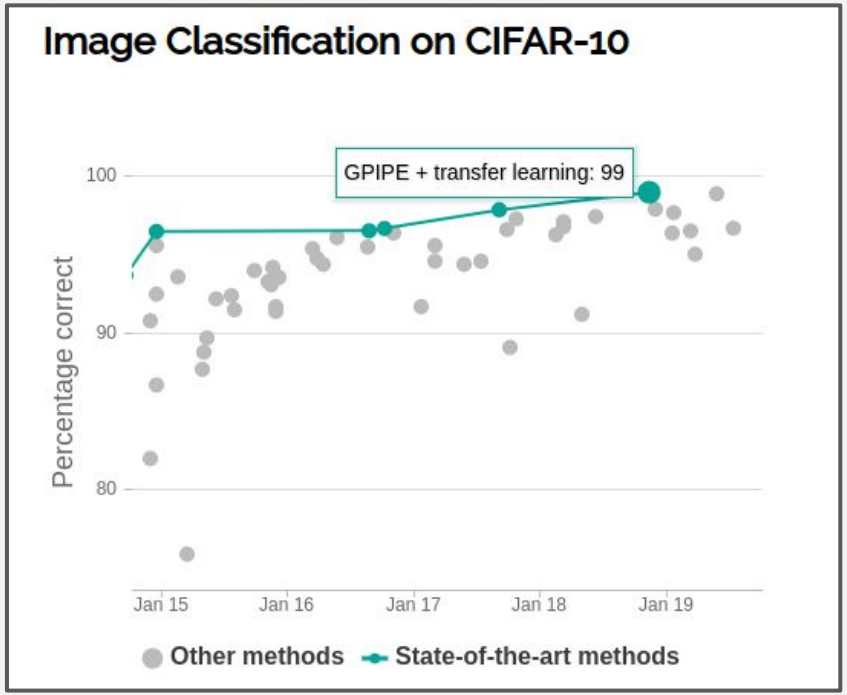
# AI for Common Good through benchmarking at scale

Laure Delisle

AI COMMONS

AI research is doing great!

# Performant



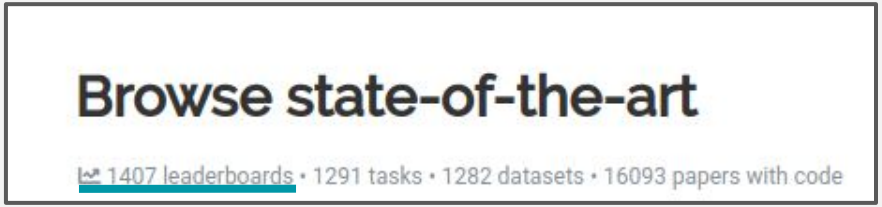

Performance

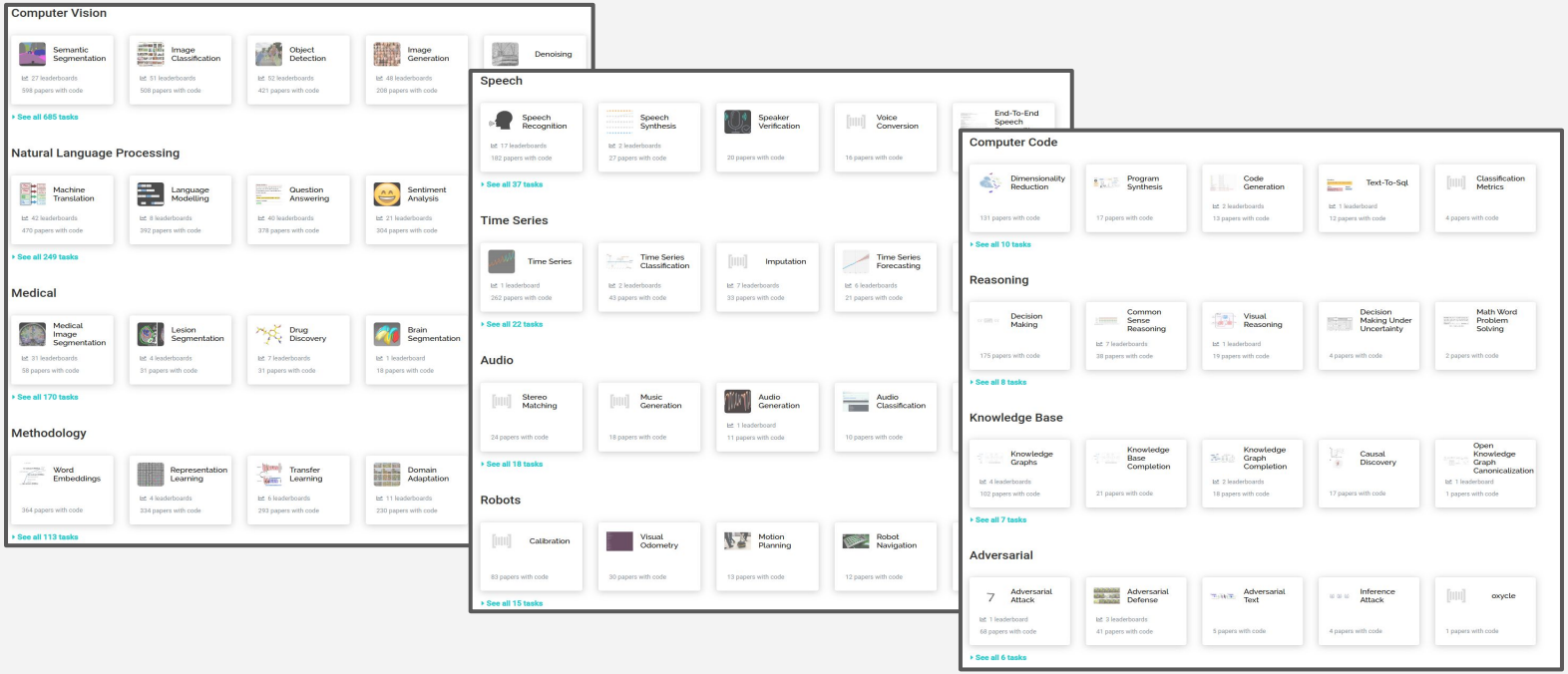
| | Comparison to best public-available results | | | | |
|---|---|---|---|---|---|
| | Model | Acc. | #Param | Our Model | Acc. |
| CIFAR-10 | NASNet-A | 98.0% | 85M | EfficientNet-B0 | 98.1% |
| CIFAR-100 | NASNet-A | 87.5% | 85M | EfficientNet-B0 | 88.1% |
| Birdsnap | Inception-v4 | 81.8% | 41M | EfficientNet-B5 | 82.0% |
| Stanford Cars | Inception-v4 | 93.4% | 41M | EfficientNet-B3 | 93.6% |
| Flowers | Inception-v4 | 98.5% | 41M | EfficientNet-B5 | 98.5% |
| FGVC Aircraft | Inception-v4 | 90.9% | 41M | EfficientNet-B3 | 90.7% |
| Oxford-IIIT Pets | ResNet-152 | 94.5% | 58M | EfficientNet-B4 | 94.8% |
| Food-101 | Inception-v4 | 90.8% | 41M | EfficientNet-B4 | 91.5% |

Benchmarking

**SOTA** - [Papers with Code]

# Prolific



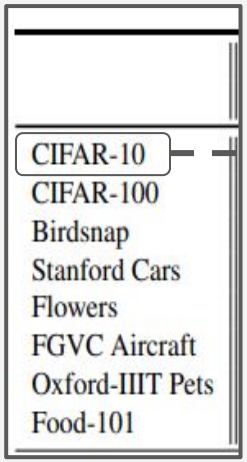



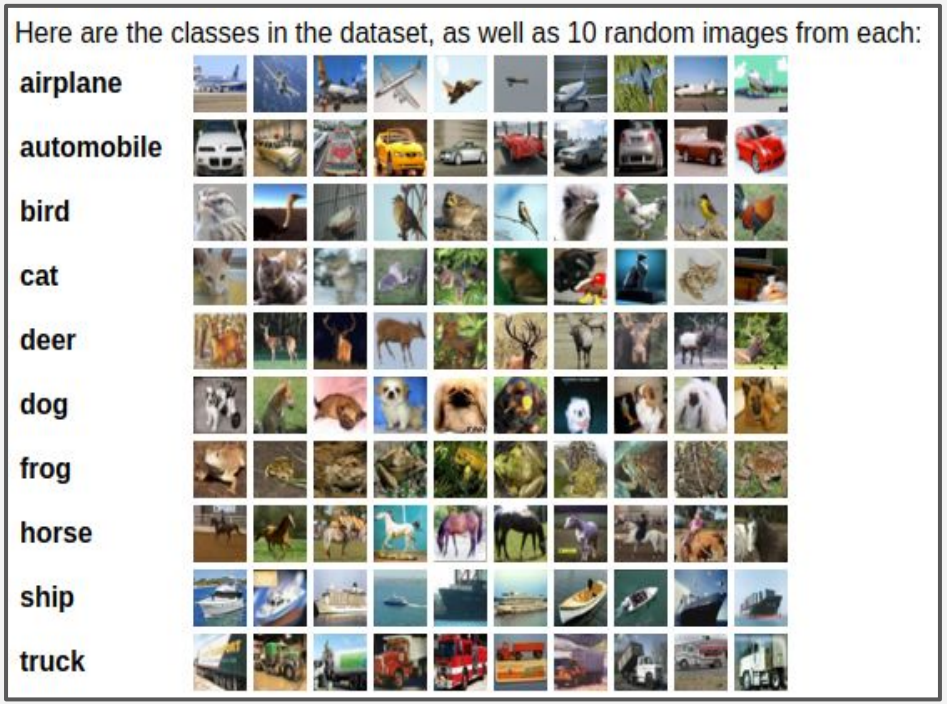

State Of The Art

**SOTA** - [Papers with Code]

# We can do even better

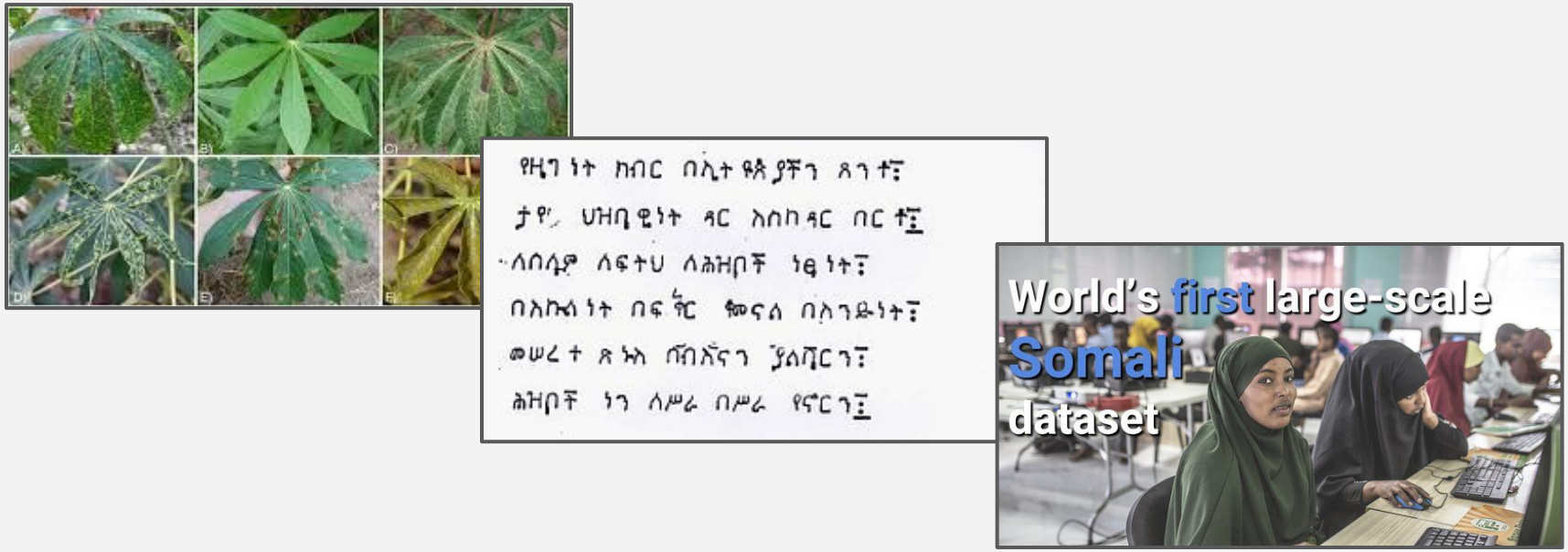
# Current benchmarking datasets...




**CIFAR-10** - Alex Krizhevsky. Learning Multiple Layers of Features from Tiny Images. 2009.

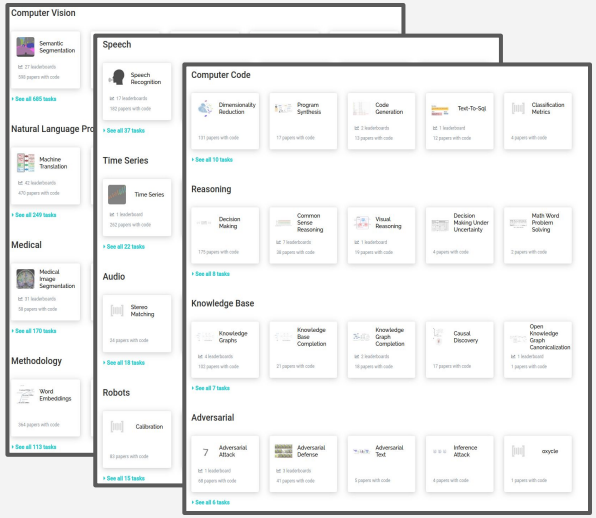
# Datasets for real world *impact*




Underserved datasets

**Cassava leaves dataset** - Ramcharan et al. Deep Learning for Image-Based Cassava Disease Detection. Frontiers in Plant Science 2017.
**Ethiopian datasets** - Knoema. Illustration: Ethiopian anthem in Ahmaric
**Somali audio transcription dataset** - Shaqodoon. 2019

# Filling the gap

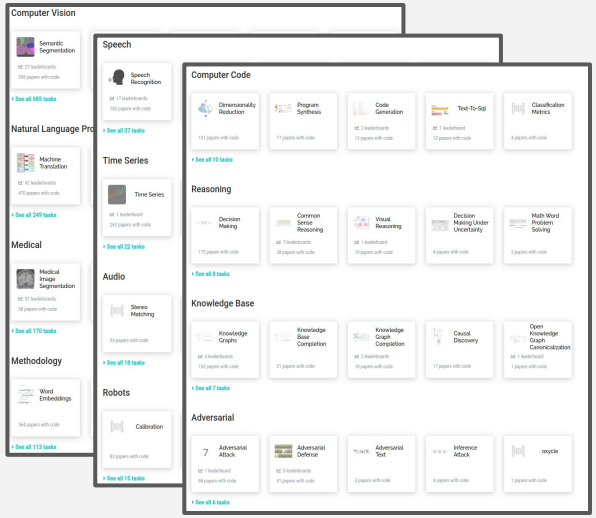



State Of The Art

?


Underserved datasets

# Filling the gap




State Of The Art

- Incentives ➤


Underserved datasets

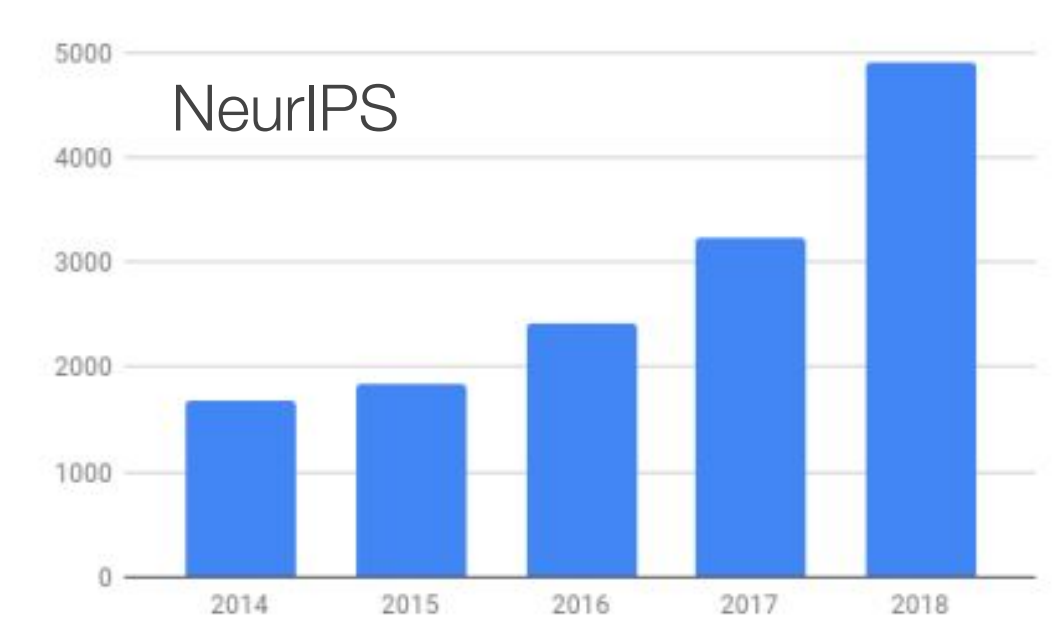
# Exploiting the system

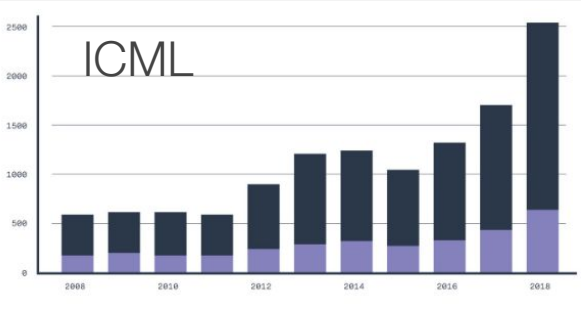
In 2018, the number of conference paper submissions has increased

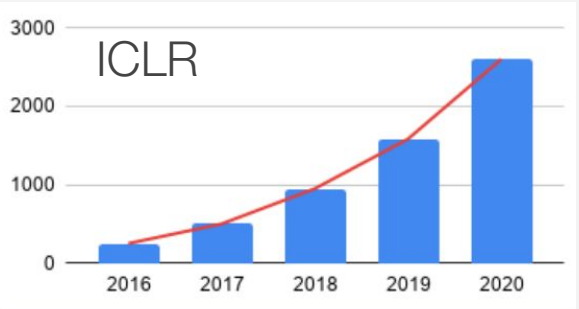
by 47% for ICML, by 50% for NeurIPS, and by almost 100% for ICLR.

D. Sculley, Jasper Snoek and Alex Wiltschko. Avoiding a Tragedy of the Commons in the Peer Review Process. Critiquing and Correcting Trends in Machine Learning, NeurIPS 2018 Workshop.

# AI papers submitted

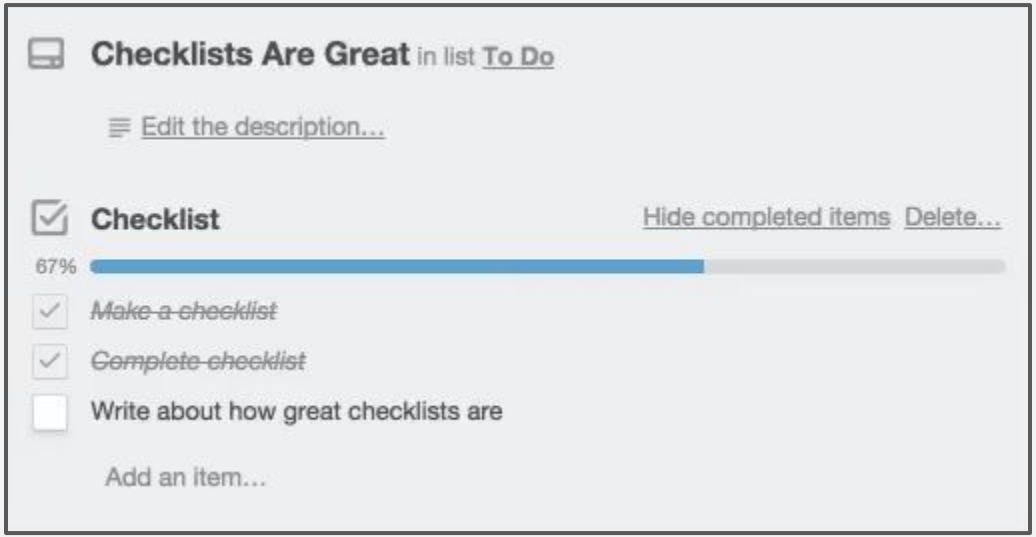


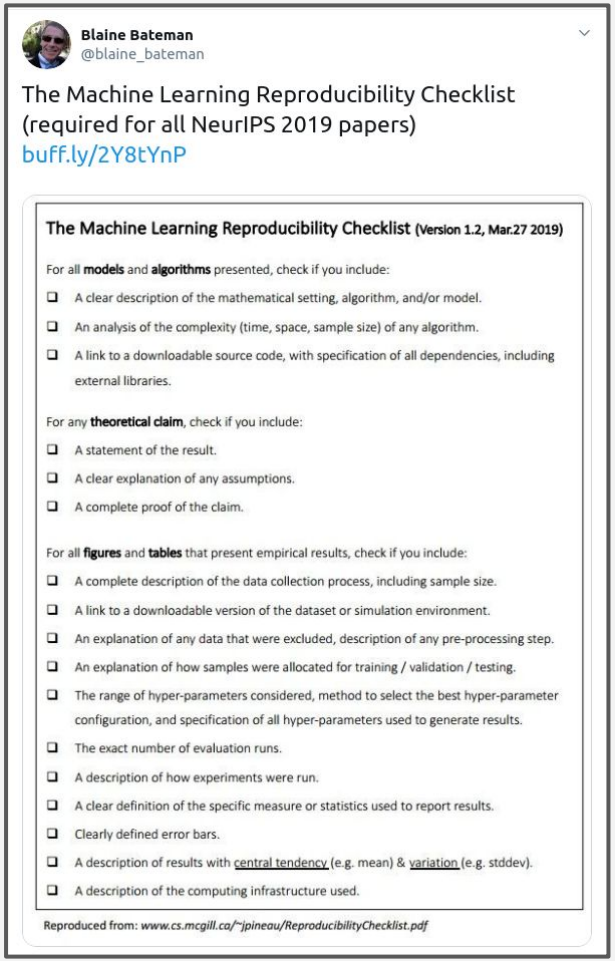




**NeurIPS** - D. Sculley, Jasper Snoek and Alex Wiltschko. Avoiding a Tragedy of the Commons in the Peer Review Process. Critiquing and Correcting Trends in Machine Learning, NeurIPS 2018 Workshop.
**ICML** - Peltarion. https://peltarion.com/article/icml-2018-an-ai-party-in-our-own-backyard
**ICLR** - @iclr_conf. Tweet on Sep 25 2019.

# Checking boxes



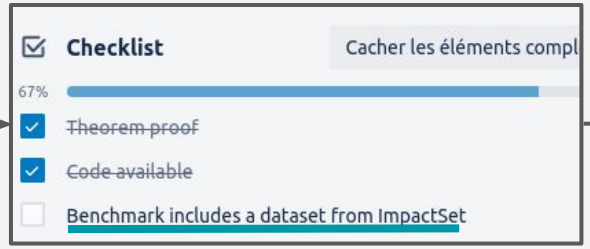




**Checklist illustration** - Trello
**NeurIPS checklist** - @blaine_bateman. Tweet on May 9 2019.

# ImpactNet

# ImpactNet by AI COMMONS


Gather the datasets




Incentivize

| | Model | Comp Acc. |
|---|---|---|
| CIFAR-10 | NASNet-A | 98.0% |
| CIFAR-100 | NASNet-A | 87.5% |
| Birdsnap | Inception-v4 | 81.8% |
| Stanford Cars | Inception-v4 | 93.4% |
| Flowers | Inception-v4 | 98.5% |
| FGVC Aircraft | Inception-v4 | 90.9% |
| Oxford-IIIT Pets | ResNet-152 | 94.5% |
| Food-101 | Inception-v4 | 90.8% |
| ImpactNet \| | | % |

AI prototype