Exploring the use of AI to predict Air Quality Index from Photos of A Polluted Sky

1Vishwa Iyer
1Student
1American High School, California

***Abstract* - There has been a lot of discussion about air pollution in India. In 2019 alone, around 1.7 million people were killed in India as a result of air pollution Most of these deaths happen within poor urban and rural communities. Members of these communities perform more outdoor labour, are unable to afford devices such as air purifiers, and are often, unaware of the AQI value around them. The vast majority of Indian districts have no air monitoring stations. Without the adequate amount of monitors, citizens wouldn't be able to identify the air quality value in their communities. This research paper proposes the construction of an AI platform called AVA that allows users to simply take a photo of the sky. My AI model then predicts the AQI in the photo, and provides public health guidance. I deployed the platform in villages and small cities across India, to very promising results.**

***keywords* - computer vision, air pollution, AQI, AI model, dataset creation**

1 INTRODUCTION

There has been a lot of discussion about air pollution in India. In 2019 alone, around 1.7 million people were killed in India as a result of air pollution [1]. In Delhi, India's capital, one in three children has impaired lungs, and nationwide, there are over 35 million chronic asthma patients. Scientists believe that India's rising asthma cases are linked directly to the rising air pollution. [2]

The 1.7 million Indians killed due to air pollution in 2019 made up 18% of all deaths in India that year. [3] Globally, 6.7 million people died from air pollution due to different illnesses, such as stroke, heart disease, lung cancer, and chronic respiratory diseases, in 2019 alone. Consequently, India accounts for more than 25% of all deaths attributed to air pollution during the year.

Most of the deaths due to poor air pollution in India happen within poor urban and rural communities. My interviews with local politicians and non-profit leaders on the ground helped me understand that this is mostly due to poor communities performing more outdoor labour. Poor communities are also not able to afford devices such as air purifiers. My interviews with local leaders helped me understand that most people don't know the importance of the AQI value, and continue to leave their houses in toxic conditions.

In the months of October, November, and December, the AQI in Delhi usually soars to toxic amounts, and regularly stays above 300. This can be extremely dangerous, so it is important that citizens are able to get a reading for the air quality, so that they know whether it is safe to go outside or not.

As of September 2019, under the guidance of CPCB, there are 200 continuous air monitoring stations operating in 20 states and covering 116 districts with at least one monitor [4]. This means that the vast majority of Indian districts (602 districts) have no air monitoring stations. Air quality also changes significantly every few kilometres, and every district should have at least a few air quality monitors. Without the adequate amount of air quality monitors in their districts, citizens wouldn't be able to identify the air quality value in their communities. Without the AQI value, they wouldn't be able to take public health measures to protect themselves and their families.

Thankfully, India finds itself in a curious position where due to easily available data (India has the cheapest data costs in the world), there are now over 800 million smartphones in India, up from 350 million just 3 years ago. [5]

My hope is to use the recent proliferation of cheap smartphones to augment people's understanding of their air quality. AVA or the Air Variance Authority allows its users to simply take a photo of the sky outside their home, and my neural network model returns them an estimated AQI value, based on the photo. AVA also provides them with public health guidance on how they can reduce the AQI value within their home, mask guidance, instructions on how to build DIY masks, and phone numbers of local hospitals and politicians.

I trained AVA by building 4 AQI buckets of air quality values, ranging from 0-50, 50-200, 200-400, and 400+. Images were sourced from Twitter and Instagram (using the popular hashtag #smogselfies), news websites, and a crowdsourcing initiative from friends and family in Delhi.

A total of 500 images were collected, individually labelled with the correct AQI value. I did so using the date and location of the photo, and the nearest AQI reading for that place and time.

My AI is able to distinguish between photos with AQI 0-50 and other buckets, as the former usually consists of bluer skies, and the latter consists of haze, smoke and significantly greyer skies. My model shows an 80% accuracy in identifying images in these buckets. False positives include blue skies being mistaken as grey skies due to the sun being overhead, too much light or over-exposure.

As the AQI increases over 400, the sky turns from grey to dark grey, orange and red. My AI is able to identify this dramatic shift, and shows a 75% accuracy in identifying these images.

We also built a different model that doesn't require the Internet. This model allows the users to simply take a photo of the sky. It extracts the RGB values of the sky from the photo, and then compares those values to a dataset of RGB values mapped to the AQI index. Using a simple k-Nearest-Neighbour classifier, I find the nearest match of the uploaded photo in my dataset. Then, I simply assign both the photos to the same AQI bucket. This simple model has shown very accuracy in my tests, getting an average of 85% accuracy.

Our goal with AVA is to allow people to understand the AQI around them, and enable them to make informed choices. In India, the rapid growth in smartphone users allows me to provide more and more people with access to basic public health information, at a click of their finger. I strongly believe that my approach outperforms the traditional AQI sensors, which although are very accurate, are also prohibitively expensive.

We deployed AVA globally, in Fremont, California (where wildfires led to AQI going over 100), to industrial belts in Uttar Pradesh and West Bengal in India. The response was surprisingly positive, with my users describing how useful the guidance was. My users also immediately started asking for more guidance, including the phone numbers for local politicians. In one of my villages, people began to lobby the local politician to shut down the factories during early mornings, so parents could drop their kids off at school and go on their morning walks.

It is important to note that air pollution is not just an Indian problem. It is a global problem. Countries like China, Bangladesh, Pakistan, and many others experience similar problems with bad air quality. There has recently been around 91% of the world's population living in areas where air quality levels exceed WHO limits [6]. People need access to accurate sensors, or millions of people will continue to die every year to air pollution, when the number of deaths can be limited if people simply had access to more public health guidance.

2 PRIOR RESEARCH

My research at AVA is built on a decade of prior work, in researching the effects of air quality, the importance of AQI sensors and predicting AQI from satellite imagery. In recent years, there has been an effort to understand the AQI in an area from alternative sources, including Twitter data and photos of the sky taken by GoPros. My work is a natural continuation of this work.

In 2015, Jiang et al. used tweets on Sina Weibo (Chinese Twitter) to detect outdoor air pollution and monitor Air Quality Index [7]. While this is a novel approach, most of my rural participants don't use Twitter or Instagram. Duong et. al [8] use a multi-source machine learning approach to approximate the local AQI scores at users' location in a big city, including geographical data, sensor data as well as the public weather data. While such an approach might work in major Indian cities (where this data is available), most small cities and villages in India don't have any such sensors. Similarly, in "Environmental Health Monitoring with Smartphone Application", Satoto et al. build sensors that can be attached to smartphones to enable AQI monitoring [9]. While this approach is indeed innovative, I am afraid that it would be hard to distribute these sensors to millions of people in rural India.

We were deeply inspired by the work done by Khanna et. al in "Air Cognizer: Predicting Air Quality with TensorFlow Lite" [10] which extracts features from user-uploaded images, and matches this information with other features such as the position of the sun, geographic information and weather conditions, to predict the PM2.5 index. This approach matches my vision the most.

Similarly, AVA is able to extract AQI information from a simple photo/selfie taken against a polluted / clear sky. This allows me to serve millions of people, who currently don't have access to this information. I do not expect that AVA will replace precise AQI sensors, however, its simplicity and smartphone readiness can help promote air pollution awareness, and alert people with serious respiratory diseases to stay away from suspected polluted air.

3 METHODOLOGY: DATA COLLECTION

The first step of my work was to build a dataset of polluted sky images. I used images on Instagram and Twitter that were tagged with popular hashtags such as #smogselfies, #delhismog, #smog among others. I also obtained photos from news reports on smog, air pollution. Finally, I crowdsourced the collection of another hundred photos from friends living in Delhi, who provided me with photos of polluted skies from their camera rolls.

These photos were, then, manually tagged with the observed AQI at the location and date tagged in the photo. I obtained historical AQI data from online resources such as aqicn.org and data.gov.in. I divided my photos into 5 AQI buckets of air quality values, ranging from 0-50, 100-200, 200-400, and 400 and above. I labelled and collected a minimum of 100 images per bucket, leading to an image corpus of 500 images of polluted skies, and their corresponding air quality indices.

3.1 Model #1 (Convolutional Neural Networks)

This corpus of 500 images was further divided into training, testing and validation sets (split in 80-10-10 ratio). My first AI, built using Google Cloud's Vision tools, used a convolutional neural network to label the test images. I noticed the following trends:

The AI is able to differentiate images in the 0-50 bucket from the other AQI buckets. However, sometimes, it mistakes an image in the 0-50 bucket as an image in the 50-200 bucket. On closer inspection, these images are taken with the sun being overhead, leading to the perceived sky color being grey, instead of blue. I need to find a way to better represent such images in my dataset, so my AI can better differentiate between overhead-sun grey skies and polluted grey skies.

The AI is unable to properly differentiate images in the 200-400 and 400 and above buckets. There could be several reasons for this. Both buckets have very polluted sky images, with colours ranging from brown, red and dark grey. However, orange sky images, for example, occur in both buckets. Polluted skies can appear brown, and yet, the AQI can be both in the 200-400 range or the 400+ range. AQI depends on several factors, outside of just the perceived colour of the sky. In fact, the majority of the time, my AI confused test images belonging in 200-400 with 400+, and test images belonging in 400+ with 200-400. Unfortunately, I don't have access to data such as wind speed, temperatures etc. in most of my villages. There is another way to fix this, including reducing the size of the AQI buckets. Most images being confused here have an AQI of 350 and above, making them very close to the AQI 400 and above bucket.

Our AI is able to separate images in the 50-200 bucket from other buckets. That's because slight air pollution looks very different from air pollution with an AQI of 200+. This is especially promising for me, as my primary goal with AVA is to provide the correct public health guidance. The public health guidance for AQIs less than 200 is different from AQIs above 200. Since my AI is able to differentiate between the two relatively well, I can use it to provide basic guidance to people in rural India.

True Label \ Predicted Label	0_50	200_400	50_200	400andabove
0_50	80%	-	20%	-
200_400	-	29%	14%	57%
50_200	-	25%	75%	-
400andabove	-	71%	-	29%

Figure 1: Confusion Matrix for my first model

3.2 Model #2 (Convolutional Neural Networks)

After trying the first model, I opted for a different, much simpler idea with my second model. I used k-Nearest Neighbor to find the nearest neighbor of the user-uploaded photo, within my dataset. Then, I simply assign the user-uploaded photo the same AQI bucket as the nearest neighbor. This simple approach can provide me with a surprising 92% accuracy.

I started by grabbing the color palette from every single image in my dataset using JavaScript [11]. I built a database mapping RGB values of the sky to the assigned AQI buckets. I added more images of overhead-sun skies to allow my model to differentiate between overhead-sun sky images and grey-polluted sky images. I also modified the AQI buckets to 3 buckets, 0-100, 100-300, and 300 and above. This also allows the model to perform better.

When a user uploads a photo, my second model simply looks at the dominant colors in the photo, finds the nearest neighbor in the database (calculated using the RGB values), and assigns the user-uploaded image the same AQI bucket as the nearest neighbor.

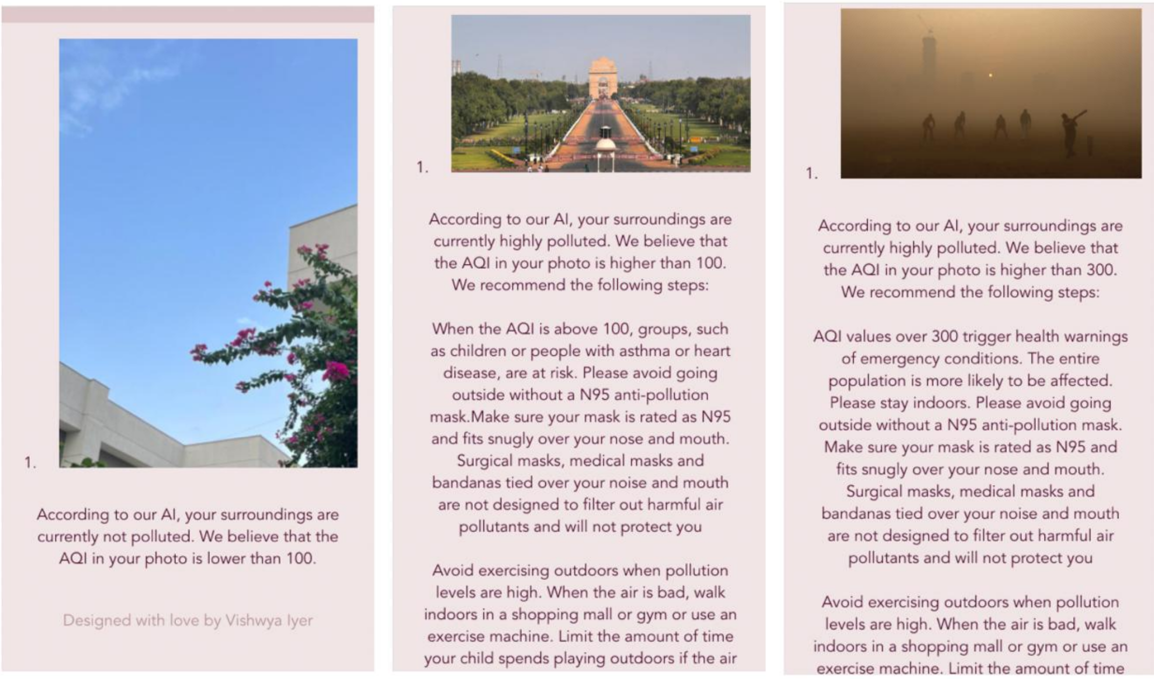

Figure 2: App screenshots, showing the different public health guidance for different AQIs

For example, the public health guidance for images with AQIs above 300 is shown below:

According to our AI, your surroundings are currently highly polluted. We believe that the AQI in your photo is higher than 300. AQI values over 300 trigger health warnings of emergency conditions. The entire population is more likely to be affected. Please stay indoors. Please avoid going outside without a N95 anti-pollution mask. Make sure your mask is rated as N95 and fits snugly over your nose and mouth. Surgical masks, medical masks and bandanas tied over your nose and mouth are not designed to filter out harmful air pollutants and will not protect you. Avoid exercising outdoors when pollution levels are high. When the air is bad, walk indoors in a shopping mall or gym or use an exercise machine. Limit the amount of time your child spends playing outdoors if the air quality is unhealthy. Please feel free to follow this helpful guide (links to a web.md article) to keep AQI low indoors. We have also curated this video (shown in app) which describes preventive measures you can take.

4 MY PILOTS

Even though AVA was designed for cities and villages in India, I started the first pilot with 15 students in California. As discussed earlier, pollution is a global problem, and during the annual wildfire season, AQI values in Fremont, California can routinely go above 100. I started distributing AVA to local community members during the wildfire season of 2021. AVA was surprisingly accurate, in providing the correct AQI value, and my users found the app to be extremely simple to use.

For the second pilot, AVA was distributed to villages in industrial belts in Uttar Pradesh, West Bengal and Bihar. I partnered with prominent local non-profits to distribute the app, and translated the app into local languages to enable my users to use it easily.

Since deployment, Ava has been used by over 5,000 unique users. Users have uploaded over 15,000 different images and received the relevant public health guidance. During the second pilot, I observed the following feedback:

4.1 People made changes to their daily schedules, due to AVA

I was surprised to see how many people changed their daily schedules, due to AVA providing them their AQI information.

"I always thought that the AQI would be low in the morning, but, due to the factory in our area starting at 8 am, the mornings are very polluted. We are going to sarpanch (village head), to ask them to delay the factory times, so we can enjoy clean mornings"

: P17, 34, male, store owner

Factories cause poor air in villages in the morning, so people decided to ask that the factory timings be delayed.

"I tried to check when the AQI was lowest. In the mornings, it was the highest. Someone said it's because all the city trucks leave at night, and their fumes are in the air in the morning. Around 5-6 pm, AQI is low. So, I go for my walk then now, instead of the morning"

:P2, 45, male, business owner

This person decided to leave their house at a different time for their daily walk as there was less pollution in the evening.

4.2 People immediately wanted more information, and ways to protect themselves

"We used the application to see the air quality in our village, and it said the AQI was above 300. We tried to follow the advice it gave us, but finding masks in my area is difficult."

: P12, man, 47, teacher

People found out that the air quality was bad, but could not do much to protect themselves. N-95 masks aren't easily available in rural areas. In the future, I would like to partner with non-profit organizations to give my users free N-95 masks.

"I wish there was more guidance on what we can do. It is hard to not leave the house, as we have to work outside. The app should help us get jobs we can do from home, when the air outside is polluted"

:P5, 19, male, student

Despite knowing that the air quality is bad, people still have to leave the house to make a living.

4.3 People were surprised at how polluted their environments were

"It was surprising to see how AQI affects me. Everyone in the village uses chulha (a traditional stove), and there's a lot of smoke. I think it is better to cook while wearing a mask, as pollution makes it very difficult to breathe"

: P10, 21, female, homemaker

This person was not aware about how bad the air in her village was, and decided to make a change based on this revelation.

"I really enjoyed the app. One day, it rained and the sky was blue, after a long time. I tried it (the app) then, and it said that the air was clean. I wish the sky was like that, every day. We should all take efforts to reduce pollution. The app should also have anti-pollution measures we can take"

: P11, 18, female, student

This person used the app and found that the air was clean that day, but believes that the app should give information about how to reduce pollution.

5 DISCUSSION AND CONCLUSION

AVA hopes to continue on the work that has already been done, and become a better and a more accurate product. Currently, the AQI buckets used in the program are quite large, so specific AQI values cannot be found. In the future, I want to reduce the size of the buckets to at least ranges of 10. This allows people to have more accurate data and make better decisions in relation to air quality.

I also want to spread AVA to more places. Right now, only people in certain parts of India and California are using the app. To start, I want to spread AVA all over India. This will require translations into more languages, and contacting villages or local governments across the country. After that is done, I want to spread AVA to as many people as possible, giving them access to an easy to use method of getting an AQI value. This will be challenging, but can be done with time and effort.

Along with making the AQI estimates more accurate, I would also like to have more accurate and more localized public health guidance. The steps that are necessary to stay safe often depend on location, so it is important to take this into account. Once again, this information can be gained working with localized governments. In addition, the guidance given currently is vague, and gives people general advice. More specific steps would allow people to stay safer.

Another problem is that people exposed to these poor air quality conditions do not have access to resources required to protect themselves. Masks are tools that are often used to protect people from breathing in poor air. Working with locals, I hope to spread resources like masks to those struggling as a result of air pollution. I also want to make connections with politicians across the world to raise awareness and lobby for the fight against air pollution. Everyone talks about how bad air pollution is, but not many people actually do something about it.

In the future, AVA hopes to allow people to live their lives normally despite poor air. Many people work jobs which require them to be doing outdoor activities. These jobs are often necessary to put money on the table for themselves and their family.

Using AVA, I hope that people will be able to find jobs which they can perform from home, until the air quality gets better. They will not have to risk their lives for a livelihood.

I believe that my work proves the possible use of a free, AI- platform to help people identify the AQI in their environments. With the right AQI, I can provide the right public health guidance to the hundreds of millions of people who live in polluted environments. I strongly encourage the creation of more technologies, like AVA, that democratise important information to people everywhere.

REFERENCES

[1] 1.7 million deaths in India were attributable to air pollution in 2019, says study. 2021. The Hindu. https://www.thehindu.com/news/national/17-million-deaths-in-india-were-attributable-to-air-pollution-in-2019-says-study/article33394600.ece.

[2] One In Three Children In Delhi Has Impaired Lungs, India Had 35 Million Asthma Patients In 2016. 2021. Moneycontrol. https://www.moneycontrol.com/news/trends/health-trends/one-in-three-children-in-delhi-has-impaired-lungs-india-had-35-million-asthma-patients-in-2016-2449013.html.

[3] Home | State of Global Air. 2021. Stateofglobalair.org. https://www.stateofglobalair.org/.

[4] India – Ambient Air Monitoring Data. 2021. Urbanemissions.info. https://urbanemissions.info/india-air-quality/india-ambient-monitoring-data/.

[5] Digital India Technology to transform a connected nation. 2021. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Digital%20India%20Technology%20to%20transform%20a%20connected%20nation/MGI-Digital-India-Report-April-2019.pdf.

[6] 92% of the world's population exposed to unsafe levels of air pollution. 2021. https://www.sciencedaily.com/releases/2016/09/160927144248.htm#:~:text=92%25%20of%20the%20world's%20population%20exposed%20to%20unsafe%20levels%20of%20air%20pollution,-Date%3A%20September%2027&text=Summary%3A,exposure%20to%20outdoor%20air%20pollution.

[7] Jiang W, Wang Y, Tsou MH, Fu X (2015) Using social media to Detect Outdoor Air Pollution and Monitor Air Quality Index (AQI): A Geo-Targeted Spatiotemporal Analysis Framework with Sina Weibo (Chinese Twitter). PLOS ONE 10(10): e0141185. https://doi.org/10.1371/journal.pone.0141185

[8] D. Q. Duong et al., "Multi-source Machine Learning for AQI Estimation," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 4567-4576, doi: 10.1109/BigData50022.2020.9378322.

[9] K. I. Satoto, E. D. Widianto and Sumardi, "Environmental Health Monitoring with Smartphone Application," 2018 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2018, pp. 281-286, doi: 10.1109/ICITACEE.2018.8576902.

[10] Air Cognizer: Predicting Air Quality with TensorFlow Lite. 2021. Blog.tensorflow.org. https://blog.tensorflow.org/2019/02/air-cognizer-predicting-air-quality.html.

[11] Dhakar, L. 2021. Color Thief. Lokeshdhakar.com. https://lokeshdhakar.com/projects/color-thief/.