

Data Science

Lecture 1-2: Social Impact of Data Science



UNIVERSITY
OF AMSTERDAM

Lecturer: Yen-Chia Hsu

Date: Feb 2026

Data science **can create a social impact** to
influence society positively!

This lecture uses my previous projects to
show the social impact of data science.

Air Quality Monitoring System

Integrating smoke videos, sensor readings, and smell reports

<http://shenangochannel.org>

The screenshot shows the ACM Digital Library homepage with a search bar and navigation links. Below the search bar, a red banner displays conference-related links: Conference, Proceedings, Upcoming Events, Authors, Affiliations, and Award Winners. The main content area shows the search results for "Community-Empowered Air Quality Monitoring System".

Home > Conferences > CHI > Proceedings > CHI '17 > Community-Empowered Air Quality Monitoring System

RESEARCH-ARTICLE • ♀

Community-Empowered Air Quality Monitoring System



Authors: Yen-Chia Hsu, Paul Dille, Jennifer Cross, Beatrice Dias, Randy Sargent, Illah Nourbakhsh

[Authors Info & Affiliations](#)

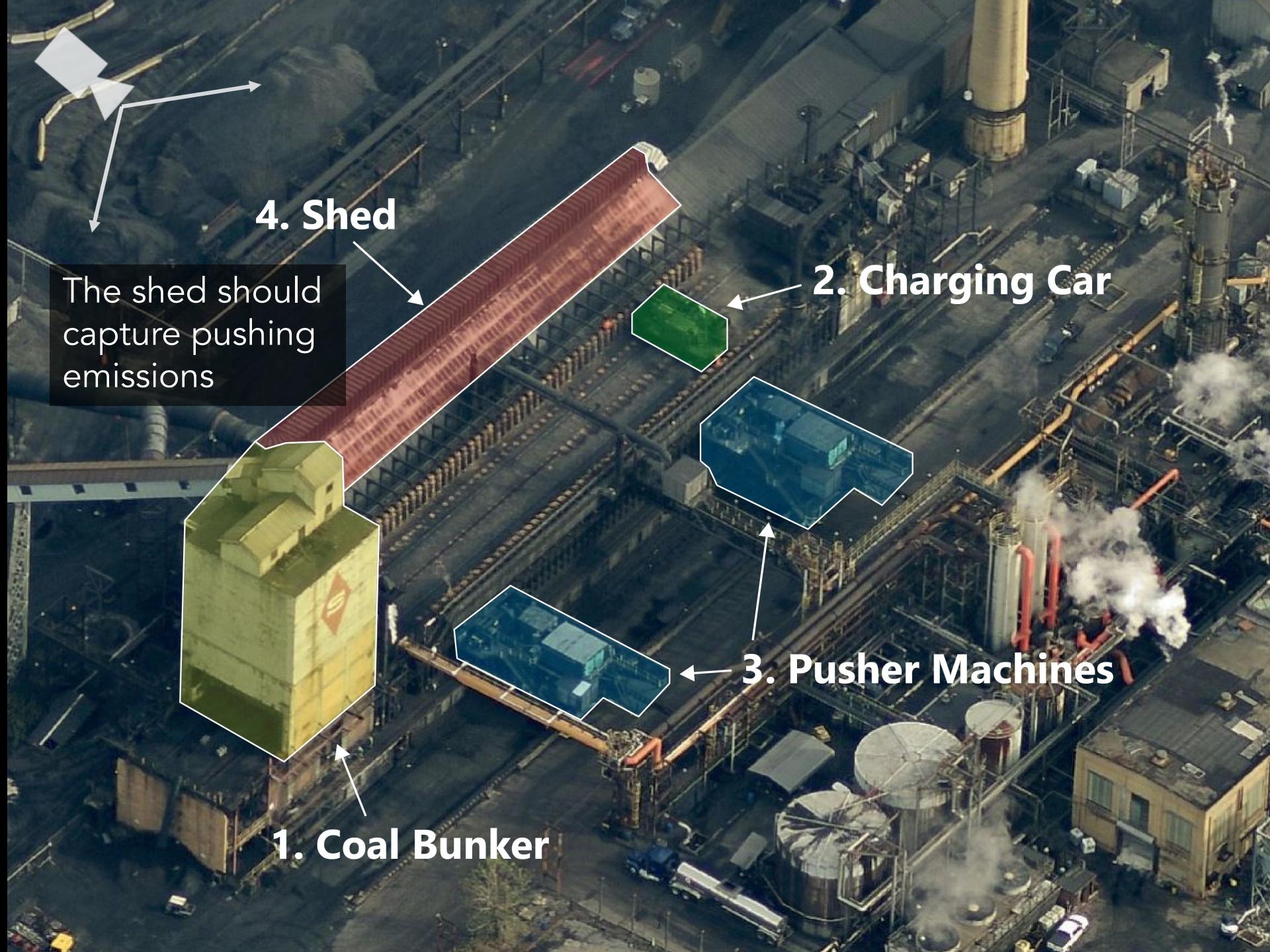
Publication: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems • May 2017

• Pages 1607–1619 • <https://doi.org/10.1145/3025453.3025853>

10 540

Get Access

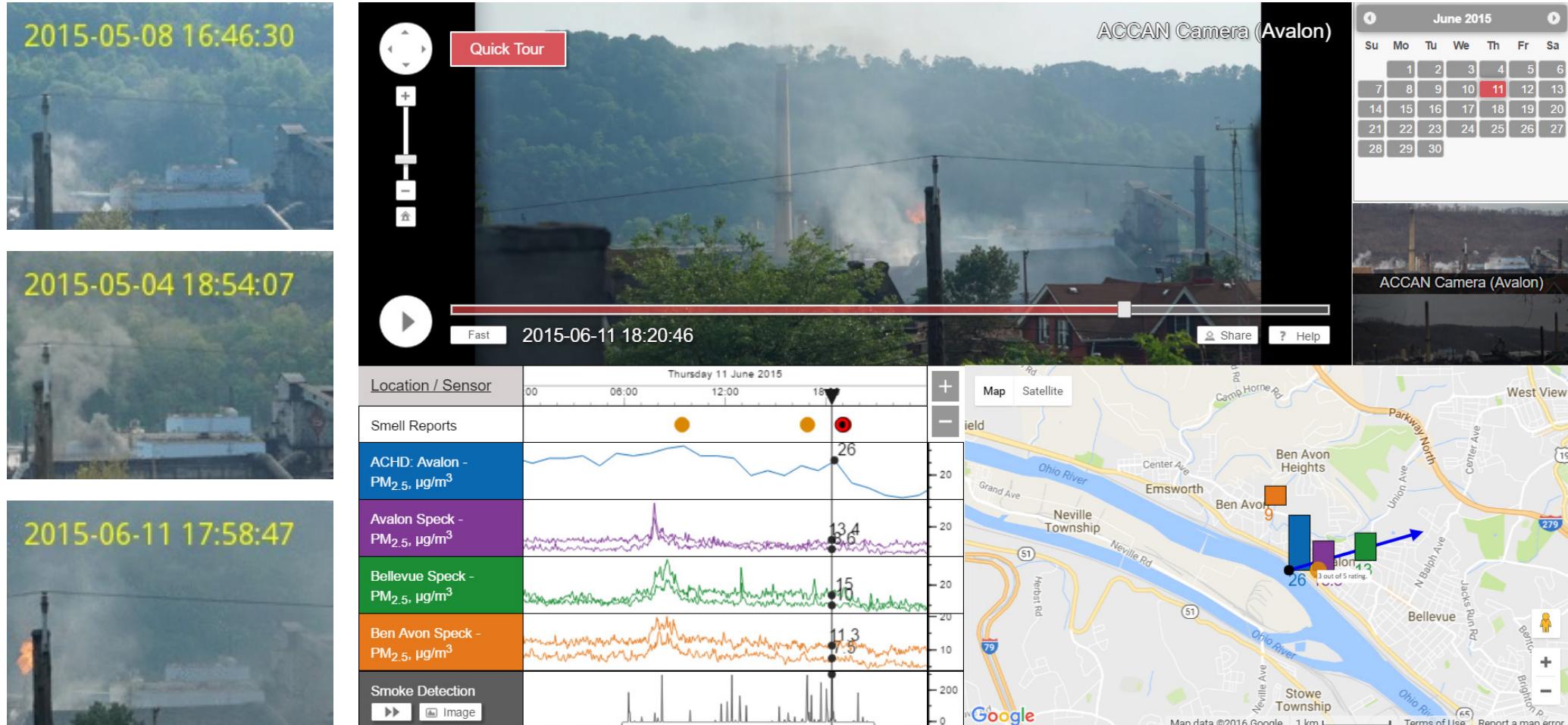






- Co-design to collect evidences of air pollution
- Present them to the government or media
- Raise the public awareness of air quality issues

Residents can integrate the evidence of air pollution and their personal experiences into a story to **influence regulators' attitudes** and **increase the community's confidence**.



“But what I see in the video,” the acting director of U.S. EPA Region III Air Protection Division said, referring to videos from the system that were projected on a screen at the front of the meeting room, “is totally unacceptable.”



2015-05-27 12:44:10



2015-10-11 17:40:42



2015-06-15 17:44:21



2015-06-11 17:54:00







Smell Pittsburgh

Crowdsourcing and visualizing pollution odors

<https://smellpgh.org>

The screenshot shows the ACM Digital Library homepage with a search bar and navigation links. Below it, a green banner lists journal categories like Journals, Magazines, Proceedings, Books, SIGs, Conferences, and People. A specific article is highlighted in the center:

Home > ACM Journals > ACM Transactions on Interactive Intelligent Systems > Vol. 10, No. 4 > **Smell Pittsburgh: Engaging Community Citizen Science for Air Quality**

RESEARCH-ARTICLE

Smell Pittsburgh: Engaging Community Citizen Science for Air Quality

Authors: Yen-Chia Hsu, Jennifer Cross, Paul Dille, Michael Tasota, Beatrice Dias, Randy Sargent, Ting-Hao (Kenneth) Huang, Illah Nourbakhsh [Authors Info & Affiliations](#)

Publication: ACM Transactions on Interactive Intelligent Systems • November 2020 • Article No.: 32
• <https://doi.org/10.1145/3369397>



REPORT AN AIR QUALITY COMPLAINT FORM

Type what you're looking for 

[More How Do I... >](#)

» Health Department > Programs > Air Quality > Report an Air Quality Complaint Form

Report an Air Quality Complaint

Use this form to send us a comment or to register a complaint with the Health Department's Air Quality Program.

Enforcement inspectors respond to every citizen complaint received via the complaint line (412-687-ACHD) or this form. Please remember to include your name and email address if you wish to receive a response. Comments or complaints cannot be acknowledged without an email address.

Please note: Be as specific as possible. When filing a complaint about open burning or foul odors, please include the time, location (neighborhood or zip code), and a brief description of the odor or smoke associated with your complaint.

An asterisk (*) denotes a required field. Name and email are suggested.

Air Quality Program Office:

301 39th St.

Building 7

Pittsburgh, PA 15201

 [Google Directions](#)

Name:

Email:

Subject:

*Time, Location, Nature of Complaint:

* Denotes a required field.

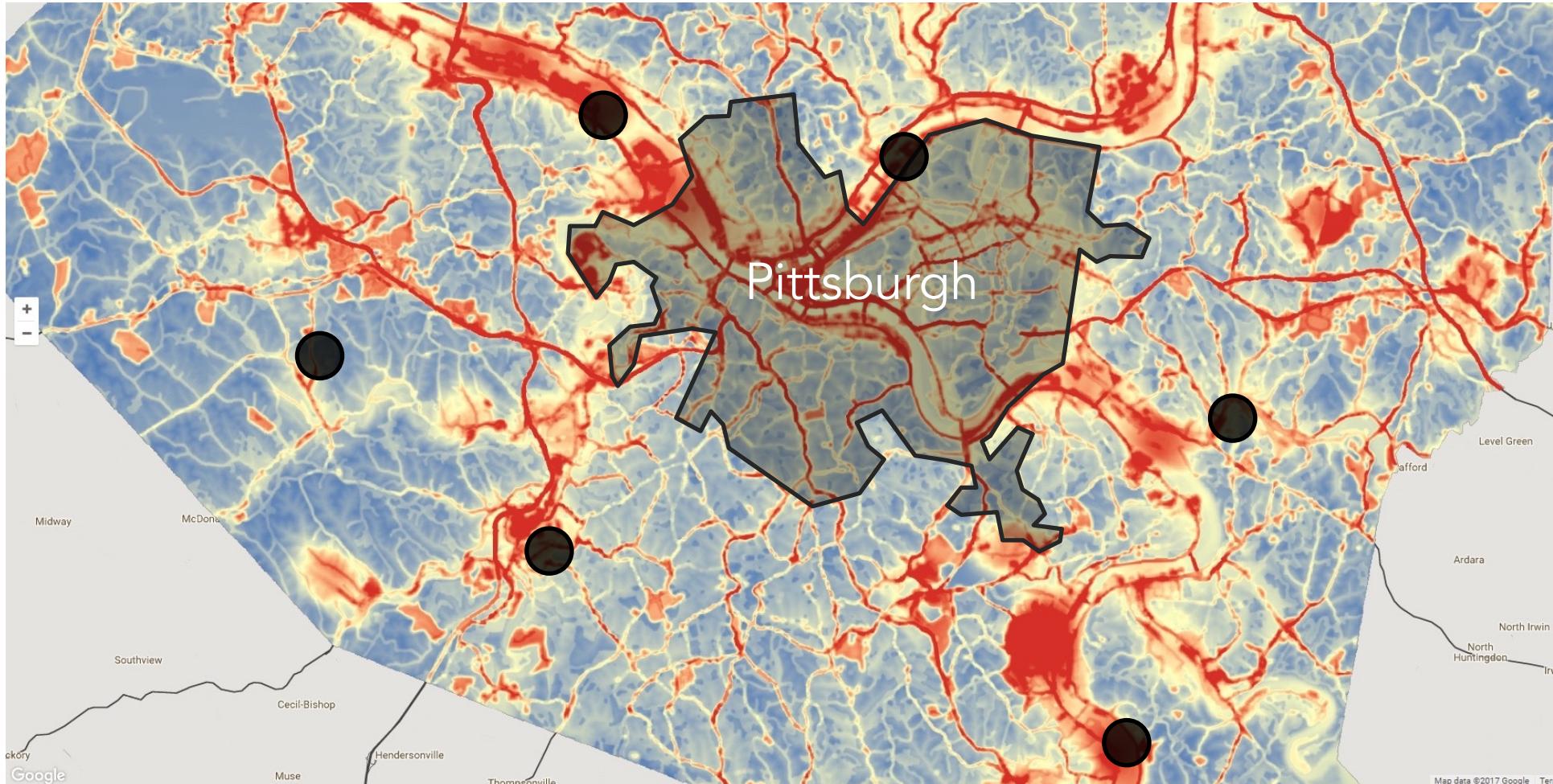
[Send Message](#)

The prior approach that asks citizens to report odor complaints post hoc via forms or phone calls suffers from:

- poor data quality
- non-transparency

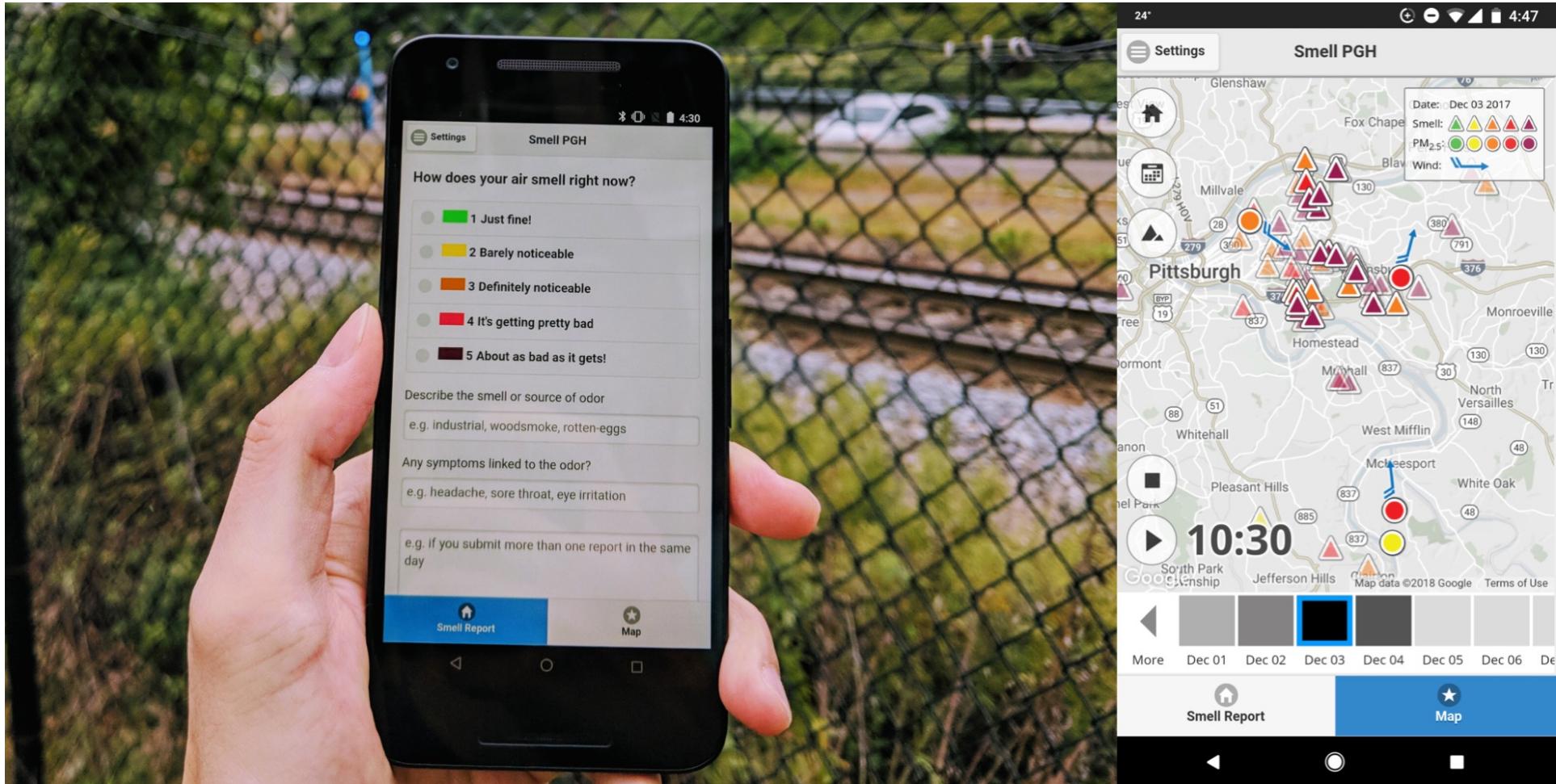
The odor reporting form -- <https://www.allegenycounty.us/Health-Department/Programs/Air-Quality/Report-an-Air-Quality-Complaint-Form.aspx>

How can we effectively collect the smell experiences on a **city-wide scale** with more than 300,000 residents and many pollution sources over many years?

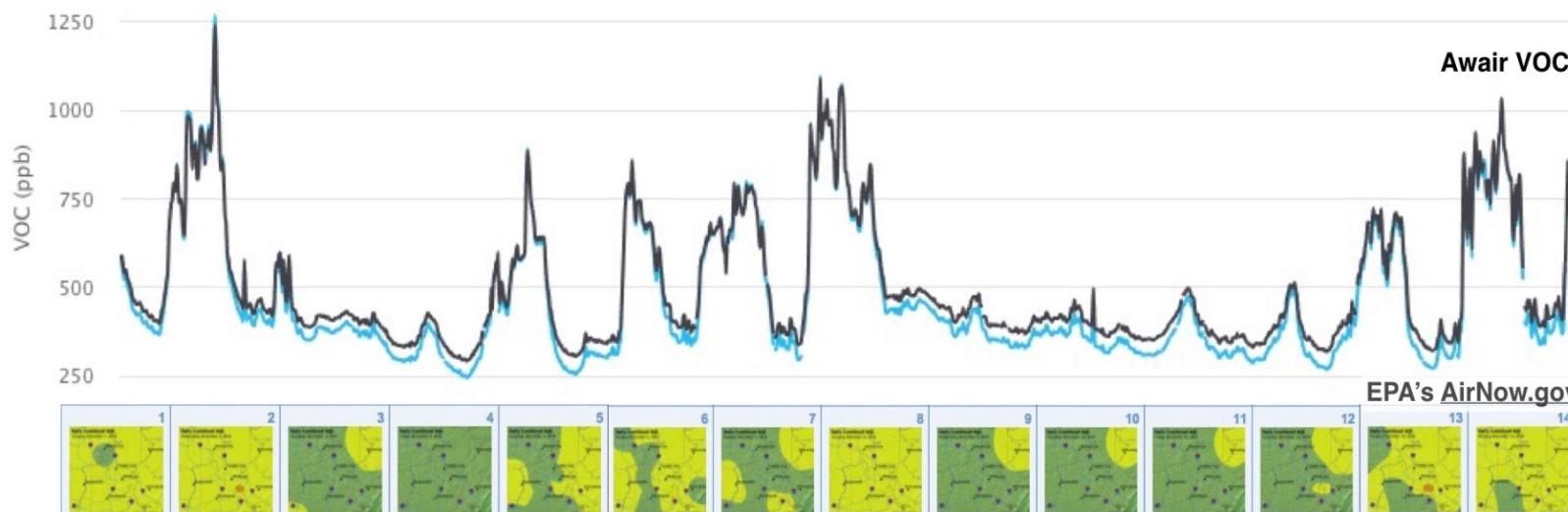
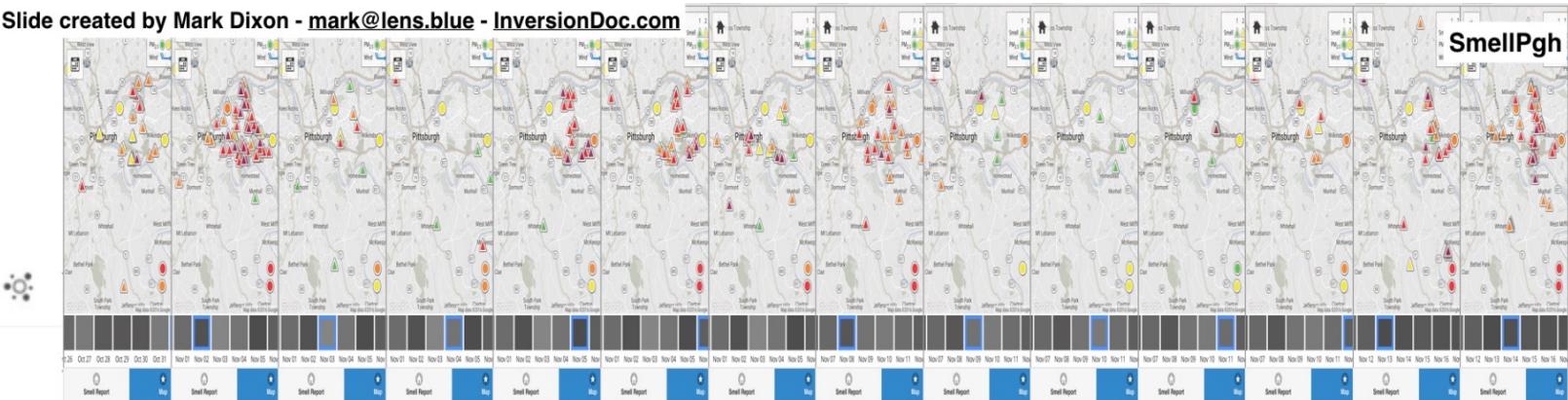


Pittsburgh pollution map -- <https://breathetheproject.org/pollution-map/>

Smell Pittsburgh enables communities to [collect data on a large scale](#). Also, visualizing multi-modal data in real-time can help communities understand local concerns.



Community members plot smell reports with self-operated VOC (volatile organic compounds) sensors to [find correlations](#) and inspect how pollution impacts them.



Decision-makers in the local health department mentioned that "Every aspect of the activity and operation of these coke plants will have a more stringent standard applied."

Monday, November 12, 2018, 12:09PM | 44°

Pittsburgh Post-Gazette

Obituaries | PGe | PG Store | Archives | Classifieds ▾

☰ MENU SUBSCRIBE LOGIN REGISTER Home | News | Local | Sports | Opinion | A&E | Life | Business | Contact Us [f](#) [t](#) [i](#) [Q](#) NEWSLETTERS



Air advocates read 'scroll of smells' at health board meeting

Photo credit Don Hopey

Table 1: Distribution of Smell Reports

Smell Rating	Description	2022	2021	2020	2019	2018	2017
1	Just fine!	333 (3.6%)	707 (5.7%)	1,565 (8.2%)	1,711 (9.5%)	1,199 (13.0%)	1,658 (20.4%)
2	Barely noticeable	287 (3.1%)	447 (3.6%)	921 (4.8%)	798 (4.4%)	497 (5.4%)	665 (8.2%)
3	Definitely noticeable	1,922 (20.8%)	2,902 (23.3%)	4,436 (23.3%)	4,305 (23.9%)	2,649 (28.8%)	2,246 (27.7%)
4	It's getting pretty bad	3,185 (34.5%)	4,258 (34.2%)	6,014 (31.6%)	5,805 (32.3%)	2,932 (31.9%)	2,171 (26.8%)
5	About as bad as it gets!	3,506 (38.0%)	4,126 (33.2%)	6,082 (32.0%)	5,358 (29.8%)	1,918 (20.9%)	1,372 (16.9%)
Sum		9,233	12,440	19,019	17,977	9,195	8,112

Table 2: User Engagement with The Smell PGH App

Number of Unique Users	2022	2021	2020	2019	2018	2017
Submitted Reports	1,242 (37.0%)	1,804 (41.5%)	2,688 (46.8%)	3,274 (50.9%)	1,769 (66.9%)	1,308 (58.4%)
Used the Map	3,054 (90.9%)	3,919 (90.2%)	5,227 (91.0%)	5,708 (88.8%)	2,248 (85.0%)	1,949 (87.0%)
Participated (N)	3,358	4,347	5,743	6,429	2,645	2,239

Is smell data useful in predicting local air pollution events and identifying patterns?

Smell Pittsburgh predicts upcoming smell events (based on the collected reports) and send push notifications to inform users and encourage engagement.



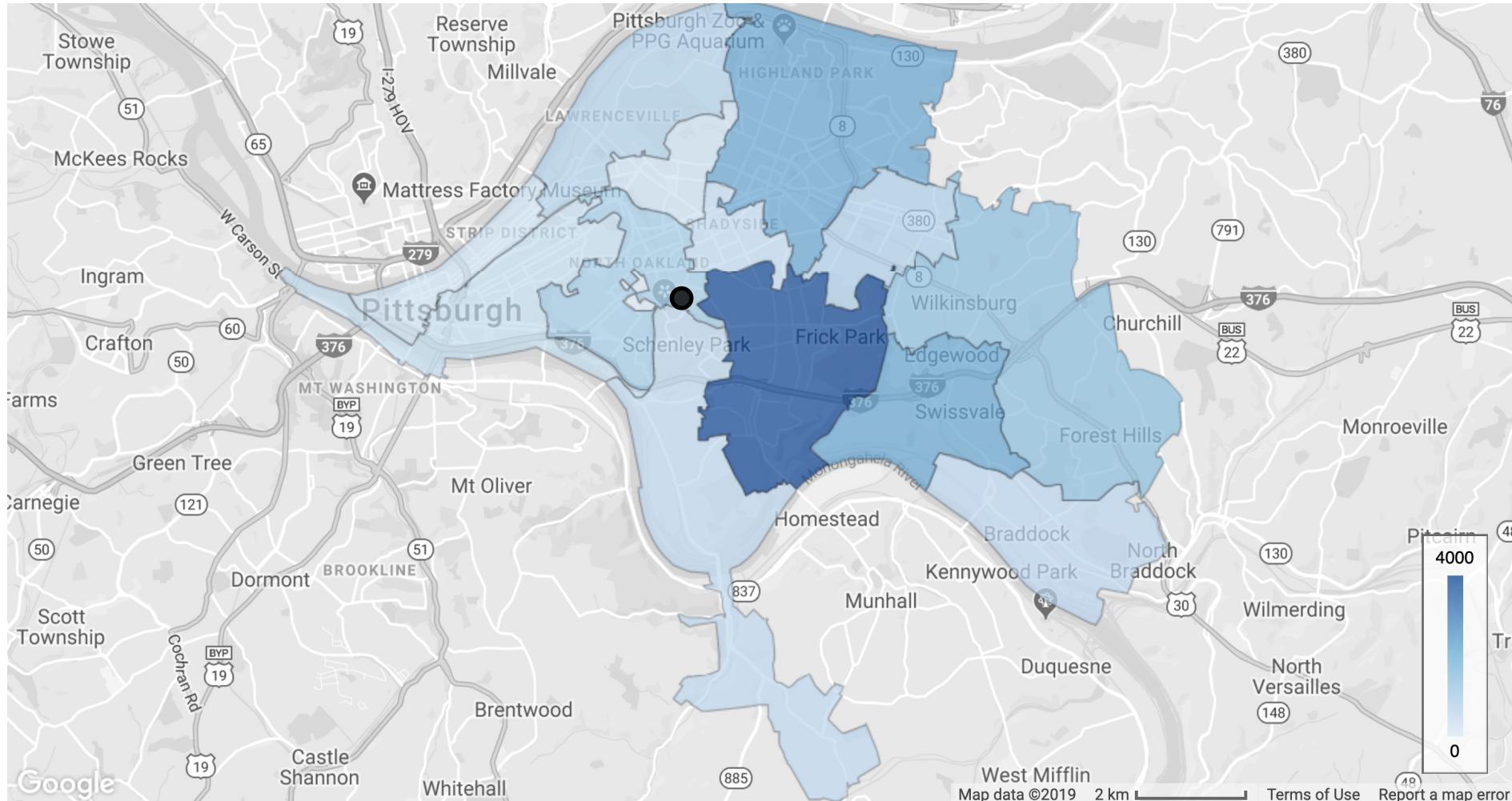
SMELL PGH

Smell Event Alert

Local weather and pollution data indicates there may be a Pittsburgh smell event in the next few hours.

Keep a nose out and report smells you notice!

A geographic region in Pittsburgh is manually selected when predicting the smell events. The black dot is Carnegie Mellon University.



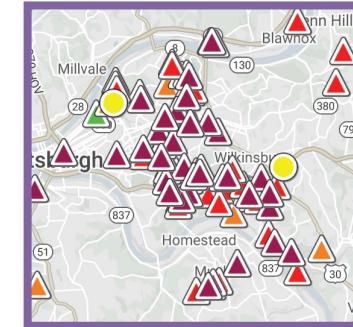
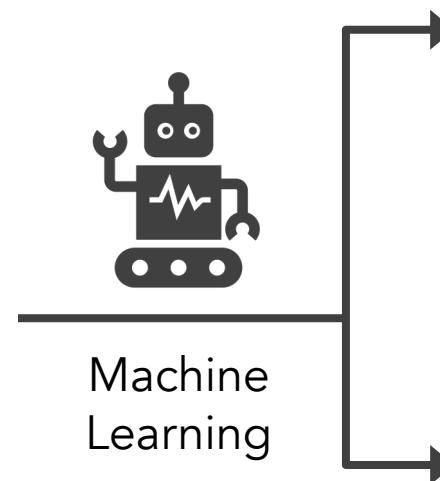
Number of smell reports aggregated by zip codes -- <https://smellpgh.org/analysis#figure9>

We use a Random Forest (a machine learning model) to predict smell events from air quality data (obtained from government-operated sensor stations).

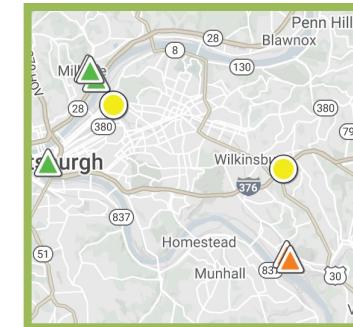
O ₃ : 26 ppb	CO: 127 ppb
H ₂ S: 0 ppb	PM _{2.5} : 9 µg/m ³
Wind: 17 deg	...
Observation 1	

O ₃ : 1 ppb	CO: 1038 ppb
H ₂ S: 9 ppb	PM _{2.5} : 23 µg/m ³
Wind: 213 deg	...
Observation 2	

⋮

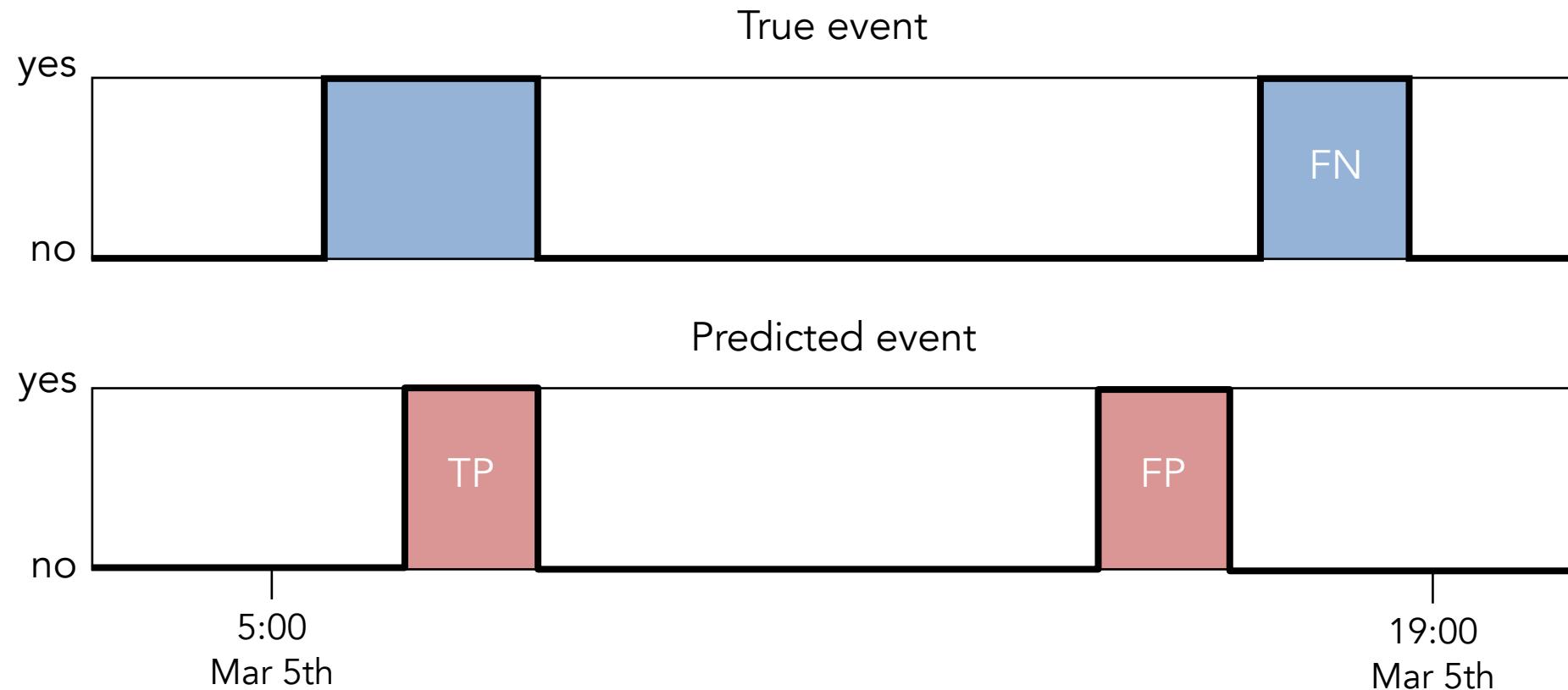


:(Has Event

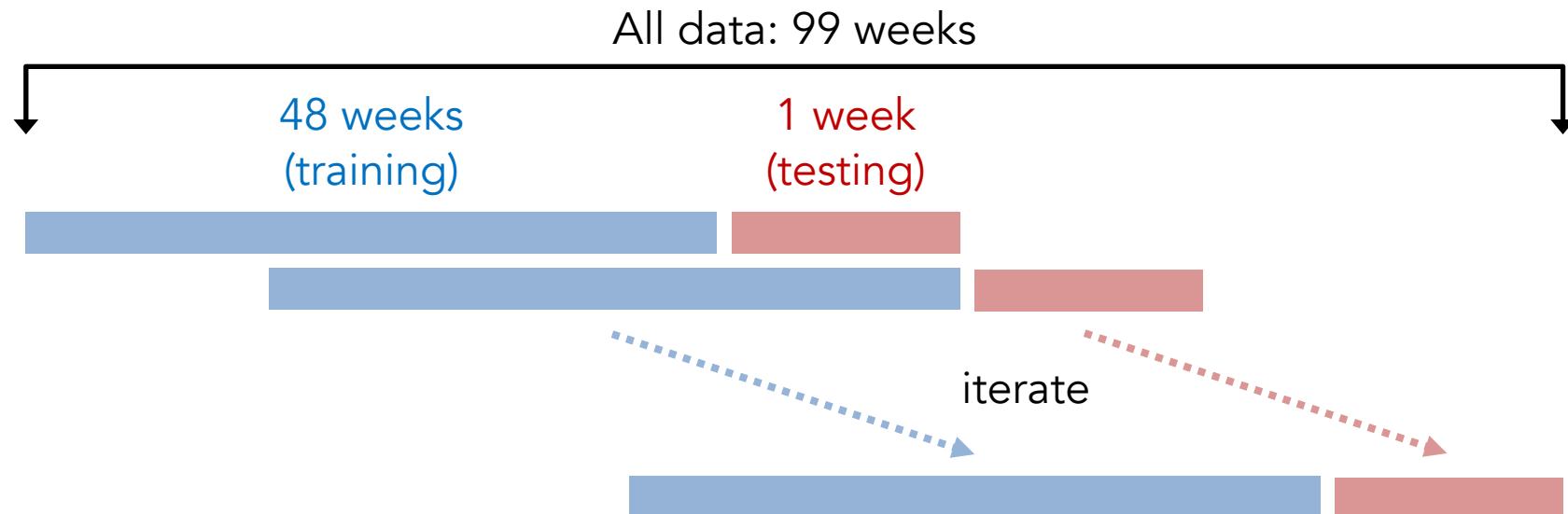


:> No Event

To evaluate models, we first compute **true positives (TP)**, **false positives (FP)**, and **false negatives (FN)** for smell events.

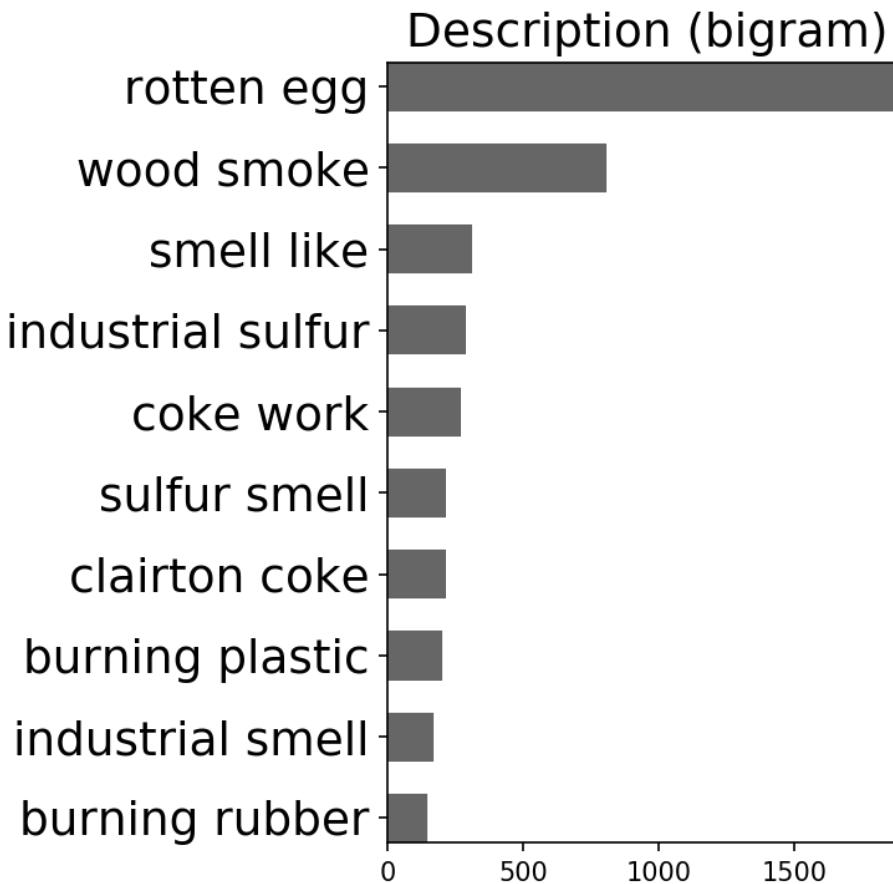
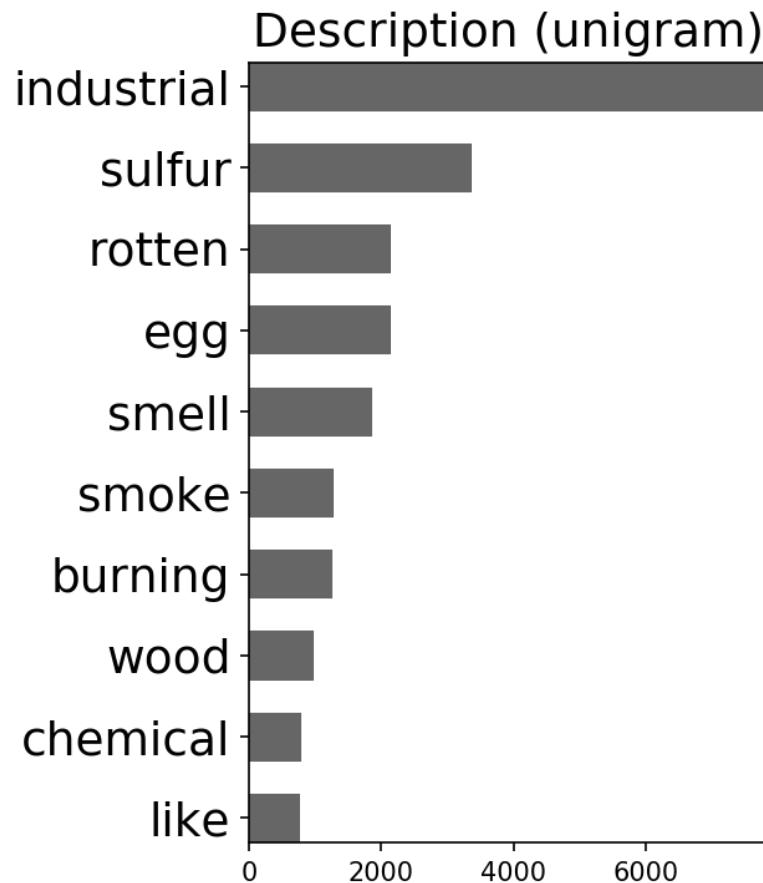


We apply time series cross-validation of several pairs of **training** and **testing** sets to evaluate the model performance.

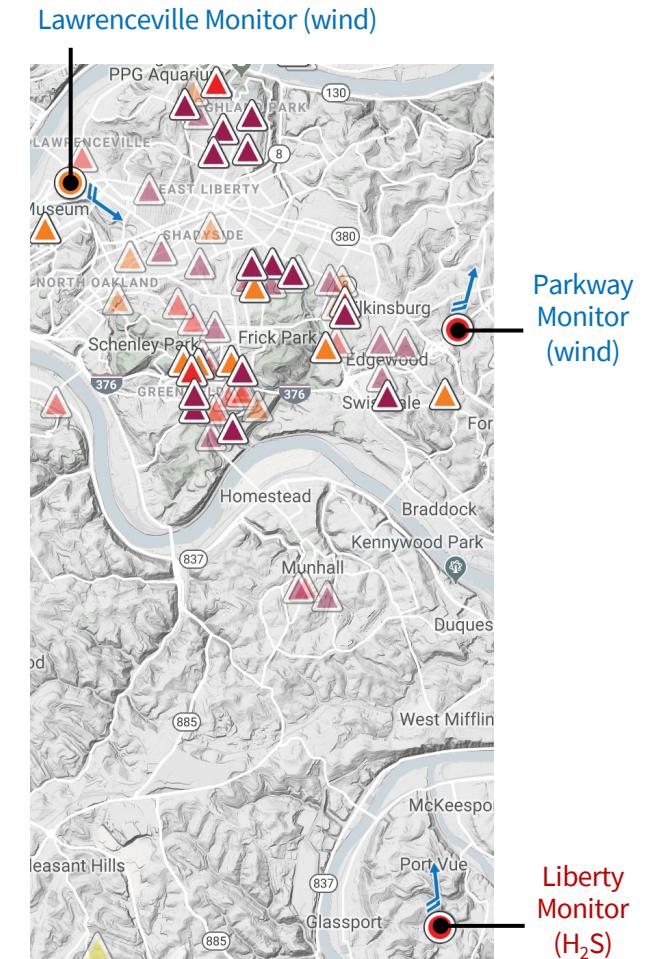
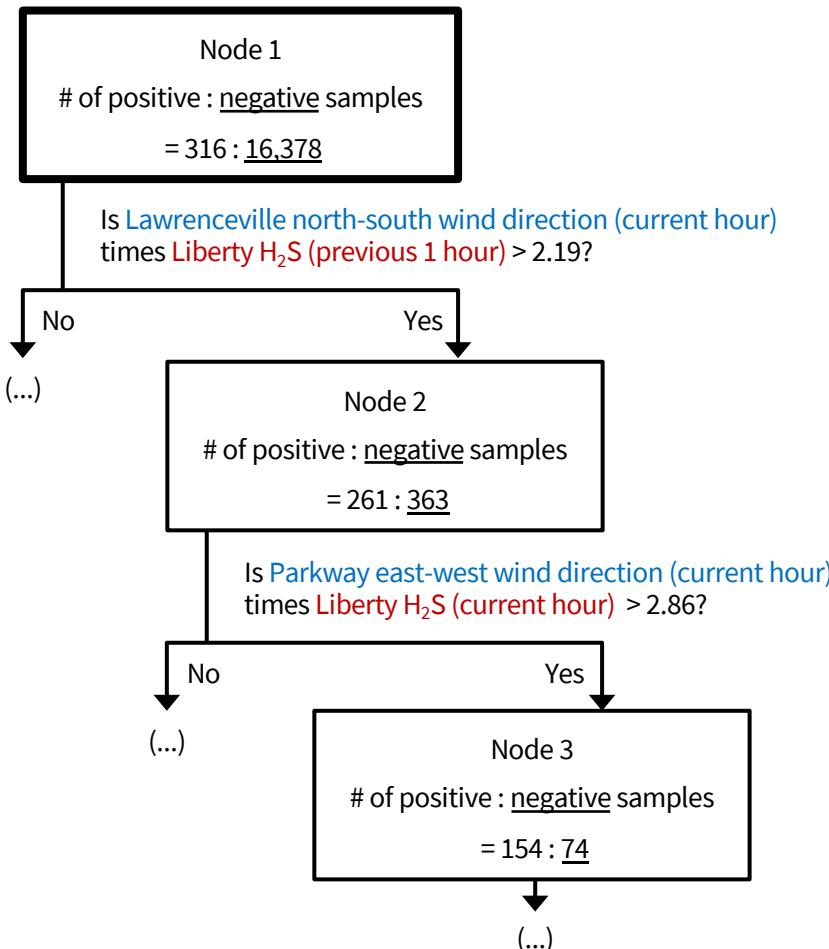


	Precision	Recall	F-score
Our best model	0.87 ± 0.01	0.59 ± 0.01	0.70 ± 0.01
Always yes	0.2	1	0.33

High-frequency words and phrases in smell reports mostly describe **industrial pollution odors**, like **hydrogen sulfide**.

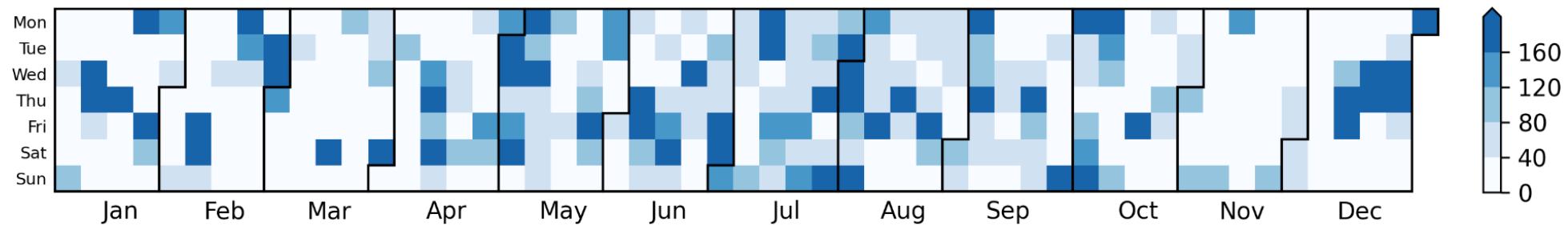


We also use Decision Tree (a machine learning model) to explain about 30% of the smell events, which is a joint effect of wind information and hydrogen sulfide.

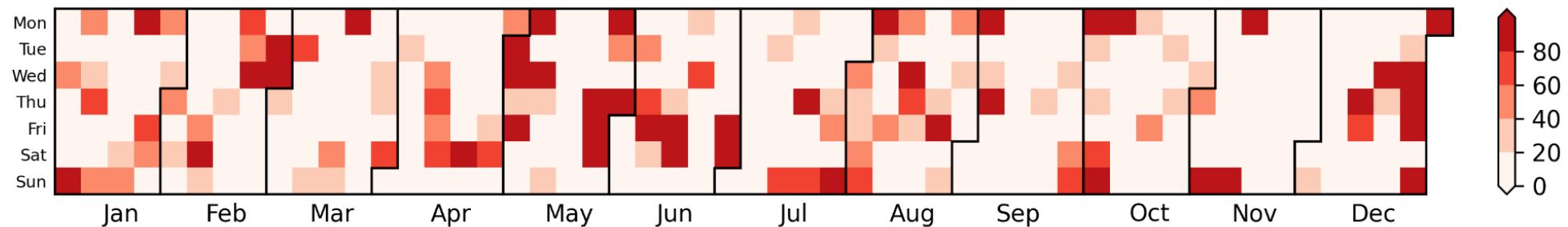


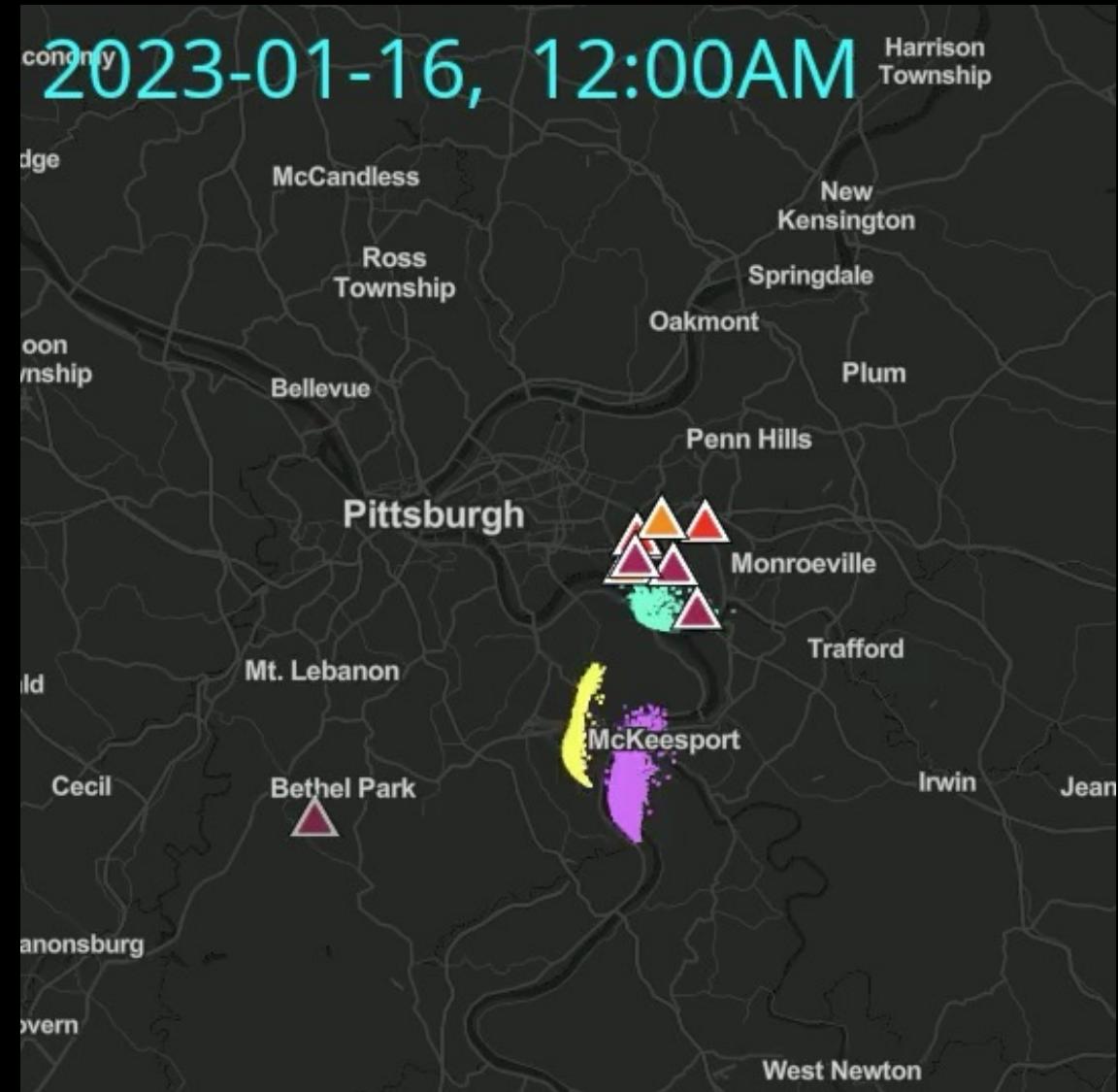
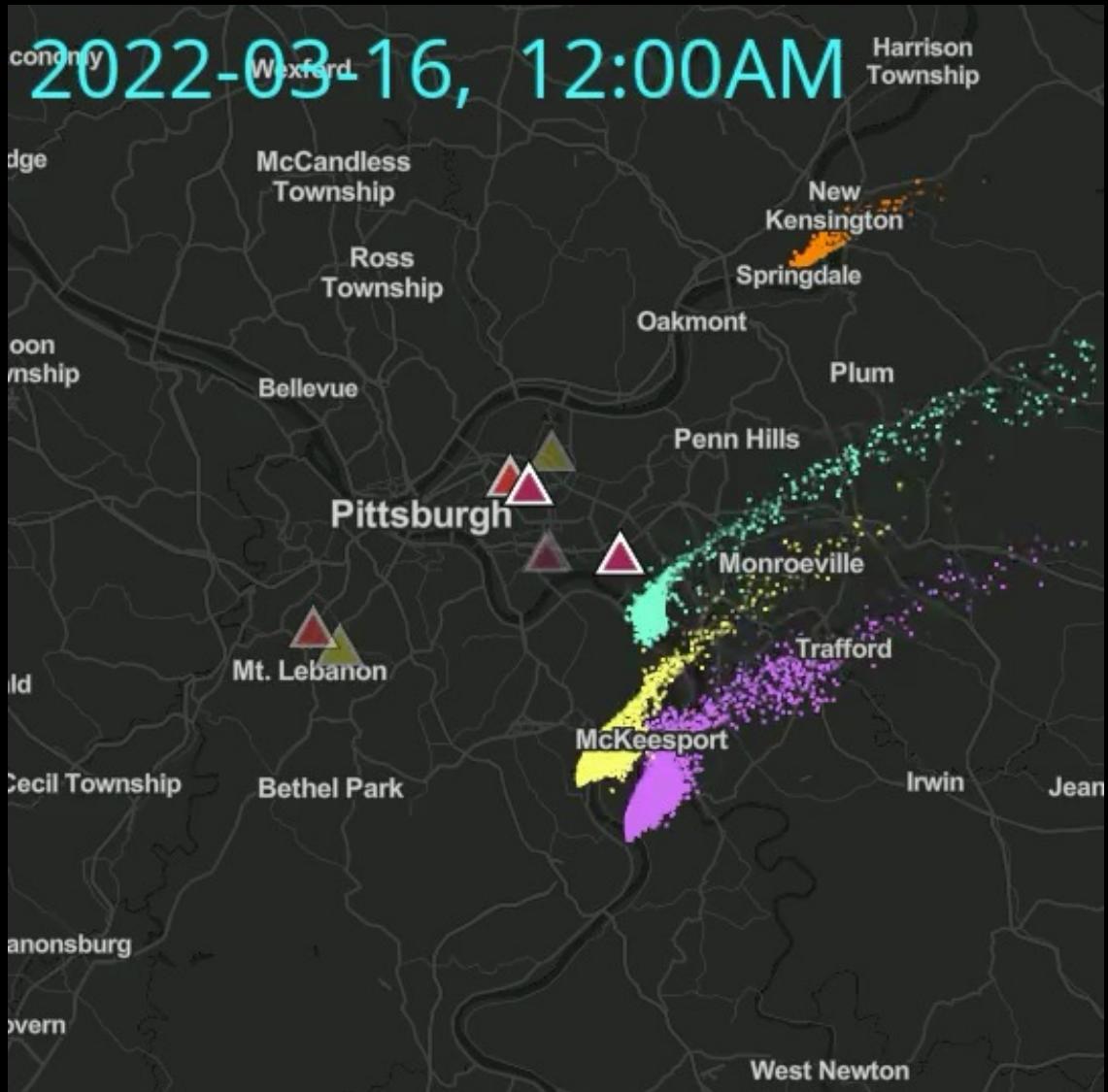
We further compute and visualize the relationship between the sum of smell ratings and the maximum concentration of weighted hydrogen sulfide for each day.

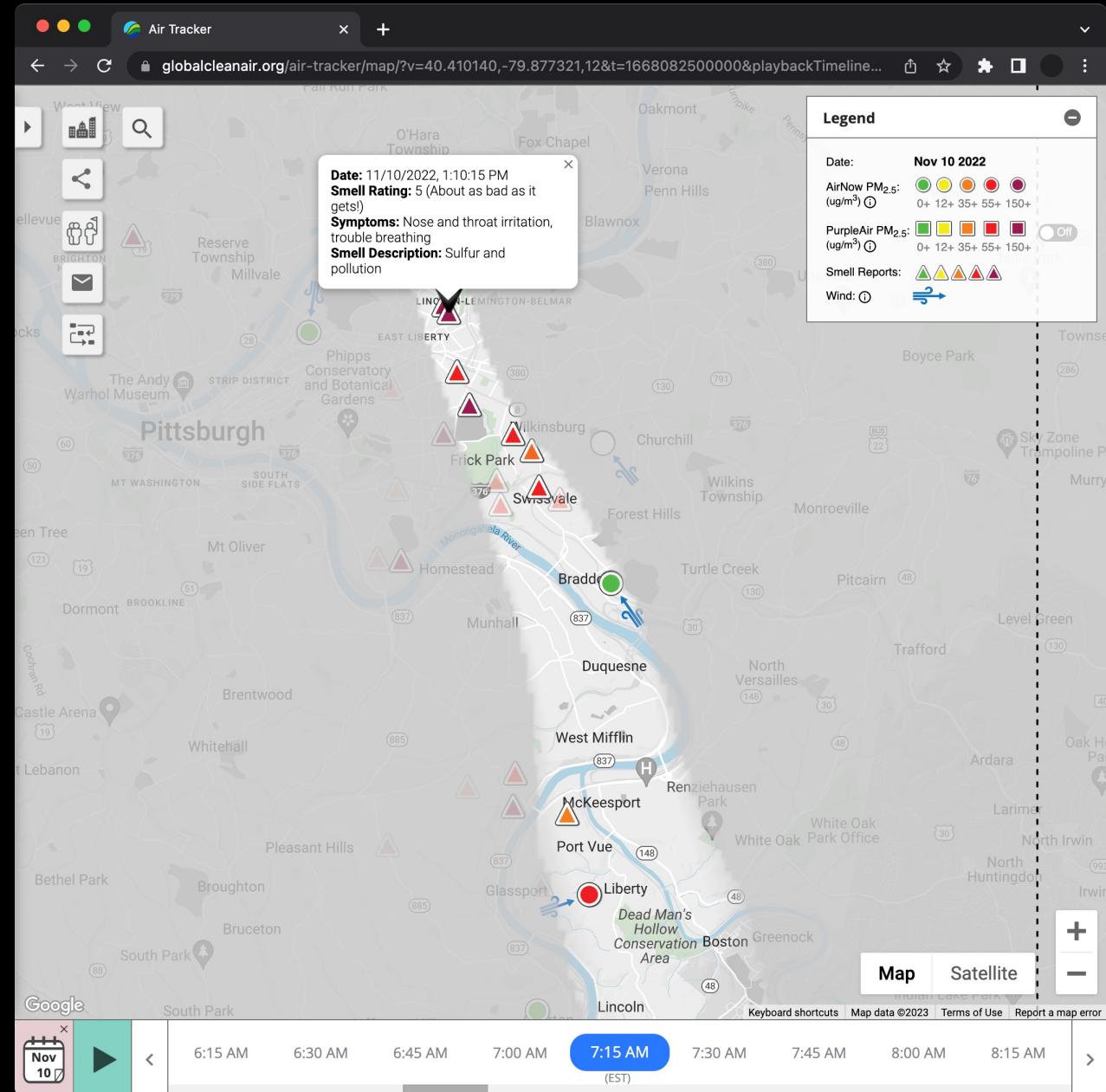
Sum of smell ratings by date (2018)



Maximum of weighted hydrogen sulfide concentration by date (2018)







Track potential air pollution sources -- <https://globalcleanair.org/air-tracker/map/>

Project RISE

Recognizing Industrial Smoke Emissions

<https://smellpgh.org>

Proceedings of the AAAI Conference on Artificial Intelligence

Current Archives About ▾

Search

Home / Archives / Vol. 35 No. 17: IAAI-21, EAAI-21, AAAI-21 Special Programs and Special Track / AAAI Special Track on AI for Social Impact

Project RISE: Recognizing Industrial Smoke Emissions

Yen-Chia Hsu
Carnegie Mellon University

Ting-Hao (Kenneth) Huang
Pennsylvania State University

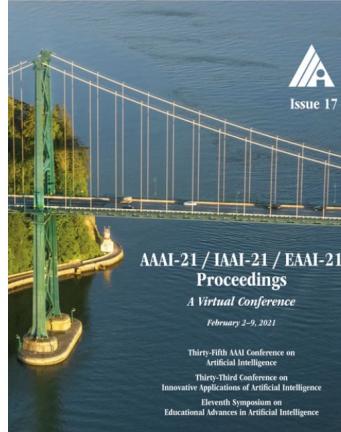
Ting-Yao Hu
Carnegie Mellon University

Paul Dille
Carnegie Mellon University

Sean Prendi
Carnegie Mellon University

Ryan Hoffman
Carnegie Mellon University

Anastasia Tsuhlares
Carnegie Mellon University



Information

For Readers
For Authors
For Librarians

Open Journal Systems

Subscription

Login to access subscriber-only resources.

PKP|PS

This project aim to [recognize industrial smoke emissions](#) automatically on the videos obtained from a camera monitoring network.



<http://smoke.createlab.org>



Has smoke



No smoke



Has smoke



Has smoke

We invite communities to annotate if the videos have industrial smoke emissions using a web-based tool.



Project RISE

So far, 22929 (23.91%) out of 95879 videos are fully labeled, and 3403 (3.55%) videos are partially labeled (learn more). You have reviewed 431 pages, of which (researcher) have passed the quality check (learn more).

Sign Out Interactive Tutorial My Contribution

Each video is 3 seconds. Click or tap to select videos that have smoke. Click or tap again to deselect. Skip a video if you are not sure whether it has smoke.

1

2

A screenshot of a web browser window titled "Project RISE" with the URL "smoke.createlab.org/label.html". The page contains instructions for labeling videos. Below the instructions are four video frames. The first frame on the left is outlined in red and has the number "1" below it. The second frame on the right is outlined in black and has the number "2" below it. The other two frames are also outlined in black. The top right corner of the browser window shows a menu icon consisting of three horizontal lines.



19
Views



The data [enables Computer Vision applications using deep neural networks](#), and our baseline model (I3D+Timeception) can find emissions with reasonable performance.



	Precision*	Recall*	F-score*
Our best I3D model	0.86	0.79	0.82
Always yes	0.41	1	0.58

*Average of the metric for all data splits (training, validation, and test)

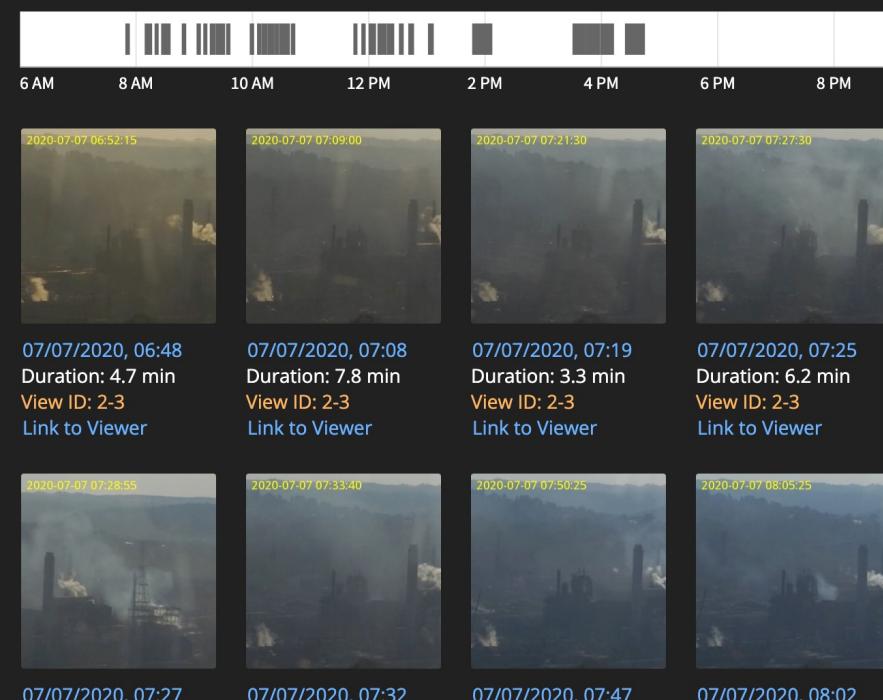
Project RISE



This page shows videos that the Artificial Intelligence model thinks **have hazardous smoke emissions**. The following image shows the camera view ID and location.



The timeline below shows smoke events ([learn more](#)), where the x-axis means hour of day. To show videos, select a date **2020-07-07** and compare the events with [Smell Pittsburgh](#). Also, select a camera ID **2** and view ID **all** to filter videos.



Dutch Context



spotdegifwolk.nl

SPOT DE GIFWOLK

gespotte gifwolken 66

over deze actie | delen | contact

15
14
13
12
11
10
9
8
7
6
5
4
3
2
1

Opgelegde dwangsommen

2023-06-12 22:28:04

Frisse Wind. Nu!

Hoogoven 7 = 110 meter hoog. Indien gif wolken geel/oranje dan o.a. gasvormig kwik.

GREENPEACE

Frisse Wind. Nu!

GIFWOLK MELDEN

HELP ?

Spot the Toxic Cloud -- <https://spotdegifwolk.nl/>

The screenshot shows a web browser window for the website [fruitpunch.ai](https://www.fruitpunch.ai). The main header features the FruitPunch AI logo, navigation links for Challenges, For Individuals, For Organizations, Publications, and a prominent pink button labeled "Community Platform". Below the header is a large background image of an industrial facility with smokestacks emitting plumes of smoke. Overlaid on this image are several text elements: a white rounded rectangle in the top-left corner contains the "FruitPunch AI" logo; a white rectangular box in the center-left says "STARTING SOON"; a large white h1 heading "AI against Toxic Clouds" is centered; a subtitle below it reads "Use Computer Vision to detect toxic clouds emitted by large factories and report the authorities on the sightings"; and a pink "Go to challenge" button in the bottom-right corner.

Help environmental authorities monitor illegal air pollution by large factories.

Illegal pollution damages the environment and public health. Old production plants often do not reach the standards of today's laws and regulations, but, due to corporate lobbying and meager funding, authorities are rarely able to uphold these standards. To combat toxic emissions, local governments can issue fines for harmful incidents. The only problem is that some companies cross the line daily, while environmental agencies do not have the manpower to monitor a facility 24/7.

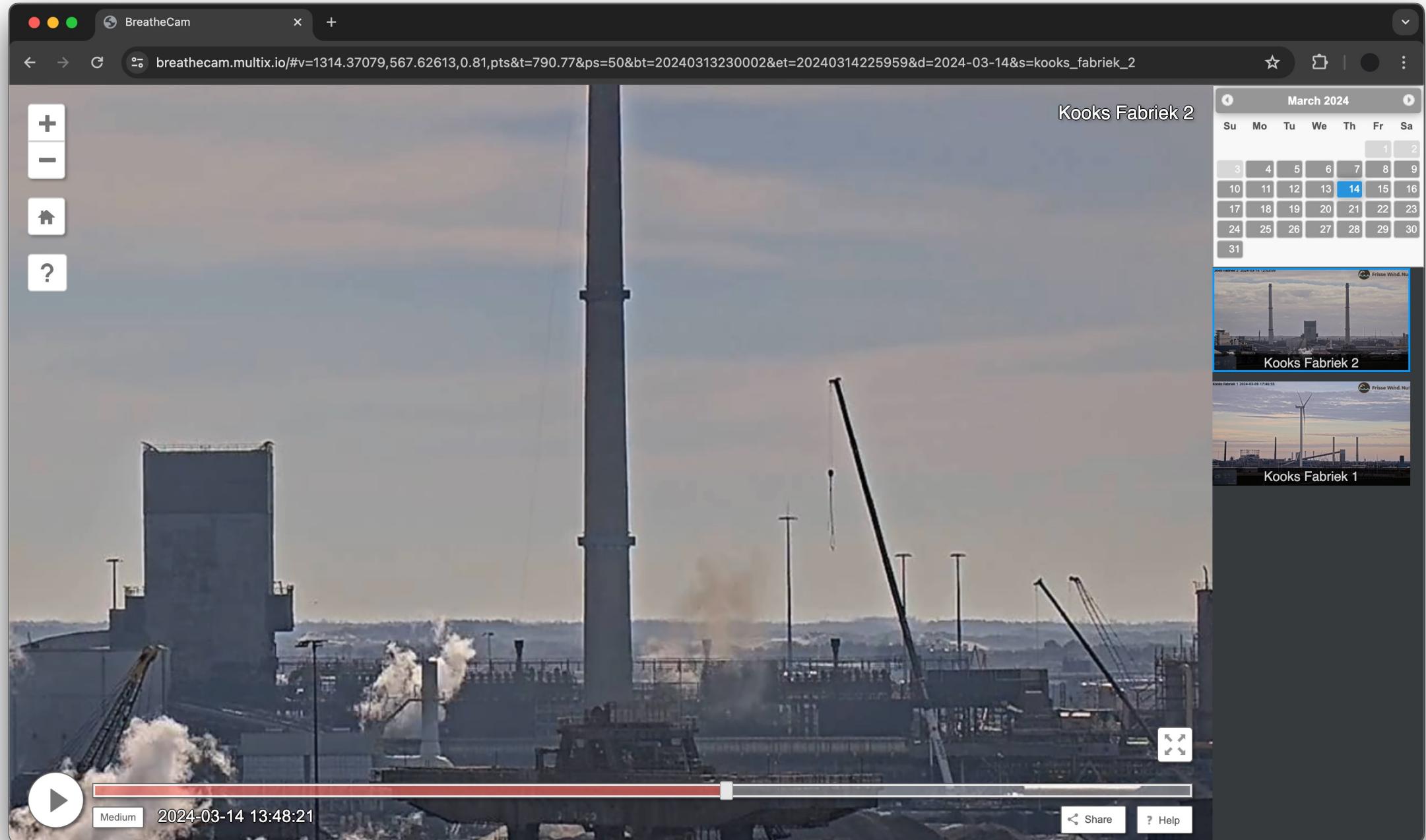
In this Challenge, we will collaborate with Greenpeace to aid in the monitoring of the Tata Steel plant in IJmuiden, the Netherlands.

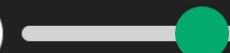
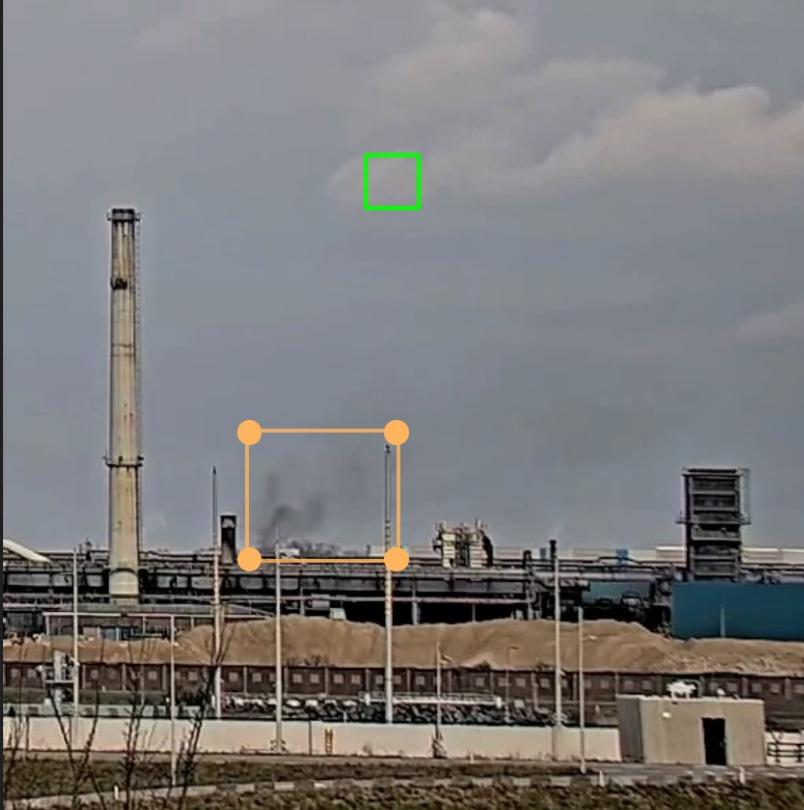
Tata Steel* is one of those sites that regularly release illegal toxic substances into the air. For this specific steel plant, we narrow down the problem to the most polluting factories, which are the two coke ovens*. These are extremely old and leak like a sieve. When processing raw coke, the factories sometimes release black clouds containing carcinogenic PAHs (polycyclic aromatic hydrocarbons), which is illegal under Dutch law. Unsurprisingly, the area around the Tata Steel plant sadly has a significantly higher incidence of lung cancer than the rest of the Netherlands.

Tata Steel is obligated to report incidents in which PAHs escape to the Dutch Environmental Services but rarely does so.

Application deadline
July 17, 2023
[Join challenge](#)

Timeline
Application Deadline: **17 July 2023**
Challenge Kick-off: **19 July 2023**
Midterm Presentations: **9 Augustus 2023**
Final Presentations: **27 September 2023**





10/36

Let op: geef alleen feedback op de afbeelding met het groene kader,
niet op de hele video.

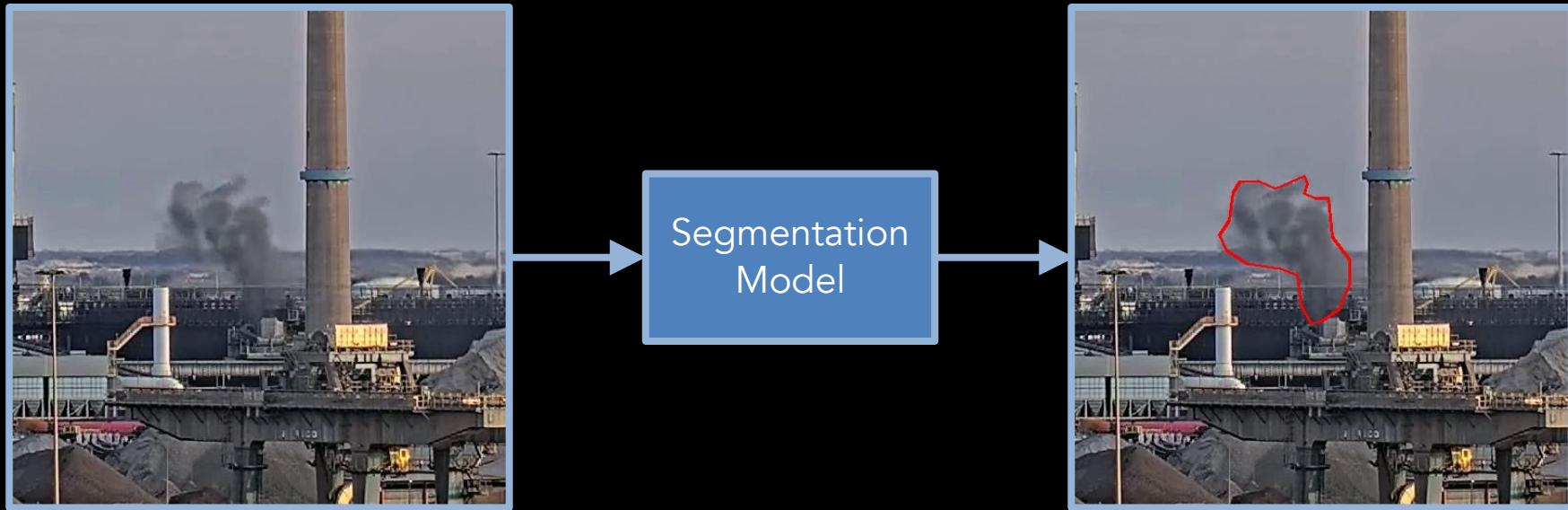
Beeld 5/8

Verwijder kader

< Vorige

Bevestigen en Volgende >

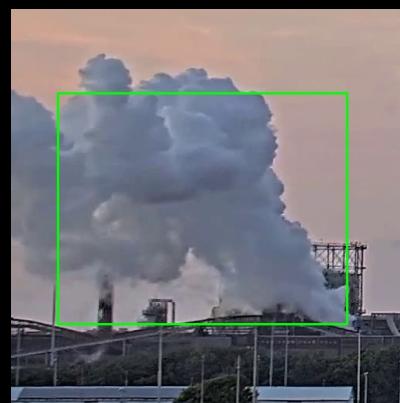




To what extent can community feedback help improve the model?



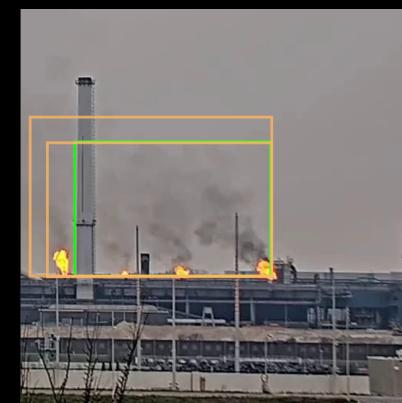
Wrong: No Smoke



Wrong: Steam



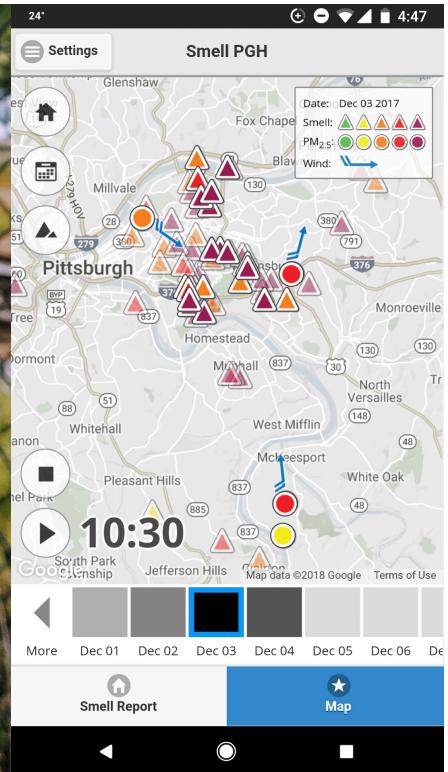
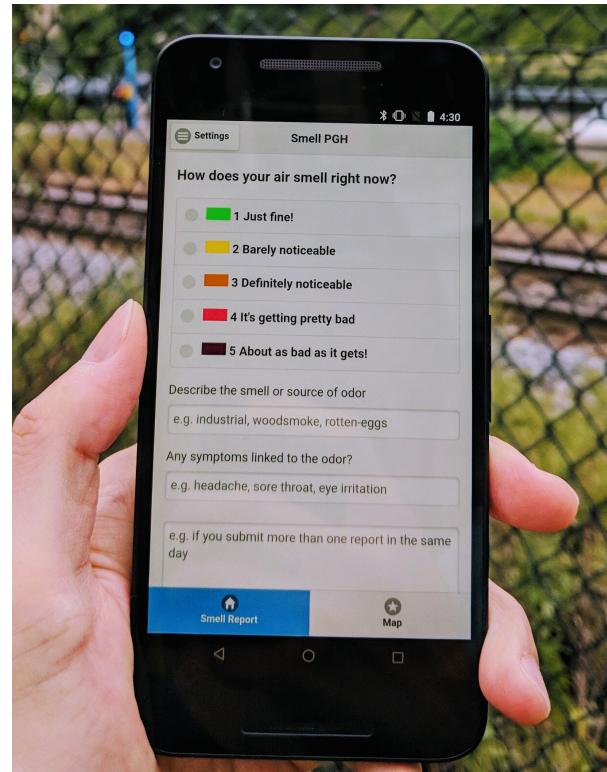
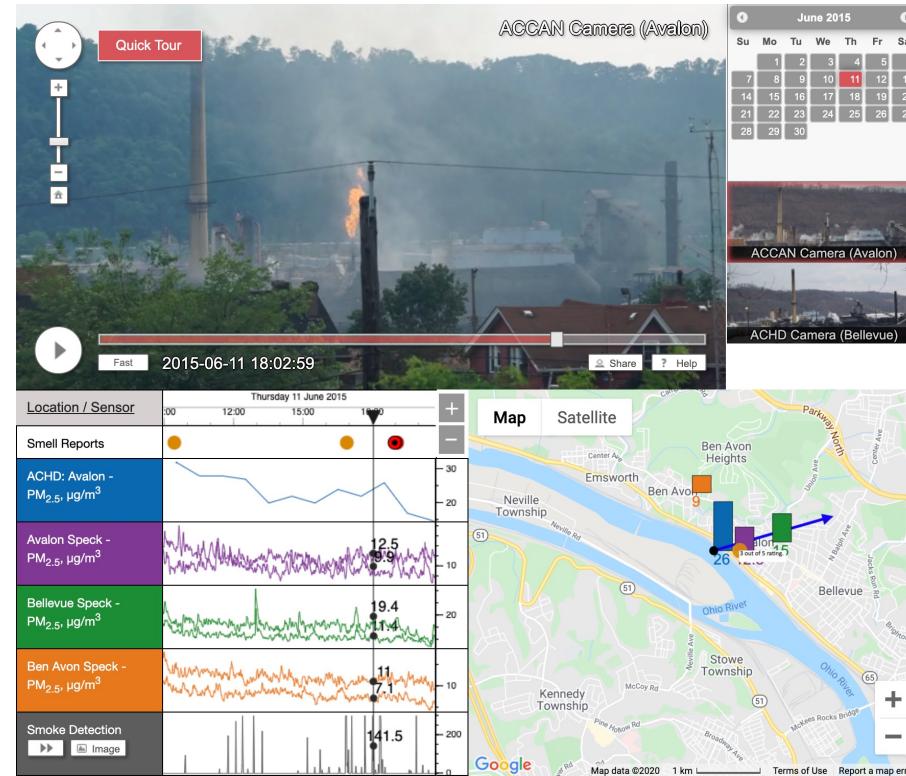
Too Small



Roughly OK

Questions?

- Yen-Chia Hsu – <http://yenchiah.me>
- Air Quality Monitoring – <http://shenangochannel.org>
- Smell Pittsburgh – <https://smellpgh.org>
- Project RISE – <https://smoke.createlab.org>
- IJmondCAM – <https://ijmondcam.multix.io/>



Project RISE

About Learn Label Gallery Event FAQ

So far, 12666 (13.21%) out of 95879 videos are fully labeled, and 11304 (11.79%) videos are partially labeled (learn more).

Sign In Interactive Tutorial My Contribution

Each video is 3 seconds. Click or tap to select videos that have smoke. Click or tap again to deselect. Skip a video if you are not sure whether it has smoke.

1

2

3

4