Beyond here and now: Evaluating pollution estimation across space and time from street view images with deep learning

Ricky Nathvani, D. Vishwanath, Sierra N. Clark, Abosede S. Alli, Emily Muller, Henri Coste, James E. Bennett, James Nimo, Josephine Bedford Moses, Solomon Baah, Allison Hughes, Esra Suel, Antje Barbara Metzler, Theo Rashid, Michael Brauer, Jill Baumgartner, George Owusu, Samuel Agyei-Mensah, Raphael E. Arku, Majid Ezzati

PII: S0048-9697(23)04793-9

DOI: https://doi.org/10.1016/j.scitotenv.2023.166168

Reference: STOTEN 166168

To appear in: *Science of the Total Environment*

Received date: 4 May 2023

Revised date: 7 August 2023

Accepted date: 7 August 2023

Please cite this article as: R. Nathvani, D. Vishwanath, S.N. Clark, et al., Beyond here and now: Evaluating pollution estimation across space and time from street view images with deep learning, *Science of the Total Environment* (2023), https://doi.org/10.1016/j.scitotenv.2023.166168

Beyond here and now: evaluating pollution estimation across space and time from street view images with deep learning

Ricky Nathvani[†1,2], Vishwanath D[†1,2], Sierra N. Clark[1,2], Abosede S. Alli[3], Emily Muller[1,2], Henri Coste[1,2], James E Bennett[1,2], James Nimo[4], Josephine Bedford Moses[4], Solomon Baah[4], Allison Hughes[4], Esra Suel[1,2,5], Antje Barbara Metzler[1,2], Theo Rashid[1,2], Michael Brauer[6], Jill Baumgartner[7,8], George Owusu[9], Samuel Agyei-Mensah[10], Raphael E. Arku[‡3], Majid Ezzati[‡1,2,11]*

** Corresponding author: r.nathvani@imperial.ac.uk*

† Joint first authors

‡ Joint senior authors

1 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

2 MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK

3 Department of Environmental Health Sciences, School of Public Health and Health Sciences, University of Massachusetts, Amherst, USA

4 Department of Physics, University of Ghana, Accra, Ghana

5 Centre for Advanced Spatial Analysis, University College London, London, UK

6 School of Population and Public Health, University of British Columbia, Vancouver, Canada

7 Institute for Health and Social Policy, McGill University, Montreal, Canada

8 Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada

9 Institute of Statistical, Social & Economic Research, University of Ghana, Accra, Ghana

10 Department of Geography and Resource Development, University of Ghana, Accra, Ghana

11 Regional Institute for Population Studies, University of Ghana, Accra, Ghana

Abstract

Advances in computer vision, driven by deep learning, allows for the inference of environmental pollution and its potential sources from images. The spatial and temporal generalisability of image-based pollution models is crucial in their real-world application, but is currently understudied, particularly in low-income countries where infrastructure for measuring the complex patterns of pollution is limited and modelling may therefore provide the most utility. We employed convolutional neural networks (CNNs) for two complementary classification models, in both an end-to-end approach and as an interpretable feature extractor (object detection), to estimate spatially and temporally resolved fine particulate matter ($PM_{2.5}$) and noise

levels in Accra, Ghana. Data used for training the models were from a unique dataset of over 1.6 million images collected over 15 months at 145 representative locations across the city, paired with air and noise measurements. Both end-to-end CNN and object-based approaches surpassed null model benchmarks for predicting $PM_{2.5}$ and noise at single locations, but performance deteriorated when applied to other locations. Model accuracy diminished when tested on images from locations unseen during training, but improved by sampling a greater number of locations during model training, even if the total quantity of data was reduced. The end-to-end models used characteristics of images associated with atmospheric visibility for predicting $PM_{2.5}$, and specific objects such as vehicles and people for noise. The results demonstrate the potential and challenges of image-based, spatiotemporal air pollution and noise estimation, and that robust, environmental modelling with images requires integration with traditional sensor networks.

Keywords: Deep learning; computer vision; air pollution; noise pollution; street-view images; environmental modelling

1) Introduction

The urban population in low- and middle-income countries (LMICs) increased from 357 million in 1950 to 3.39 billion in 2020, with the majority of the LMIC population now living in cities (United Nations, Department of Economic and Social Affairs, & Population Division, 2019). While cities offer their inhabitants better access to infrastructure, services and economic opportunity (Ezzati et al., 2018), factors such as road transport and residential and commercial

energy generation can also increase hazardous environmental exposures, including air and noise pollution (Kammen and Sunter, 2016; Kelly and Zhu, 2016). Although some sources of urban pollution in LMICs, such as vehicular traffic, are similar to those of many high income countries, there are also differences in the sources, and in their spatial and temporal patterns (Alli et al., 2021; Amegah and Agyei-Mensah, 2017; Clark et al., 2021; Deng et al., 2020; Ebare et al., 2011; Weagle et al., 2018; Zhou et al., 2013) such as seasonal Saharan Desert dust storms (Zhou et al., 2013), burning biomass fuels for cooking and heating, and the use of diesel generators where there are intermittent electricity outages (Dionisio Kathie L. et al., 2010).

Data on the patterns of air and noise pollution and their sources across space and time are needed to identify and evaluate mitigation measures and policies. However, collecting such data is challenging in resource-constrained settings (Bauer et al., 2019; Clark et al., 2020; Khan et al., 2018). Recent methodological advances in image processing and analysis, particularly in the form of deep convolutional neural networks, have demonstrated that street-level images can help with predicting air and noise pollution levels (Ganji et al., 2020; Hong et al., 2020; Qi and Hankey, 2021; Weichenthal et al., 2019), contingent on initial data measurements needed to develop the image-based pollution estimation models. So far, image-based pollution models have largely been developed for East Asia (Chakma et al., 2017; Feng et al., 2021; Gu et al., 2019; Liu et al., 2016, 2015; Wang et al., 2022; Won et al., 2022; Zhang et al., 2018) and North America (Ganji et al., 2020; Hong et al., 2020; Qi and Hankey, 2021), typically based on a few weeks' observation at selected locations, asynchronous or spatially distant from pollution measurements. Few studies have sought to predict spatially and temporally resolved pollution from images, and none in Africa, the world's fastest urbanising region (United Nations, Department of Economic and Social Affairs, & Population Division, 2019).

We developed and evaluated machine learning models to predict temporally and spatially varying noise and fine particulate matter ($PM_{2.5}$; particles <2.5 μm in diameter, with known human health impacts (Pope and Dockery, 2006)) levels from street-level images in Accra, Ghana. We used deep convolutional neural networks (CNNs), which learn robust and hierarchical feature representations that give them superior performance for many image-processing tasks (Schmidhuber 2015; Gu et al. 2018), in two complementary strategies. The first used a CNN, without a priori assumptions on the image features relevant for prediction, and another used gradient boosted machines, applied to previously extracted, interpretable image features in the form of object counts, obtained from applying an object-detection CNN to each image. These models were applied to a bespoke dataset of over 1.6 million time-lapsed images co-located with $PM_{2.5}$ and noise measurements at 145 representative locations over 15 months (Clark et al., 2020). Models were trained and evaluated on subsets of data specifically to interrogate their temporal and spatial generalisability and in order to compare and contrast strategies for data collection with fixed resources when developing such models. We further assessed model performance for both the day and night time, different seasons, and types of urban land use.

2) Data and Methodological Context and Contributions

Some studies have predicted pollution from visual elements of the environment. Two studies, also from Accra, recorded $PM_{2.5}$ and PM_{10} in selected neighbourhoods, in a multi-week measurement campaign (Dionisio Kathie L. et al., 2010; Rooney et al., 2012), together with researcher observations and census data on environmental factors, such as biomass fuels and

unpaved roads, to predict pollution levels. Some studies have also predicted pollution using remote sensing data, which differs from our study, not only in the view of the city, but also in spatial and temporal scales and the observable features in images (Sorek-Hamer et al., 2022; Wei et al., 2020; Weigand et al., 2019).

Other studies used terrestrial images for predicting air pollution (Chakma et al., 2017; Feng et al., 2021; Ganji et al., 2020; Gu et al., 2019; Hong et al., 2020; Liu et al., 2016, 2015; Qi and Hankey, 2021; Wang et al., 2022; Won et al., 2022; Zhang et al., 2018), and one for noise (Hong et al., 2020) based on images and pollution measurement data though none had spatiotemporally linked image and pollution data during the night time, as we do. Previously adopted approaches span a variety of experimental configurations, making a unifying, quantitative comparison among studies infeasible. The specific metric of pollution (e.g., black carbon vs PM), timescales on which pollution is predicted (single measurement in time vs variation across day), spatial resolution (city-wide vs local), images used (static vs time-varying), data inputs (solely images vs inclusion of meteorological variables), temporal range (<~1 week vs multiple months of observation), synchrony between data sources (pollution and images <~5 min apart vs >~1 year apart), modelling approach (regression of continuous pollution data vs classification into different classes) and model inputs (specific features vs entire images), vary from study to study. Furthermore, within studies that used images as model inputs, a variety of features and feature extraction methods (object detection vs segmentation) were used, including in relation to stationarity of features in time (e.g., buildings and trees vs vehicles and pedestrians). The majority of studies used a single configuration from such choices depending on the available data, generating prediction tasks that are easier or more difficult relative to others. We outline the different experimental setups for previous studies in Appendix Table A. In the specific case of

cities in Africa, one study used street-view images to predict $PM_{2.5}$ and NO_2 across several cities, including Accra (Suel et al., 2022). Data used for model training were derived from modelled estimates of annual average pollution level with a model only evaluated, not trained, on data from Accra.

Our work advances the state of knowledge in a number of ways. Our dataset is much larger and was collected over a longer duration than most previous image-based studies, comprising 145 locations and a total (prior to merging with our pollution data) of 2.1 million images over 15 months. We co-captured both air pollution and noise data with images in both day and night time. We predicted air pollution concentrations and noise levels at finer classification intervals, i.e. with classes that each encompass a smaller and more precise range as described in Section 3.3, than comparable previous classification-based studies. We systematically evaluated both the spatial and temporal generalisability of models which is relevant for designing an optimal digital surveillance strategy and guiding data collection. Our study is unique in the use of both end-to-end CNN (outcome-driven) and object based (feature-driven) models, which both inform model selection and enhance model interpretability. Finally, to our knowledge, this is the first use of images for predicting both air and noise pollution in the context of an African city.

3) Materials and Methods

3.1) Data

We collected co-located time-lapsed images at 5-minute intervals and $PM_{2.5}$ and noise measurements averaged and recorded at 1-minute intervals in a field campaign from April 2019 to June 2020, details of which are described in Appendix A and the study protocol paper (Clark

et al., 2020). We had ten fixed sites where data were collected over 15 months, and 135 rotating sites where data was collected for one week. The fixed sites provided information for assessing temporal generalisability of models, and both fixed and rotating sites for assessing spatial generalisability. Sites were grouped into four land-use classes: commercial, business, industrial (CBI); informal, mostly high-density, settlements and slums; formal, mostly low- and medium-density, residential areas; and "other" areas that are often peri-urban or rural, and can have dense vegetation (i.e., forest, grassland) or barren land (i.e., sand, soil, dirt). The classes for each fixed site are detailed in Appendix Table B.

3.2) Research questions

We developed two types of models that used images to predict noise and air pollution. We analysed how well our models' prediction generalise across time and space, through the following research questions:

1a) Temporal generalisability: How well do models trained on images taken from a single location predict noise and $PM_{2.5}$ at different, random times at the same location?

1b) Spatial generalisability: How well do models trained in 1a), which are based on a single location, generalise to another unseen location?

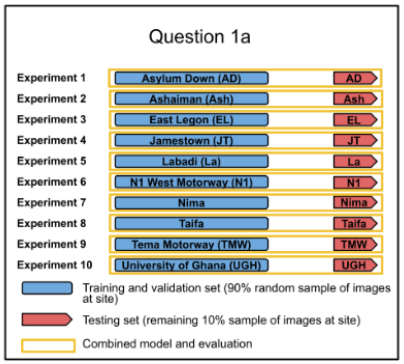
2a) Spatial generalisability: How well do models trained using an abundance (~1,000,000 total across sites) of images from a set of nine (long-term) fixed sites, predict noise and $PM_{2.5}$ at the remaining (10^{th}) unseen location?

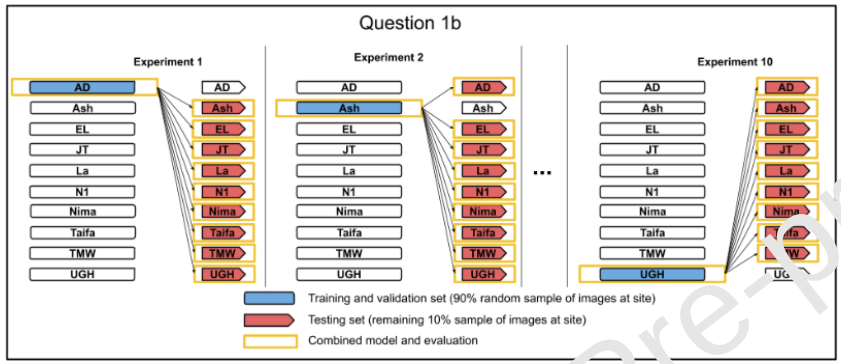
2b) Spatial generalisability: How well can models trained using fewer images (~100,000 total across sites) from ~90% of our 135 rotating sites, predict noise and $PM_{2.5}$ at the remaining ~10% of sites?

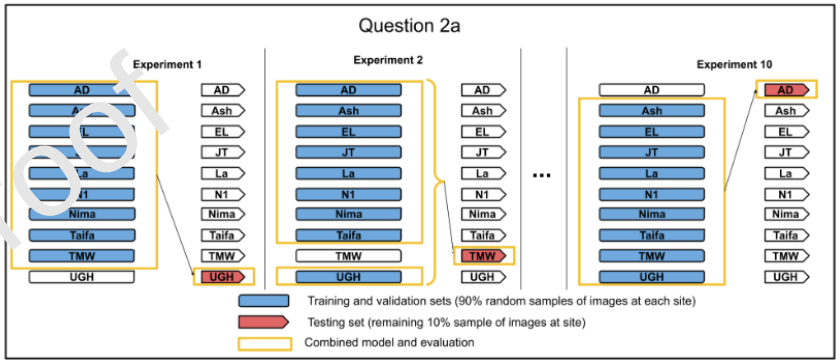
The fixed sites, due to their extended data collection period, comprised seven times as much data as rotating sites in total. Since in-situ pollution measurements are resource intensive, especially in quantities needed to train a CNN (Sun et al., 2017), there is a need to optimally allocate the use of cameras and pollution measurement hardware, as well as personnel time. Therefore we also investigated whether models trained using more data from a smaller number of (fixed) sites, or fewer data from a greater number of (rotating) sites led to more spatially generalisable CNN models:

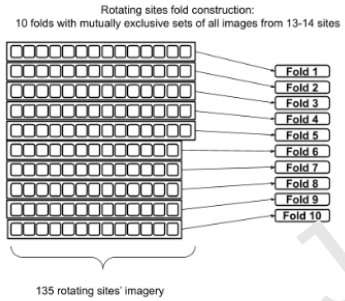
2c) Comparison of model types from 2a) and 2b): Do models perform better on multiple, unseen locations (remaining ~10% of rotating sites) when given an abundance of images from a few locations, or fewer images across a variety of locations?

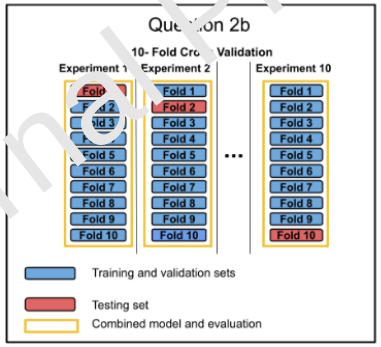
For each question, we divided our data into subsets for training and testing, as illustrated in Figure 1. The number of images belonging to each of the datasets is given in Appendix Table B.











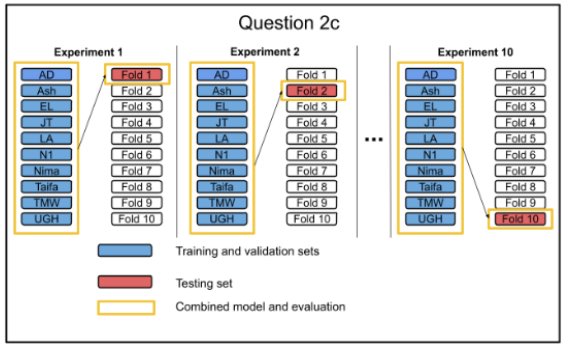


Figure 1. Data use for training and testing of models.

Each panel shows how data from fixed and rotating sites were allocated to training and testing sets, for each question posed in Section 3.2. For the training sets, indicated in blue, each block was further divided into a 75-25 split with the latter being used as a validation set during training configuration and hyperparameter determination. Final models were trained on the entire training set (including the validation set) and evaluated on the testing set, indicated in red.

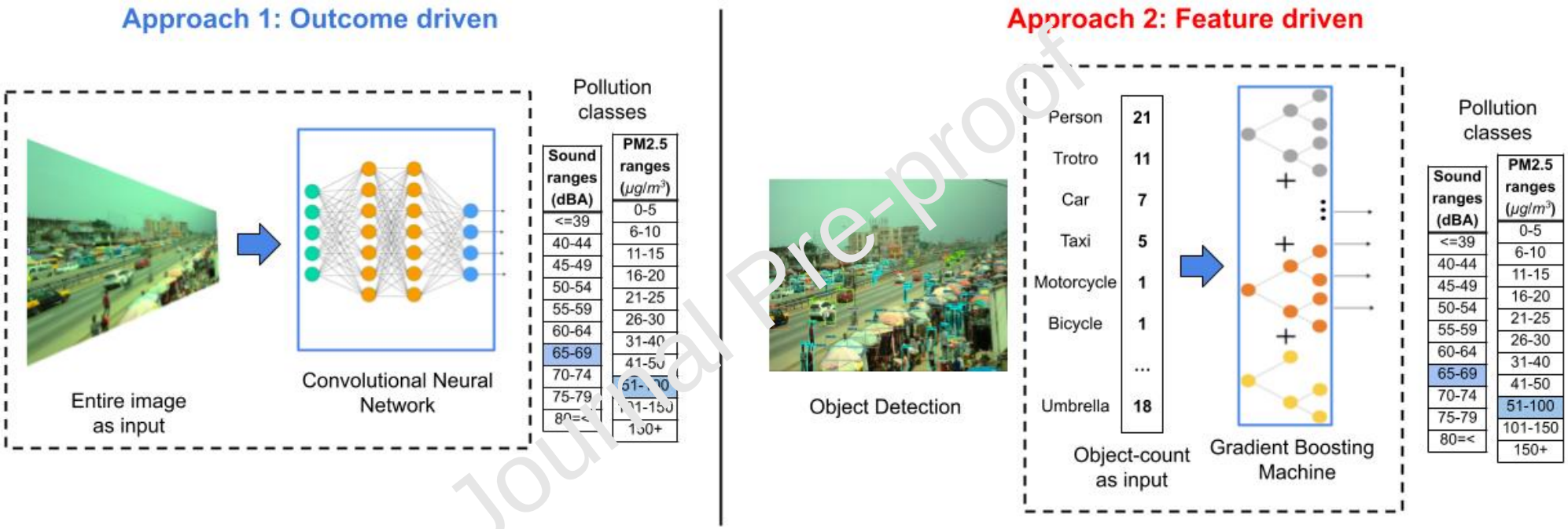


Figure 2. End-to-end (CNN) and object-based (GBM) modelling approaches.

3.3) Modelling

For all research questions, we trained models to predict levels of noise (dBA) or $PM_{2.5}$ ($\mu g/m^3$) at a given time (1-minute averaged interval) and location from a single image taken within 30 seconds of pollution measurement. We framed the prediction task as a classification problem, i.e. the models predict specific ranges (classes) in which noise and $PM_{2.5}$ fall rather than as a continuous value, for two reasons. First, policy targets and guidelines, such as those of the World Health Organization (Basner and McGuire, 2018; World Health Organization, 2021), tend to be formulated based on discrete levels. Second, a preliminary analysis indicated that models trained explicitly for classification outperformed regression models trained for continuous value prediction, as detailed in Appendix C. The classes for noise were: <=39, 40 to <45, 45 to <50, 50 to <55, 55 to <60, 60 to <65, 65 to <70, 70 to <75, 75 to <80, >=80 dBA. The classes for $PM_{2.5}$ were: 0 to <5, 5 to <10, 10 to <15, 15 to <20, 20 to <25, 25 to <30, 30 to <40, 40 to <50, 50 to <100, 100 to <150, >=150 $\mu g/m^3$.

For both forms of pollution, we produced two classification models (Figure 2). The first, referred to as end-to-end classification, used an entire unprocessed image, with red, green and blue pixel channels, as input to a CNN to predict pollution class. No assumptions were made on relevant image features, which were learned from the data. The second group of models used counts of objects detected from images as input for classification via gradient boosted machines (Friedman, 2001) (GBM). Other approaches to feature extraction from images, such as semantic segmentation, could also have been employed to provide model inputs for pollution estimation, as used in a North American study (Qi and Hankey, 2021). We used objects in our second approach since the data needed to train a model, namely objects, were less resource intensive to

generate within our bespoke dataset with bounding boxes (Nathvani et al., 2022) as compared with pixel-level annotation, which may also be explored in future work. The object counts were obtained from training an object detection CNN, described in detail in previous work (Nathvani et al., 2022), for object categories relevant to the local environmental context: persons, market vendor (a person carrying a container over their heads which is a common scene in African markets), car, taxi, pick-up truck, bus, lorry, van, tro-tro (mini buses used for public transportation), motorcycle, bicycle, market stall, loudspeaker, umbrella (commonly used to protect market and roadside vendors from the sun and rain), cookstove, cooking pot/bowl (which frequently contain wares for sale in the marketplace), food, trash, (piece of) debris, and animal. All object categories are those which may vary over time at a given place, since although other static features, such as buildings or trees, may also affect noise and air pollution, their unchanging presence over daily timescales is less informative for predicting temporal variation in pollution at a single location (such as those models developed in 1a and 1b). The accuracy with which these objects could be detected in our images is given in Appendix Table C and described in previous work (Nathvani et al., 2022). In this analysis, we did not use counts of cookstove, loudspeakers, market vendors or buses, due to their sparse presence in our data (<10 counts of each in 2.1 million images). The end-to-end and feature-driven approaches are complementary with respect to flexibility and feature agnosticism versus prior assumption and interpretability (Zhang and Zhu, 2018).

3.4) Data preparation.

We prepared our data in the following manner for the purpose of training and evaluating both CNN and object-GBM models. First, due to the Covid-19 pandemic and associated lockdown in

Accra from March 30th to April 20th 2020, we excluded images and pollution data from March 23rd to May 11th 2020, when we were unable to attend to the regular maintenance of monitoring hardware, and therefore data collection was incomplete and uneven across sites.

A small number of cameras experienced internal failure of their clocks, resetting to a factory default of January 2017 at the start of their deployment, which led to images recorded with incorrect timestamps. We corrected the timestamps for these images by re-assigning the initial timestamp based on the start of the monitoring period, which was recorded on a log-sheet when visiting every site. Since each image thereafter was captured at regular five minute intervals, subsequent images were assigned time stamps at five minute intervals. Finally, a small fraction (<1%) of images and pollution data were corrupted and hence unreadable by code. These data were excluded.

Images and pollution data were combined by assigning each image the pollution observation nearest in time, with a requirement that the pollution value was recorded within +/- 30 seconds of image capture. Where two cameras were placed at a site, both images are assigned pollution data based on this procedure. Some images did not have corresponding pollution values due to a lack of measurements when monitors failed or were unstable. For noise, 83-98% of fixed site images and 99% of rotating site images were assigned pollution data. For $PM_{2.5}$, 68-89% of fixed site images and 79% of rotating site images were assigned pollution data. Full details are provided in Appendix Figure C. As mentioned in section 3.3 we applied a previously developed object detection CNN to all 2.1 million images in our unmerged data set to obtain information on the counts of different object categories within each image, which are used as numerical inputs for

our GBM models. Examples of the detected objects within our images may be seen in Appendix Figure D.

1.6 million images were assigned corresponding $PM_{2.5}$ values and 1.9 million images with noise values. Each dataset was divided into training, validation and test sets, as shown in Figure 1. The test set was 10% of all data. Training and validation sets (a 25% holdout of the remainder of data) were used to train the model and set hyperparameters. After determining hyperparameters, the final models were trained on the combined training and validation sets and evaluated on the test set. Both CNN and GBM models used the exact same sets of images, with the former using the entire image and the latter the counts of objects in each image.

3.5) Model training

3.5.1) End-to-end CNN

We used a Residual Network with 101 layers (ResNeXt 101) (Xie et al., 2017) as the convolutional neural network (CNN) architecture for classifying noise and $PM_{2.5}$ levels from an entire image. The algorithm was implemented and trained in PyTorch (Paszke et al., 2019) and was pretrained on ImageNet data to enable the CNN to recognise low level features, e.g., edges, which improved model performance, as seen in other computer vision tasks (Huh et al., 2016).

During training, CNNs were given images resized to 224 × 224 pixels with an layer depths of 3 to accommodate the Red, Green and Blue channels, with Z-score normalisation applied across all images to assist with the gradient descent process of learning (LeCun et al., 2012). A modified

cross-entropy loss function with a log-barrier constraint (Belharbi et al., 2020) was used to account for the ordinal nature of pollution classes, as described in Appendix B. Training was performed on two NVIDIA Quadro RTX 6000 GPUs (48GB memory), with models taking approximately 1 hour per epoch and lasted for 30 epochs. The batch size was 32 images, using stochastic gradient descent with an initial learning rate of 0.001, a momentum of 0.9 and a step size of 40. Final models were those which performed best on the validation set during the training process, ranging from the 16-25th epoch.

At training time, data augmentation was used to improve model generalisability and mitigate overfitting to the data (Shorten and Khoshgoftaar, 2019), by uniformly, randomly cropping the image borders, with the central 90% area of the image always preserved, random rotations of the image between 10° anti-clockwise to 10° clockwise, and evenly random flipping of the image in the horizontal plane. These transformations correspond to the variance seen between camera images and the placement at different sites, which had different fields of view and camera orientations.

3.5.2) Object-based Gradient Boosted Machines

We used Gradient Boosting Machines or GBMs as the algorithm for classifying noise and $PM_{2.5}$ from specific, interpretable features, which were counts of objects detected within each image from a separate CNN (Nathvani et al., 2022). GBMs are ensemble tree-based models which use "boosting", i.e. adaptively changing the weights of data points in the training distribution during the learning process to improve performance on less easily predicted data (Friedman, 2001), which were implemented XGBoost (Chen and Guestrin, 2016) in Python and Scikit-Learn

(Pedregosa et al., 2011). GBMs have high efficacy across many problem domains with structured data inputs, due to their ability to learn non-linear relationships between features with robustness to outliers in a flexible and scalable manner (Chen and Guestrin, 2016). They offer advantages compared to linear models, which are more biased in complex data domains, computationally expensive models such as Support Vector Machines, and Artificial Neural Networks which are typically more cumbersome to optimise. Furthermore, in a preliminary analysis, GBMs had better performance, as measured by classification accuracy, than comparable tree-based methods such as decision trees and random forests.

The input to the GBM models were vectors representing the counts of different objects in each image, e.g., (cars: 2, people: 3, umbrellas: 0…). The model hyperparameters were determined with Bayesian optimisation, with the validation set being used for fixed sites' data and with 3-fold cross validation when training on 9 folds of rotating site data (Figure 1) and a cross-entropy loss function. The search range for the parameters is given in Appendix Table D. Training was performed during the Bayesian hyperparameter tuning process and was stopped when 5 iterations of tuning yielded no further improvements in overall class prediction accuracy on the independent validation set.

3.6) Model evaluation

We compared the performance of both end-to-end and feature-driven approaches to infer plausible contributors to how they predict pollution. We calculated classification accuracy for exact class prediction as well as for when the model classified into the ground truth or adjacent class (shown as "same and ±1 class accuracy"). We evaluated all our models against a null

model which measures whether the models do better than simply taking the average from a distribution of training data. We also calculated the models' accuracy for specific subsets of data under different environmental conditions, including day and night time, and the dry and dusty Harmattan season (November–February). During the Harmattan season dust from the Saharan Desert is carried by trade winds (Adetunji et al., 1979), and there is haze and "redness"(Adetunji et al., 1979; Anuforom, 2007; Ette and Olorode, 1988; McTainsh, 1980; Ochei and Adenola, 2018; Pinker et al., 1994), caused by absorption and scattering of light (Groblicki et al., 1981; Waggoner and Weiss, 1980) and the dust itself, which has a red brown colour (Breuning-Madsen and Awadzi, 2005; Lafon et al., 2004). These changes in visibility can inform air pollution estimation (Hyslop, 2009; Ozkaynak et al., 1985).

For GBM models in 1a, we quantified the importance of each object for prediction via its permutation importance, which calculates the reduction in the model's accuracy on the test set, before and after randomly shuffling the values of an input feature (in our case, counts in a given object category) across images. Object counts in our data were correlated amongst different object categories across images (Nathvani et al., 2022), e.g., people and cars. Therefore a given object's importance score might be lower when multiple objects are used for prediction than when single objects are used, because other correlated objects capture some of the same information.

4) Results and Discussion

For noise, classification accuracy at different times in the same location (i.e. Question 1a) for both the end-to-end (CNN) and object-based (GBM) models ranged from 40-70% across sites,

considerably outperforming their null models (Figure 3). Accuracy increased to 80-90% for neighbouring (±1) class classification. The performance of the two models was similar, with CNNs slightly outperforming GBM models. Predictions at sites with high road-traffic, such as Asylum Down, Tema Motorway and N1 West Motorway, had higher accuracy than those at other sites. Noise predictions using CNN models were often more accurate in the daytime (59.9% average classification accuracy across all sites) than night time (49.8%) (Appendix Figure E) which may result from predictive features such as people, traffic and marketplace indicators (e.g., umbrellas) being present, and more visible in the day, as in Appendix Figure D, since street lighting conditions vary across our sites.

$PM_{2.5}$ classification had lower accuracy than that of noise in most model and site combinations (Figure 3). There was also a larger discrepancy between the predictive performance of CNN and GBM models for $PM_{2.5}$, with CNN models achieving 30-55% classification accuracy and GBM models 15-25%, though both outperformed null model benchmarks. Sites with higher classification accuracy for noise performed less well for air, and vice versa; for example, the three poorest performing sites for noise (University of Ghana, Ashaiman and East Legon) had the greatest accuracy for CNN models when predicting $PM_{2.5}$. Accuracy of CNN models reached 70-90% for neighbouring class classification, and that of GBM models 30-50%. Unlike noise, performance of $PM_{2.5}$ classification using CNN models differed little between day (40.2% average classification accuracy across all sites) and night time (39.6%) and had higher accuracy during the Harmattan period (57.7%) than in other times (35.1%) (Appendix Figure E), whereas GBM models had no consistent advantage during the Harmattan (17.5%) than in other times (19.9%) .

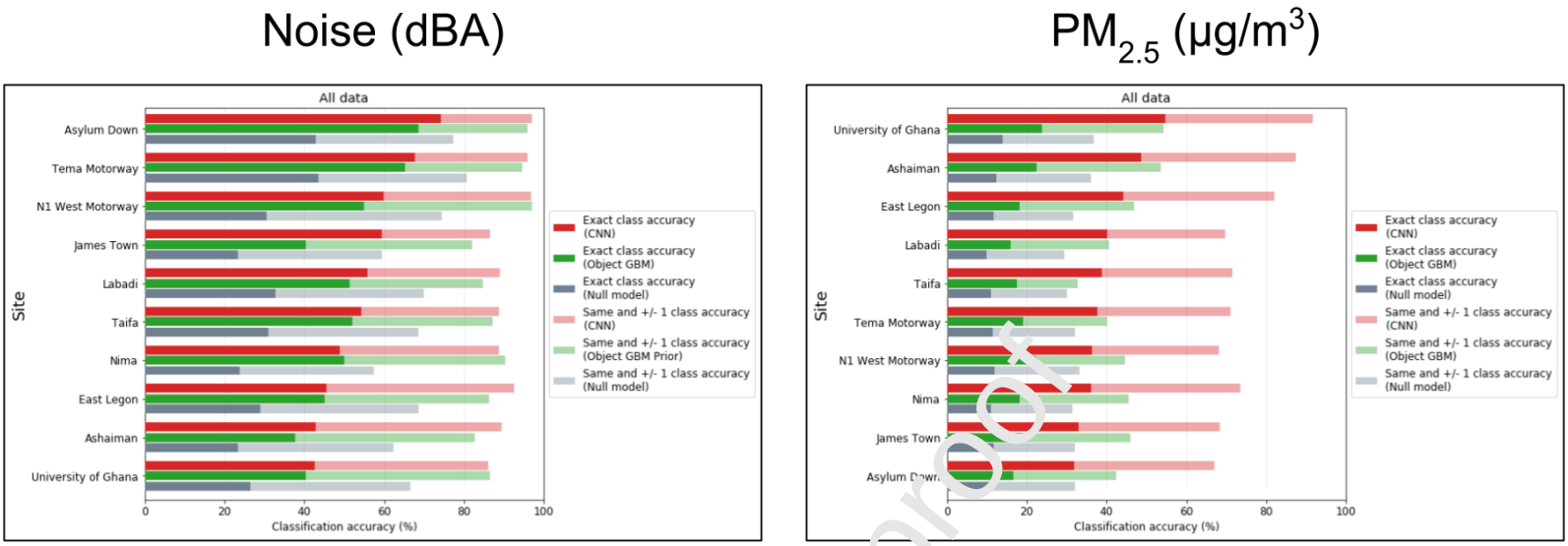


Figure 3. The classification accuracy achieved by CNN and GBM models trained and tested on images from the same fixed site (Question 1a) is shown for noise and $PM_{2.5}$.

When trained at one fixed site and tested at a different fixed site (Question 1b), accuracy dropped compared with same-site testing and CNN models for noise and $PM_{2.5}$, and performed similar to null model benchmarks, demonstrating the inability of models to generalise from the measurements at a single site (Figure 4). For noise, the variation in accuracy of GBM models was greater than that of CNNs, but on average achieved greater improvement over null model benchmarks (+2.4%) than did CNNs (+0.9%). For $PM_{2.5}$ accuracy ranged 7-20% for both GBM and CNN models, with the latter achieving greater improvement over the null model benchmarks (+0.9%) than did GBMs (+0.7%). In all cases, accuracy and null model performance broadly increased with Bhattacharya coefficient, a measure of the similarity of pollution distributions

between training and testing site (Bhattacharyya, 1943).

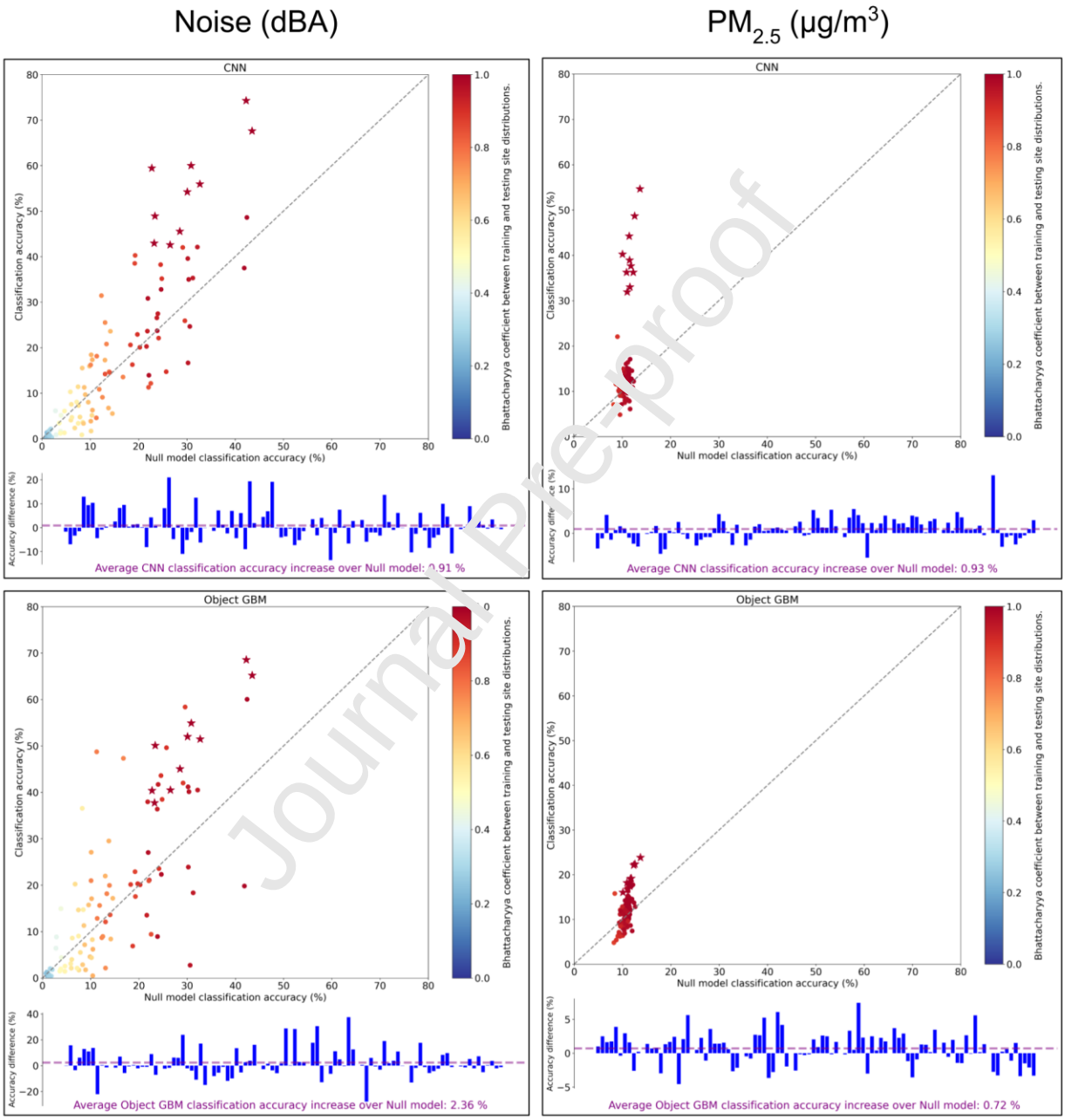


Figure 4. The classification accuracy achieved by the CNN (left) and GBM (right) models trained and tested from images at one fixed site and tested at a different fixed site (Question 1b) for both noise (top) and $PM_{2.5}$ (bottom) prediction. Points are coloured by the Bhattacharya

coefficient between the pollution distributions between the training and testing sites, which is a measure of the overlap between the distributions. Data points with a star indicate testing and training performed at the same site, as in 1a. All null models are from the fixed site used for training. Below each scatterplot the relative improvements in classification accuracy over the null model accuracy is given for each data point (i.e. the vertical distance between the round points and the dashed diagonal line in the scatterplot); the purple dashed line shows the average across all data points, illustrating whether models achieved improvement over their benchmarks overall.

Training on nine fixed sites with abundant (~1,000,000) data (Question 2a) produced similar results for generalising to a single, unseen fixed site as for models trained at single fixed sites (Appendix Figure F). For noise, accuracies were at best 30-40% for both CNN and GBM models, and in some instances similar to the null model; neither CNN or GBM models had a distinct advantage. For $PM_{2.5}$, both CNN and GBM model accuracy remained similar to null model performance. CNN and GBM models trained using fewer images (~100,000) from ~90% of rotating sites (121-122 sites) for classifying noise and $PM_{2.5}$ at the remaining rotating sites (Questions 2b) outperformed their respective null models (Figure 5). As in temporal transferability (Question 1a), models performed better in classifying noise than $PM_{2.5}$, but the advantage of CNN over GBM models disappeared. Noise models had 25% accuracy for exact class and 65% when allowing for neighbouring class classification, with little variance (± 2.5%) between folds, whilst $PM_{2.5}$ models had 17.5% accuracy, and 47.5% allowing for neighbouring class classification. For noise, but not for $PM_{2.5}$, CNN accuracy for daytime images was significantly higher than on night time images (Appendix Figure G). CNN models had similar

performance across all land use categories for both forms of pollution (Appendix Figure G).

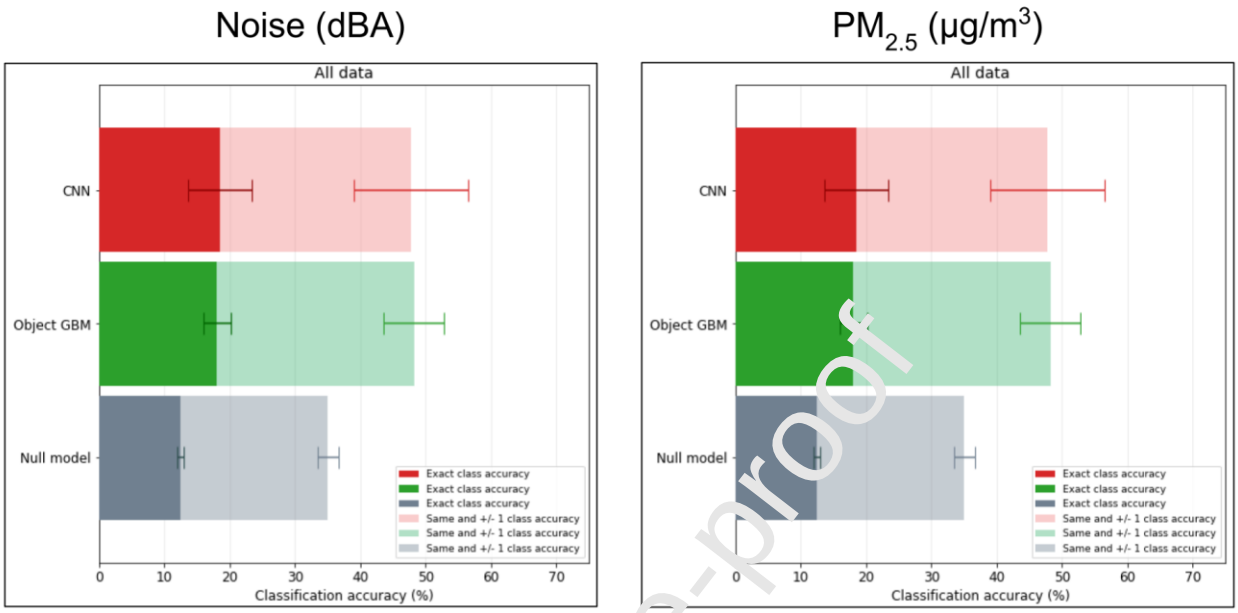


Figure 5. Classification accuracy achieved by CNN and GBM models trained and tested on images from rotating sites (Question 2b) for both noise and $PM_{2.5}$ prediction. Accuracies are shown as the average over the folds of training data, as shown in Figure 1. The bars show the standard deviation of the accuracy across different folds.

When comparing the approaches in Questions 2a (fewer sites with more data per sites) and 2b (more sites with fewer data per sites) with consistent test data (Question 2c), models trained on a smaller amount of data from many (rotating) sites performed better than those trained on many times more data from a smaller number of (fixed) sites, for both noise and air pollution models (Appendix Figure H). This suggests that with finite monitoring capacity, a diversity of locations for data gathering is more likely to produce spatial generalisability for pollution prediction using CNN models than long-term capture of data at fewer locations, highlighting the importance of optimising the spatial as well as temporal representativeness of data within cities for pollution

modelling, as seen in other domains of computer vision (Schat et al., 2020).

4.1) Object feature importance

In GBM models developed in 1a, cars, people, taxis, umbrellas and tro-tros contributed most to predictions for both noise and $PM_{2.5}$ (Figure 5). For $PM_{2.5}$, debris and trucks also contributed to prediction accuracy. These object categories were frequently detected in fixed site images. We calculated the Spearman correlation between object counts and noise and $PM_{2.5}$ levels across images at each fixed site (Appendix Table E) in order to test whether this correlation explained an object's permutation importance. The explained variance, calculated as the square of Pearson correlation between object-pollution correlations and the permutation importance for each object was 0.86 for noise and 0.76 for $PM_{2.5}$. Heuristically, the greater the correlation of an objects' counts with that of pollution, the greater its contribution to model prediction accuracy, which may also indicate why noise models in 1a performed better than those for $PM_{2.5}$. Since objects visible in the images had greater impact on the accuracy of noise prediction compared to $PM_{2.5}$, CNN models for noise may have learned to rely on the same features as those used by the GBM models prediction. This may in turn explain the similar performance between the two models in 1a, which we further examine in the section below.

As shown in Figure 6, the objects with the highest permutation importance were various types of vehicles, consistent with previous research on the significant contributions from road traffic to both air and noise pollution (Dionisio Kathie L. et al., 2010; Onuu, 2000; Rooney et al., 2012). In addition, umbrellas were frequently present in images due to their extended use in the daytime to protect market vendors and their merchandise from the sun and rain. Markets also attract high

levels of vehicular traffic (Agyapong and Ojo, 2018), people (Asante, 2020; Asante and Mills, 2020), and roadside cooking and food vending, which collectively increase noise and air pollution (Alli et al., 2021; Clark et al., 2021). Furthermore, for $PM_{2.5}$ models, debris had higher than average feature importance. Although not a source, debris is more visible and readily detected by our object detection algorithm during daylight hours, serving as a proxy for time of day, and for sites with diurnal patterns of $PM_{2.5}$. In addition, debris may be more visible when unobscured by other objects, acting as an implicit indicator for a lack of crowds or traffic, and instances where the road surface, from which dust particles may be resuspended, is exposed.

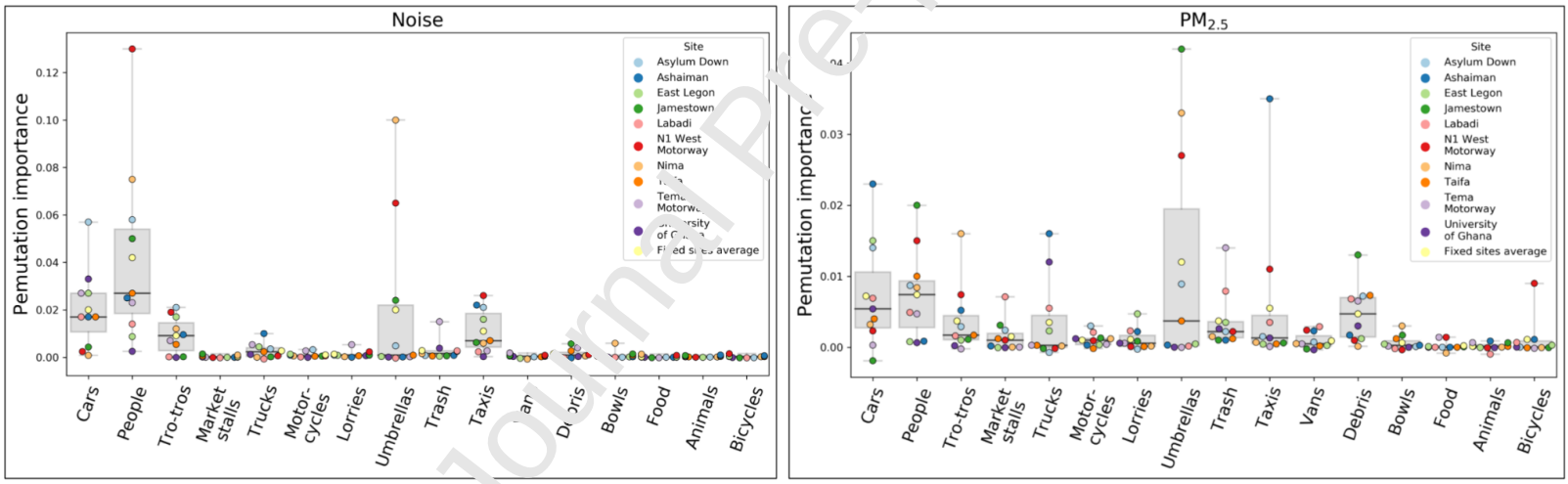


Figure 6. Permutation importance for each object used as inputs to the noise and $PM_{2.5}$ GBM models in Question 1a, for each fixed site. Permutation importance is calculated on the test set for each model, as shown in Figure 1.

4.2) Harmattan influence

To probe why CNN models in 1a performed better during the Harmattan season, when $PM_{2.5}$ levels were much higher, we derived characteristics of images related to changes in hue and haze

between Harmattan and non-Harmattan periods, since previous work has demonstrated that changes due to Harmattan dust, such as an increase in "redness" and haze are indicators and predictive factors for pollution (Adetunji et al., 1979; Anuforom, 2007; Ette and Olorode, 1988; McTainsh, 1980; Ochei and Adenola, 2018; Pinker et al., 1994), as well as light scattering from dust (Groblicki et al., 1981; Waggoner and Weiss, 1980). Other approaches for predicting pollution from images have also used these features (Feng et al. 2021; Wang et al. 2022; Liu et al. 2015; Ganji et al, 2020) and we therefore created feature metrics which relate to qualities of hue and haze in our images in order to study our models. For daytime images, we compared mean pixel intensity in each colour channel between these periods: red, green and blue. For night-time images in single-channel grayscale, we used mean pixel intensity and pixel intensity standard deviation (SD). Red pixel intensity was greater during the Harmattan period, whilst the opposite was seen for blue (Appendix Figure I). Furthermore, night-time pixel intensity was higher at eight of ten sites during Harmattan, while pixel SD was lower at seven sites. To infer to what extent this information was used by our CNN models for $PM_{2.5}$ classification, we calculated Spearman correlations between mean red and blue pixel intensity and each image's associated $PM_{2.5}$ value (Appendix Table F). The average across sites was 0.11 for red pixel intensity and -0.21 for blue pixel intensity. Similarly, grayscale pixel intensity tended to be positively correlated with air pollution, 0.10, whilst grayscale pixel SD was negatively correlated, -0.18, consistent with light sources appearing more diffuse in hazy conditions due to light scattering. The magnitude of these correlations was also greater during the Harmattan period, during which CNN models in 1a had greater accuracy (Appendix Table F). The Pearson correlation across all 10 sites between the aforementioned Spearman correlations for red pixel intensity and the accuracy of a CNN model at that site in 1a, was 0.73. In heuristic terms, the greater the

correspondence between redness of image and air pollution, the better the model performed on average.

4.3) Model interpretability across time and space

Our results suggest that specific features, such as the objects selected for the GBM model, are better attuned at predicting noise, whilst predictive performance for $PM_{2.5}$ is somewhat improved by leveraging more complex visual features from images such as red pixel channel intensity and haziness. This is supported by the discrepancy between day and night image accuracy for CNN models for noise in 1a and 2b, which may result from objects being more visible and present in the day than night, allowing CNN models to make more accurate predictions of noise, but not $PM_{2.5}$, using daytime images.

These observations also suggest that features learned by the CNN for noise classification are likely similar to the objects used by the GBM models, whilst being of less importance for CNN models for $PM_{2.5}$. To investigate this, we generated a sample of gradient class activation maps (Selvaraju et al., 2017) (Grad-CAMs) for our CNN models in 1a. Grad-CAMs use gradient descent to work backwards from a network's class prediction from a given image to the regions in the image itself which most contributed to that prediction. Figure 7 shows that noise models focus either on visible objects or the location of main thoroughfares, whilst $PM_{2.5}$ models tend to focus on either fixed features of the built environment, or the sky, supporting the possibility of the object-driven nature of noise models and the reliance of $PM_{2.5}$ CNN models on complex visual features. This may be due to the fact that noise is transitory and spatially linked to sources or their proxies such as market stalls and their associated umbrellas. By contrast, $PM_{2.5}$ persists

at a location longer than its sources, and is composed of both local (e.g., traffic) and non-local (dust, neighbouring regions' emissions) sources.

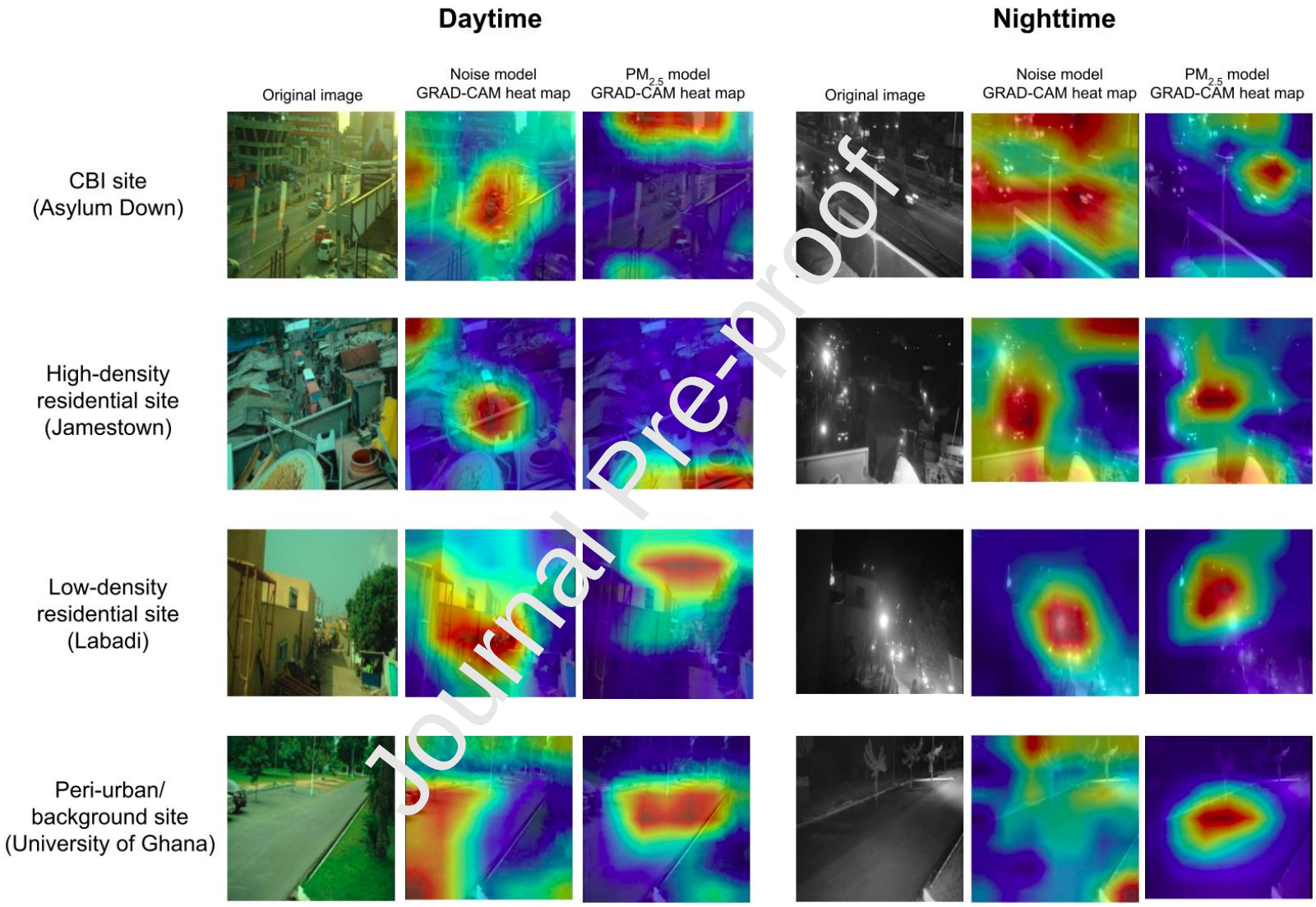


Figure 7. Grad-CAM visualisation heat maps for four fixed sites of different land-use categories, obtained from CNN models. The highlighted regions in each image (red) indicate the features most salient to the model's prediction, for that image. The same image is shown as used by the corresponding noise and $PM_{2.5}$ models, developed in Question 1a.

5) Conclusions

We used a unique dataset of co-located images and pollution measurements, and two different modelling approaches, to investigate how images predict spatially and temporally resolved noise and air pollution measurements in a major city in Africa. Most models developed in this work surpassed null model accuracy baselines, indicating that models can learn information latent to images, beyond extrapolation from outcome data. Models had similar performance across different land-use categories within the city and across both day and night time, with a slight advantage to model accuracy in the daytime for noise. However, even when training and testing models at single sites, classifying noise and $PM_{2.5}$ with either CNN or GBM models had moderate performance. No model surpassed 80% classification accuracy, and performance was considerably lower when testing on previously unseen locations. This may be due to a combination of factors including the static nature of images, which may fail to capture transitory sources, e.g., emergency vehicle sirens, compared with exposures which are averaged over one minute's observation. Furthermore some pollution sources (and predictors) are non-local or out of the field of view of the corresponding image's camera. Since the models developed in questions 1b to 2c rely on a single images from consumer-grade digital cameras to predict pollution in an unseen location, our dataset and methodology also informs on the viability of pollution modelling from comparable camera technology, including CCTV networks which are increasingly deployed in African cities and mobile phone camera capture, as has been used elsewhere in the literature. In addition, we intentionally tested models under the challenging condition of estimating pollution from a single image in time and space, in order to independently and conservatively assess the additional benefit images may confer above data extrapolation. Future work could improve model accuracy by making use of CNN architectures

with multiple image inputs across time, or in the case of feature-driven prediction, with object counts across a series of time-points prior to the moment in time whose pollution levels are estimated. Similarly, image and/or object counts from neighbouring sites might also help predict pollution at a given location. These may prove especially beneficial for the prediction of $PM_{2.5}$, whose presence is more persistent than many of its transitory sources, such as travelling cars.

Where pollution sources are complex and vary widely across small spatial scales, the inclusion of data from many sites improved the spatial generalisability of CNN models in comparison to an abundance of data from a small number of locations. Our results highlight the importance of optimising the spatial as well as temporal representativeness of data gathered within cities for pollution modelling, as seen in other domains of computer vision (61). Although spatially representative data improved model performance, for CNNs model accuracy also varied under particular times of day and seasons, related to time-varying environmental factors. We also find that modelsare capable of making comparably accurate estimates for nighttime air and noise pollution, particularly for $PM_{2.5}$ estimation, when such data is gathered and used for modelling, alongside corresponding images. This shows the need for temporally diverse, paired pollution and image data that capture urban environmental change on both short (< 1 hour) and long (~1 year) timescales.

Overall, our results show that inference of noise and $PM_{2.5}$ from imagery is a feasible but challenging task, especially as the spatial and temporal scale of prediction becomes smaller, which is relevant for detailed policy formulation, such as dynamic smart congestion pricing, and evaluation of impacts on human exposure. Therefore, accurate and generalisable estimates of

short timescale pollution in cities continues to require primary data collection at representative and diverse scales to support these efforts.

AUTHOR INFORMATION

Corresponding Author

Ricky Nathvani (r.nathvani@imperial.ac.uk)

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported by the Pathways to Equitable Healthy Cities grant from the Wellcome Trust [209376/Z/17/Z]. This work was also supported by a GCRF Digital Innovation for Development in Africa network grant from UKRI [EP/T029145/1]. SC, ABM and TR were supported by the Imperial College President's PhD scholarship. SC was supported by a Canadian Institutes for Health Research (CIHR) Foreign Study Doctoral Scholarship. For the purpose of Open Access, the author has applied a CC BY public copyright licencelicense to any Author Accepted Manuscript version arising from this submission.

Acknowledgement

We thank Giulia Mangiameli and Abeer Arif for project management and coordination of activities and Professor John Spengler and Professor George Thurston for their comments, references and insight.

Data availability

Our analysis code, trained models, object count data and site metadata can be downloaded from http://globalenvhealth.org/code-data-download/ and http://equitablehealthycities.org/data-download/ upon publication of the paper. Requests for re-analysis of images should be sent to the corresponding authors.

References

Adetunji, J., McGregor, J., Ong, C.K., 1979. Harmattan Haze. Weather 34, 430–436. https://doi.org/10.1002/j.1477-8696.1979.tb03389.x

Agyapong, F., Ojo, T.K., 2018. Managing traffic congestion in the Accra Central Market, Ghana. J. Urban Manag. 7, 85–96. https://doi.org/10.1016/j.jum.2018.04.002

Alli, A.S., Clark, S.N., Hughes, A., Nimo, J., Bedford-Moses, J., Baah, S., Wang, J., Vallarino, J., Agyemang, E., Barratt, B., Beddows, A., Kelly, F., Owusu, G., Baumgartner, J., Brauer, M., Ezzati, M., Agyei-Mensah, S., Arku, R.E., 2021. Spatial-temporal patterns of ambient fine particulate matter (PM2.5) and black carbon (BC) pollution in Accra. Environ. Res. Lett. 16, 074015. https://doi.org/10.1088/1748-9326/ac074a

Amegah, A.K., Agyei-Mensah, S., 2017. Urban air pollution in Sub-Saharan Africa: Time for action. Environ. Pollut. Barking Essex 1987 220, 738–743. https://doi.org/10.1016/j.envpol.2016.09.042

Anuforom, A.C., 2007. Spatial distribution and temporal variability of Harmattan dust haze in sub-Sahel West Africa. Atmos. Environ. 41, 9079–9090. https://doi.org/10.1016/j.atmosenv.2007.08.003

Asante, L.A., 2020. Urban governance in Ghana: the participation of traders in the redevelopment of Kotokuraba Market in Cape Coast. Afr. Geogr. Rev. 39, 361–378. https://doi.org/10.1080/19376812.2020.1726193

Asante, L.A., Mills, R.O., 2020. Exploring the Socio-Economic Impact of COVID-19 Pandemic in Marketplaces in Urban Ghana. Afr. Spectr. 55, 170–181. https://doi.org/10.1177/0002039720943612

Basner, M., McGuire, S., 2018. WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Effects on Sleep. Int. J. Environ. Res. Public. Health 15, E519. https://doi.org/10.3390/ijerph15030519

Belharbi, S., Ayed, I.B., McCaffrey, L., Granger, E., 2020. Non-parametric Uni-modality Constraints for Deep Ordinal Classification. ArXiv191110720 Cs Stat.

Bhattacharyya, A., 1943. On a measure of divergence between two statistical populations defined by their probability distributions. Bull Calcutta Math Soc 35, 99–109.

Brauer, M., Guttikunda, S.K., K A, N., Dey, S., Tripathi, S.N., Weagle, C., Martin, R.V., 2019. Examination of monitoring approaches for ambient air pollution: A case study for India. Atmos. Environ. 216, 116940. https://doi.org/10.1016/j.atmosenv.2019.116940

Breuning-Madsen, H., Awadzi, T.W., 2005. Harmattan dust deposition and particle size in Ghana. CATENA 63, 23–38. https://doi.org/10.1016/j.catena.2005.04.001

Chakma, A., Vizena, B., Cao, T., Lin, J., Zhang, J., 2017. Image-based air quality analysis using deep convolutional neural network, in: 2017 IEEE International Conference on Image Processing (ICIP). Presented at the 2017 IEEE International Conference on Image Processing (ICIP), pp. 3949–3952. https://doi.org/10.1109/ICIP.2017.8297023

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Clark, S.N., Alli, A.S., Brauer, M., Ezzati, M., Baumgartner, J., Toledano, M.B., Hughes, A.F., Nimo, J., Moses, J.B., Terkpertey, S., Vallarino, J., Agyei-Mensah, S., Agyemang, E., Nathvani, R., Muller, E., Bennett, J., Wang, J., Beddows, A., Kelly, F., Barratt, B., Beevers, S., Arku, R.E., 2020. High-resolution spatiotemporal measurement of air and environmental noise pollution in Sub-Saharan African cities: Pathways to Equitable Health Cities Study protocol for Accra, Ghana. BMJ Open 10, e035798. https://doi.org/10.1136/bmjopen-2019-035798

Clark, S.N., Alli, A.S., Nathvani, R., Hughes, A., Ezzati, M., Brauer, M., Toledano, M.B., Baumgartner, J., Bennett, J.E., Nimo, J., Bedford Moses, J., Baah, S., Agyei-Mensah, S., Owusu, G., Croft, B., Arku, R.E., 2021. Space-time characterization of community noise and sound sources in Accra, Ghana. Sci. Rep. 11, 11113. https://doi.org/10.1038/s41598-021-90454-6

Deng, L., Kang, J., Zhao, W., Jambrošić, K., 2020. Cross-National Comparison of Soundscape in Urban Public Open Spaces between China and Croatia. Appl. Sci. 10, 960. https://doi.org/10.3390/app10030960

Dionisio Kathie L., Rooney Michael S., Arku Raphael E., Friedman Ari B., Hughes Allison F., Vallarino Jose, Agyei-Mensah Samuel, Spengler John D., Ezzati Majid, 2010. Within-Neighborhood Patterns and Sources of Particle Pollution: Mobile Monitoring and Geographic Information System Analysis in Four Communities in Accra, Ghana. Environ. Health Perspect. 118, 607–613. https://doi.org/10.1289/ehp.0901365

Ebare, M.N., Omuemu, V.O., Isah, E.C., 2011. Assessment of noise levels generated by music shops in an urban city in Nigeria. Public Health 125, 660–664. https://doi.org/10.1016/j.puhe.2011.06.009

Ette, A.I.I., Olorode, D.O., 1988. Technical note The effects of the harmattan dust on air conductivity and visibility at Ibadan, Nigeria. Atmospheric Environ. 1967 22, 2625–2627. https://doi.org/10.1016/0004-6981(88)90499-4

Ezzati, M., Webster, C.J., Doyle, Y.G., Rashid, S., Owusu, G., Leung, G.M., 2018. Cities for global health. BMJ 363, k3794. https://doi.org/10.1136/bmj.k3794

Feng, L., Yang, T., Wang, Z., 2021. Performance evaluation of photographic measurement in the machine-learning prediction of ground PM2.5 concentration. Atmos. Environ. 262, 118623. https://doi.org/10.1016/j.atmosenv.2021.118623

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29, 1189–1232. https://doi.org/10.1214/aos/1013203451

Ganji, A., Minet, L., Weichenthal, S., Hatzopoulou, M., 2020. Predicting Traffic-Related Air Pollution Using Feature Extraction from Built Environment Images. Environ. Sci. Technol. 54, 10688–10699. https://doi.org/10.1021/acs.est.0c00412

Groblicki, P.J., Wolff, G.T., Countess, R.J., 1981. Visibility-reducing species in the denver "brown cloud"—I. Relationships between extinction and chemical composition. Atmospheric Environ. 1967, Plumes and Visibility Measurements and Model Components-Supplement 15, 2473–2484. https://doi.org/10.1016/0004-6981(81)90063-9

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T., 2018. Recent advances in convolutional neural networks. Pattern Recognition 77, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Gu, K., Qiao, J., Li, X., 2019. Highly Efficient Picture-Based Prediction of PM2.5 Concentration. IEEE Trans. Ind. Electron. 66, 3176–3184. https://doi.org/10.1109/TIE.2018.2840515

Hong, K.Y., Pinheiro, P.O., Weichenthal, S., 2020. Predicting outdoor ultrafine particle number concentrations, particle size, and noise using street-level images and audio data. Environ. Int. 144, 106044. https://doi.org/10.1016/j.envint.2020.106044

Huh, M., Agrawal, P., Efros, A.A., 2016. What makes ImageNet good for transfer learning? ArXiv160808614 Cs.

Hyslop, N.P., 2009. Impaired visibility: the air pollution people see. Atmos. Environ., Atmospheric Environment - Fifty Years of Endeavour 43, 182–195. https://doi.org/10.1016/j.atmosenv.2008.09.067

Kammen, D.M., Sunter, D.A., 2016. City-integrated renewable energy for urban sustainability. Science 352, 922–928. https://doi.org/10.1126/science.aad9302

Kelly, F.J., Zhu, T., 2016. Transport solutions for cleaner air. Science 352, 934–936. https://doi.org/10.1126/science.aaf3420

Khan, J., Ketzel, M., Kakosimos, K., Sørensen, M., Jensen, S.S., 2018. Road traffic air and noise pollution exposure assessment - A review of tools and techniques. Sci. Total Environ. 634, 661–676. https://doi.org/10.1016/j.scitotenv.2018.03.374

Lafon, S., Rajot, J.-L., Alfaro, S.C., Gaudichet, A., 2004. Quantification of iron oxides in desert aerosol. Atmos. Environ. 38, 1211–1218. https://doi.org/10.1016/j.atmosenv.2003.11.006

LeCun, Y.A., Bottou, L., Orr, G.B., Müller, K.-R., 2012. Efficient BackProp, in: Montavon, G., Orr, Geneviève B., Müller, K.-R. (Eds.), Neural Networks: Tricks of the Trade: Second Edition, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 9–48. https://doi.org/10.1007/978-3-642-35289-8_3

Liu, C., Tsow, F., Zou, Y., Tao, N., 2016. Particle Pollution Estimation Based on Image Analysis. PLOS ONE 11, e0145955. https://doi.org/10.1371/journal.pone.0145955

Liu, X., Song, Z., Ngai, E., Ma, J., Wang, W., 2015. PM2:5 monitoring using images from smartphones in participatory sensing, in: 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Presented at the 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 630–635. https://doi.org/10.1109/INFCOMW.2015.7179456

McTainsh, G., 1980. Harmattan dust deposition in northern Nigeria. Nature 286, 587–588. https://doi.org/10.1038/286587a0

Nathvani, R., Clark, S.N., Muller, E., Alli, A.S., Bennett, J.E., Nimo, J., Moses, J.B., Baah, S., Metzler, A.B., Brauer, M., Suel, E., Hughes, A.F., Rashid, T., Gemmell, E., Moulds, S., Baumgartner, J., Toledano, M., Agyemang, E., Owusu, G., Agyei-Mensah, S., Arku,

R.E., Ezzati, M., 2022. Characterisation of urban environment and activity across space and time using street images and deep learning in Accra. Sci. Rep. 12, 20470. https://doi.org/10.1038/s41598-022-24474-1

Ochei, M.C., Adenola, E., 2018. Variability of Harmattan Dust Haze Over Northern Nigeria. J. Pollut. 1, 8.

Onuu, M.U., 2000. Road Traffic Noise in Nigeria: Measurements, analysis and evaluation of nuisance. J. Sound Vib. 233, 391–405. https://doi.org/10.1006/jsvi.1999.2832

Ozkaynak, H., Schatz, A.D., Thurston, G.D., Isaacs, R.G., Husar, R.B., 1985. Relationships between Aerosol Extinction Coefficients Derived from Airport Visual Range Observations and Alternative Measures of Airborne Particle Mass. J. Air Pollut. Control Assoc. 35, 1176–1185. https://doi.org/10.1080/00022470.1985.10466020

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pinker, R.T., Idemudia, G., Aro, T.O., 1994. Characteristic aerosol optical depths during the Harmattan Season on sub-Sahara Africa. Geophys. Res. Lett. 21, 685–688. https://doi.org/10.1029/93GL03547

Pope, C.A., Dockery, D.W., 2006. Health Effects of Fine Particulate Air Pollution: Lines that Connect. J. Air Waste Manag. Assoc. 56, 709–742. https://doi.org/10.1080/10473289.2006.10464485

Qi, M., Hankey, S., 2021. Using Street View Imagery to Predict Street-Level Particulate Air Pollution. Environ. Sci. Technol. 55, 2695–2704. https://doi.org/10.1021/acs.est.0c05572

Rooney, M.S., Arku, R.E., Dionisio, K.L., Paciorek, C., Friedman, A.B., Carmichael, H., Zhou, Z., Hughes, A.F., Vallarino, J., Agyei-Mensah, S., Spengler, J.D., Ezzati, M., 2012. Spatial and temporal patterns of particulate matter sources and pollution in four communities in Accra, Ghana. Sci. Total Environ. 435–436, 107–114. https://doi.org/10.1016/j.scitotenv.2012.06.077

Schat, E., Schoot, R. van de, Kouw, W.M., Veen, D., Mendrik, A.M., 2020. The data representativeness criterion: Predicting the performance of supervised classification based on data set similarity. PLOS ONE 15, e0237009. https://doi.org/10.1371/journal.pone.0237009

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Networks 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, in: 2017 IEEE International Conference on Computer Vision (ICCV). Presented at the 2017 IEEE International Conference on Computer Vision (ICCV), pp. 618–626. https://doi.org/10.1109/ICCV.2017.74

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. J. Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0

Sorek-Hamer, M., Von Pohle, M., Sahasrabhojanee, A., Akbari Asanjan, A., Deardorff, E., Suel, E., Lingenfelter, V., Das, K., Oza, N.C., Ezzati, M., Brauer, M., 2022. A Deep Learning Approach for Meter-Scale Air Quality Estimation in Urban Environments Using Very High-Spatial-Resolution Satellite Imagery. Atmosphere 13, 696. https://doi.org/10.3390/atmos13050696

Suel, E., Sorek-Hamer, M., Moise, I., von Pohle, M., Sahasrabhojanee, A., Asanjan, A.A., Arku, R.E., Alli, A.S., Barratt, B., Clark, S.N., Middel, A., Deardorff, E., Lingenfelter, V., Oza, N.C., Yadav, N., Ezzati, M., Brauer, M., 2022. What You See Is What You Breathe? Estimating Air Pollution Spatial Variation Using Street-Level Imagery. Remote Sens. 14, 3429. https://doi.org/10.3390/rs14143429

Sun, C., Shrivastava, A., Singh, S., Gupta, A., 2017. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era, in: 2017 IEEE International Conference on Computer Vision (ICCV). Presented at the 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, pp. 843–852. https://doi.org/10.1109/ICCV.2017.97

United Nations, Department of Economic and Social Affairs, & Population Division, 2019. . World urbanization prospects: the 2018 revision.

Waggoner, A.P., Weiss, R.E., 1980. Comparison of fine particle mass concentration and light scattering extinction in ambient aerosol. Atmospheric Environ. 1967 14, 623–626. https://doi.org/10.1016/0004-6981(80)90098-0

Wang, X., Wang, M., Liu, X., Zhang, X., Li, R., 2022. A PM2.5 concentration estimation method based on multi-feature combination of image patches. Environ. Res. 211, 113051. https://doi.org/10.1016/j.envres.2022.113051

Weagle, C.L., Snider, G., Li, C., van Donkelaar, A., Philip, S., Bissonnette, P., Burke, J., Jackson, J., Latimer, R., Stone, E., Abboud, I., Akoshile, C., Anh, N.X., Brook, J.R., Cohen, A., Dong, J., Gibson, M.D., Griffith, D., He, K.B., Holben, B.N., Kahn, R., Keller, C.A., Kim, J.S., Lagrosas, N., Lestari, P., Khian, Y.L., Liu, Y., Marais, E.A., Martins, J.V., Misra, A., Muliana, U., Pratiwi, R., Quel, E.J., Salam, A., Segev, L., Tripathi, S.N., Wang, C., Zhang, Q., Brauer, M., Rudich, Y., Martin, R.V., 2018. Global Sources of Fine Particulate Matter: Interpretation of PM2.5 Chemical Composition Observed by SPARTAN using a Global Chemical Transport Model. Environ. Sci. Technol. 52, 11670–11681. https://doi.org/10.1021/acs.est.8b01658

Wei, X., Chang, N.-B., Bai, K., Gao, W., 2020. Satellite remote sensing of aerosol optical depth: advances, challenges, and perspectives. Crit. Rev. Environ. Sci. Technol. 50, 1640–1725. https://doi.org/10.1080/10643389.2019.1665944

Weichenthal, S., Hatzopoulou, M., Brauer, M., 2019. A picture tells a thousand…exposures: Opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology. Environ. Int. 122, 3–10. https://doi.org/10.1016/j.envint.2018.11.042

Weigand, M., Wurm, M., Dech, S., Taubenböck, H., 2019. Remote Sensing in Environmental Justice Research—A Review. ISPRS Int. J. Geo-Inf. 8, 20. https://doi.org/10.3390/ijgi8010020

Won, T., Eo, Y.D., Sung, H., Chong, K.S., Youn, J., Lee, G.W., 2022. Particulate Matter Estimation from Public Weather Data and Closed-Circuit Television Images. KSCE J. Civ. Eng. 26, 865–873. https://doi.org/10.1007/s12205-021-0865-4

World Health Organization, 2021. WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health

Organization.
Xie, S., Girshick, R., Dollar, P., Tu, Z., He, K., 2017. Aggregated Residual Transformations for Deep Neural Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, pp. 5987–5995. https://doi.org/10.1109/CVPR.2017.634
Zhang, C., Yan, J., Li, C., Wu, H., Bie, R., 2018. End-to-end learning for image-based air quality level estimation. Mach. Vis. Appl. 29, 601–615. https://doi.org/10.1007/s00138-018-0919-x
Zhang, Q., Zhu, S., 2018. Visual interpretability for deep learning: a survey. Front. Inf. Technol. Electron. Eng. 19, 27–39. https://doi.org/10.1631/FITEE.1700808
Zhou, Z., Dionisio, K.L., Verissimo, T.G., Kerr, A.S., Coull, B., Arku, R.E., Koutrakis, P., Spengler, J.D., Hughes, A.F., Vallarino, J., Agyei-Mensah, S., Ezzati, M., 2013. Chemical composition and sources of particle pollution in affluent and poor neighborhoods of Accra, Ghana. Environ. Res. Lett. 8, 044025. https://doi.org/10.1088/1748-9326/8/4/044025

Author contributions

RN, VDG, SNC, EM, MB, ES and ME conceptualised the study. SNC, EM, ASA, JN, JBM, SB, AH, REA and ME designed and implemented the field campaign to collect data. RN, VDG, JEB, HC, and ME developed analytical methods. RN and VDG implemented methods and conducted analyses. RN, VDG, SNC, and ME developed the presentation of results. RN, VDG and ME wrote the original draft. RN, VDG, EM, ES, JB, ABM, MB, AH, TR, EG, SM, JEB, EA, GO, SA-M and REA provided input and revisions. AH, SA-M, REA and ME supervised data collection and analysis.

Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Highlights

- Street-view image based deep learning models can extend pollution estimatation
- Image and feature-based models are complimentary in flexibility and interpretability
- Noise and air models use specific features (e.g. market umbrellas and haze)
- Images and sensor networks can broaden pollution monitoring in African cities
- Data collection for model development should prioritise spatial representativeness