Journal of Student-Scientists' Research

Improving PM2.5 Sensor Calibration Using Transformer Model Architecture

Ryan Guo
NSF Spatiotemporal Innovation Center, Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA

Anusha Srirenganathanmalarvizhi
NSF Spatiotemporal Innovation Center, Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA

Chaowei Yang
NSF Spatiotemporal Innovation Center, Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA

DOI: https://doi.org/10.13021/jssr2025.5358

PUBLISHED

2025-09-25

ISSUE

Vol. 7 (2025)

SECTION

College of Science: Department of Geography and Geoinformation Science

LICENSE

ABSTRACT

Air pollution has a significant effect on human health and the environment, with PM2.5 (fine particulate matter with diameter less than 2.5 μm) causing respiratory issues due to its ability to enter an individual's lungs and blood stream. Low-cost PM2.5 sensors, such as those from Clarity and integrated via the OpenAQ platform, offer widespread spatial coverage but often lack the precision of regulatory-grade monitors. Recent studies for PM2.5 calibration have used statistical methods or neural networks, but these approaches are either too simple and can't notice non-linear, complex relationships, or are risk for overfitting. Transformer models have had success in natural language processing, but remain underexplored in the context of PM2.5 and sensor calibration. This project addresses this gap by implementing a Transformer model for calibrating low-cost PM2.5 sensor data. The data contains meteorological data, such as temperature, humidity, wind speed, and PM2.5, as well as temporal features such as date and time. The model uses positional encoding for time-series data, multi-head attention to find trends and patterns, and an Encoder-Decoder structure for aligning sequences. In addition to evaluating model accuracy, the study aimed to determine how much impact the length of historical data (24-hour vs. 48-hour input times) had on model performance. Model performance was assessed using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) metrics. This research fills the gap by providing a robust Transformer

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

model approach for accurately calibrating low-cost PM2.5 sensor data, which can be impactful in environmental monitoring.

A publication of the Aspiring Scientists Summer Internship Program a program of the College of Science, George Mason University.

Published with support from the George Mason University Libraries' Mason Publishing Group.

Platform &
workflow by
OJS / PKP