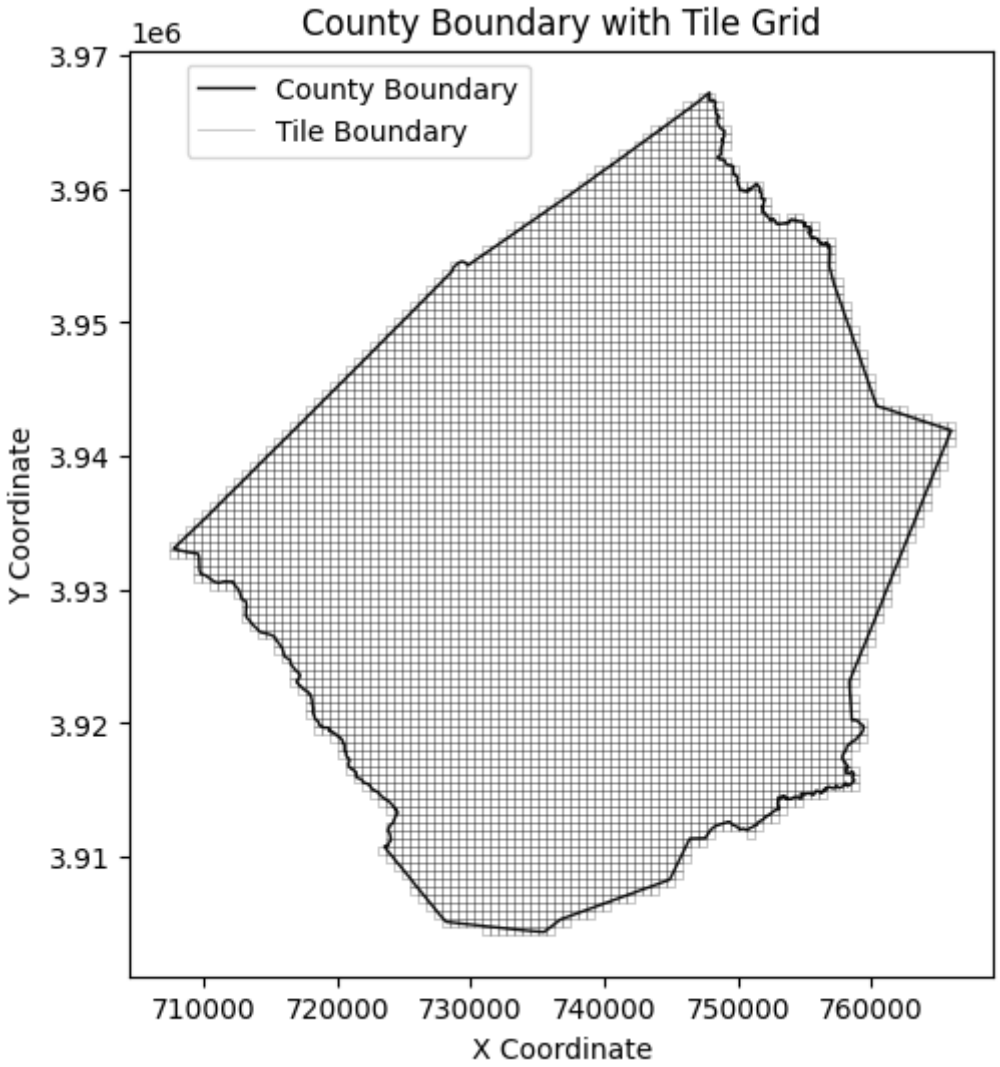

Figure 2: Map displaying the boundary of Johnston County, North Carolina, overlaid with a 600x600-meter tile grid.

Urban imperviousness data from the National Land Cover Database (NLCD) were prepared to align with the spatial framework defined by the 600×600-meter tiles across the study area. Initially, the data were clipped to the study area and reprojected to EPSG:32617 for consistency with other datasets. To maintain a balance between data granularity and computational efficiency, the urban imperviousness data was then resampled from a 30-meter to a 200-meter resolution using bilinear interpolation. This resampling method provided smooth transitions between pixel values, preserving key spatial patterns while adapting the data to a coarser resolution suitable for modeling.

Selecting a 200-meter resolution was specifically determined based on the modeling objectives. Aggregating the data to the full 600×600-meter extent of each tile would result in the loss of fine-scale details essential for accurately capturing urban density variations. Conversely, retaining the original finer resolution would require significantly more computational resources. By selecting 200 meters, the resampling yielded a manageable 3×3 matrix for each tile, which was used as the target for the deep learning model. This level of precision struck an optimal balance, providing sufficient spatial detail for robust urban growth analysis while remaining computationally feasible for large-scale processing.

Urban imperviousness data were extracted for each tile and associated with the corresponding grid tile, enabling localized assessments of urban density within the study area. This approach facilitated the integration of urban imperviousness data with the Sentinel-2 multispectral imagery, creating a structured dataset for urban growth analysis across Johnston County.

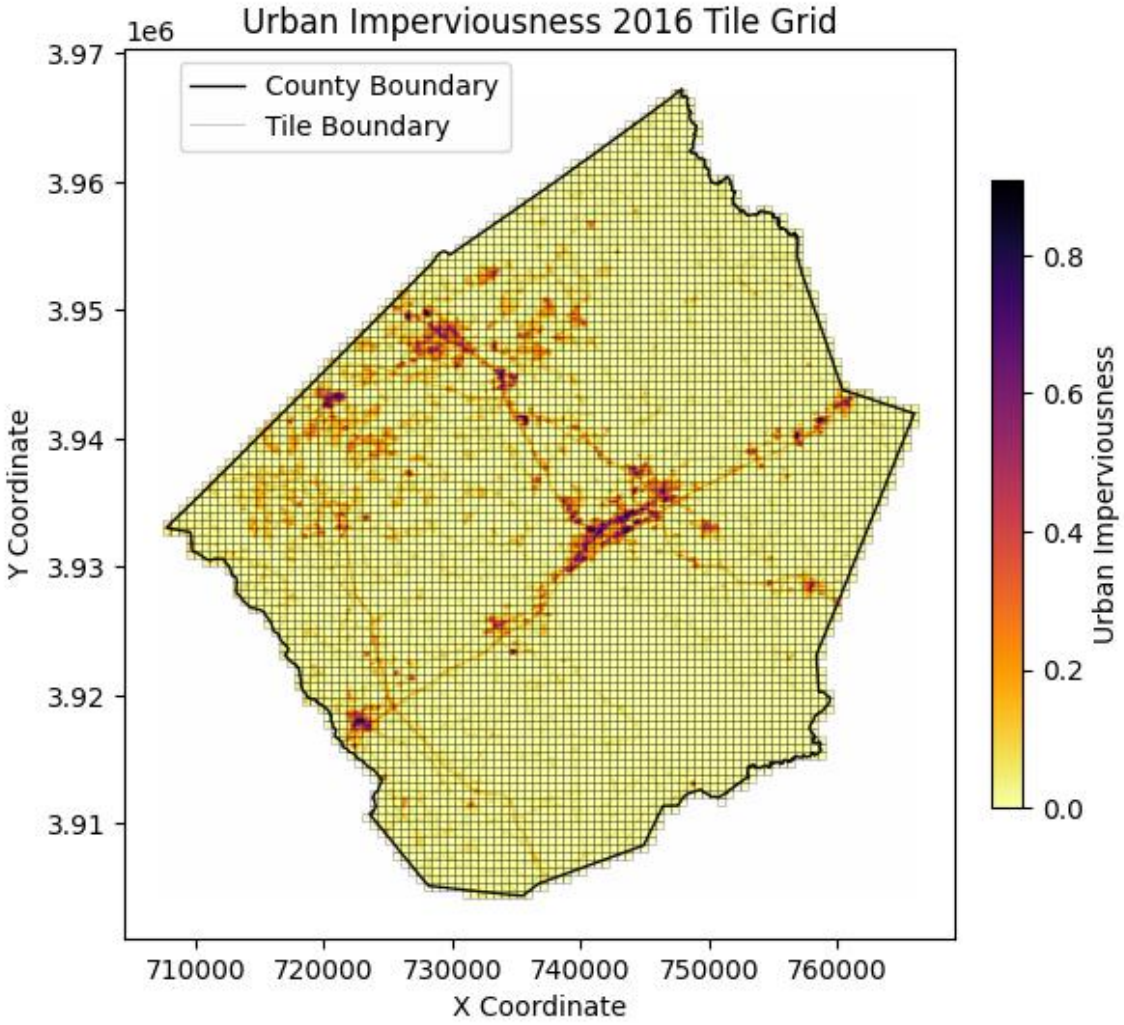

Figure 3: Urban imperviousness across Johnston County, North Carolina, in 2016, with a 600x600-meter tile grid overlay and the county boundary.

Sentinel-2 imagery was retrieved from the Microsoft Planetary Computer and AWS Earth Search APIs, providing multispectral data for the study area. Each image captured detailed spectral information relevant to urban growth analysis. Specifically, four spectral bands—Blue (B02), Green (B03), Red (B04), and Near Infrared (NIR, B08)—were downloaded at a 10-meter spatial resolution to maximize detail. Each raster was organized into 600×600-meter tiles in the EPSG:32617 coordinate reference system, facilitating manageable processing units for spatial analysis. Images were chosen based on quality criteria, primarily focusing on scenes with less than 1% cloud cover to ensure minimal interference in the analysis. The goal was to collect a consistent set of images over time, representing different seasonal conditions that could influence vegetation and other land features. While an ideal quarterly sampling interval was targeted from 2016 to 2024, certain images were unavailable due to various constraints, such as temporary data gaps or weather-related omissions. This resulted in an adjusted schedule with images retrieved approximately three times per year, yielding a total of 20 images covering the study period.

To optimize the temporal distribution of the selected dates, a filtering and date-selection process was implemented. This process involved buffering around unavailable dates to avoid excessive temporal clustering and ensured that the selected images represented the best possible seasonal coverage for each year. Additionally, adjustments were made to avoid redundant or less informative dates, yielding a final dataset with a balanced temporal spread across the study period. Each 600×600-meter tile, aligned to EPSG:32617, was thus associated with urban imperviousness data and multitemporal spectral imagery, creating a robust, high-resolution dataset for analyzing urban growth. This selection process enabled the

creation of a multispectral dataset that accounted for seasonal variability and minimized data redundancy, supporting the extraction of spatial features necessary for the deep learning model's urban growth analysis.

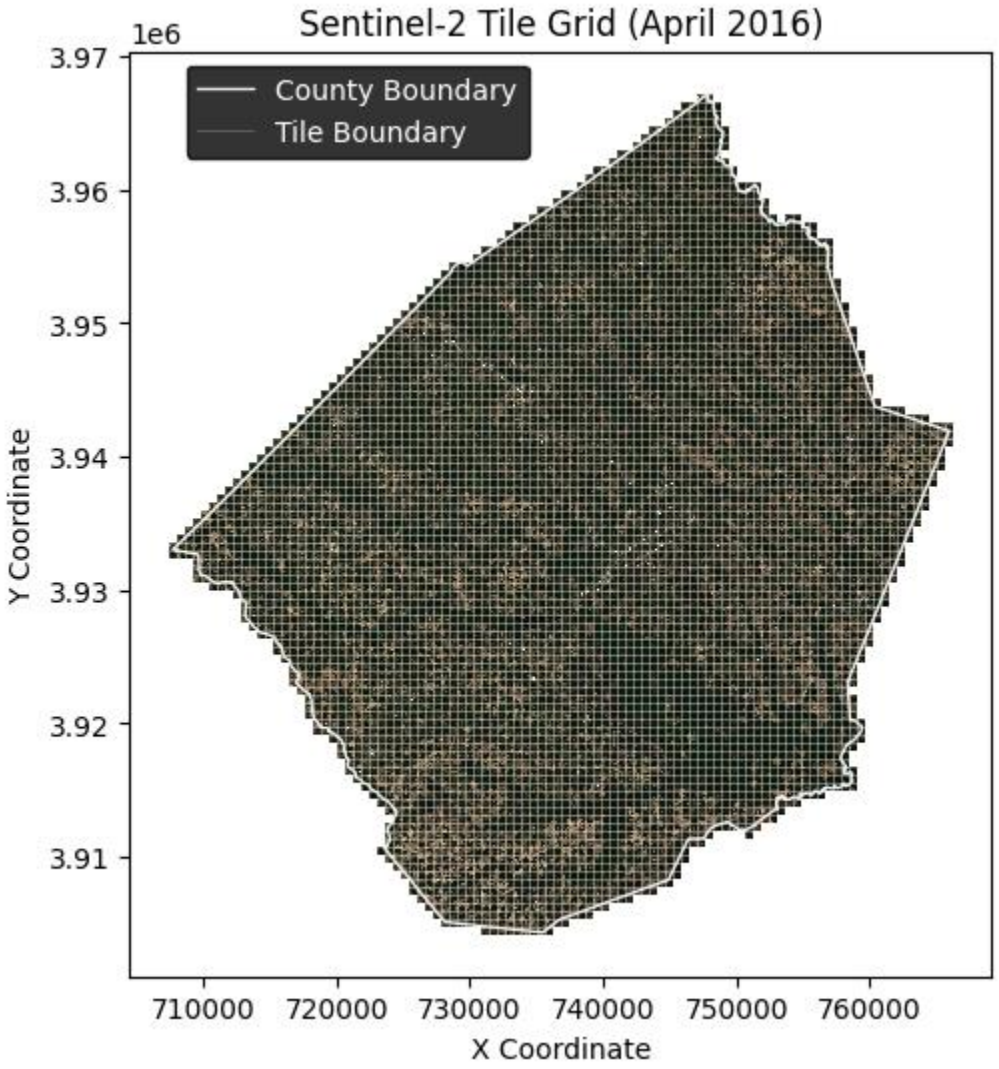

Figure 4: Sentinel-2 imagery of Johnston County in April 2016, overlaid with a 600x600-meter tile grid used and the county boundary.

To leverage the multispectral Sentinel-2 dataset for urban growth analysis, the Clay Foundation Model was employed—a pretrained deep learning framework specifically designed for Earth observation applications (Clay Foundation, 2023). This model was utilized to extract spatial embeddings from the Sentinel-2 imagery, effectively capturing complex spectral features relevant to urban imperviousness. Prior to embedding extraction, the Sentinel-2 images were preprocessed to align with the model's input requirements, including normalization of pixel values and the integration of temporal and spatial context.

Each Sentinel-2 image was normalized by applying band-specific means and standard deviations, ensuring consistent input across the dataset. To incorporate temporal information, temporal embeddings were generated for each image date using sine and cosine transformations of the date values, capturing seasonal patterns that might influence spectral signatures. Spatial context was integrated by calculating normalized latitude and longitude embeddings based on the centroid coordinates of each 600×600-meter tile, allowing the model to account for geographic variations within the study area.

Using the preprocessed data, data cubes were constructed for each tile and image date by combining the spectral bands with the temporal and spatial embeddings. The Clay Foundation Model was then applied

to these data cubes to extract spatial embeddings for each tile, resulting in a rich set of features that encapsulated both spectral and contextual information pertinent to urban imperviousness. These embeddings were stored alongside the associated dates and tile identifiers, forming a comprehensive dataset for subsequent modeling.

Concurrently, the urban imperviousness data was processed to serve as the target variable for model training. The resampled 200-meter resolution imperviousness data for each tile was normalized by dividing by 100 to represent percentage values between 0 and 1. This resulted in a 3×3 matrix of imperviousness values for each tile, providing a spatially detailed target for the model. By associating the extracted embeddings with the corresponding imperviousness matrices, a dataset suitable for supervised learning was established, where the model aimed to predict urban imperviousness based on the spectral and contextual features captured in the embeddings.

The final dataset was organized into a geospatial data structure, aligning the embeddings, temporal and spatial features, and imperviousness targets for each tile and date. This structured approach facilitated efficient data handling and allowed for the implementation of deep learning models to analyze urban growth patterns across Johnston County. The dataset was divided into training and testing subsets to evaluate the model's performance, ensuring robust assessment of its predictive capabilities.

To evaluate predictive performance, the dataset was partitioned into training, validation, and testing subsets based on temporal and spatial considerations. The data were first filtered to include only imagery and imperviousness data from before 2017 for model training and validation, reserving post-2017 data for future predictions and assessment. Unique tile indices were shuffled and divided according to a 70:10:20 ratio, allocating 70% of the tiles to the training set, 10% to the validation set, and 20% to the testing set. This stratification ensured a representative spatial distribution across the study area and prevented spatial autocorrelation from biasing the evaluation.

The feature set comprised the flattened spatial embeddings extracted from the Clay Foundation Model, which encapsulated the spectral and contextual information for each tile and date. The target variable was the corresponding 3×3 matrix of urban imperviousness values for each tile, reshaped into a one-dimensional array to align with the model's output requirements. Both features and targets were converted into numerical arrays and then into PyTorch tensors to facilitate efficient computation during model training. Data preprocessing steps, including normalization and reshaping, were applied consistently across all subsets to maintain compatibility with the deep learning framework and to optimize the learning process.

To establish a benchmark for predictive performance, a baseline model was developed using the mean urban imperviousness calculated across the training dataset. This baseline prediction served as a reference point against which the performance of more complex models could be compared. Evaluating the neural networks relative to this baseline allowed assessment of the extent to which they captured the spatial variability and intricate patterns of urban imperviousness beyond the average values.

Four distinct neural network architectures were implemented to predict urban imperviousness from the extracted spatial embeddings: a simple fully connected neural network (SNN), a deeper fully connected neural network (DNN), a convolutional neural network (CNN), and a long short-term memory (LSTM) network. Each model was designed to explore different aspects of the data and assess how effectively the embeddings captured relevant information for predicting urban imperviousness.

Simple Neural Network (SNN)

The simple neural network served as a baseline neural model to assess the predictive capacity of the embeddings with a minimal network structure. It consisted of an input layer, a single hidden layer with nonlinear activation functions, and an output layer producing the predicted imperviousness values. This model was chosen for its simplicity and to establish whether the embeddings contained sufficient information for prediction without complex transformations.

Mathematically, the SNN can be represented as:

$$\mathbf{h} = \text{ReLU}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)$$

$$\hat{\mathbf{y}} = \sigma(\mathbf{W}_2\mathbf{h} + \mathbf{b}_2)$$

where:

- $\mathbf{x} \in \mathbb{R}^{768}$ is the input feature vector (flattened embeddings).
- $\mathbf{W}_1 \in \mathbb{R}^{128\times768}$ and $\mathbf{W}_2 \in \mathbb{R}^{9\times128}$ are weight matrices.
- $\mathbf{b}_1 \in \mathbb{R}^{128}$ and $\mathbf{b}_2 \in \mathbb{R}^9$ are bias vectors.
- ReLU is the Rectified Linear Unit activation function:

$$\text{ReLU}(\mathbf{x}) = \max(0, \mathbf{x})$$

- σ is the sigmoid activation function ensuring outputs between 0 and 1:

$$\sigma(\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{x}}}$$

- $\hat{\mathbf{y}} \in \mathbb{R}^9$ is the predicted imperviousness vector.

Deep Neural Network (DNN)

The deep neural network extended the architecture by adding an additional hidden layer between the input and output layers. This increased depth allowed the model to learn more complex representations and capture nonlinear relationships within the data. By incorporating multiple layers, the DNN could hierarchically extract features from the embeddings, potentially improving predictive performance over the simpler architecture.

The DNN is mathematically expressed as:

$$\mathbf{h}_1 = \text{ReLU}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)$$

$$\mathbf{h}_2 = \text{ReLU}(\mathbf{W}_2\mathbf{h}_1 + \mathbf{b}_2)$$

$$\hat{\mathbf{y}} = \sigma(\mathbf{W}_3\mathbf{h}_2 + \mathbf{b}_3)$$

where the additional hidden layer $\mathbf{h}_2$ allows for deeper feature extraction.

Convolutional Neural Network (CNN)

The convolutional neural network was designed to leverage the spatial hierarchies inherent in the embeddings by applying convolutional operations. The CNN architecture included convolutional layers that applied filters to local regions of the input data, capturing spatial patterns and dependencies that might be relevant for predicting urban imperviousness. This approach is particularly effective for data where spatial locality and structure play crucial roles in the underlying phenomena.

Mathematically, the CNN can be represented as:

$$\mathbf{h}_{\text{fc}} = \text{ReLU}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)$$

$$\mathbf{h}_{\text{conv}} = \text{ReLU}\big(\text{Conv1D}(\mathbf{h}_{\text{fc}})\big)$$

$$\hat{\mathbf{y}} = \sigma(\mathbf{W}_2\mathbf{h}_{\text{conv}} + \mathbf{b}_2)$$

where:

- $\mathbf{h}_{\text{fc}}$ is the output of the initial fully connected layer.
- Conv1D represents the one-dimensional convolution operation applied to $\mathbf{h}_{\text{fc}}$.
- $\mathbf{h}_{\text{conv}}$ captures the spatial features extracted by the convolutional filters.

Long Short-Term Memory Network (LSTM)

The LSTM network was employed to explore the potential temporal dynamics captured within the spatial embeddings, despite the primary focus being spatial analysis. LSTM networks are adept at modeling sequential data and can capture long-term dependencies through their gated architecture. By processing the embeddings through LSTM layers, the model might uncover temporal patterns or sequences embedded within the data that could enhance prediction accuracy.

The LSTM model can be described mathematically as:

$$\mathbf{h}_{\text{fc}} = \text{ReLU}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)$$

$$\mathbf{h}_{\text{lstm}},\ \mathbf{c} = \text{LSTMCell}(\mathbf{h}_{\text{fc}},\ \mathbf{h}_{t-1},\ \mathbf{c}_{t-1})$$

$$\hat{\mathbf{y}} = \sigma(\mathbf{W}_2\mathbf{h}_{\text{lstm}} + \mathbf{b}_2)$$

where:

- $\mathbf{h}_{\text{lstm}}$ is the hidden state, and $\mathbf{c}$ is the cell state of the LSTM.
- $\mathbf{h}_{t-1}$ and $\mathbf{c}_{t-1}$ are the previous hidden and cell states.
- LSTMCell represents the computations within an LSTM unit.

Table 2: Summary of Neural Network Models.

Model	Architecture	Activation Function(s)	Parameters
Baseline	Mean of training data	N/A	N/A
Simple NN	Input (768) → Hidden Layer (128) → Output (9)	ReLU, Sigmoid	
Deep NN	Input (768) → Hidden Layers (128, 128) → Output (9)	ReLU, Sigmoid	
CNN	Input (768) → FC Layer (128) → Conv1D Layer → Output (9)	ReLU, Sigmoid	
LSTM	Input (768) → FC Layer (128) → LSTM Layer → Output (9)	ReLU, Sigmoid	

All models were trained using supervised learning techniques, optimizing the mean squared error (MSE) loss function to minimize the difference between the predicted ($\hat{\mathbf{y}}$) and actual ($\mathbf{y}$) imperviousness values:

$$\text{MSE} = \frac{1}{\mathbf{n}}\sum_{i=1}^{n}(\hat{\mathbf{y}}_i - \mathbf{y}_i)^2,$$

where $\mathbf{n}$ is the number of samples. Early stopping was employed as a regularization technique to prevent overfitting; training was halted if the validation loss did not improve over a predefined number of epochs. This approach helped in maintaining the model's generalizability to unseen data.

Model performance was evaluated on the test dataset using standard regression metrics, including MSE and mean absolute error (MAE). The MAE is calculated as:

$$\text{MAE} = \frac{1}{\mathbf{n}}\sum_{i=1}^{n}|\hat{\mathbf{y}}_i - \mathbf{y}_i|,$$

Additionally, residual analyses were conducted to examine the distribution and magnitude of prediction errors, providing insights into model biases and areas for potential improvement. This involved plotting residuals and assessing whether errors were randomly distributed or exhibited patterns indicating model deficiencies.

By comparing the performance of these models against the baseline, the study aimed to determine which neural network architecture most effectively leveraged the embeddings to predict urban imperviousness. The inclusion of various architectures allowed for a comprehensive evaluation of different modeling strategies in capturing the complex relationships within the data.

Results

- Present and explain the results qualitative and quantitative, tables, graphs, maps/images; compare with results from other studies – confirms previously observed phenomena, shows something new, which questions remain unresolved.

Discussion

Conclusion

- Summary of the most important findings including advances in methodology, future work

References

1. Ayush, K., Uzkent, B., Meng, C., Kumar, T., Burke, M., Lobell, D., & Ermon, S. (2021). Geography-Aware Self-Supervised Learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 10181–10190). 10.1109/ICCV48922.2021.01002
2. Clay Foundation. (2023). *Clay Foundation Model: An Open Source AI Model for Earth*. Retrieved from https://www.clay.earth
3. Copernicus Sentinel-2 (processed by ESA). (2021). *MSI Level-1C TOA Reflectance Product. Collection 1. European Space Agency.* https://doi.org/10.5270/S2_-742ikth
4. Dionelis, N., Fibaek, C., Camilleri, L., Luyts, A., Bosmans, J., & Le Saux, B. (2024). *Evaluating and Benchmarking Foundation Models for Earth Observation and Geospatial AI*. arXiv preprint, arXiv:2406.18295. https://doi.org/10.48550/arXiv.2406.18295
5. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). "Masked Autoencoders Are Scalable Vision Learners." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16000-16009. doi:10.1109/CVPR52688.2022.01553
6. Jean, N., Wang, S., Samar, A., Azzari, G., Lobell, D., & Ermon, S. (2019). *Tile2Vec: Unsupervised representation learning for spatially distributed data.* Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 3967–3974. doi:10.1609/aaai.v33i01.33013967
7. Johnston County Department of GIS. (2011). "Johnston County GIS Data." Retrieved from https://www.johnstonnc.com/gis2/content.cfm?PD=data. Metadata last reviewed on 2011-02-01. Contact: Craig Franklin, GIS System Analyst, Johnston County Department of GIS, 212 East Market Street, Smithfield, NC 27577, USA
8. Mai, G., Cundy, C., Choi, K., Hu, Y., Lao, N., & Ermon, S. (2022). *Towards a foundation model for geospatial artificial intelligence*. Proceedings of the 30th International Conference on Advances in Geographic Information Systems (Article No. 106, pp. 1-4). https://doi.org/10.1145/3557915.3561043
9. Microsoft Open Source, McFarland, M., Emanuele, R., Morris, D., & Augspurger, T. (2022, October 28). microsoft/PlanetaryComputer: October 2022 (Version 2022.10.28) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.7261897
10. R. Goetzke, M. Braun, H. -P. Thamm and G. Menz (2008). *Monitoring and Modeling Urban Land-Use Change with Multitemporal Satellite Data*. IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 2008, pp. IV - 510-IV - 513, doi: 10.1109/IGARSS.2008.4779770
11. STAC Specification. (2023). *SpatioTemporal Asset Catalog (STAC) API: OpenAPI Definition (Version 1.0.0)*. Retrieved from http://stacspec.org

12. U.S. Geological Survey. (2021). "National Land Cover Database (NLCD) 2019 Products." *U.S. Department of the Interior*. Retrieved from https://www.mrlc.gov/data
13. Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). "Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources." *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36. doi:10.1109/MGRS.2017.2762307

Appendix

- Workflows, commands, scripts, metadata, software-specific issues

Pseudo-code For Initial Data Processing

1. **Setup Libraries and Directories**
    - Import required libraries for geospatial, raster, and deep learning processing.
    - Define the directory paths to store tiles, boundaries, and processed data.
2. **Load and Preprocess County Boundary Data**
    - Load `county_boundary.shp` as a GeoDataFrame (`GDF`).
    - Reproject to **EPSG:32617 (UTM Zone 17N)** to match Sentinel-2 data CRS.
    - Calculate bounding coordinates of the boundary with `minx`, `miny`, `maxx`, and `maxy` values and align these to a tile size (600m).
3. **Generate Grid Tiles for the County**
    - Create a grid of 600x600-meter tiles that cover the county boundary.
    - Store these as polygons within a new `GDF`.
    - Ensure each tile intersects with the county boundary; keep full tile geometry.
    - Initialize columns in `GDF` to track whether each tile has been processed and its associated data files.
    - Save the tiles as a GeoJSON file.
4. **Create Unified Boundaries for All Tiles**
    - Combine all tiles into a single unified boundary polygon.
    - Save this as a **shapefile** and create a **raster mask** with a **10-meter pixel resolution**.
    - Define a transform for the raster that aligns with the bounds and CRS of the tiles.
5. **Load and Mask Urban Imperviousness Data**
    - Load **NLCD urban imperviousness raster** and clip it to the county's bounding box.
    - Save this clipped data as an intermediate TIFF file.
    - Resample the clipped raster to **10m and 200m** resolutions, reprojecting it to **EPSG:32617** with **bilinear resampling**.
    - Save both resampled TIFF files.
6. **Extract Urban Imperviousness Data for Each Tile**
    - For each tile, extract a 200m-resampled urban imperviousness raster corresponding to that tile’s bounding box.
    - Save each extracted tile raster as a TIFF, updating the `GDF` with the path to each tile’s raster file.

7. **Query Available Sentinel-2 Dates**
    - Set up a query for **Sentinel-2 imagery** within a specified date range (e.g., 2016-01-01 to 2024-08-31).
    - Filter for images with **<1% cloud cover** using Microsoft Planetary Computer STAC API.
    - Save the available dates to a pickle file.
8. **Select Optimal Dates for Data Collection**
    - Group the available dates by year and select a specified number of dates per year (e.g., quarterly).
    - Store the selected dates as the main temporal dataset.
9. **Download and Save Sentinel-2 Image Tiles**
    - For each tile and each selected date, query STAC items for Sentinel-2 imagery in the tile's bounding box.
    - Extract and save each band (Blue, Green, Red, NIR) as TIFF files, ensuring alignment with the tile bounds and CRS.
    - Update the `GDF` with the file paths to each downloaded image file.
10. **Generate Mosaics and RGB Composites**
    - Create RGB mosaics for the tiles using bands B02 (Blue), B03 (Green), and B04 (Red).
    - Save the merged mosaics as TIFF files, creating RGB composites for visualization.
11. **Merge Subdivided Urban Data Tiles into a Single Raster**
    - Collect all 200m-resampled urban tiles and merge them into a single **GTiff** raster.
    - Ensure the final merged raster retains the CRS and spatial alignment of the tiles.
12. **Assign Data Files to Dates for Each Tile**
    - For each tile, associate its downloaded Sentinel-2 data with specific dates.
    - Update the `GDF` to track which dates are associated with available data files, facilitating data retrieval for modeling.
13. **Error Handling and Cleanup**
    - Identify missing files, incomplete downloads, and tiles with multiple files for the same date.
    - Re-query missing data and remove redundant or incomplete files.

Model Embeddings and Train/Test Data Setup Pseudocode

1. Setup Environment

- Import required libraries for geospatial processing, deep learning, data handling, and utilities.
- Set up directories and file paths for data, tiles, and model checkpoints.
- Define parameters such as Sentinel-2 bands, spatial resolution (10m), and coordinate reference system (EPSG:32617).

2. Load Data

- Load the geospatial data files (tiles_with_dates.geojson and county_boundary.shp) as GeoDataFrames.

- Define data directories and mapping for the Sentinel-2 bands to color labels (e.g., B02 -> Blue).
- Load available dates from a pre-saved pickle file to check available imagery for each tile.
- Load the 20 selected dates used by the Sentinel-2 data (~3 per year, 2016-2024).

4. Initialize Clay Model for Embeddings

- Define and load the Clay Foundation Model from the checkpoint file.
- Set the model to evaluation mode and load it onto the available device (GPU or CPU).

5. Retrieve and Stack Sentinel-2 Image Data for Each Tile

- Define a function to retrieve image data (bands) for each tile based on its geometry.
- Stack multispectral image data for each selected date, aligning with the Sentinel-2 bands.
- Normalize data based on band-specific means and standard deviations.

6. Generate Additional Features

- Calculate the temporal embedding for each date using sine and cosine transformations to encode week and hour information.
- Calculate normalized latitude and longitude embeddings based on each tile's centroid.

7. Prepare Data Cubes for Embedding Extraction

- Create a data cube for each tile by combining spectral, temporal, and spatial information.
- Normalize pixel values and generate lat/lon and temporal embeddings as per the model's expected format.

8. Extract Embeddings for Each Tile and Save

- For each tile, run the model to generate spatial embeddings for the date-specific data cube.
- Store embeddings and associated dates within the GeoDataFrame for future retrieval.
- Periodically save the updated GeoDataFrame to a pickle file to avoid data loss during long processing runs.

9. Process Urban Imperviousness Data

- Load each tile's 200m urban imperviousness data as matrices and normalize values (e.g., divide by 100 for percentages).
- Save the urban imperviousness data for each tile as a separate GeoDataFrame.

10. Combine Embeddings and Urban Imperviousness Data

- For each tile, extract embeddings, coordinates, and urban imperviousness matrices.
- Flatten embeddings and urban imperviousness matrices, aligning with the spatial and temporal features.
- Save the final merged GeoDataFrame with all relevant features, embeddings, and labels.

Train/Test Split and Model Setup Pseudocode

1. Load and Preprocess Data

- Load the merged GeoDataFrame containing embeddings and urban imperviousness matrices.
- Filter data by date, designating pre-2017 data for training/testing and post-2017 data for future predictions.

2. Define Train, Validation, and Test Splits

- Shuffle unique tile indices and split them into training, validation, and test sets based on defined ratios (e.g., 70% train, 10% validation, 20% test).
- Filter the GeoDataFrame by tile indices to create separate DataFrames for each set (train, validation, test).

3. Define Feature and Target Columns

- Identify the columns representing model features (e.g., feature_0 to feature_767) and target (urban imperviousness matrix).
- Reshape the urban imperviousness data into 3x3 matrices for each tile to be used as model targets.

4. Prepare Data for Deep Learning Model

- Convert features and targets for each dataset (train, validation, test) into NumPy arrays.
- Reshape urban imperviousness matrices to a flattened 9-element vector per tile for compatibility with model input.

5. Convert Data to PyTorch Tensors

- Convert the feature arrays and target arrays into PyTorch tensors.
- Flatten the 3x3 matrices to 1D arrays of 9 elements to simplify the training process.

6. Save Data for Model Training

- Print dimensions to confirm shapes align with model input requirements.
- Finalize and save the processed data as tensors in PyTorch format, ready for training and evaluation.