USING MACHINE LEARNING AND GOOGLE EARTH ENGINE TO UNDERSTAND LAND USE AND LAND COVER CLASSIFICATIONS AND NO_2 LEVELS IN CALIFORNIA

Author(s): Benyamain Yacoob, Ethan Scheys, Eyiara Oladipo, Andre Price
Advisor: Dr. Shadi Banitaan
University of Detroit Mercy
College of Engineering and Science (ECECS)

Overview

- Introduction
- Methodology
- Dataset(s)
- Results
- Implications

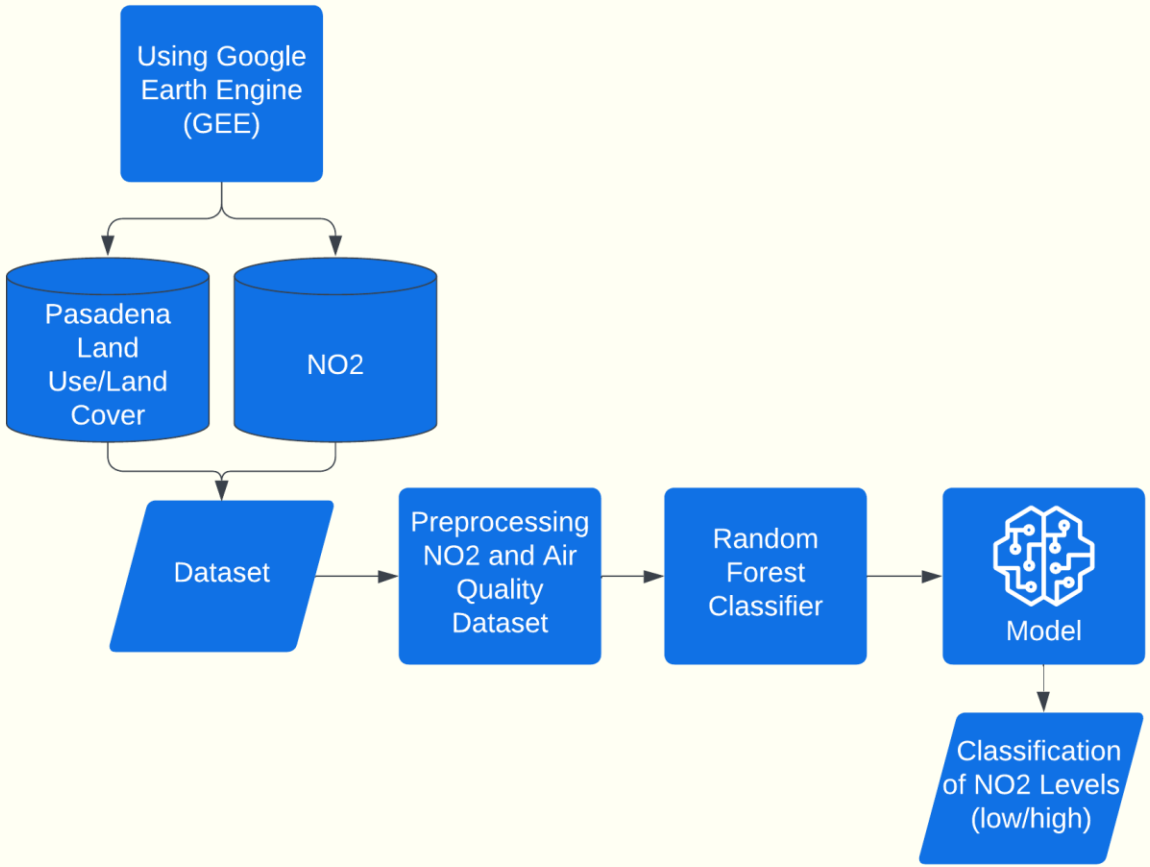
Introduction

- **Picture This**: Harnessing satellite imagery and AI to decode the air we breathe and revolutionize urban planning
- **Why It Matters**: Unveiling the hidden connections between how we use land and the quality of our environment
- **Join Us**: Discover the methods that could transform our cities into healthier, greener spaces

Methodology

- Focused on Pasadena, California as the study area
- Used Google Earth Engine (GEE) to prepare land cover and air quality datasets
- Sampling strategies and data distribution across attributes were described
- Random Forest (RF) and Decision Tree (DT) classifiers were used for analysis

Methodology

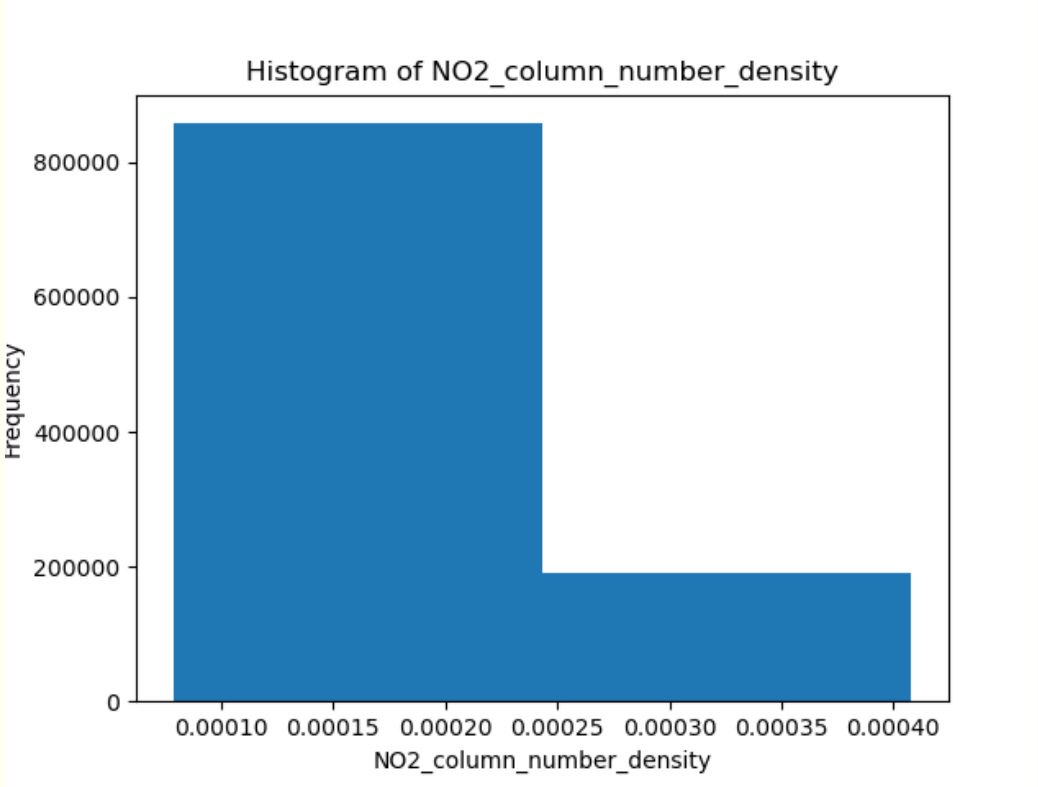


Dataset(s)

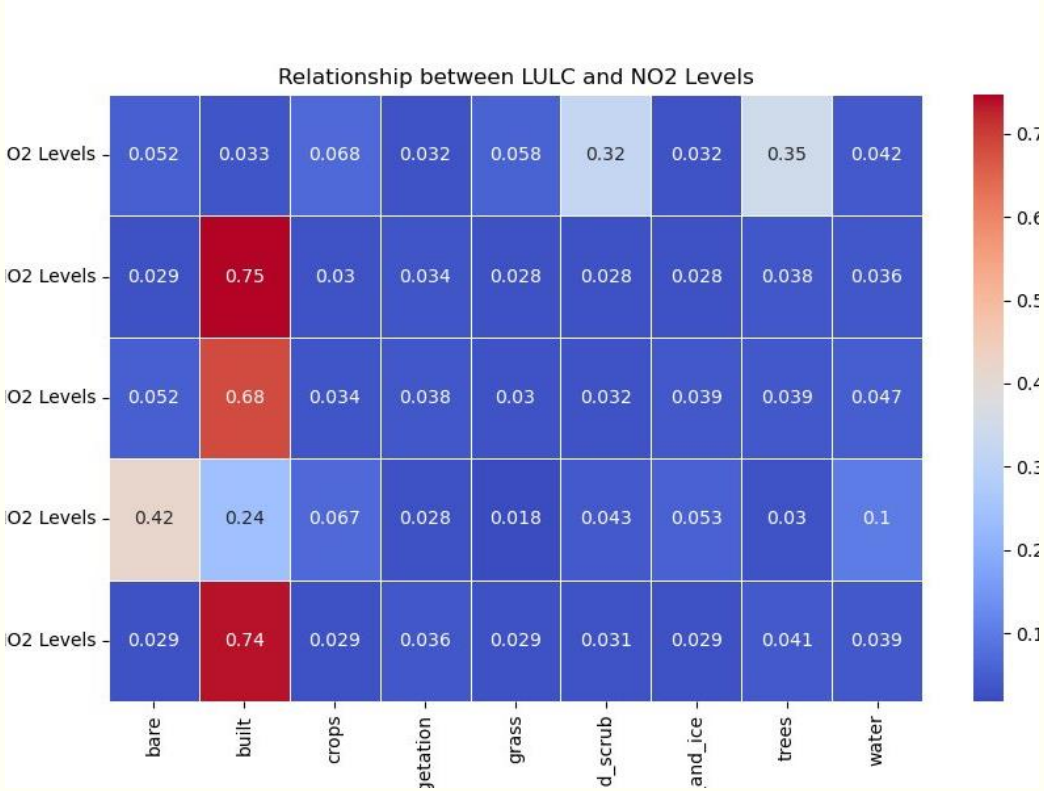
- Land Cover Datasets:
  - Used Google Dynamic World Dataset - GOOGLE/DYNAMICWORLD/V1
  - Data filtered monthly from January 1, 2021, to December 1, 2021 (199 images)
  - Used "sampleRegions" function to sample pixel values and obtained 1,048,576 elements
- Air Quality Datasets:
  - Used Sentinel-5P NRTI NO_2 dataset for near real-time high-resolution imagery of nitrogen dioxide concentrations
  - Used "NO_2 column number density" band (measured in mol/m^2)
  - Equal frequency discretization was used to create a binary classification task with "Low NO_2 Levels" and "High NO_2 Levels"

Dataset(s)

Histogram



Heatmap

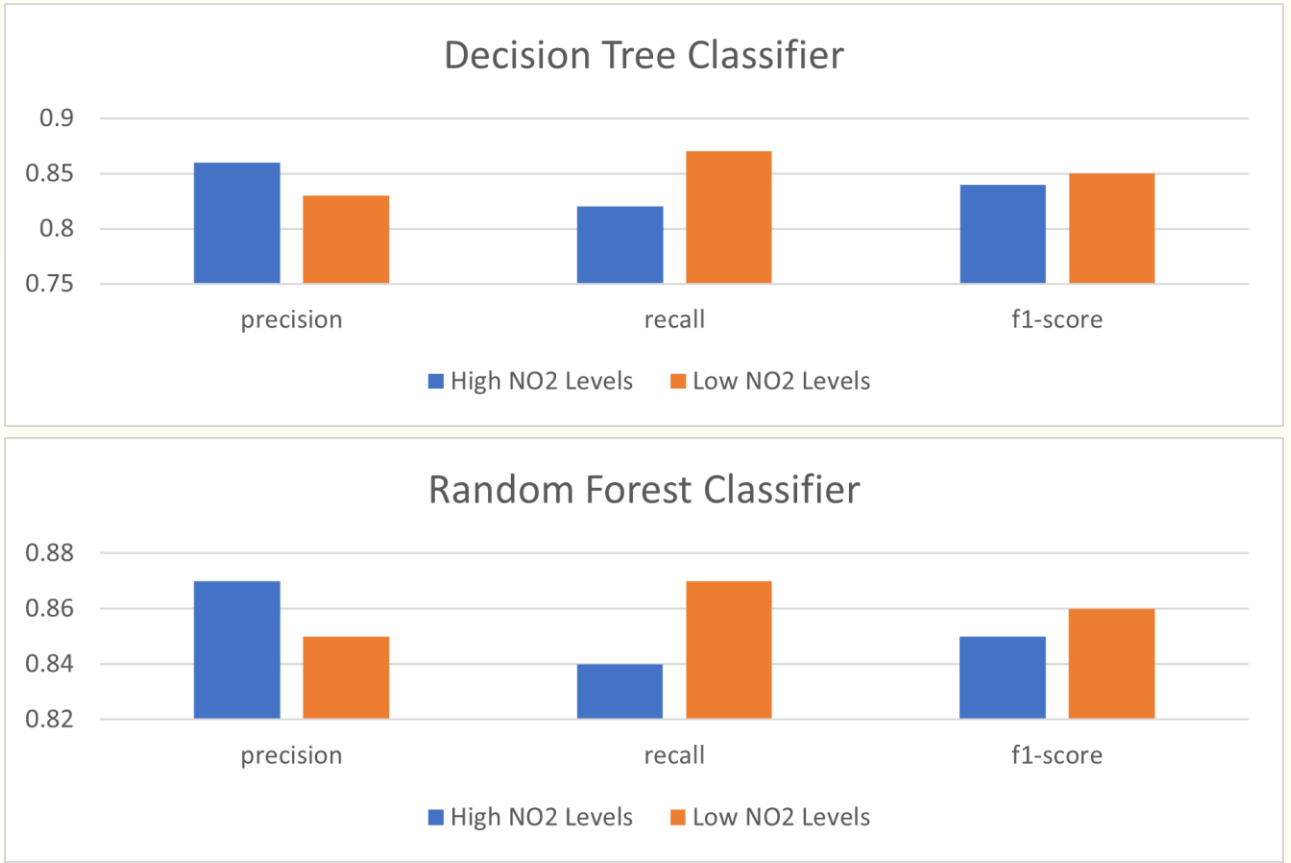


Results

- Feature importances and classification report: Internal Python programming language library

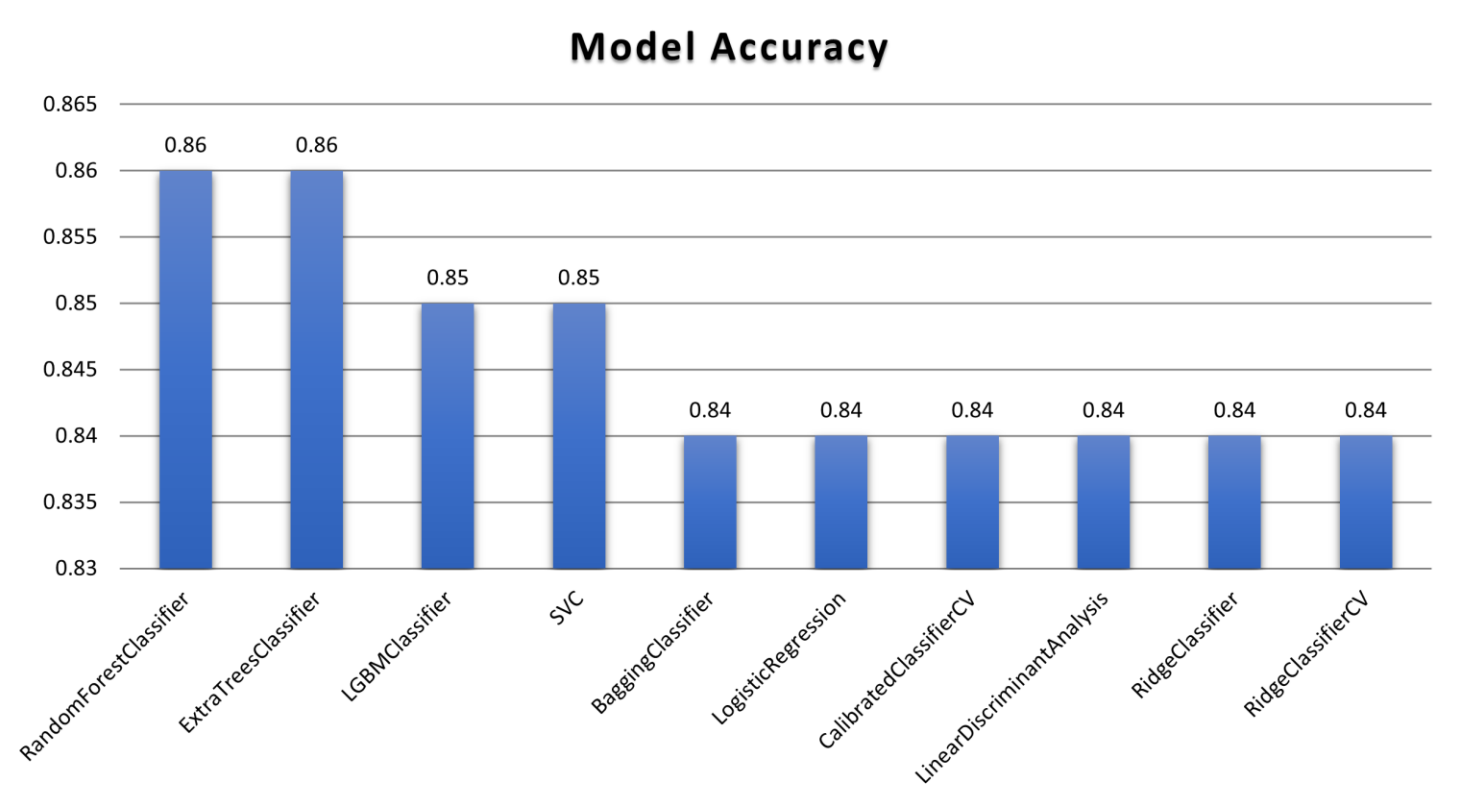
TABLE I: Feature Importance of Attributes Retrieved from LULC.

Attribute Type	Importance Score
Built	0.25
Shrub and Scrub	0.19
Trees	0.11
Grass	0.11
Water	0.07
Bare	0.07
Crops	0.07
Snow and Ice	0.06
Flooded Vegetation	0.06



Results

- Decision Tree (DT) accuracy: 85%
- Random Forest (RF) accuracy: 86%
- 10-fold cross-validation was used, achieving a mean accuracy of 86%
- The strong association between LULC and air quality (specifically NO_2 presence) was concluded based on the high accuracy



Implications

- The presence of NO_2 diminishes quality of life and standard of living, affects production of crops
- Incorporate greenery spaces to counteract anthropogenic activities present in urban areas
- Different regions of interest (variety in natural landscapes, man-made structures)
- Environmental policy regulations