Medical Research Archives

European Society of Medicine

OPEN ACCESS

Published: January 31, 2024

Citation: Fernando BA, Talebi S, et al., 2024. Environmental Health: Unraveling Particulate Matter Trends with Biometric Signals, Medical Research Archives, [online] 12(1). https://doi.org/10.18103/mra.v12i1.4899

DOI
https://doi.org/10.18103/mra.v12i1.4899

ISSN: 2375-1924

RESEARCH ARTICLE

Data-Driven Environmental Health: Unraveling Particulate Matter Trends with Biometric Signals

Bharana Ashen Fernando, Shawhin Talebi, Lakitha Wijeratne, John Waczak, Vinu Sooriyaarachchi, Shisir Ruwali, Prabuddha Hathurusinghe, David J. Lary*, John Sadler, Tatiana Lary, Matthew Lary, Adam Aker

Department of Physics, University of Texas at Dallas, 800 W Campbell Rd, Richardson TX 75080, USA

***Corresponding author:** David.Lary@utdallas.edu

ABSTRACT

Human physiology is known to react to various environmental stimuli over different time frames. Prolonged exposure to elements such as heat, air pollution, and volatile organic compounds negatively affects health, as established in previous research. Our earlier work demonstrated that autonomic responses of the human body, recorded through biometric sensors on a single individual, could empirically predict levels of inhalable particulate matter in their immediate environment. This current study extends this finding to observations from multiple participants. Subjects cycled on stationary bikes outdoors, equipped with a range of biometric sensors, while environmental sensors simultaneously captured data on their surroundings. Using this expanded data set, machine learning models achieved a high degree of accuracy (R^2=0.97) in predicting concentrations of particulate matter ($PM_{2.5}$) using a few readily available biometric features, including skin temperature, heart rate, and respiration rate. This research underscores the importance of physiological responses as markers of exposure to particulate matter, laying the foundation for the use of biometric data in environmental health surveillance and real-time pollution assessment.

Keywords: particulate matter; autonomic responses; machine learning

1. Introduction

The environment continuously exerts an influence on human health on multiple temporal[1,2,3] and spatial scales. There has been significant literature on long-term exposure to air pollutants such as nitrogen oxides,[4] sulfur dioxide, volatile organic compounds (VOC), and PM[5] that produce and exacerbate adverse health effects such as cardiovascular diseases,[6] damage to nervous systems,[7] and even cancer.[8] Studies of short-term exposures and effects are less prevalent, but still outline effects such as wildfire smoke on cognitive abilities after a 24-hour delay from the time of exposure.[9]

Among the air pollutants mentioned above, PM, in particular, can affect human physiology to varying degrees. The depth of penetration into the human body through inhalation is inversely related to the size of PM.[10] For example, PM_{10} can only travel through the trachea, while PM_1 can penetrate through the alveoli. Historically, PM_{10} and $PM_{2.5}$ have been the most prominent metrics for quantifying habitable air quality, but since smaller PM particles are capable of reaching deeper into respiratory systems, it is important to characterize their effects on health as well.

To characterize the short-term effects of PM on human physiology, we observe the autonomic responses of a human subject. Autonomic responses such as heart rate, galvanic skin response (GSR), pupillometry, and electroencephalography (EEG) are controlled by the autonomic nervous system (ANS) and regulate involuntary bodily functions.[11] Thus, by simultaneously measuring the environmental context alongside the subject's biometric variables, we may relate how PM may affect human physiology in extremely short temporal and spatial scales. Previous research has found that healthy non-smoking women exposed to $PM_{2.5}$ traffic during cycling were associated with a decrease in heart rate variability (HRV) 3 hours after exposure.[12] Among patients with chronic obstructive pulmonary disease (COPD), short-term exposure to $PM_{2.5}$ can result in decreased blood oxygen saturation (SpO_2) in the order of minutes.[13] This manuscript describes an extension of a previous study,[14] and contains biometric data from four subjects collected in four cases in October 2021, January, and February 2022. We observe that with additional data from increased subjects, machine learning (ML) models can better predict smaller PM concentrations from a handful of biometric features (skin temperature, heart rate, and respiratory rate).

2. Materials and Methods

The data collection methodology is inherited from the previous iteration of this study.[14] Although the fundamental methodologies for data collection, described in Sections 2.1 and 2.2, have remained consistent, the main differences and improvements of the study presented are outlined below.

Previously, data acquisition involved a single subject participating in multiple trials of an outdoor bicycle riding along a public bike trail, while an accompanying electric vehicle equipped with an array of sensors captured the surrounding environmental context. This study involved the collection of data from four distinct participants, each on a stationary bicycle, also placed outdoors next to the WSTC building at the University of Texas at Dallas in Richardson, TX, on four separate occasions, spanning October 2021 to February 2022. Again, the environmental context was acquired through sensors placed immediately adjacent to the participants.

Although a significant portion of the biometric and environmental attributes recorded in both experiments remain consistent, substantial enhancements have been made to the temporal resolution of the data. Specifically, we now have access to biometric and environmental data in the order of 1 Hz, whereas previously the data for analysis were sampled every 30 seconds, which allows the ML models to capture rapid changes in both the environment and physiological responses to the environment.

a. Biometric Data

To comprehensively capture the autonomic responses exhibited by the participants, a wide variety of biometric variables were captured from a range of sensors. The objective of this approach was to build ML models without preconceived notions of the physiological responses that are being explored.

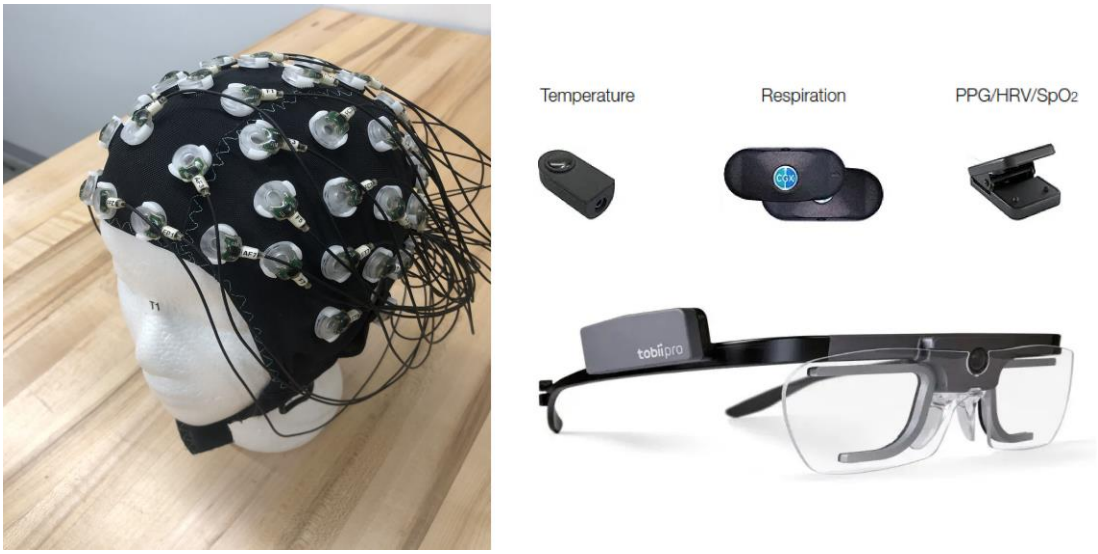

Figure 1: Left: Cognionics (CGX) 64-electrode system used to measure EEG data. **Top right**: CGX AIM2 sensors measure biometrics such as skin temperature, rate of respiration. **Bottom right**: Tobii Pro Glass 2 used to collect pupillometric data.

Electroencephalography (EEG) signals were recorded using a CGX Mobile-64 system [1], shown in Figure 1, which consists of 64 electrodes that are in contact with the scalp of the subject. The electrodes were placed according to the 10-20 system, which describes the distances between the electrodes.[15] Millions of neuronal excitations emanating from the brain produce electric fields that are captured by the electrodes.[16] A conductive gel is applied on the scalp of the subject to improve the detection of electrical activity on the electrode. These signals, captured in units of microvolts as a time series, can provide valuable information on the cognitive state of an individual.

In addition to the EEG headset, we also incorporate the CGX AIM2 sensing system, which collects a rich spectrum of biometrics, such as skin temperature, respiration rate, blood oxygen concentration, electrocardiography (ECG), and galvanic skin response (GSR), a measure of autonomic sweat gland activity.

[1]https://www.cgxsystems.com/

Tobii Pro Glass 2 eye tracking system is used to measure pupillometric data from subjects. From this sensor, we have access to metrics such as pupil diameters, gaze directions, and pupil center positions, all collected at 100 Hz. Furthermore, video streams of the participant's eyes and their point of view are recorded at 50 and 25 frames per second (fps).

b. Particulate Matter

The Fidas Frog [2], a fine dust measurement device, is the sensor used to measure particulate matter near participants. It is an optical aerosol spectrometer that uses the scattering of light from single particles to determine their sizes.

The sensor has a time resolution of 1 s and is capable of measuring the concentrations of PM_1, $PM_{2.5}$, PM_4, PM_{10}, (particle matter of size 1, 2.5, 4 and 10 μm, respectively) in addition to the concentrations of particles within the range of 0.18 – 93 μm. To serve as a direct comparison to the previous study with a single subject, we also only considered the effects of particulate matter concentrations on biometric features.

c. Data Collection

We collected physiological data from four subjects on four separate days in October 2021, January, and February 2022. Before stationary cycling began, there was a calibration period of two minutes each in which subjects had their eyes open and closed, respectively, to establish baseline biometrics. Each participant was equipped with a set of sensors, shown in Figure 3, in such a way that they could unobtrusively cycle on the stationary bike for a period of time at their discretion. Each data collection instance lasted approximately 10 minutes.

The environmental sensors were located in the back of an electrical vehicle, which collects ambient air from an input for analysis, parked approximately two meters from the participant. The Fidas Frog PM measurement device was placed adjacent to the subject of the participant.

[2]https://www.palas.de/en/product/fidasfrog

d. Data Processing

To explore the relationship between biometric and environmental data in multiple people, we frame the problem as a supervised ML problem. Biometric data will serve as features of the ML model, and the environmental context, measured by particle concentrations for PM sizes, is the target.

As part of data processing, raw EEG data, measured in μV, are transformed from the time domain to the frequency domain using the Welch method,[17] which calculates power spectral densities at different frequencies. These can be further subdivided into frequency bands, namely: Delta (0.5-4Hz), Theta (4-8 Hz), Alpha (8-13 Hz), Beta (13-30 Hz), and Gamma (30-100 Hz).[16] Each frequency band is associated with specific states of brain activity, such as deep sleep, calm wakefulness, and attentiveness. The spectral density of each frequency band can be calculated for each of the 64 electrodes, resulting in 320 feature columns.

To compare the data collected from various sensors at disparate frequencies, they need to be appropriately synchronized. The three main sources of data -biometric, pupillometric, and environmental are collected at 500 Hz,

Figure 2: Left: Fidas Frog PM measurement unit. **Right**: AIRMAR Weatherstation 220WX used to collect environmental quantities such as pressure, air temperature.

Figure 3: Participants are adorned with biometric sensors and ride a stationary bike next to environmental sensors.

100 Hz, and approximately 1 Hz, respectively. Thus, all sources are merged and resampled at a rate of 1 Hz, the smallest measurement frequency from the sensor suite. After the data processing was completed, including the removal of the data from the T7 electrode due to poor signal quality, there were a total of 1,383 unique timestamp records in 324 columns of biometric features, measured from four participants.

e. Feature Selection and Machine Learning Model

Initially, a decision tree-based algorithm known as XGBoost[18] was chosen for the development of the model. XGBoost is used for supervised ML tasks, such as regression, and since we are attempting to predict ambient PM from biometric signatures from individuals, XGBoost is well suited for the problem. Builds decision trees sequentially, iteratively minimizing the loss function, which in this case is the mean squared error (MSE).

Similarly, to a standard ML workflow, training was performed on 80% of the data, while the model was tested on the remaining 20%. For explainability, we employ SHapley Additive exPlanations (SHAP) values,[19] which attempt to quantify the importance of a particular feature by assigning a score to the model when the feature is removed from the predictions. Furthermore, it attempts to score how the model reacts when a particular feature value increases or decreases, which can provide an intuition on how the features contribute to the model. It should be noted that these contributions should not be interpreted as causal, but merely as how the features are correlated to the model.

Given the large number of feature columns, we also attempted to build a more robust learning environment by selecting only the most optimal, information-rich features. If the selected features do not contain useful and generalizable information, the resulting model is an overfitted model that does not perform well when faced with data the model was not trained on or one that learns the noise of the data instead of the meaningful characteristics of features. For that reason, Pearson's correlation coefficients and mutual information[20] were considered potential methods to extract information-rich characteristics. Pearson's correlation coefficients are known to capture linear relationships between features, while mutual information can quantify non-linear relationships between variables. For two continuous variables X and Y, mutual information I is given as:

$$I(X;Y) = \int\int p(x,y) \cdot \log\left(\frac{p(x,y)}{p(x) \cdot p(y)}\right) dx\, dy \qquad (1)$$

where $p(x)$ and $p(y)$ are marginal probability density functions of X and Y, and $p(x, y)$ is the joint probability density function of X and Y. In other words, mutual information measures the dependency between features using an estimate of the entropy from the distances of the k nearest neighbors and produces a nonnegative score; the higher the score, the higher the mutual information between the two variables and vice versa.

Following the premise that features with high mutual information with the target variable are good predictors of the said target, the ML model can be made more robust by selecting a smaller number of features based on their mutual information score to create a more general ML model.

Furthermore, in an attempt to investigate easily obtainable biometric variables, we developed ML models to predict PM concentrations exclusively using skin temperature, respiration rate, blood oxygen concentration (SpO_2), changes in blood volume (PPG) and heart rate. For these models, all biometric features that are less readily collected from the more resource-intensive sensors such as EEG |.

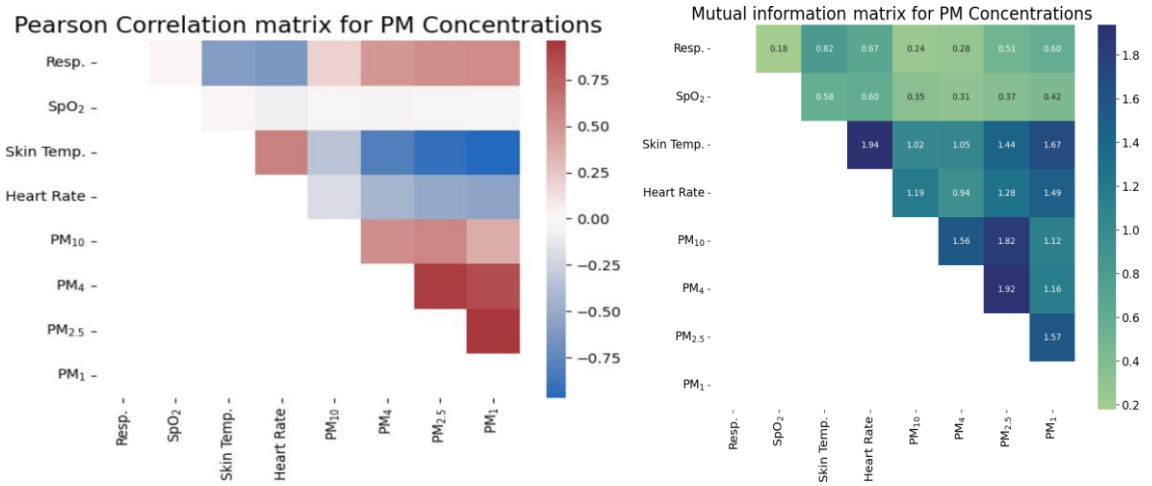

Figure 4: The correlation (**left**) and mutual information (**right**) matrices calculated between a subset of the features used in our empirical regression models and the target dataset of PM concentrations.

f. A Causal Perspective for Feature Selection

A fundamental principle for the interpretation of statistical data is 'correlation does not imply causation'.[21] When choosing what features to use in an empirical machine learning regression model, it is, therefore, helpful to use a causal perspective to try and draw a directed graph. The graph is based on observations alone and describes the likely relationship between the ambient environment and the physiological response of the human body. We then used only the features that have a causal relationship with the variable we are seeking to estimate. The Occam Razor is a principle that states 'Entities should not be multiplied without necessity' or 'The simplest solution is most likely the right one'.[22] Applying Occam's Razor in regression modeling, with an emphasis on including only causally related variables, leads to models that are likely to be more accurate, reliable, interpretable, and generalizable. Simpler models with fewer parameters and less complexity are less likely to overfit the training data. This means that they are more likely to perform well on unseen data, which is the ultimate goal of a machine learning model. For clarity, it is useful to note that the causal graph does not utilize the machine learning model, it is employed at the feature selection stage before we train the models. First, we calculate a 'causal effect' between the target (in this example,

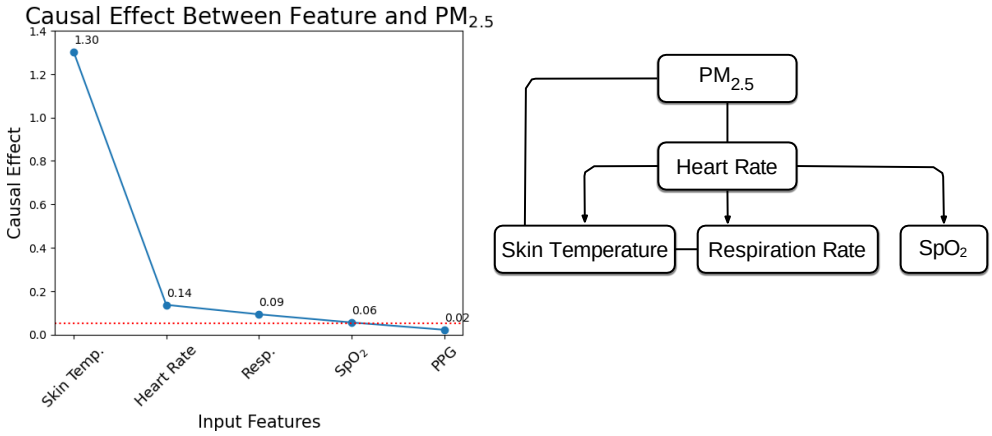

Figure 5: Left: Causal effect between input feature and $PM_{2.5}$ for easily accessible biometric features. **Right:** A directed graph created using causal analysis of the collected biometric and environmental data to assess the interaction between the ambient PM abundance and the autonomic responses in humans.

$PM_{2.5}$) and the biometric features, as proposed by.[23] This method suggests that the degree to which a feature can reduce the uncertainty of a target variable, calculated by Shannon entropy, can quantify the causal effect of a feature on a target. While the algorithm failed to execute on the full biometric dataset due to memory constraints, we were able to successfully calculate causal effects on the easily collectible biometric features (skin temperature, heart rate, and so on). Then we use the approach described by[24,25] to produce the directed graph shown in Figure 5. In the Discussion section, we explore the physiological relationships between $PM_{2.5}$ and heart rate, skin temperature, respiration rate, and SPO_2.

3. Results

The first ML model to predict PM concentrations from biometric data from all participants included all biometric characteristics, a total of 324 columns, ranging from EEG data to pupillometrics and metrics such as heart rate, skin temperature, and galvanic skin response. The results of this model can be seen on the top left of Figure 6.

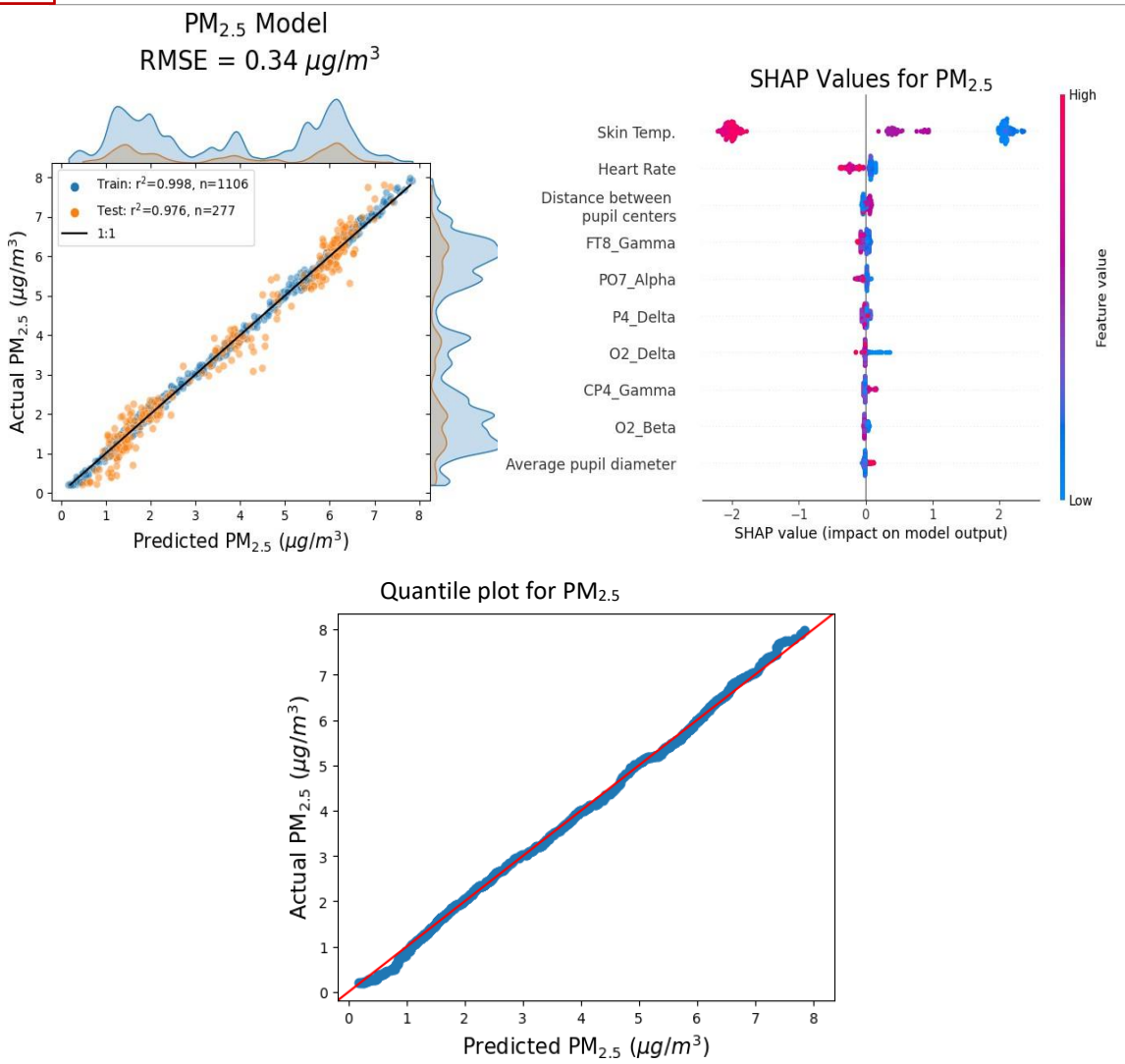

Figure 6: Top left: Scatter plot with the goodness of fit of the ML model. **Top right**: SHAP values of the model, which provides both a ranking of feature importance as well insight into how each variable affects the model. **Bottom**: quantile distribution of actual and prediction $PM_{2.5}$ values.

For $PM_{2.5}$, the correlation coefficient reports an $r^2 = 0.976$ as the goodness of fit on the test dataset, an improvement over the previous study, and suggests a relationship between biometric context and PM concentrations in an extremely short time scale (on the order of seconds) between multiple participants. The axes in the scatter plot show the distribution of training and testing data, which shows the subgroups of data from different participants. The plot on the top right in Figure 6 shows the results for 10 features with the highest SHAP values from the model. We note that skin temperature is ranked as the most important feature, followed by heart rate, distance between the center of the pupils, and a host of frequency band information from EEG electrodes. The SHAP values suggest that, among participants, given a higher value of skin temperature, the model is more likely to predict a lower concentration of $PM_{2.5}$ and vice versa.

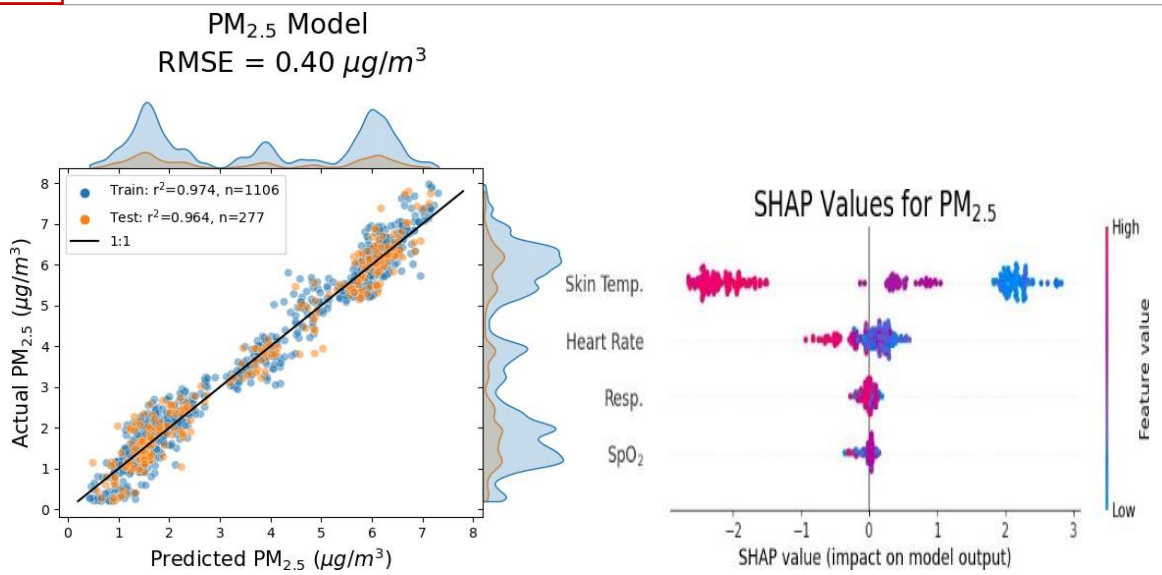

Figure 7: With a handful of (relatively) easily-collectible biometric features, the ML model is able to predict ambient $PM_{2.5}$. The RMSE error had increased from 0.34 to 0.40 $\mu g/m^3$

To explore viable simplifications of the ML model, we calculated the mutual information between the target and the biometric data. In Figure 4, we show the mutual information matrix of a handful of variables (from the initial 324) that contain the most information and are accessible in terms of data collection. Rebuilding a machine learning model to predict ambient $PM_{2.5}$ concentration with this high mutual information biometric features yields results comparable to the previous model, which included all biometric features collected and derived. The respective scatter plot and SHAP values are shown in Figure 7.

4. Discussion

Empirical ML models that predict PM concentration on multiple participants extend the results that predicted PM concentration from the biometric signature of a single participant.[14] The first model, Figure 6, included all biometric characteristics collected and derived, represented by 324 columns, ranging from EEG frequency spectral densities to pupillometrics to skin temperature and heart rate. This is an improvement over the previous results from a single subject, which had a goodness of fit of $r^2 = 0.91$ between the predicted and actual PM concentrations for the test data set. It should also be noted that the models trained and tested on data from a single subject are unable to predict on the test data set with a comparable degree of precision, as shown in Figure 8. Multiple trials may improve the predictive power of the model, similar to the previous single-subject experiment.[14]

Through the usage of SHAP values, it is also revealed that skin temperature and heart rate hold the most importance for this empirical ML model. Research has shown that there is a relationship between skin temperature and heart rate in children,[26] but the short-term effects of PM concentration on physiological responses such as skin temperature and heart rate are still being explored. It is interesting that the sweat response, measured through GSR, is not an important metric even though the skin temperature is highly correlated with the ambient PM concentration.

The performance of the empirical models created from a handful of easily collectible biometric variables also sets an important paradigm. It suggests that to model ambient PM concentration, one may not necessarily need to invest in a resource-intensive sensor suite that captures EEG or pupillometric data. Although the data from these sensors have helped, biometrics such as skin temperature, heart rate, and blood oxygen concentration can model the ambient PM concentration equally well.

The benefits of linking skin temperature, heart rate, and particulate matter concentration are numerous. Measurement of core body temperature is a fundamental body temperature measurement but is challenging to make, as it usually involves a rectal probe, so we have instead measured the skin temperature. Skin temperature is influenced by the core temperature since the body regulates heat loss or retention through blood flow to the skin,[27,28] but it is also subject to external environmental factors, making it an imperfect proxy of the core

temperature. Heart rate and core temperature are intrinsically linked through thermoregulation and cardiovascular physiology. As the core temperature increases, such as during fever or exercise, the body initiates thermoregulation processes such as vasodilation and sweating, which can lead to an increase in heart rate to facilitate heat loss. In contrast, a drop in core temperature can trigger vasoconstriction and shivering, which also affect heart rate. This bidirectional influence suggests a complex relationship that

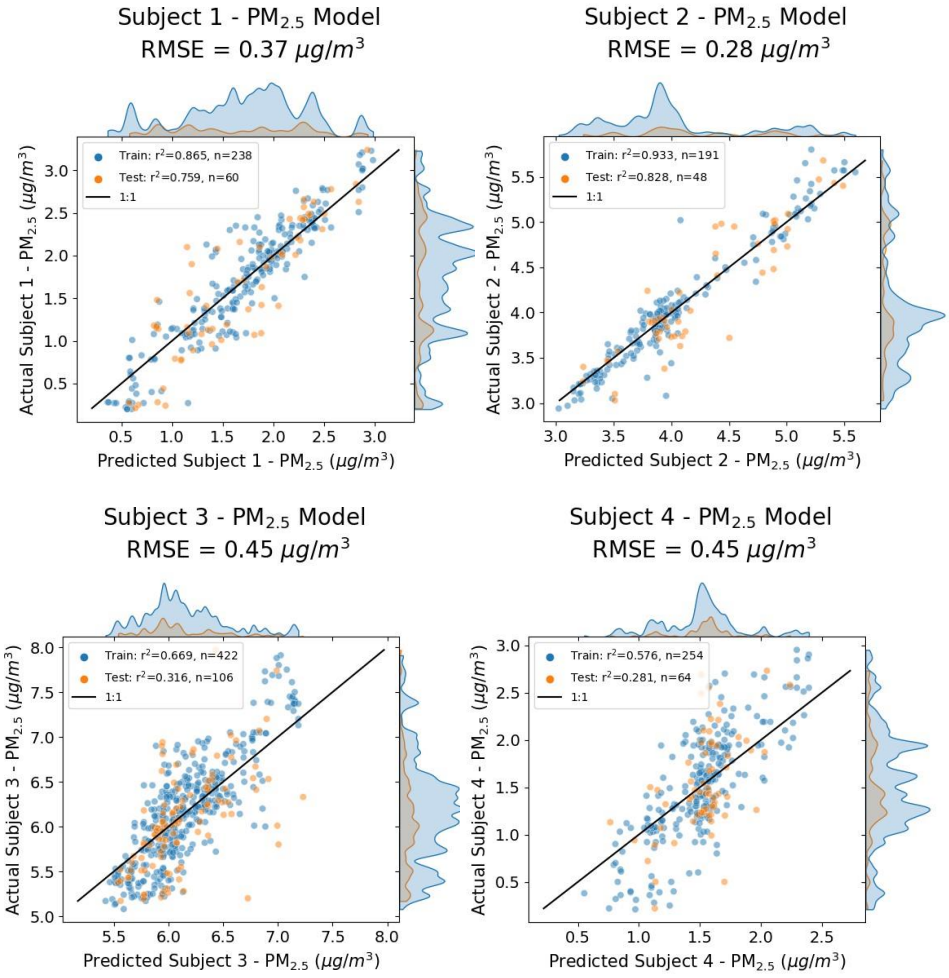

Figure 8: Each scatter plot shows predictive power of a model trained only on a single subject's biometric data and their ambient $PM_{2.5}$ concentration. None of the models show performance comparable to when data from all is part of a feedback system that maintains homeostasis.

When the heart rate increases, usually during physical exertion or stress, the blood flow to the skin also increases, potentially increasing the temperature of the skin.[29,30] Conversely, in a relaxed state, a lower heart rate is associated with a decrease in blood flow to the skin, which can decrease the temperature of the skin. Both heart rate and skin temperature are influenced by a variety of other factors, including ambient environmental conditions such as temperature, physical activity, emotional state, and general health.

The physiological relationship between heart rate and respiration rate is closely coordinated through the autonomic nervous system, where both rates increase to meet the increasing demand for oxygen and nutrients in the body and to remove carbon dioxide and waste products.[31,32] This coordination is known as cardiorespiratory coupling. Each can be independently influenced by various factors such as physical activity, emotional state, and overall health.

Heart rate and blood oxygen saturation (SpO_2) are related in the context of cardiovascular and pulmonary function; Heart rate determines how

often blood is cycled through the lungs, where oxygenation occurs, and SpO_2 indicates the percentage of oxygen-saturated hemoglobin in blood.[33,34] Elevated heart rate can increase oxygen delivery to tissues by circulating blood more rapidly, but does not necessarily increase SpO_2, which is generally maintained at a high level by homeostatic mechanisms until pulmonary or cardiovascular function is compromised. The relationship is a complex interaction, especially under stress or pathological conditions.

Of all the biometric variables we measured, skin temperature has the strongest response to the change in PM abundance (Figures 6 and 7). Ambient PM can potentially influence skin temperature through a variety of mechanisms, each of which is likely to operate on different timescales. The shortest timescales of seconds or less are those that we are likely to be observing in our observations. Response times are likely to vary between individuals depending on the individual's sensitivity, the concentration of PM, and the duration of exposure.

1. **Vascular Responses**: PM can cause immediate (within seconds to minutes) vasoconstriction due to the release of endothelial signaling molecules such as endothelin, resulting in decreased blood flow to the skin and potentially reduced skin temperature.[35]
2. **Autonomic Nervous System Disruption**: The effects of PM on the autonomic nervous system can occur as PM-induced irritation of the respiratory tract can alter the autonomic balance, affecting sweat production and dilation of blood vessels.[36]
3. **Inflammation and Oxidative Stress**: Inflammatory and oxidative stress responses to PM can develop in hours to days, as the body mobilizes immune cells and generates reactive oxygen species in response to particulates, which can affect the regulation of skin temperature.[37]
4. **Endocrine Changes**: Endocrine responses to PM, such as disruptions in thyroid function, can take days to weeks, as alterations in hormone levels can affect metabolic processes and consequently, skin temperature over time.[38]

The relationship between ambient particulate concentrations and heart rate could involve an intricate interplay of multiple physiological mechanisms, each with its own timescale.

1. **Autonomic Imbalance**: Exposure to PM can alter the balance of the autonomic nervous system (in seconds to minutes) by enhancing sympathetic activity.[39] This immediate response can occur due to the activation of neural receptor reflexes in the respiratory tract after inhalation of particulates.
2. **Oxidative Stress**: Particulates can induce oxidative stress that leads to changes in vascular tone and heart rate due to the effects of reactive oxygen species on the cardiovascular system.[37]
3. **Inflammatory Response**: Over a period of hours to days, PM can induce systemic inflammation, potentially increasing heart rate through increased sympathetic activity.[35] The release of cytokines and other inflammatory mediators into the bloodstream can increase heart rate by influencing the electrical activity and contractility of the heart.[40]
4. **Endothelial Dysfunction**: This can occur over a longer period, from days to weeks, as chronic exposure to particulates leads to endothelial dysfunction, affecting heart rate by altering blood vessel reactivity.[41,42]
5. **Direct Effect on Heart Cells**: Fine particles can enter the blood-stream and exert direct effects on cardiac cells.[40]
6. **Cardiovascular Reflexes**: Irritation caused by particulates in the lungs can provoke reflexive cardiovascular responses, including increasedheart rate.[42]
7. **Hypoxemia**: Particulate matter has been found to affect gas exchange and subsequently have effects on heart rate.[43]

5. Conclusion

This study builds on the earlier study of[14] by using a much larger data set amassed from a diverse cohort of participants, substantially increasing the number of data points available for the development of machine learning models, and expands the analysis to include a causal perspective and derive a directed graph based on the collected data that describes the relationship between variables. This causal perspective highlights the importance of physiological responses as indicators of exposure to particulate matter, providing a basis for the use of biometric data in environmental health surveillance and real-time pollution assessment. The results are encouraging, as we shown that, between multiple participants, the concentration of ambient $PM_{2.5}$ can be predicted with a high degree of precision, initially from many biometric features, and then using only a handful of easily collectible features such as skin temperature, heart rate, blood oxygen and respiration rate. We believe that the study can and should be further strengthened by increasing the sample size to a more numerous and representative

population. Data collected from demographics of various statures that would facilitate the analysis of autonomic physiological responses that are disproportionately affected by air pollution.

Supplementary Materials

Both the data and code have been made publicly available at: https://github.com/mi3nts/DUEDARE_multiple_participants and https://zenodo. org/records/10152548 (accessed on November 17th, 2023).

Acknowledgments

Support from the University of Texas at Dallas Office of Sponsored Programs, Dean of Natural Sciences and Mathematics, and Chair of the Physics Department is gratefully acknowledged.

Ethics Statement

The Institutional Review Board at The University of Dallas approved all experimental protocols, and study participants provided informed consent.

Author Contributions

Methodology, D.J.L., B.A.F., S.T. and T.L.; software, B.A.F., V.S.; formal analysis, D.J.L., B.A.F., V.S., and T.L.; data curation, B.A.F., S.T., D.J.L., L.O.H.W., T.L., M.L., J.S., A.S., A.A., Y.Z. and J.W.; writing—original draft preparation, B.A.F., D.J.L.; writing—review and editing, B.A.F., D.J.L., J.W., and S.R.; visualization, B.A.F.,V.S.; supervision, D.J.L.

Funding

This research was funded by the following grants: The US Army (Dense Urban Environment Dosimetry for Actionable Information and Recording Exposure, U.S. Army Medical Research Acquisition Activity, BAA CDMRP Grant Log #BA170483). EPA 16th Annual P3 Awards Grant Number 83996501, entitled Machine Learning Calibrated Low-Cost Sensing. The Texas National Security Network Excellence Fund award for Environmental Sensing Security Sentinels. SOFWERX award for Machine Learning for Robotic Teams. The analysis was partially supported via an MOU with the Health Outcomes of Military Exposures (HOME) office of the War Related Illness & Injury Study Center, at the Washington DC US Department of Veterans Affairs Medical Center

Competing Interests

The author has declared that no competing interests exist.

References

[1] K. A. Miller, D. S. Siscovick, L. Sheppard, K. Shepherd, J. H. Sullivan, G. L. Anderson, and J. D. Kaufman, "Long-term exposure to air pollution and incidence of cardiovascular events in women.," *The New England journal of medicine*, vol. 356 5, pp. 447–58, 2007.

[2] J. O. Anderson, J. G. Thundiyil, and A. l. Stolbach, "Clearing the air: A review of the effects of particulate matter air pollution on human health," *Journal of Medical Toxicology*, vol. 8, pp. 166–175, 2012.

[3] J. Hu, X. Xue, M. Xiao, W. Wang, Y. Gao, H. Kan, J. Ge, Z. Cui, and R. Chen, "The acute effects of particulate matter air pollution on ambulatory blood pressure: a multicenter analysis at the hourly level," *Environment International*, vol. 157, p. 106859, 2021.

[4] H. Tian, R. Xu, J. G. Canadell, R. L. Thompson, W. Winiwarter, P. Suntharalingam, E. A. Davidson, P. Ciais, R. B. Jackson, G. Janssens-Maenhout, *et al.*, "A comprehensive quantification of global nitrous oxide sources and sinks," *Nature*, vol. 586, no. 7828, pp. 248–256, 2020.

[5] R. T. Burnett, H. Chen, M. Szyszkowicz, N. L. Fann, B. Hubbell, C. A. Pope, J. S. Apte, M. Brauer, A. J. Cohen, S. Weichenthal, J. S. Coggins, Q. Di, B. Brunekreef, J. J. Frostad, S. S. Lim, H. dong Kan, K. D. Walker, G. D. Thurston, R. B. Hayes, C. C. Lim, M. C. Turner, M. Jerrett, D. Krewski, S. M. Gapstur, W. R. Diver, B. Ostro, D. E. Goldberg, D. L. Crouse, R. V. Martin, P. Peters, L. L. Pinault, M. Tjepkema, A. van Donkelaar, P. J. Villeneuve, A. B. Miller, P. Yin, M. Zhou, L. Wang, N. A. H. Janssen, M. Marra, R. W. Atkinson, H. Tsang, T. Q. Thach, J. B. Cannon, R. T. Allen, J. E. Hart, F. Laden, G. Cesaroni, F. Forastiere, G. Weinmayr, A. Jaensch, G. Nagel, H. Concin, and J. V. Spadaro, "Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, pp. 9592– 9597, 2018.

[6] C. A. Pope, R. T. Burnett, G. D. Thurston, M. J. Thun, E. E. Calle, D. Krewski, and J. Godleski, "Cardiovascular mortality and long-term exposure to particulate air pollution: Epidemiological evidence of general pathophysiological pathways of disease," *Circulation: Journal of the American Heart Association*, vol. 109, pp. 71–77, 2003.

[7] Şermin Genç, Z. F. Zadeoğluları, S. H. Fuss, and K. Genc, "The adverse effects of air pollution on the nervous system," *Journal of Toxicology*, vol. 2012, 2012.

[8] A. C. Pope, R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, and G. D. Thurston, "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution.," *JAMA*, vol. 287 9, pp. 1132–41, 2002.

[9] S. E. Cleland, L. H. Wyatt, L. Wei, N. Paul, M. L. Serre, J. J. West, S. B. Henderson, and A. G. Rappold, "Short-term exposure to wildfire smoke and $pm_{2.5}$ and cognitive performance in a brain-training game: A longitudinal study of u.s. adults," *Environmental Health Perspectives*,nvol. 130, no. 6, p. 067005, 2022.

[10] Y. Bai, Y. Zhang, J. Zhang, Q. Mu, W. Zhang, M. Butlin, Y. Guo, and X. Wang, "Particulate matter and human health: Toxicological assessment and importance of size and composition of particles for oxidative damage and carcinogenic mechanisms," *Journal of Environmental Science and Health, Part C*, vol. 33, pp. 489–520, 2015.

[11] N. L. Strominger, R. J. Demarest, and L. B. Laemle, *Noback's human nervous system, seventh edition structure and function*. Humana Press, 2012.

[12] S. Weichenthal, M. Hatzopoulou, and M. S. Goldberg, "Exposure to traffic-related air pollution during physical activity and acute changes in blood pressure, autonomic and micro-vascular function in women: a cross-over study," *Particle and fibre toxicology*, vol. 11, no. 1, pp. 1–16, 2014.

[13] X. Xia, H. Qiu, T. Kwok, F. W. Ko, C. L. Man, and K.-F. Ho, "Time course of blood oxygen saturation responding to short-term fine particulate matter among elderly healthy subjects and patients with chronic obstructive pulmonary disease," *Science of The Total Environment*, vol. 723, p. 138022, 2020.

[14] S. Talebi, D. J. Lary, L. O. Wijeratne, B. Fernando, T. Lary, M. Lary, J. Sadler, A.

Sridhar, J. Waczak, A. Aker, and et al., "Decoding physical and cognitive impacts of particulate matter concentrations at ultra-fine scales," *Sensors*, vol. 22, no. 11, p. 4240, 2022.

[15] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," *NeuroImage*, vol. 34, p. 1600–1611, Feb 2007.

[16] P. L. Nunez and R. Srinivasan, *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006.

[17] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[18] T. Chen and C. Guestrin, "Xgboost," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[19] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 4765–4774, Curran Associates, Inc., 2017.

[20] B. C. Ross, "Mutual information between discrete and continuous data sets," *PLoS ONE*, vol. 9, no. 2, 2014.

[21] R. A. Fisher, *The Design of Experiments*. Edinburgh, Scotland: Oliver and Boyd, 1935. This book discusses the principles of experimental design and the distinction between correlation and causation.

[22] D. J. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge, UK: Cambridge University Press, 2003. The book discusses the application of Occam's Razor in the context of information theory and statistical modeling.

[23] Y.-N. Sun, W. Qin, J.-H. Hu, H.-W. Xu, and P. Z. Sun, "A causal model-inspired automatic feature-selection method for developing datadriven soft sensors in complex industrial processes," *Engineering*, vol. 22, pp. 82–93, 2023.

[24] J. D. Ramsey, K. Zhang, M. Glymour, R. S. Romero, B. Huang, I. EbertUphoff, S. Samarasinghe, E. A. Barnes, and C. Glymour, "Tetrad—a toolbox for causal discovery," in *8th international workshop on climate informatics*, 2018.

[25] Y. Zheng, B. Huang, W. Chen, J. Ramsey, M. Gong, R. Cai, S. Shimizu, P. Spirtes, and K. Zhang, "Causal-learn: Causal discovery in python, *arXiv preprint arXiv:2307.16405*, 2023.

[26] P. Davies and I. Maconochie, "The relationship between body temperature, heart rate and respiratory rate in children," *Emergency Medicine Journal*, vol. 26, no. 9, pp. 641–643, 2009.

[27] W. D. McArdle, F. I. Katch, and V. L. Katch, *Exercise Physiology: Nutrition, Energy, and Human Performance*. Wolters Kluwer Health, 8 ed., 2015.

[28] L. Sherwood, *Human Physiology: From Cells to Systems*. Cengage Learning, 9 ed., 2015.

[29] J. T. Cacioppo and L. G. Tassinary, *Principles of psychophysiology: Physical, social, and inferential elements*. Cambridge University Press, 1990.

[30] W. T. Roth, "Cardiovascular behavioral medicine," *Handbook of Psychophysiology*, vol. 2, p. 4, 1984.

[31] A. C. Guyton and J. E. Hall, *Textbook of Medical Physiology*. Elsevier Saunders, 11 ed., 2006.

[32] S. K. Powers and E. T. Howley, *Exercise Physiology: Theory and Application to Fitness and Performance*. McGraw-Hill, 8 ed., 2012.

[33] G. J. Tortora and B. Derrickson, *Principles of Anatomy and Physiology*. John Wiley Sons, 12 ed., 2009.

[34] D. E. Mohrman and L. J. Heller, *Cardiovascular Physiology*. McGrawHill Education, 8 ed., 2018.

[35] R. D. Brook, B. Franklin, W. Cascio, Y. Hong, G. Howard, M. Lipsett,
R. Luepker, M. Mittleman, J. Samet, S. C. Smith, *et al.*, "Cardiovascular effects of air pollution," *Circulation*, vol. 109, no. 21, pp. 2655–2671, 2004.

[36] C. A. Pope III, D. W. Dockery, R. E. Kanner, G. M. Villegas, and J. Schwartz, "Heart rate variability, air pollution, and cardiovascular disease risk," *Air Quality, Atmosphere Health*, vol. 113, no. 4, pp. 324– 332, 2006.

[37] J. Lelieveld, A. Haines, and A. Pozzer, "Air pollution and oxidative stress in

cardiovascular diseases," *Frontiers in Public Health*, vol. 7, p. 307, 2019.

[38] M. Boas, U. Feldt-Rasmussen, and K. M. Main, "Environmental endocrine disruptors: an evolutionary perspective," *The Journal of Clinical Endocrinology Metabolism*, vol. 91, no. 6, pp. 2074–2080, 2006.

[39] C. A. Pope III and D. W. Dockery, "Air pollution and heart rate variability," *Circulation*, vol. 109, no. 21, pp. e211–e211, 2004.

[40] A. Bhatnagar, "Environmental cardiology: studying mechanistic links between pollution and heart disease," *Circulation research*, vol. 99, no. 7, pp. 692–705, 2006.

[41] W. J. Paulus and C. Tschope, "Endothelial dysfunction: a pathophysiologic factor in heart failure," *Circulation*, vol. 116, no. 19, pp. 2034–2046, 2007.

[42] R. D. Brook, S. Rajagopalan, C. A. Pope III, J. R. Brook, A. Bhatnagar, A. V. Diez-Roux, F. Holguin, Y. Hong, R. V. Luepker, M. A. Mittleman, *et al.*, "Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the american heart association," *Circulation*, vol. 121, no. 21, pp. 2331–2378, 2010.

[43] J. B. West, *Pulmonary Pathophysiology: The Essentials*. Lippincott Williams Wilkins, 8 ed., 2012.