2024 IEEE International Conference on Electro Information Technology (eIT) | 979-8-3503-3064-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/eIT60633.2024.10609851

Using Machine Learning and Google Earth Engine to Understand Land Use and Land Cover Classifications and NO_2 Levels in California

Benyamain Yacoob, Ethan Scheys, Eyiara Oladipo, Andre Price, and Shadi Banitaan

Dept. of Electrical & Computer Engineering & Computer Science
University of Detroit Mercy
Detroit, MI, United States
(yacoobby, scheysej, oladipea, pricean2, banitash)@udmercy.edu

***Abstract*—Anthropogenic activities release pollutants into the air, which can negatively affect human health and the environment. One such pollutant is nitrogen dioxide (NO_2), which can contribute to smog formation, decreased crop growth and yield, and respiratory damage. This study aimed to find a relationship between land use/land cover (LULC) classifications and NO_2 levels in the air. We used Google Earth Engine (GEE) to collect LULC and air quality data using the Google Dynamic World and the Sentinel-5P NRTI NO_2 datasets. We focused on Pasadena, California, as it provided a good demonstration of an urban area surrounded by greenery, allowing for an adequate analysis of both forms of landscape and their impact on air quality. Random forest (RF) and decision tree (DT) classifiers were used on the provided datasets, with the estimated probability of complete coverage for each LULC type being the input features and the NO_2 density being the output label, measured in mol/m^2. Our output labels were then discretized, classifying the categories into high and low NO_2. The machine learning classifier found a correlative relationship between LULC and NO_2 levels, as signified by our modeled accuracy outputting a value of 85%, with an average f1 score of 86%. We performed 10-fold cross-validation to enhance the reliability of model evaluation. The results from this study suggest that machine learning models can be used to predict the changes in air quality based on changes in LULC from anthropogenic activities. With future studies confirming this relationship, inner-city green spaces may benefit mental and physical well-being.**

***Index Terms*—land use, land cover, air quality, nitrogen dioxide, machine learning, random forest classifier, decision tree classifier**

I. Introduction

Land use/land cover (LULC) is a widely studied area of research regarding mitigation or prevention solutions related to man-made structures and green spaces. In addition to understanding the impacts of anthropogenic activities, researchers often attempt to find solutions for the negative effects that can stem from these activities. Nitrogen dioxide (NO_2) has been linked to many lasting effects when present in large quantities. These effects include airway inflammation and increased risk for lung cancer [1]. While NO_2 is produced in natural combustion processes and leaves the atmosphere by rain, an excess is produced by anthropogenic activities, exacerbating the effects seen when humans and environments are exposed to this chemical. Nitric oxide (NO) is a product of gas-powered vehicles, and when released into the air, it pairs with ozone (O_3), forming NO_2. When paired with larger cities with majority car infrastructure, one can assume the NO_2 levels would be higher, thus decreasing the air quality. The most beneficial action we can take to decrease NO_2 is to use less nitrogenous fuel sources. By implementing green spaces in high-density cities, we can expect to see a decrease in NO_2 levels. The characteristics of NO_2 as an air pollutant, its sources from combustion processes, particularly those related to gas-powered vehicles, underscore the importance of investigating the correlation between LULC and NO_2 levels in our study.

Google Earth Engine (GEE) is a cloud-based planetary-scale environmental data analysis platform. It provides access to a vast archive of satellite imagery and geospatial datasets, as well as powerful analysis and visualization capabilities. GEE is designed to enable scientists, researchers, and developers to monitor and analyze changes on Earth at unprecedented scales.

GEE has several features that make it a powerful tool for environmental data analysis. These include:

- Dataset collection: GEE provides access to over 100 petabytes of satellite imagery and geospatial datasets, including Landsat, Sentinel, MODIS, and VIIRS. This data is updated regularly, providing users with the most up-to-date information on Earth's changing environment.
- Powerful analysis and visualization capabilities: GEE provides various tools for analyzing and visualizing environmental data. These tools include image

979-8-3503-3064-9/24/$31.00©2024 EEE

processing, geospatial analysis, and raster mapping. Users can use these tools to create maps, charts, graphs, and other visuals to help them understand their data.

- A built-in code editor: GEE has a built-in code editor that allows users to write and run code to manipulate, analyze, and export their data. This code can be written in JavaScript or Python.

GEE is a powerful tool that allows scientists, researchers, and developers to better understand and monitor Earth's changing environment. With petabytes of open-source datasets made available to the public, there is no end to the possibility of research that can stem from using GEE. In this study, GEE was instrumental in gathering LULC classifications and air quality data from the Google Dynamic World and Sentinel-5P NRTI NO_2 datasets, respectively. The latter provides near real-time high resolution imagery of NO_2 concentrations. This created a critical foundation that allowed investigation of the relationship between LULC and NO_2 levels in the air.

In the field of machine learning, computers utilize various techniques to learn from data and make predictions based on the information provided. Depending on the specific classification technique used, a classification model can be trained to predict categorical class labels in the form of input data points based on past observations. In exploring the relationship between NO_2 and LULC, the application of machine learning introduces a novel and impactful dimension to environmental research. The traditional methods of studying these associations often face limitations in handling complex, non-linear relationships within diverse datasets. Machine learning algorithms offer a promising avenue for uncovering nuanced connections between NO_2 concentrations and the complexities of LULC. The capacity of these models to discern subtle patterns and interactions, beyond the scope of conventional statistical approaches, brings a new level of precision to the analysis. By leveraging machine learning in this context, we aim to enhance predictive accuracy and gain insights into the dynamics governing the influence of land use/land cover on NO_2 levels, ultimately contributing to a more comprehensive understanding of environmental dynamics and possibly aiding policymakers in crafting effective environmental policies aimed at mitigating NO_2 emissions and promoting sustainable land use practices.

Classification models utilize supervised learning, a type of approach where the model learns from a dataset containing input data linked to specific output labels (apriori labeling). When these data are analyzed, the model can then make predictions for new datasets by mapping them to predicted output values. Our research uses a decision tree (DT) and a random forest (RF) classifier. A DT makes decisions based on input features, while an RF is an ensemble of such trees, combining their outputs to improve prediction accuracy through randomness and diversity. RF and DT are considered nonparametric techniques, which tend to yield better results, as they do not rely on specific assumptions or hypotheses [2]. RF is versatile and capable of handling numerical and categorical data with minimal preprocessing. The ensemble nature of RF, consisting of multiple decision trees, provides resilience against overfitting, enhancing generalization to new data and robustness against outliers. While logistic regression is straightforward, it may struggle with non-linear relationships, a challenge overcome by DT and RF. In comparison to SVM, which might be sensitive to the choice of kernel function, the simplicity and robustness of DT and RF become apparent. Lastly, RF was chosen because it has been shown to report the highest model accuracy when dealing with LULC datasets [5].

The strength and accuracy of a classification model provide the foundation for the results of an experiment. A low model accuracy output can imply either underfitting or the present lack of a relationship. In contrast, a higher model accuracy number can imply overfitting or complex patterns can be found to establish such a relationship. The model's results can be interpreted depending on the purpose of the research. We used the classification metrics of accuracy, precision, recall, and f1 score to test our hypothesis.

The purpose of this paper is to examine the relationship between land use/land cover (LULC) and nitrogen dioxide (NO_2) levels in the urban region of Pasadena, California. Utilizing Google Earth Engine (GEE) for data extraction, we use machine learning techniques, specifically a decision tree (DT) and a random forest (RF) classifier, to analyze the correlation between LULC patterns and NO_2 concentrations. By implementing these classification models, we aim to enhance predictive accuracy and gain insights into the dynamics governing the influence of land use/land cover on NO_2 levels.

The following sections provide a comprehensive examination of our study. The "Related Work" section summarizes existing research contributing to our understanding of the relationship between land use/land cover and nitrogen dioxide levels. The "Methodology" section details our study area, the data preparation process for land cover and air quality datasets, and the methodologies used, including sampling strategies and data distribution across attributes. The "Results" section presents our findings, and the "Discussion" section interprets and contextualizes these findings. The "Conclusion" section summarizes our key insights and contributions, and the "Future Work" section outlines potential avenues for future research.

II. Related Work

In recent years, satellite remote sensing has been a growing field that offers beneficial information for understanding and visualizing the surface of our planet. Many tools are available to the public that provide access to satellite data and a way to analyze the data. As mentioned, we decided to use GEE for our satellite imagery analysis due to its easy-to-use code interface, the ability to run complex algorithms, and the free access to petabytes of datasets [2]. The use of ML models in GEE, such as RF classifier or support vector machine (SVM) classifier, is relatively common [3].

Oo et al. [4] implement a maximum likelihood algorithm as their supervised classification approach. In doing this, they can determine the maximum for a given statistic from a known class of distributions. While also using a supervised classification technique, we decided to use RF and DT classifications, as they are more applicable to the project.

As seen in a study by Talukdar et al. [5] and Gao et al. [6] on LULC classification by machine learning classifiers for satellite observations, we see RF algorithms have been widely applied for solving environmental problems, such as water resource management and natural hazard management. It has also been shown to be beneficial when used in satellite imagery analysis because it is a combination of ensemble regression and classification trees. Our study which uses satellite imagery datasets was built off of RF classification to evaluate the relationship between LULC and NO_2 levels regarding air quality.

Moreover, Prasai et al. [7] make use of a confusion matrix to evaluate overall model accuracy. This allowed an assessment of the user's accuracy (the number of correctly classified pixels divided by the total number of pixels predicted within that LULC class) and the producer's accuracy (the number of correctly classified pixels divided by the total number of pixels truly in that LULC class). We used a similar approach, evaluating the precision, recall, and f1 score, which makes use of a confusion matrix, allowing us to have a basis for properly evaluating our model.

The need for research on preventable measures against the ever-growing air quality crisis is addressed in Zou et al. [8]. Their project addressed the impacts of LULC on air quality in an urban setting and focused on PM_{10}, referring to inhalable particles with a diameter of 10 micrometers or less, and found that there is a decrease in PM_{10} concentrations in newly developed built-up areas. This prompts questions about how large of an effect urban development and LULC classifications can have on air quality. A reason why this paper strives to answer this question is the potential policy advocacy for more green space development and preservation.

III. Methodology

This section elaborates on our research methodology, which includes the study area, data preparation, and the methods used in our study. We focus on Pasadena, California, as our region of interest because it is an urban center with a dense population and green spaces. This allows us to understand the dynamic relationship between urban development and natural elements. We use GEE to prepare land cover datasets and air quality datasets. We then describe our sampling strategies and data distribution across attributes, which sets the stage for the subsequent subsections that delve into each aspect of our approach.

A. Study Area

Our research was focused on Pasadena, California. Due to its dense population, topographical features, and heavy transportation, California has built a reputation for being one of the most polluted states in the United States. Pasadena, a city in the Los Angeles metropolitan area, illustrates the environmental issues faced by urban centers. It presents a clear representation of urban areas surrounded by green spaces, which can help us highlight the dynamic relationship between urban development and the presence of natural elements. Our inclusion of both urban and green areas aims to emphasize the relationship between human-made structures and natural environments on the presence of nitrogen dioxide in the air. Using GEE, we selected our specific region of interest in the form of a coordinate polygon:

$$\begin{array}{l}(-118.22580057194779, 34.062493521042164),\\ (-117.91269022038529, 34.062493521042164),\\ (-117.91269022038529, 34.31354553281858),\\ (-118.22580057194779, 34.31354553281858),\\ (-118.22580057194779, 34.062493521042164)\end{array} \quad (1)$$

As seen in Fig. 1, these coordinates map out our region of interest for this section of Pasadena, California.

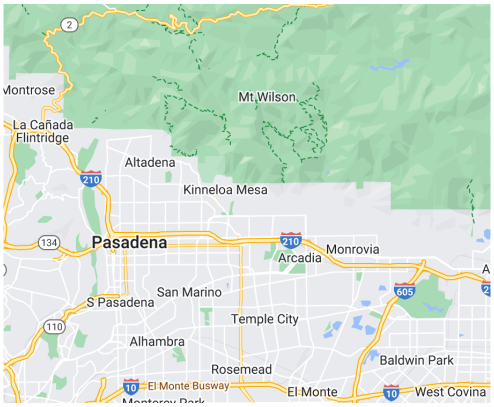

Fig. 1: Region of Interest.

B. Land Cover Datasets

This research uses the Google Dynamic World Dataset - GOOGLE/DYNAMICWORLD/V1. Dynamic world data is gathered and processed in near real-time, providing a continuous stream of LULC predictions produced using deep learning techniques on 10m Sentinel-2 imagery [9]. The data retrieved was filtered monthly from January 1, 2021, to December 1, 2021, for a total of 199 images. This data was also restricted to our region of interest using the "filterBounds" function provided by Google Earth Engine.

Each month, multiple satellite scans, or images, are taken. Each image contains 10 bands of raster layers, including information such as the estimated probability of complete coverage by water, crops, trees, and so on, with all probability bands adding up to 1. To reduce the size of these images, we calculated the mean of the values for each band at every pixel across the entire month. This resulted in one image per month range, which allows us to properly sample the pixel values for each band to produce our data.

The "sampleRegions" function provided by GEE is a powerful tool that can be used to reduce the size of large images or to extract data from specific areas of an image. It is an essential tool for working with satellite imagery, as it can help to improve the efficiency and performance of image processing tasks. We used the "sampleRegions" function to sample pixel values from the given satellite scans based on the scale and collection given. We used a scale of 100, and we specified the collection as the region of interest, which meant that we only sampled pixels from the area of interest. This gave us the band values from the sampled pixels for the image retrieved per month.

Using this sampling strategy, we obtained 97,440 elements, or pixel values, that contained the aforementioned probability bands for each month range. For example, we obtained 97,440 scans for January–February, February–March, and so on. This gave us a total of 1,071,840 elements. We then exported this data to Excel using GEE's export feature. Each element was a row, with the appropriate land cover classification as the column and the estimated probability of complete coverage by those classifications as the data contained within those columns. As discussed later, this data was limited by the Excel row limit, allowing us to use only 1,048,576 elements.

Lastly, we deleted the label column provided by the Dynamic World dataset, as the averaging process and alterations to the order of the row made it unreliable.

C. Air Quality Datasets

The Sentinel-5P NRTI NO_2 dataset was used to provide near real-time high-resolution imagery of nitrogen dioxide concentrations. This chemical compound is produced mainly through anthropogenic activities, such as fossil fuel combustion and biomass burning [1]. This research used the "NO2_column_number_density" band provided by this dataset, which represents the total vertical column of NO_2 (ratio of the slant column density of NO_2 and the total air mass factor, measured in mol/m^2). As shown in Fig. 2, a histogram is shown graphing the frequency of occurrence for each record present and categorizing them into two bins. By having two bins, we create a binary classification task with two labeled outputs to be "Low NO_2 Levels" and "High NO_2 Levels". Before preprocessing, the air quality dataset used the equal width discretization technique, but it did not produce good findings because most of the dataset lay in the "Low NO_2 Levels", making it the majority class and potentially biasing the model in its predictions. This issue was causing our other label to output an f1 score that was not on par with its counterpart, 0.85 compared to 0.65, respectively, to the classified labels mentioned above. A proposed solution was incorporating custom weights for each label, with the minority class having a higher weight. When the model was trained, the intention would be to mitigate that bias and allow the smaller bin to be equally chosen as the majority class. However, that might introduce potential problems that we could not foresee. Therefore, we changed our approach, considered equal frequency, and still have it as a binary classification problem. After doing so, the bins would have an equal amount of records, although the range would be dynamic to account for half of the dataset in each respective bin. Our f1 score for both output labels was almost the same, which was 0.86, indicating that the approach had worked in our favor and had better results than the former method.

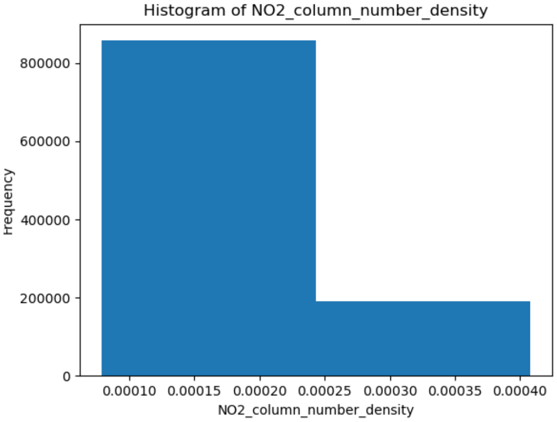

Fig. 2: Discretizing the NO_2 Attribute.

D. Sampling Strategies

The Dynamic World dataset provided classification based on ecosystem types, which was divided into 9 categories: "crops", "trees", "grass", "shrubs and scrubs",

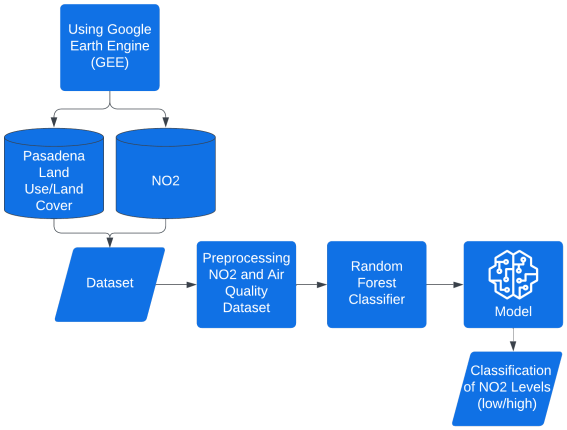
Using Google Earth Engine (GEE)
Pasadena Land Use/Land Cover
NO2
Dataset
Preprocessing NO2 and Air Quality Dataset
Random Forest Classifier
Model
Classification of NO2 Levels (low/high)

Fig. 3: The Proposed Framework.

"flooded vegetation", "water bodies", "built-up areas", "bare land", and "snow and ice". This data was aggregated with the Sentinel-5p NRTI NO_2 dataset, with each pixel value for each band corresponding with a pixel on the "NO2_column_number_density" band.

Using the GEE "sampleRegions" function, sample points (pixels) were taken from our region of interest, which spanned areas of natural land and high human activity. Using Python's "Pandas" library, we used its "qcut" function to bin the numerical data of nitrogen dioxide evenly into discrete intervals.

Random forest is an ensemble learning technique that builds multiple decision trees from randomly sampled data and features. The final prediction is determined by a majority vote in classification or an average in regression. RF was run on the extracted data, explained in the results section. Fig. 3 highlights a summary of the approach.

E. Data Distribution Across Attributes

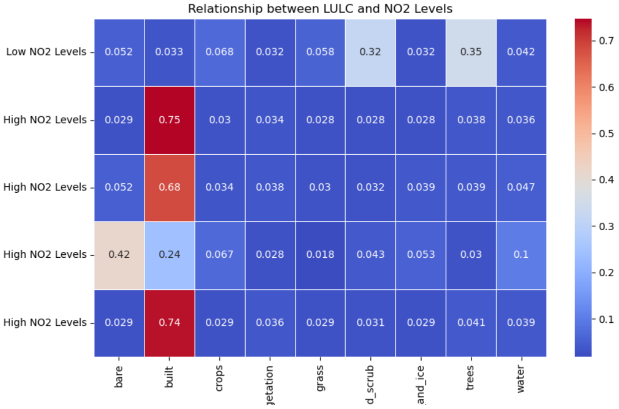

Fig. 4: Feature-Label Association Heatmap.

To understand the relationship across our respective categorical dimensions, a heatmap, as shown in Fig. 4, was utilized to visually interpret the intensity of data points in our matrix. Each cell within the matrix corresponds to an intersection of our input features compared to our discretized output labels. The color of each cell indicates the magnitude and intensity of the association between the input and target label. The intensity between the input features and the target labels was substantial enough to further experiment with this approach. Eventually, we have come to integrate different classifiers to validate this association.

F. Classifiers Chosen

Random forest (RF) and decision tree (DT) classifiers were chosen for this study due to several reasons. Firstly, they are non-parametric techniques that do not rely on specific assumptions or hypotheses, which tends to yield better results for complex data like LULC and air quality datasets [2]. Not to mention, RF is an ensemble learning method that combines multiple decision trees, making it robust against overfitting and outliers compared to single classifiers. Also, previous studies on LULC classification using machine learning have shown that RF often reports the highest model accuracy [5]. Finally, RF can handle both numerical and categorical data with minimal preprocessing, making it well-suited for our diverse feature set.

IV. Results

We ran a decision tree and a random forest classifier on our data. The DT had an accuracy of 85%, while the RF classifier had an accuracy of 86%. We also calculated the precision, recall, and f1 score values (Equations 2, 3, and 4), both popular classification metrics used to evaluate various aspects of model performance, including accuracy. Besides the two classifiers we fine-tuned, an external tool, "LazyPredict" was utilized to find other classifiers that could perform better. The results of these classification metrics are shown in Fig. 7. The consistency between the identified high-performing models by "LazyPredict" and our DT and RF classifiers reinforces the validity of our choices. Moreover, we measured model accuracy using the cross-validation technique, which is a resampling method that uses k sections of the data to test and train a model on each iteration of k. Using k = 10, or 10 iterations, the models achieved a mean accuracy of 86%. The formulas for each metric are displayed below, while Fig. 5 and Fig. 6 illustrate the results, respectively. The metrics for the outsourced tool mentioned above align with the consistency of the classifier that was best for this dataset.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

The classifier configurations for which the model produced this precision can be seen in the following for both DT and RF. Various experiments were conducted to finetune the optimal hyperparameters for the two classifiers. GPU configurations were unavailable using the packages associated with these classifiers, so training the models takes some time. The estimated time to have a trained model for RF can be decreased if “n_jobs” parameter is changed to -1, helping to use all available CPU cores to speed up the training process.

```
DecisionTreeClassifier(criterion='gini
    ', splitter='best', max_depth=10,
    min_samples_split=10,
    min_samples_leaf=5, max_features=
    None, class_weight=None)
```

```
RandomForestClassifier(n_estimators
    =100, criterion='gini', max_depth=
    None, min_samples_split=2,
    min_samples_leaf=1,
    min_weight_fraction_leaf=0.0,
    max_features='sqrt',
    max_leaf_nodes=None,
    min_impurity_decrease=0.0,
    bootstrap=True, oob_score=False,
    n_jobs=None, random_state=None,
    verbose=0, warm_start=False,
    class_weight=None, ccp_alpha=0.0,
    max_samples=None)
```

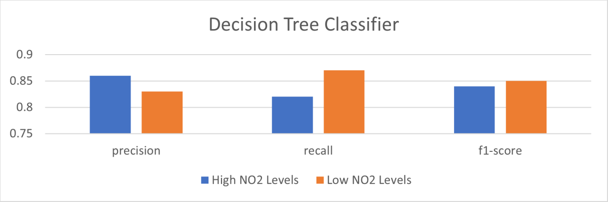

Fig. 5: Classification Report for Decision Tree Classifier.

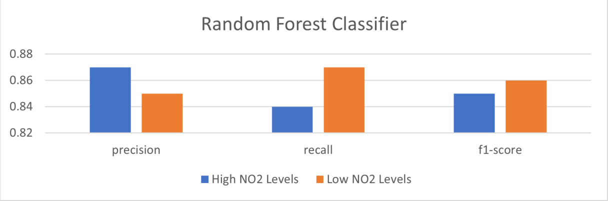

Fig. 6: Classification Report for Random Forest Classifier.

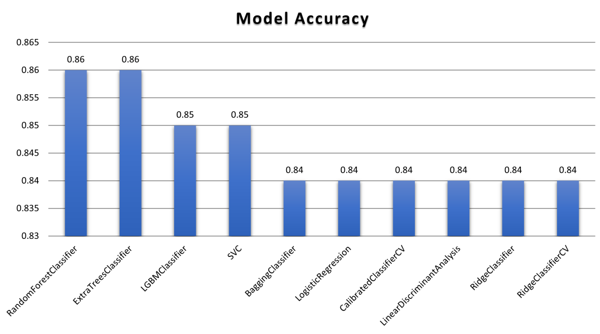

Fig. 7: Classifier Performance Comparison Based on Accuracy.

Using a feature importance score, we determine the percentage that each feature contributed to our final model. This can be seen in Table I.

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

V. Discussion

An accuracy of 85% can lead us to strongly conclude that there is a strong association between LULC and air quality, specifically the presence of nitrogen dioxide. This also means that we can use the generated model to predict how changes in land cover over time due to deforestation, urban development, or wildfires can affect pollution. As nitrogen dioxide is mainly produced through anthropogenic activities [1], we believe that the link between an increase in urban landscapes and an increase in NO_2 is logical, as anthropogenic activities present themselves mostly in highly urban areas, which contain factories, numerous means of transportation, workplaces, and so on.

These technological advances, despite improving different parts of our quality of life and standard of living, can also contribute to a lower quality of living through dangers like reduced air quality can lead to respiratory damage, lower crop yields, acid rain, smog, etc [1]. As more companies begin to place a focus on carbon neutrality and to phase out the use of fossil fuels, this can decrease the presence of NO_2 in the air. In urban landscapes, incorporating green spaces with trees can effectively counterbalance air pollution by allowing trees to absorb nitrogen dioxide, leading to improved air quality and mitigated negative effects [10], [11].

We speculate that future research can incorporate bigger regions of interest, allowing for better and broader distinctions between LULC areas. Our data was extracted using a scale of 100, and we are curious to see what changes or improvements in data quality result from increasing the scale of our data. We would also be interested in seeing the relationship between increased NO_2 presence in the air and crop yields in large-scale farming units surrounded by urban areas.

Research has shown that greenery in urban settings is linked to improved mental health and cognitive function. Studies have also shown that people who spend time in green spaces have better attention, memory, and problem-solving skills [12]. This is likely because nature can provide a stimulating environment that encourages us to be active and engaged. This research can contribute to the growing body of research on the relationship between LULC, air quality, pollution, and how the constantly evolving urban landscape can continue to thrive by incorporating these green spaces.

Regarding the replicability of the findings obtained using GEE, it is important to remember that the selected coordinates were crucial because they offered a strong distinction between natural landscapes and human-made structures in an urban environment. Those who wish to replicate our findings in the future, specifically using different cities, should bear in mind a region of interest comprised solely of vegetation can result in the model not providing valuable data, as the majority of the data is skewed toward one land type, disregarding the other nine types. The same can be said for a region of interest comprised solely of built land or water.

VI. Conclusion and Future Work

Through our findings, we tested two machine learning classifiers to establish a relationship between our targeted attributes, LULC and NO_2 levels. We implemented random forest and decision tree classifiers, which attempted to find complex patterns between our targeted attributes through model training and validation. From our experiments, we found that the difference in these model accuracies outputted was negligible. Our accuracy was around 0.85, with a mean cross-validation score of 0.86. We found that having 10 subsets of the dataset through cross-validation was appropriate given our dataset's large amount of data, close to 1.05 million instances. Our approach implemented these classifiers on a specific region of interest in California, that being the city of Pasadena. Our codebase utilized the results from Google Earth Engine to supplement our machine learning research and answer our goals and questions. The results of our experiment shed light on the importance of continuing LULC and air quality research and how their association can be applied to solutions designed to counteract the negative effects of anthropogenic activities.

With an accuracy of 85%, we can conclude that there is a strong correlation between LULC and air quality. This brings into question the utility of this research such as how could the results stemming from this research be used for the environment and environmental policy? The results, showcasing human structures emitting NO_2 into the atmosphere, can be utilized by experts, scientists, and lawmakers to determine whether any further restrictions over current environmental regulations should be implemented alongside creating more designated green spaces.

Although we were able to provide a machine learning classifier that addressed our research goals, we encourage those who are reading to contextually apply our findings to the California standard for NO_2 levels, either by unit conversion or some other technique, to further contribute research to the relationship between LULC and its relationship with the presence of NO_2. It benefits research to indicate the relevance of these patterns found by the classifier to validate their results. These findings would not only serve as evidence of the model's general usability in environmental awareness of what impacts air quality but also potentially contribute to the reduction of it. Another avenue of exploration that can be considered is finding other regions of interest similar to the one analyzed in this article to give more insight into that chosen region's environment and distinguish within that region its different types of land classifications. The scalability of our system can be demonstrated by the results of our codebase, which can be adapted to different types of data and assert with strong confidence that a relationship is present between LULC and air quality. The source code for this paper can be found here: https://github.com/Benyamain/lulc-air-quality.

References

[1] Ritz et al., "The Effects of Fine Dust, Ozone, and Nitrogen Dioxide on Health," *Deutsches Arzteblatt International*, *51-52*(51-52), 881–886, 2019, doi: https://doi.org/10.3238/arztebl.2019.0881.

[2] Vizzari et al., "PlanetScope, sentinel-2, and sentinel-1 data integration for object-based land cover classification in google earth engine," *Remote Sensing*, vol. 14, Art. no. 11, 2022, doi: https://doi.org/10.3390/rs14112628.

[3] Yang et al., "Google earth engine and artificial intelligence (AI): A comprehensive review," *Remote Sensing*, vol. 14, Art. no. 14, 2022, doi: https://doi.org/10.3390/rs14143253.

[4] Arunrat et al., "Comparing four machine learning algorithms for land cover classification in gold mining: A case study of kyaukpahto gold mine, northern myanmar," *Sustainability*, vol. 14, Art. no. 17, 2022, doi: https://doi.org/10.3390/su141710754.

[5] Talukdar et al., "Land-use land-cover classification by machine learning classifiers for satellite Observations—A review," *Remote Sensing*, vol. 12, Art. no. 7, 2020, doi: https://doi.org/10.3390/rs12071135.

[6] Gao et al., "Detailed and automated classification of land use/land cover using machine learning algorithms in Google Earth Engine," *Geocarto International*, vol. 37, Art. no. 18, 2022, doi: https://doi.org/10.1080/10106049.2021.1917005.

[7] Prasai et al., "Application of Google earth engine python API and NAIP imagery for land use and land cover classification: A case study in Florida, USA," *Ecological Informatics*, vol. 66, p. 101474, 2021, doi: https://doi.org/10.1016/j.ecoinf.2021.101474.

[8] Zou et al., "Effect of Land Use and Cover Change On Air Quality in Ubran Sprawl," *Sustainability*, vol. 8, no. 7, p. 677, 2016, doi: 10.3390/su8070677.

[9] Salvatore et al., "Integrated use of Sentinel-1 and Sentinel-2 data and open-source machine learning algorithms for land cover mapping in a Mediterranean region," *Remote Sensing*, vol. 55, Art. no. 1, 2022, doi: https://doi.org/10.1080/22797254.2021.2018667.

[10] Gong et al., "Assessment of NO_2 Purification by Urban Forests Based on the i-Tree Eco Model: Case Study in Beijing, China," *Forests*, 2022, doi: https://doi.org/10.3390/f13030369.

[11] Zhang et al., "Improving air quality by nitric oxide consumption of climate-resilient trees suitable for urban greening," *Frontiers in Plant Science*, vol. 11, 2020, doi: https://doi.org/10.3389/fpls.2020.549913.

[12] Pasanen et al., "Urban green space and mental health among people living alone: The mediating roles of relational and collective restoration in an 18-country sample," *Environmental Research*, vol. 232, 2023, doi: https://doi.org/10.1016/j.envres.2023.116324.