Efficient Sensing of PAH Concentrations Using Fluorescence Imaging

Gaurav Mahamuni, Anamol Pundle

Mechanical Engineering, University of Washington

CSE 546: Machine Learning

ABSTRACT

- **Polycyclic Aromatic Hydrocarbons (PAHs)** present in combustion particulate matter (PM) e.g. cigarette smoke, can contribute to cardiovascular and respiratory diseases.
- Determining PAH concentration in air pollution samples helps determine toxicity of PM.
- Excitation Emission Matrix (EEM) image analysis is cheap and efficient, and can be used for PAH sensing.
- EEM + Data Analysis can be used to develop data driven sensors for air pollution exposure monitoring.

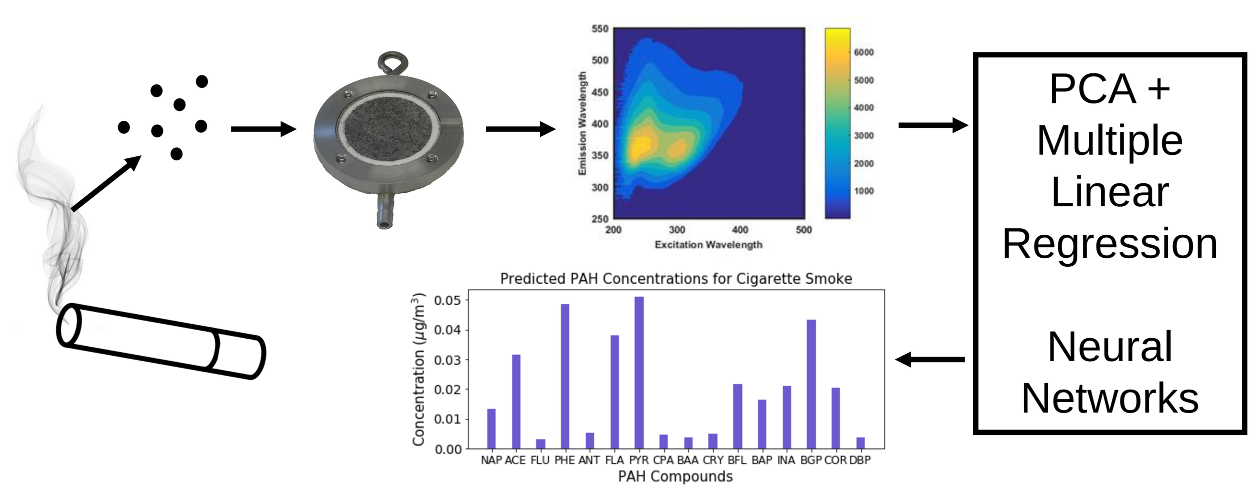


Figure 1. EEM images of PM from sources like cigarette smoke are used to predict concentration of PAH compounds to determine level of toxicity of source particles

DATA COLLECTION METHODS

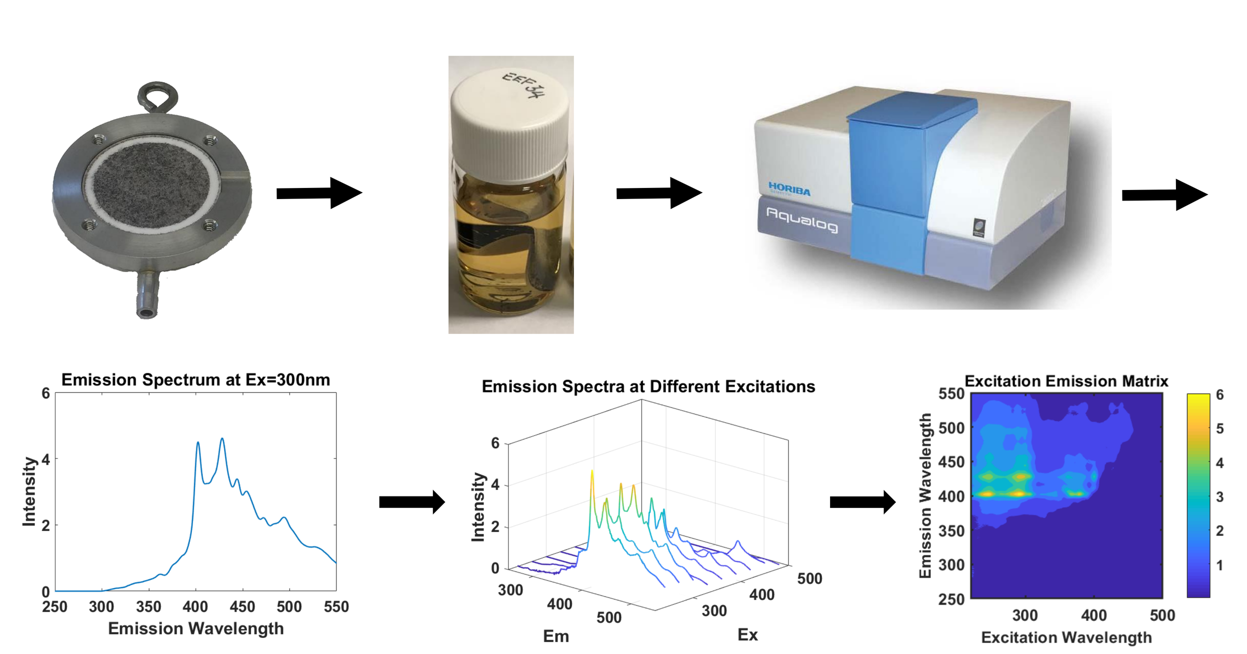


Figure 2. Excitation Emission Matrix (EEM) image sensing has advantages over chromatographic methods as it is quick and requires small volumes of the sample

Figure 3. GCMS is a cumbersome and expensive lab based technique to determine PAH concentrations but is very accurate.

Training model on accurate data using EEM images can help develop data driven sensing without the need of GCMS measurements

GENERATING DATA FROM EXISTING DATA

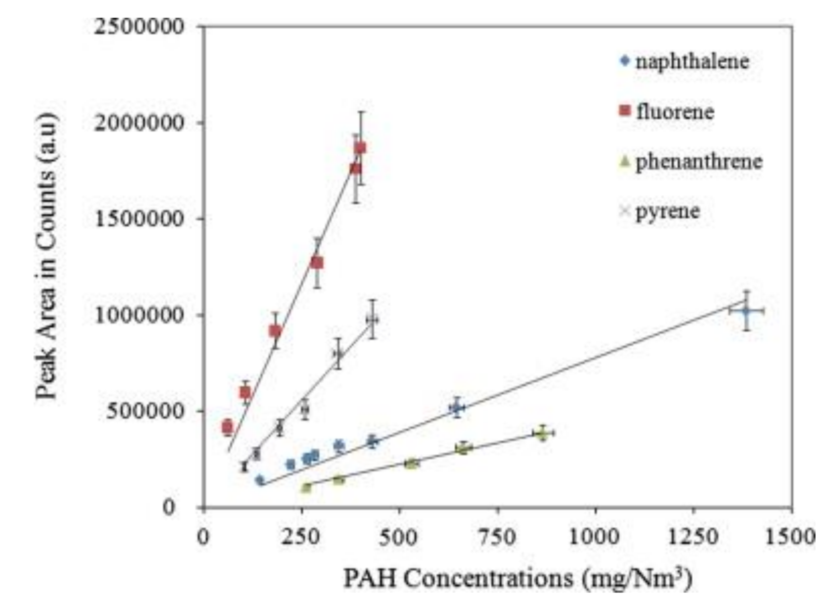


***Figure 4.** Fluorescence Intensity of EEM images is directly proportional to concentration of PAHs

$$F_p = \sum_{i=1}^{16} k_i C_i + b_p$$

F_p is the fluorescence intensity at pixel p of EEM, C_i is concentration of *i'th* PAH, k_i is the weight for *i'th* PAH and b_p is pixel constant. Note: constant vary between EEMs

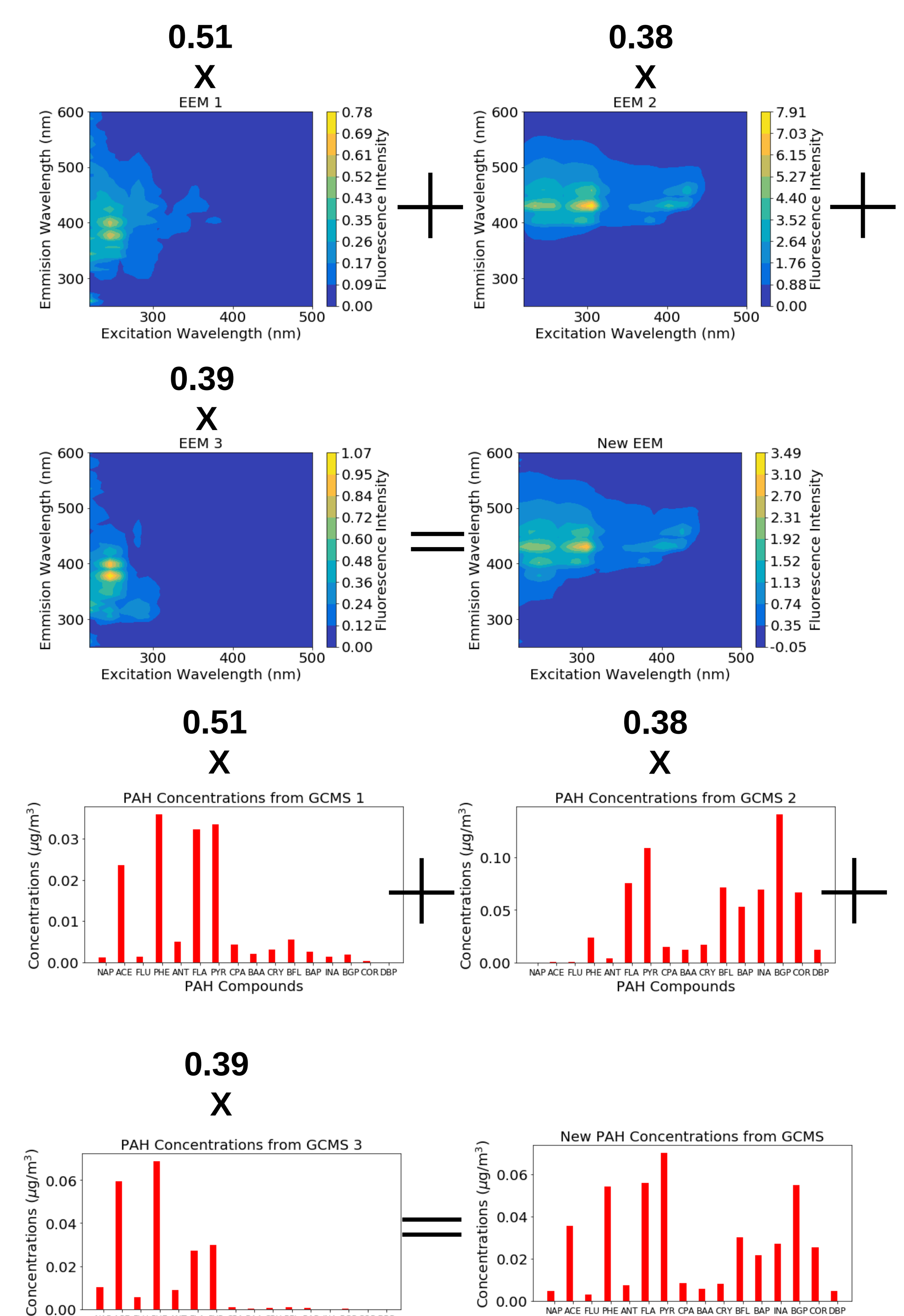
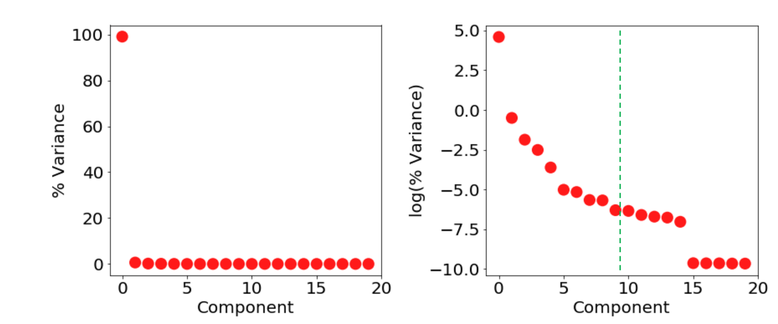


Figure 5. We create 600 training images and labels using 15 original images and labels. We create 150 test images and labels using 5 original images and labels

*Sun, Renhui, et al. "Analysis of gas-phase polycyclic aromatic hydrocarbon mixtures by laser-induced fluorescence." *Optics and Lasers in Engineering* 48.12 (2010): 1231-1237.

PRINCIPAL COMPONENT ANALYSIS +MULTIPLE LINEAR REGRESSION



→ MLR linear_model sklearn →

$$\% \; error_j = 100 \; X \; \frac{\sum_{i=1}^{n} |ypred_i^j - y_i^j|}{\sum_{i=1}^{n} y_i^j}$$

For *j'th* PAH (j=1,..,16), % error is calculated over all n samples

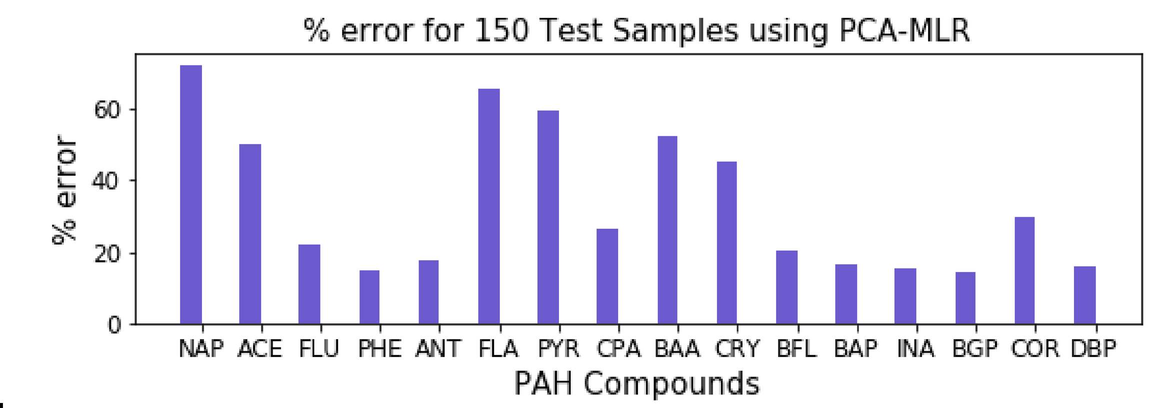


Figure 6. We choose d=10 from variance plots. The average %error over all PAHs is 34%

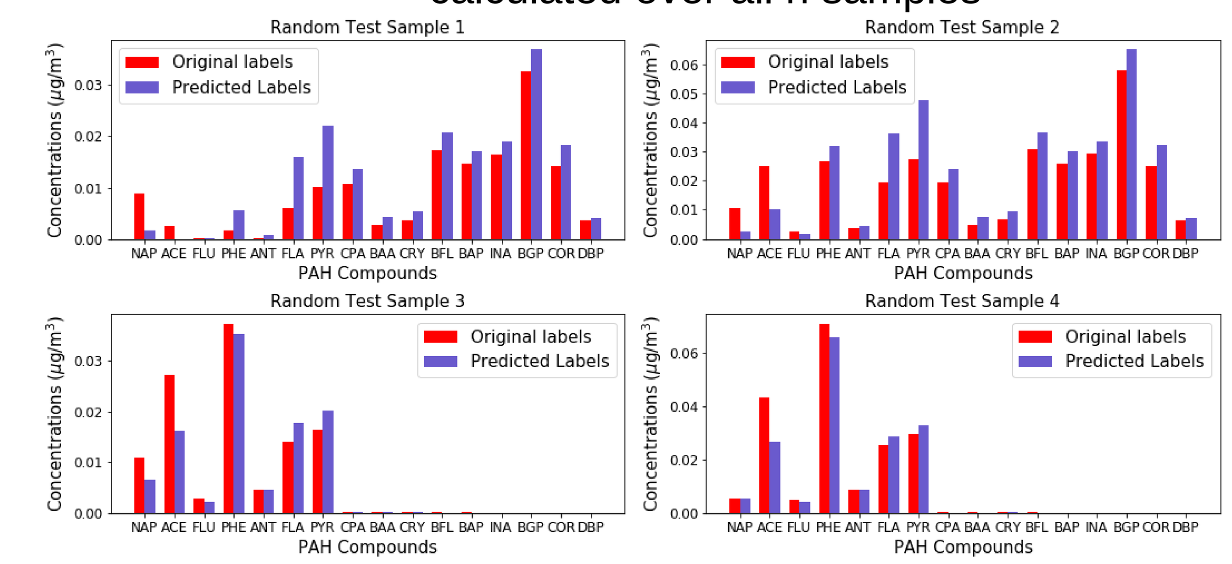


Figure 7. Comparison of predicted vs original PAHS concentrations for 4 random test samples

KERNEL REGRESSION

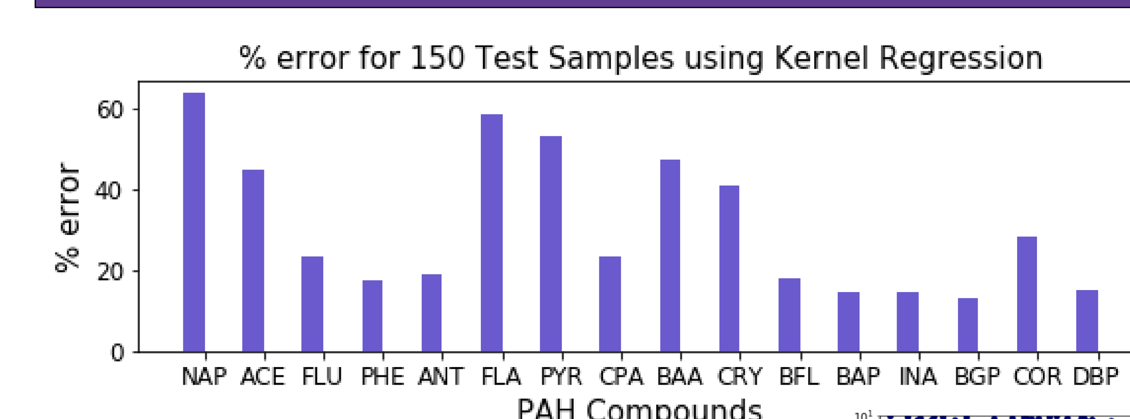


Figure 8. The average %error over all PAHs is 31% (top)

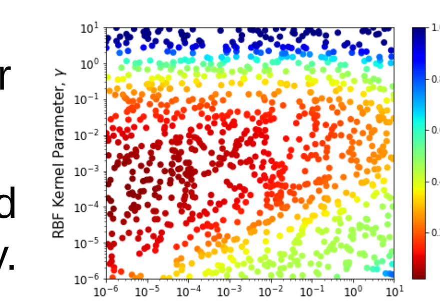


Figure 9. Hyperparameter plot. Red indicates better validation accuracy. (right)

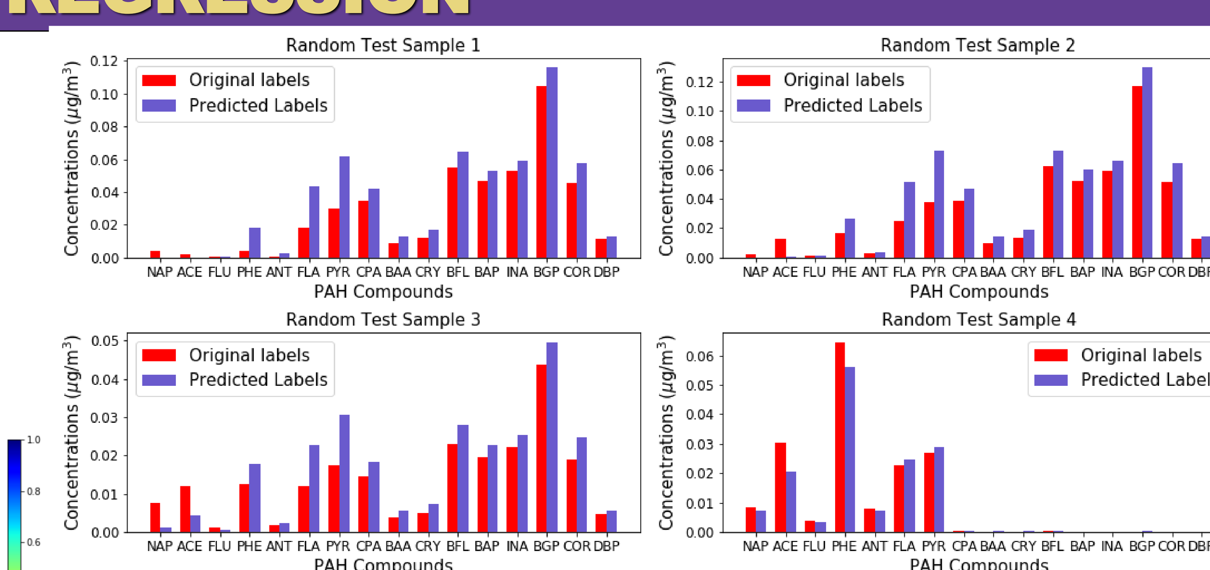


Figure 10. Comparison of predicted vs original PAHS concentrations for 4 random test samples

NEURAL NETWORKS

INPUT $\xrightarrow{\text{n x 3300}}$ HIDDEN LAYER $\xrightarrow{\text{n x M}}$ OUTPUT LAYER $\longrightarrow$ n x 16

Hyperparameters tuned: M, learning rate and momentum

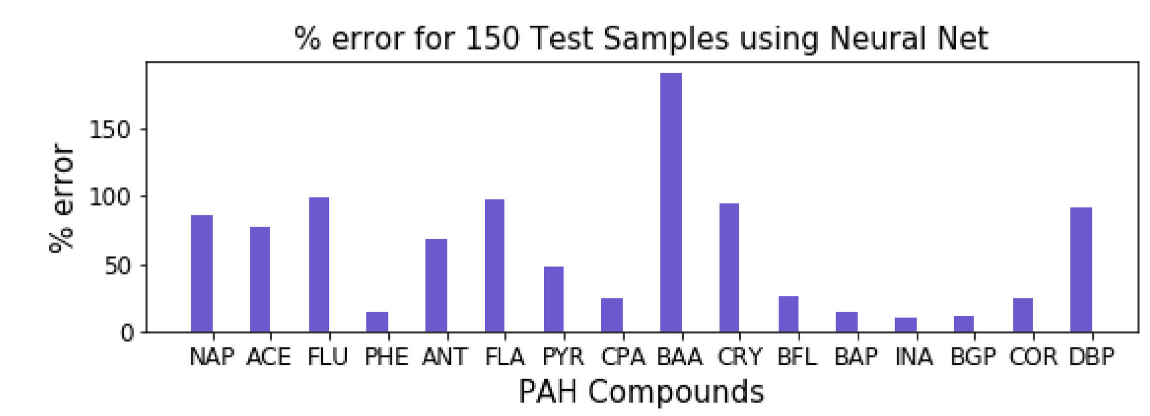


Figure 11. The average %error over all PAHs is 61%

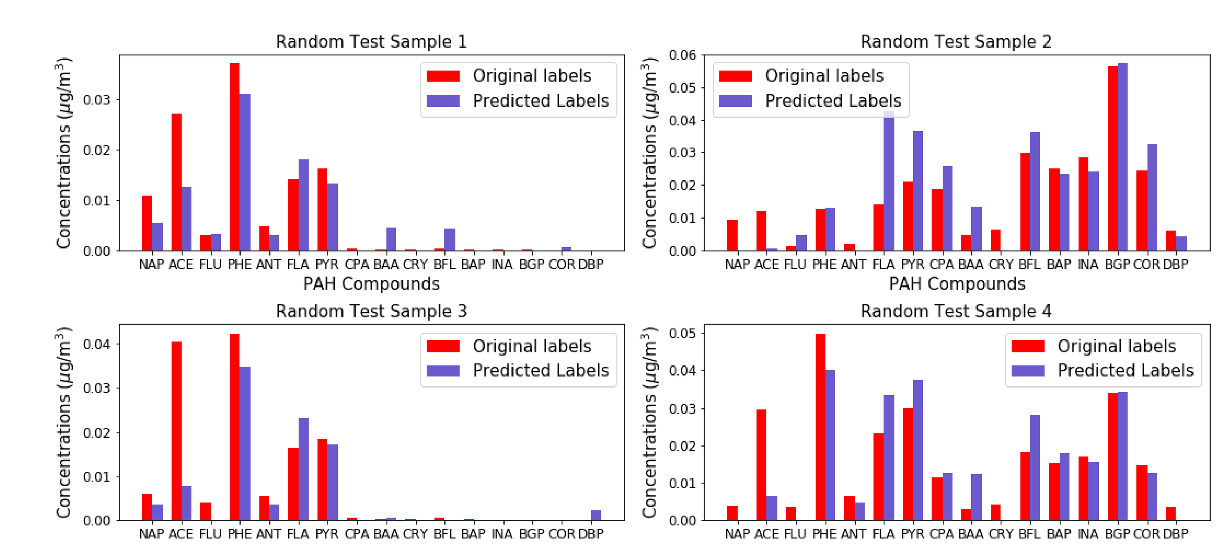


Figure 12. Comparison of predicted vs original PAHS concentrations for 4 random test samples

CONCLUSIONS

- 3 models trained and tested on EEM image data to predict PAH concentrations obtained from GCMS
- Average %error on test set
  - PCA+MLR: 34%
  - KR: 31%
  - Single Layer Neural Networks: 61%
- EEM is indicative of PAH concentration in combustion generated aerosols and can be used to develop low cost, compact toxicity sensors

FUTURE WORK

- Implement EEM + Machine Learning on real mixed soot samples to predict PAH concentrations
- Implement EEM + Machine Learning on data from different sources

ACKNOWLEDGMENTS

This project was a part of CSE546 Machine Learning Course. We thank Prof. Kevin Jamieson for his feedback.

Contact: Gaurav Mahamuni gauravsm@uw.edu Anamol Pundle apundle@uw.edu