USING STREET VIEW IMAGES FOR AIR QUALITY INDEX PREDICTION MODELLING USING DEEP AND MACHINE LEARNING

Seung Jun Choi

Community & Regional Planning
School of Architecture
The University of Texas at Austin

ABSTRACT

Air pollution has significant adverse health effects. Recently, evaluating air quality became one of the most significant tasks for planners with rising concerns of environmental challenges due to climate change. Microsoft has been installing an IoT sensor named Project Eclipse to measure real-time climate pollutants in Chicago. Land use and built environment effects on air quality have been long questioned pertinent to air quality prediction modeling. Here, we incorporated street view images. Conventionally, land use inventory and satellite images were used. Scene features are extracted from it. Different land use features were dimensionally scaled using principal component analysis. The scaled feature score was used for air quality index prediction modeling. We modeled linear, ensemble machine learning models, long-short term memory (LSTM) neural networks and computed them. The analysis results show that scene features extracted from street view images share similar aspects to conventional land use features. For air quality prediction, LSTM outperformed. Street view images have the potential to improve climate prediction modeling.

1. INTRODUCTION

Air pollution has significant adverse health effects such as respiratory and cardiovascular diseases, cancer, and even death [1]. 3 billion people die due to outdoor air pollution, according to World Health Organization (WHO) report in 2012. Evaluating air quality and exploring the relationship between building environment and air quality has become critical concerns. Urban built infrastructures and urban inhabitants are related to urban air quality [2]. Urban environments provide researchers and urban planners with indicators to evaluate and predict urban air quality. Three data sources can be considered to get land use features; 1) the scene features extracted from google street view images; 2) urban land use features extracted from satellite images; and 3) land use inventories/building footprints.

Our research focused on testing these three data sources for urban air quality evaluation and prediction. Urban scene features are extracted from street view images, such as official buildings, parking lots, and streets. These features are then classified into different land use features and compared with features extracted from satellite images and land use maps. Ground truth air quality is gathered by Project Eclipse Microsoft sensors in Chicago [3].

Research questions include the following:

(1) Propose a framework to incorporate street view images for air quality prediction modelling.
(2) Evaluate scene features extracted from street view images with conventional land use inventory and satellite images.
(3) Compute different machine learning and deep learning models on microscopic IoT sensor-based air quality prediction

2. BACKGROUND

In the year 2020, the Urban Innovation team in Microsoft partnered with JCDecaux and Array of Things to deploy over 100 low-cost air pollution sensors across Chicago. The custom-designed air quality sensor can measure CO2, NO2, SO2, and O3, PM1, PM2.5, and PM10, along with temperature, humidity, and pressure [3]. It also offers AQI (Air Quality Index). The greater the value, the greater the likelihood of a health concern. Data is publicized as an interactive map. Using OpenAPI, users can pull out Project Eclipse data. The device is powered by solar and can collect data at 60 seconds and 15 minutes intervals. These sensors are citywide coverage using a stratified random sampling approach and are equitable coverage with oversampling in environmental justice neighborhoods [3]. The sensor installation is named Project Eclipse. As of August 2021, 110 Project Eclipse sensors are installed throughout the city of Chicago. Figure 1 describes the deployment of Project Eclipse sensors and compares the average hourly temperature of 9 AM and 9 PM of August 2021. It shows that different regions experience variance even if they are close to one another.

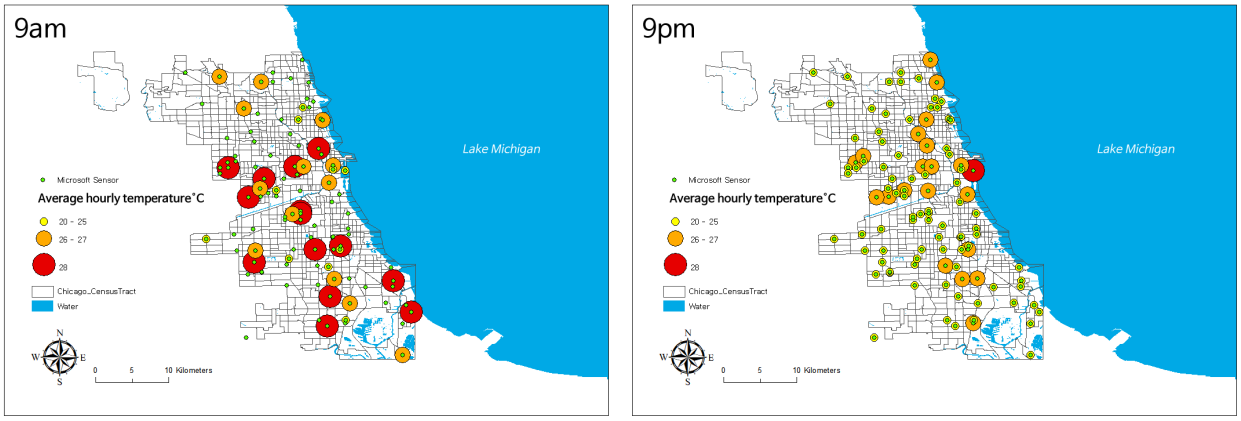


Figure 1: Average hourly temperature in Chicago in 9am v. 9pm of August

Google street view (GSV) images provide us with an efficient source to recognize our urban built environment, such as landscape assessment [4-6], demographic makeup [7], neighborhood environment [8-10], and urban development [11]. Neural Networks, such as YOLO and ResNet, are widely used to detect objects [12-14] and image classification [15-17]. These classification models efficiently extract urban-built information features from GSV.

Google Street View static API provides a powerful tool to download street view images automatically. The necessary parameters are location (latitude and longitude value of the image) or pano (the specific panorama ID), output size of the image, key of API, and other optional parameters, such as the heading (compass heading of the camera), fov (horizontal field of view of the image) and pitch (the up and down angle of the camera relative to the street view vehicle).

Conventionally, land use inventory and Landsat based satellite images were used for modeling air quality prediction. Land features are generally used to see its effect on improving or worsening the air pollution issues. However, the potential of incorporating street view images pertinent to IoT sensor air quality data has not been deeply explored.

3. DATA & VARIABLES

3.1. LAND USE FEATURE SCORE CALCULATION

Four Google Street View (GSV) Imageries at heading angle of 0°, 90°, 180° and 270° for each street node are generated using the Google Street View static API. We downloaded street view images (no early than the year 2019) pertinent to Project Eclipse sensor locations in Chicago. Figure 2 illustrates some examples of our downloaded street view images.

For each node, we useed Resnet50 to extract scene features from GSV. The structure of our Resnet is shown as Figure 3. The Resnet50 is pretrained by the Places365 dataset, which classifies images into 365 categories of scenes, such as highway, forest, field, street, church, plaza and so on. The weight of the Resnet50 model is trained by 18 million images and predicts the class of each image with high accuracy. Features in all heading angles are generated, $X_i' = \left[X_{i,90^\circ}', X_{i,180^\circ}', X_{i,270^\circ}' \right]$, where X_i' is the list of features at each of the four heading angles of Node i. Then, the average value of each location is calculated, $X_{i,j} = \frac{1}{4}\left(X_{i,0^\circ,j}' + X_{i,90^\circ,j}' + X_{i,180^\circ,j}' + X_{i,270^\circ,j}'\right)$, where $X_{i,j}$ is the feature of Node i at dimension j, and $i \in N, j \in D$. We use the Resnet50 to extract scene features because it employs an identity shortcut connection mechanism to skip one or more layers when training the model, so that more layers could be included in the model resulting in higher accuracy. Resnet50 requires relatively fewer parameters and less computational cost to extract features from images than other methods.

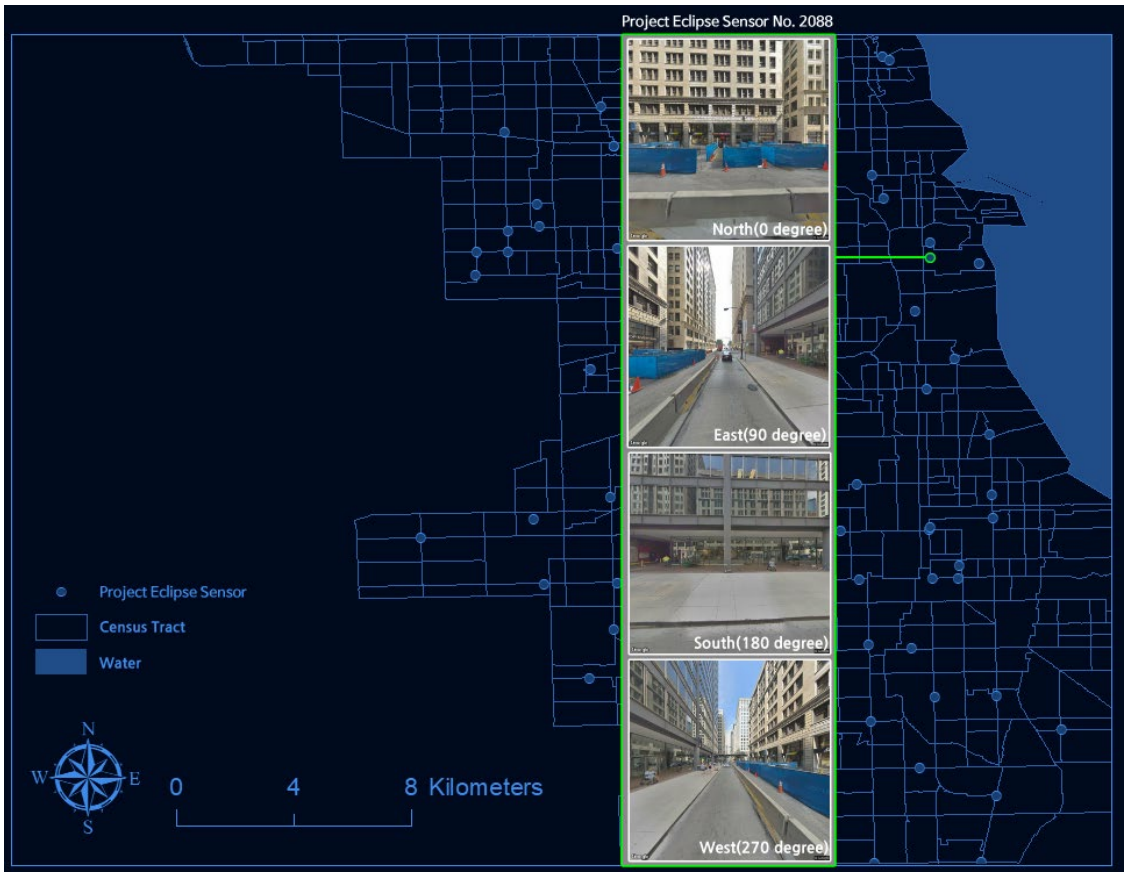


Figure 2. Examples of our downloaded street view images

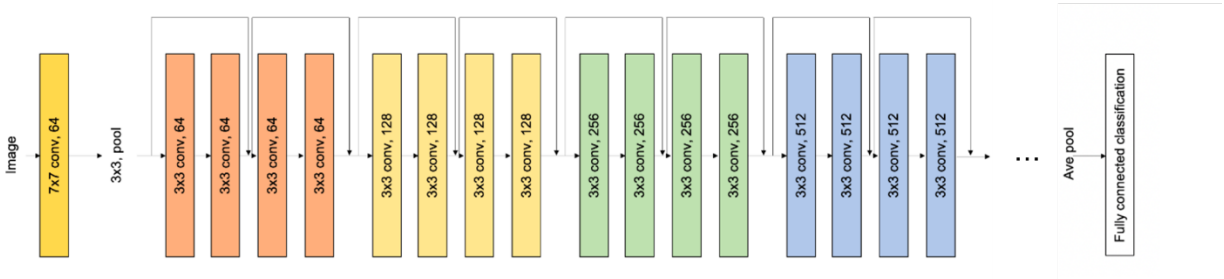


Figure 3. Structure of Resnet50

In total, 365 categories of the scene are extracted. Each category contains the probability ratio. We classified the probability pertinent to related land-use types. We measured four land-use types; residential, green open space, transportation, and office. Residential land use type refers to the probability of apartment building. Green open space is the sum of probability of park, yard, roof garden, botanical garden, and formal garden. Transportation land-use types contain bridges, bus stations, crosswalks, driveway, parking garage, parking lots, railroad track, street, and highway. Office land-use types include office, office buildings, and office cubicles.

3.2. CONVENTIONAL LAND USE FEATURES EXTRACTION

The land use inventory map and satellite image were reprocessed in GIS (Geographic Information System). For technical support, we used ArcMap 10.8.

We used the land use inventory map of 2015 in Chicago. It was obtained from the government of Illinois. First, we created a buffer of 400 meters of each deployed Project Eclipse sensors. Four hundred meters is a walking distance of 5 minutes. Second, we intersected buffers with land use maps. Third, the ratio of land use types was measured by dissolving the intersected layer by Project Eclipse sensors. Fourth, we summed up the calculated land use types ratio by primary land use type based on zoning and building codes. It includes urbanized (1000), agriculture (2000), Open Space (3000), Vacant (4000), Water (5000), Non-Parcel Area (6000), 9999 (Unknown).

For satellite image extraction, we used LandSat8 of September 2021. It was the latest one under condition, for having the cloud in minimum (less than 10%) and during daytime. ArcMap offers supervised classification. Using training samples through the image classification toolbar it measures the classes and their statistics pertinent to training samples [18]. Using band combination (also named as NIR (Near Infrared) composite) in LandSat8, we classified the land use types by waterbody, vegetation, dense settlement, and barren land as described in Figure 4. Then, similar

to land use inventory extraction, the land use classification was clipped to 400-meter buffers of deployed Project Eclipse sensors.

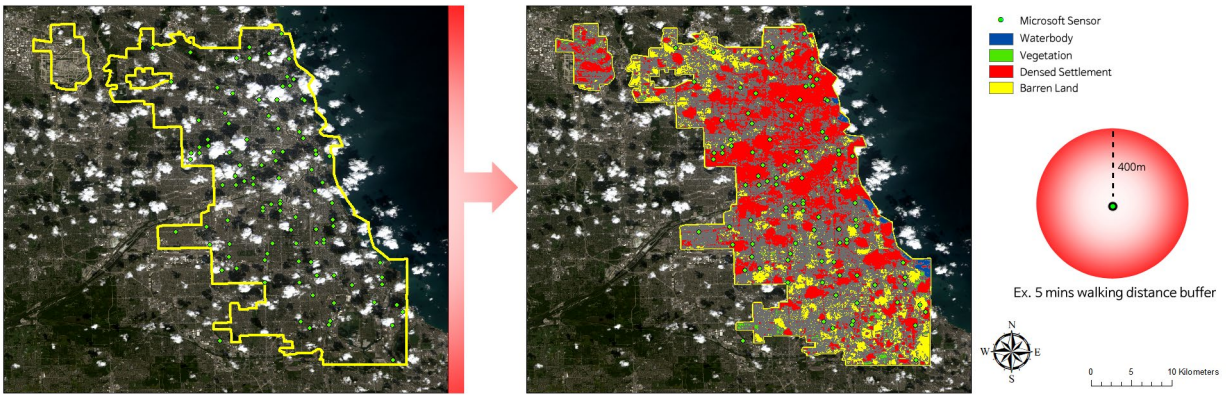


Figure 4. Supervised land use classification using LandSat8

3.3. VARIABLES

The three land use features are extracted via three different data; street view, land use inventory, and satellite image. However, as the three models are from the same city of Chicago, the issue of multicollinearity needs to be resolved. Therefore, we calculated the VIF score and trimmed the variables. The chosen land use features have a VIF score lower than five, as shown in Figure 5.

	feature	VIF
0	SV_Residential	2.284813
1	SV_GreenOpenSpace	1.780560
2	SV_Transportation	3.423327
3	SV_Office	2.347830
4	LU_2000_Agriculture	1.074981
5	LU_3000_OpenSpace	1.721780
6	LU_4000_Vacant_UnderConstruction	1.689181
7	LU_5000_Water	1.097989
8	SA_Waterbody_1	1.582868
9	SA_Vegetation_7	2.780150
10	SA_BarrenLand_45	2.195629

Figure 5. VIF score measurement

For the dependent variable, hourly daily AQI measured by Project Eclipse sensors throughout the whole month of August 2021 were used. Project Eclipse sensors provide data every 15 minutes. The daily hourly average AQI was calculated using Python. However, some sensors react as an outlier and troubleshoot, for falling short when compared with neighboring sensors, street view images are unavailable or located outside the boundary of the city of Chicago. To remove the outliers, we removed six Project Eclipse Sensors. Among 110, a total of 104 sensors were used. Then, we took a log on the dependent variable. Figure 6 shows a normalized distribution of AQI. And Table 1 discerns descriptive statistics.

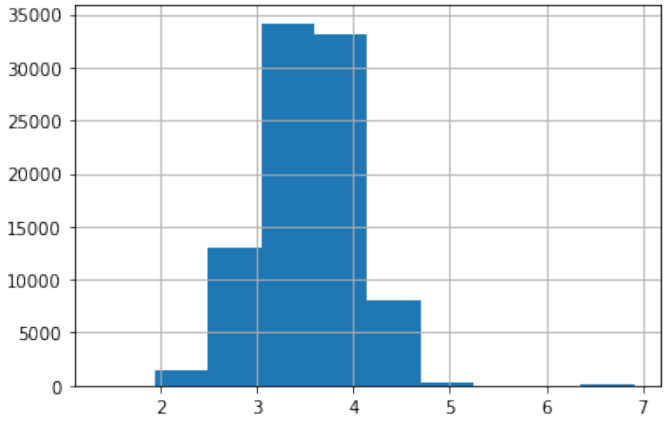


Figure 6. Normalized distribution of dependent variable

Table 1. Descriptive Statistics

	AQI	SV_Residential	SV_GreenOpenSpace	SV_Transportation	SV_Office	LU_2000_Agriculture	LU_3000_OpenSpace	LU_4000_Vacant_UnderConstruction	LU_5000_Water	SA_Waterbody_1	SA_Vegetation_7	SA_BarrenLand_45
count	90230	90230	90230	90230	90230	90230	90230	90230	90230	90230	90230	90230
mean	37.75004	0.002893	0.007879	0.474793	0.009065	0.000057	0.048151	0.119434	0.000847	0.000055	0.060326	0.235401
std	30.17454	0.012303	0.014656	0.176217	0.015289	0.000668	0.125962	0.115345	0.006388	0.00048	0.057998	0.182722
min	3	0.000018	0.000195	0.025461	0.000085	0	0	0	0	0	0	0
25%	24.33333	0.000246	0.000724	0.347681	0.001248	0	0	0.015316	0	0	0.017796	0.087875
50%	34.16667	0.000256	0.003019	0.55048	0.001355	0	0.002475	0.066517	0	0	0.033821	0.250491
75%	45.45455	0.001267	0.003558	0.552716	0.009544	0	0.03443	0.291555	0	0	0.088581	0.281009
max	1000	0.125361	0.085527	0.916433	0.072615	0.00791	0.905643	0.308502	0.072821	0.005145	0.325831	0.985334

The study period is the month of August 2021. The unit of analysis is each deployed Project Eclipse sensors. The time of analysis is an hour per daily basis.

4. METHODS

Eleven land use features from scene features, land use inventory, and satellite images were scaled using PCA. We name the scaled feature variances as land feature scores. Rather than using the raw data, a land feature score was used for AQI prediction. Similarly, Amato et al. (2021) argue that interpolation of PCA for Spatio-temporal prediction of environmental data can be a valuable source [19]. Figure 7 summarizes the typology.

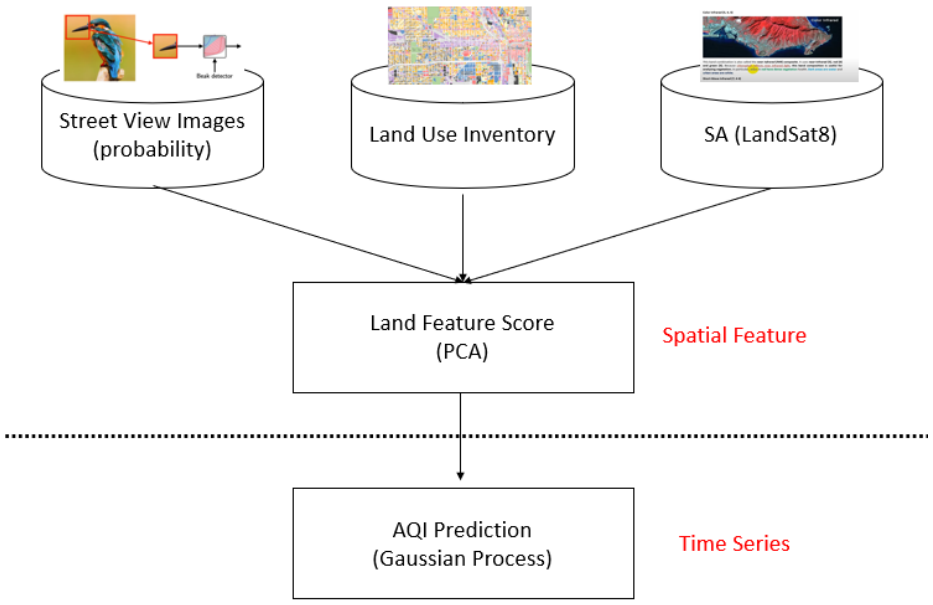


Figure 7. Study typology

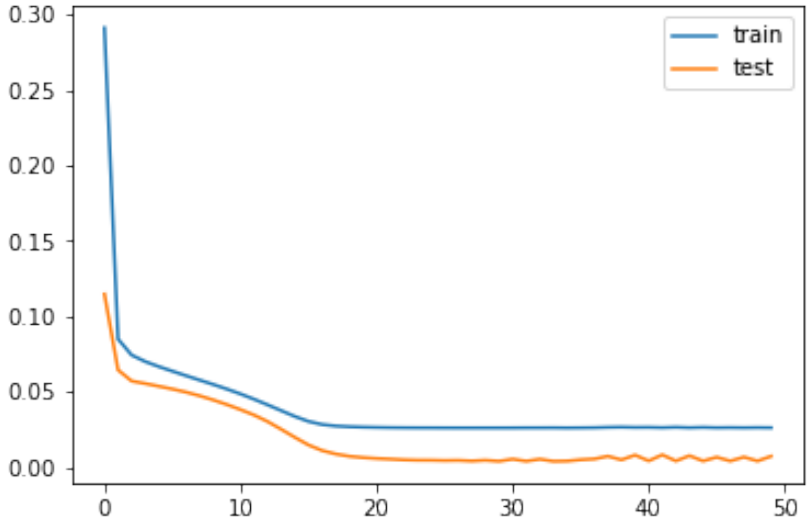
Conventionally, the neural network was most preferred in air quality prediction modeling and followed regression, ensemble, and hybrid models [20]. We built linear regression, ensemble machine learning models, and a Long Short-term Memory neural network. Ensemble machine learning regression models includes bagging-based Random Forest (RF), and boosting-based GradientBoosting (GB), XGBoost (XGB), and LightGBM (LGBM). LSTM is a deep learning model that uses RNN (Recurrent Neural Network) structures; however, it adds additional gates refer to forget gate with input and output gate. Recent studies argue that LSTM tends to outperform conventional RNN. The models were computed by calculating RMSLE (Root Mean Squared Log Error).

Ensemble models went through an optimization process using GridSearch modules. We tested several estimators from 10 to 100 and calculated a Negative Mean Absolute Error (NMAE). The NMAE closer to 0 indicates the better the model is. It was found that a number of estimators as 90 works the best and have the minimum NMAE of (-)0.3691. 80% of data was

used for training, and 20% was used for testing. Similarly, linear regression uses 80% of data for training and 20% of data for testing.

LSTM neural networks use 12 inputs and predict one output (AQI-log). We set 50 epochs and 2,496 batch sizes, transmitting one daily Project Eclipse data (104 sensors * 24 hours). Within one month, 25 days of data (104 sensors * 24 hours * 25 days) was used for training, and the rest were used for testing. Train loss and test loss were compared, and as described in Figure 8, the model is slightly overfitting. For ensemble learning models and linear regression machine learning models, we used Scikit-learn modules. LSTM neural networks use Keras.

Figure 8. Train loss. v. test loss in LSTM



5. RESULTS

5.1. PCA ANALYSIS

Figure 9 discerns the eigenvalue, contribution, and cumulative contributions of eleven PCA variances. All eigenvalues of PCA were lower than 1; however, four or five PCA components soften the slope of explained variances. Figure 10 describes the correlation matrix of PCA score with original independent variables.

	Eigenvalue	Contribution	Cumulative Contribution
pca1	4.596612e-02	0.399788	0.399788
pca2	3.864558e-02	0.336118	0.735906
pca3	1.971513e-02	0.171471	0.907377
pca4	6.689958e-03	0.058186	0.965563
pca5	3.274214e-03	0.028477	0.994040
pca6	3.410854e-04	0.002967	0.997007
pca7	1.808450e-04	0.001573	0.998580
pca8	9.285386e-05	0.000808	0.999387
pca9	6.968442e-05	0.000606	0.999993
pca10	5.655121e-07	0.000005	0.999998
pca11	1.980885e-07	0.000002	1.000000

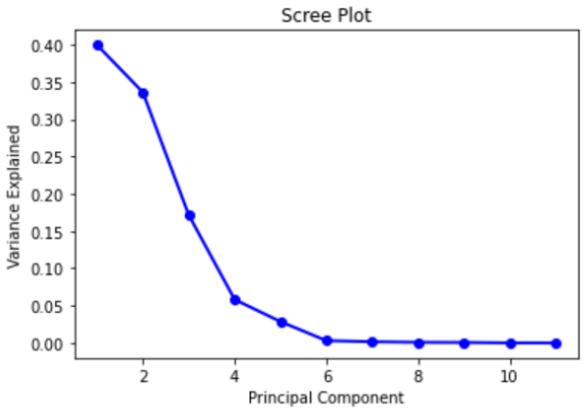


Figure 9. PCA variances

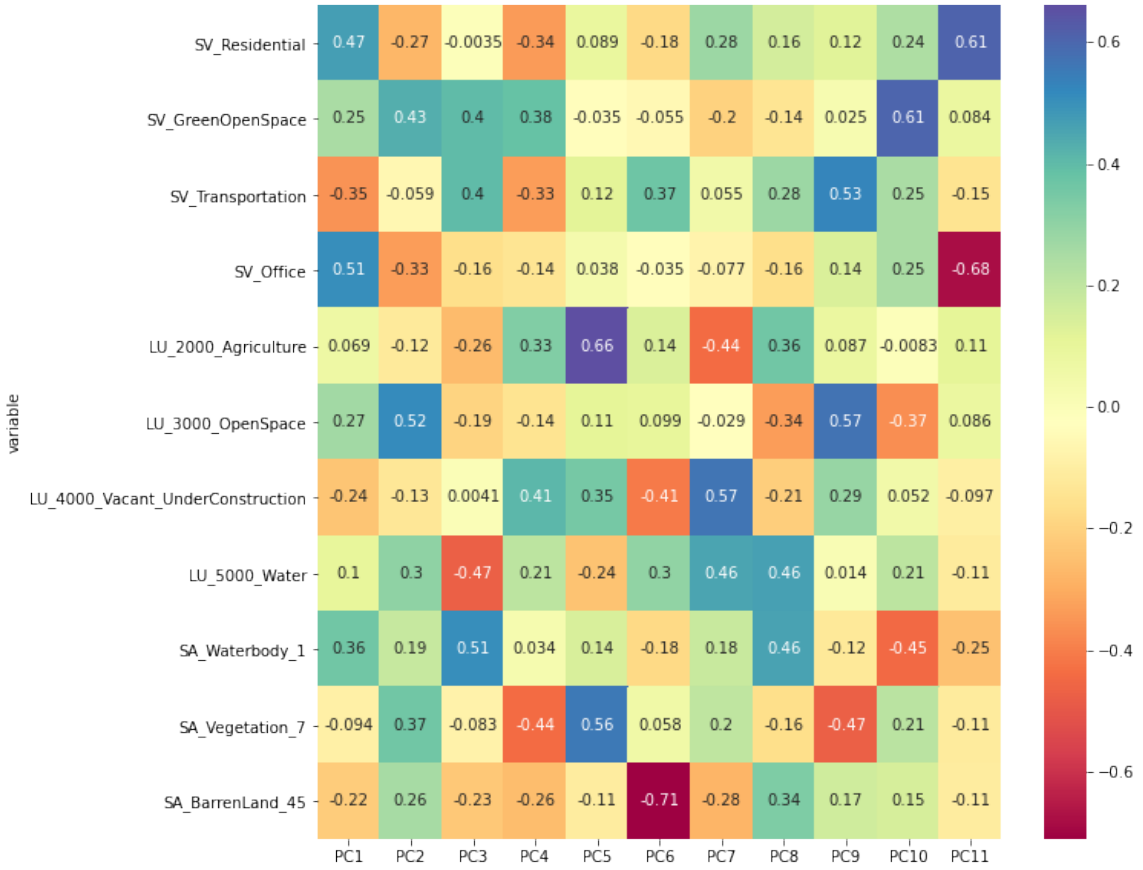


Figure 10. Original variables compared with PCA score

5.2. VALIDATION

Figure 11 summarizes how deployed Project Eclipse sensors in Chicago can be explained using two-three land feature scores pertinent to original land use features. A positive and negative value of variables only implicates the direction of vectors, so comparing the variance score is not favored. Green open space features from street view and open space features from land use are heading toward the same direction. Satellite image-based waterbody and land use water share similar directions. Residential and office land use from scene features share similar aspects. Transportation of scene features and vacant and under construction land use show similar patterns.

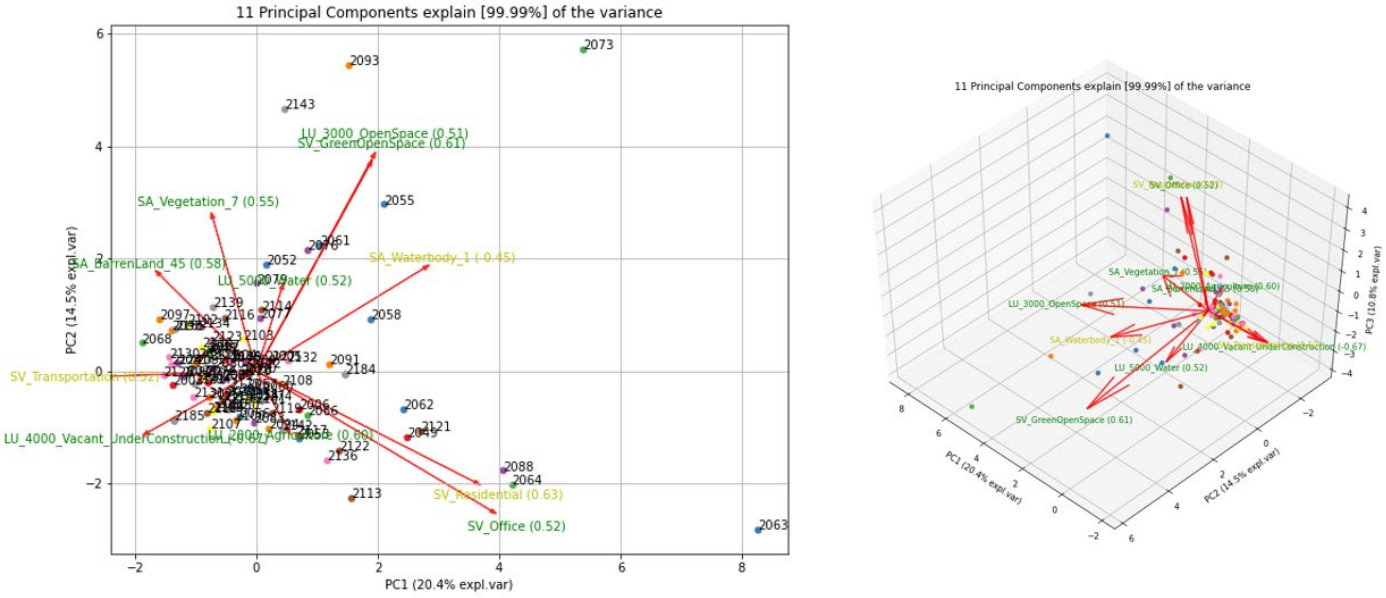


Figure 11. Biplot

Now the question is whether Project Eclipse sensor No. 2073 and 2093 do on land use with high green open space or near the lakeside. Looking at satellite images as described in Figure 12, these sensors are located within the high resolution of green open space or adjacent to *Lake Michigan*.

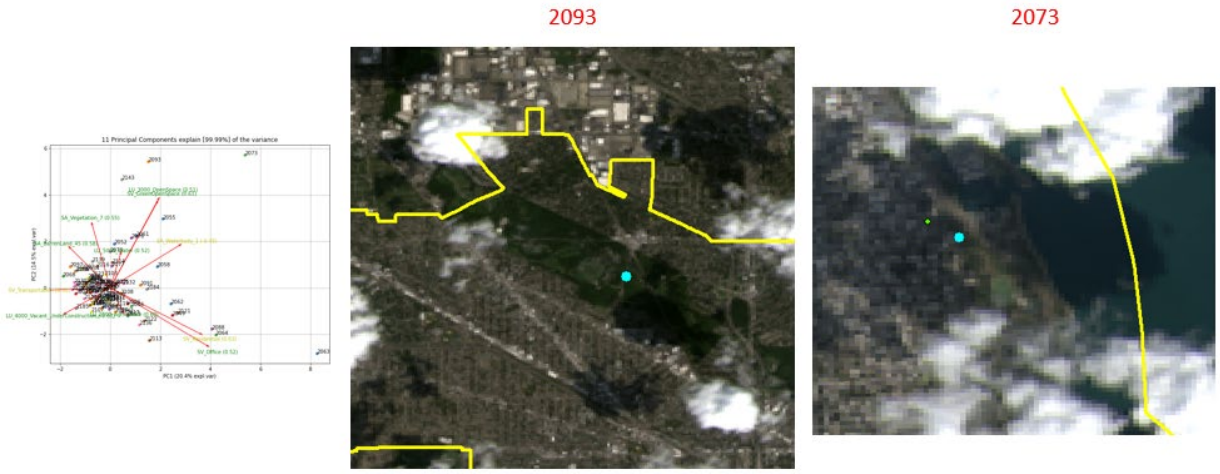


Figure 12. Project Eclipse sensors No. 2093 & 2073

5.3. PREDICTION RESULT

Table 2 summarizes RMSLE and compares the results. LSTM outperformed for having the minimum RMSLE of 0.013. Simple linear regression follows by 0.107, and ensemble learning catches up. RF performed better than boosting-based methods by scoring 0.459. Figure 13 compared AQI prediction to actual value (testing dataset). Overall, the prediction generally goes in line with the testing dataset.

Table 2. Calculated RMSLE

Model		**RMSLE**
Regression	OLS	0.107
Ensemble	RF	0.459
	GB	0.470
	XGB	0.469
	LGBM	0.469
Neural Networks	LSTM	0.013

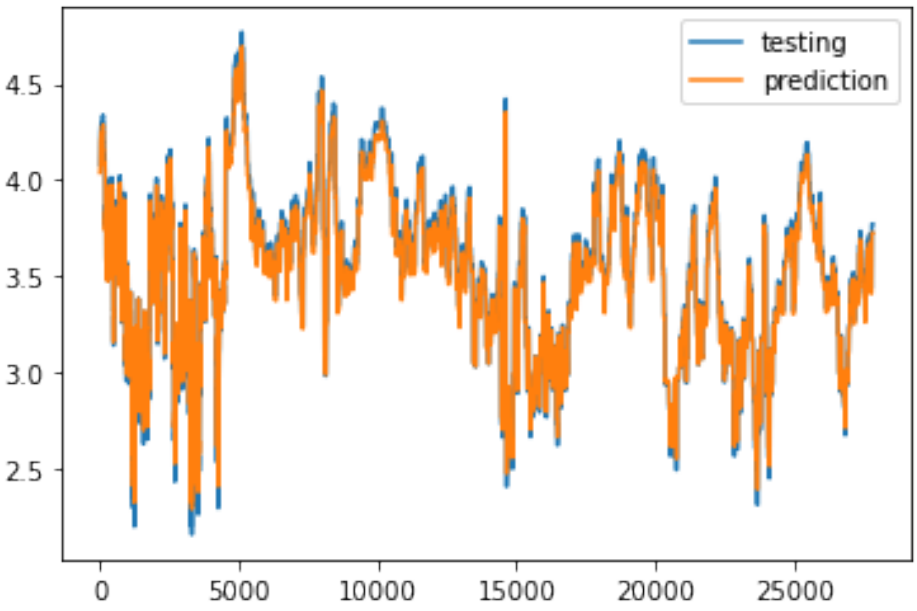


Figure 13. LSTM AQI Prediction

6. CONCLUSION

The study extracted scene features from street view images, combined with conventional land use features, including land use inventory and satellite image pertinent to Project Eclipse sensor deployment status. Three land use features went through dimensional scaling using PCA to calculate land use feature score. The study used land use feature score to model AQI prediction in Chicago. The study has several contributions. First, we suggested a framework to incorporate scene features extracted from street view images for air quality modeling. Second, PCA found that scene features are similar to conventional land use features concerning Project Eclipse sensor locations. Third, LSTM neural networks outperformed other machine learning models for air quality prediction. Bagging-based RF had a lower error score among ensemble machine learning models than boosting models. Perhaps, incorporating scene features for other climate features forecast modeling such as temperature, or particulate matter (PM 2.5, 10) may improve the performances.

Nevertheless, the study has several limitations. First, due to time constraints, other linear models, including the Gaussian process, have not taken place. Second, LSTM was slightly overfitting. Third, the number of installed Project Eclipse sensors is relatively small for PCA analysis. We leave these tasks behind for future studies.

REFERENCES

[1] J. Rovira, J. L. Domingo, and M. Schuhmacher, "Air quality, health impacts and burden of disease due to air pollution (PM10, PM2.5, NO2 and O3): Application of AirQ+ model to the Camp de Tarragona County (Catalonia, Spain)," *Sci. Total Environ.*, vol. 703, p. 135538, Feb. 2020, doi: 10.1016/j.scitotenv.2019.135538.

[2] J. Gallagher, R. Baldauf, C. Fuller, P. Kumar, L. Gill, and A. McNabola, "Passive methods for improving air quality in the built environment: A review of porous and solid barriers," *Atmos. Environ.*, vol. 120, pp. 61–70, Aug. 2015, doi: 10.1016/j.atmosenv.2015.08.075.

[3] "Project Eclipse," *Microsoft Research*. https://www.microsoft.com/en-us/research/project/project-eclipse/ (accessed Dec. 01, 2021).

[4] D. R. Richards and P. J. Edwards, "Quantifying street tree regulating ecosystem services using Google Street View," *Ecol. Indic.*, vol. 77, pp. 31–40, Jun. 2017, doi: 10.1016/j.ecolind.2017.01.028.

[5] X. Li, C. Zhang, W. Li, R. Ricard, Q. Meng, and W. Zhang, "Assessing street-level urban greenery using Google Street View and a modified green view index," *Urban For. Urban Green.*, vol. 14, no. 3, pp. 675–685, Jan. 2015, doi: 10.1016/j.ufug.2015.06.006.

[6] L. Yin and Z. Wang, "Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery," *Appl. Geogr.*, vol. 76, pp. 147–153, Nov. 2016, doi: 10.1016/j.apgeog.2016.09.024.

[7] T. Gebru *et al.*, "Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States," *Proc. Natl. Acad. Sci.*, vol. 114, no. 50, pp. 13108–13113, Dec. 2017, doi: 10.1073/pnas.1700035114.

[8] K. Hara, V. Le, and J. Froehlich, "Combining crowdsourcing and google street view to identify street-level accessibility problems," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, Apr. 2013, pp. 631–640. doi: 10.1145/2470654.2470744.

[9] A. G. Rundle, M. D. M. Bader, C. A. Richards, K. M. Neckerman, and J. O. Teitler, "Using Google Street View to Audit Neighborhood Environments," *Am. J. Prev. Med.*, vol. 40, no. 1, pp. 94–100, Jan. 2011, doi: 10.1016/j.amepre.2010.09.034.

[10] P. D. Isola, J. N. Bogert, K. M. Chapple, S. Israr, T. L. Gillespie, and J. A. Weinberg, "Google Street View assessment of environmental safety features at the scene of pedestrian automobile injury," *J. Trauma Acute Care Surg.*, vol. 87, no. 1, pp. 82–86, Jul. 2019, doi: 10.1097/TA.0000000000002338.

[11] L. Ilic, M. Sawada, and A. Zarzelli, "Deep mapping gentrification in a large Canadian city using deep learning and Google Street View," *PLOS ONE*, vol. 14, no. 3, p. e0212814, Mar. 2019, doi: 10.1371/journal.pone.0212814.

[12] Hendry and R.-C. Chen, "Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning," *Image Vis. Comput.*, vol. 87, pp. 47–56, Jul. 2019, doi: 10.1016/j.imavis.2019.04.007.

[13] R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec. 2018, pp. 2503–2510. doi: 10.1109/BigData.2018.8621865.

[14] W. Lan, J. Dang, Y. Wang, and S. Wang, "Pedestrian Detection Based on YOLO Network Model," in *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, Aug. 2018, pp. 1547–1551. doi: 10.1109/ICMA.2018.8484698.

[15] J. Kang, M. Körner, Y. Wang, H. Taubenböck, and X. X. Zhu, "Building instance classification using street view images," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 44–59, Nov. 2018, doi: 10.1016/j.isprsjprs.2018.02.006.
[16] H. Cui, G. Yuan, N. Liu, M. Xu, and H. Song, "Convolutional neural network for recognizing highway traffic congestion," *J. Intell. Transp. Syst.*, vol. 24, no. 3, pp. 279–289, May 2020, doi: 10.1080/15472450.2020.1742121.
[17] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Nov. 2015, pp. 730–734. doi: 10.1109/ACPR.2015.7486599.
[18] Image classification using the ArcGIS Spatial Analyst extension—ArcMap | Documentation [WWW Document], n.d. URL https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/image-classification/image-classification-using-spatial-analyst.htm (accessed 12.4.21).
[19] Amato, F., Guignard, F., Robert, S., Kanevski, M., 2020. A novel framework for spatio-temporal prediction of environmental data using deep learning. Sci Rep 10, 22243. https://doi.org/10.1038/s41598-020-79148-7
[20] Iskandaryan, D., Ramos, F., Trilles, S., 2020. Air Quality Prediction in Smart Cities Using Machine Learning Technologies based on Sensor Data: A Review. Applied Sciences 10, 2401. https://doi.org/10.3390/app10072401