Geospatial Analysis for the Land Trust of Tennessee

Product Managers: Manda Li
Team Members: Isaac Liu, Steven Yi, Ann Mathew, Teo Kitanovski

VANDERBILT DATA SCIENCE

THE LAND TRUST FOR TENNESSEE

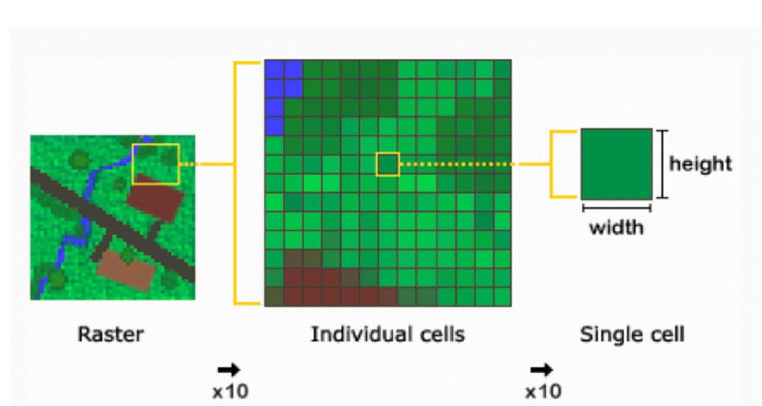
01 Introduction

Project Mission Statement:
The Land Trust of Tennessee seeks to identify potential land conversion from natural or farmland to developed lands, to better focus conservation efforts.

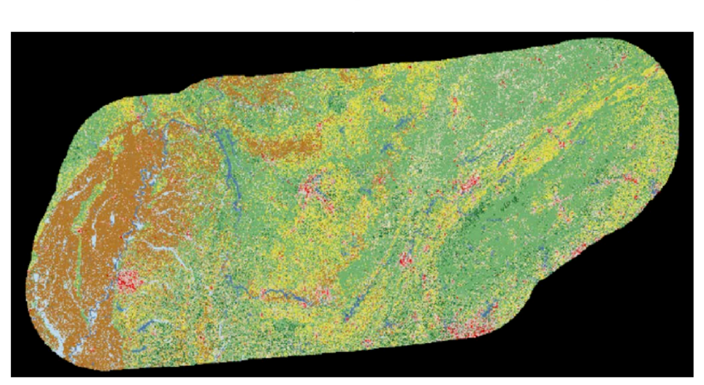
Project Goals:
Provide a geospatial raster dataset depicting future land cover change likelihoods in future years (2030, 2040, 2050, etc.) within the targeted region.

02 Our Data

Raster data is a type of digital image comprised of a matrix of cells, each representing specific information, such as color or intensity, commonly used in satellite imagery and geographic mapping.



Area of Interest: Tennessee
Our raster dataset (2001, 2004, 2006, 2008, 2011, 2013, 2016, 2019, 2021) is collected from the NLCD under the U.S Geological Survey, and is composed of 30-meter by 30-meter pixels.



Class 21-24 represents the developed lands that natural lands will change to.

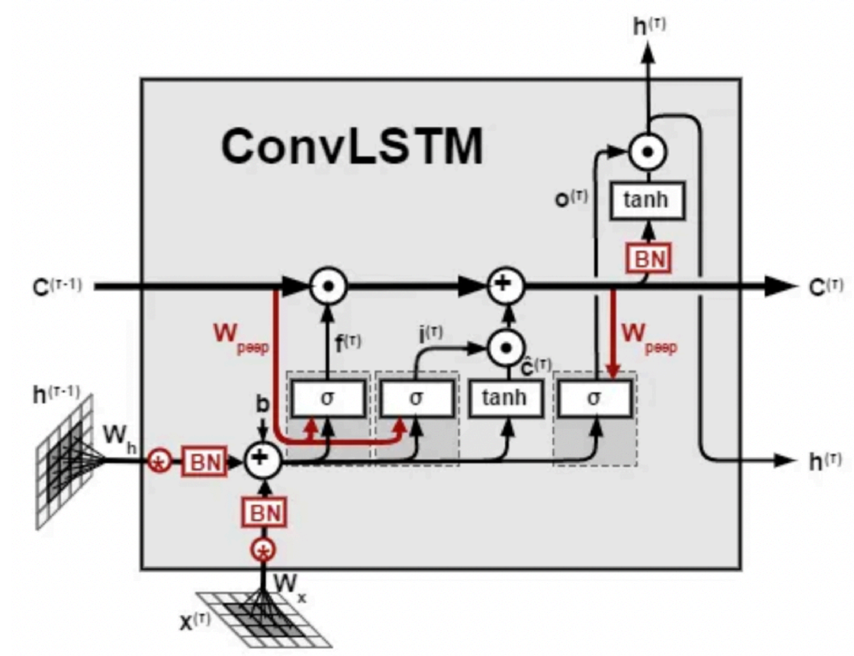
NAME	FROM	TO
Deciduous Forest	41	21-24
Evergreen Forest	42	21-24
Mixed Forest	43	21-24
Shrub/Scrub	52	21-24
Grassland/Herbaceous	71	21-24
Pasture/Hay	81	21-24
Cultivated Crops	82	21-24
Woody Wetlands	90	21-24
Emergent Herbaceous Wetlands	95	21-24

03 Approach

In the study, raster images are preprocessed using rasterio to extract relevant 64x64 pixel patches. This preprocessing includes filtering based on 2D prefix sum arrays, ensuring only dynamic and relevant patches are used, optimizing the training process.

The ConvLSTM model is chosen for its unique ability to analyze spatial and temporal data concurrently, essential for time-series raster data. It combines spatial pattern recognition with temporal sequence understanding, making it ideal for predicting land type changes.

The training process involves feeding sequences of these processed patches to the ConvLSTM, allowing it to learn the evolution of land types. This method ensures the model captures both spatial and temporal dynamics.



Model: "sequential"

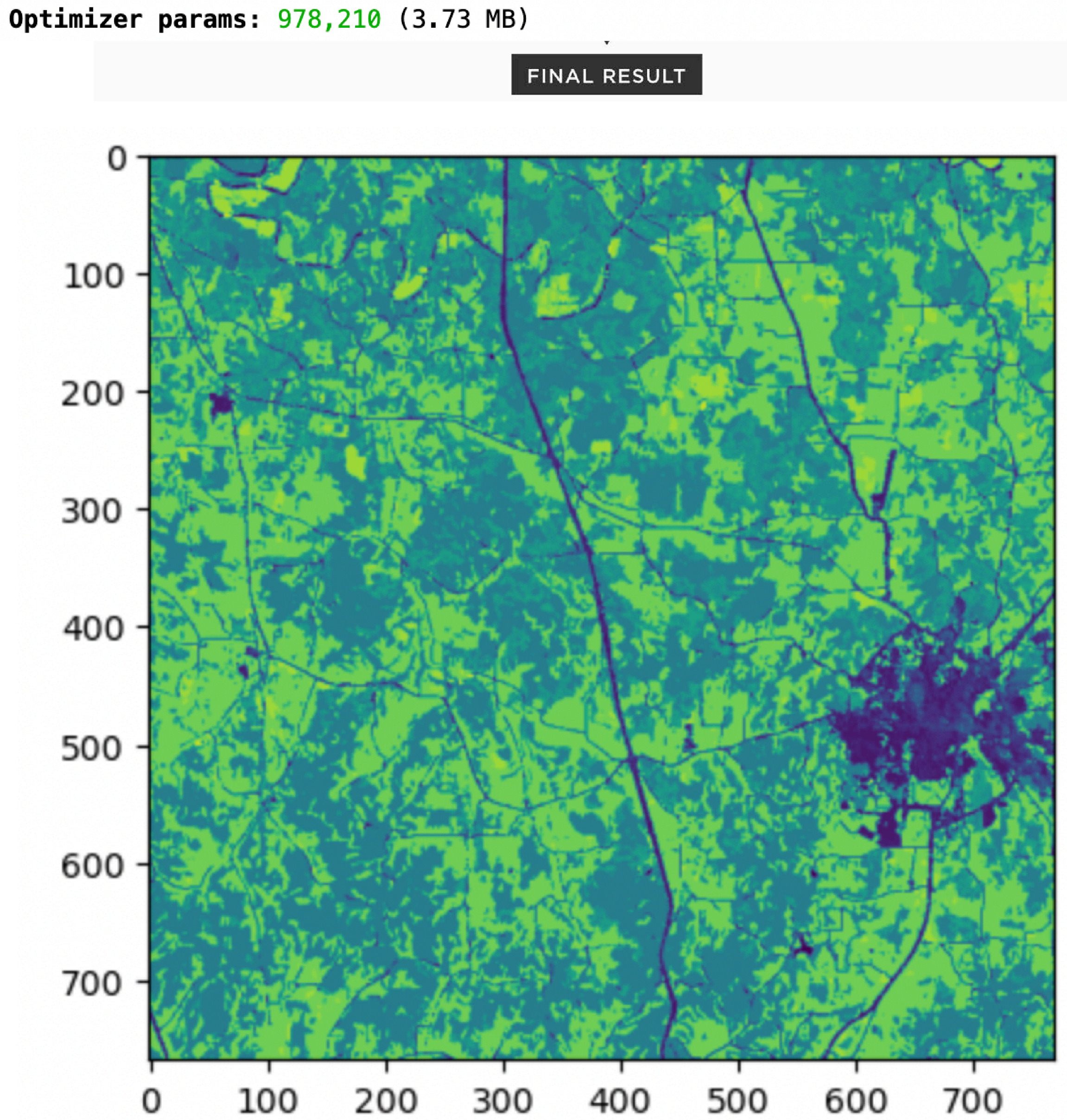
Layer (type)	Output Shape	Param #
conv_lstm2d (ConvLSTM2D)	(None, 5, 64, 64, 64)	184,576
batch_normalization (BatchNormalization)	(None, 5, 64, 64, 64)	256
conv_lstm2d_1 (ConvLSTM2D)	(None, 64, 64, 64)	295,168
conv2d (Conv2D)	(None, 64, 64, 16)	9,232

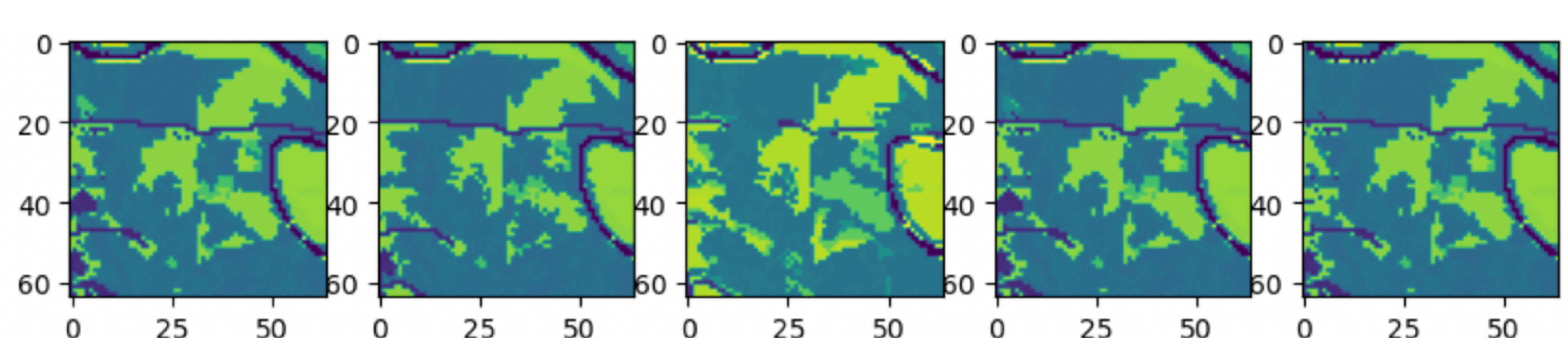
Total params: 1,467,442 (5.60 MB)
Trainable params: 489,104 (1.87 MB)
Non-trainable params: 128 (512.00 B)
Optimizer params: 978,210 (3.73 MB)

04 Final Product

Predictions are made by loading the previously trained ConvLSTM model. The process involves setting specific parameters for generating a comprehensive prediction map. This is executed by applying the model to numerous consecutive 64x64 patches, representing different years, initially without overlap.

Subsequently, a more refined map is created with overlapping patches to ensure continuity and coverage across the temporal dimension. At each pixel, probability distributions from all overlapping patches are aggregated before applying the argmax function to determine the most likely land type.





05 Limitations

Considering the conversion of raster data to vector data for use with classification or regression models was an initial strategy. However, converting raster data to vector data using ArcGIS and Rasterio results in significantly large datasets. The size of these datasets complicates concatenation of all vector data and the addition of useful features.

Challenges were also encountered in processing unfamiliar file types. Additionally, due to the lack of comprehensive features in the converted dataset(slope, floodplain, wetlands, etc), relying solely on classification models may not yield accurate predictions.

06 Future Directions

Incorporating comprehensive geographical and demographic factors into the ConvLSTM model has the potential to significantly enhance its predictive accuracy, especially in the context of land conversion. By integrating features such as slope, floodplain status, wetland presence, excluded areas, and proximity to key infrastructure like transportation, schools, and other essential services, we can provide a more nuanced and contextually rich dataset for the model. Moreover, demographic factors like projected population growth and soil types further refine the predictive capabilities by offering insights into land suitability and development potential.

Our future direction aims to enrich the ConvLSTM model by integrating these diverse factors, which will allow for a more detailed and accurate analysis of land conversion patterns. By systematically including these variables, we expect not only to improve the model's performance but also to provide stakeholders with more robust tools for land use planning and environmental conservation.