Rebuttal- Reviewer 2

Dear Reviewer,

Thank you for giving us valuable suggestions to improve the presented research work. As per the given suggestions, we have incorporated all the changes and highlighted the changes in the blue color and figure, and table-caption changes have been represented using yellow color.

Q-1 How far into the future, if at all, are you predicting AQI values? Predicting AQI values from actual measured concentrations is literally what the AQI is – a mathematical conversion of concentration values. If your major contribution is that you can use low cost sensors to accurately produce the same AQI values as very expensive monitors (like regulatory monitors), then that needs to be made very clear. If you are predicting *future* values and creating a forecast, then that also needs to be clear too (or instead). Or, if neither, and you are simply stating an interesting finding that one can model AQI in this unique set of ways (which I think is truly interesting), then state as much early on in the introduction and in the Discussion section.

Thank you for giving us valuable suggestions. In the presented research work, we have used machine learning and deep neural network-based approaches to predict AQI values of various air pollutants such as PM 2.5, PM 10, CO, NO2, and O3. The presented approach can function lifelong and can be retrained according to the updated dataset.

Q-2 Has anyone else presented on this interesting finding – that linear regression of low-cost sensor data can be used to "predict" AQI with quite minimal error?

Thank you for giving us valuable inputs. According to my knowledge and based on the conducted literature work, we can assure that such an approach has not been used for the presented experiments. This was the primary motivation for conducting the presented research work.

Q-3 Section 2.2.3 needs much more information on the collection of data over 90 days. Where are data for PWP collected – which city and under which seasons? Where is/are monitoring station(s) located? Is this part of a larger data collection effort that is reported on elsewhere? This is critical for understanding the context for interpretation and validity of the model you present.

Thank you for giving us valuable inputs. We have made the required changes. In the presented research work, we have presented a PWP system that is configured with pollution sensing units such as SDS021, MQ07-CO, NO2-B43F, and Aeroqual Ozone(O3). These sensing units were utilized to collect and measure various pollutant levels such as PM 2.5, PM10, CO, NO2, and O3 for 90 days at Symbiosis International University, Pune, Maharashtra, India. The data collection was carried out between the duration of December 2019 to February 2020 during Winter.

Q-4 What is considered "truth" for these predictions? Are you comparing data against government reported AQI values at some specific location?

Thank you for giving us valuable inputs. In the presented research, we have formulated customized approach for calculating AQI values as described in section 2.2.3.

Q-5 The introduction is still long (it should be reduced considerably) and jumps around from topic to topic too much. Each topic should be its own paragraph or cluster of sentences. E.g., the health effects of different air pollutants. I suggest five paragraphs or paragraph clusters:

1. **Why air pollution is bad and, relative to other risk factors, how does air pollution exposure in general and for each specific pollutant that you will study compare.**
2. **What the AQI is (a metric for normalizing and communicating the risks associated with concentrations of different pollutants – for each pollutant and in aggregate).**
3. **What your study does (predict AQI) and why you chose the AQI as your metric of interest.**
4. **What it adds to the relevant literature – major contributions to the literature e.g. starting at Line 200**
5. **Where it succeeds or adds where other studies have failed or omitted. Perhaps this entire bit, which is roughly Lines 95 – 177, belongs as part of the Discussion.**

Discussion of cross-sensitivities of each sensor used would add context to potential external biasing factors in/of your models.

Thank you for giving us suggestions. As per the given suggestions, we have divided introduction section in 5 separate paragraphs and made the required changes.

Q-6 Line 38: "… is less of development and more of a risk…" seems pedantic and maybe overstated. I'd recommending saying something like "…development provides benefits but it also poses risks…"

Thank you for giving us suggestions. We have made the required changes.

Q-7 Line 42: the GBD is most typically referred to as "the Global Burden of Disease study" – no pluralization on "disease," and no inclusion of "Injuries, and Risk Factors"

Thank you for giving us suggestions. We have made the required changes.

Q-8 Line 42-46: GBD requires citation.

Thank you for giving us suggestions. We have added the required citation.

Q-9 Line 45: "daily" should be "DALYs"

Thank you for giving us formatting related suggestions. We have made the required changes.

Q-10 Line 49: "worse" should be "worst"

Thank you for giving us formatting related suggestions. We have made the required changes.

Q-11 Line 54: "hair"?

Thank you for giving us formatting suggestions. We have made the required changes.

Q-12 Line 59: Suggested edit: "…*can* be microscopic and reach the…"

Thank you for giving us formatting suggestions. We have made the required changes.

Q-13 Line 60: Vary in what kind of range? Presumably this means "can vary in size"?

Thank you for giving us formatting suggestions. We have made the required changes.

Q- 14 Line 69: Kyrkilis study needs more background – what 18 cities? Where were the cities – in India? 5 out of 18 cities had higher pollution than what? Than the other 13 cities? And those other 13 studies also had bad air quality? This is a confusing addition.

Thank you for giving us valuable suggestions. We have made the required changes.

Q-15 Line 76-78. Both the point and the meaning of this addition are unclear. Specifically, "…, which then results in knowing the formation process of the pollutant depending on the methodology." Perhaps this citation belongs in a different paragraph?

Thank you for giving us valuable suggestions. We have made the required changes.

Q-16 Lines 80-82. Do not belong in this paragraph, and may be removed completely.

Thank you for giving us valuable suggestions. We have made the required changes.

Q-17 Line 102. "Despite…" can start a new paragraph.

Thank you for giving us valuable suggestions. We have made the required changes.

Q-18 Line 106: For which pollutants is AQI calculated? Are any missing from your sensor array?

Thank you for giving us valuable suggestions. We have calculated AQI values for a variety of pollutants such as PM 2.5, PM 10, CO, NO2, and O3. We have verified that we are not missing any of the presented pollutants.

Q-19 Line 112 Kim et al seems to not fit in this paragraph. If you really want to discuss studies of health impacts of air pollutants, it should be its own paragraph or later in the Discussion.

Thank you for giving us valuable suggestions. We have made the required changes and shifted the citation in an appropriate paragraph.

Q-20 Line 145 – bracketed citations for recent high-profile mobile monitoring should be added

Thank you for giving us valuable suggestions. We have added the required citations.

Q-21 Generally: the use of "the proposed approach" and "the presented approach" in this write-up is confusing, because it typically refers to the paper that you are writing (not a paper that you are citing)

Thank you for giving us valuable suggestions. We have made the required changes. We have made the use of words such as analysed, discussed, researched, studied, examined, conducted study etc. We have highlighted it with yellow color throughout the paper.

Q- 22 Lines 185-186: prediction of what? AQI as it relates to those pollutants? Specific concentrations of those pollutants?

Thank you for giving us valuable suggestions. We have made the required changes.

Q- 23 Line 204: how far into the future, if at all, are you predicting AQI values?

Thank you for giving us valuable suggestions. Yes, we are predicting AQI values.

Q- 24 Please check Tables 1-5 for correctness. For example, I found the following inconsistencies and wonder if there may be other errors in these tables:

Thank you for giving us valuable suggestions. We have made the required changes and highlighted table and figure captions with yellow color.

Q-25 Line 247 states an operating range of 0.5 ppm, but Table 5 indicates a max of 1000 ppm.

Thank you for giving us valuable suggestions. We have made the required changes.

Q-26 Line 265 states an operating range of 0.5 ppm, but Table 5 indicates a max of 10 ppb.

Thank you for giving us valuable suggestions. We have made the required changes.

Q-27 Line 304: "Pollutants present are mainly…" should be "Pollutants measured are…". There are many more pollutants "present" in the air, and a list of "main" pollutants would probably differ based on which scientists you asked.

Thank you for giving us valuable suggestions. We have made the required changes.

Q-28 Line 353 / Equation 1: the first term cancels out to 1 as currently written. Pmax is not included in the equation as currently written.

Thank you for giving us valuable suggestions. We have made the required changes.

Q-29 Equations 1, 2, and 3 and Table 6 should have citations.

Thank you for giving us valuable suggestions. We have made the required citations in equations 1 to 3), and Table 6.

Q-30 Line 368: Kelvin does not use degrees (°)

Thank you for giving us valuable suggestions. We have made the required changes.

Q-31 Figure 5. Why does AQI go to 5000 in part b? Also in part b, CO does not show up very well. For parts a and b, why do different pollutants extend to different values along the Y (and x) axes? As an in-silica representation (or is this not the case?) I would expect all would go at least as far as the end of the frame or plot outline.

Thank you for giving us valuable suggestions. We have made the required changes in Figure 5.

Q-32 Line 427: "o-7 µm" should be "0-7 µm" (zero is the letter air)

Thank you for giving us valuable suggestions. We have made the required changes.

Q-33 Figure 6: The top 2 plots are the same, are they not? Also, these values are low for PM10 and PM2.5, and may very well be below the actual/practical detection limit of the low-cost PM sensors used. Also, are the data here from your low cost sensors or somewhere else? Where were they taken and what context do they represent (rush hour in some city downtown? etc.)

Thank you for giving us valuable suggestions. We have made the required changes in Figure 6. In the presented research work, we have presented a PWP system that is configured with pollution sensing units such as SDS021, MQ07-CO, NO2-B43F, and Aeroqual Ozone(O3). These sensing units were utilized to collect and measure various pollutant levels such as PM 2.5, PM10, CO, NO2, and O3 for 90 days at Symbiosis International University, Pune, Maharashtra, India. The data collection was carried out between the duration of December 2019 to February 2020 during Winter.

Q-34 Section 2.2.4: again, are these future values being predicted or are you comparing low-cost sensor-based AQI estimates to high-cost monitor based AQI calculations?

Thank you for giving us valuable suggestions. Here, the future values are being predicted.

Q-35 Line 503: Is this supposed to include the unit that each pollutant uses, too?

Thank you for giving us valuable suggestions. We have made the required changes.

Q-36 Lines 554-573: The analysis of the multi-pollutant model is interesting and seems reasonable if analysing pollutant-specific models (vs. the model that predicts total AQI from all pollutants in aggregate). Perhaps it should be reconsidered and re-written to account for the

fact that more than just concentration is at play – it is (At least) both the concentration of each pollutant *as well as* the frequency of variation and overall background presence of that pollutant.

Thank you for giving us valuable suggestions. Furthermore, in the conducted experiments, we have only considered a relation between quantities of pollutants and AQI values of pollutants at an outdoor site (Symbiosis International University, Pune, Maharashtra, India) for 90 days. We have not considered variations such as places with high altitude, geographical variations, and Spatio-temporal variations. These are the future directions for fellow researchers.

Q-37 Line 557: Does "x-intercept" mean "slope" or "model coefficient", or is it actually the point at which y (AQI) = 0 for each pollutant?

Thank you for asking important questions. It is our model coefficient.

Q-38 Line 637 "pollutant"

Thank you for giving us valuable suggestions. We have made the required changes.

Q-39 Line 650: should "86%" be "84%" – based on Table 9?

Thank you for giving us valuable suggestions. We have made the required changes.

Q-40 Line 665: The fact that you explored all of these methods should be included in the paper's Methods section. This is the first mention of them outside of the abstract.

Thank you for giving us valuable suggestions. We have made the required changes.

Q-41 Section 4: This is probably better suited as supplemental text rather than a section after the Results and Discussion.

Thank you for giving us valuable suggestions. We have made the required changes.

Q-42 Table 12 belongs in the Introduction

Thank you for giving us valuable suggestions. We have shifted Table 12 to introduction section and captioned as Table 1.

Q-43 Lines 739 & 741: By "some increases" do you mean "slight increases" or "small relative increases"?

Thank you for giving us valuable suggestions. We have made the required changes.

Q-44 Section 5 should also include the intriguing finding that linear regression can be used to accurately "predict" AQI. Again, what you mean by "predict" should be very clear here and in the rest of the paper.

Thank you for giving us valuable suggestions. We have made the required changes. In the presented research work, we have used machine learning and deep neural network-based approaches to predict AQI values of various air pollutants such as PM 2.5, PM 10, CO, NO2, and O3.