ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Online | Mobile | Global

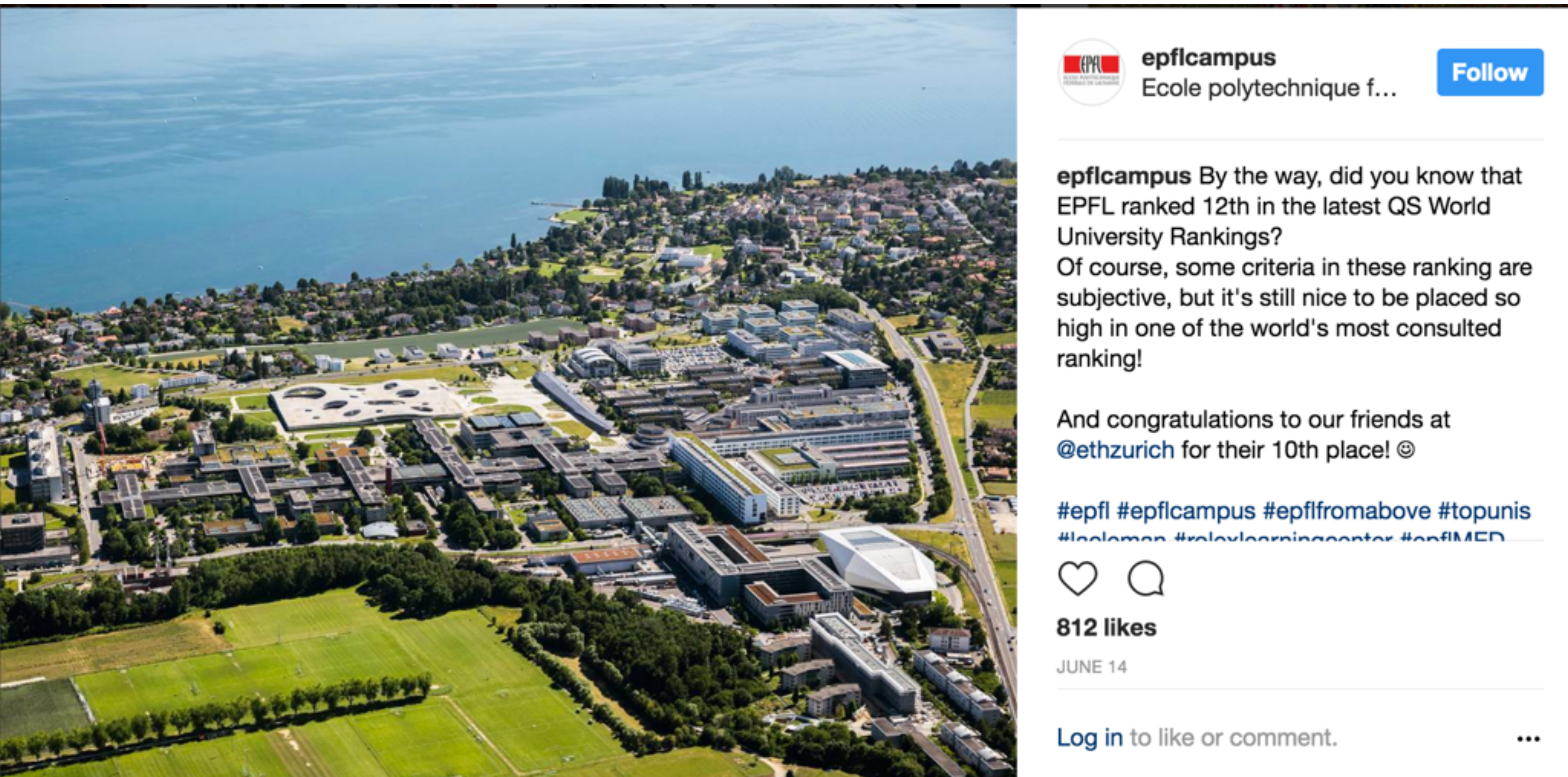
Digital Health & Epidemiology

Prof. Marcel Salathé, Digital Epidemiology Lab, EPFL

@marcelsalathe

EPFL

Swiss Federal Institute of Technology



The future is already here

it's just not very evenly distributed.

William F. Gibson

The past 10 years:

Internet + Big Data

	2007	2017

	2007	2017
	~0.5 Million	

	2007	2017
	~0.5 Million	~328 Million

	2007	2017
	~0.5 Million	~328 Million
	~20 Million	
		

	2007	2017
	~0.5 Million	~328 Million
	~20 Million	~2,000 Million

	2007	2017
	~0.5 Million	~328 Million
	~20 Million	~2,000 Million
	cloud?	
		

	2007	2017
	~0.5 Million	~328 Million
	~20 Million	~2,000 Million
amazon web services	cloud?	CLOUD!

	2007	2017
	~0.5 Million	~328 Million
	~20 Million	~2,000 Million
	cloud?	CLOUD!
		

	2007	2017
	~0.5 Million	~328 Million
	~20 Million	~2,000 Million
	cloud?	CLOUD!
		~3 B smartphones ~ 1 B tablets

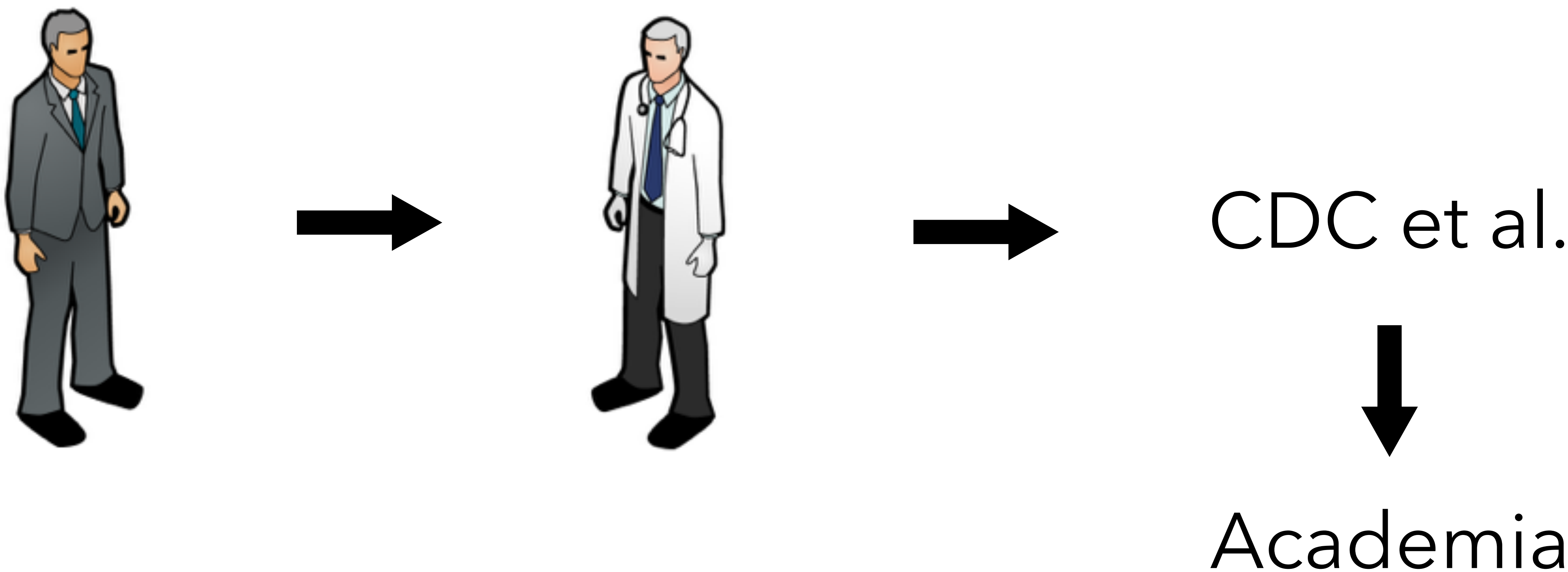
Digital Economy

March 2006		July 2017	
ExxonMobil	362.5B		756.3B
	348.5B	Alphabet	668.2B
Microsoft	279.0B	Microsoft	561.9B
citi	230.9B	amazon	471.8B
bp	225.9B		463.6B

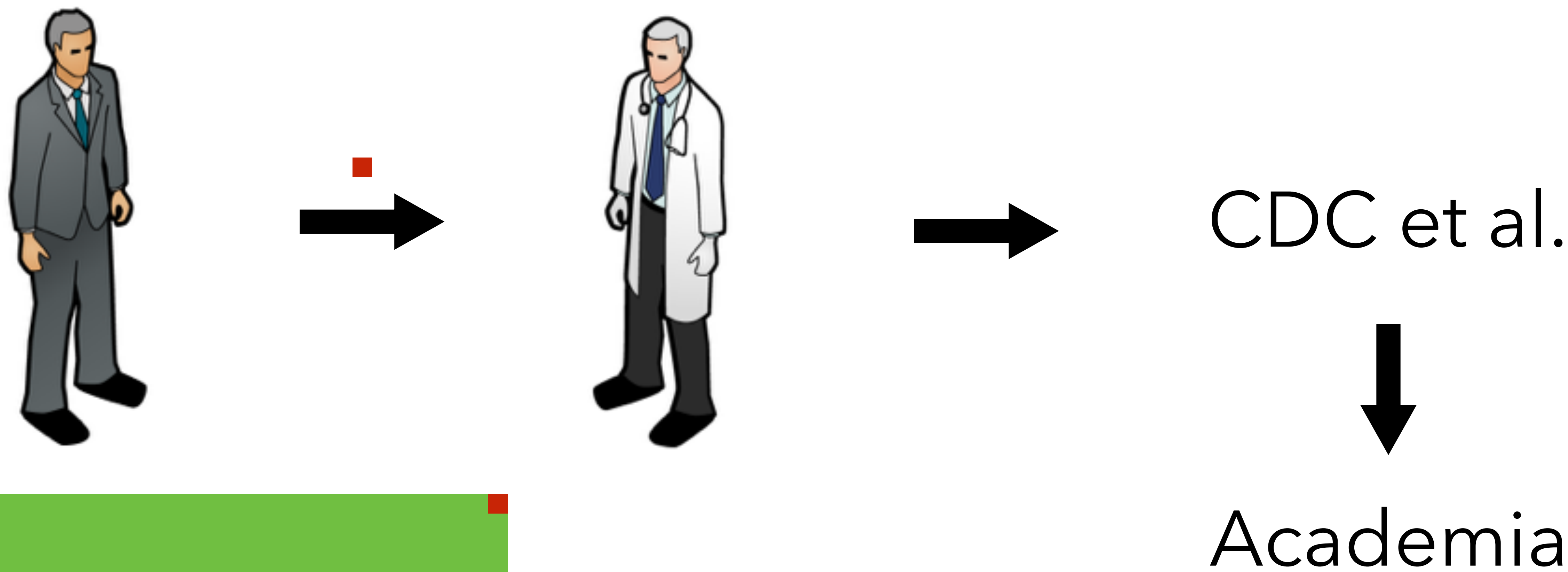
Digital Economy

March 2006	July 2017	
+1200%		756.3B
+500%	Alphabet	668.2B
+100%	Microsoft	561.9B
+3000%	amazon	471.8B
N/A (+450% since 2012)		463.6B

Traditional Epidemiology



Traditional Epidemiology

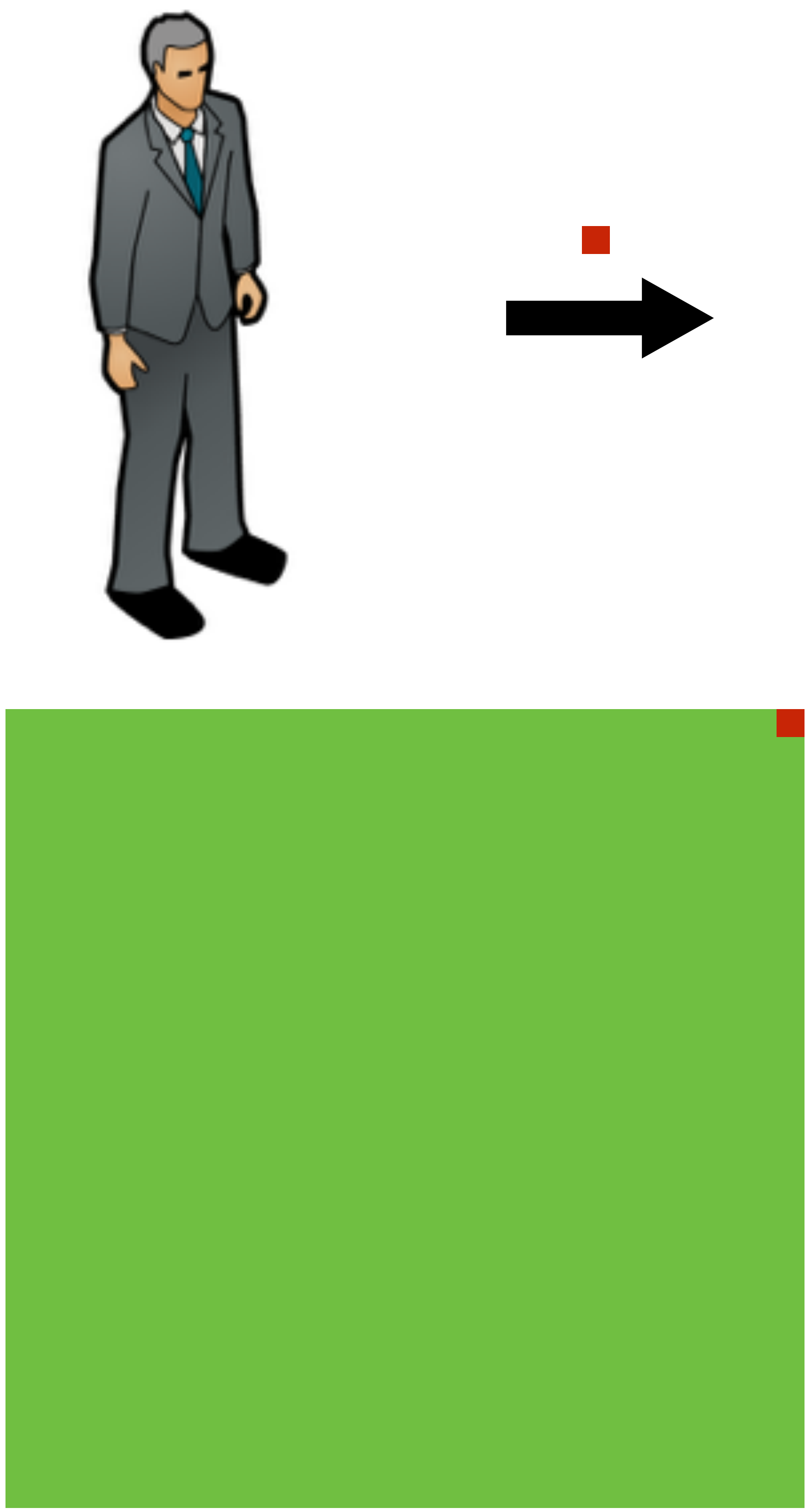


Traditional Epidemiology

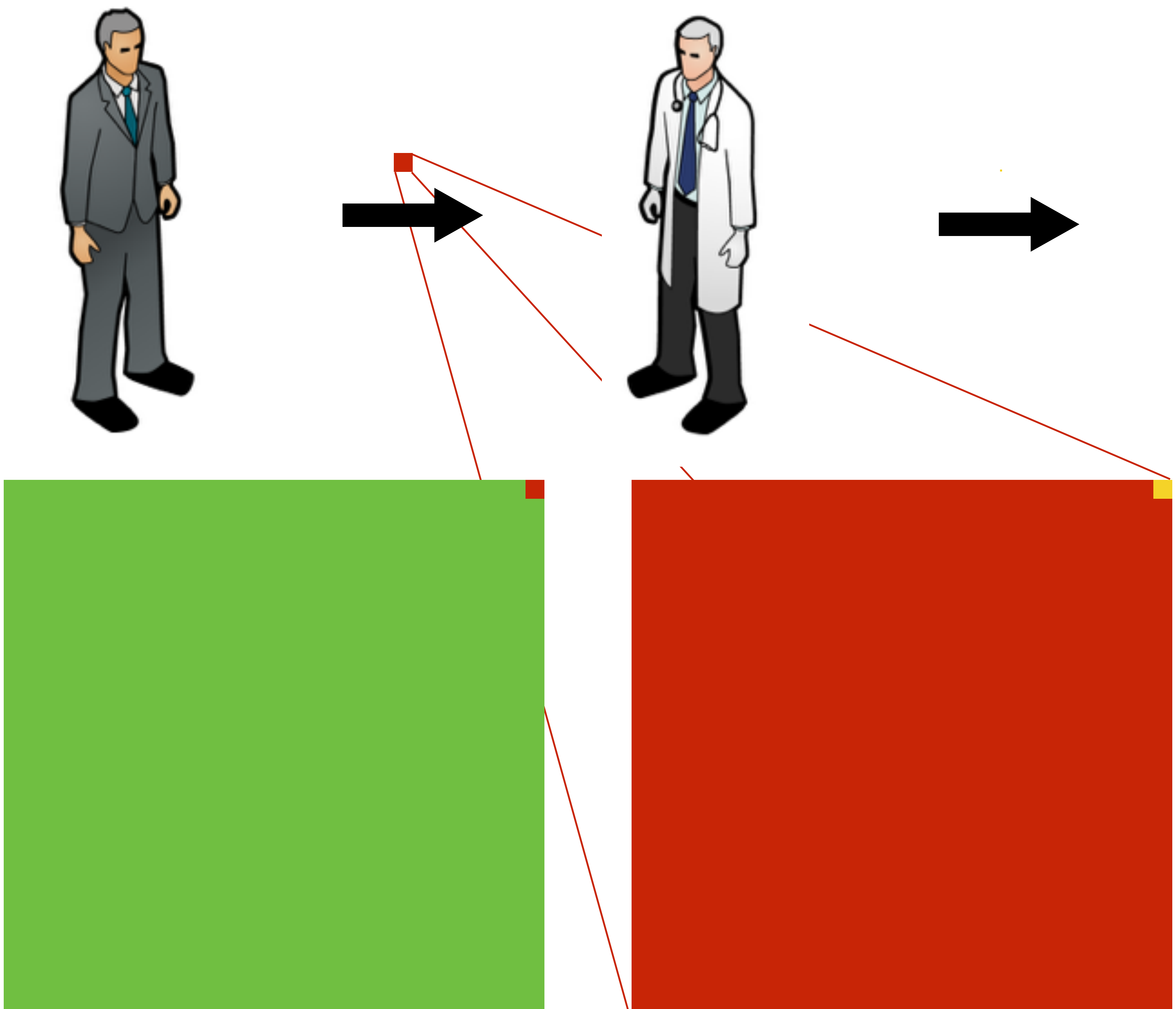
CDC et al.

Academia

Traditional Epidemiology



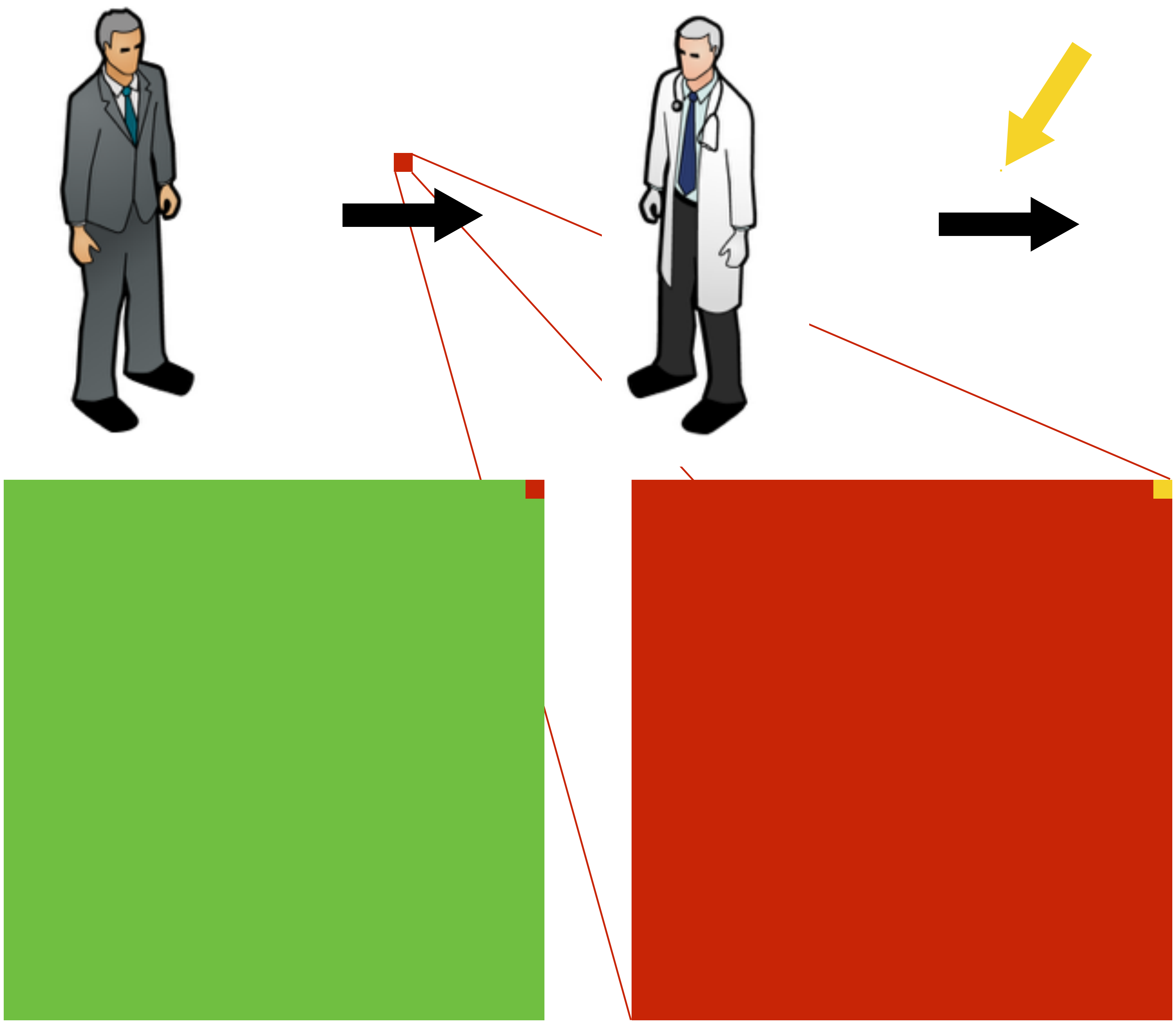
Traditional Epidemiology



CDC et al.

Academia

Traditional Epidemiology



CDC et al.

Academia

Digital Epidemiology: new data streams

"Got my flu shot this morning and now my throat is sore."

"Stomach flu & normal flu in the same month. I'm officially a germaphobe."

Text
Images
Videos
Sounds
Location
Biological data
etc.

"My weight: 170.1 lb. 10.1 lb to go. #raceweight @Withings scale auto-tweets my weight once a week http://withings.com"

"Such an upset stomach today. I hope it's just a bug and not the Truvada."

From Personalized Health…

- my DNA
- my *omics data
- my location data
- my activity data (heart rate, etc.)
- my lab tests
- what I ate
- how I slept
- how I feel
- my health history
- etc.

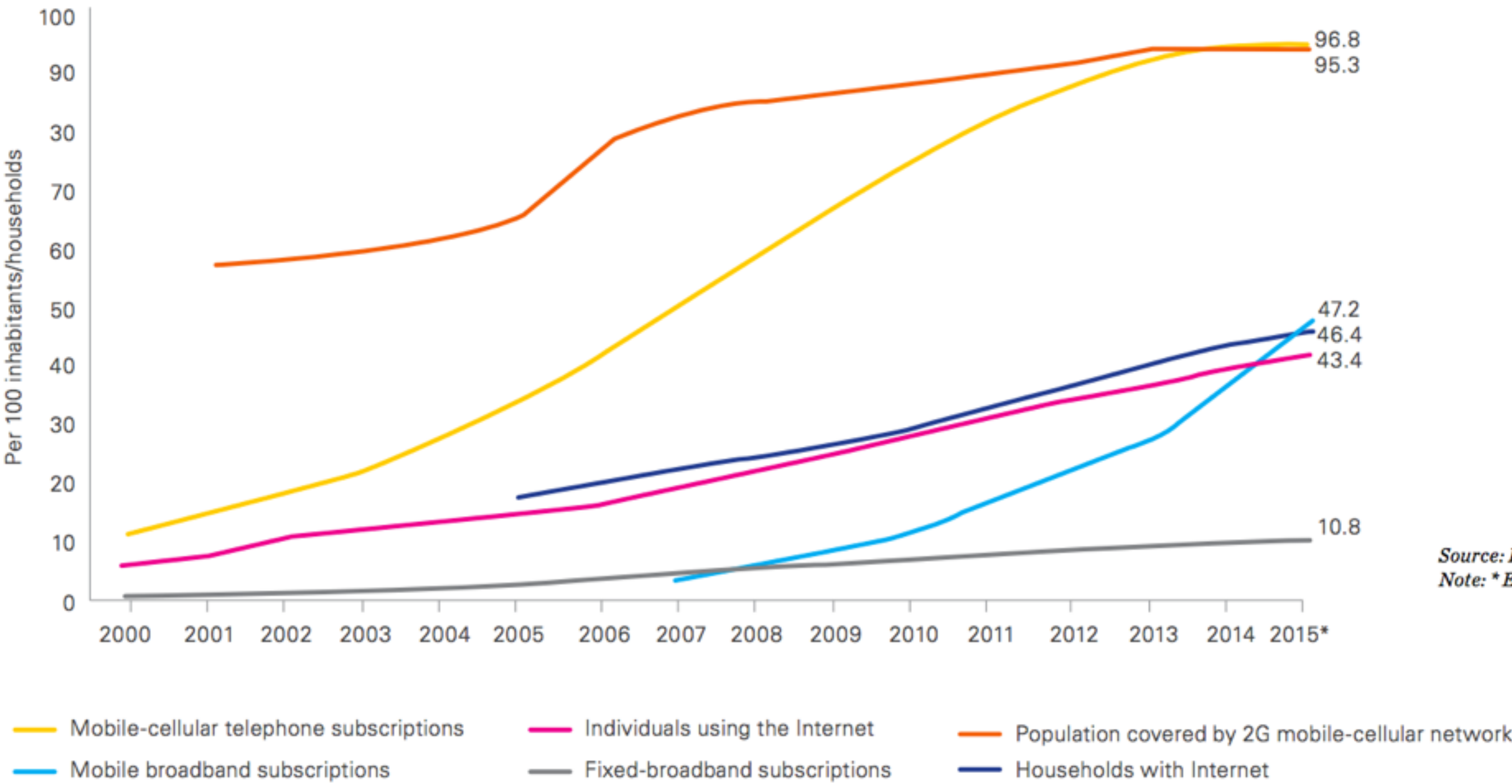
"The patient will see you now"

…To Truly Global Health

shared
via mobile
apps

2021:
6+ Billion
Smartphones

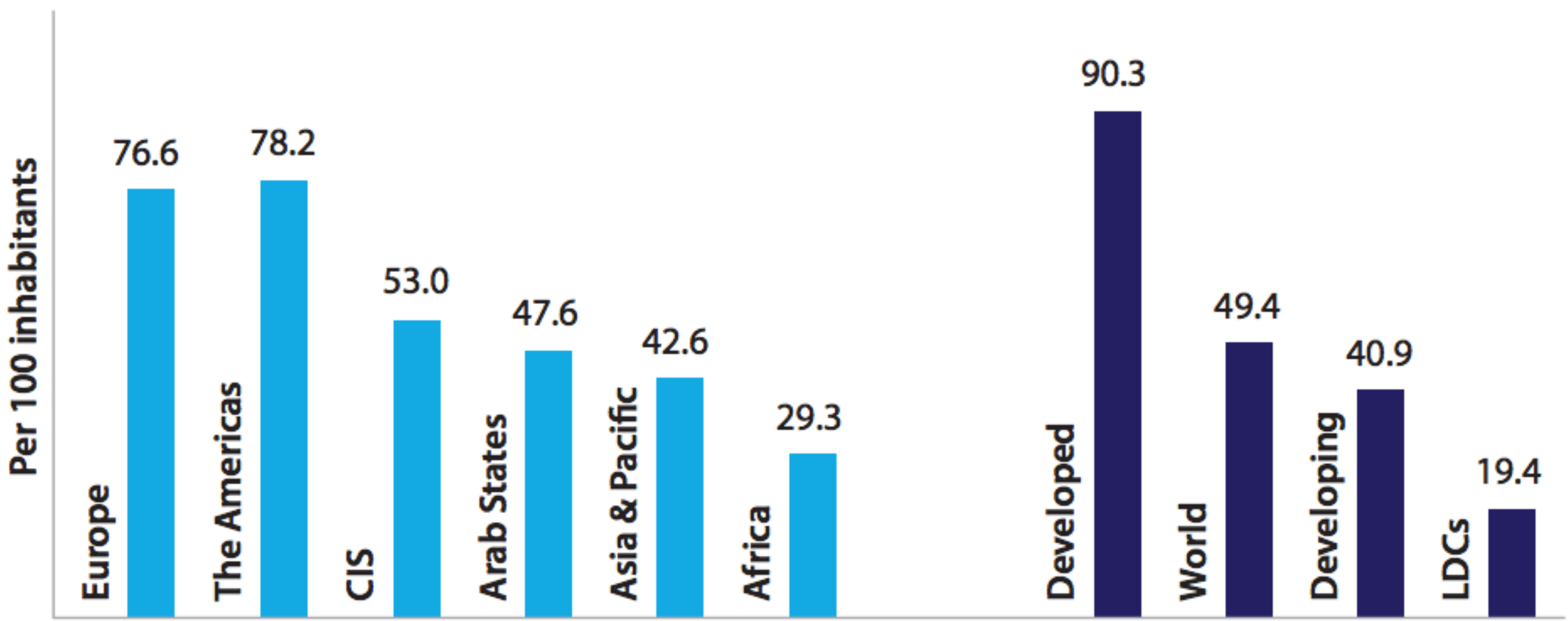
Mobile broadband



Source: ITU.
*Note: * Estimates.*

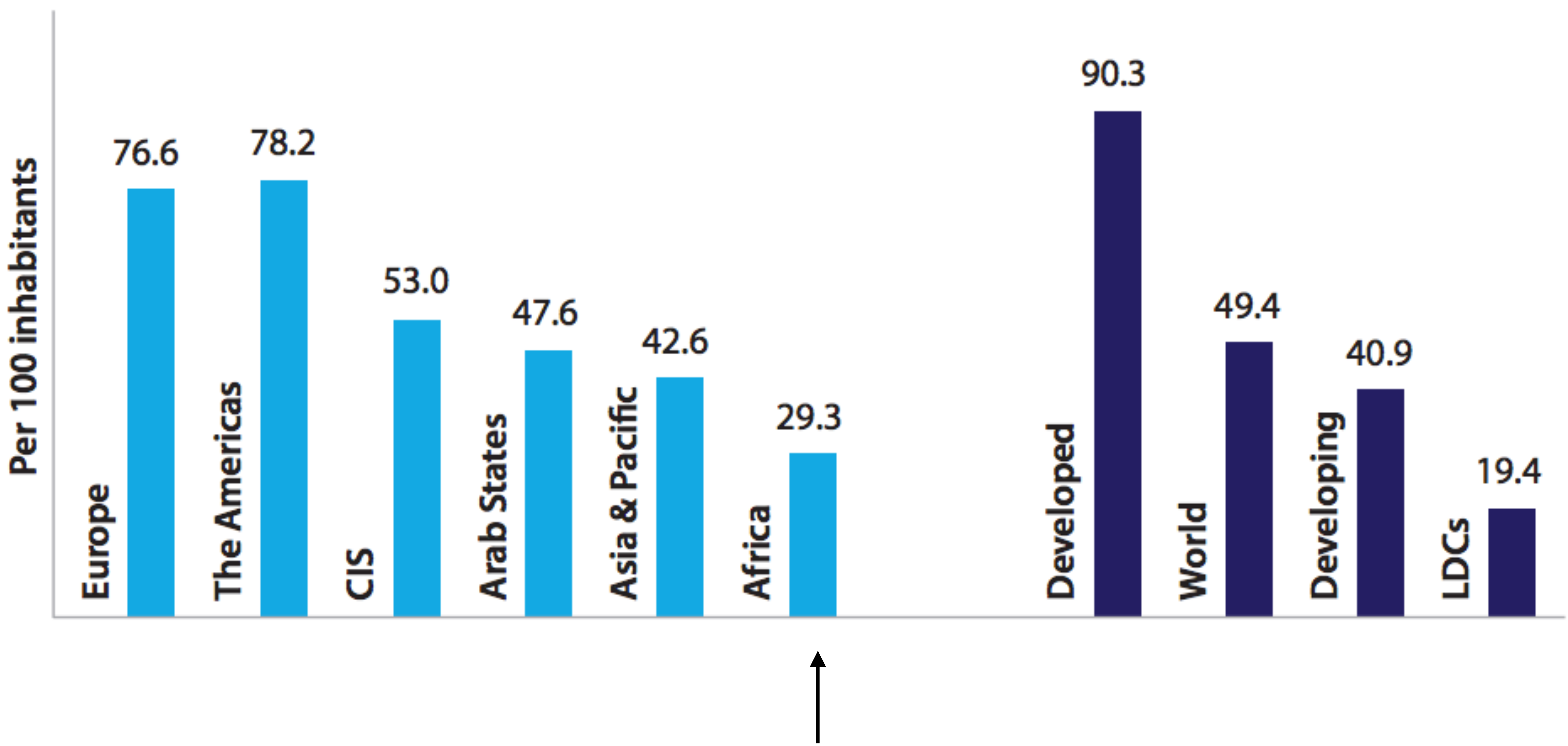
https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf

Mobile broadband (2016)



http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf

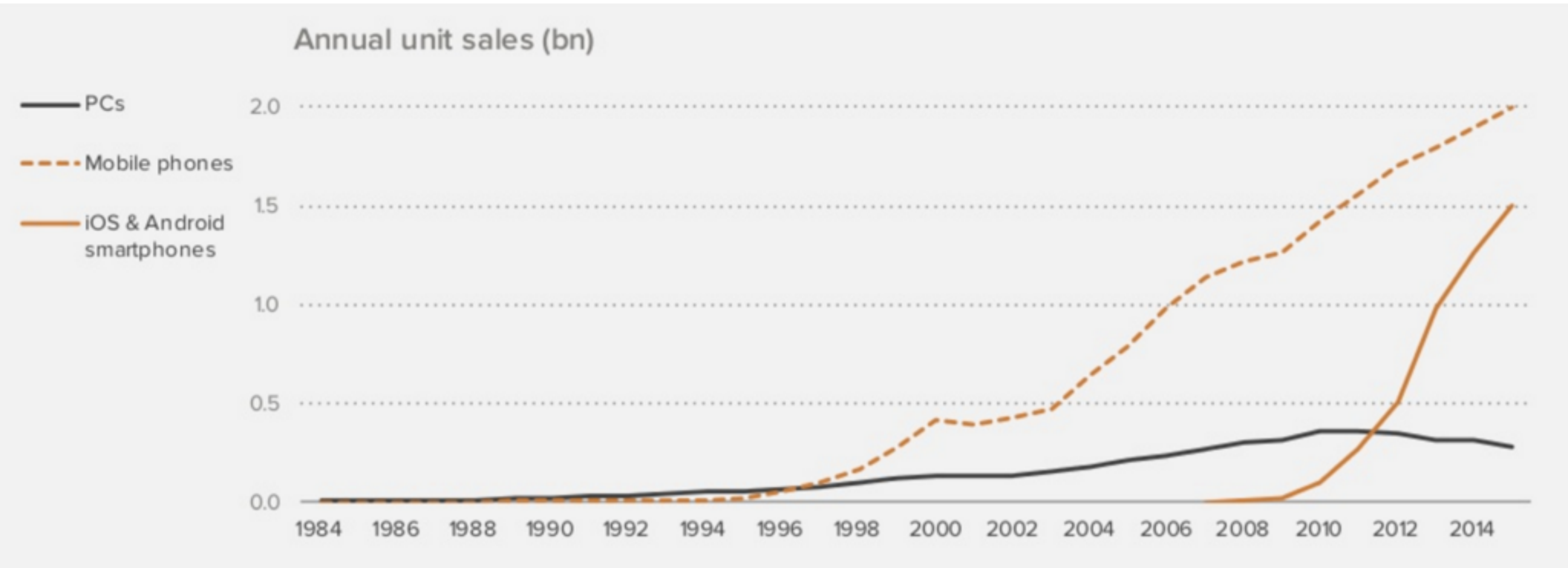
Mobile broadband (2016)



2015: 17.4

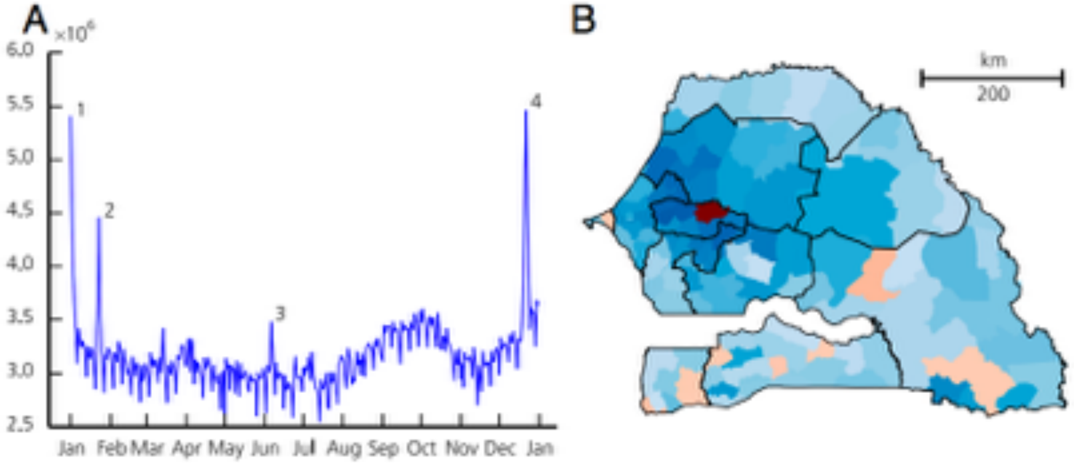
http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf

Mobile phone & smart phone revolutions dwarf the PC revolution



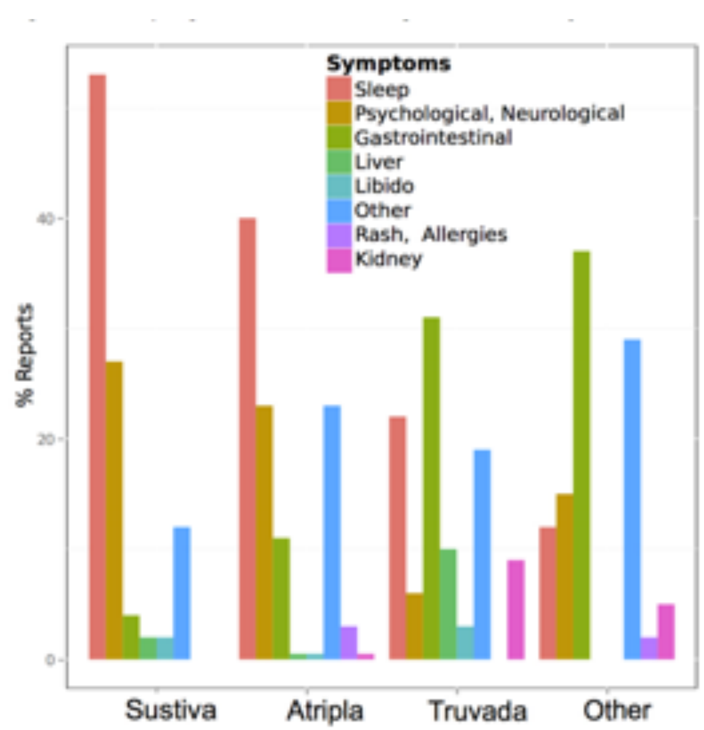
http://ben-evans.com/benedictevans/?tag=Presentations

Mobile phones - Track disease outbreaks
Finger et al 2016

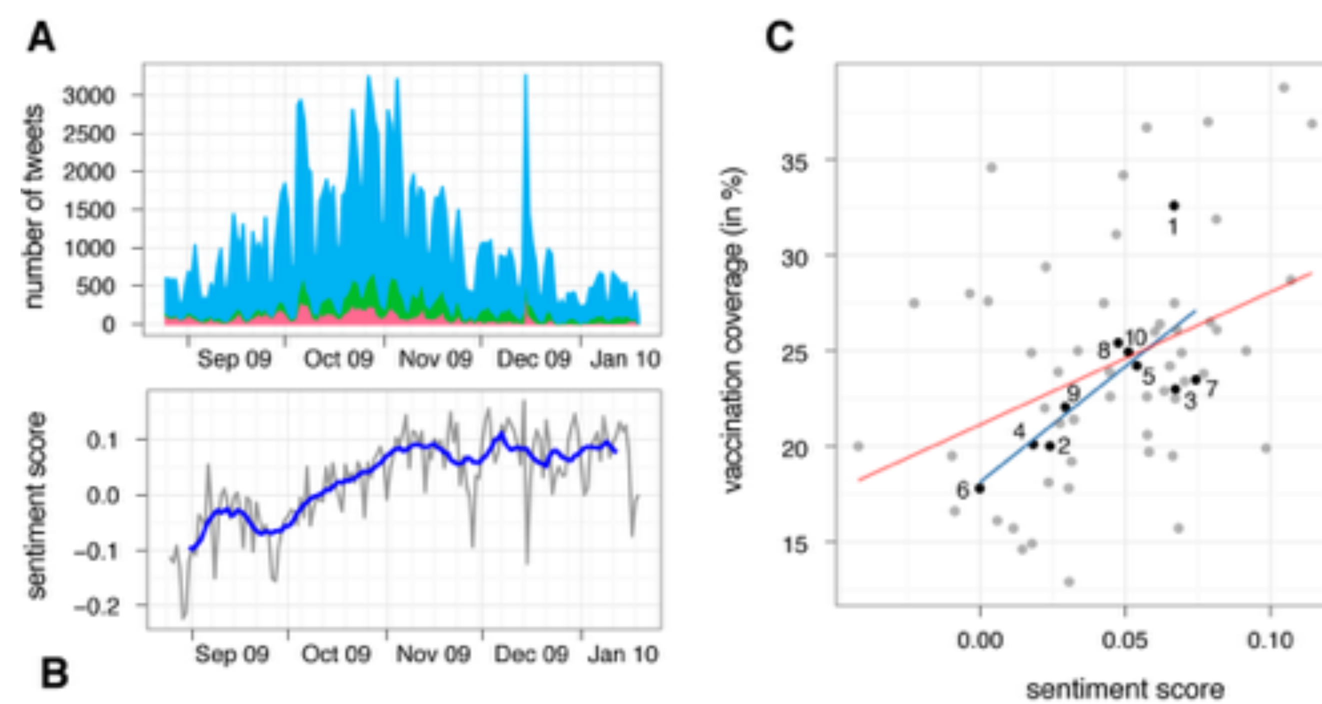


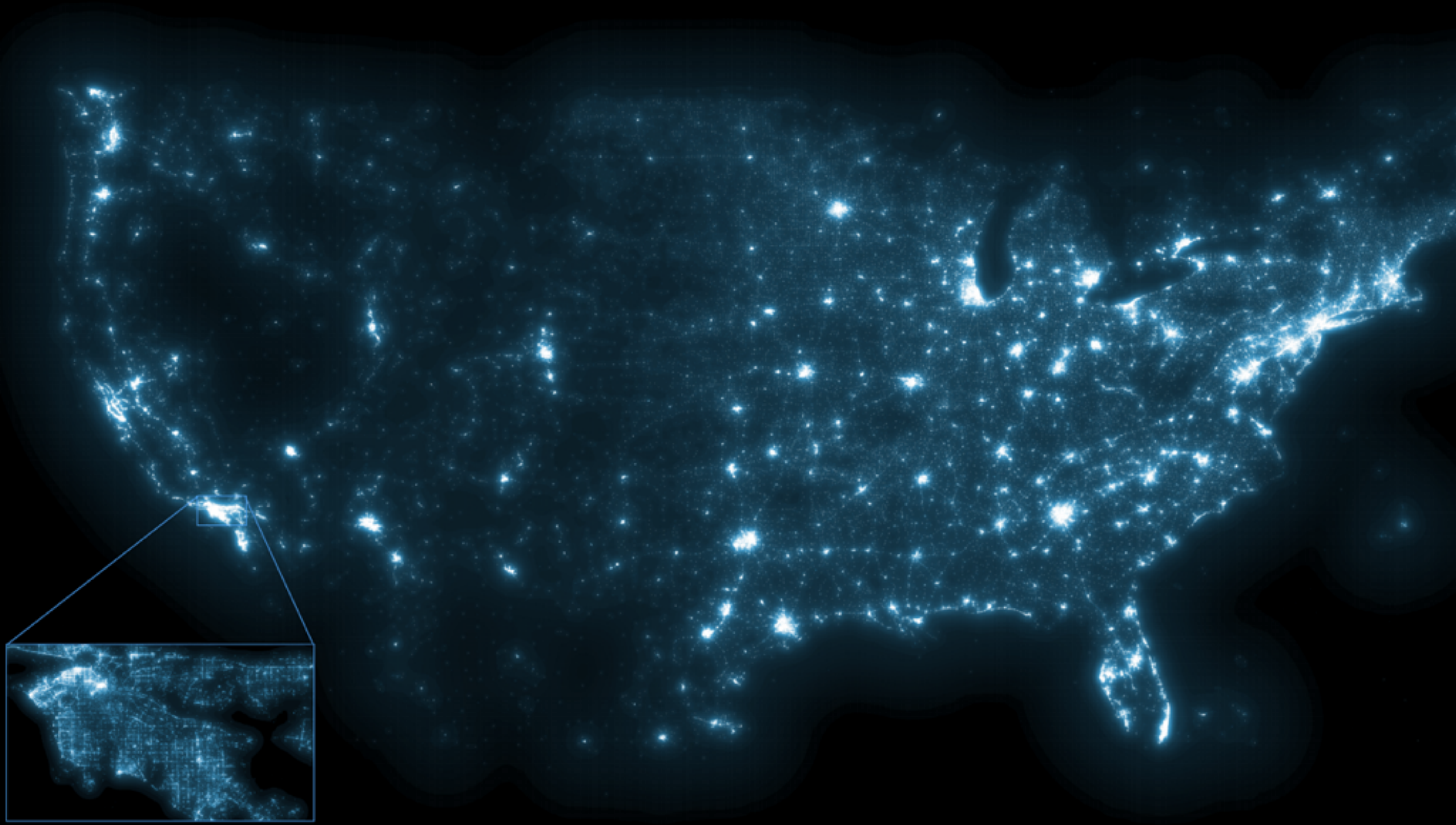
Wikipedia - Influenza forecasting
McIver & Brownstein 2014

Twitter - Pharmacovigilance
Adrover et al 2015

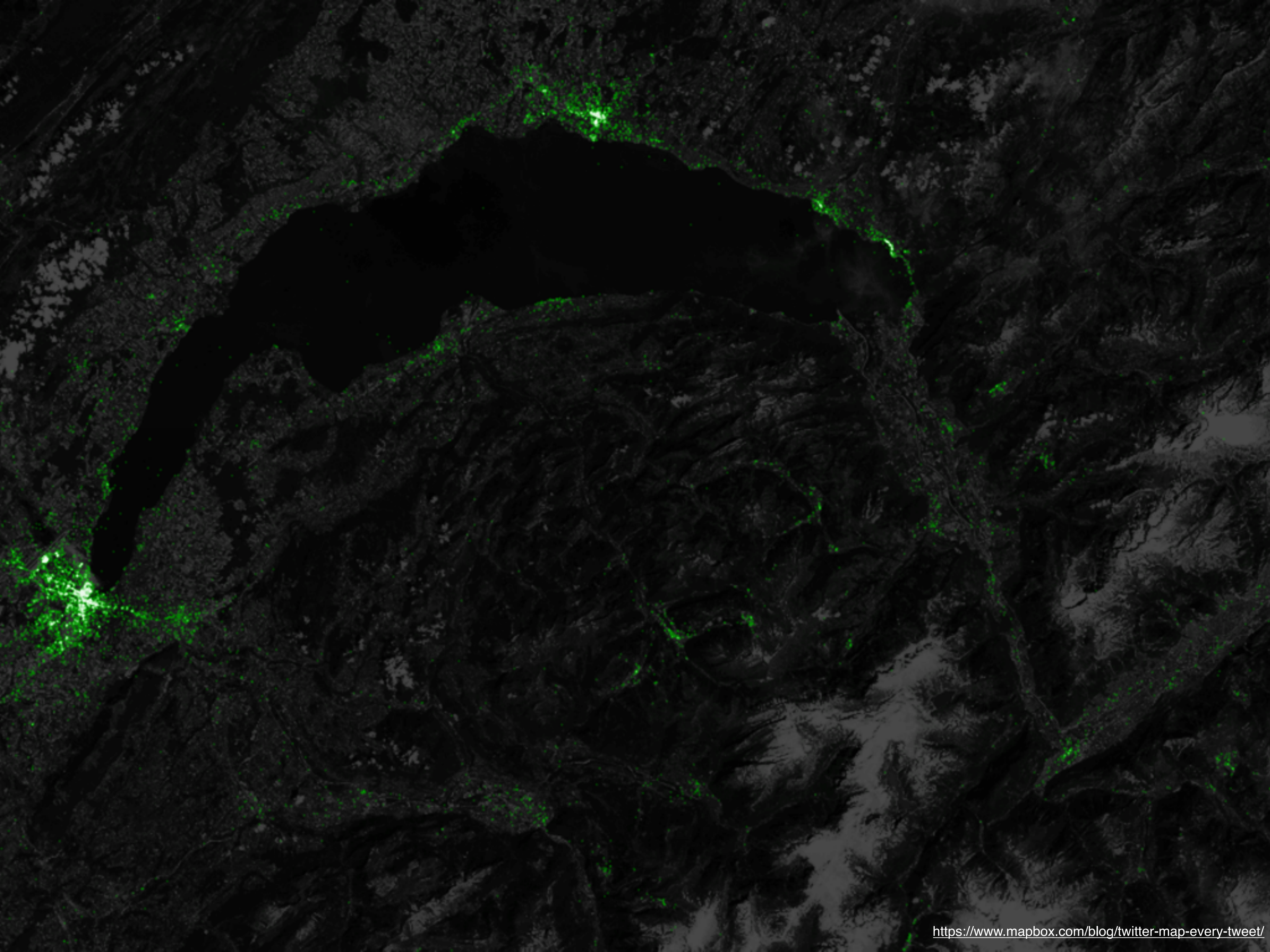


Twitter - Vaccine uptake
Salathé & Khandelwal, 2011





Salathé et al. PLOS Computational Biology 2012



https://www.mapbox.com/blog/twitter-map-every-tweet/

https://www.mapbox.com/blog/twitter-map-every-tweet/

The next 10 years:

Machine Learning

(and IoT, blockchain, wearables, etc.)

PlantVillage: Machine Learning for Disease Recognition

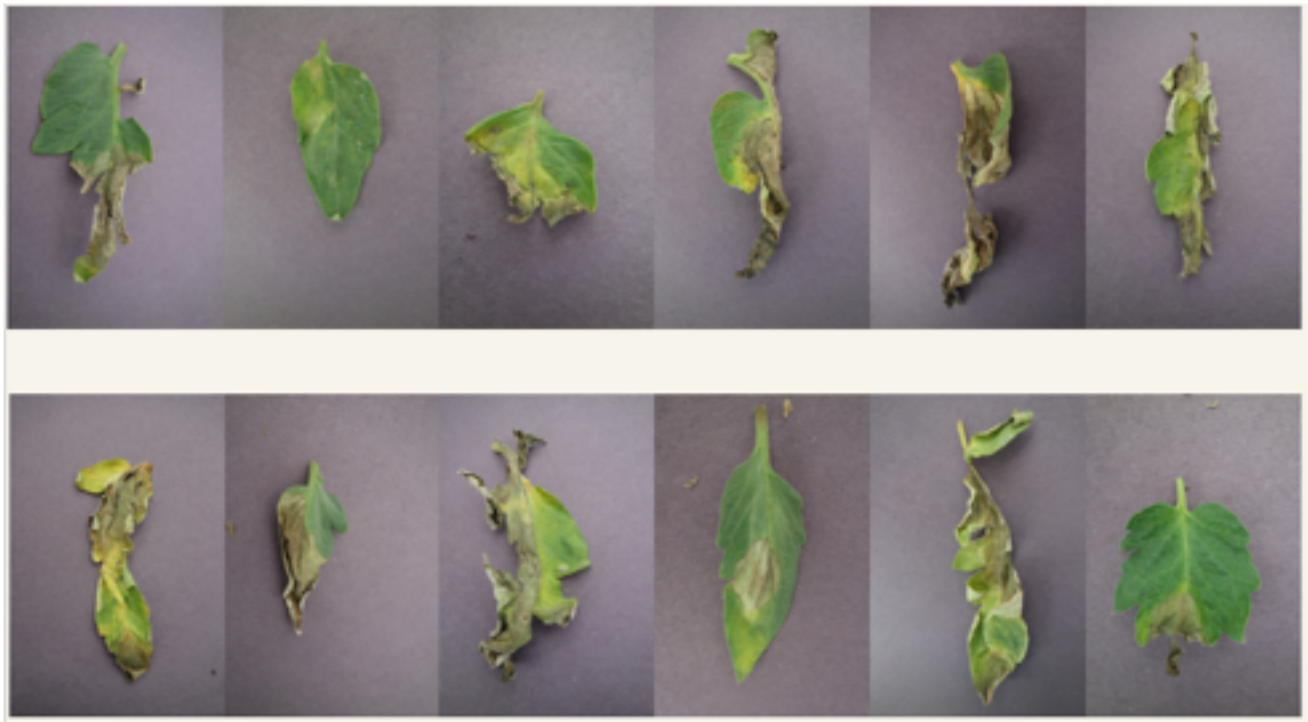


Image Data

Collecting 1M+ labelled images as training set for machine learning algorithm development

Machine Learning

Crowdsourced, open machine learning competitions based on open access images

[Collaboration with Penn State, Prof. David Hughes]

METHODS ARTICLE

Front. Plant Sci. | doi: 10.3389/fpls.2016.01419

Using Deep Learning for Image-Based Plant Disease Detection

Sharada P. Mohanty[1], David P. Hughes[2] and Marcel Salathé[1*]

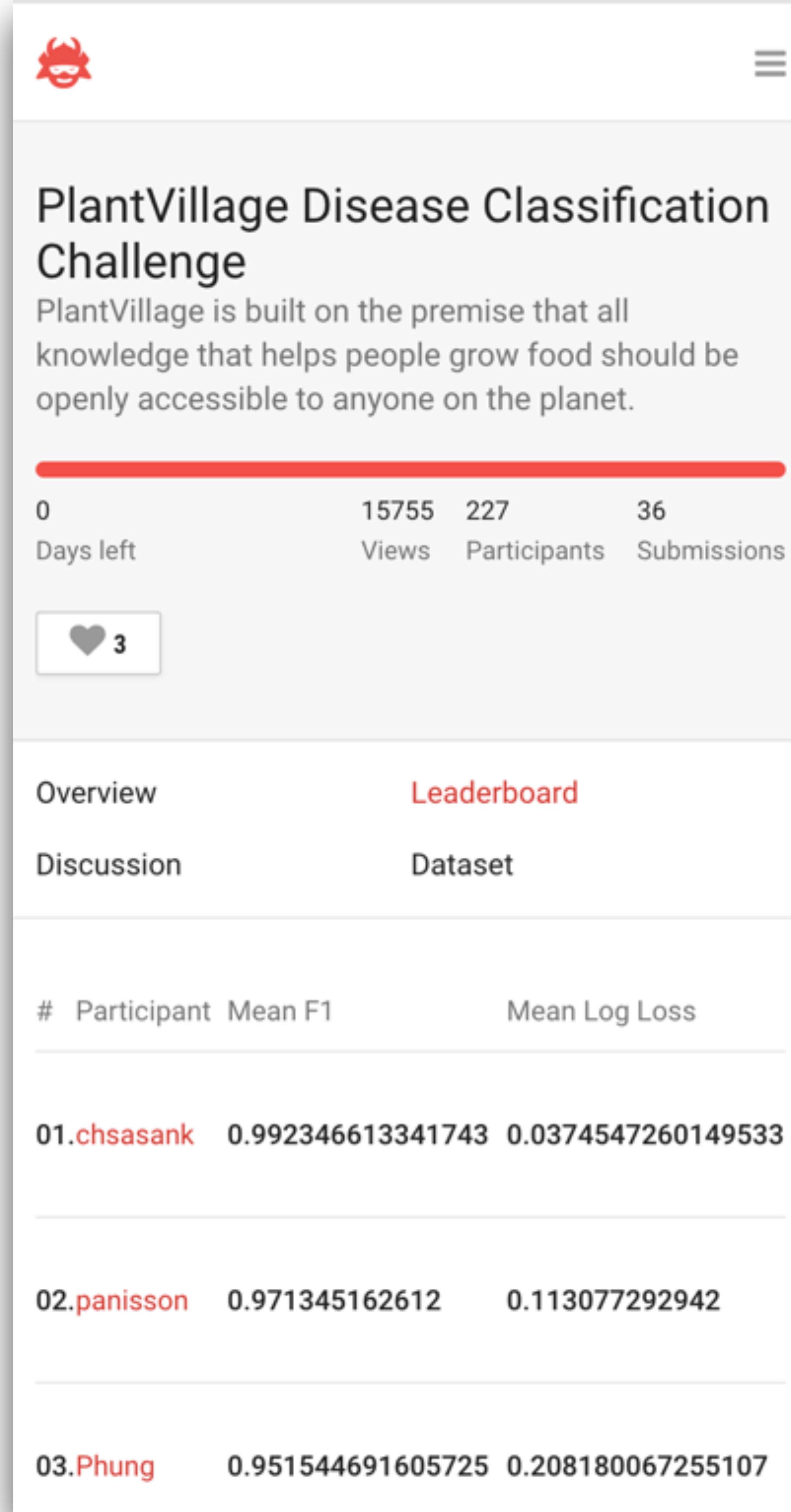
[1]EPFL, Switzerland

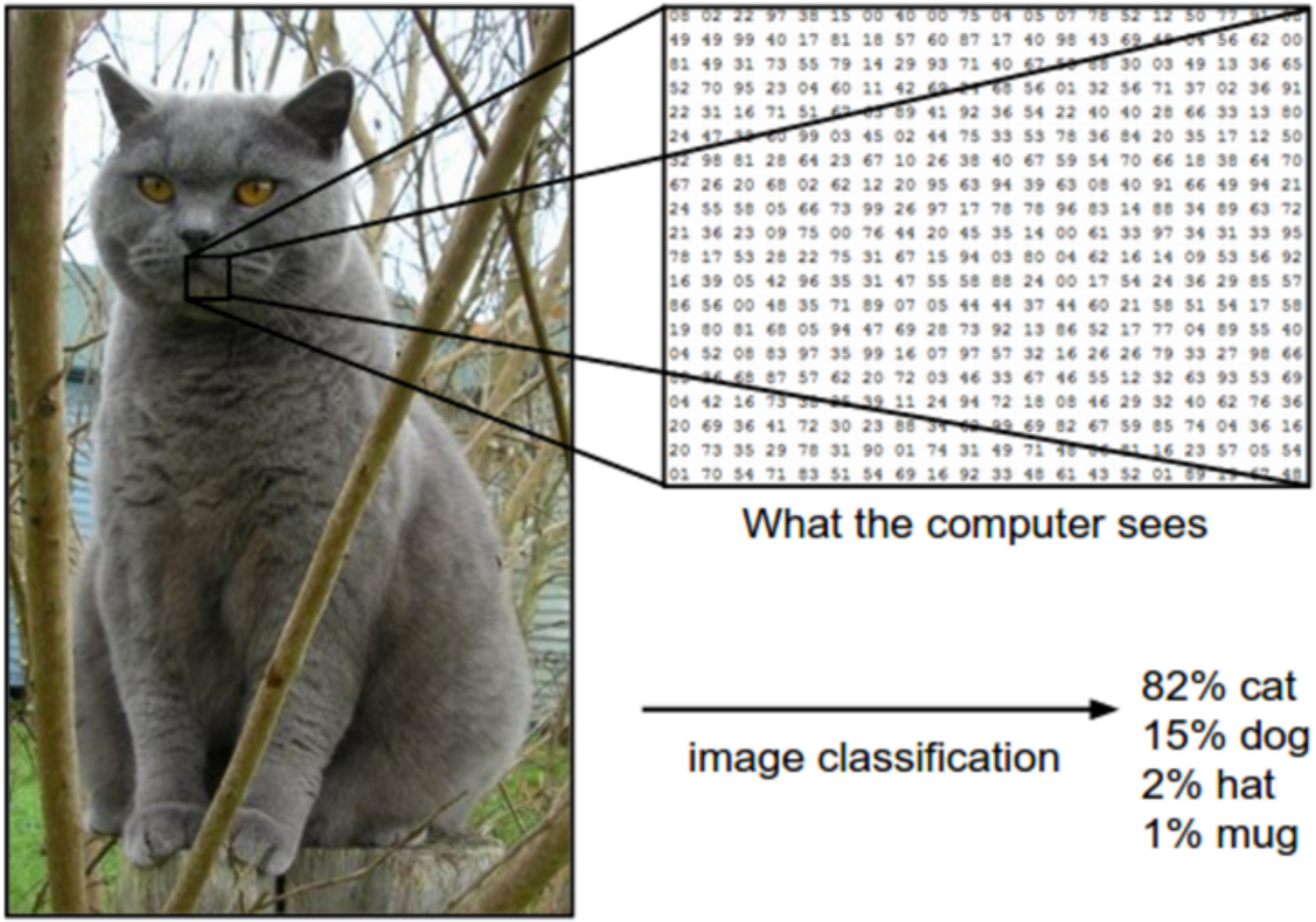
[2]Penn State University, USA

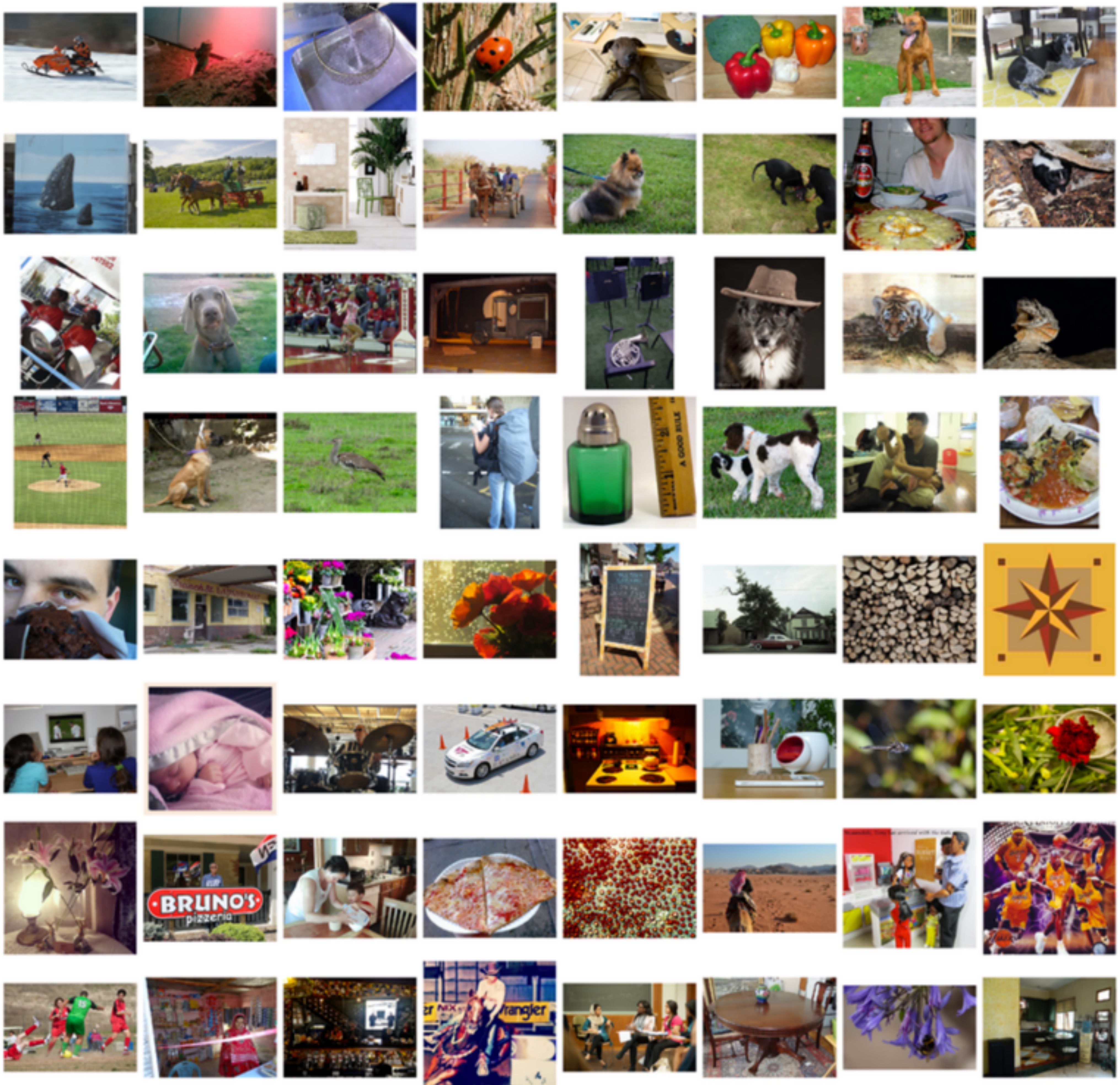
Crop diseases are a major threat to food security, but their rapid identification remains difficult in many parts of the world due to the lack of the necessary infrastructure. The combination of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning has paved the way for smartphone-assisted disease diagnosis. Using a public dataset of 54,306 images of diseased and healthy plant leaves collected under controlled conditions, we train a deep convolutional neural network to identify 14 crop species and 26 diseases (or absence thereof). The trained model achieves an accuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach. Overall, the approach of training deep learning models on increasingly large and publicly available image datasets presents a clear path towards smartphone-assisted crop disease diagnosis on a massive global scale.

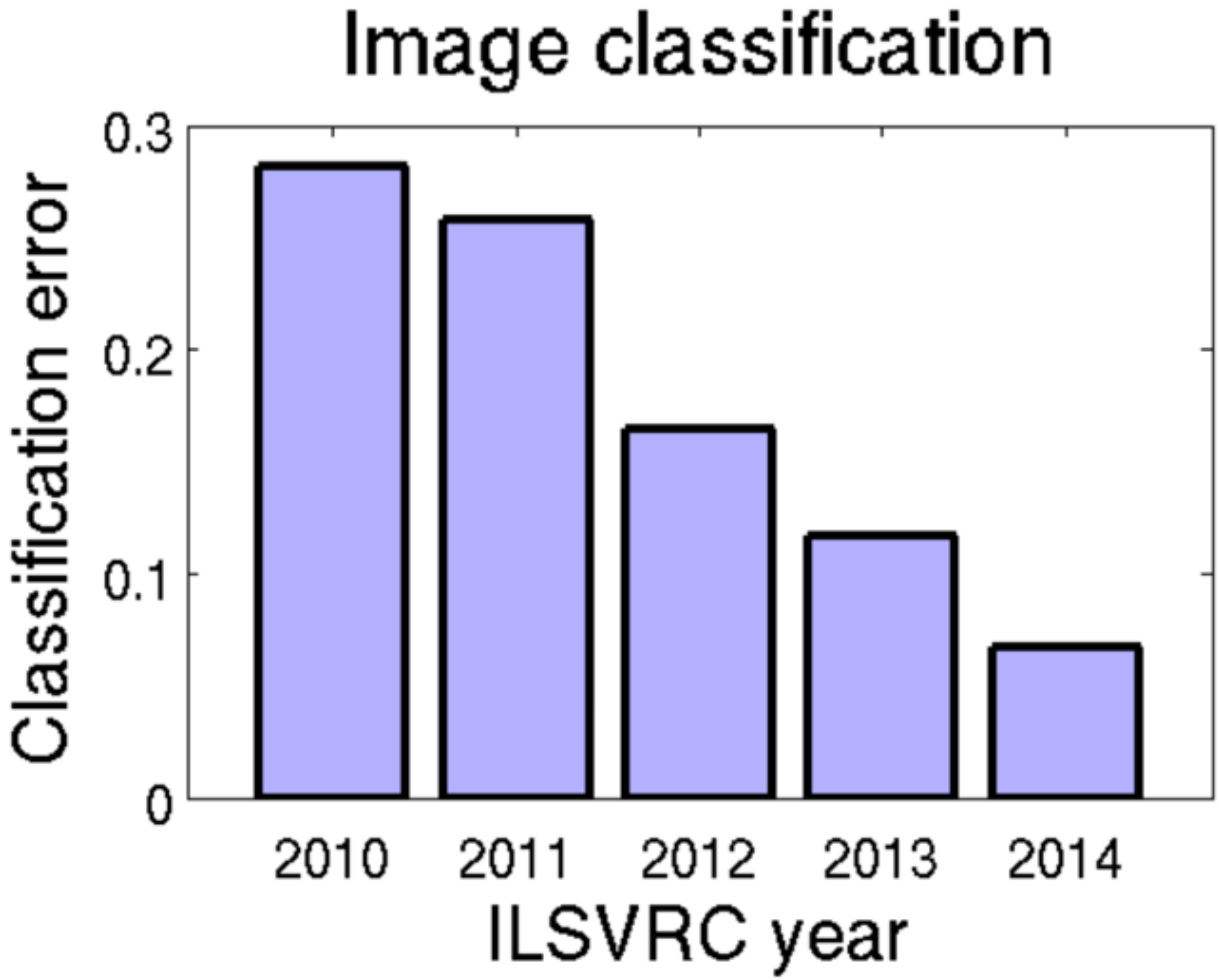
Crowdsourced, open machine learning competitions based on open access images (www.crowdAI.org)

matched published results →

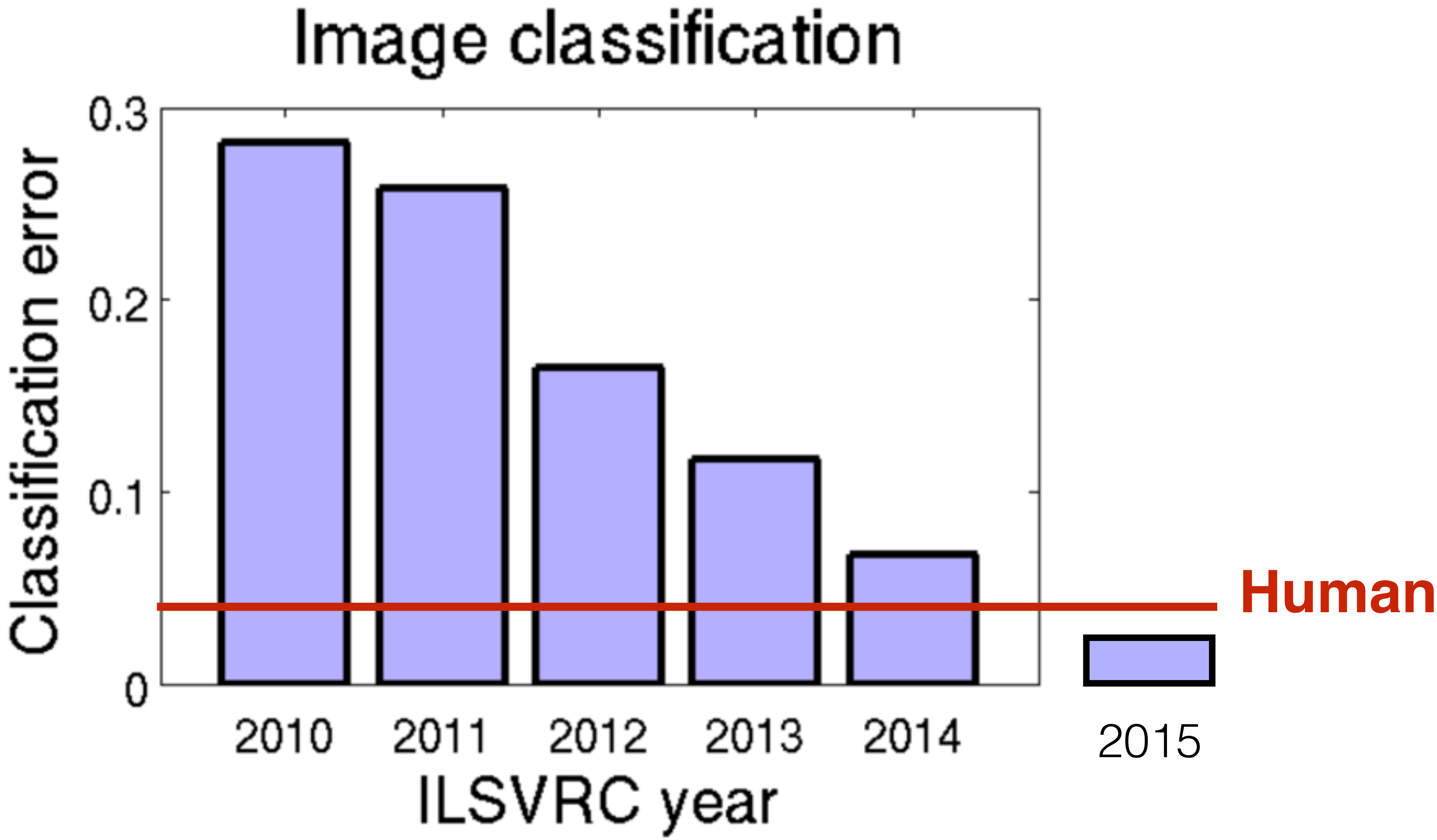



What the computer sees
image classification
82% cat
15% dog
2% hat
1% mug


BRUNO'S
pizzeria

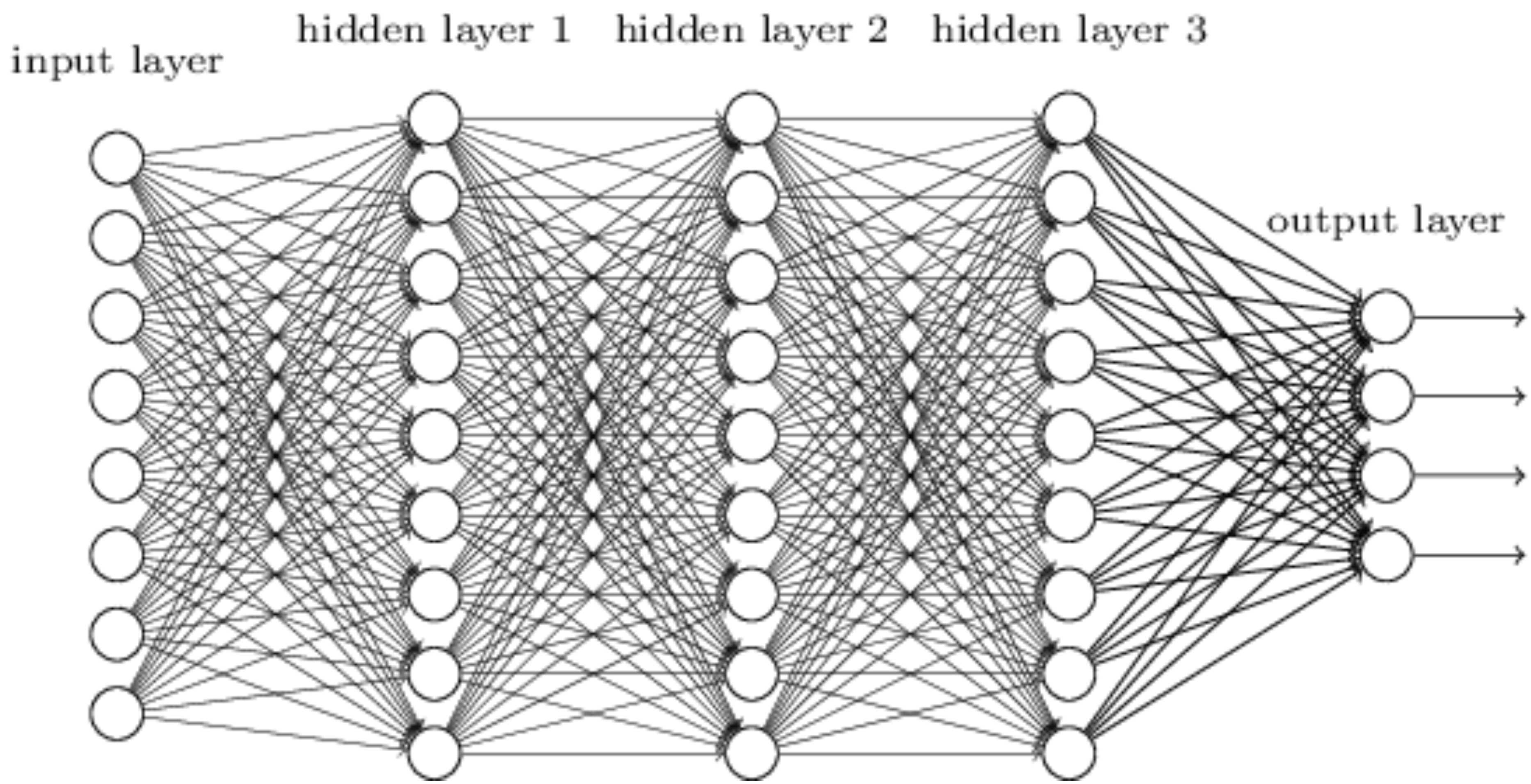


https://arxiv.org/pdf/1409.0575v3



https://arxiv.org/pdf/1409.0575v3

What is deep learning

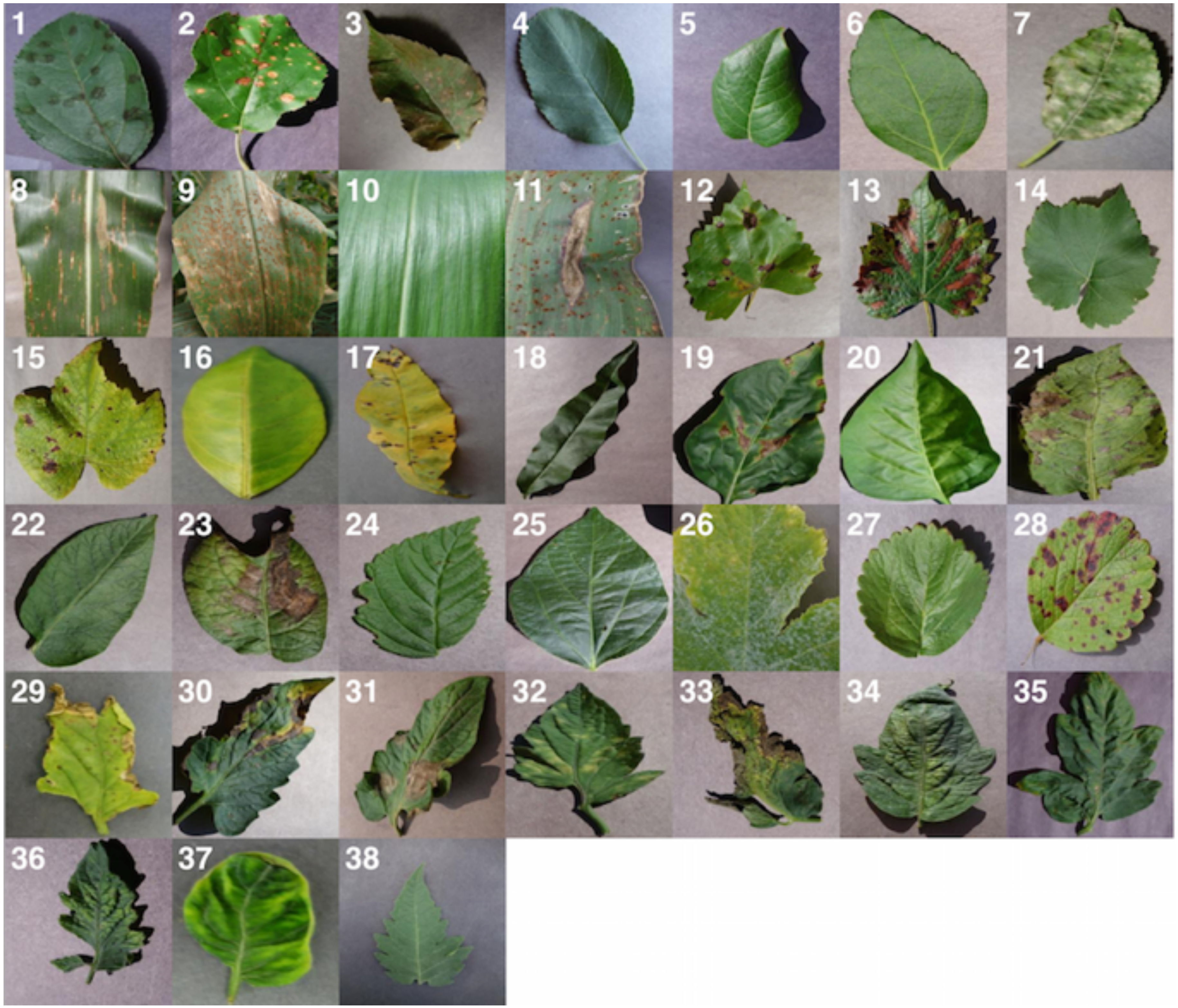


What is deep learning

- The more data, the better
- Training is computationally very expensive - multiple hours or days on GPU clusters
- Once model is trained, it runs in ~1-2 seconds on CPU, no internet connection required.

Table 1. Examples of ML Approaches in Plant Species for Stress Phenotyping

ML Algorithm Application in HTSP	ML Algorithm Type	Sensor	Plant Species	Trait(s) Phenotyped	Stress Type	Refs
Identification	SVM with a linear kernel	Thermal and stereo visible light	Tomato (*Solanum lycopersicum* L.)	Powdery mildew	Disease	[31]
Identification	SAM	Remote sensing	Sugar beet (*Beta vulgaris* L.)	*Heterodera schachtii* and *Rhizoctonia solani*	Pest and disease	[44]
Identification	None Preprocessing via segmentation	Kinect RGB depth images	Apple (*Malus domestica* Borkh.)	Apple scab	Disease	[70]
Identification	SVM and Gaussian processes classifier (GPC)	Visible and thermal images	Spinach (*Spinacia oleracea* L.)	Drought/water stress	Abiotic stress	[71]
Identification	Bayes factor and DAR	Hyperspectral images	Barley (*Hordeum vulgare* L.)	Rust, net blotch, and powdery mildew	Disease	[11]
Identification	SVM	Fluorescence imaging spectroscopy	Citrus [*Citrus sinensis* (L.) Osbeck]	Huanglongbing (HLB)	Disease	[36]
Identification	OBIA-based classification	UAV-based RGB images and multispectral image	Sunflower (*Helianthus annuus* L.)	Weed	Biotic stress	[21]
Identification	None Preprocessing via segmentation	RGB images	Cotton (*Gossypium hirsutum* L.)	Southern green stink bug, bacterial angular and *Ascohyta* blight	Disease and insect	[39]
Identification	SVM, linear kernel, quadratic kernel (QP), radial basis function (RBF), multilayer perceptron (MLP), and polynomial kernel	RGB images	Tomato	Tomato yellow leaf curl virus and tomato yellow leaf curl disease	Disease	[40]
Identification	ANN variant	RGB images	Orchid (*Phalaenopsis*)	Bacterial soft rot, *Phythopthora* black rot, bacterial brown spot	Disease	[42]
Identification	SVM	UAV- and aircraft-based sensors	Citrus	Huanglongbing (HLB)	Disease	[37]



54,306 images

14 crops

26 diseases

38 classes

Now:

70,000+ images

17 crops

33 diseases

47 classes

Mohanty et al. 2016

PlantVillage: Machine Learning for Disease Recognition

crowdsourcing

Image Recognition
(Machine Learning)

Diagnosis
Treatment Suggestions

LETTER

doi:10.1038/nature21056

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva[1]*, Brett Kuprel[1]*, Roberto A. Novoa[2,3], Justin Ko[2], Susan M. Swetter[2,4], Helen M. Blau[5] & Sebastian Thrun[6]

Skin cancer, the most common human malignancy[1–3], is primarily diagnosed visually, beginning with an initial clinical screening and followed potentially by dermoscopic analysis, a biopsy and histopathological examination. Automated classification of skin lesions using images is a challenging task owing to the fine-grained variability in the appearance of skin lesions. Deep convolutional neural networks (CNNs)[4,5] show potential for general and highly variable tasks across many fine-grained object categories[6–11]. Here we demonstrate classification of skin lesions using a single CNN, trained end-to-end from images directly, using only pixels and disease labels as inputs. We train a CNN using a dataset of 129,450 clinical images—two orders of magnitude larger than previous datasets[12]—consisting of 2,032 different diseases. We test its performance against 21 board-certified dermatologists on biopsy-proven clinical images with two critical binary classification use cases: keratinocyte carcinomas versus benign seborrheic keratoses; and malignant melanomas versus benign nevi. The first case represents the identification of the most common cancers, the second represents the identification of the deadliest skin cancer. The CNN achieves performance on par with all tested experts across both tasks, demonstrating an artificial intelligence capable of classifying skin cancer with a level of competence comparable to dermatologists. Outfitted with deep neural networks, mobile devices can potentially extend the reach of dermatologists outside of the clinic. It is projected that 6.3 billion smartphone subscriptions will exist by the year 2021 (ref. 13) and can therefore potentially provide low-cost universal access to vital diagnostic care.

There are 5.4 million new cases of skin cancer in the United States[2] every year. One in five Americans will be diagnosed with a cutaneous

images (for example, smartphone images) exhibit variability in factors such as zoom, angle and lighting, making classification substantially more challenging[23,24]. We overcome this challenge by using a data-driven approach—1.41 million pre-training and training images make classification robust to photographic variability. Many previous techniques require extensive preprocessing, lesion segmentation and extraction of domain-specific visual features before classification. By contrast, our system requires no hand-crafted features; it is trained end-to-end directly from image labels and raw pixels, with a single network for both photographic and dermoscopic images. The existing body of work uses small datasets of typically less than a thousand images of skin lesions[16,18,19], which, as a result, do not generalize well to new images. We demonstrate generalizable classification with a new dermatologist-labelled dataset of 129,450 clinical images, including 3,374 dermoscopy images.

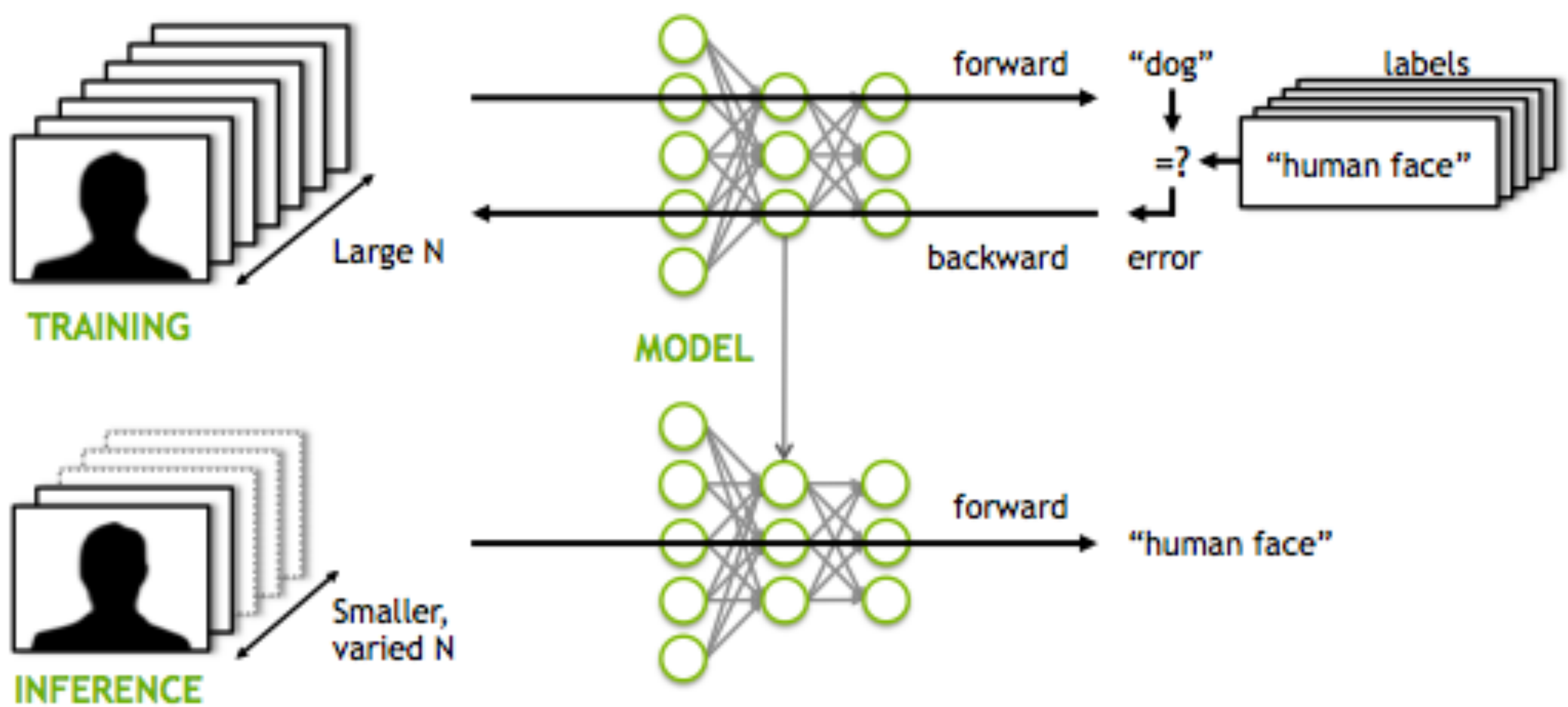
Deep learning algorithms, powered by advances in computation and very large datasets[25], have recently been shown to exceed human performance in visual tasks such as playing Atari games[26], strategic board games like Go[27] and object recognition[6]. In this paper we outline the development of a CNN that matches the performance of dermatologists at three key diagnostic tasks: melanoma classification, melanoma classification using dermoscopy and carcinoma classification. We restrict the comparisons to image-based classification.

We utilize a GoogleNet Inception v3 CNN architecture[9] that was pre-trained on approximately 1.28 million images (1,000 object categories) from the 2014 ImageNet Large Scale Visual Recognition Challenge[6], and train it on our dataset using transfer learning[28]. Figure 1 shows the working system. The CNN is trained using 757 disease classes. Our dataset is composed of dermatologist-labelled images organized in a

And beyond…

AI & Society

Deep Learning



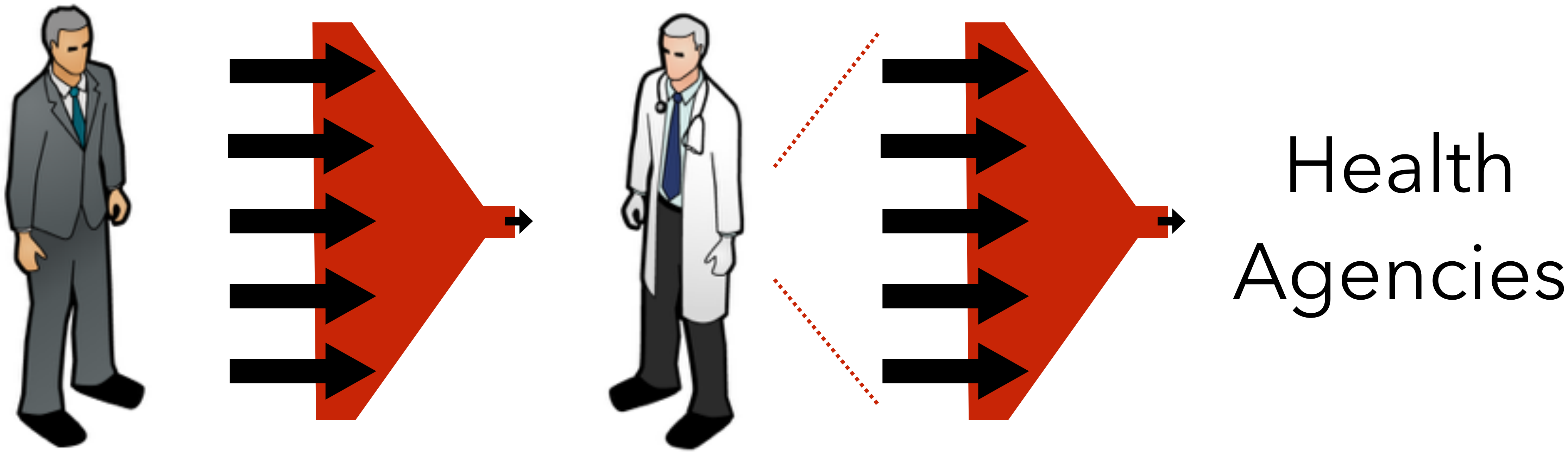
https://developer.nvidia.com/deep-learning

- The more data, the better
- "Anything humans can do in 0.1 sec, the right big 10-layer network can do too" (Jeff Dean, Google)
- Software is widely available, many opensource libraries
- Bottleneck is data. Machine Intelligence / AI requires lots of data.

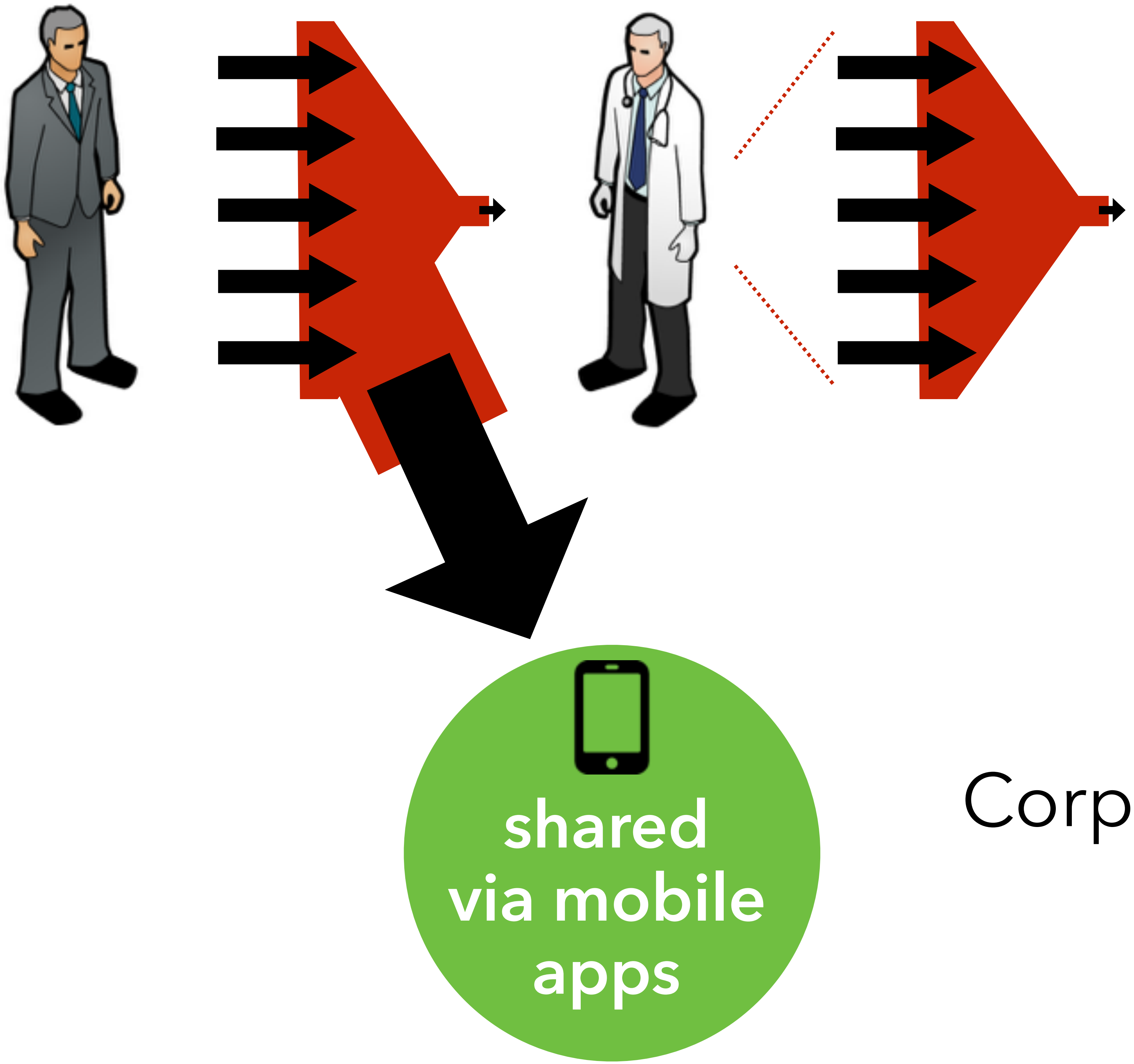
The challenge

- Massive streams of data from traditional and non-traditional sources
- Can be used for predictive modeling / AI-enhanced diagnosis using deep learning
- **Artificial intelligence (AI): can be developed by those who have the data; algorithms are no longer limiting factor!**

Follow the data…



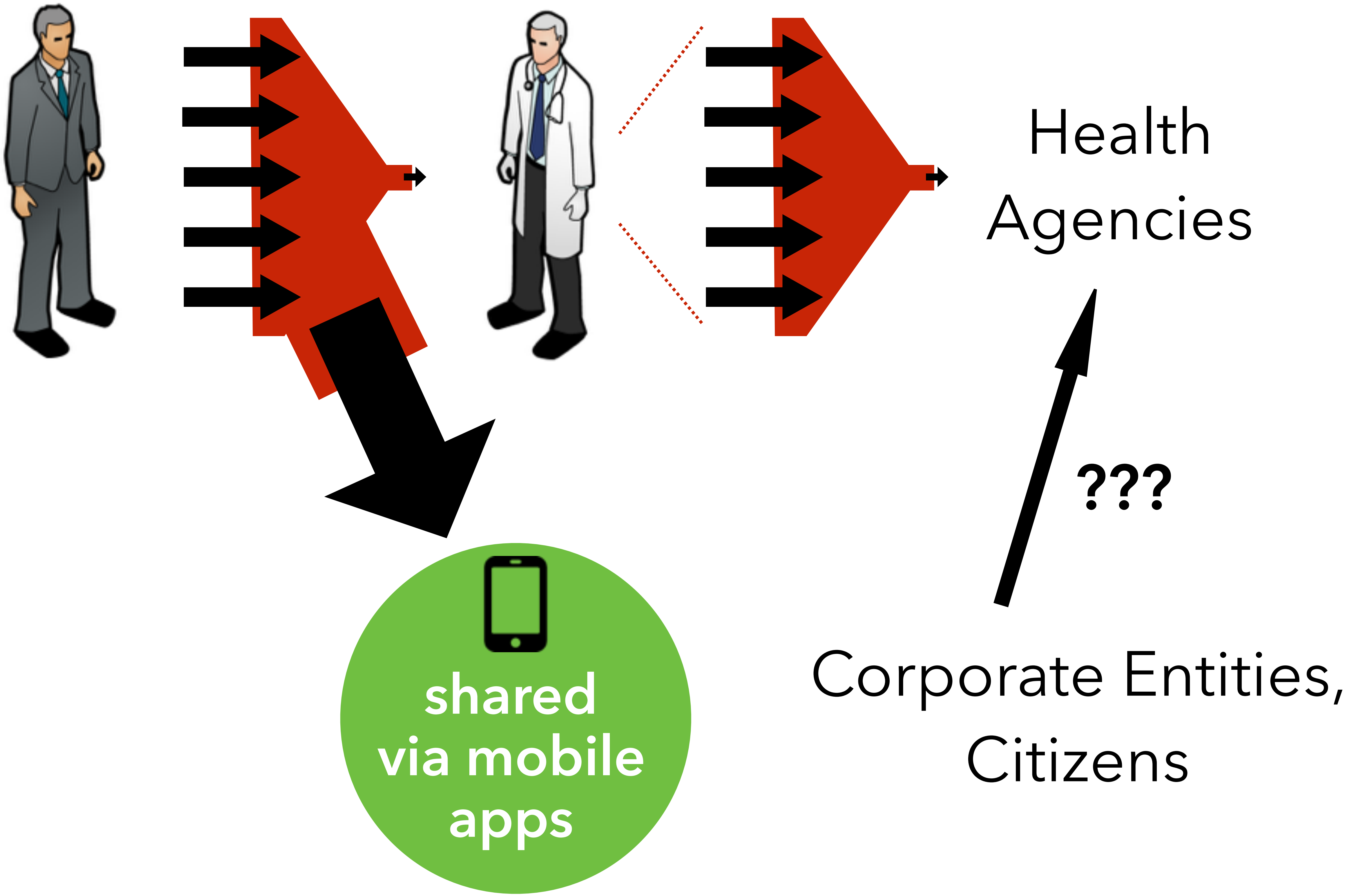
Follow the data...



Health
Agencies

Corporate Entities,
Citizens

Follow the data…



On AI vs MD

SECTIONS

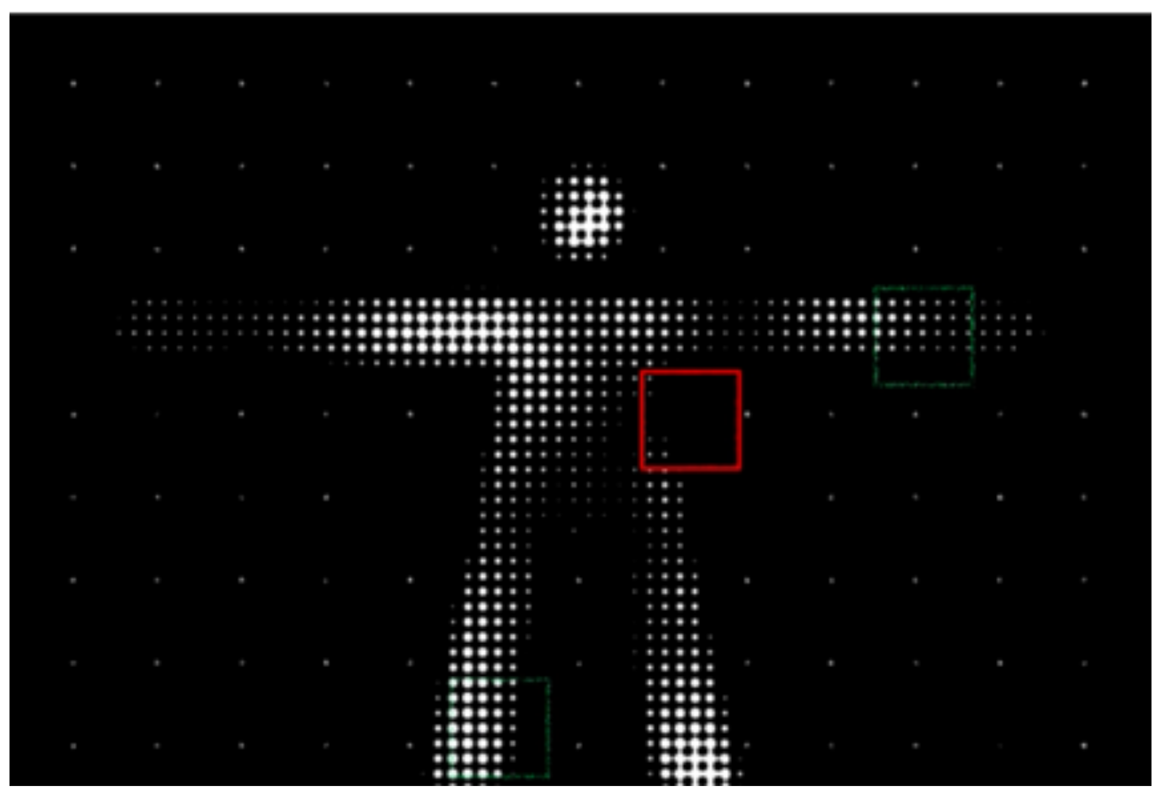
THE NEW YORKER

ANNALS OF MEDICINE APRIL 3, 2017 ISSUE

A.I. VERSUS M.D.

What happens when diagnosis is automated?

By Siddhartha Mukherjee



On AI vs MD

- **AI vs X is false premise. AI + X = enhanced X.**
- Enormous potential for AI in medical field. Much of it will come from industry. But current crop of AI is black box.
- Therefore, we must make sure that we keep the knowledge in the public domain. Collaborations between medical field and research has important role to play.

doi:10.1038/nature21056

Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva[1]*, Brett Kuprel[1]*, Roberto A. Novoa[2,3], Justin Ko[2], Susan M. Swetter[2,4], Helen M. Blau[5] & Sebastian Thrun[6]

Skin cancer, the most common human malignancy[1–3], is primarily diagnosed visually, beginning with an initial clinical screening and followed potentially by dermoscopic analysis, a biopsy and histopathological examination. Automated classification of skin lesions using images is a challenging task owing to the fine-grained variability in the appearance of skin lesions. Deep convolutional neural networks (CNNs)[4,5] show potential for general and highly variable tasks across many fine-grained object categories[6–11]. Here we demonstrate classification of skin lesions using a single CNN, trained end-to-end from images directly, using only pixels and disease labels as inputs. We train a CNN using a dataset of 129,450 clinical images—two orders of magnitude larger than previous datasets[12]—consisting of 2,032 different diseases. We test its performance against 21 board-certified dermatologists on biopsy-proven clinical images with two critical binary classification use cases: keratinocyte carcinomas versus benign seborrheic keratoses; and malignant melanomas versus benign nevi. The first case represents the identification of the most common cancers, the second represents the identification of the deadliest skin cancer. The CNN achieves performance on par with all tested experts across both tasks, demonstrating an artificial intelligence capable of classifying skin cancer with a level of competence comparable to dermatologists. Outfitted with deep neural networks, mobile devices can potentially extend the reach of dermatologists outside of the clinic. It is projected that 6.3 billion smartphone subscriptions will exist by the year 2021 (ref. 13) and can therefore potentially provide low-cost universal access to vital diagnostic care.

There are 5.4 million new cases of skin cancer in the United States[2] every year. One in five Americans will be diagnosed with a cutaneous images (for example, smartphone images) exhibit variability in factors such as zoom, angle and lighting, making classification substantially more challenging[23,24]. We overcome this challenge by using a data-driven approach—1.41 million pre-training and training images make classification robust to photographic variability. Many previous techniques require extensive preprocessing, lesion segmentation and extraction of domain-specific visual features before classification. By contrast, our system requires no hand-crafted features; it is trained end-to-end directly from image labels and raw pixels, with a single network for both photographic and dermoscopic images. The existing body of work uses small datasets of typically less than a thousand images of skin lesions[16,18,19], which, as a result, do not generalize well to new images. We demonstrate generalizable classification with a new dermatologist-labelled dataset of 129,450 clinical images, including 3,374 dermoscopy images.

Deep learning algorithms, powered by advances in computation and very large datasets[25], have recently been shown to exceed human performance in visual tasks such as playing Atari games[26], strategic board games like Go[27] and object recognition[6]. In this paper we outline the development of a CNN that matches the performance of dermatologists at three key diagnostic tasks: melanoma classification, melanoma classification using dermoscopy and carcinoma classification. We restrict the comparisons to image-based classification.

We utilize a GoogleNet Inception v3 CNN architecture[9] that was pre-trained on approximately 1.28 million images (1,000 object categories) from the 2014 ImageNet Large Scale Visual Recognition Challenge[6], and train it on our dataset using transfer learning[28]. Figure 1 shows the working system. The CNN is trained using 757 disease classes. Our dataset is composed of dermatologist-labelled images organized in a buoy Technology Team Blog Log in Sign up Check My Symptoms

Hello, I'm Buoy.
I can help you decide what
to do when you're sick.

What brings you in?

Check My Symptoms

Log in or Sign up

emerged with a solution. As opposed to Web 1.0 symptom checkers that use simple, pre-determined question logic, Eddie explained that our algorithm would crunch countless statistics simultaneously, ranking diagnoses and questions in real-time, and choosing the next question based on statistical reasoning. If we could feed the program with the right statistics about symptoms, risk factors, and diseases, the program would be able to reason its way to a set of diagnoses, just like a real doctor. Eddie started building the algorithm, and I began pulling the necessary data from clinical papers.

Today, Buoy covers over 1,600 diagnoses, knows 30,000 interview questions, and relies on data from over 18,000 clinical papers covering 5,000,000 patients. In our first quality control trial, where patients used the program in the waiting room before seeing their doctor, we agreed with the doctors' diagnoses 90.9% of the time.

What we learned from our users' feedback.

The Buoy journey only starts with the algorithm. My two other cofounders, Adam Lathram and Nate Ren, took the lead on designing Buoy's look and feel, which raised even more questions. What could Buoy do to make people feel listened to? How could a computer program convey empathy and compassion? What

POWER

POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER
POWER

AI

AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI AI

AI

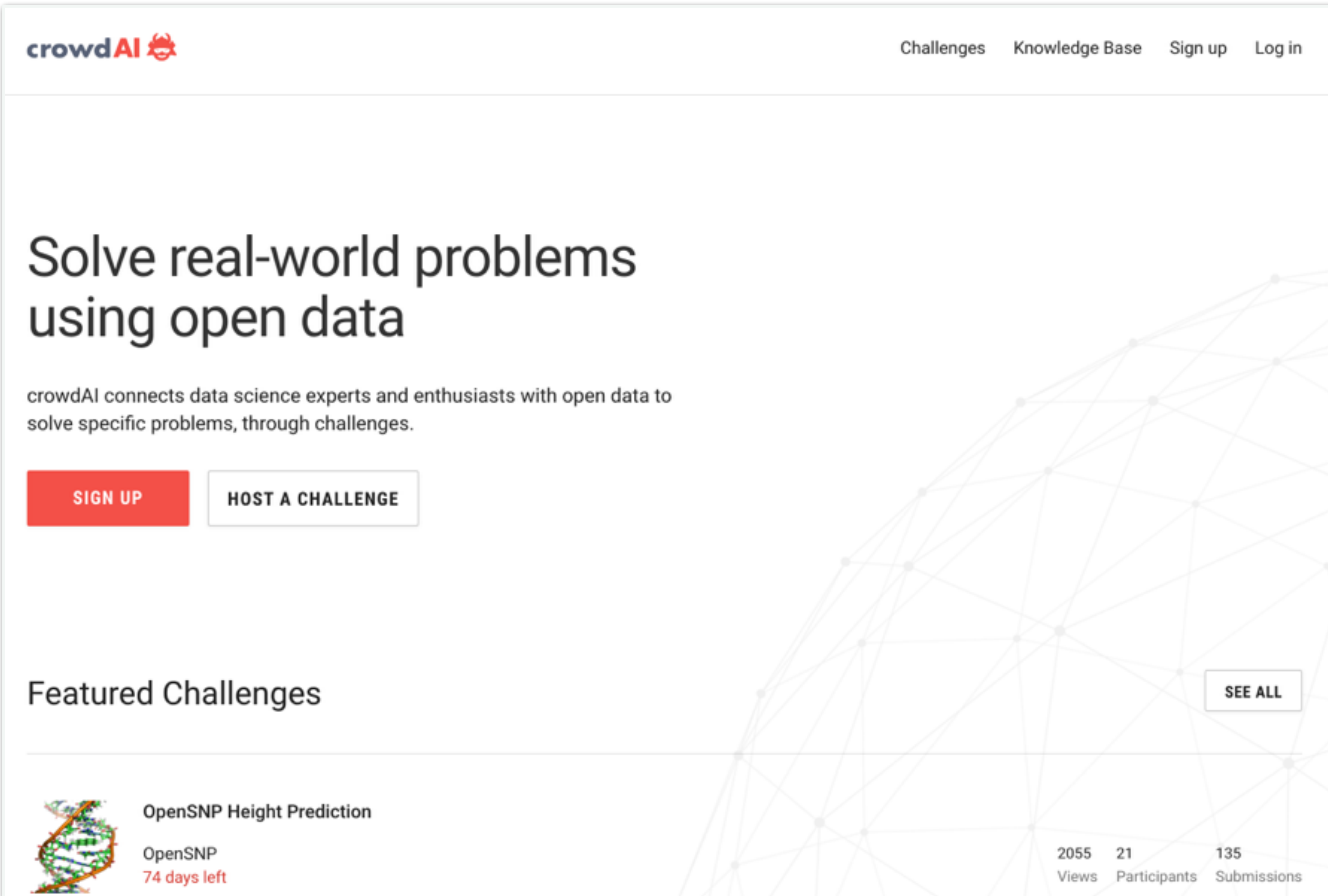
=

ALGORITHM

+

DATA

www.crowdAI.org



www.crowdAI.org

crowdAI

Challenges Knowledge Base Admin

NIPS 2017: Learning to Run

Reinforcement learning environments with musculoskeletal models

By Stanford Neuromuscular Biomechanics Laboratory

68 Days left

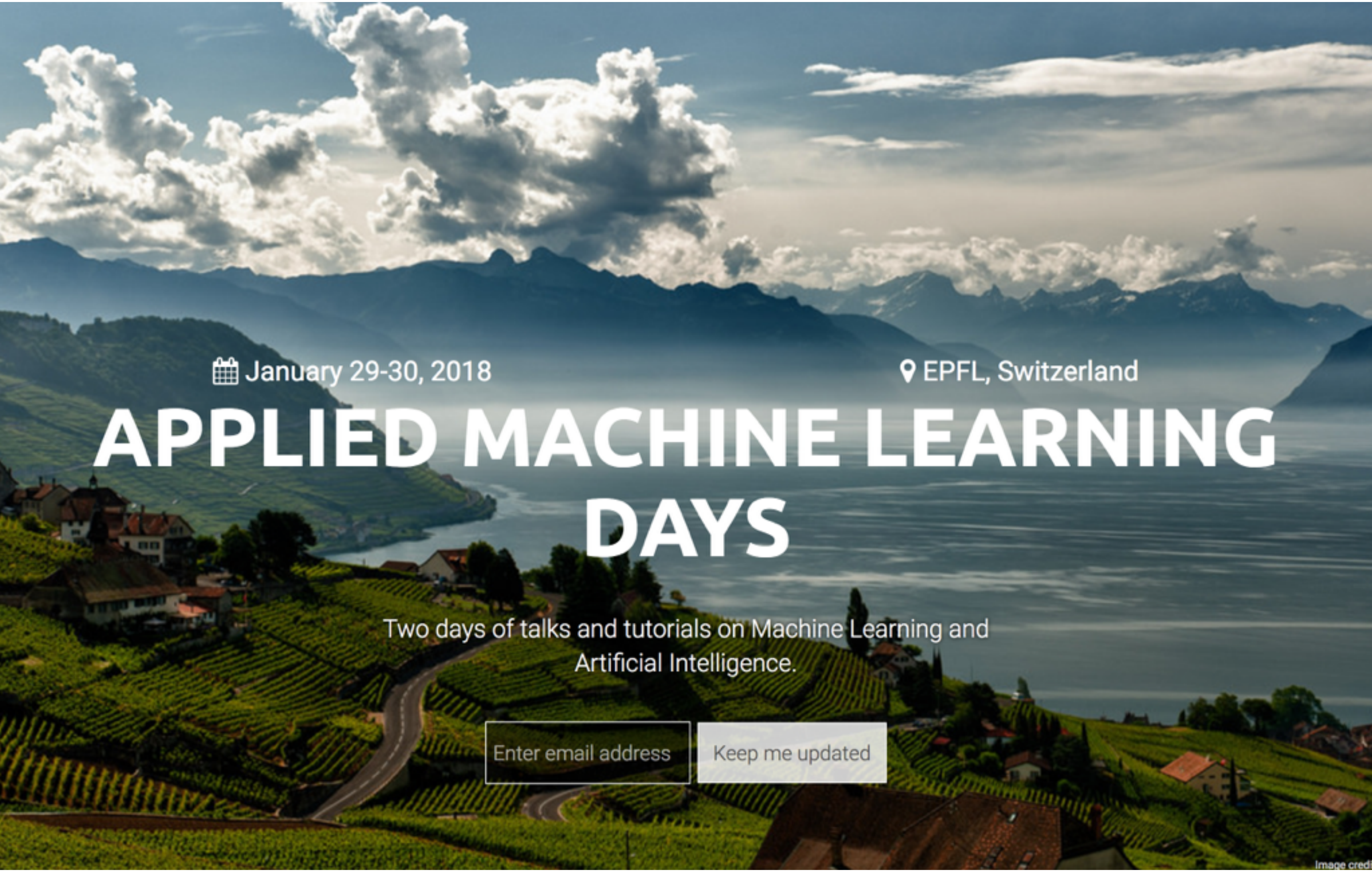
8278 Views | 36 Participants | 69 Submissions

53

Overview | Leaderboard | Discussion | Dataset

Overview

Welcome to **Learning to Run**, one of the 5 official challenges in the NIPS 2017 Competition Track In this competition, you are tasked with developing a controller to enable a physiologically-based human model to navigate a complex obstacle course as quickly as possible. You are provided with a human musculoskeletal model and a physics-based simulation environment where you can synthesize physically and physiologically accurate


January 29-30, 2018
EPFL, Switzerland
APPLIED MACHINE LEARNING DAYS
Two days of talks and tutorials on Machine Learning and Artificial Intelligence.
Enter email address
Keep me updated

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Online | Mobile | Global

Digital Health & Epidemiology

Prof. Marcel Salathé, Digital Epidemiology Lab, EPFL

@marcelsalathe