QuantumBlack Hackathon

Monitoring methane emissions with AI

Team 6 - Adel Remadi, Ian Moon, Lasse Schmidt, Marina Pellet,
Łukasz Pszenny, Clara Besnard

CleanR

Without a drastic reduction in methane emissions we will not reach the 1.5°C target

Methane has

84x

greater global warming potential than CO_2

Methane is responsible for

30%

of Global Warming

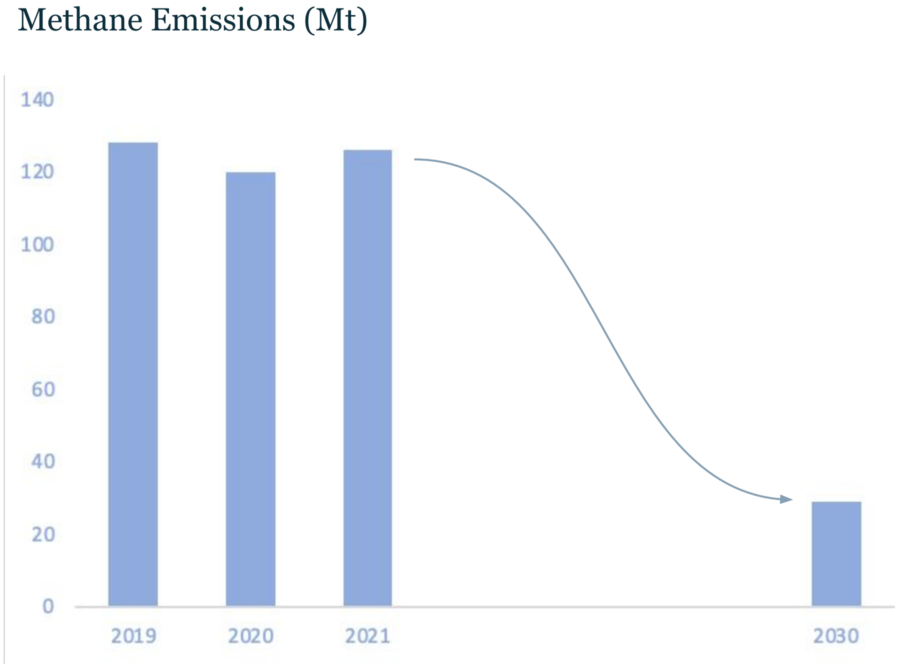
Human activities are at the origin of

60%

of global methane emissions

CleanR

Increasing regulatory pressure will require stronger monitoring & reporting



Recent regulations on methane emissions:

- **Global Methane Pledge 2021**
  30% reduction by 2030 (compared to 2020)
- **EU regulation on CH_4 emissions reduction 2021**
  55% reduction by 2030 (compared to 2015)
- **US Environmental Protection Agency**
  40% reduction by 2030 (compared to 2013)

Graph 1: Global methane emissions from fossil fuels in the Net Zero Scenario *(source: IEA)*

CleanR

NGOs and governments need a proactive strategy to meet goals

Needs

- **Identify** major emission events.
- **Notify** relevant stakeholders.
- **Support and track** evolution of mitigation strategies.

Pain Points:

- Strategy partly relies on companies **self-reporting**.
- Auditing and inspecting companies is **costly and time-consuming**.
- Studies show that **companies are emitting more methane than are being reported** by government agencies.

CleanR

Oil and Gas - the highest potential for rapid, cost-effective reductions

Oil and Gas Company

70 Mt = 18%

of global methane emissions

Financial Pain Points

- Estimated **$30B** of methane released into atmosphere.
- Lack of reliable reporting leads to poor management of **carbon credits**.

Operational Pain Points

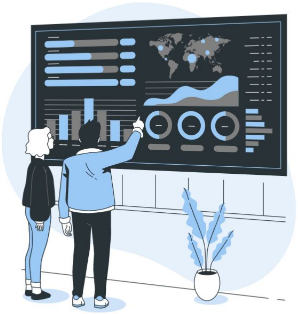
- Increased regulation on identification, reporting, and fixing methane leaks ⇒ costly **sensors** deployed on site and more frequent **manual walkthroughs** of methane sites.

On a Mission to foster collaboration around CH4 reduction

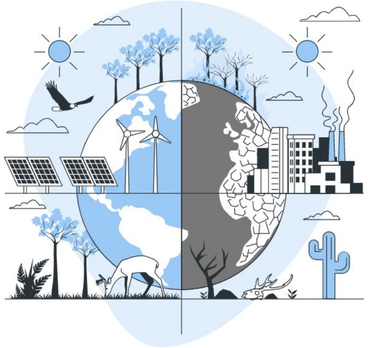
CleanR’s **mission** is to…

Leverage Satellite imaging, AI and Advanced **Analytics**

To power a **seamless monitoring and reporting tool** for methane emitters…



…and allow **regulators** to **track global** methane **emissions**



In order to meet international goals and **maintain sustainable life on earth!**

CleanR

CleanR - How it Works

Define sites to monitor and the frequency of analysis

Automatically pull satellite images via API

Dashboard to build, view, and report on methane emissions

Apply computer vision to detect methane leaks

CleanR applies advanced computer vision to satellite images to empower companies to **automatically monitor** their methane emissions and **seamlessly run reports** and analytics.

CleanR solves key pain points for companies and watchdogs

By leveraging open-source satellite images and computer vision, companies and NGOs can automatically monitor sites with potential methane leaks in a scalable manner.

Real-time view of global state of methane emissions

Fast and cost-effective deployment

Improved compliance through earlier detection of methane leaks

Seamless monitoring and reporting via automated data syncs

Modeling

Issues with Data Set - Our Solution - Model Performance

Dataset Overview

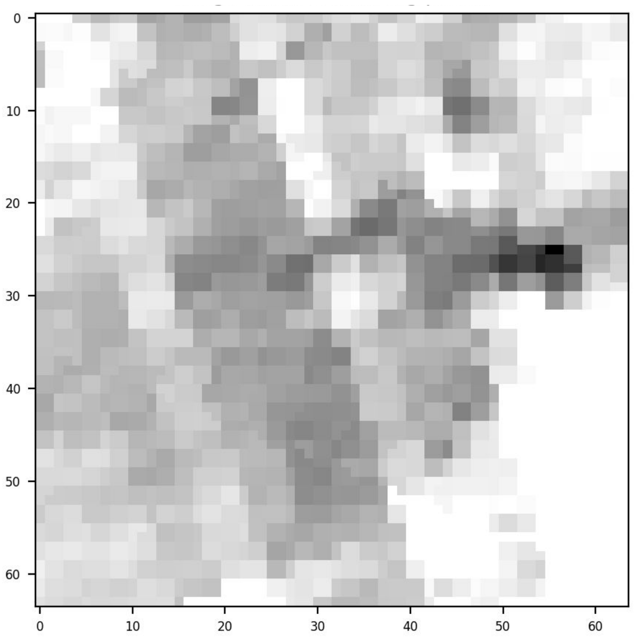
Metadata of data set

Data Provider: [Netherlands Institute for Space Research]

Geography: Worldwide

Year of Collection: 2023

Figure 1 Example of an image with plume*



*20230307_methane_mixing_ratio_id_8701.tif

CleanR

Dataset Overview

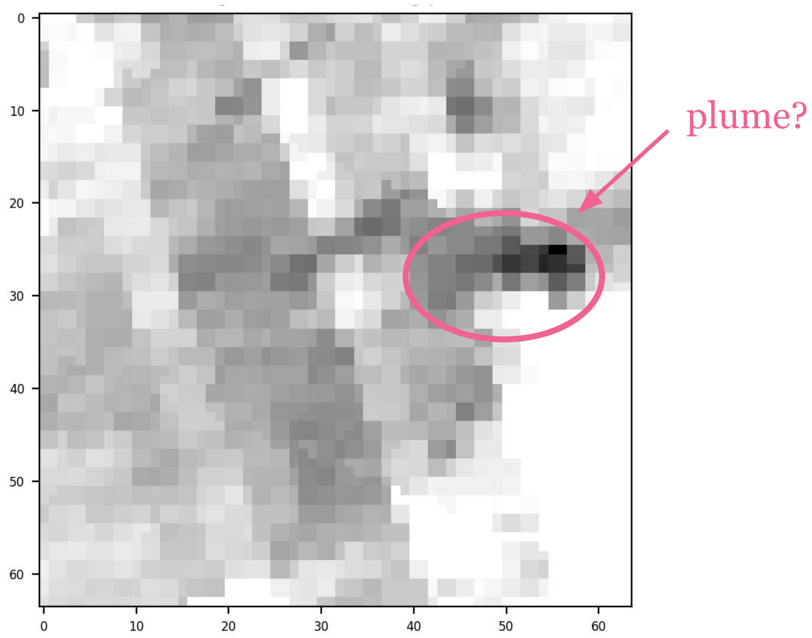
Metadata of data set

Data Provider: [Netherlands Institute for Space Research]

Geography: Worldwide

Year of Collection: 2023

Figure 1 Example of an image with plume*



*20230307_methane_mixing_ratio_id_8701.tif

Spatial Analysis of Dataset

Figure 1 Geographical Overview of Dataset
(each point is a unique location for which we can have up to 21 images)

Consequences of Imbalance on our Model Training and Selection Process

Learnings

Imbalanced labels : All 97 images from India have a methane plume (same issue with all other locations).

Consequences

Unwanted bias!

(do not pass location to model)

CleanR

Consequences of Imbalance on our Model Training and Selection Process

Learnings

Imbalanced labels : All 97 images from India have a methane plume (same issue with all other locations).

Skewed concentration : Some locations have many more images than others (25% of all images come from India).

Consequences

Unwanted bias!

(do not pass location to model)

Careful train-validation-test split.
→ to meaningfully select best hyperparameters
→ to accurately estimate our model

(to avoid data leakage in validation and test)

CleanR

Consequences of Imbalance on our Model Training and Selection Process

Learnings

Imbalanced labels : All 97 images from India have a methane plume (same issue with all other locations).

Skewed concentration : Some locations have many more images than others (25% of all images come from India).

Not a lot of data available : 430 images not a lot for Computer Vision.

Consequences

Unwanted bias!

(do not pass location to model)

Careful train-validation-test split.
→ to meaningfully select best hyperparameters
→ to accurately estimate our model

(to avoid data leakage in validation and test)

Apply heavy random transformations on training data to artificially enrich dataset.

CleanR

Our solution to Imbalance : Stratified 5-fold cross validation procedure

Table 1 Illustration of location-based, stratified 5-fold cross validation

Location	1	2	3	4	...	100	101
Nbr Images	2	11	17	7	...	4	2
Fold 1	Train	Validation	Train	Train		Test	Test
Fold 2	Validation	Train	Train	Train		Test	Test
⋮							
Fold 5	Train	Train	Validation	Validation		Test	Test

CleanR

Our solution to Imbalance : Stratified 5-fold cross validation procedure

Table 1 Illustration of location-based, stratified 5-fold cross validation

Location	1	2	3	4	...	100	101
Nbr Images	2	11	17	7	...	4	2
Fold 1	Train	Validation	Train	Train		Test	Test
Fold 2	Validation	Train	Train	Train		Test	Test
⋮							
Fold 5	Train	Train	Validation	Validation		Test	Test

Methodology:

1. Test set contains roughly 15% of all images.
2. Each validation set contains roughly 20% of all images and non-test locations appear exactly once in validation set across all 5 folds.
3. In each fold, we select the best performing model using AUC on validation set.
4. We run best model of each fold on the test set to get accurate performance estimate of our ensemble.

CleanR

Model Performance Overview

Table 2 Overview of Model Performances (ranked by Test AUC descending)

No.	Model	Train AUC	Val AUC	Test AUC	Training Time (min)	Model Size (mb)
1	ResNet34 + 2 Layer Linear Classifier	97.4	91.4	91.3	15	83
2	MobileNet_V3_Large	90.7	92.19	90.22	15	16
3	MobileNet_V3_Small	87.6	91.7	89.4	12	5
4	DenseNet 121	87.8	90.4	87.3	24	28
5	DenseNet 161	83.6	87.8	85.9	30	105
Baseline	ResNet18 + 1 Layer Linear Classifier	95.2	90.0	84.6	10	43

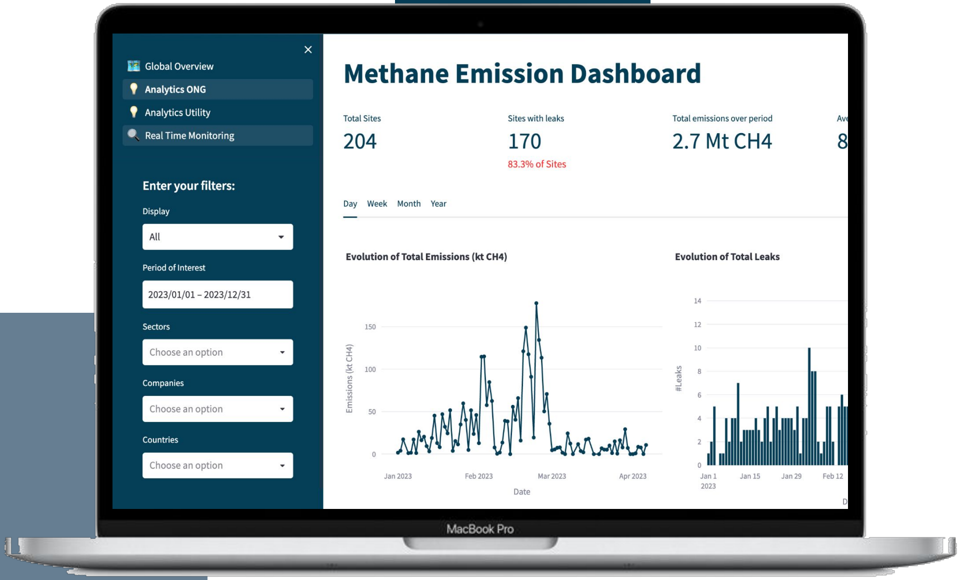
Model Performance Overview

Table 2 Overview of Model Performances (ranked by Test AUC descending)

No.	Model	Train AUC	Val AUC	Test AUC	Training Time (min)	Model Size (mb)
1	ResNet34 + 2 Layer Linear Classifier	97.4	91.4	91.3	15	83
2	MobileNet_V3_Large	90.7	92.19	90.22	15	16
3	MobileNet_V3_Small	87.6	91.7	89.4	12	5
4	DenseNet 121	87.8	90.4	87.3	24	28
5	DenseNet 161	83.6	87.8	85.9	30	105
Baseline	ResNet18 + 1 Layer Linear Classifier	95.2	90.0	84.6	10	43

Final Decision: Use Model #2! “Only” second best performance on our custom test set but can perform inference much faster thanks to 4x smaller model size compared to Model #1 (16mb). Also very stable results across Train, Val, and Test. Model #1 shows more variance in its performance.

CleanR



CleanR

Product Demo

Conclusion

Problems

Clear gap between **sustainability targets** and **current emission reduction** efforts.

Increase in regulatory pressure to **decrease** methane **emissions faster**.

Monitoring and reporting process to detect and reduce methane emissions is **costly, manual, and unscalable**.

Solutions

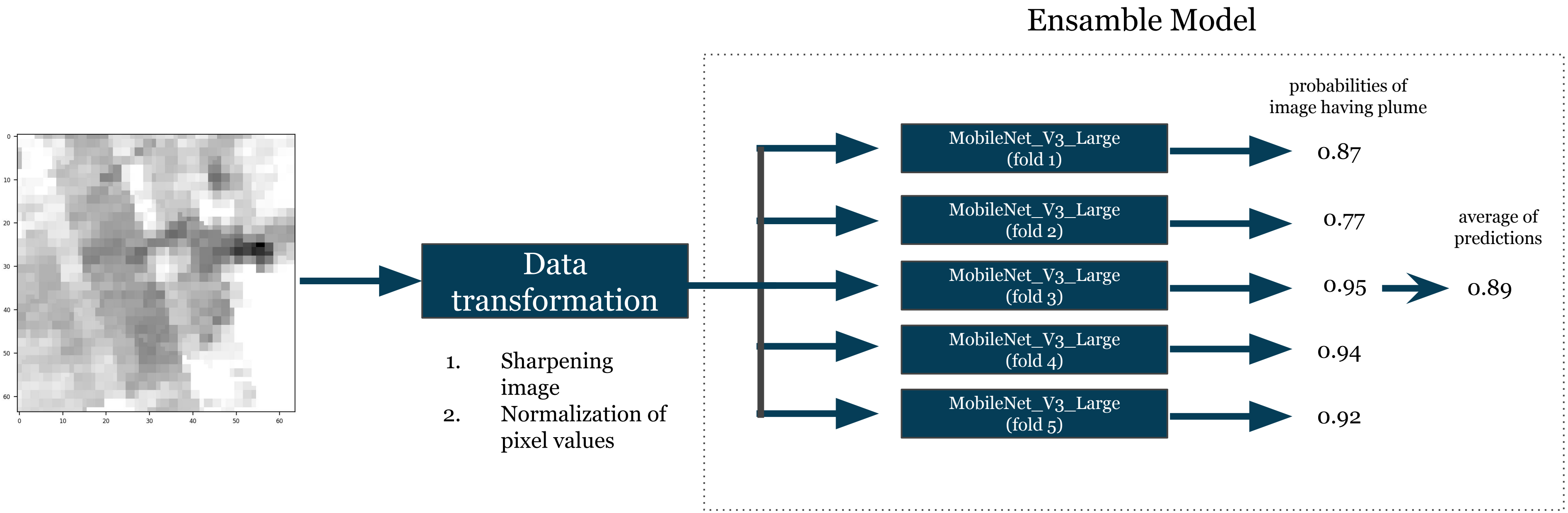
Easy-to-deploy tool that runs in the background

Early detection of leaks

Comparative analytics to assess ESG performance

Appendix

Final data flow



Dataset Overview

Metadata of data set

Data Provider: [Netherlands Institute for Space Research]

Geography: Worldwide

Year of Collection: 2023

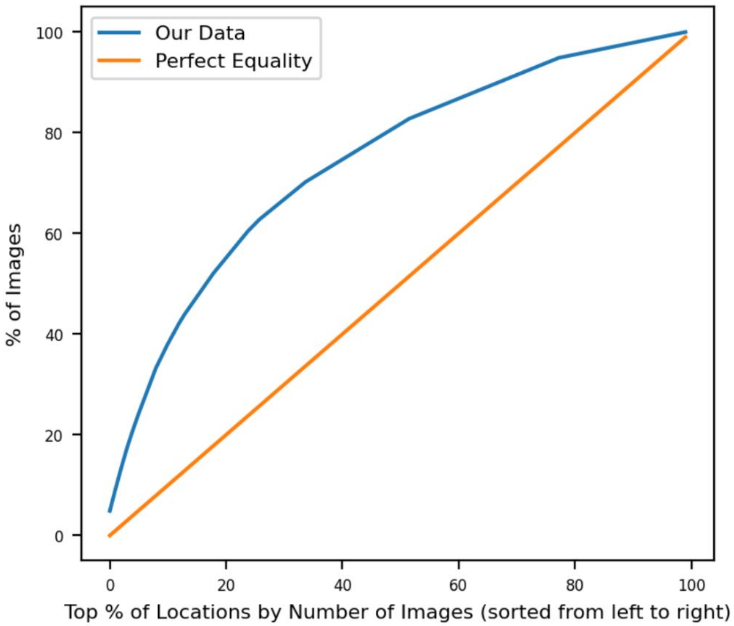
Table 1 Example of a datapoint

data type	feature name	values
categorical	date	20230307
	set	train
geographical	city	Kahna
	country	Pakistan
	country_code	PK
	id_coord	id_8701
	lat	31.443333
	lon	74.316667
numerical	coord_x	26
	coord_y	47
string	path	images/plume/20230307_meth...
target	plume	yes

Quantitative Analysis of Dataset Imbalance across sub-geographies

Figure 1 Geographical Overview of Dataset

Graph 2 Inequality of Number of Images by Location*



*Location refers to dataset feature "coord_id"

CleanR

Model Descriptions

ResNet architecture uses identity shortcut connections to allow the model to skip one or more layers. This allows for a deeper architecture without affecting performance.

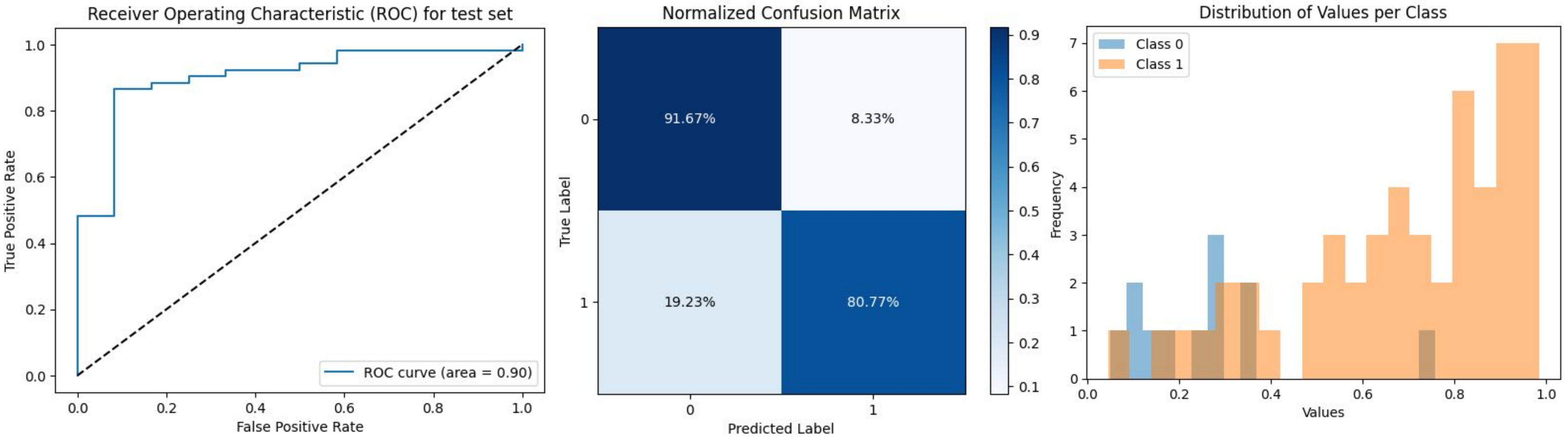
MobileNet uses depth-wise separable convolutions to construct a more lightweight architecture.

DenseNet "connects each layer to every other layer in a feed-forward fashion. For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers."

EfficientNet architecture uses scaling methods to uniformly scale all dimensions

SqueezeNet architecture is designed to reduce the number of parameters by using fire modules that squeeze parameters to 1x1 convolutions.

Performance report on artificial test set



Classification report

	precision	recall	f1-score	support
0	0.52	0.92	0.67	12
1	0.98	0.81	0.88	52
accuracy			0.83	64
macro avg	0.75	0.86	0.78	64
weighted avg	0.89	0.83	0.84	64

CleanR

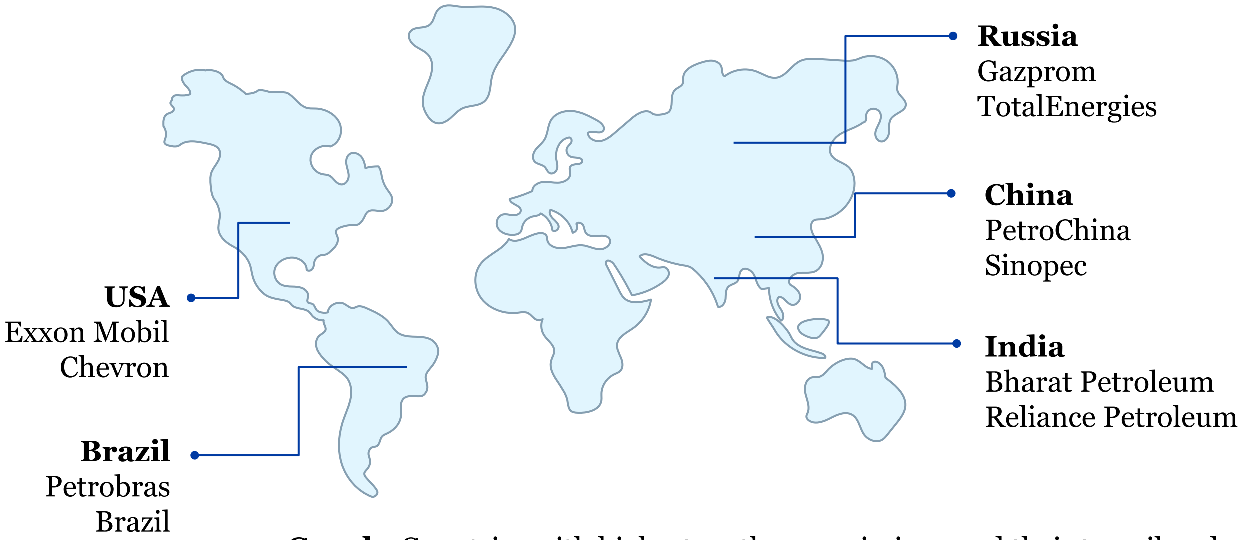
Deployment Strategy

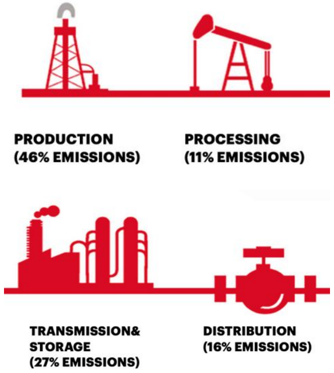
<u>NGOs/Regulators</u>

1. Build database of top methane sites with historical data
2. Set up ongoing pulls of new satellite images

<u>Companies</u>

1. Target countries and sectors with the highest emissions => Oil and Gas
2. Leverage sector averages to target companies who are below average on methane leak fixes



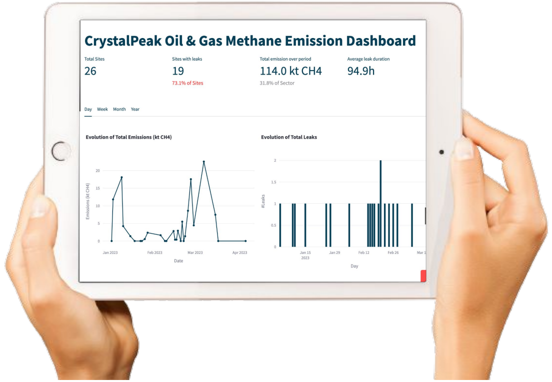


Graph: Countries with highest methane emissions and their top oil and gas companies

Pricing

Cost

- Data
    - Satellite images: NASA API provides free satellite images (rate limits); paid services
    - Data Warehouse - cloud provider
- Tool
    - Engineers and Hardware
- Operations/Customer Support
- Sales/Marketing
    - Conferences and Sales Team



Price

- NGO/Regulators - Use NASA API to provide a lower cost solution
- Companies - Recurring fee based on # of sites to monitor and reporting capabilities

CleanR will apply for EU funding to get money for developers and data engineers to build database for NGOs and continue to build features for the tool

Imbalance in labels and skewed concentration in sub-geographies

Learnings

Imbalanced labels : All 97 images from India have a methane plume (same issue with all other locations).

Skewed concentration : Some locations have many more images than others (25% of all images come from India).

Not a lot of data available : 430 images not a lot for Computer Vision.

Figure 1 Geographical Overview of Dataset
(each point is a unique location for which we can have up to 21 images)

Sources

https://edgar.jrc.ec.europa.eu/dataset_ghg70

https://www.iea.org/reports/methane-emissions-from-oil-and-gas-operations

https://www.edf.org/sites/default/files/documents/Fueling%20a%20Digital%20Methane%20Future_FINAL.pdf

https://rhg.com/research/untapped-potential-reducing-global-methane-emissions-from-oil-and-natural-gas-systems/

https://iea.blob.core.windows.net/assets/465cb813-5bf0-46e5-a267-3be0ccf332c4/Driving_Down_Methane_Leaks_from_the_Oil_and_Gas_Industry.pdf

https://www.hindawi.com/journals/misy/2020/7602384/

https://arxiv.org/abs/1608.06993

https://arxiv.org/abs/1905.11946

https://arxiv.org/abs/1602.07360

https://edition.cnn.com/2023/04/17/us/methane-oil-and-gas-epa-climate/index.html

https://www.accenture.com/us-en/blogs/accenture-utilities-blog/methane-measurement#:~:text=Apart%20from%20the%20safety%20risk,the%20atmosphere%20than%20carbon%20dioxide