From Space to Health: Deep Learning Based Prediction of Health Outcomes from Satellite Imagery in Fine Grained Geographical Areas

Author: Yuming Shi | **UCL Geography – MSc Social and Geographic Data Science**
UCL Supervisor: Dr Stephen Law | **Co-supervisors**: Dr Sanja Šćepanović and Dr Daniele Quercia (Nokia), Dr Barbara Metzler (Turing Institute)

Introduction

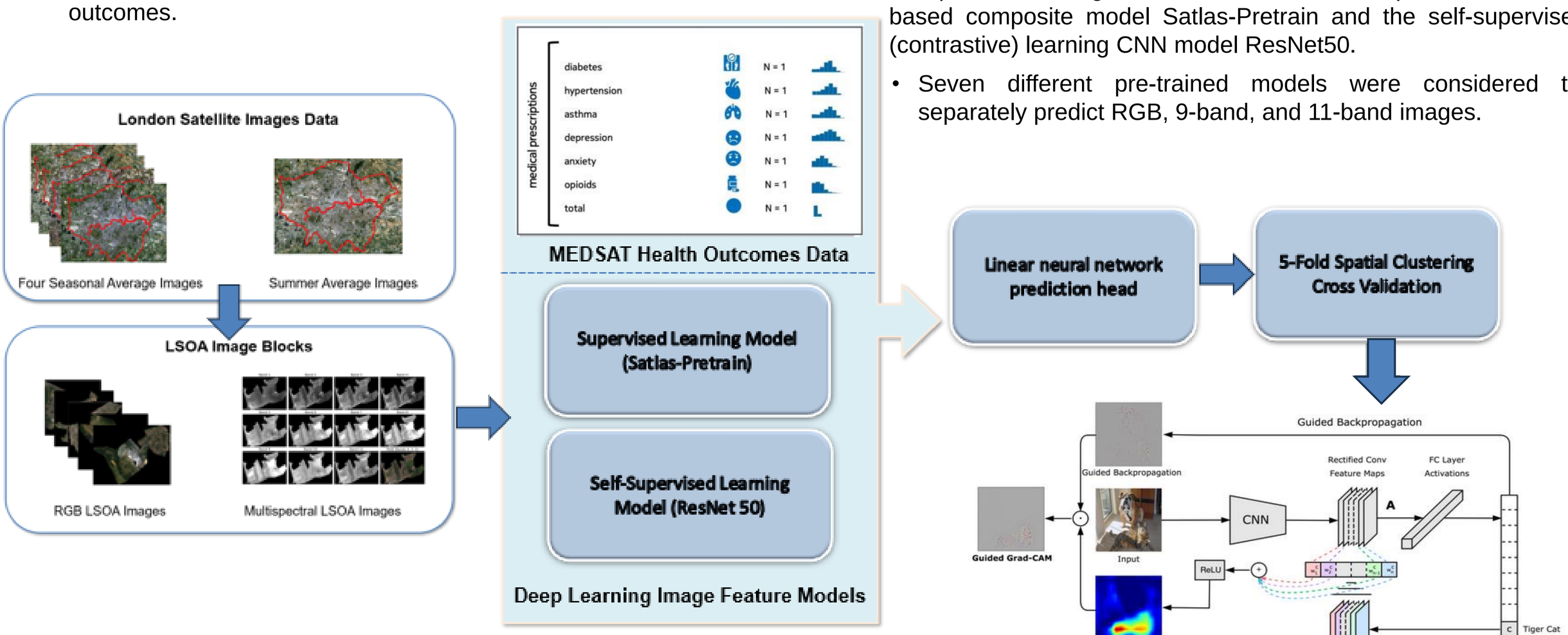
The critical role of the environment in shaping public health outcomes has become increasingly apparent (WHO, 2023). Continuous analysis of environmental factors is crucial for maintaining and improving public health. Satellite imagery contains a wealth of environmental information; however, due to the unique characteristics of satellite data, it is not fully leveraged in environmental health research. This study evaluates the feasibility of using LSOA-level satellite imagery to predict and interpret health outcomes based on the latest MEDSAT dataset for satellite images and health outcomes, with London as the case study.

Main Contribution

- A comparative analysis of deep learning models, data sources, and health outcomes.
- Customized spatial clustering cross-validation method for assessing spatial generalization of prediction results.
- Achieving an accuracy surpassing the baseline image features.
- Using XAI techniques to reveal key image features related to health outcomes.

Data and Models for Comparison

- **4,659** satellite images of London LSOAs and the corresponding per capita prescription number for hypertension, asthma, diabetes, depression, and anxiety.
- Summer/Annual(four season) average images; RGB/Multispectral images were included for comparison.
- Compared the image feature models of the supervised learning-based composite model Satlas-Pretrain and the self-supervised (contrastive) learning CNN model ResNet50.
- Seven different pre-trained models were considered to separately predict RGB, 9-band, and 11-band images.



General Workflow

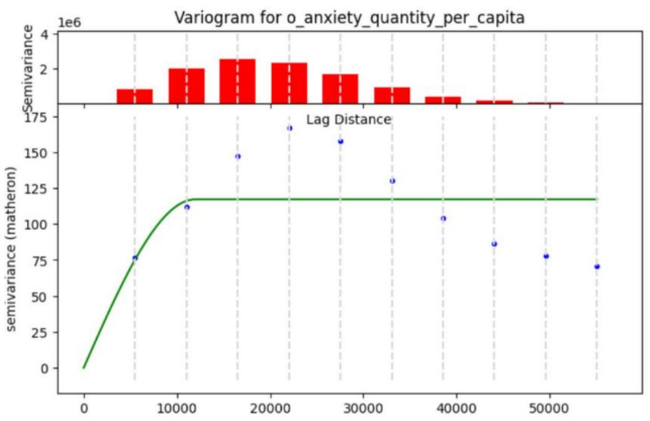
Grad-CAM

Evaluation (Spatial CV)

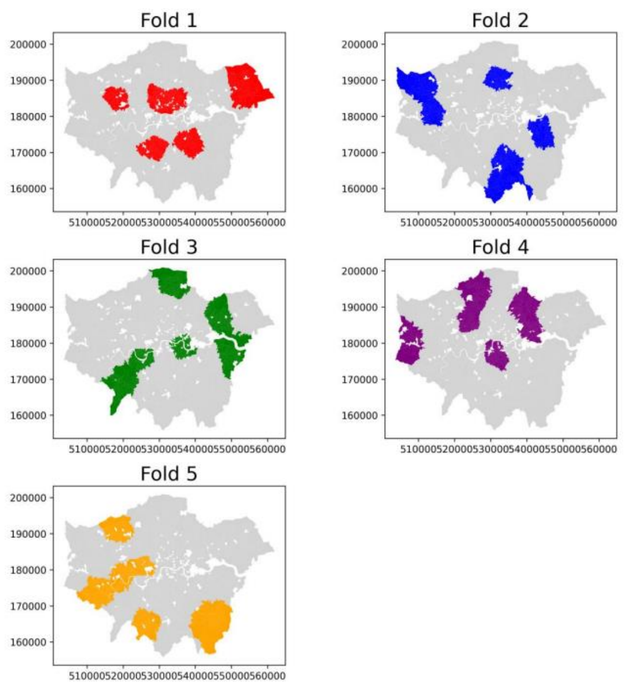
The final best prediction results were evaluated based on the 5-fold clustering cross-validation method, following the validation frameworks of Roberts et al. (2017) and Mahoney et al. (2023). First, the semi-variogram of the health outcomes data was calculated, and the spatial autocorrelation range was estimated based on the convergence distance of the semi-variance. Then, using the spatial autocorrelation range, the training and test-validation sets were divided based on K-means spatial clustering to ensure sufficient distance between the training and testing sets, thereby reducing the risk of overestimating the results due to spatial autocorrelation.

$$\gamma(h) = \frac{1}{2}\sum_{i=1}^{N(h)}[(H(x_i) - H(x_i + h))]^2$$

The formula of the semi-variance



Example semi-variogram



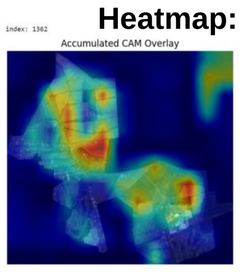
Spatial segmentation for each fold

Result

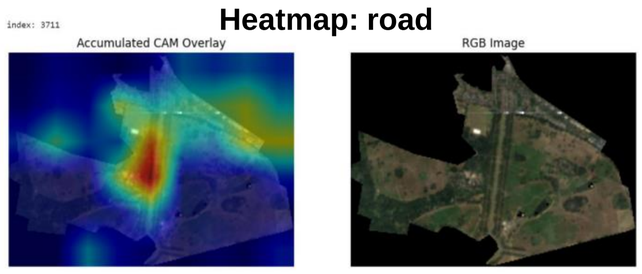
Best prediction results were obtained using summer average RGB images, with visual features extracted from SeCo, a foundational contrastive learning model. Grad-CAM visualization of the top 5 features showed that exposed soil may contribute to higher prescription rates in LSOAs, while road-like features seem linked to lower rates.

R^2	*diabetes*	*asthma*	*hypertension*	*depression*	*anxiety*	*Average*
Annual Average Image						
Satlas RGB	-0.015	0.205	0.153	0.177	0.168	0.176
MoCo RGB	0.022	0.203	0.241	0.197	0.186	0.207
SeCo RGB	0.074	0.202	0.230	0.183	0.190	0.201
Satlas 9 Bands	-0.042	0.224	0.151	0.201	0.180	0.189
Satlas 11 Bands	0.005	0.240	0.201	0.216	0.205	0.216
MoCo 11 Bands	-0.051	0.184	0.219	0.199	0.185	0.197
SimCLR 11 Bands	-0.132	0.108	-0.001	0.066	0.056	0.057
Summer Average Image						
Satlas RGB	-0.004	0.199	0.168	0.195	0.176	0.185
MoCo RGB	0.100	0.201	0.279	0.187	0.210	0.220
SeCo RGB	0.076	0.249	0.283	0.210	0.208	0.232
Satlas 9 Bands	-0.008	0.233	0.134	0.178	0.160	0.176
Satlas 11 Bands	0.013	0.226	0.178	0.209	0.188	0.200
MoCo 11 Bands	-0.020	0.209	0.222	0.172	0.179	0.196
SimCLR 11 Bands	-0.108	0.074	-0.018	0.058	0.036	0.038

Heatmap: exposed soil



Heatmap: road



Reference:

World Health Organization (WHO), 2023. Climate Change. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health [Accessed 19 July 2024].

Roberts, D. R. et al. (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography (Copenhagen). [Online] 40 (8), 913–929.

Mahoney, M. J. et al. (2023) Assessing the performance of spatial cross-validation approaches for models of spatially structured data. arXiv.org. [Online]