# Distinguishing Urban versus Non-Urban Land Masses to Predict Urban Expansion using Satellite Imagery

Samuel Nelson

May 2025

## 1 Introduction

Satellite imagery provides spatial data to observe Earth's surface from above, and the collection and processing of this data contribute to the advancement of global Geographical Information Systems (GIS). The development of dynamic GIS is essential for innovating how we interact with our environment and surroundings.

The use of satellite imagery is especially prevalent within the study of Urban Expansion (UE) – a major global phenomenon impacting infrastructure, the environment, and resource allocation. Urban planners rely on critical information including the categorization of diverse land cover and land use to make informed decisions and guide infrastructure development to suit Earth's rapidly growing population.

The EuroSAT dataset, derived from the European Space Agency's Sentinel-2 mission, offers a plethora of labeled satellite images for 10 categories of land cover across regions of Europe. EuroSAT is widely used for land cover classification. A state-of-the-art convolutional neural network (CNN) benchmark on this dataset reached an overall accuracy of 98.57% for the original 10-class classification task.

This project reframes that task into a binary classification problem, predicting Urban versus Non-Urban land cover to evaluate whether machine learning models can effectively distinguish urban areas by using only RGB satellite images.

To support this analysis, I implemented a three-layer CNN and compared its performance with a random forest classifier benchmark. Both models were trained with EuroSAT RGB satellite imagery to classify each image as Urban or Non-Urban. The three-layer CNN achieved an overall accuracy of 93.85% whereas the random forest baseline achieved 83.45% accuracy.

## 2 EuroSAT Dataset

The EuroSAT RGB subset was used to support this work which comes from the European Space Agency's Sentinel-2 mission. It contains 27,000 labeled and geo-referenced images, covering 10 distinct land cover categories across regions of Europe. A diagram of the 10 different land cover categories can be seen in Figure 1 below. The categories of land cover

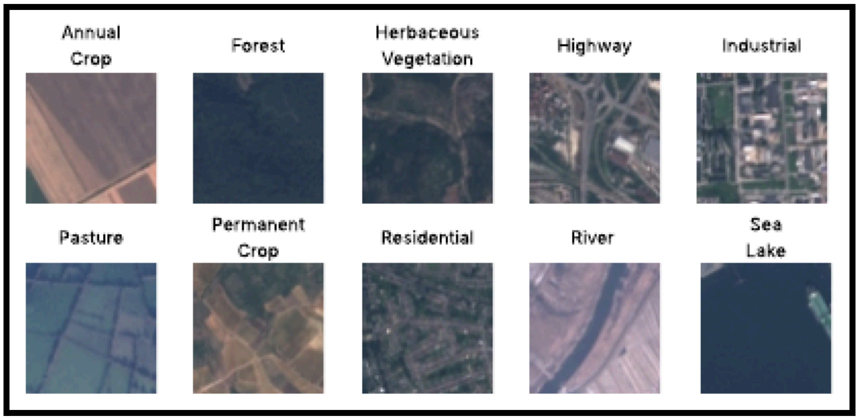


**Figure 1: Examples of each of the 10 categories of land cover.**

include AnnualCrop, Forest, HerbaceousVegetation, Highway, Industrial, Pasture, PermanentCrop, Residential, River, and SeaLake.

The original land cover classes were relabeled into binary classes. The Urban class included Highway, Industrial, and Residential samples. The Non-Urban class included AnnualCrop, Forest, HerbaceousVegetation, Pasture, PermanentCrop, River, and SeaLake samples. Since Urban images included 3 of the original 10 land cover classes, it only comprised ~22% of the dataset. To address this class imbalance, I applied random downsampling of the Non-Urban class in the training set. Both models utilized a 70% training, 15% validation, and 15% testing split to ensure equal class distribution.

# 3 Baseline Random Forest Classifier

I implemented a random forest classifier as a baseline for comparison. To balance model stability and speed, I used 100 decision trees. This helped to reduce training time and capture enough variance to improve the model's generalizability. Due to the balanced nature of the data after downsampling, I used Gini impurity as the criterion, which performs better on balanced data. Features were extracted by flattening each 3x64x64 input image into a 12,288-dimensional vector, removing spatial structure.

# 4 CNN

I utilized a three-layer CNN for further analysis of classification performance, and the corresponding architecture diagram can be found in Figure 2 below. The network consists of three convolutional blocks. Each block includes a 2D convolutional layer (with kernel size 3x3 and padding of 1), followed by batch normalization, a ReLU activation function, and a 2x2 max pooling layer. The first convolutional layer has 32 output channels, the second has 64, and the third has 128. After the final block, an adaptive average pooling layer reduces the feature map to a fixed 4x4 spatial size, regardless of the input dimensions. The resulting tensor is flattened and passed through a fully connected layer of 256 units with ReLU activation, followed by a dropout layer with a dropout rate of 0.5 for regularization. Finally, a single-unit output layer provides the binary classification logit. The network was trained using the binary cross-entropy with logit loss function, optimized with Adam at a learning rate of 0.0001 for 20 epochs. The model with the lowest validation loss was saved for evaluation.

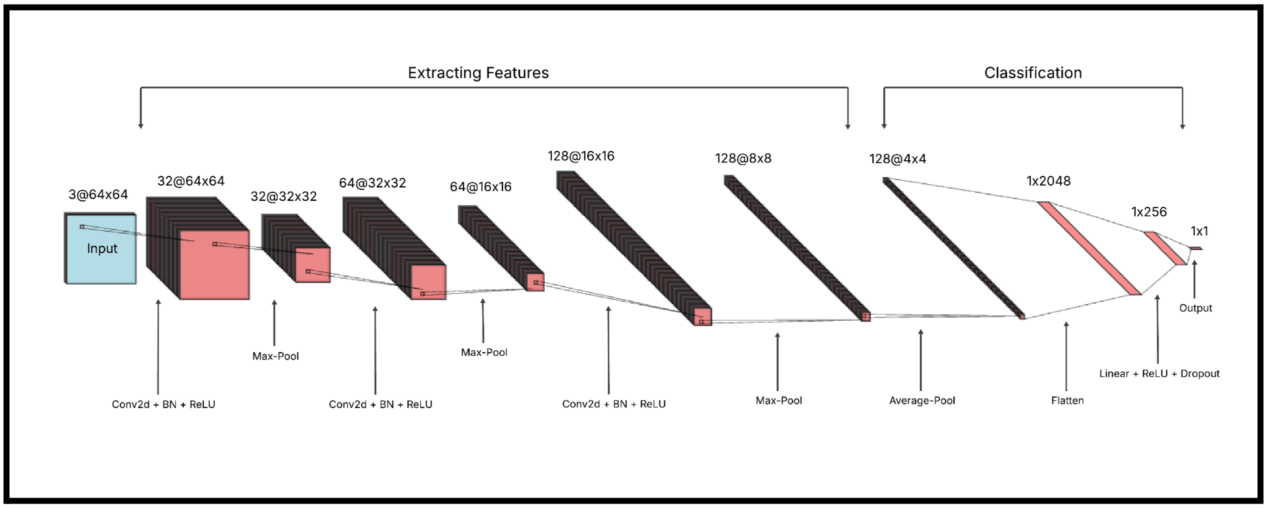


**Figure 2: Architecture of the three-layer CNN.**

# 5 Results

To evaluate the performance of the baseline random forest classifier and the three-layer CNN, I created a classification report, detailing the overall test accuracy from both models, and their corresponding precision, recall, and F1-score for both the Urban and Non-Urban classes. I also detail the resulting AUC score of both models and their respective ROC curves on the same subplot.

| Urban | | |
|---|---|---|
| | CNN | RF |
| Precision | 0.87 | 0.67 |
| Recall | 0.94 | 0.91 |
| F1-Score | 0.90 | 0.77 |

**Figure 3: Classification report for the Urban class.**

| Non-Urban | | |
|---|---|---|
| | CNN | RF |
| Precision | 0.97 | 0.95 |
| Recall | 0.94 | 0.80 |
| F1-Score | 0.95 | 0.87 |

**Figure 4: Classification report for the Non-Urban class.**

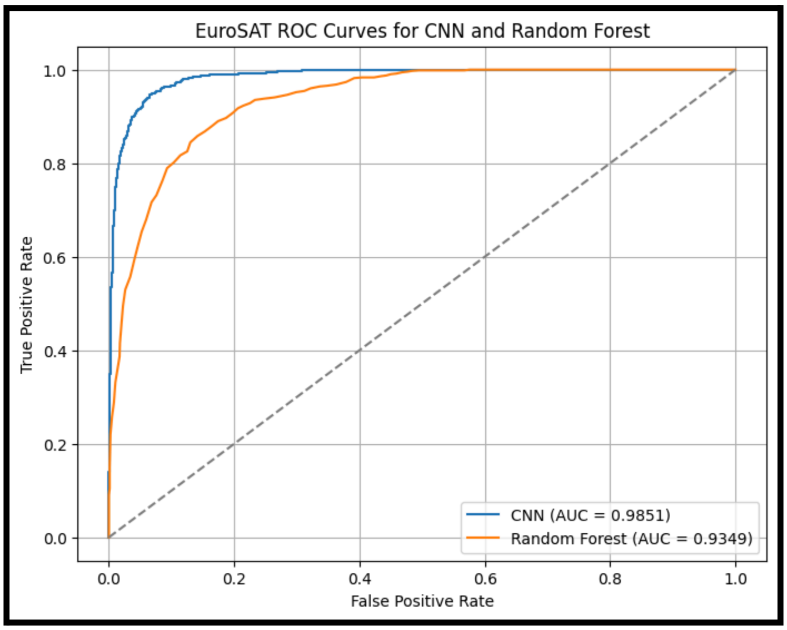


**Figure 5: ROC Curves for both models.**

## 5.1 Random Forest Classifier

With the random forest classifier using 100 decision trees, I achieved an overall accuracy of 83.45% on the test set. Random forest had a low F1-score (0.77) on the Urban class, mainly due to its poor recall (0.67). This suggests that it was fairly good at capturing actual Urban examples, but at the cost of more false positives because of how often it was predicting Urban. Random forest had a far greater F1-score (0.87) on the Non-Urban class with a precision of 0.95, likely due to its carefulness in predicting the Non-Urban class. It also had an AUC score of 0.93, which is good, but still far from perfect because of its inability to recognize spatial context.

## 5.2 CNN

I was able to achieve 93.85% accuracy with the three-layer CNN on the test set. The CNN had an AUC score of 0.99, meaning that it had almost perfect separation between the Urban and Non-Urban classes. It was particularly good at distinguishing the Non-Urban class, having an F1-score of 0.95 with a balanced score for both precision (0.97) and recall (0.94). Similarly to the random forest classifier, it lacked precision (0.87) for the Urban class, which could likely be due to confusion in spatial textures such as rooftops or buildings resulting from the somewhat shallow architecture.

# 6 Discussion

In this project, I compared a random forest classifier and a three-layer CNN to predict Urban and Non-Urban land areas using a dataset consisting of RGB satellite images. The CNN had an edge over the random forest classifier because it can handle features more effectively due to convolutional layers. The shallow architecture of the CNN proved to be expressive, but displayed some disadvantages when predicting the Urban class. The random forest classifier was limited by its flattened input representation, and removing spatial context likely affected its separation of the Urban and Non-Urban classes.

In hindsight, data augmentation could have helped the models' generalization by simulating cloud cover or different satellite positions. Future rework could include a more robust labeling strategy, especially when labeled categories overlap – for example, bodies of water are sometimes near cities and urban areas. EuroSAT also includes a multi-spectral subset with satellite images containing 13 spectral bands, which could be used to further improve the CNN's ability to recognize spatial ambiguities.

# 7 References

[1] Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. Patrick Helber, Benjamin Bischke, Andreas Dengel, Damian Borth. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019.

[2] Introducing EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. Patrick Helber, Benjamin Bischke, Andreas Dengel. 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018.