Developing and Evaluating an Integrated Machine Learning Approach to Estimating Near Real-Time Local Level Concentrations of Outdoor Ultrafine Particles using Street-Level Images and City Sounds

Scott Weichenthal*,[1] Alessya Venuta,[1] Marshall Lloyd,[1] Shaainee Ramnawaz,[1] Marianne Hatzopoulou,[2] Doug Petican,[3] Blake Laufer[4]

[1]McGill University, Department of Epidemiology, Biostatistics, and Occupational Health, Montreal, Quebec, Canada; [2]University of Toronto, Department of Civil and Mineral Engineering, Toronto, Ontario, Canada; [3]Consultant, Hamilton, Ontario, Canada; [4]Mistall Insight Inc, Hamilton, Ontario, Canada

A. Capture Images and Spectrograms

Figure 1. Platform installation

Figure 2. Platform components

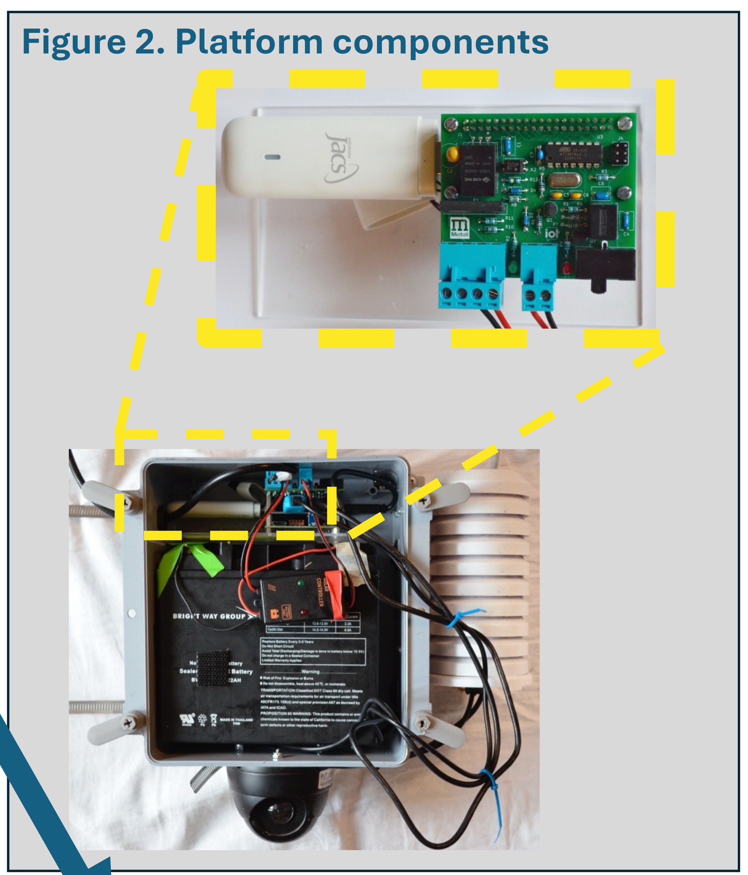
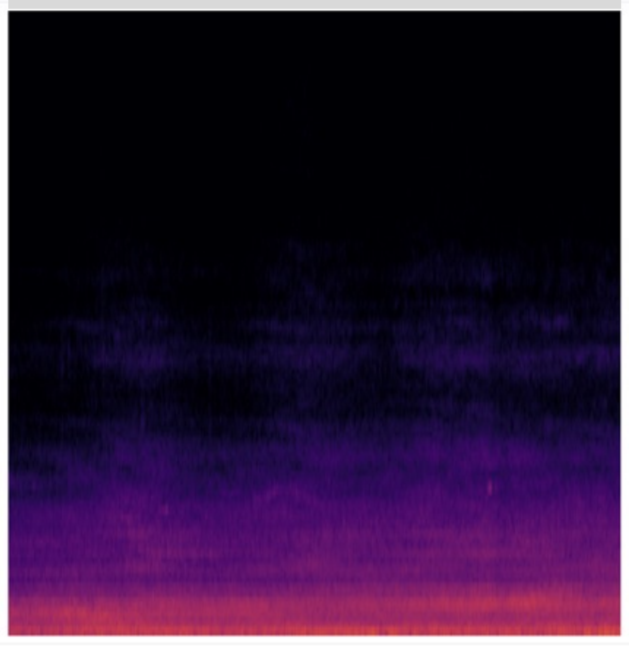


Figure 3. Street-level image

Figure 4. Ten-second spectrogram



B. Measure UFPs and Noise

Figure 5. UFP monitor

Figure 6. Noise monitor

C. Train Separate Deep Convolutional Neural Network (CNN) Models

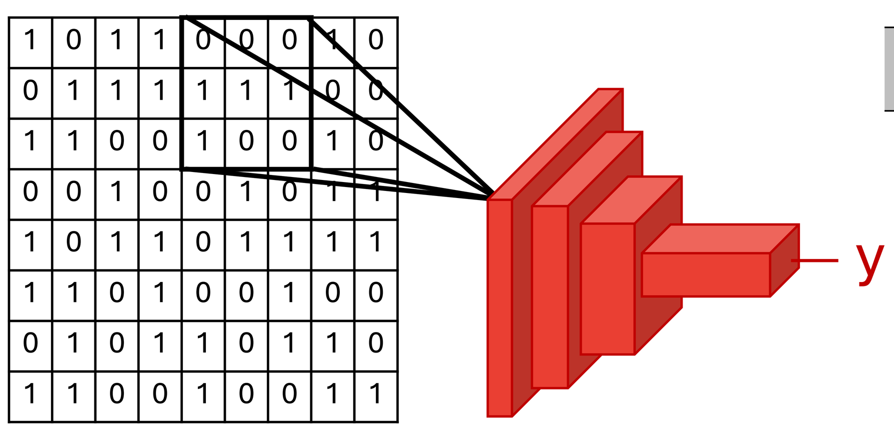


Table 2. CNN models

CNN Model #	Predictor		Predicted Exposure
1	Images	~	log(UFP Concentration)
2	Images	~	log(UFP Size)
3	Images	~	log(Noise)
4	Spectrograms	~	log(UFP Concentration)
5	Spectrograms	~	log(UFP Size)
6	Spectrograms	~	log(Noise)

D. Combine CNN Predictions and Regional Atmospheric Conditions in XGBoost Models

$$\boldsymbol{log(UFP\ Concentration)} \sim xgboost(UFP\ Conc_{CNN\ \#1}, UFP\ Conc_{CNN\ \#4}, UFP\ Size_{CNN\ \#2}, UFP\ Size_{CNN\ \#5}, Temp, Wind\ Speed, PM_{2.5})$$

$$\boldsymbol{log(Noise)} \sim xgboost(Noise_{CNN\ \#3}, Noise_{CNN\ \#6}, UFP\ Size_{CNN\ \#2}, UFP\ Size_{CNN\ \#5}, Temp, Wind\ Speed, PM_{2.5})$$

E. Evaluate Model Performance in Test Set

Figure 7. UFP model performance

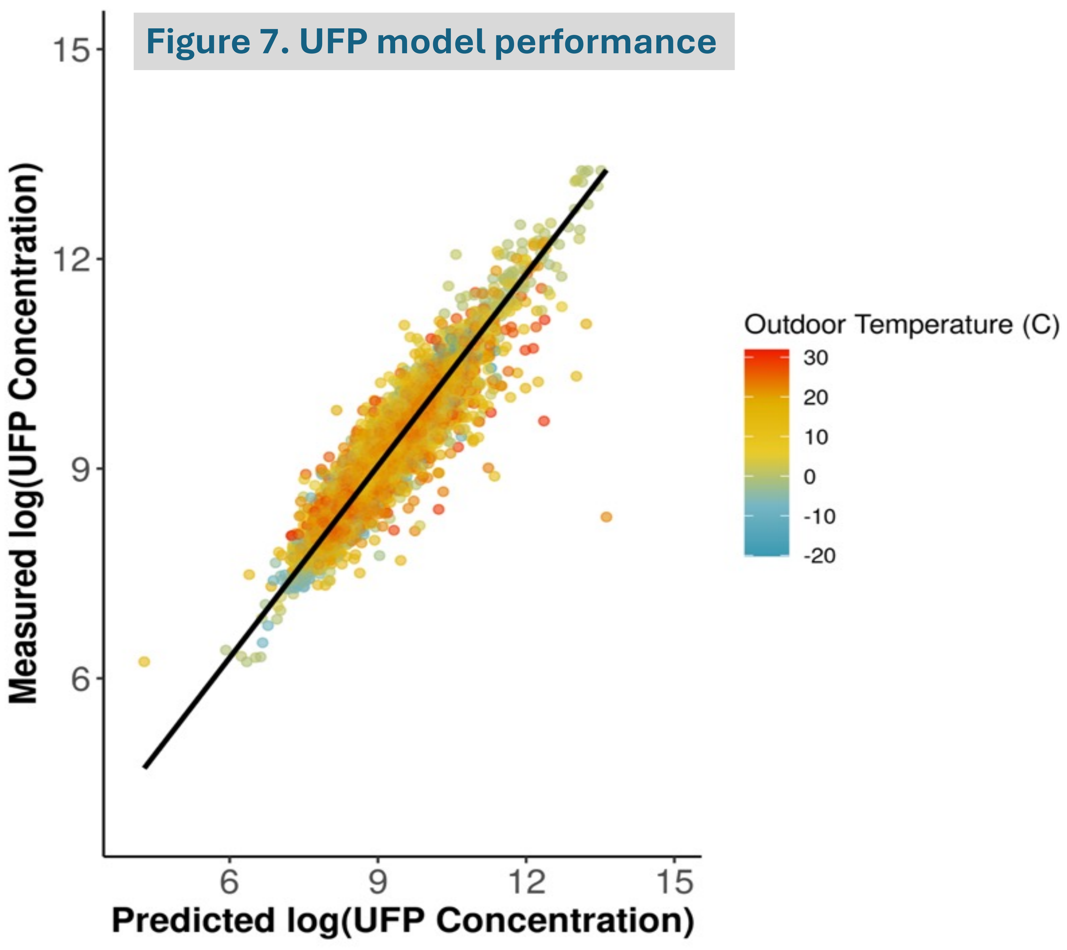


Figure 8. Noise model performance

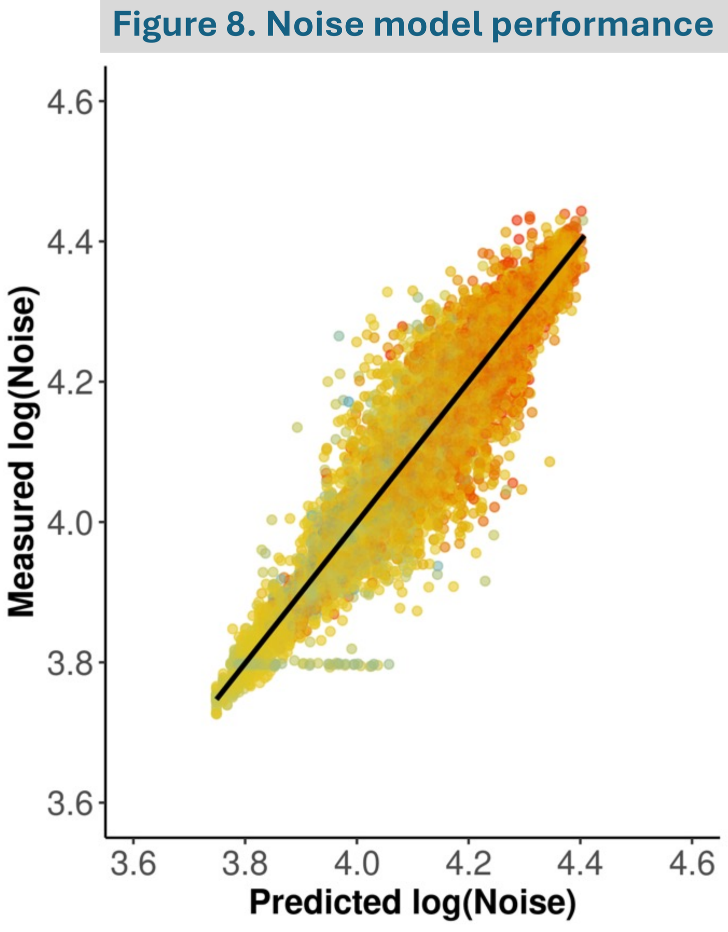


Table 2. Model performance in test set

Exposure	Relationship between Measured vs. Predicted Values in Random Test Set				
	Slope (95% CI)	Intercept (95% CI)	R^2	RMSE	Mean Difference on Original Scale (95% CI)
log(UFP Concentration)	1.012 (1.009, 1.126)	-0.120 (-0.151, -0.089)	0.926	0.195	447 (342, 552)
log(Noise)	1.004 (1.000, 1.007)	-0.017 (-0.031, -0.003)	0.913	0.036	0.039 (0.014, 0.064)

Introduction

- Traffic-related air pollution and noise are known to have adverse impacts on human health
- Most exposure modelling studies focus on long-term exposures, there is a gap in near real-time local level exposures

Study Aim: Develop a new platform that captures street-level images and spectrograms to generate near real-time predictions of outdoor ultrafine particle (UFPs; particulate matter < 100 nm) concentrations and noise (dB(A)).

Summary

A. Platform:

- Solar powered device (Figure 1) mounted on city street pole
- Raspberry Pi computer and modem (Figure 2) transmitted street-level images (Figure 3) and 10-second spectrograms (Figure 4)

B. Exposures:

- UFP number concentrations and UFP size measured every 1 second using Naneos Partector 2 (Figure 5)
- Noise measured every 1 second using Convergence Instruments NSRT Mk4 (Figure 6)

Monitoring Campaign:

- Images, spectrograms, UFPs, and noise were measured at 11 road-side locations across Montreal, Canada during 2021-2022
- Approximately 300,000 samples were collected

Data Preparation:

- A 10-second moving average and log transformations were applied to UFP concentration, UFP size, and noise
- Data randomly split into train, validate, and test sets (80-10-10)

C. Convolutional Neural Network (CNN) Model Training:

- Separate models were trained on images and spectrograms to train UFP number concentration, UFP size, and noise (Table 1)
- Xception or ResNet architectures were used

D. Temporal Adjustment:

- CNN Predictions were combined with regional weather and fine particulate matter concentrations ($PM_{2.5}$) using XGBoost models in the train and validate sets

E. Model Performance in Test Set:

- CNN predictions using images and spectrograms were used to generate XGBoost model predictions in the test set
- Models explained 92.6% of the variance in the observed UFP number concentrations (Figure 7 and Table 2)
- Models explained 91.3% of the variance in the observed noise (Figure 8 and Table 2)

Impact

- New platform offers an efficient means of predicting local-level noise and UFP concentrations in near real-time
- Could be used to support future epidemiological analyses or applied in occupational environments where noise and vehicle emissions are a concern