Computational Tools for Understanding Air Pollution

Rishiraj Adhikary
rishiraj.a@iitgn.ac.in
Indian Institute of Technology (IIT)
Gandhinagar, Gujarat

Nipun Batra
nipun.batra@iitgn.ac.in
Indian Institute of Technology (IIT)
Gandhinagar, Gujarat

BACKGROUND

Ambient fine particulate ($PM_{2.5}$) is the most significant risk factor for premature death, shortening life expectancy at birth by 1.5 to 1.9 years [2]. 91% of the world's population lives in areas where air pollution exceeds safety limits[1]. 99% of the people in countries like India, Pakistan, Nepal, and Bangladesh experience ambient exposures of $PM_{2.5}$ exceeding 75 $\mu g/m^3$ to 100 $\mu g/m^3$ [3]. My Ph.D. thesis will be on understanding the perception of air pollution among people using social media data. I also intend to develop a wearable air pollution exposure monitor and design an air pollution visualisation tool to reduce the entry barrier for air pollution research.

CCS CONCEPTS

• **Human-centered computing** → *User studies.*

KEYWORDS

air pollution wearable; visualisation; air pollution perception; social media

ACM Reference Format:
Rishiraj Adhikary and Nipun Batra. 2020. Computational Tools for Understanding Air Pollution . In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers (UbiComp/ISWC '20 Adjunct), September 12–16, 2020, Virtual Event, Mexico.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3410530.3414327

1 PROBLEM STATEMENTS

1.1 Scalable Air Quality Perception

Motivation:
People are at risk if they do not perceive the threat of air pollution correctly. The efficacy of air pollution mitigation strategies depends on public participation and response. Both public participation and response are highly dependent on the public's perception of air pollution. Existing techniques, like questionnaire-based surveys, are not scalable to understand air pollution perception. Understanding the public perception can help policymakers engage with the public, appropriately educate and introduce mitigation strategies, and estimate the odds of success for a mitigation strategy.

Problem Statement: Scalable sensing of public air quality perception using NLP and social media.
Our goal is to measure the public sentiment towards the untested mitigation strategies by classifying the relevant posts into three classes: i) positive ii) negative, and iii) neutral. By positive, we mean positive towards untested mitigation strategies and likewise.
Events like policy announcement bring about an episodic concern among people about air pollution. Episodic discussions are ignored as they are short-lived. We investigate the causal relationship between air pollution levels and social media discussion volume.

1.2 Wearable for Air Quality and Health Sensing

Motivation:
Air pollution is the cause and aggravating factor of many respiratory diseases like chronic obstructive pulmonary disease (COPD) [11, 16], asthma [15, 16], and lung cancer [21, 24]. Ambient and non-ambient air pollution are not highly correlated [10, 26] and thus, it is essential to measure an individual's exposure to air pollution. Current wearables on air pollution monitoring are not continuous and measure some components of Volatile Organic Compound (VOC) and Ozone but not particulate matter ($PM_{2.5}$) [18].Thus monitoring exposure air pollution and lung health is of utmost importance. Primary health conditions of *blue-collar* workers can be improved using continuous monitoring of $PM_{2.5}$ using wearable exposure monitor.
A spirometer is a device that measures lung function, but given its accessibility challenge, its adaptation outside the clinical setting is limited. Retrofitting lung sensing capabilities on ubiquitous wearables can eliminate this effort.

Problem Statement:Design a wearable air pollution exposure monitor and smart mask.
The wearable air pollution monitor should have the following attributes, i) Ease of use with minimal user intervention, ii) low power or harvested energy as a power source, iii) lightweight, iv) monitors $PM_{2.5}$, Volatile organic compound, temperature and humidity. The objective of the smart mask is to monitor lung parameter using a microphone and act as a proxy for spirometry test [17, 19].

1.3 Tools for Visualisation and Eco-Feedback

Motivation:
Visualising air quality and related parameters is often the first step towards developing solutions, advancing research and engaging stakeholders. A few visualisations tools like OpenAir [6], arcGIS [14] and QGIS [23] are currently used by the community. But

[1] https://www.who.int/health-topics/air-pollution

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
UbiComp/ISWC '20 Adjunct, September 12–16, 2020, Virtual Event, Mexico
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-8076-8/20/09…$15.00
https://doi.org/10.1145/3410530.3414327

existing tools pose a high entry barrier for new researchers and enthusiasts in the domain of air pollution.
Air pollution is not a visible phenomenon. Most people lack awareness about how their everyday behaviours such as driving to work contributes to air pollution and exposes them to poor air quality. There is a need for persuasive technology intervention using existing sensing system and interactive display to feedback data to the users.

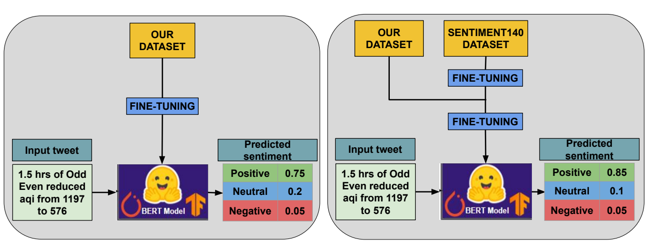
Problem Statement: Develop tools for air pollution visualisation and eco-feedback
We propose to design an air quality visualisation toolkit that addresses the challenges of current visualisation tools followed by interactive displays and mobile application for eco-feedback exclusively on air pollution

2 METHODOLOGY

2.1 Scalable Air Quality Perception

To gauge public perception and investigate the relationship between pollution level and social media discussion volume, we curate two datasets. i) The air pollution ($PM_{2.5}$) dataset and ii) The tweet dataset consisting of 7.2M tweet out of more than 40 keywords on air pollution.

Sentiment Analysis: We require an accurate and diverse language model to represent the semantics of the tweet text. BERT (Bidirectional Encoder Representation from Transformers) [9] is a state-of-the-art language model. BERT is trained on Wikipedia data and not on social media data. Thus, it is essential to contextualise the embeddings to the domain.
We use our tweet dataset to fine-tune the BERT layers, as described in Figure 1(a). During the process, all the attention layers try to learn a representation with a specific context. In the end, we have a fully connected softmax layer to output the prediction probability of each of the three classes (positive, negative and neutral). We learn Twitter-specific embeddings by fine-tuning on a subset of 'sentiment140' dataset [12]. Next, we fine-tune the model for the specific context with our tweet dataset, as described in Figure 1(b).
Topic Modeling: An important aspect of air pollution is to understand i) what topics of air pollution does each tweet represent and ii) how do the most common topics evolve.It is practically impossible to look into each tweet and assign a topic label due to the dataset size. We apply an unsupervised machine learning technique known as topic modelling [5] to determine the topics for each tweet.



(a) (b)

Figure 1: Representation of: (a) BERT fine-tuned model; (b) BERT twice fine-tuned model.

2.2 Wearable for Air Quality and Health Sensing

We propose to design a wearable air pollution monitor to measure exposure air pollution and a smart mask to measure lung health.
Generic Wearable: The primary sensing parameter for air pollution exposure monitor is $PM_{2.5}$, volatile organic compound and humidity. These parameters are chosen as the target users are *blue-collar* workers who work in an extreme environment like a construction site. Construction sites involve welding and painting activities which emit particulate matter and organic compounds respectively as air pollutants [1, 25]. Wearables are meant to be low weight and low energy consuming device. We plan to harvest energy on a LiPo battery using $4V$, $100mA$ solar panels combined with energy harvesting circuit to power our wearables. The wearables will be connected via LoRa to an MQTT server [13] which will minimize human intervention in data collection.
Smart Mask: We propose to design a smart mask that would measure lung function like, forced vital capacity (FVC) and forced expiratory flow (FEC) to perform spirometry. Also, an IR sensor will be used to measure the temperature [27]. The smart mask's primary sensor will be a MEMS-based microphone which will record the sound of breathing. The audio signal will be processed to output FVC and FEC. An LED will indicate vital lung function, and the detailed signal will be transmitted to the smartphone using BLE. Figure 2 shows a representative smart mask.

Figure 2: A representative figure showing the retrofitted mask and the associated mobile application.

2.3 Tools for Visualisation and Eco-Feedback

Visualisation Toolkit: To discover the challenges of visualising air pollution, we conducted an hour-long formative interview on six personnel working on various aspect of air pollution. We had two policy-makers, three researchers and one visualisation journalist. While we are interviewing more people and preparing a web-based survey, we believe that the six interviewees represent a diverse sample in terms of i) domain expertise; ii) job profile; iii) programming background; iv) educational background; v) work country.
Our questions were categorised into different groups, such as programming and data challenges, tools used for air quality visualisations, impediments faced while creating visualisations etc.Our

future work would be primarily in two directions. First, to conduct more interviews to understand the pain points better, and correspondingly develop more functionality. Second, to conduct an exploratory first user study to understand the efficacy of our toolkit.

Eco Feedback: We will feedback the data collected from the wearable air pollution monitor to the user. A representative diagram of how the retrieved data will be visualised is shown in Figure 3. The evaluation of such a system will be under controlled environment and rely on self-reported behaviour changes by the user.

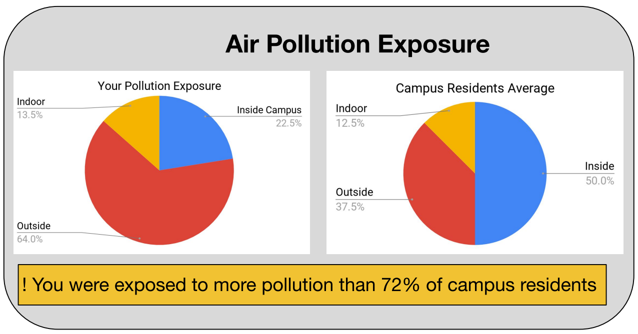


Figure 3: Social-pressure moves people. An eco-feedback display showing the contribution to air pollution from a household and comparing it with an immediate neighbour.

3 EXPERIMENTS AND RESULTS

3.1 Scalable Air Quality Perception

Experiment 1:

Governments across the world have proposed and implemented several strategies for reducing air pollution. These strategies include, but, are not limited to: i) using higher-grade fuel for vehicles ii) cutting emissions from power plants among other. Several studies [8, 20] suggest that these strategies will not help reduce air pollution. One such highly debated strategy is installing outdoor air purifiers called "Smog Tower" [7]. In Delhi, India, a "Smog Tower" was inaugurated in the first week of January 2020. However, a think-tank suggests we need 2.5 million of such towers given the air shed of a city like Delhi[2]. Another mitigation measure implemented in Delhi (India) is a vehicle rationing scheme called "Odd-Even". A study [8] suggest that "Odd-Even" may have been ineffective. We choose Delhi as our testbed because i) We have local expertise about air pollution in Delhi; ii) Delhi is often called the most polluted city in the world; iii) Delhi has a dense population, and thus potentially millions of people are at health risk; iv) Various mitigation strategies have been proposed so far in Delhi, and people discuss these strategies in social media.

We collected 600 tweets on "Smog Tower". Two investigators labelled 80% of these tweets either as a positive, neutral or negative sentiment. We used multiple state of the art baselines and the BERT [9] machine learning classifier fine-tuned on Sentiment140 dataset to classify the tweets.

[2]https://twitter.com/CEEWIndia/status/1229702994372919296

Result 1:

We achieved an F1 score of 0.71, which is comparable to other states of the art methods [4, 22]. The result of sentiment over time is shown in Figure 4. We observe from the plot that at any point in time, the positive sentiment associated with the "Smog Tower" is more than negative sentiment. Similarly, we got comparable results on "Odd Even" and abstain from mentioning it due to space constraints.

Experiment 2:

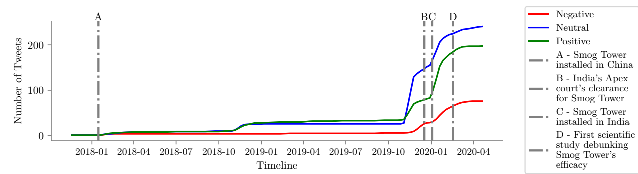


Figure 4: The figure shows the cumulative sum of positive, neutral and negative tweets for a particular mitigation strategy over time. The rate of increase in the number of positive and neutral tweets is greater than negative tweets.

We created the largest to-date dataset on Delhi air pollution containing 2.23 million unique tweets from the year 2016 to March 2020. This dataset is a subset of the original 7.2M tweet dataset on air pollution. We performed topic modeling on this data using Latent Dirichlet Allocation (LDA) [5] to mine topics of discussion.

Result 2:

We found twelve topics of discussion. Some of them are, i) **OddEven**: The vehicle rationing scheme in Delhi. ii) **Stubble Burning**: A significant cause of pollution in Delhi is due to stubble burning in nearby states. iii) **COPD**: Chronic obstructive pulmonary disease is a common lung disease as a result of air pollution. iv) **Diwali Crackers**: Cracker burning during the winter period deteriorates pollution levels. Other topics are shown in Figure 5. We also observed the topic of discussion among air pollution protagonists. Social media protagonist (with a considerable number of followers) talk sparingly about air pollution.

Experiment: 3

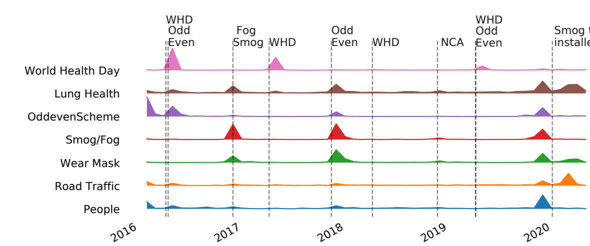


Figure 5: Topics (in Y-axis) evolution over time. Peaks represents discussion density. Vertical lines represent events, i)"Odd-Even", ii) "World Health Day (WHD)", iii)"Fog/Smog" iv) "National Climate Assessment (NCA)", and v) "Smog Tower Installation".

Having seen the sparsity in discussions about Delhi air pollution, we hypothesised that air quality is a year-long problem, but Twitter discussions are not. We verified it by visualising the quantity of air pollution and tweets over time.

Result 3:
Figure 6 demonstrates that air toxicity is indeed a year-long phenomenon in Delhi. Figure 7 shows the number of tweets and $PM_{2.5}$ values in the same plot. An interesting observation was the spike in air pollution-related tweets on June 5, 2018, i.e., World Environment Day. Similar comments apply for the timeline of odd-even vehicle restriction policy.

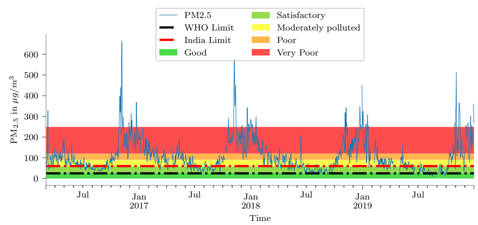


Figure 6: Yearly $PM_{2.5}$ Level In Delhi. Colours corresponds to Indian air quality standards. The WHO and India limit on $PM_{2.5}$ is also shown.

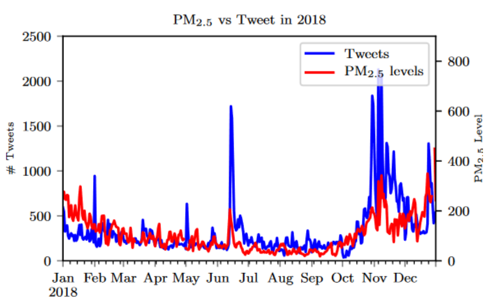


Figure 7: Twitter discussions on Delhi Air Pollution raised on June 5th during World Environment Day. It grew again in November, smog makes air pollution visible in Delhi.

3.2 Wearable for Air Quality and Health Sensing

Experiment 1: Design of the first prototype of wearable air pollution monitor.
Result 1:
Figure 8 shows the wearable air pollution monitor that measures $PM_{2.5}$, CO_2 and Volatile organic compound. The current prototype relies on an external battery for power and Bluetooth low energy for data transfer. Our future prototype will adhere to the attributes as mentioned in Section 2.

Experiment 2: Air Pollution Monitoring in Kitchen Mess, Construction Site and mobile monitoring.
Result 2:
A monitor placed at the dormitory kitchen (catering to more than 2000 students) reported $PM_{2.5}$ value between 350 $\mu g/m^3$ to 400 $\mu g/m^3$ during cooking activity. We found that construction sites have $PM_{2.5}$ values between 400 $\mu g/m^3$-1000 $\mu g/m^3$. The concentration of CO_2 in indoor environment reaches a hazardous level of ≥1000

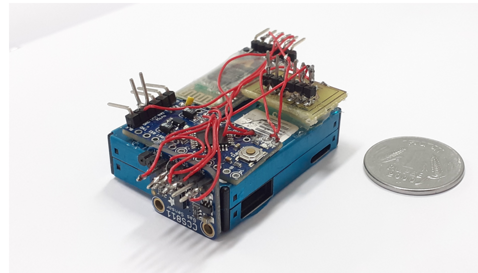


Figure 8: Prototype of our wearable air quality monitor

ppm. A sensor placed in a bus revealed that the $PM_{2.5}$ value outside the campus is at least two times higher than the inside. We observe that even in a highly localized environment, the exposure can vary significantly from the ambient concentration of air pollution. High temporal variance in exposure was observed when people are in transit, and air quality of meeting rooms deteriorate as the number of people increases, thus creating an unhealthy environment.
Experiment 3: Design of the prototype of a smart mask.
Result 3:
We designed a smark mask that uses a MEMS microphone to record the breathing sound of an individual. The raw audio is processed to calculate the forced vital capacity (FVC). The FVC value processed from the raw audio data was evaluated using a handheld spirometer and was found to be 100% accurate. We are currently working on measuring the amount of air an individual can force out of her lungs in one second (known as FEC or Forced Expiratory Volume). FEC, in combination with FVC, gives a quantitative measure of the health of the lung.

3.3 Tools for Visualisation and Eco-Feedback

We reflected on the challenges uncovered by our interviews to formulate a set of design goals for our *Vayu* prototype:

- **D1:** Minimise the extrinsic complexities of software installation/configuration and on-disk data file management.
- **D2:** Make visualisation accessible to the colour blind.
- **D3:** Provide automated data pre-processing pipelines.
- **D4:** Provide scaffolding for annotation and sharing of plots in the browser via URL.

Experiments on eco-feedback are left as future work.

4 CONCLUSION

Our work on social media is currently limited to Delhi, India. In the future, we plan to extend our analysis to multiple geographies. For our wearable air quality monitor, we plan to do a more exhaustive study and reduce the form factor and energy needs for our sensors to scale the study. Once we develop such a wearable, we would plan to conduct studies by measuring the exposure of different campus residents. We also plan to study the exposure of other kinds of blue-collar job workers. For the work on air quality visualisation, our future work would be primarily in two directions. First, to conduct more interviews to understand the pain points better, and correspondingly develop more functionality into the toolkit. Second, to conduct an exploratory first user study to understand the efficacy of the toolkit.

REFERENCES

[1] AA Adu, OJ Aderinola, and V Kusemiju. 2012. An assessment of soil heavy metal pollution by various allied artisans in automobile, welding workshop and petrol station in Lagos State, Nigeria. *Science Journal of Environmental Engineering Research* 2012 (2012).

[2] Joshua S Apte, Michael Brauer, Aaron J Cohen, Majid Ezzati, and C Arden Pope III. 2018. Ambient PM2. 5 reduces global and regional life expectancy. *Environmental Science & Technology Letters* 5, 9 (2018), 546–551.

[3] Joshua S Apte and Pallavi Pant. 2019. Toward cleaner air for a billion Indians. *Proceedings of the National Academy of Sciences* 116, 22 (2019), 10614–10616.

[4] Noureddine Azzouza, Karima Akli-Astouati, and Roliana Ibrahim. 2019. Twitter-BERT: Framework for Twitter Sentiment Analysis Based on Pre-trained Language Model Representations. In *International Conference of Reliable Information and Communication Technology*. Springer, 428–437.

[5] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[6] David C Carslaw and Karl Ropkins. 2012. Openair—an R package for air quality data analysis. *Environmental Modelling & Software* 27 (2012), 52–61.

[7] K Chandramouli, N Pannirselvam, D Vijaya Kumar, Sagar Reddy Avuthu, and V Anitha. 2019. A STUDY ON SMOG FILTERING TOWER. *Journal of Advanced Cement & Concrete Technology* 2, 1, 2 (2019).

[8] Sourangsu Chowdhury, Sagnik Dey, Sachchida Nand Tripathi, Gufran Beig, Amit Kumar Mishra, and Sumit Sharma. 2017. "Traffic intervention" policy fails to mitigate air pollution in megacity Delhi. *Environmental science & policy* 74 (2017), 8–13.

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[10] Stefanie T Ebelt, William E Wilson, and Michael Brauer. 2005. Exposure to ambient and nonambient components of particulate matter: a comparison of health effects. *Epidemiology* (2005), 396–405.

[11] Annunziata Faustini, Massimo Stafoggia, Paola Colais, Giovanna Berti, Luigi Bisanti, Ennio Cadum, Achille Cernigliaro, Sandra Mallone, Corrado Scarnato, and Francesco Forastiere. 2013. Air pollution and multiple acute respiratory outcomes. *European Respiratory Journal* 42, 2 (2013), 304–313.

[12] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford* 1, 12 (2009), 2009.

[13] Urs Hunkeler, Hong Linh Truong, and Andy Stanford-Clark. 2008. MQTT-S—A publish/subscribe protocol for Wireless Sensor Networks. In *2008 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE'08)*. IEEE, 791–798.

[14] Kevin Johnston, Jay M Ver Hoef, Konstantin Krivoruchko, and Neil Lucas. 2001. *Using ArcGIS geostatistical analyst*. Vol. 380. Esri Redlands.

[15] Anna Karakatsani, Antonis Analitis, Dimitra Perifanou, Jon G Ayres, Roy M Harrison, Anastasia Kotronarou, Ilias G Kavouras, Juha Pekkanen, Kaarle Hämeri, Gerard PA Kos, et al. 2012. Particulate matter air pollution and respiratory symptoms in individuals having either asthma or chronic obstructive pulmonary disease: a European multicentre panel study. *Environmental Health* 11, 1 (2012), 75.

[16] FJ Kelly and JC Fussell. 2011. Air pollution and airway disease. *Clinical & Experimental Allergy* 41, 8 (2011), 1059–1071.

[17] Ronald J Knudson, Ronald C Slatin, Michael D Lebowitz, and Benjamin Burrows. 1976. The maximal expiratory flow-volume curve: normal standards, variability, and effects of age. *American Review of Respiratory Disease* 113, 5 (1976), 587–600.

[18] Balz Maag, Zimu Zhou, and Lothar Thiele. 2018. W-air: Enabling personal air pollution monitoring on wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–25.

[19] Martin R Miller, JATS Hankinson, V Brusasco, F Burgos, R Casaburi, A Coates, R Crapo, Pvd Enright, CPM Van Der Grinten, P Gustafsson, et al. 2005. Standardisation of spirometry. *European respiratory journal* 26, 2 (2005), 319–338.

[20] Dinesh Mohan, Geetam Tiwari, Rahul Goel, and Paranjyoti Lahkar. 2017. Evaluation of odd–even day traffic restriction experiments in Delhi, India. *Transportation Research Record* 2627, 1 (2017), 9–16.

[21] Sonja N Sax, Ke Zu, and Julie E Goodman. 2013. Air pollution and lung cancer in Europe. *The Lancet Oncology* 14, 11 (2013), e439–e440.

[22] Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter sentiment analysis with deep convolutional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 959–962.

[23] QGIS Delopment Team et al. 2016. QGIS geographic information system. *Open Source Geospatial Foundation Project* (2016).

[24] Paolo Vineis, Gerard Hoek, Michal Krzyzanowski, Federica Vigna-Taglianti, Fabrizio Veglia, Luisa Airoldi, Herman Autrup, Alison Dunning, Seymour Garte, Pierre Hainaut, et al. 2006. Air pollution and risk of lung cancer in a prospective study in Europe. *International journal of cancer* 119, 1 (2006), 169–174.

[25] G1 Wieslander, D Norbäck, E Björnsson, C Janson, and G Boman. 1996. Asthma and the indoor environment: the significance of emission of formaldehyde and volatile organic compounds from newly painted indoor surfaces. *International archives of occupational and environmental health* 69, 2 (1996), 115–124.

[26] William E Wilson, David T Mage, and Lester D Grant. 2000. Estimating separately personal exposure to ambient and nonambient particulate matter for epidemiology and risk assessment: why and how. *Journal of the Air & Waste Management Association* 50, 7 (2000), 1167–1183.

[27] KS Xiahou, XJ Zeng, LK Hu, GP Li, and RH Ye. 2011. Design of human-body temperature monitoring system based on MLX90614 and ZigBee. *Automation and instrumentation* 11 (2011), 23–26.

A BIOGRAPHICAL SKETCH

In August 2020, Rishiraj Adhikary[3] will start his second year as a Ph.D. student at the Department of Computer Science, Indian Institute of Technology (IIT) Gandhinagar, Gujarat, India. Dr Nipun Batra[4] at the same department, advises him. Given the early stage of PhD, participating in the doctoral colloquium at UbiComp 2020 will allow him to receive feedback on the proposed work. Rishiraj expects to graduate by August of 2024.

Rishiraj is actively interested in the interdisciplinary domain of Electronics and Computer Science to solve societal issues. In the first semester of Ph.D., he was involved in a large scale sensor deployment in the campus to measure various parameters like water flow, air pollution etc. In the second semester, he started to analyze social media data to gauge public perception in social media to help policymakers. Currently, he is also designing a solar-powered wearable for blue-collar workers to measure exposure air pollution. He solicits feedback from the Ubicomp community on the use of ubiquitous computing to solve societal problems.

[3]https://rishi-a.github.io/

[4]https://nipunbatra.github.io/