

## Article

# Decoding Physical and Cognitive Impacts of Particulate Matter Concentrations at Ultra-Fine Scales

Shawhin Talebi , David J. Lary , Lakitha O. H. Wijeratne , Bharana Fernando , Tatiana Lary, Matthew Lary, John Sadler, Arjun Sridhar, John Waczak, Adam Aker and Yichao Zhang

Hanson Center for Space Sciences, University of Texas at Dallas, Richardson, TX 75080, USA; david.lary@utdallas.edu (D.J.L.); lhw150030@utdallas.edu (L.O.H.W.); ashen.fernando@utdallas.edu (B.F.); tatiana.lary@utdallas.edu (T.L.); matthew.lary@utdallas.edu (M.L.); jcs170001@utdallas.edu (J.S.); arjun.sridhar@utdallas.edu (A.S.); john.waczak@utdallas.edu (J.W.); adam.ra1@utdallas.edu (A.A.); yxz154930@utdallas.edu (Y.Z.)

\* Correspondence: shawhintalebi@gmail.com

**Abstract:** The human body is an incredible and complex sensing system. Environmental factors trigger a wide range of automatic neurophysiological responses. Biometric sensors can capture these responses in real time, providing clues about the underlying biophysical mechanisms. In this prototype study, we demonstrate an experimental paradigm to holistically capture and evaluate the interactions between an environmental context and physiological markers of an individual operating that environment. A cyclist equipped with a biometric sensing suite is followed by an environmental survey vehicle during outdoor bike rides. The interactions between environment and physiology are then evaluated through the development of empirical machine learning models, which estimate particulate matter concentrations from biometric variables alone. Here, we show biometric variables can be used to accurately estimate particulate matter concentrations at ultra-fine spatial scales with high fidelity ( $r^2 = 0.91$ ) and that smaller particles are better estimated than larger ones. Inferring environmental conditions solely from biometric measurements allows us to disentangle key interactions between the environment and the body. This work sets the stage for future investigations of these interactions for a larger number of factors, e.g., black carbon, CO<sub>2</sub>, NO/NO<sub>2</sub>/NO<sub>x</sub>, and ozone. By tapping into our body's 'built-in' sensing abilities, we can gain insights into how our environment influences our physical health and cognitive performance.

**Keywords:** holistic sensing; particulate matter; physiology; machine learning



**Citation:** Talebi, S.; Lary, D.J.; Wijeratne, L.O.H.; Fernando, B.; Lary, T.; Lary, M.; Sadler, J.; Sridhar, A.; Waczak, J.; Aker, A.; et al. Decoding Physical and Cognitive Impacts of Particulate Matter Concentrations at Ultra-Fine Scales. *Sensors* **2022**, *22*, 4240. <https://doi.org/10.3390/s22114240>

Academic Editor: Loris Nanni

Received: 6 April 2022

Accepted: 30 May 2022

Published: 2 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Over four million premature deaths worldwide were attributed to outdoor air pollution in 2016 [1]. In 2019, 99% of the global population resided in areas that fell short of the World Health Organization (WHO) air quality guidelines [1]. There is mounting evidence that poor air quality negatively impacts respiratory, cardiovascular, and cerebrovascular health [2–7]. Further, there is emerging evidence on the impact of poor air quality on neurological outcomes including chronic diseases (e.g., Alzheimer's disease and dementia) [2,8,9] and acute cognitive impairment [10–14].

Although several large-scale epidemiological studies show the negative effects of air pollution on physical and cognitive health [2–7], these studies largely focused on coarse spatial (~10 miles) and temporal (~1 day) scales. Much less research focuses on ultra-fine spatial (~1 m) and temporal (~10 s) scales that make simultaneous environmental and holistic biometric observations of the human physiological responses.

Before an extreme result such as a disease occurs, poor air quality already negatively impacts human physical and cognitive performance [10–14]. Through this work, we investigate how air pollution impacts human health and performance by examining the relationship between environmental air quality measurements and automatic physiological responses at ultra-fine scales. Additionally, this work establishes the groundwork for future

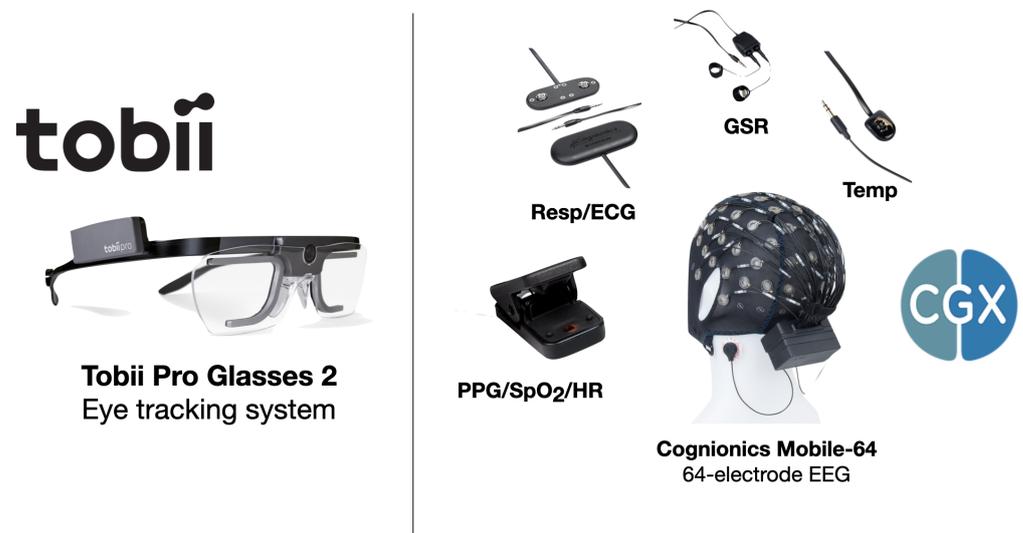
investigations by developing an experimental paradigm based on two main ingredients: holistic sensing and machine learning. Holistic sensing aims to capture all the relevant information about a system of interest. Machine learning is a framework that allows computers to learn by example and enables the development of high-fidelity empirical models [15].

This pilot study extends past works that examined interactions of cardiovascular variables such as heart rate (HR), heart rate variability (HRV), and blood pressure (BP) with air quality on fine scales [16–18]. The main contribution of this prototype study is that we augment cardiovascular markers with other biometrics, including electroencephalography (EEG), pupillometry, galvanic skin response (GSR), body temperature, oxygen saturation (SpO<sub>2</sub>), and respiration rate (RR). This extended set of variables captures both the cardiovascular and cognitive status of the participant. A study of air quality and human physiology at the ultra-fine level may shed light on the biophysical mechanisms that underlie their interactions.

## 2. Materials and Methods

### 2.1. Holistic Sensing

The data in this pilot study are a subset of a holistic biometric and environmental sensing paradigm. The goal of holistic sensing is to capture all relevant information about a system of interest. The full sensor array includes biometric monitors such as electroencephalography (EEG), eye tracking glasses, electrocardiography (ECG), galvanic skin response (GSR), body temperature, blood oxygen saturation (SpO<sub>2</sub>), and heart rate (HR), in addition to environmental factors such as particulate matter (PM), chemical composition of air, temperature, pressure, humidity, visible light spectrum, and more. The full array of biometric and environmental sensing systems are shown in Figures 1 and 2, respectively. After processing raw sensor recordings, this full sensor array has a feature space approaching 20,000 variables (~16,500 biometric and ~2000 environmental). In the present study, we focus on a relatively small subset, consisting of 329 biometric and 51 environmental variables.



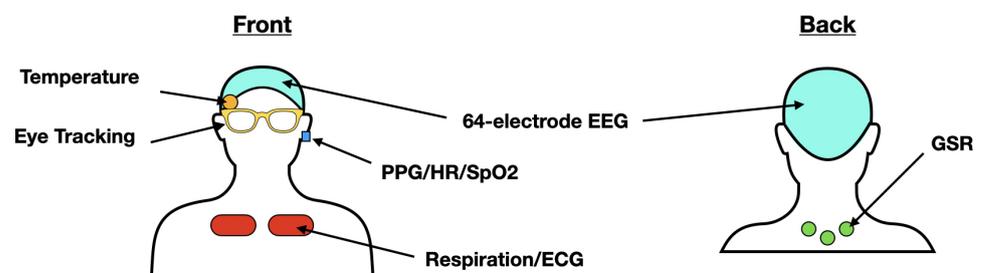
**Figure 1.** Biometric sensing systems. (Left) Tobii Pro Glasses 2 eye tracking system. This instrument performs eye tracking data, pupillometry, and provides two videos streams of the participant’s POV and eyes, respectively; (right) Cognionics Mobile-64 and AIM2 systems. Sensing suite includes 64-electrode EEG, PPG which measures SpO<sub>2</sub> and HR, respiration/ECG sensors, GSR, and temperature probe.



**Figure 2.** Images of full environmental sensing suite. Fidas<sup>®</sup> Frog Fine Dust Monitoring System measures particulate matter concentrations at 100 different size bins. The AIRMAR 220WX WeatherStation<sup>®</sup> Instrument samples barometric pressure, wind speed and direction, ambient temperature, and more. The 2B Technologies Black Carbon Photometer measures atmospheric black carbon particulates using long-path photometry. The 2B Technologies Model 205 Dual Beam Ozone sensor is a UV-based ozone monitor. The Konica Minolta CL-500A Illuminance Spectrometer measures the spectral irradiance from 360 to 780 nm at every nanometer. The portable mass spectrometer was constructed by the UNT Laboratory of Imaging Mass Spectrometry and measures charge mass ratios ranging 1–300 amu. The 2B Technologies Model 405 nm NO<sub>2</sub>/NO/NO<sub>x</sub> Monitor<sup>™</sup> directly measures atmospheric Nitrogen Dioxide (NO<sub>2</sub>) and Nitric Oxide (NO). The LI-COR LI-850 Gas Analyzer measured CO<sub>2</sub> and water vapor in the air.

The biometric sensing suite used in this research aims to comprehensively capture the physiological and cognitive status of the participant without restricting the participant's actions, movements, or decision-making. The goal is to gather the maximum amount of information with minimal interruption of normal behaviors. Biometric sensors are placed on the participant in such a way to allow for unrestricted mobility (Figure 3). Sensor recording units and other devices are organized in a backpack worn by the participant that all together weighs less than 10 lbs (left panel in Figure 4).

## Biometric Sensor Placement



**Figure 3.** Schematic of biometric sensor placement on participant. (**Left**) Cartoon of front participant view. The 64-electrode EEG sits on the participant's head. A temperature probe is placed under the EEG cap on the right temple. Eye tracking glasses are carefully placed on participant, avoiding EEG electrodes. PPG sensor is secured to left ear lobe. Respiration sensors are placed near the top of the chest. (**Right**) Cartoon of back participant view. GSR sensors are placed below the back of the neck.



**Figure 4.** Data collection images. (**Left**) custom-made backpack to house biometric devices and recording computer; (**middle**) participant and environmental survey vehicle riding in tandem during data collection; (**right**) environmental sensors organized in trunk of electric survey vehicle.

Over 100 biometric markers are measured at sampling rates of 500 Hz and 100 Hz. These quantities are processed to derive over 329 variables for the present analysis. This holistic biometric sensing suite integrates two independent sensing systems which are shown in Figure 1. Eye tracking is recorded 100 times a second using the Tobii Pro Glasses 2. Data from the glasses produced average pupil diameter, the difference in pupil diameter between left and right eyes, and the 3D spatial distance between pupil centers. All other biometric data are measured 500 times a second using the Cognionics Mobile-64 and AIM2 systems. These systems include a 64-electrode EEG, temperature sensor, respiration sensor, photoplethysmogram (PPG), and galvanic skin response (GSR) measurement. Heart rate and SpO<sub>2</sub> values are automatically computed by the AIM2 system using the PPG. Heart rate variability (HRV) and respiration rate (RR) are derived from respiration sensor data with a custom MATLAB script. All biometric data were down-sampled to 1/30 Hz (every 30 s) to match particulate matter recordings.

A holistic evaluation of an environmental setting is the ultimate goal of the sensing suite used in this study. This suite brings together several sensing packages, including fine dust monitoring from the Fidas<sup>®</sup> Frog, temperature, humidity, pressure, and wind speed and direction recorded with the AIRMAR Weatherstation 220WX; the full spectrum of visible light (360–780 nm) captured by the Konica Minolta Illuminance Spectrophotometer; dedicated gas monitors for black carbon, ozone, NO/NO<sub>2</sub>/NO<sub>x</sub>, and CO<sub>2</sub>/H<sub>2</sub>O; as well as a portable mass spectrometer (Figure 2). However, due to its significant societal relevance, for this pilot study, we focus on particulate matter (PM) concentrations recorded using the Fidas<sup>®</sup> Frog fine dust monitoring system. This instrument simultaneously measures PM mass fractions of PM<sub>1</sub>, PM<sub>2.5</sub>, PM<sub>4</sub>, PM<sub>10</sub>, and a distribution within a size range of 0.18–100 micrometers, as well as the total particle count density (dCn). PM data were recorded at sampling rate of 1 Hz and down-sampled to 1/30 Hz (every 30 s).

## 2.2. Data Collection

Biometric data collection was restricted to a single participant due to logistical constraints arising from the COVID-19 pandemic. However, future works will include data from multiple participants. The small population size in the present study is mitigated by two factors. First, data were collected over three separate days, providing a range of contexts. Additionally, the participant circled the same trail multiple times, offering multiple observations of identical positions and 360-degree changes in wind-direction angles.

Data were collected while the participant rode a bicycle in a dynamic outdoor setting. An electric survey vehicle equipped with a suite of environmental sensors followed safely behind the participant during all rides (middle image in Figure 4). Although several dimensions of the environmental context were sampled (e.g., ambient light, temperature, pressure, mass spectra, etc.), here, we focus on the relationship between particulate matter values and biometric variables. Additional relationship will be explored in future works.

Data collection took place in May and June of 2021 at Breckenridge Park located in Richardson, TX over three separate days, which included four to five trials per day. The first two trials consisted of two minutes of eyes closed and eyes open baseline biometric measurements, respectively. The third trial consisted of a “warm-up” ride, where the participant cycled

to a public bike trail in tandem with the electric survey vehicle. Additional trials consisted of the participant repeatedly cycling a one-mile loop on a public bike trail. The participant was free to stop cycling at their discretion. Data collection was halted whenever cycling stopped. If the participant chose to continue, a new data collection trial was initiated.

The complete dataset consists of 188 data records collected every 30 s (total time of about 1.5 h) with 329 biometric predictor variables and 51 PM target variables. Biometric predictor variables include: delta (1–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (13–25 Hz), and gamma (25–70 Hz) band power densities for each of the 64 EEG electrodes, body temperature, GSR, HR, HRV, RR, SpO<sub>2</sub>, average pupil diameter, difference between left and right pupil diameters, and the 3D spatial distance between left and right pupil centers. Environmental PM target variables include: PM<sub>1</sub>, PM<sub>2.5</sub>, PM<sub>4</sub>, PM<sub>10</sub>, PM<sub>Total</sub>, and 45 different PM size bins ranging of 0.18–10 µm measured in µg/m<sup>3</sup>, as well as particle count density (dCn) measured in P/cm<sup>3</sup>. The data are publicly available at the Zenodo datastore: <https://zenodo.org/record/6326357#.Yieu4RPMJb8>, accessed on 29 May 2022 (see Supplementary Materials).

Ethical approval declarations: All experimental protocols were approved by The University of Texas at Dallas Institutional Review Board and informed consent was obtained from the study participant.

### 2.3. Model Development

All models of PM concentration are obtained by an ensemble of decision trees for regression with a hyperparameter optimization process [19–24]. Ninety percent of the data is used for training, while 10% is assists as an independent validation dataset. Scripts for model training are freely available at the GitHub repository: <https://github.com/mi3nts/DUEDARE>, accessed on 29 May 2022 (see Supplementary Materials).

## 3. Results and Discussions

In this work, we used a data-driven experimental paradigm to develop and explore several empirical machine learning models which describe the connection between ambient air particulate matter (PM) concentrations and the biometric variables of an individual breathing that air. Due to logistical constraints imposed by the COVID-19 pandemic, we were only able to collect data from one participant. Additional participants will be included in future research. Two factors, however, mitigate the limited population size in this pilot study. First, the data collection took place over three days, which allowed for contextual variability. Furthermore, the participant repeatedly circled the same trail, allowing for multiple observations of identical spatial positions and 360-degree changes in wind-direction angles.

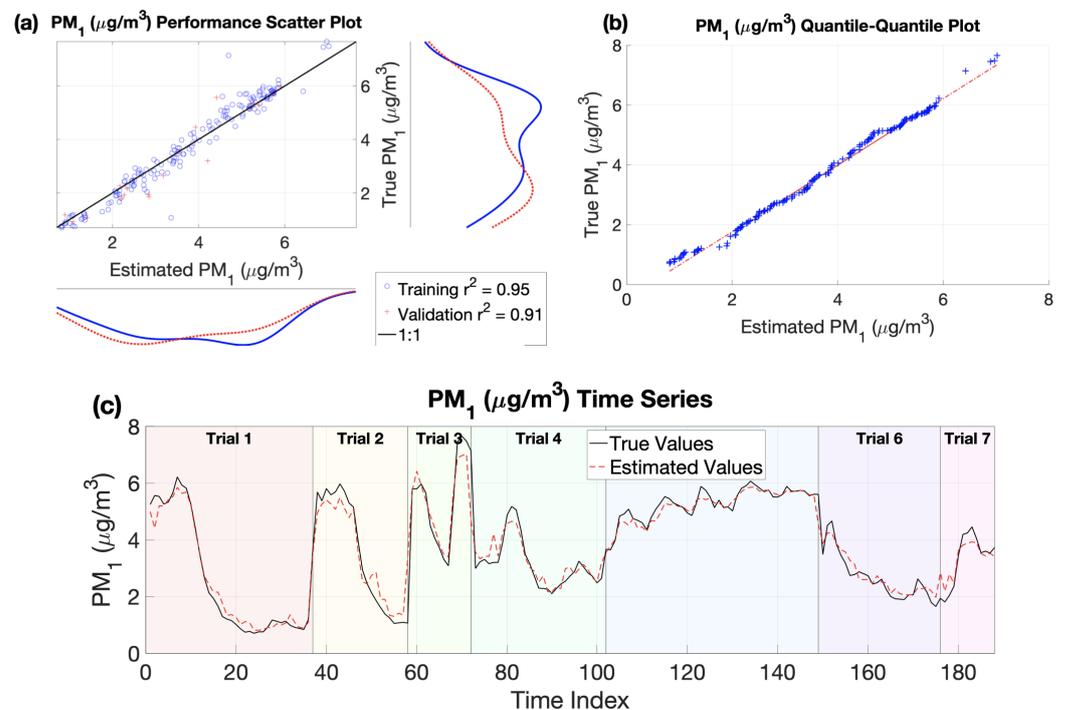
The estimated PM values included: PM<sub>1</sub>, PM<sub>2.5</sub>, PM<sub>4</sub>, PM<sub>10</sub>, PM<sub>Total</sub>, and 45 different PM size bins ranging of 0.18–10 µm measured in µg/m<sup>3</sup>, as well as particle count density (dCn) measured in particles per m<sup>3</sup>. For model development, 329 biometric predictor variables were available. Each machine learning model used was a trained ensemble of decision trees for multi-variate, non-linear, non-parametric regression with full hyperparameter optimization [19–24]. The empirical models are evaluated using two key metrics. First, the model accuracy was assessed using the squared correlation coefficient ( $r^2$ ) between the model prediction and the true PM values. Second, a ranking of predictor variable importance was obtained as the weighted average importance of each predictor across the ensemble.

We first evaluated six machine learning models for particulate matter (PM) values, which estimated the particle count density (dCn), PM<sub>1</sub>, PM<sub>2.5</sub>, PM<sub>4</sub>, PM<sub>10</sub>, and PM<sub>Total</sub>. Three hundred and twenty-nine biometric predictor variables were used as model inputs, including delta (1–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (13–25 Hz), and gamma (25–70 Hz) band power densities for each of the 64 EEG electrodes, body temperature, galvanic skin response (GSR), heart rate (HR), heart rate variability (HRV), respiration rate (RR), blood oxygen saturation (SpO<sub>2</sub>), average pupil diameter, the difference between the left and right pupil diameters (anisocoria), and the 3D spatial distance between the left and right pupil centers (vergence eye movement). Then, using an Occam's razor principle,

the top nine important biometric predictor variables were used to train an additional six models for the same PM variables.

Two subsets of nine biometric predictor variables were used to train two different sets of empirical machine-learning models. The first subset includes EEG variables, and the second subset does not. This first subset was obtained via the Occam's razor principle mentioned previously, while the second included all nine non-EEG biometric variables from the 329 available biometric predictors. The cognitive effects of air quality can be identified by evaluating predictive models with and without EEG quantities.

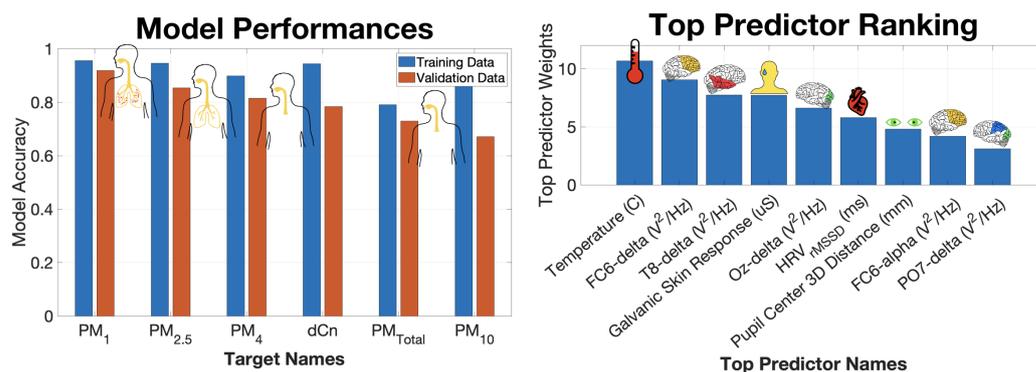
The best performing model using the top nine EEG and non-EEG biometric predictors was for PM<sub>1</sub>. This model had the highest accuracy with a validation dataset  $r^2 = 0.91$ . Comparison plots between estimated and ground truth PM<sub>1</sub> values are given in Figure 5. In the top-left plot, the estimated and true PM<sub>1</sub> concentrations in both the training (blue circles) and validation (red pluses) datasets closely following the perfect fit (black) line. In the top-right plot, the quantile–quantile comparison shows the distribution of measured PM<sub>1</sub> values closely resembles the distribution of estimated PM<sub>1</sub> values. Finally, in the bottom plot, the time series of the estimated PM<sub>1</sub> values (dashed red line) tracks very closely to the true values (solid black line) over seven different trials spanning three separate days.



**Figure 5.** Top performing model (PM<sub>1</sub>) plots comparing predicted and ground truth values. (a) Scatter plot of true versus predicted PM<sub>1</sub> values. A perfect fit is indicated by the 1:1 line shown in black. Training data are shown as blue circles and validation data are plotted as red pluses. (b) Quantile–quantile plot of true versus predict PM<sub>1</sub> values. Identical true and predicted distributions would result in a perfect  $y = x$  line. (c) Time-series plot of true PM<sub>1</sub> values (solid black line) and predicted PM<sub>1</sub> values (dashed red line). Background color indicates the trial number associated with each time period. Trials 1–3 were collected on 26 May 2021; trials 4–5 were collected on 9 June 2021; and trials 6–7 were collected on 10 June 2021.

The performances of the PM<sub>1</sub> and five other PM models in this cohort are ranked in the left panel of Figure 6. The training and independent validation dataset performances are plotted in blue and orange, respectively, and sorted in descending order of independent validation performance. As previously discussed, PM<sub>1</sub> measured in  $\mu\text{g}/\text{m}^3$  was best reproduced by the nine biometric predictors (validation  $r^2 = 0.91$ ). The empirical models based on the same biometric predictors were less able to accurately estimate the larger PM<sub>10</sub>

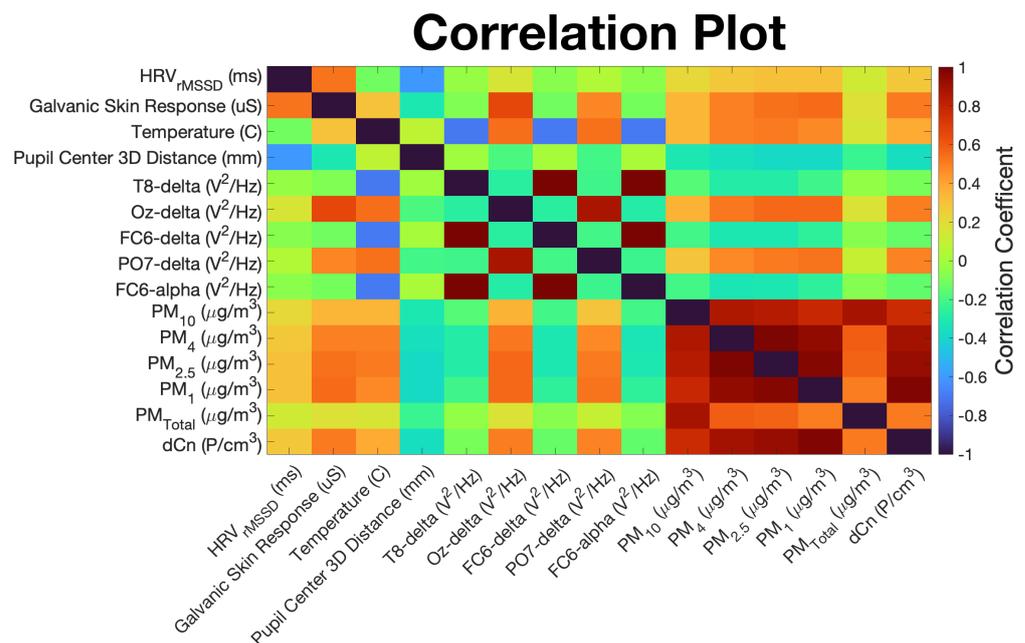
(validation  $r^2 = 0.67$ ) values and  $PM_{Total}$  (validation  $r^2 = 0.72$ ), which are dominated by  $PM_{10}$  due to the larger masses. The poor performance of these models could be explained by the fact that there are significantly fewer large particles than small particles, and thus the larger particles are not as well mixed as the far more numerous and well-mixed smaller particles. Because of their greater bulk, larger particles settle more quickly. As a result, the concentrations of large particles collected by the survey vehicle and those inhaled by the subject a few meters away are likely to differ more than for the smaller particles. Second, it is possible that the larger particles have less of an impact on the participant's physical and cognitive state because they are less likely to penetrate deeply into the respiratory and circulatory systems [25].



**Figure 6.** Summary of empirical PM concentration models estimated from 9 EEG and non-EEG biometric predictor variables. (Left) Ranking of model performance defined as squared correlation coefficient between predicted and true PM values. Training and validation dataset performances for each model are shown in blue and orange, respectively. Sorting is based on validation dataset performance. Overlaid graphics indicate the deposition of the respective PM size bins in the airways [25]. (Right) Predictor importance ranking aggregated across all 6 models.

Each of the six empirical machine learning models has an associated predictor importance ranking, which quantifies the role of individual input predictor variables in estimating the respective PM target variable. The aggregated ranking of top predictors, shown in the right plot in Figure 6, elucidates which biometric variables are most helpful to the empirical models in discerning PM values. The most important predictor variable in estimating PM values was the body temperature measured at the participant's right temple. Surprisingly, the cardiovascular variable, HRV, played less of a role. Other important biometrics included GSR and the distance between the pupil centers of the eyes. GSR is a strong correlate of body temperature. The distance between the pupil centers is a proxy for vergence eye movements, which have been associated with attentional load and are a strong predictor of cognitive status [26,27]. The delta band (1–3 Hz) power densities for the FC6, T8, and Oz electrodes were found to play an important role in estimating PM values. FC6 is above the frontal cortex on the right side of the head, T8 corresponds to the right temporal lobe, and Oz sits on top of the primary visual cortex.

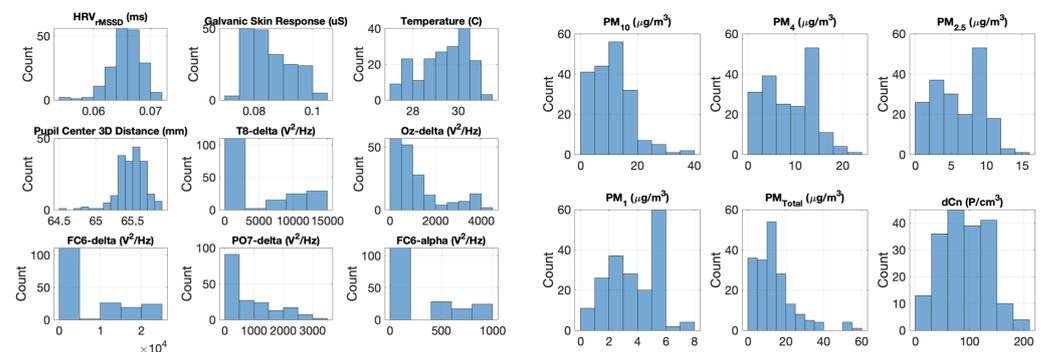
Correlations between predictor and target variables are visualized as a color-filled correlation plot in Figure 7. Strong positive correlations are indicated by dark red squares, strong negative correlations are shown by dark blue squares, and the lack of correlation is indicated by green squares. From this plot, we can see HRV, GSR, body temperature, and the delta band power densities of the Oz and PO7 electrode signals have strong positive correlations with all target variables except  $PM_{Total}$ . In other words, as these predictor variables increase, so do the corresponding PM target variables. PM target variables show the greatest negative correlation with the 3D spatial distance between left and right pupil centers, suggesting that the pupils tend to converge with an increase in PM concentrations. Lastly, of all the target variables,  $PM_{Total}$  is most strongly correlated with  $PM_{10}$  values, which reflects the strong contribution of  $PM_{10}$  particles to  $PM_{Total}$ .



**Figure 7.** Correlation plot of top 9 EEG and non-EEG biometric predictor variables, along with 6 target PM variables. Positively correlated variable pairs are indicated by a red box, negatively correlated pairs are shown by blue boxes, and non-correlated pairs have green boxes.

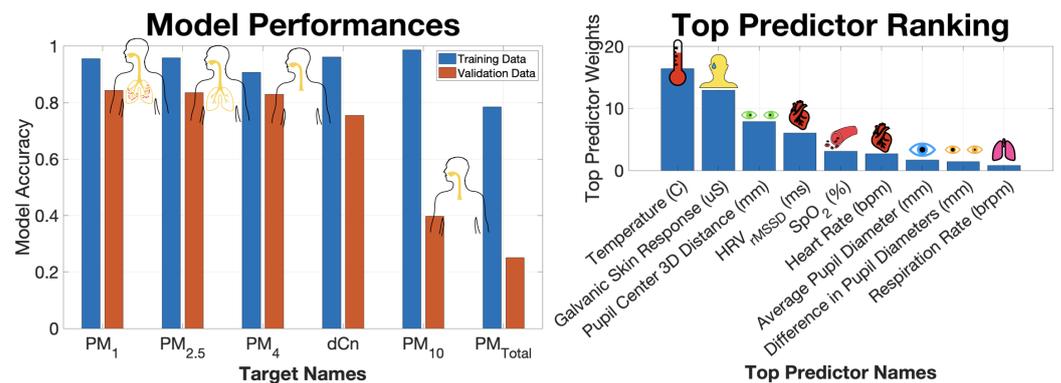
Histograms for both predictor and target variables are displayed in Figure 8. Plots are titled by the variable name and its respective physical units. From the target PM variable histograms in the right plot of Figure 8, the mass scales of different particle sizes are evident. Namely, the larger-sized  $PM_{10}$  particles vary over a much larger range ( $0\text{--}40\ \mu\text{g}/\text{m}^3$ ) than the smaller  $PM_4$  ( $0\text{--}20\ \mu\text{g}/\text{m}^3$ ),  $PM_{2.5}$  ( $0\text{--}15\ \mu\text{g}/\text{m}^3$ ), and  $PM_1$  particles ( $0\text{--}8\ \mu\text{g}/\text{m}^3$ ). This further explains the strong influence of  $PM_{10}$  values on  $PM_{\text{Total}}$ .

Next, an additional set of six empirical machine learning models for the same set of PM targets ( $dCn$ ,  $PM_1$ ,  $PM_{2.5}$ ,  $PM_4$ ,  $PM_{10}$ , and  $PM_{\text{Total}}$ ) were evaluated, except this time the PM targets were estimated from nine non-EEG biometric predictor variables (body temperature, GSR, HR, HRV, RR,  $SpO_2$ , average pupil diameter, difference between left and right pupil diameters, and the 3D spatial distance between left and right pupil centers).



**Figure 8.** (Left) Histograms of 9 EEG and non-EEG predictor variables. Plots are titled by variable name and its physical units. (Right) Histograms of 6 different PM target variables variables. Plots are titled by variable name and its physical units.

The model performance ranking for the six empirical PM models estimated from the nine non-EEG biometric predictor variables is shown in the left panel of Figure 9. We see that the smaller particles are better estimated by the non-EEG biometrics. Again, this result may be due to better mixing of smaller particles or to deeper penetration of those particles into the respiratory system or both.



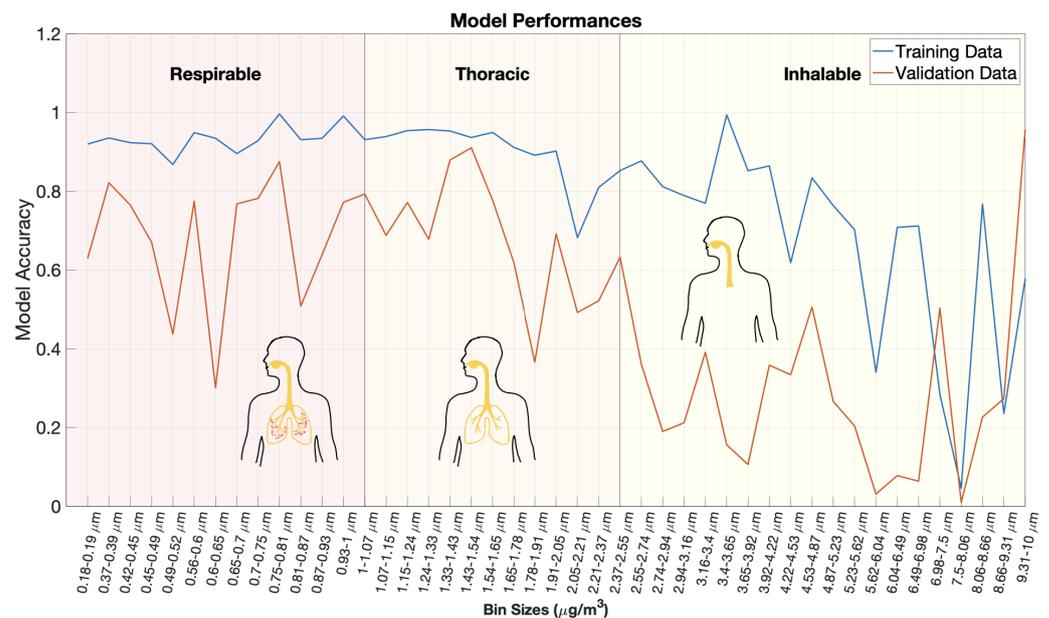
**Figure 9.** Summary of empirical PM concentration models estimated from 9 non-EEG biometric predictor variables including eye tracking, respiratory, and other physiological variables. **(Left)** Ranking of model performance defined as squared correlation coefficient between predicted and true PM values. Training and validation dataset performances for each model are shown in blue and orange, respectively. Sorting is based on validation dataset performance. Overlaid graphics indicate the deposition of the respective PM size bins in the airways [25]. **(Right)** Predictor importance ranking aggregated across all 6 models.

Comparing the performance rankings in Figures 6 and 9, there are clear changes in model accuracies. All models with the exception of PM<sub>4</sub> exhibit a drop in performance. The largest drop occurs for the already poorly performing PM<sub>Total</sub> (drop in validation  $r^2 = 0.47$ ) and PM<sub>10</sub> (drop in validation  $r^2 = 0.28$ ) models.

There is overlap between the importance rankings of Figures 6 and 9. In both cases, body temperature is the most significant predictor of the PM values. Additionally, GSR maintains its order in the ranking as the 2nd most important non-EEG predictors. Although respiratory variables such as HRV and HR appear in the top six of the importance ranking, these variables trail behind temperature, GSR, and the distance between the eye pupil centers.

The observation that smaller particles are better estimated than larger-sized particles is explored further by evaluating model performances for a finer particulate size resolution. Here, 45 models were trained to estimate different PM size bins ranging from 0.18 to 10 micrometers using the nine non-EEG biometrics listed above. Model accuracy is plotted against bin size in Figure 10. Training and validation accuracies are plotted as blue and orange lines, respectively. The regional depositions of each particle size bin are indicated by a label and background shading [25,28]. The smallest particles (PM<sub>1</sub>) are classified as respirable and can penetrate to the alveoli. The next smallest size bin is thoracic (PM<sub>2.5</sub>) which consists of particle penetrating into the bronchioles. The largest size bin are the inhalable particles (PM<sub>10</sub>) which can enter into the nose, mouth, and trachea.

There is a clear drop in both training and validation dataset accuracies for size bins between 2 and 3 micrometers, corresponding to thoracic and inhalable particles. For particle size bins above this drop, there is large degree of variation in model performances; however, most have a poor performance with a validation  $r^2$  below 0.4. While the results may imply that smaller particles have a greater impact on physiological systems due to their deeper deposition, that conclusion cannot be reached based upon the present data. The drop in performance for larger particles may be explained in part or completely by the fact that smaller particles are more plentiful and better mixed. An evaluation of the relative contributions of each of these factors requires further investigation.



**Figure 10.** Model accuracies plotted against bin size. Forty-five separate PM models were trained for size bins ranging from 0.18 to 10 micrometers. PM values were estimated solely from 9 non-EEG biometric variables. Training dataset performance is plotted as a blue line and validation dataset performance is plotted in orange. A clear drop in model performance is observed between 2–3 micrometers. Overlaid graphics indicate the deposition of the respective PM size bins in the airways [25,28].

#### 4. Conclusions

The human body and its environment form a complex ecosystem. An important aspect of this system is air quality and the impact it has on the human body. Environmental factors trigger physiological responses that can be detected by holistic biometric sensing. In this prototype study, we used an ultra-fine holistic sensing paradigm to demonstrate how particulate matter concentrations in the ambient environment can be accurately estimated using only nine biometric variables. In addition, smaller particles were found to be more accurately estimated. Two potential causes may explain this result. First, smaller particles are much more abundant and well mixed in the ambient environment than larger ones, thus resulting in a greater similarity between particles inhaled by the participant and collected by the survey vehicle. Secondly, smaller particles can deposit into the respiratory system more deeply, and may have a greater impact on the body. Further investigation is needed to assess the relative contributions, if any, of these two factors, since they are not mutually exclusive.

The largest limitation of this work is that data collection was restricted to a single participant. While it is not clear if the observations from this pilot study will extend to a broader population, we have laid a foundation for future investigations of environmental impacts on human physiology on ultra-fine scales. Future research will include data from multiple participants. Additionally, several other environmental variables were collected (e.g., ambient light, temperature, black carbon, ozone, NO/NO<sub>2</sub>/NO<sub>x</sub>, etc.) and will be evaluated for their physiological interactions. By understanding the key interactions between the environment and the human body, health and performance can be improved across a variety of domains.

**Supplementary Materials:** The data and code have been made publicly available. The full data set is available at the Zenodo data store: <https://zenodo.org/record/6326357#.Yieu4RPMJb8> (accessed on 29 May 2022) and code is available at the GitHub: <https://github.com/mi3nts/DUEDARE> (accessed on 29 May 2022).

**Author Contributions:** Methodology, D.J.L., S.T. and T.L.; software, S.T.; formal analysis, D.J.L., S.T., T.L. and B.F.; data curation, S.T., D.J.L., L.O.H.W., B.F., T.L., M.L., J.S., A.S., A.A., Y.Z. and J.W.; writing—original draft preparation, S.T. and D.J.L.; writing—review and editing, S.T., D.J.L., B.F., T.L. and J.S.; visualization, S.T.; supervision, D.J.L. All authors have read and agreed to the published version of the manuscript.

**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.

**Institutional Review Board Statement:** All experimental protocols were approved by The University of Texas at Dallas Institutional Review Board.

**Informed Consent Statement:** Informed consent was obtained from the study participant.

**Data Availability Statement:** The data used in this study are publicly available at the Zenodo datastore: <https://zenodo.org/record/6326357#.Yieu4RPMJb8>, accessed on 29 May 2022.

**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. The authors acknowledge the OIT-Cyberinfrastructure Research Computing group at the University of Texas at Dallas and the TRECIS CC\* Cyberteam (NSF 2019135) for providing HPC resources that contributed to this research (<https://utdallas.edu/oit/departments/circ/>, accessed on 22 February 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

WHO	World Health Organization
HR	Heart Rate
BP	Blood Pressure
EEG	Electroencephalography
GSR	Galvanic Skin Response
ECG	Electrocardiography
HRV	Heart Rate Variability
PPG	Photoplethysmogram
RR	Respiration Rate
PM	Particulate Matter
dCn	Total Particle Count Density
SpO <sub>2</sub>	Blood Oxygen Saturation

## References

1. WHO. Ambient (Outdoor) Air Pollution. Retrieved 24 January 2022. Available online: [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-airquality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-airquality-and-health) (accessed on 29 May 2022).
2. Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and Health Impacts of Air Pollution: A Review. *Front. Public Health* **2020**, *8*, 14. [CrossRef] [PubMed]
3. Orellano, P.; Reynoso, J.; Quaranta, N.; Bardach, A.; Ciapponi, A. Short-term exposure to particulate matter (PM<sub>10</sub> and PM<sub>2.5</sub>), nitrogen dioxide (NO<sub>2</sub>), and ozone (O<sub>3</sub>) and all-cause and cause-specific mortality: Systematic review and meta-analysis. *Environ. Int.* **2020**, *142*, 105876. [CrossRef] [PubMed]
4. Daellenbach, K.R.; Uzu, G.; Jiang, J.; Cassagnes, L.E.; Leni, Z.; Vlachou, A.; Stefanelli, G.; Canonaco, F.; Weber, S.; Segers, A.; et al. Sources of particulate-matter air pollution and its oxidative potential in Europe. *Nature* **2020**, *587*, 414–419. [CrossRef] [PubMed]
5. Pope, C.A.; Brook, R.D.; Burnett, R.T.; Dockery, D.W. How is cardiovascular disease mortality risk affected by duration and intensity of fine particulate matter exposure? An integration of the epidemiologic evidence. *Air Qual. Atmos. Health* **2010**, *4*, 5–14. [CrossRef]

6. Brook, R.D.; Rajagopalan, S.; Pope, C.A.; Brook, J.R.; Bhatnagar, A.; Diez-Roux, A.V.; Holguin, F.; Hong, Y.; Luepker, R.V.; Mittleman, M.A.; et al. Particulate Matter Air Pollution and Cardiovascular Disease. *Circulation* **2010**, *121*, 2331–2378. [[CrossRef](#)] [[PubMed](#)]
7. Brook, R.D.; Bard, R.L.; Kaplan, M.J.; Yalavarthi, S.; Morishita, M.; Dvonch, J.T.; Wang, L.; Yang, H.Y.; Spino, C.; Mukherjee, B.; et al. The effect of acute exposure to coarse particulate matter air pollution in a rural location on circulating endothelial progenitor cells: Results from a randomized controlled study. *Inhal. Toxicol.* **2013**, *25*, 587–592. [[CrossRef](#)] [[PubMed](#)]
8. Schikowski, T.; Altuğ, H. The role of air pollution in cognitive impairment and decline. *Neurochem. Int.* **2020**, *136*, 104708. [[CrossRef](#)] [[PubMed](#)]
9. Kelly, F.J.; Fussell, J.C. Air pollution and public health: Emerging hazards and improved understanding of risk. *Environ. Geochem. Health* **2015**, *37*, 631–649. [[CrossRef](#)] [[PubMed](#)]
10. Lavy, V.; Ebenstein, A.; Roth, S. The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation. 2014. Available online: [https://www.nber.org/system/files/working\\_papers/w20648/w20648.pdf](https://www.nber.org/system/files/working_papers/w20648/w20648.pdf) (accessed on 29 May 2022).
11. Chatzidiakou, L.; Mumovic, D.; Dockrell, J. *The Effects of Thermal Conditions and Indoor Air Quality on Health, Comfort and Cognitive Performance of Students*; University college London Institute for Environmental Design and Engineering: London, UK, 2015. [[CrossRef](#)]
12. Zhang, X.; Chen, X.; Zhang, X. The impact of exposure to air pollution on cognitive performance. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9193–9197. [[CrossRef](#)] [[PubMed](#)]
13. Shehab, M.A.; Pope, F.D. Effects of short-term exposure to particulate matter air pollution on cognitive performance. *Sci. Rep.* **2019**, *9*, 8237. [[CrossRef](#)] [[PubMed](#)]
14. Künn, S.; Palacios, J.; Pestel, N. Indoor Air Quality and Cognitive Performance. IZA Discussion Paper No. 12632. Available online: <http://dx.doi.org/10.2139/ssrn.3460848> (accessed on 29 May 2022).
15. The Royal Society. Machine learning: The power and promise of computers that learn by example. In *Report by the Royal Society*; The Royal Society: London, UK, 2017; Volume 66. [[CrossRef](#)]
16. Buteau, S.; Goldberg, M.S. A structured review of panel studies used to investigate associations between ambient air pollution and heart rate variability. *Environ. Res.* **2016**, *148*, 207–247. [[CrossRef](#)] [[PubMed](#)]
17. Amoabeng, Nti, A.A.; Robins, T.G.; Mensah, J.A.; Dwomoh, D.; Kwarteng, L.; Takyi, S.A.; Acquah, A.; Basu, N.; Batterman, S.; Fobil, J.N. Personal exposure to particulate matter and heart rate variability among informal electronic waste workers at Agbogbloshie: A longitudinal study. *BMC Public Health* **2021**, *21*, 2161. [[CrossRef](#)] [[PubMed](#)]
18. Santos, U.d.P.; Braga, A.L.F.; Giorgi, D.M.A.; Pereira, L.A.A.; Grupi, C.J.; Lin, C.A.; Bussacos, M.A.; Zanetta, D.M.T.; Saldiva, P.H.d.N.; Filho, M.T. Effects of air pollution on blood pressure and heart rate variability: A panel study of vehicular traffic controllers in the city of São Paulo, Brazil. *Eur. Heart J.* **2005**, *26*, 193–200. [[CrossRef](#)] [[PubMed](#)]
19. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Wadsworth, Inc.: Belmont, CA, USA, 1984.
20. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
21. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844. [[CrossRef](#)]
22. Friedman, J.H. *Greedy Function Approximation: A Gradient Boosting Machine*; The Annals of Statistics; Institute of Mathematical Statistics: London, UK, 2001; Volume 29, pp. 1189–1232. Available online: <http://www.jstor.org/stable/2699986> (accessed on 29 May 2022).
23. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]
24. Probst, P.; Wright, M.N.; Boulesteix, A.L. Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1301. [[CrossRef](#)]
25. Poh, T.Y.; Ali, N.A.T.B.M.; Aogáin, M.M.; Kathawala, M.H.; Setyawati, M.I.; Ng, K.W.; Chotirmall, S.H. Inhaled nanomaterials and the respiratory microbiome: Clinical, immunological and toxicological perspectives. *Part. Fibre Toxicol.* **2018**, *15*, 46. [[CrossRef](#)] [[PubMed](#)]
26. Balaban, C.D.; Kiderman, A.; Szczupak, M.; Ashmore, R.C.; Hoffer, M.E. Patterns of Pupillary Activity during Binocular Disparity Resolution. *Front. Neurol.* **2018**, *9*, 990. [[CrossRef](#)] [[PubMed](#)]
27. Huang, M.X.; Li, J.; Ngai, G.; Leong, H.V.; Bulling, A. Moment-to-Moment Detection of Internal Thought from Eye Vergence Behaviour. *arXiv* **2019**, arxiv:1901.06572.
28. World Health Organization. Occupational and Environmental Health Team. Hazard prevention and Control in the Work Environment: Airborne Dust. World Health Organization. 1999. Available online: <https://apps.who.int/iris/handle/10665/66147> (accessed on 29 May 2022).