Practical Deep Learning for Healthcare

Sooheon Kim, Paul Warren

2019/11/7

sooheon@factor.ai
paul@factor.ai

Table of Contents

Why Deep Learning?

nature

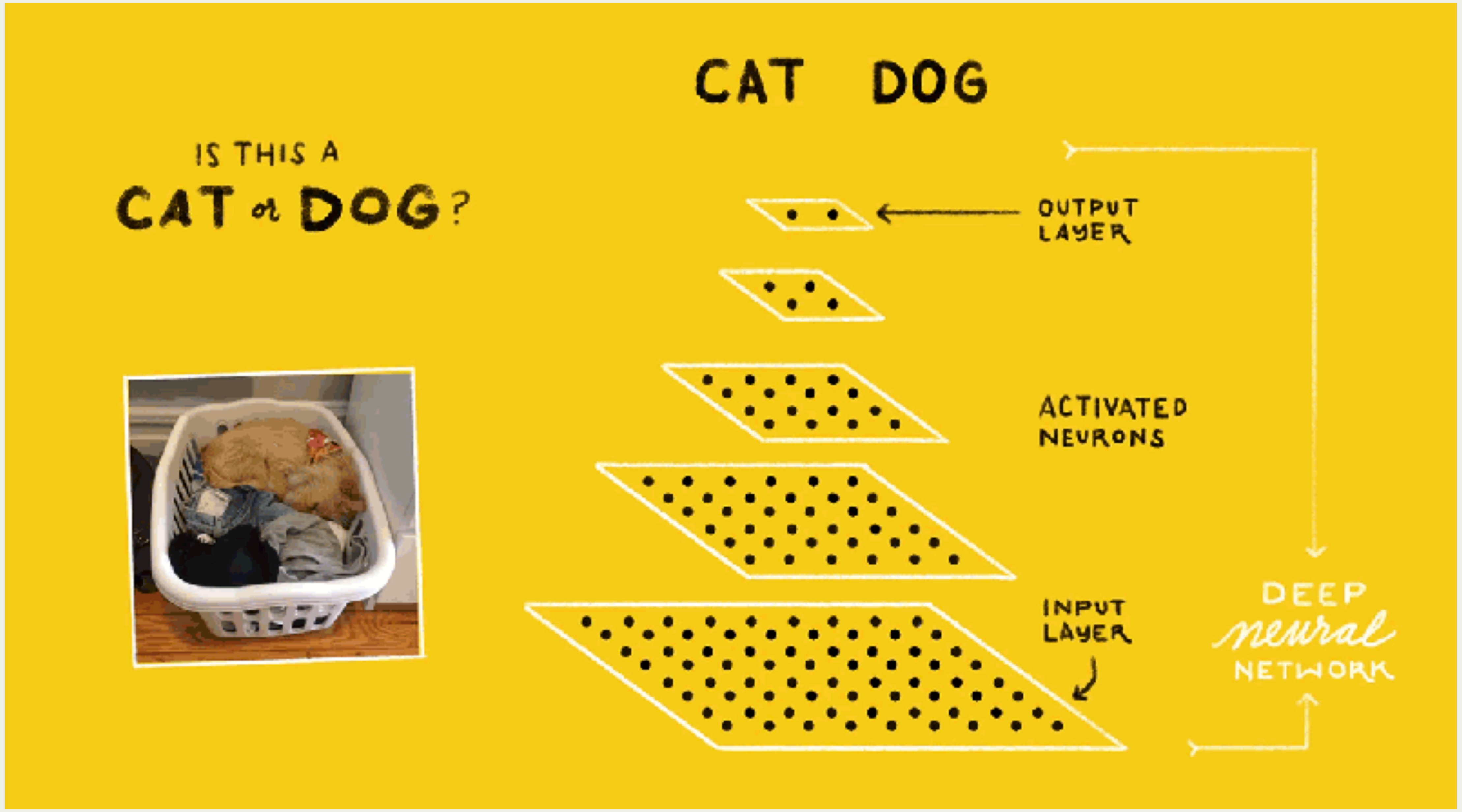
Article | Published: 30 October 2019

Grandmaster level in StarCraft II using multi-agent reinforcement learning

Oriol Vinyals, Igor Babuschkin, […] David Silver

Nature (2019) | Cite this article

1. Understanding Deep Learning Through Computer Vision



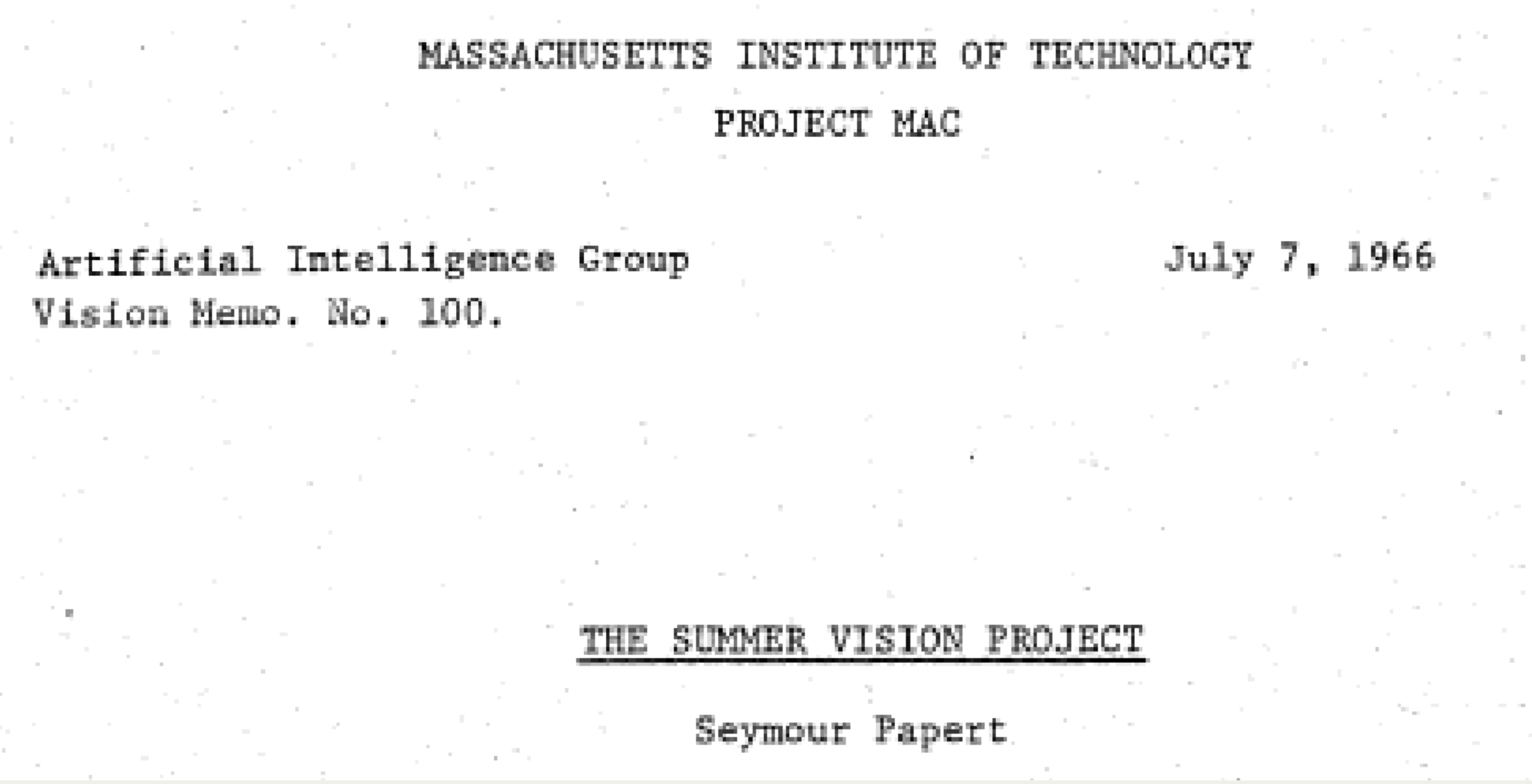
 Factor AI. 2019.

Original Task: 1966

- MIT professor:

  "Teach a computer to interpret images, it should take a summer"

- Resulted in AI winter



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

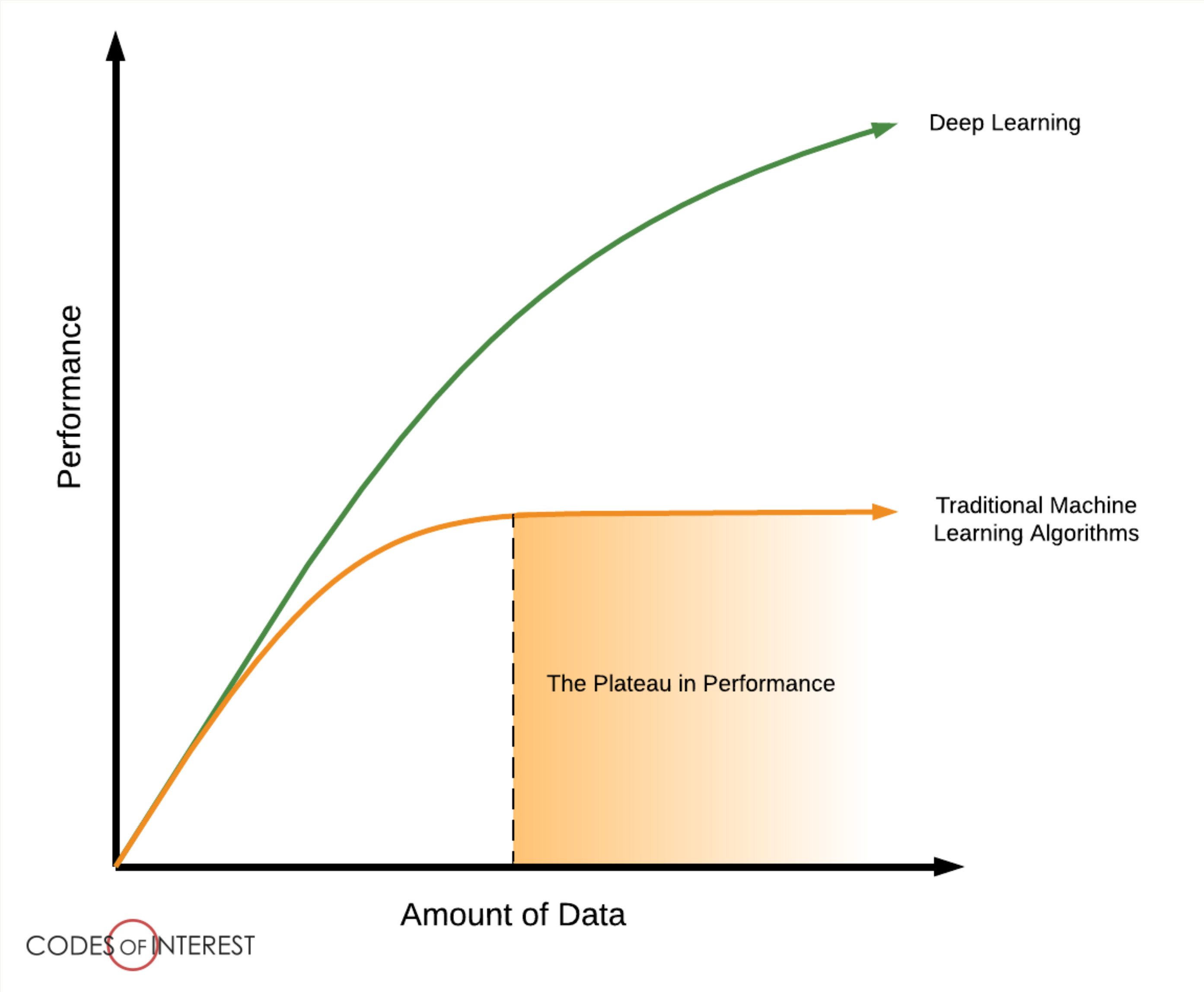
THE SUMMER VISION PROJECT

Seymour Papert

 Factor AI. 2019.

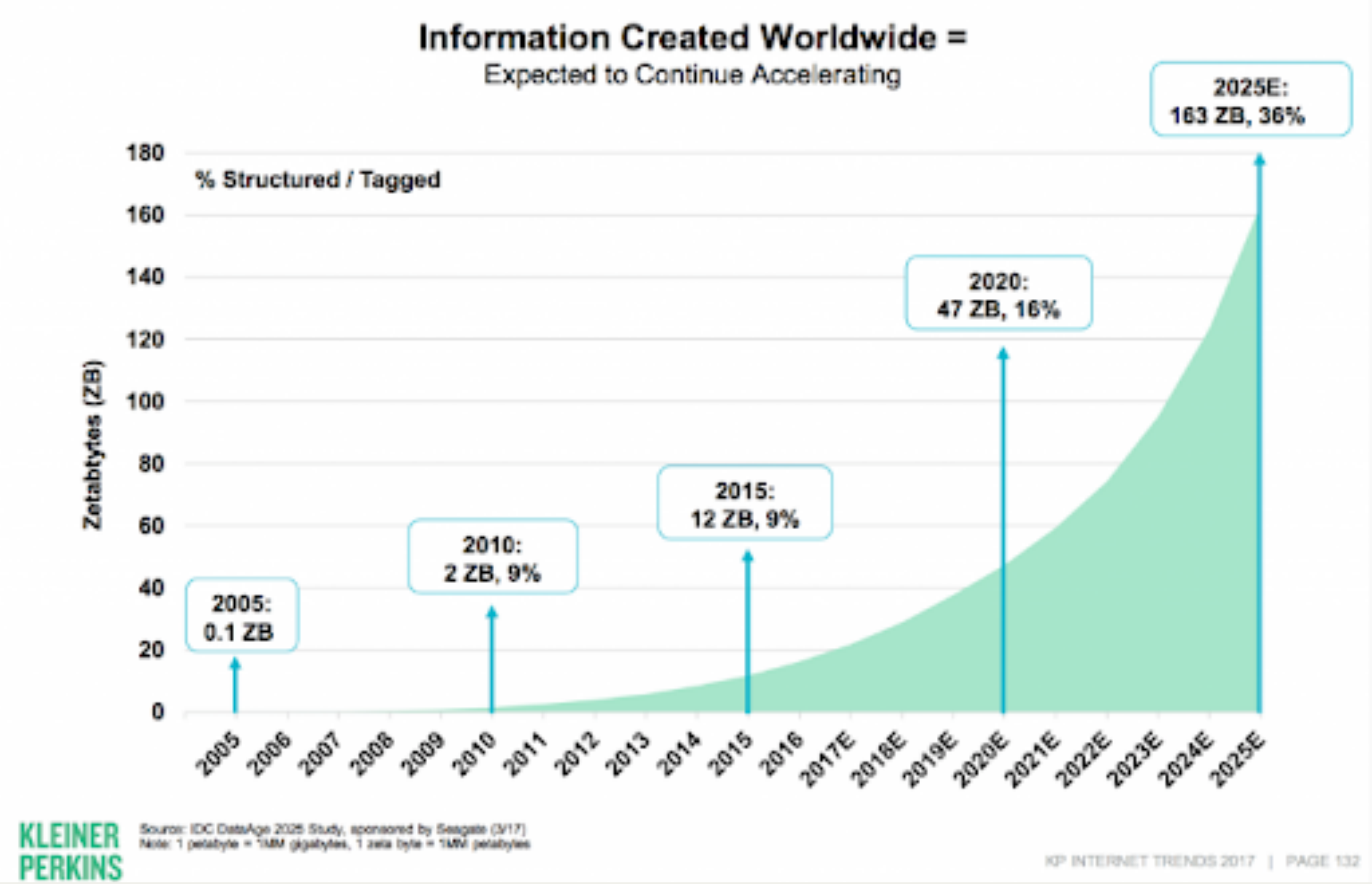
ImageNet

- 2012: human level performance for first time



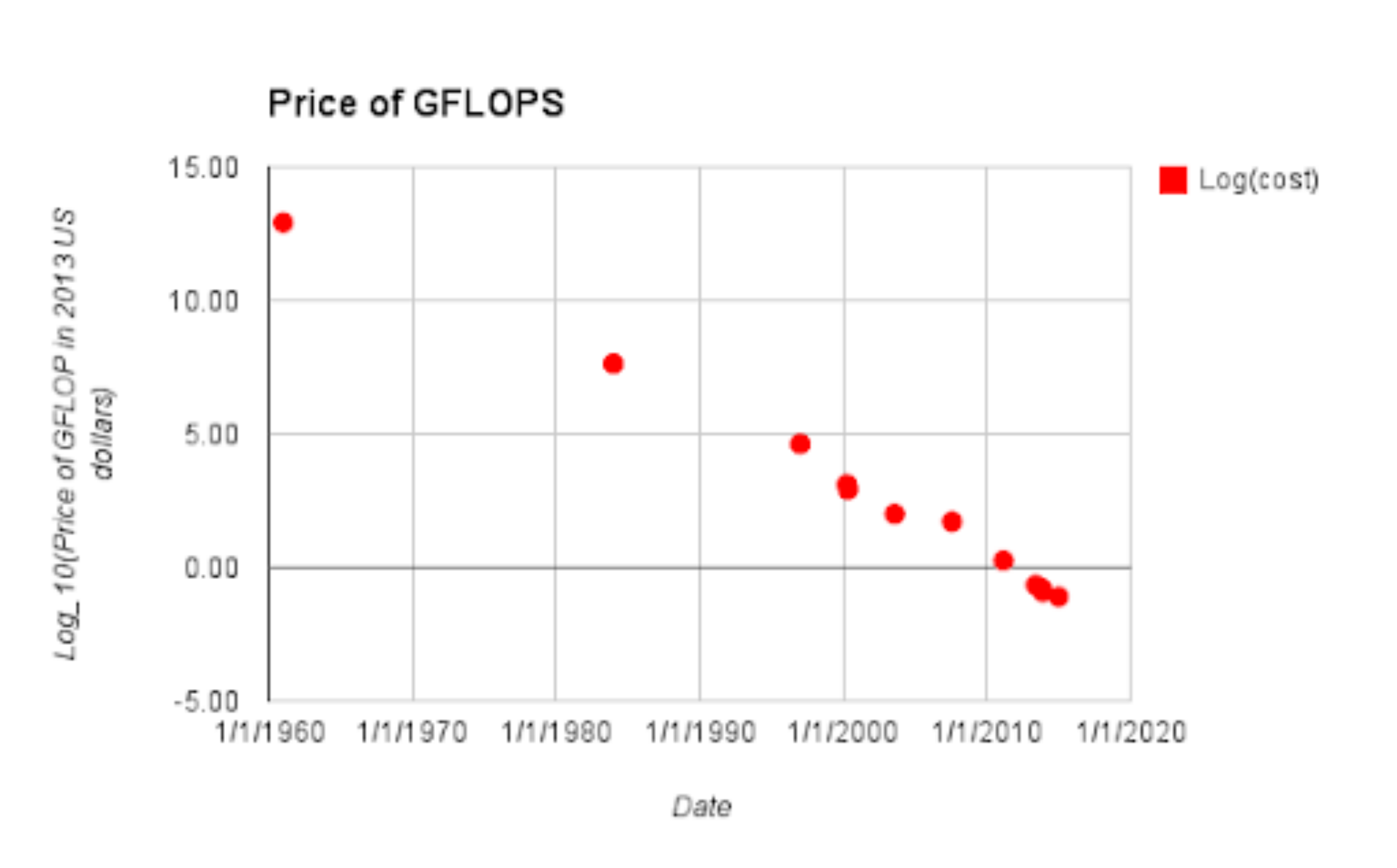
Trends

1. Big Data
2. Better Computers
3. Smarter Algorithms



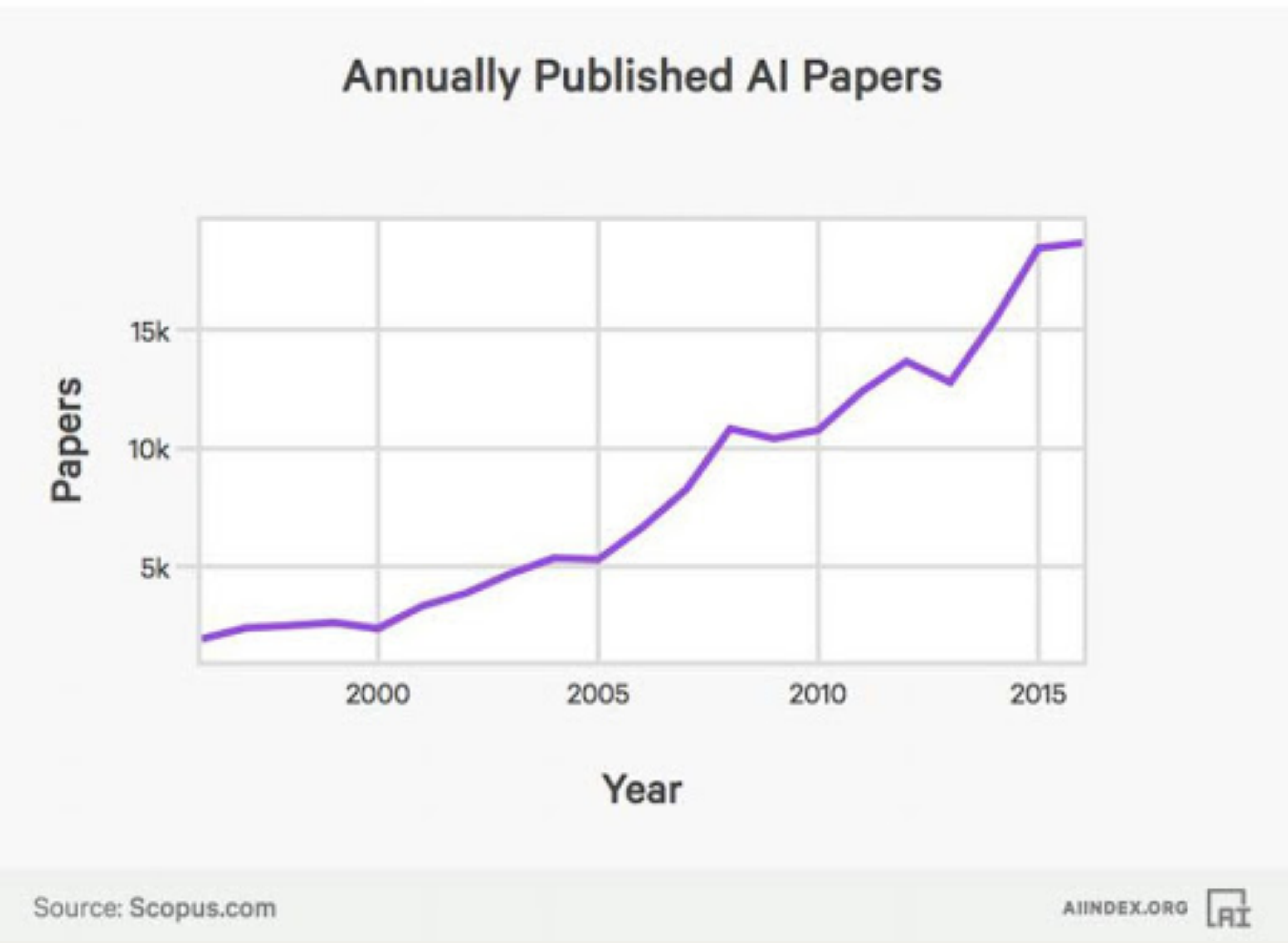
Trends

1. Big Data
2. Better Computers
3. Smarter Algorithms

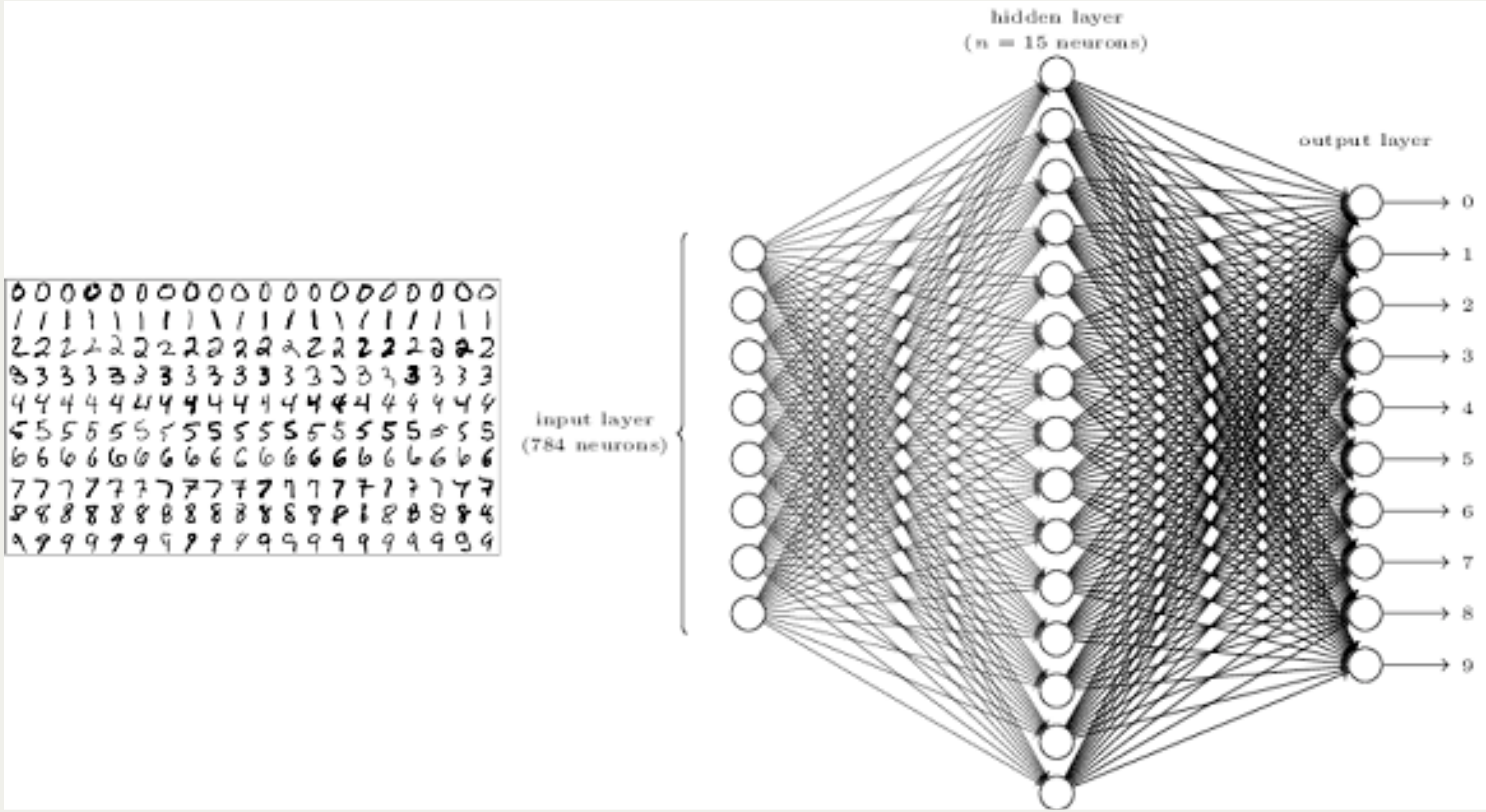


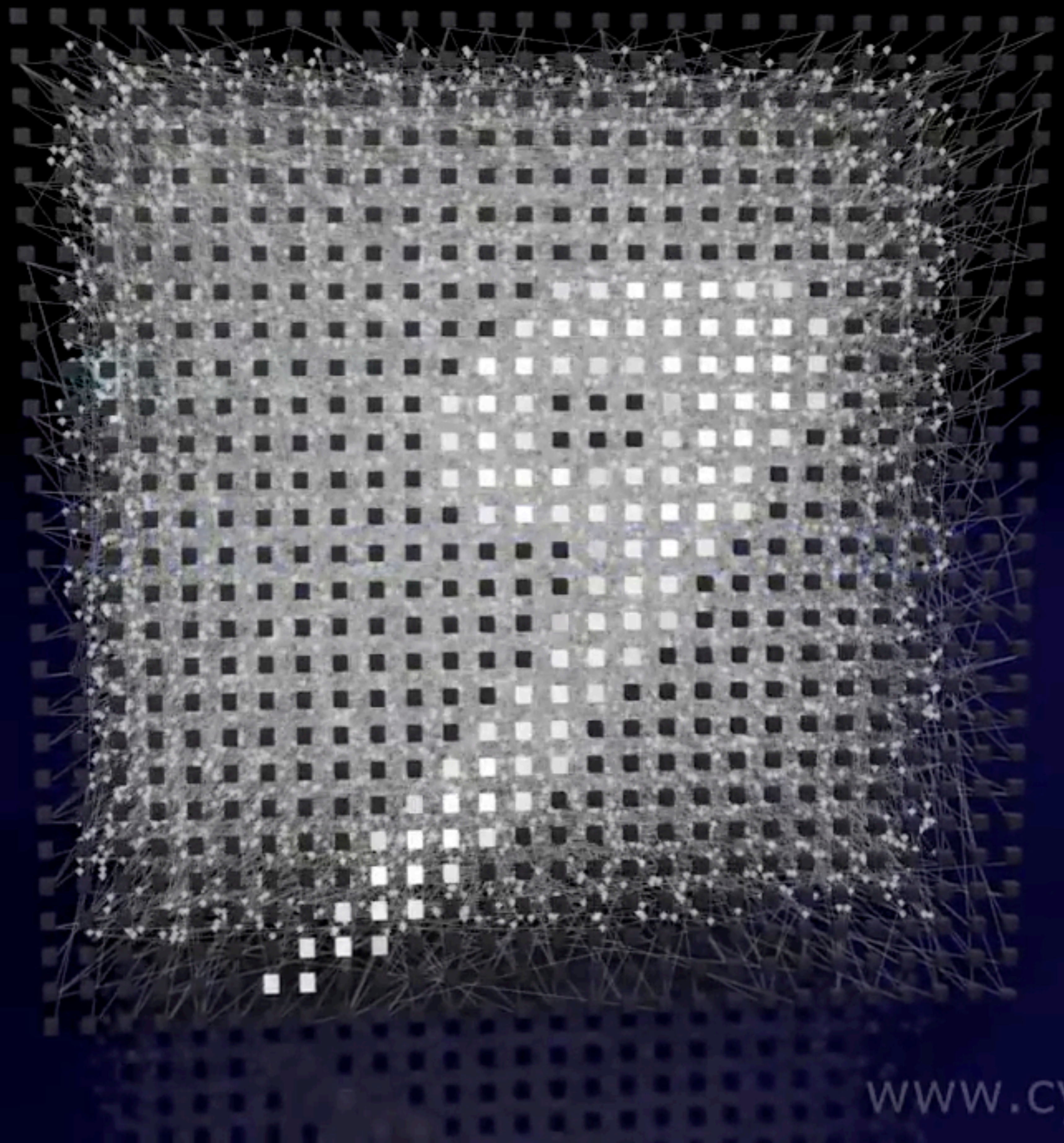
Trends

1. Big Data
2. Better Computers
3. Smarter Algorithms



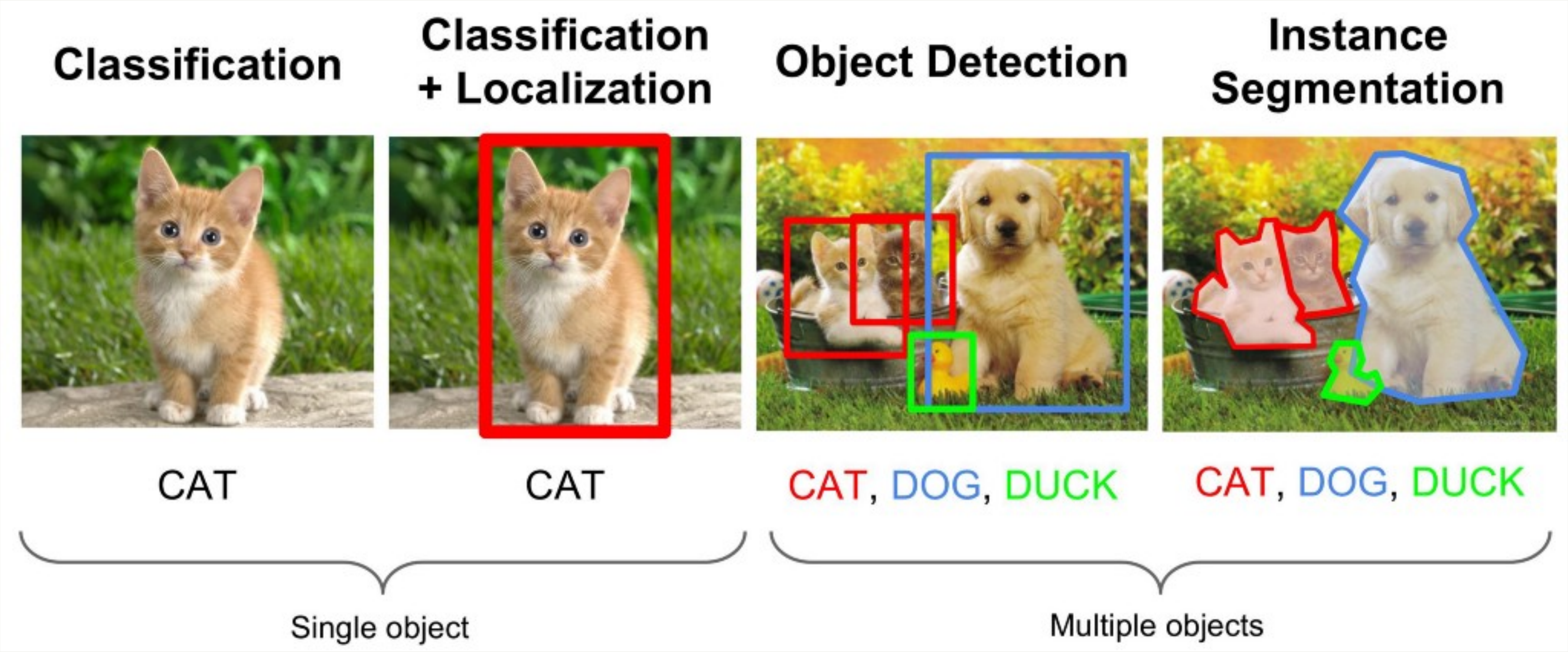
What do Deep Learning Models Look Like?



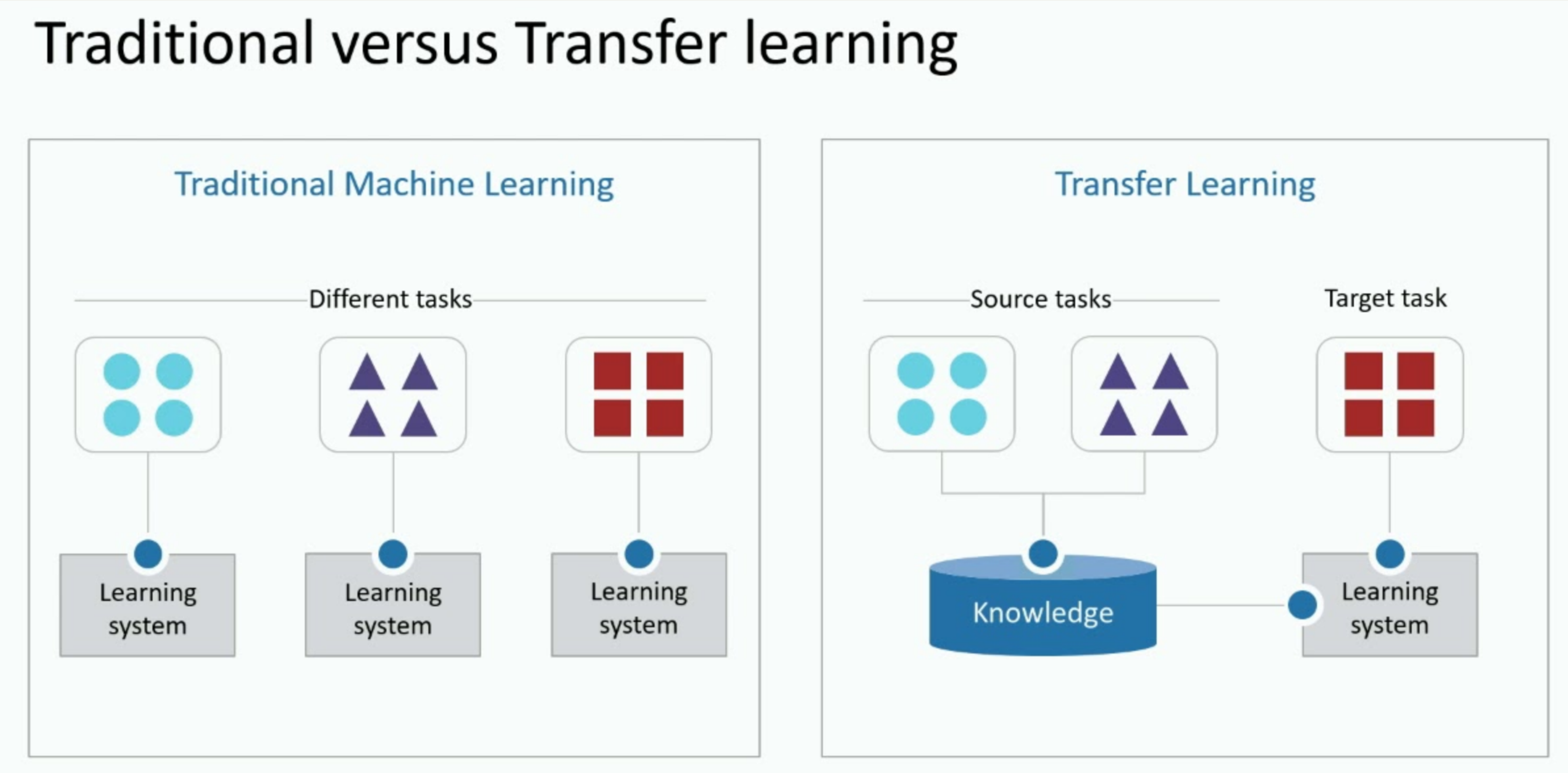


 Factor AI. 2019.

Evolution of CV Tasks



2. Deep Learning's Superpower: Transfer Learning



 Factor AI. 2019.

Types of Transfer learning

Sequential

- Big data → small dataset
- Common task → unique task
- Curriculum learning

Parallel

- Multi-task learning

Transfer Learning By Examples

Computer Vision

- ImageNet → IGV tumor classification

Regression

- QSAR multi-task prediction

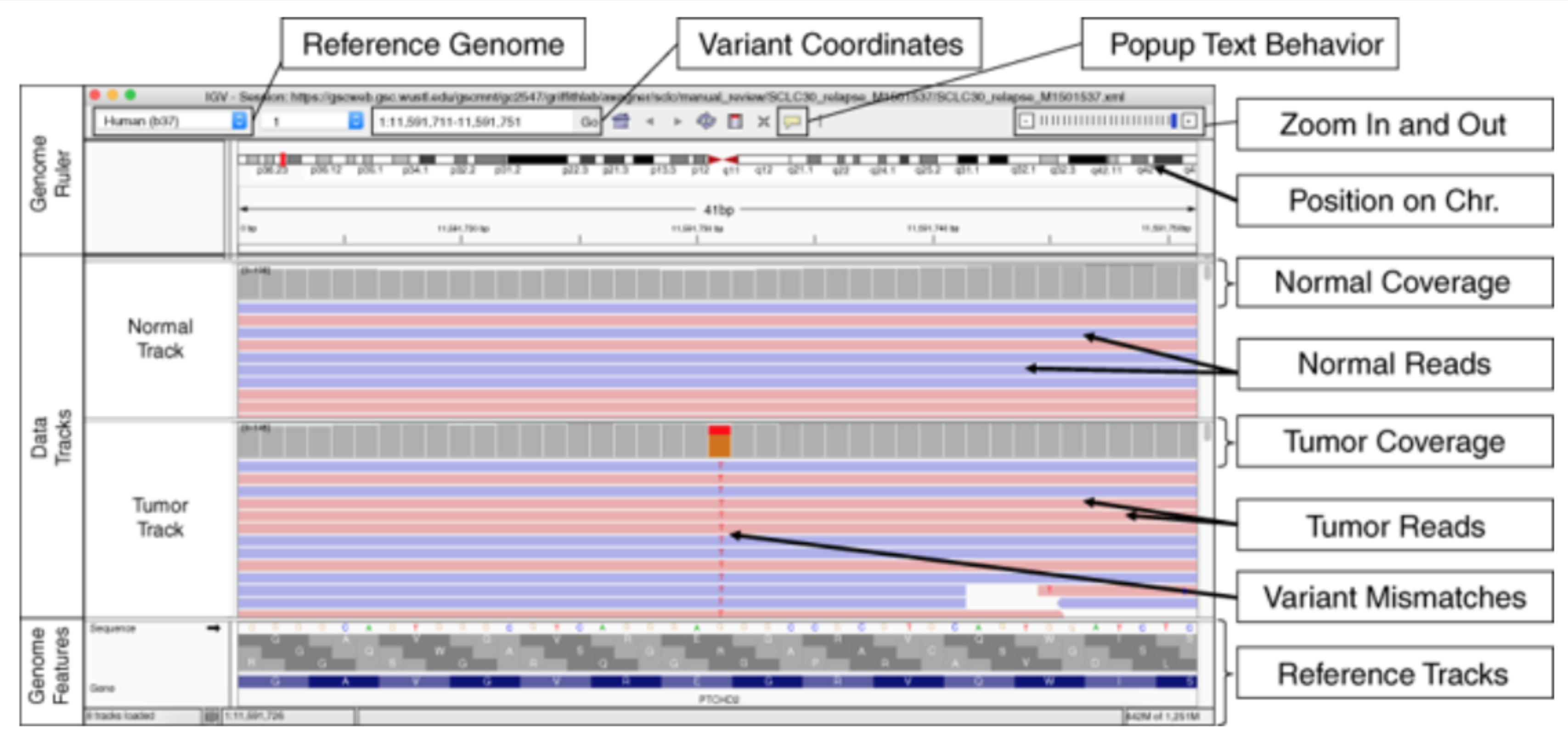
Natural Language Processing

- WikiText103 → Oncologist Language Model
- WikiText103 → Oncologist Language Model → Gene Mutation Classifier

Integrative Genomics Viewer (IGV)

Manual procedure for tumor-normal calling[i]

Using transfer learning from ImageNet, possible to achieve 93% accuracy with 20 lines of code.[igv]



[i] Genet Med 21, 972–981 (2019) doi:10.1038/s41436-018-0278-z
[igv] Tumor-normal sequencing: is this variant real? Alena Harley. 2018

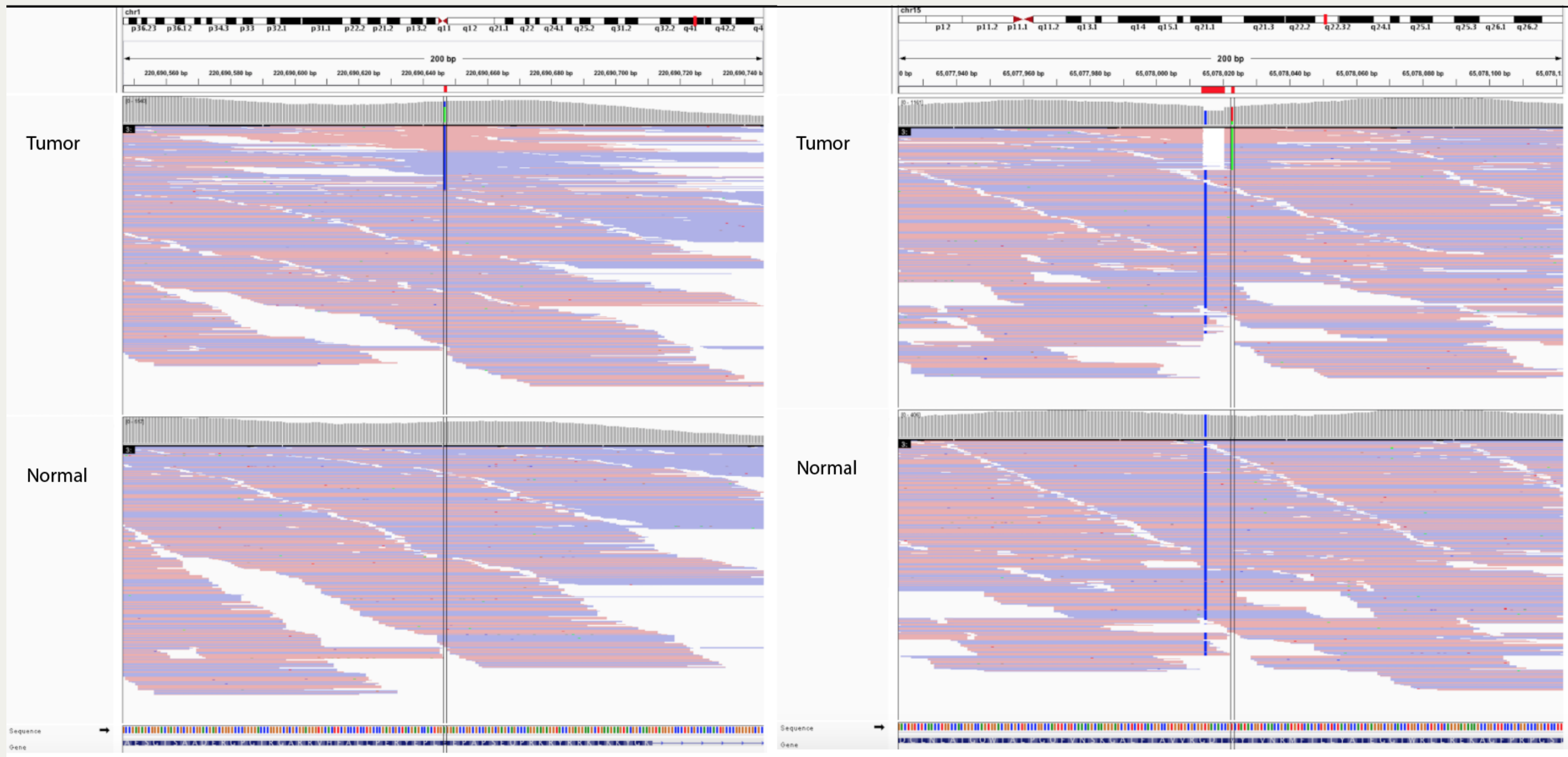
IGV Tumor Classification

```
from fastai import *
from fastai.vision import *

data = ImageDataBunch.from_folder(
  path, size=512, bs=32
  ds_tfms=get_transforms(
    do_flip=False,max_rotate=None,
    max_zoom=1, max_lighting=None,
    max_warp=None,p_affine=0,p_lighting=0
  )
)
data.normalize(imagenet_stats)

learn = create_cnn(data, models.resnet34, metrics=accuracy)
lrf = learn.lr_find()
learn.fit_one_cycle(4, max_lr=0.01)

lrf = learn.lr_find()
lr = 0.0002
lrs=np.array([lr/100,lr/10,lr])
learn.fit_one_cycle(4, max_lr=lrs)
```



Example: QSAR

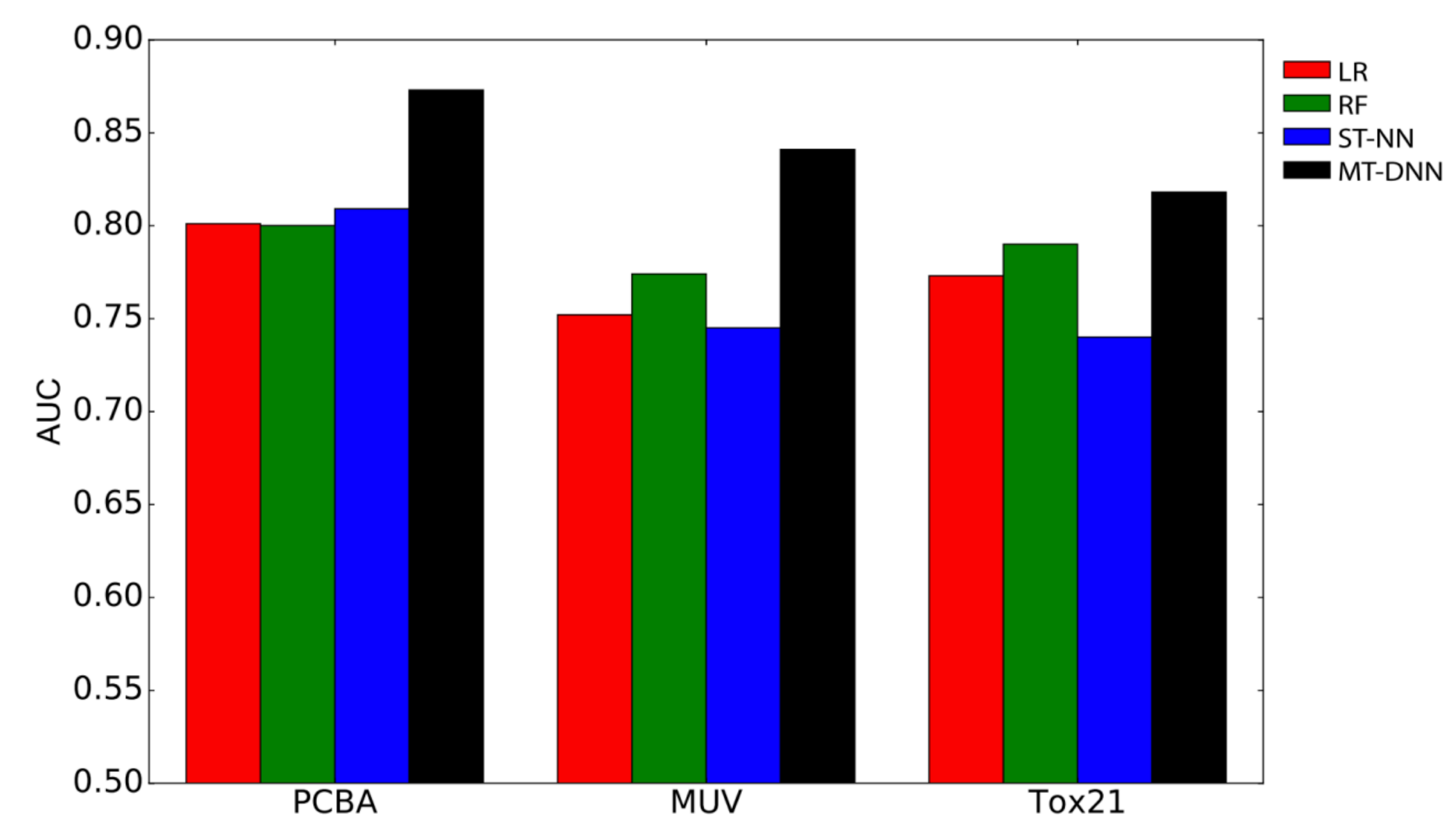
SVM, RF, Linear Regression, and single-layer NNs have been used for decades. Do DNNs improve?

Methodology

Compare performance of multi-task DNNs with single task traditional algorithms.

Results

Multi-task DNNs outperform traditional ML (linear regression, RF, SVMs)[4]

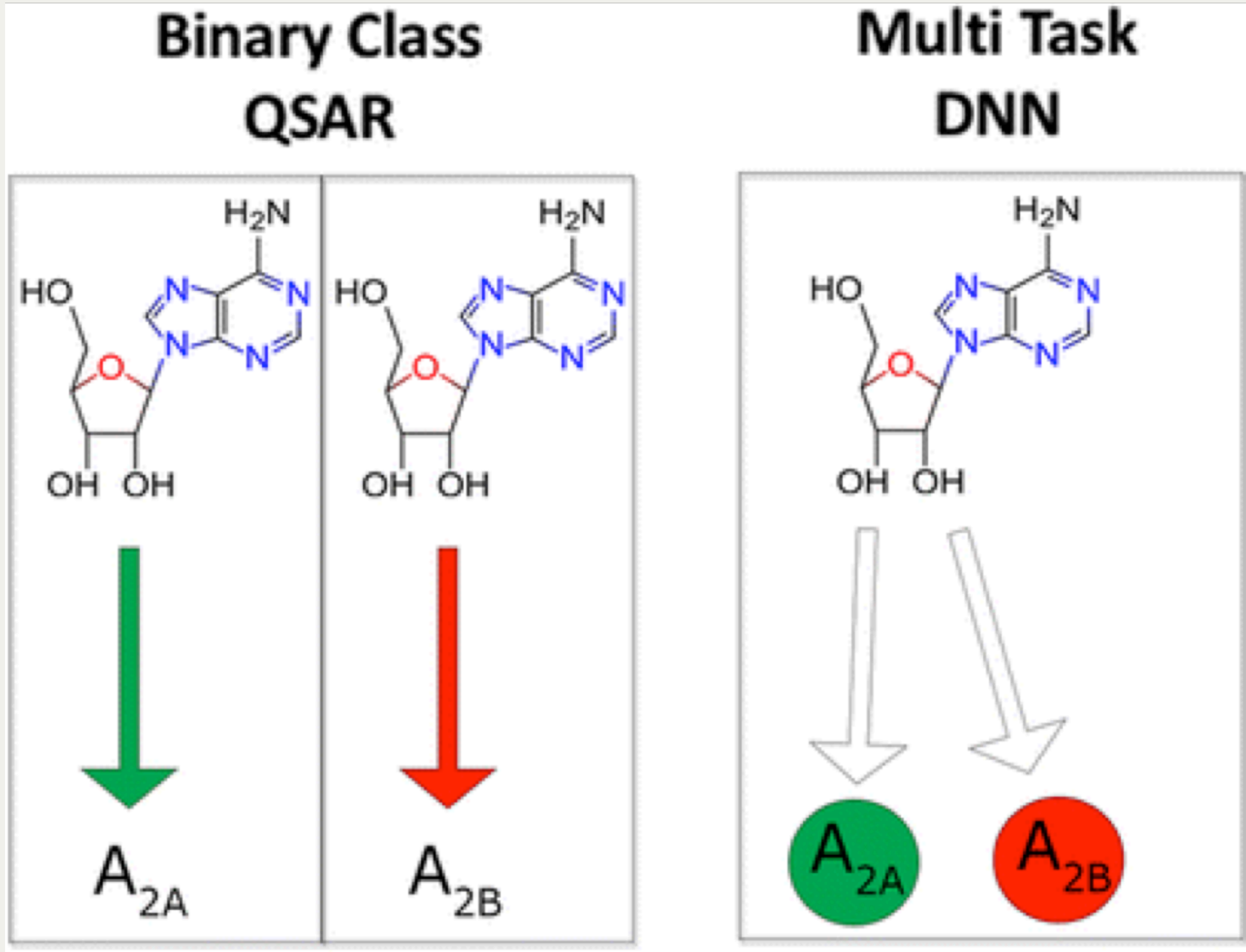


[4] Ramsundar, B.; Kearnes, S.; Riley, P.; Webster, D.; Konerding, D.; Pande, V. arXiv:1502.02072

Example: QSAR

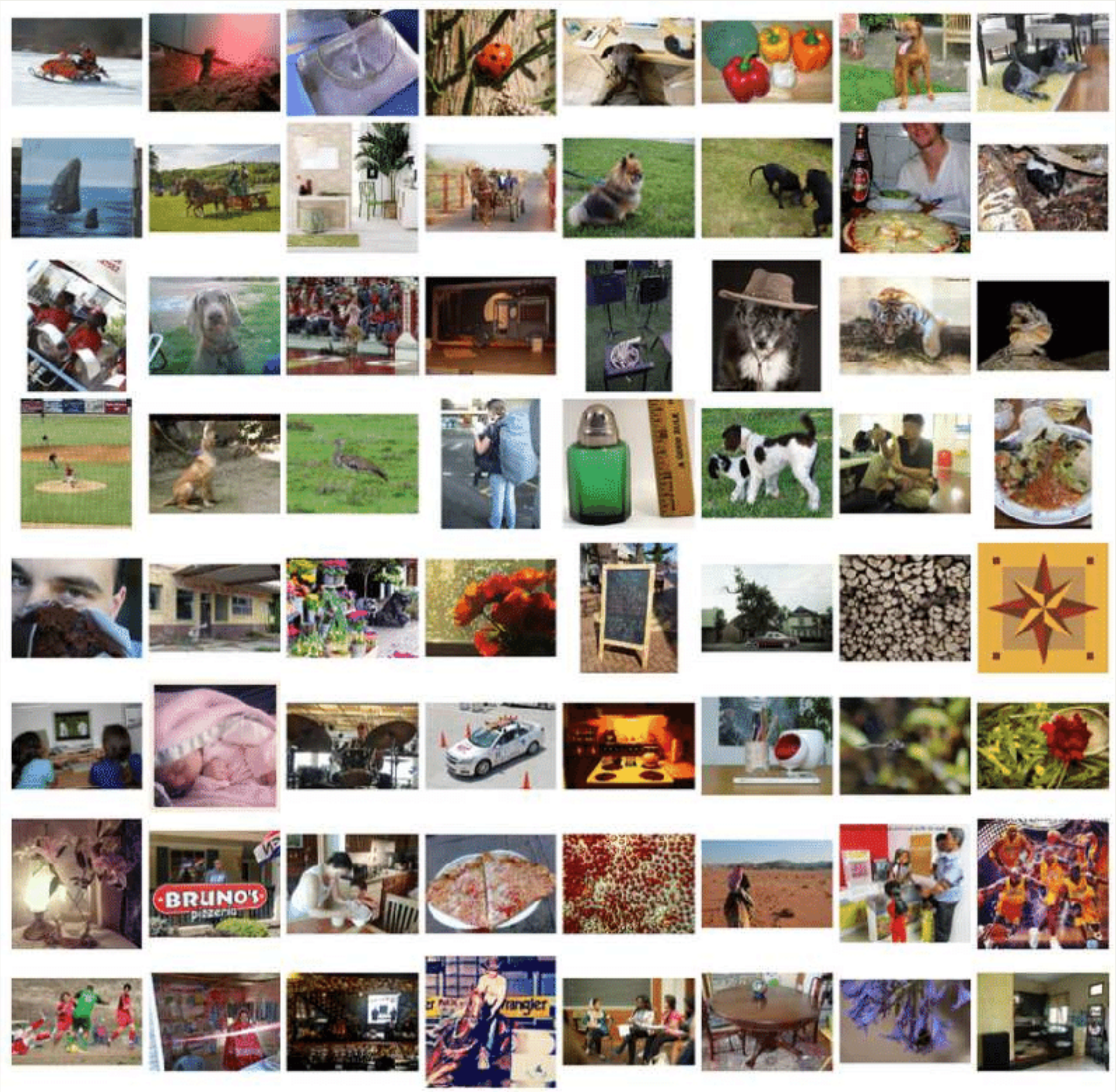
Hypothesis

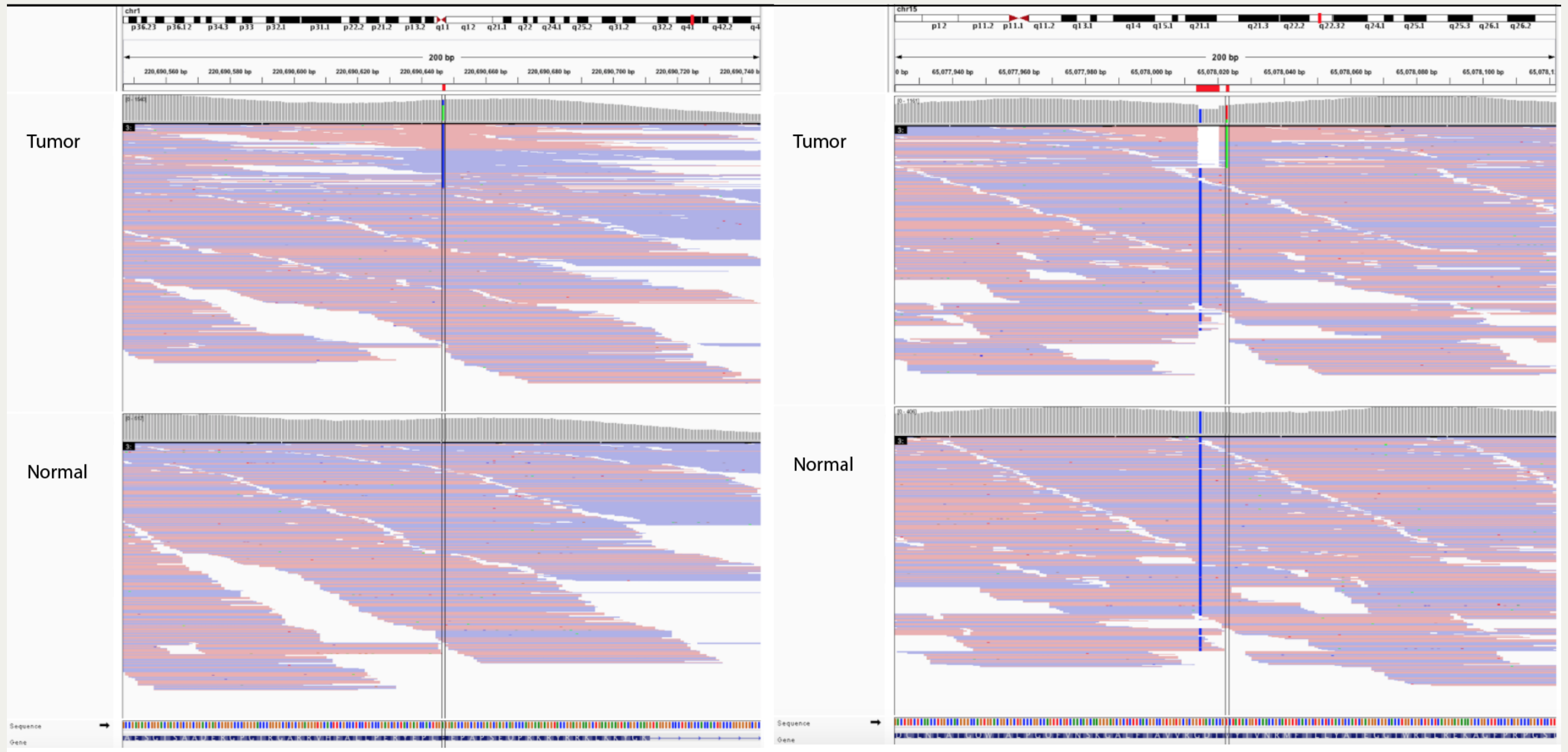
- Shared hidden representation works as form of transfer learning[5]
- Weights of each task



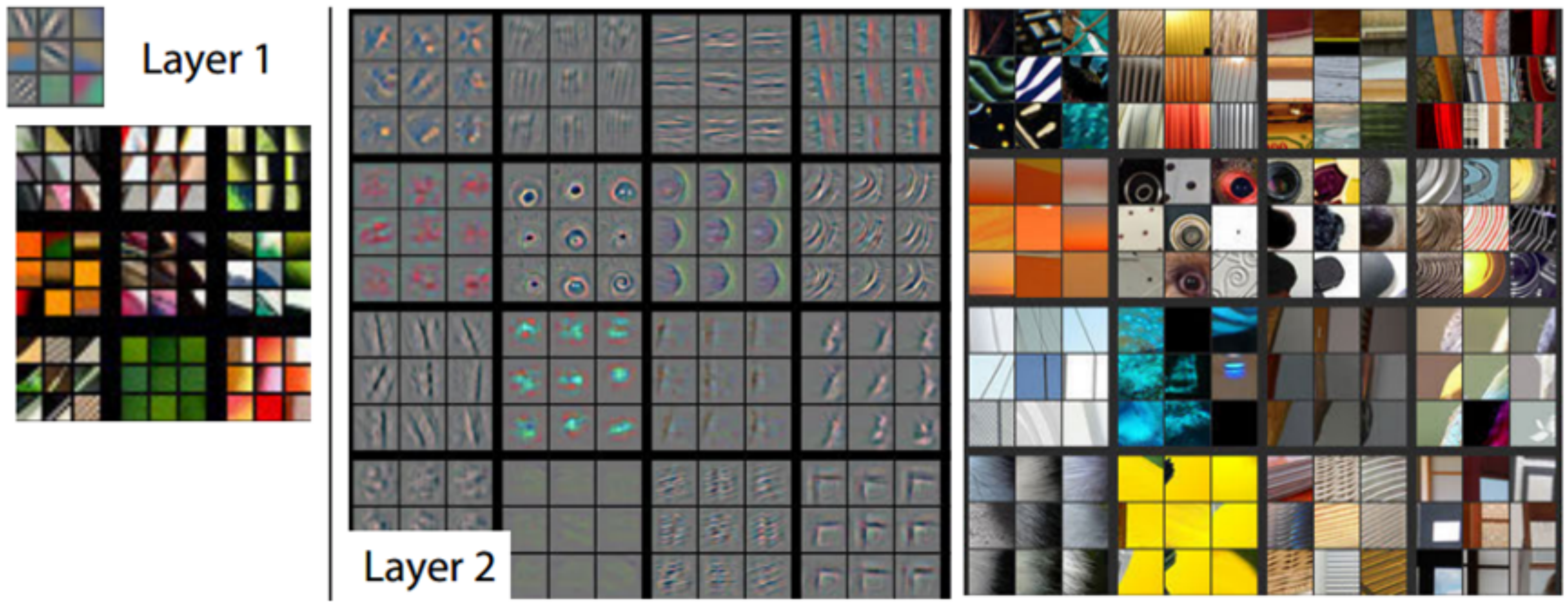
[5] J Cheminform 9, 45 (2017) doi:10.1186/s13321-017-0232-0

Shared Hidden Representations?





Generalized Hidden Representations



Visualizations of Layer 1 and 2. Each layer illustrates 2 pictures, one which shows the filters themselves and one that shows what part of the image are most strongly activated by the given filter. For example, in the space labled Layer 2, we have representations of the 16 different filters (on the left)

Visualizing and Understanding Convolutional Networks. Zeiler and Fergus.

Example: Natural Language Processing (Text)

WikiText103 → Oncologist Language Model → Gene Mutation Classifier

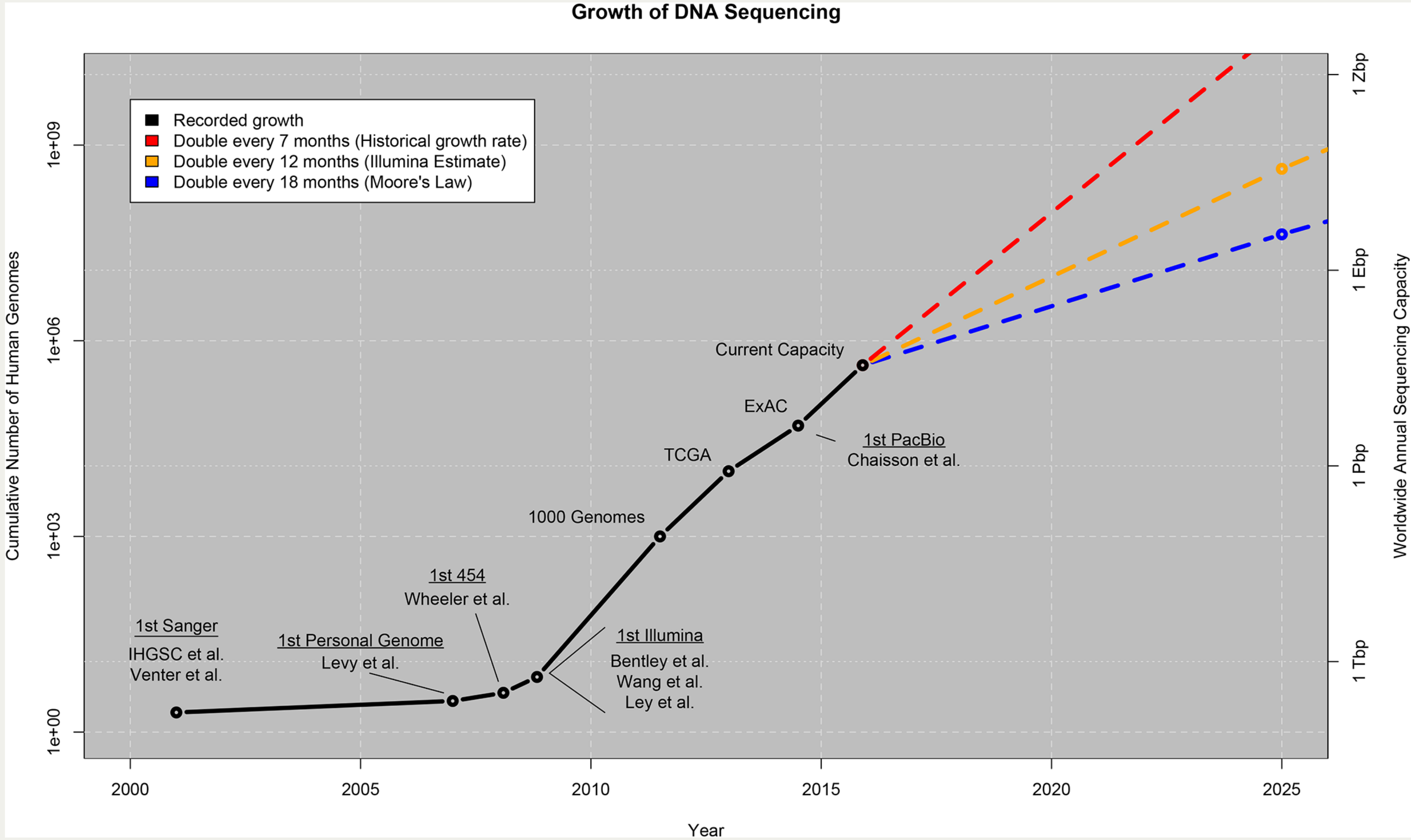
3. Data Stewardship

Biological Big Data

- DNA sequencing grows faster than computation capacity.
    - Will surpass data growth in SNS or videos by 2025[3]
- TCGA, ARCHS4, ChEMBL, Protein Data Bank, DrugBank...

Traditional ML methods do not scale well with data size, feature complexity, and task count

DNNs scale well with big data, multi-task modeling, and unengineered features.



[3] https://doi.org/10.1371/journal.pbio.1002195

Data

Lifeblood of AI.

Quantity and quality of data is competitive advantage.

Security and privacy concerns hinder collaboration.

Bottlenecks

Access, quantity, quality

Data Stewardship

Case Study

Harvard Undiagnosed Diseases Network (UDN)

- Mapping disease ↔ mutation correlations
  - 35% diagnosis rate, 31 new syndromes defined
- 32% of participants already had exome sequencing[udn]

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease

RESULTS

A total of 1519 patients (53% female) were referred to the UDN, of whom 601 (40%) were accepted for evaluation. Of the accepted patients, 192 (32%) had previously undergone exome sequencing. Symptoms were neurologic in 40% of the applicants, musculoskeletal in 10%, immunologic in 7%, gastrointestinal in 7%, and rheumatologic in 6%. Of the 382 patients who had a complete evaluation, 132 received a diagnosis, yielding a rate of diagnosis of 35%. A total of 15 diagnoses (11%) were made by clinical review alone, and 98 (74%) were made by exome or genome sequencing. Of the diagnoses, 21% led to recommendations regarding changes in therapy, 37% led to changes in diagnostic testing, and 36% led to variant-specific genetic counseling. We defined 31 new syndromes.

[udn] N Engl J Med. 2018 November 29; 379(22): 2131–2139. doi:10.1056/NEJMoa1714458

Data Stewardship

Case Study

Numer.ai

- Public hedge fund, run by prediction markets
    - Over $7.8m prize mony paid out
- Needs to hide data while preserving signal (homomorphic encryption)
- Uses proprietary Adversarial Neural Cryptography[he]
    - High cost to implement

id	era	feature1	...	feature310	target
n2b2e3dd163cb422	era1	0.75	...	0.00	0.25
n177021a571c94c8	era1	1.00	...	0.25	0.75
n7830fa4c0cd8466	era1	0.25	...	1.00	0.00
nc584a184cee941b	era1	0.25	...	0.00	1.00
nc5ab8667901946a	era1	0.75	...	0.25	0.25
n84e624e4714a7ca	era1	0.00	...	0.75	1.00

[he] Learning to Protect Communications with Adversarial Neural Cryptography. Martín Abadi, David G. Andersen.

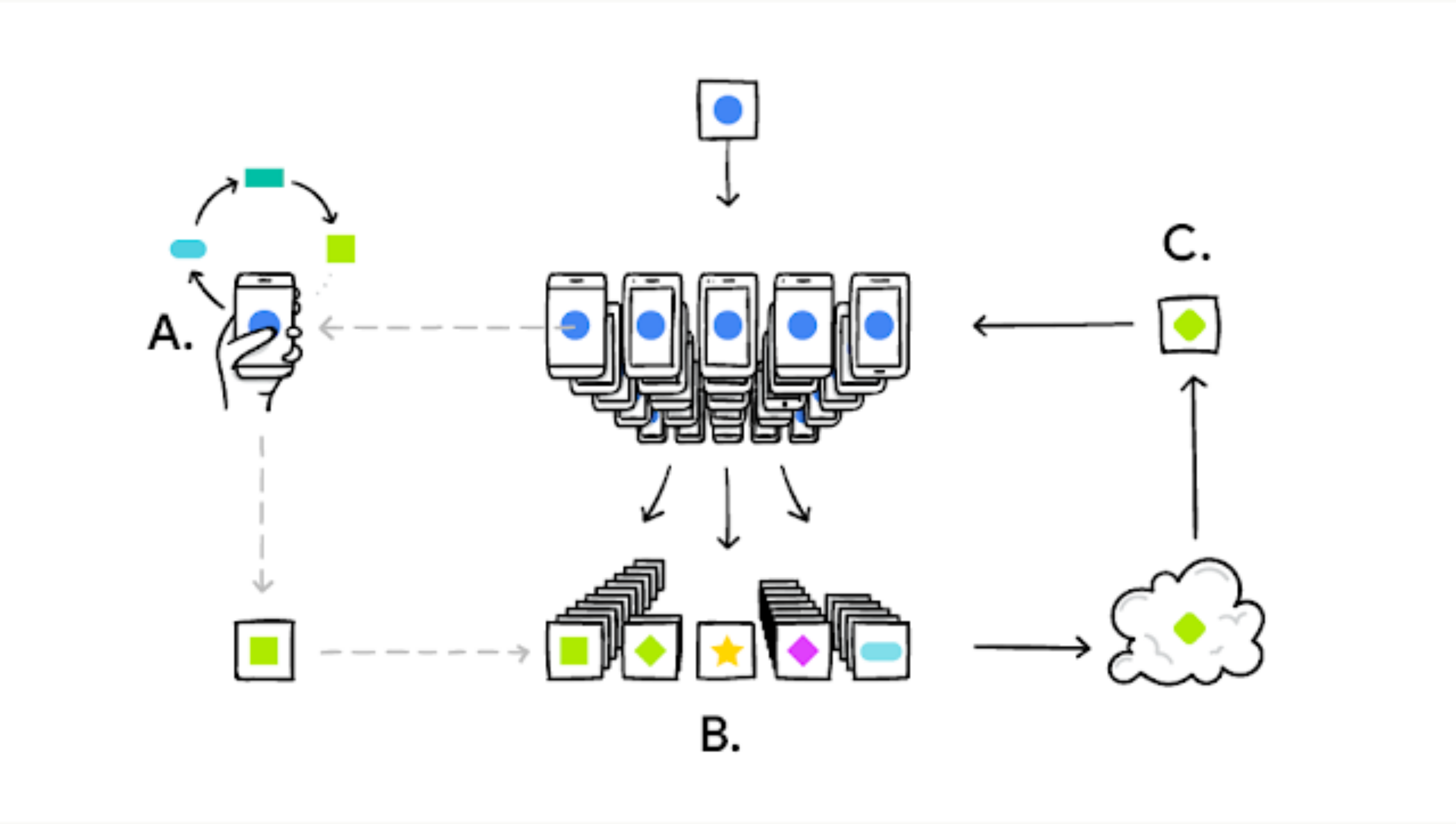
Data Stewardship: TensorFlow Federated

Problem

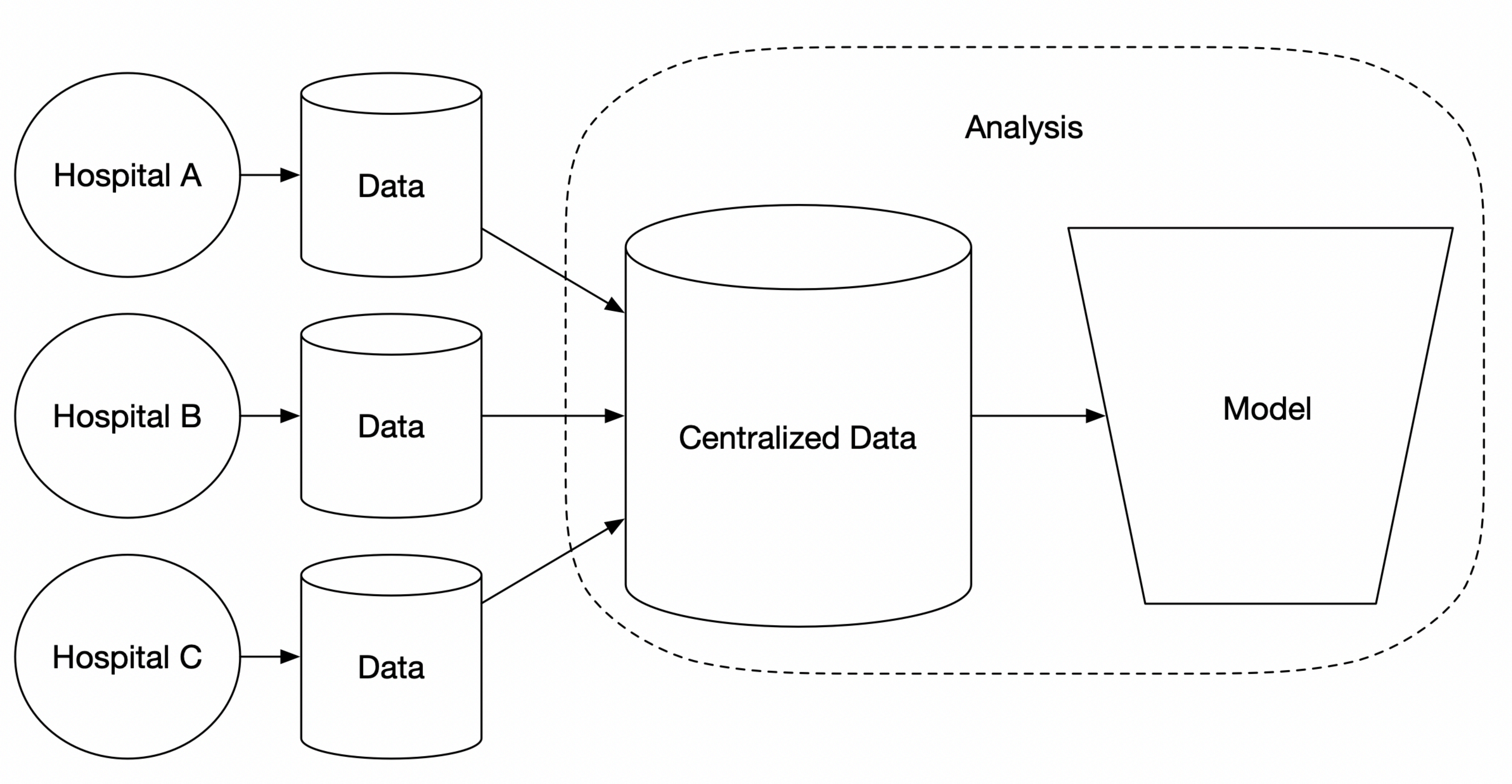
Improve predictive keyboard input, without sharing private data

Solution

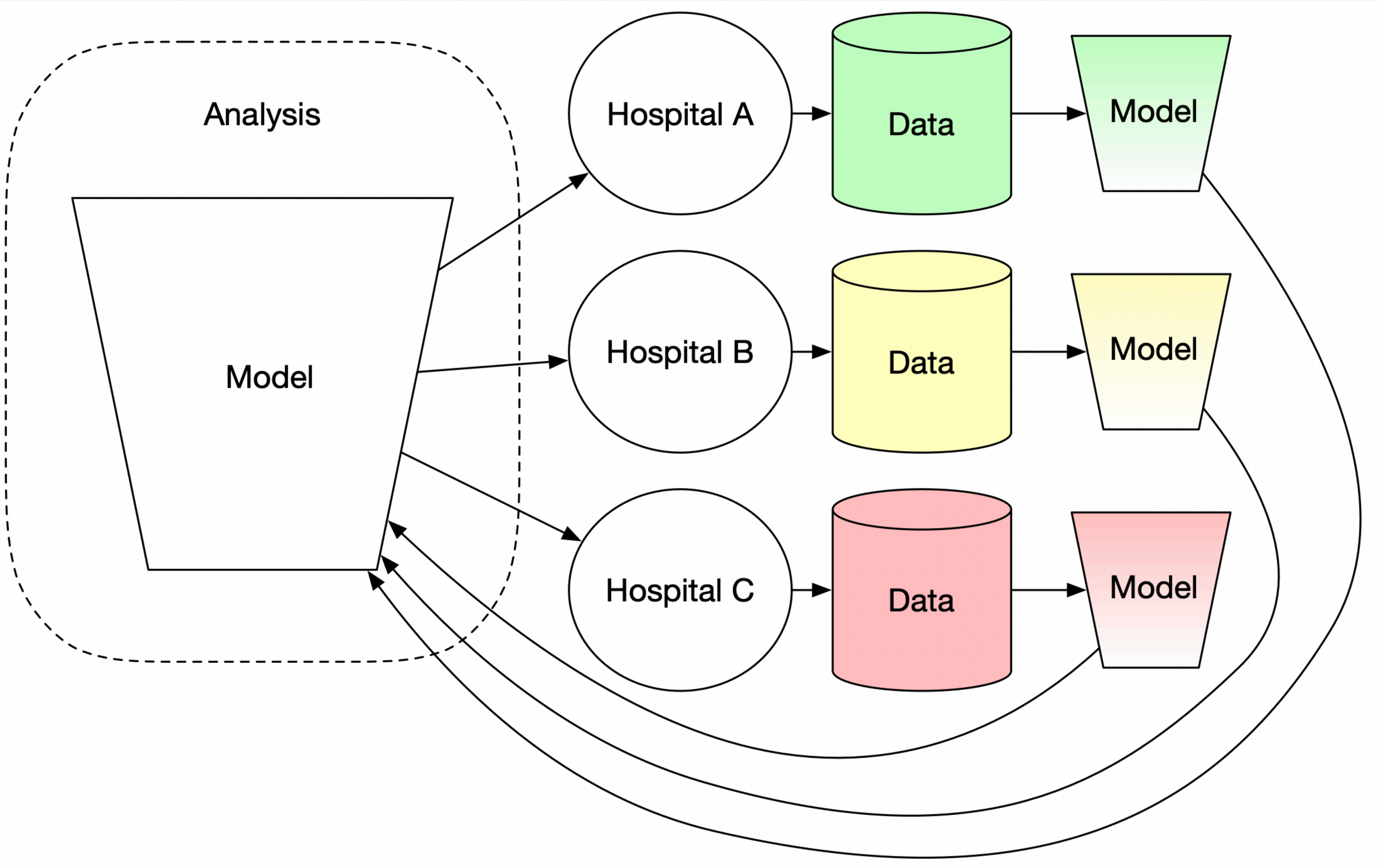
A. Keep data on phone, train locally
B. Merge trained models
C. Update consensus model in cloud



Data Stewardship: Centralized Data & Training



Data Stewardship: Federated Data & Training



 Factor AI. 2019.

Q&A

sooheon@factor.ai
paul@factor.ai

Possible Future Applications

- Variational autoencoder (VAE): chemical structure generation[6]
- One-shot learning: drug discovery (novel targets, small data)
- Hypothesis generation from learned features

[6] Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. Front. Env. Sci. 2016, 3, 1.

Types of Analysis

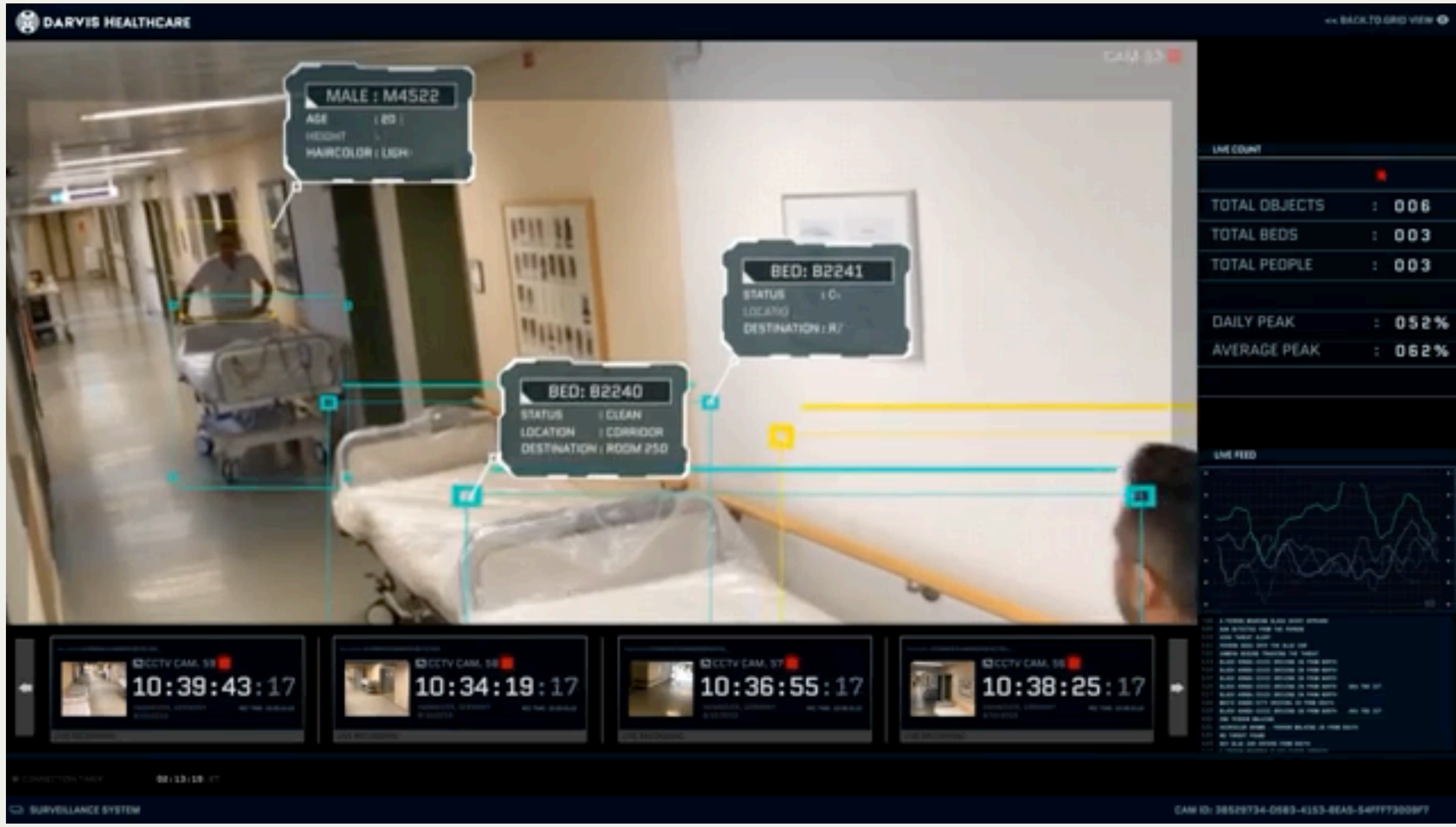
Type	Descriptive	Predictive	Prescriptive
Goal	Answer 5Ws	Predict the unknown	Prescribe the best course of action
Medical	Diagnosis	Prognosis	Treatment

Computer Vision

Hospital Bed Detection

- Computer vision on CCTV cameras to detect unclean, idle hospital beds
- Alert situation room for nurse dispatch
- Goal: ~2 more surgeries scheduled per day

Pilot under development in Germany



"All classifications in this world lack sharp boundaries, and all transitions are gradual."

– Alexandr Solzhenitsyn

When to use Deep Learning?

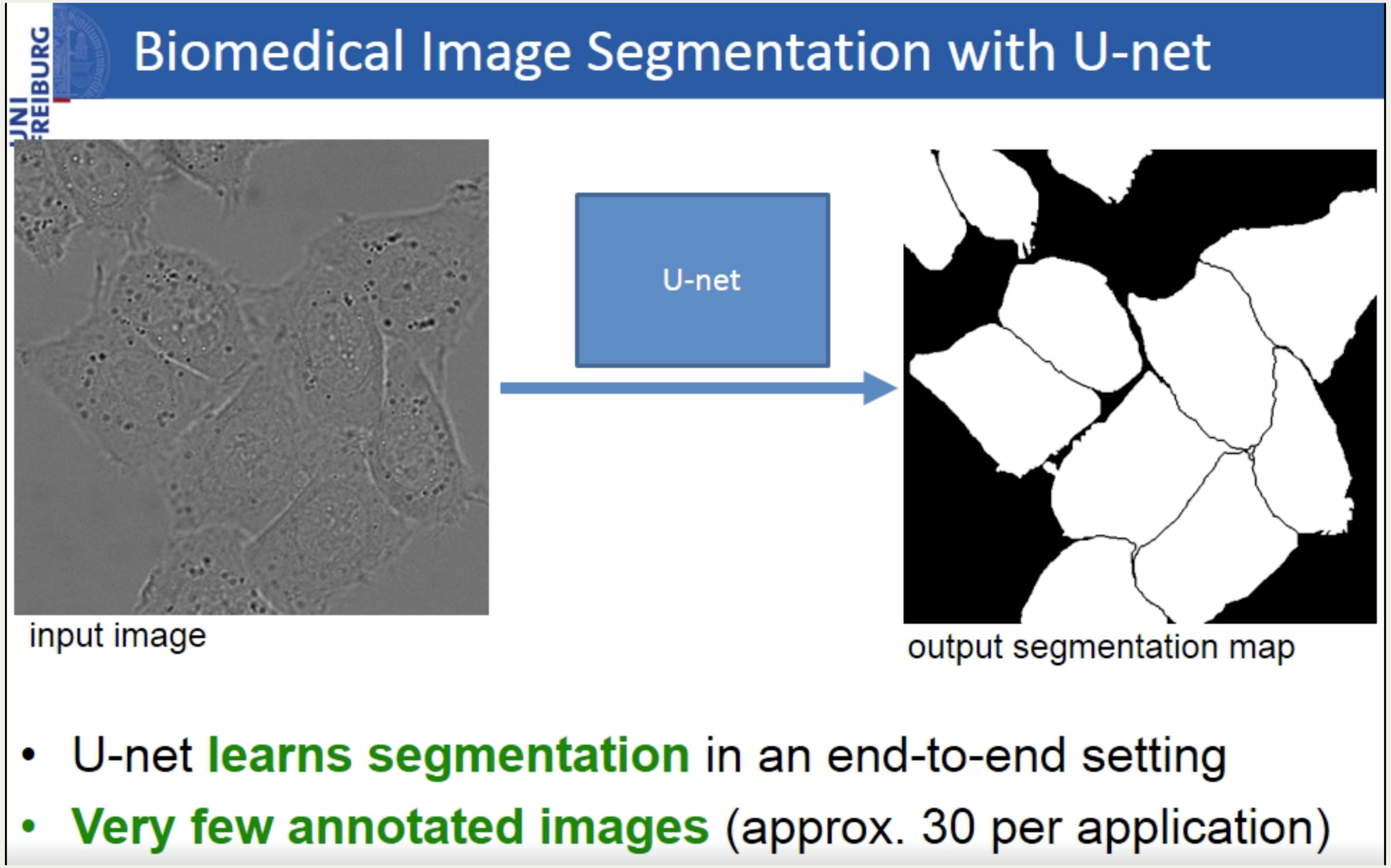
1. You have lots of data
2. Data is complex
3. ??? 🤔

Deep Learning scales well with…

1. Data quantity
    - Quantity of dataset, or similar datasets
2. Task diversity
    - The number of different inferences you can make on the same data
3. Feature complexity
    - Difficulty of engineering features traditionally
    - Sparsity, noise

Why? Deep learning is teaching by example. Robust models learn from many, diverse, examples.

U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015



Clinically applicable deep learning for diagnosis and referral in retinal disease[r]

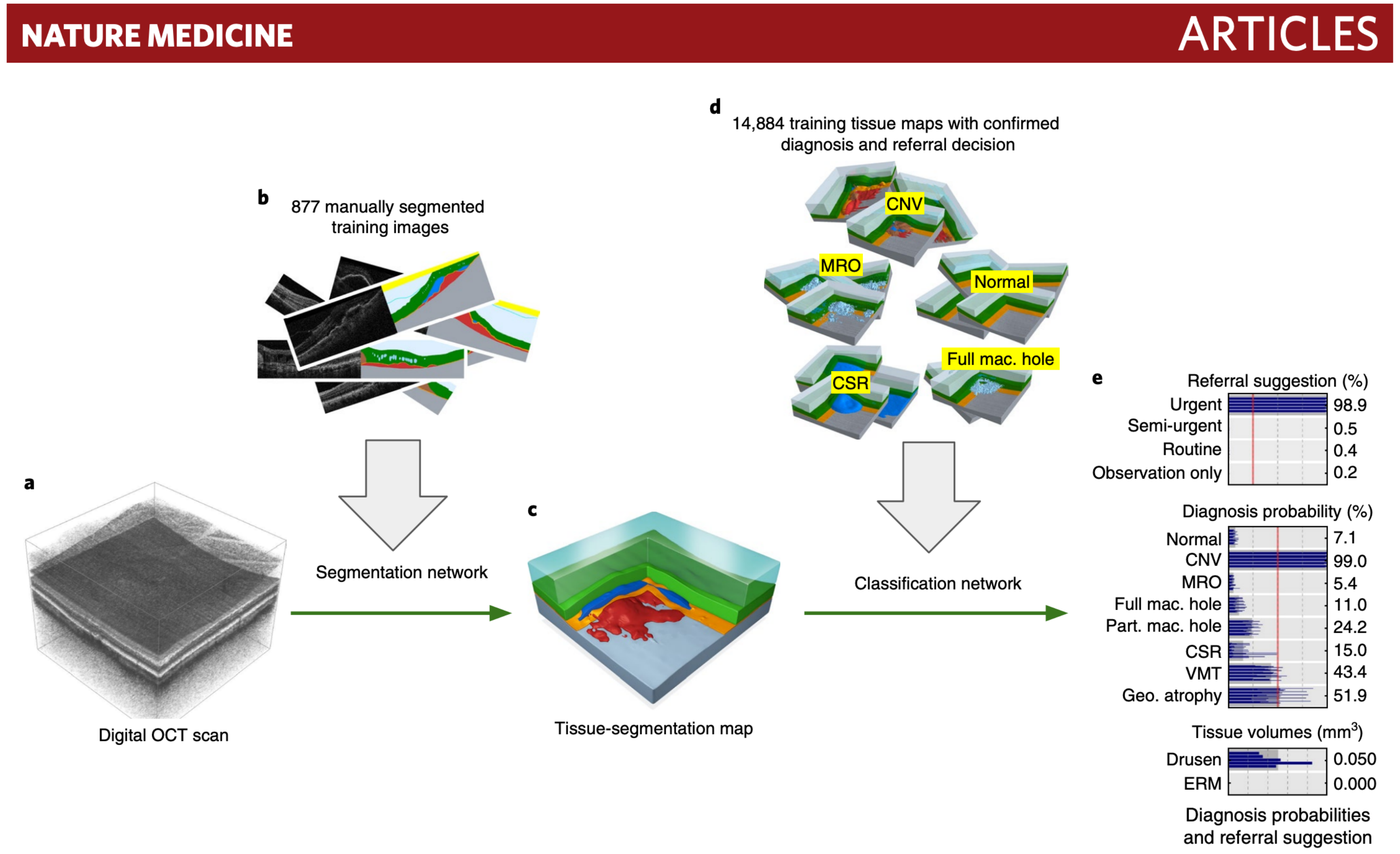


Fig. 1 | Our proposed AI framework. **a**, Raw retinal OCT scan ($6 \times 6 \times 2.3\,mm^3$ around the macula). **b**, Deep segmentation network, trained with manually segmented OCT scans. **c**, Resulting tissue segmentation map. **d**, Deep classification network, trained with tissue maps with confirmed diagnoses and optimal referral decisions. **e**, Predicted diagnosis probabilities and referral suggestions.

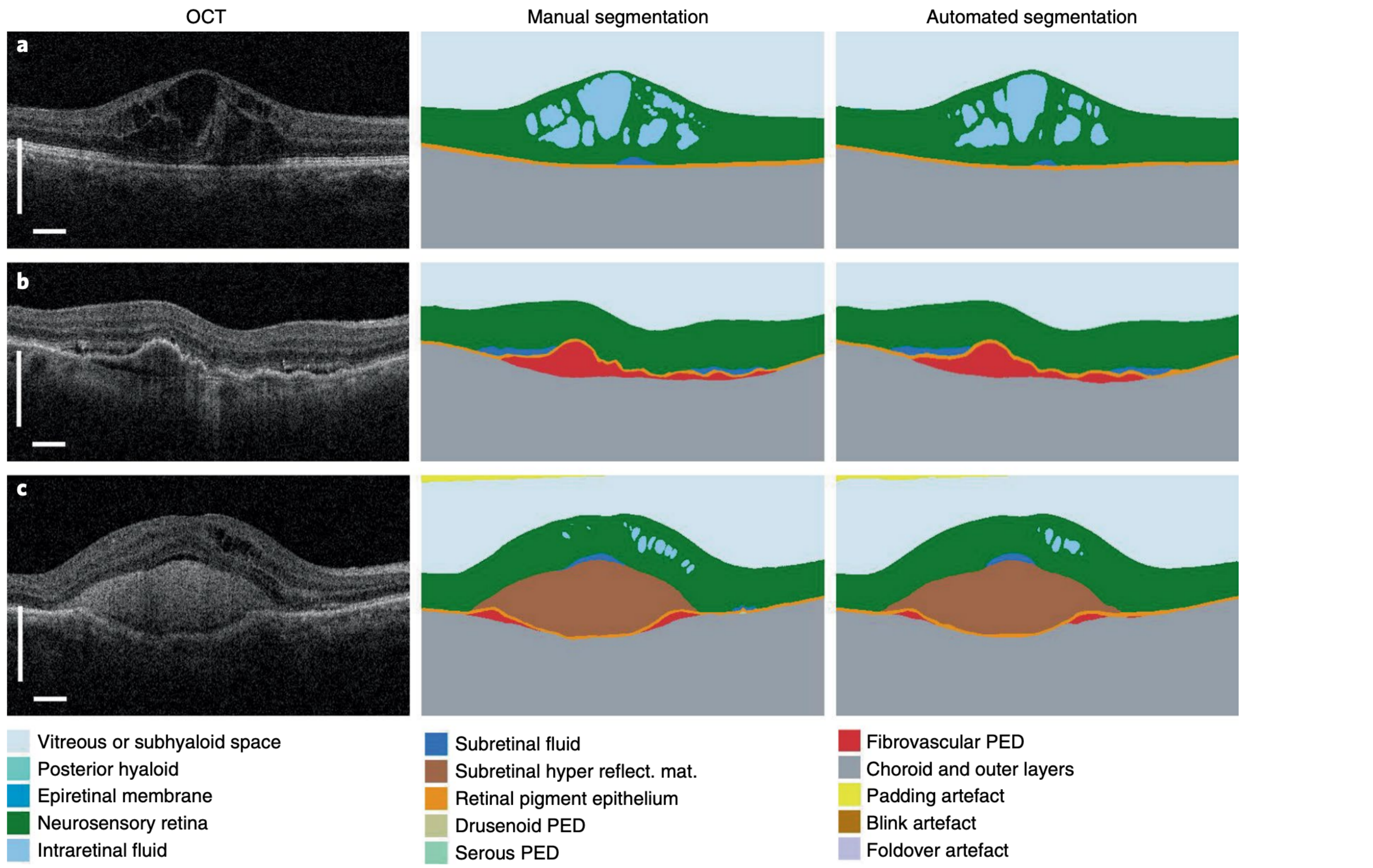


Fig. 2 | Results of the segmentation network. Three selected two-dimensional slices from the $n = 224$ OCT scans in the segmentation test set (left) with manual segmentation (middle) and automated segmentation (right; detailed color legend in Supplementary Table 2). **a**, A patient with diabetic macular edema. **b**, A patient with choroidal neovascularization resulting from age-related macular degeneration (AMD), demonstrating extensive fibrovascular pigment epithelium detachment and associated subretinal fluid. **c**, A patient with neovascular AMD with extensive subretinal hyperreflective material. Further examples of the variation of pathology with model segmentation and diagnostic performance can be found in Supplementary Videos 1–9. In all examples the classification network predicted the correct diagnosis. Scale bars, 0.5 mm.

[r] Ronneberger et al., Aug 2018

Clinically applicable deep learning for diagnosis and referral in retinal disease[r]

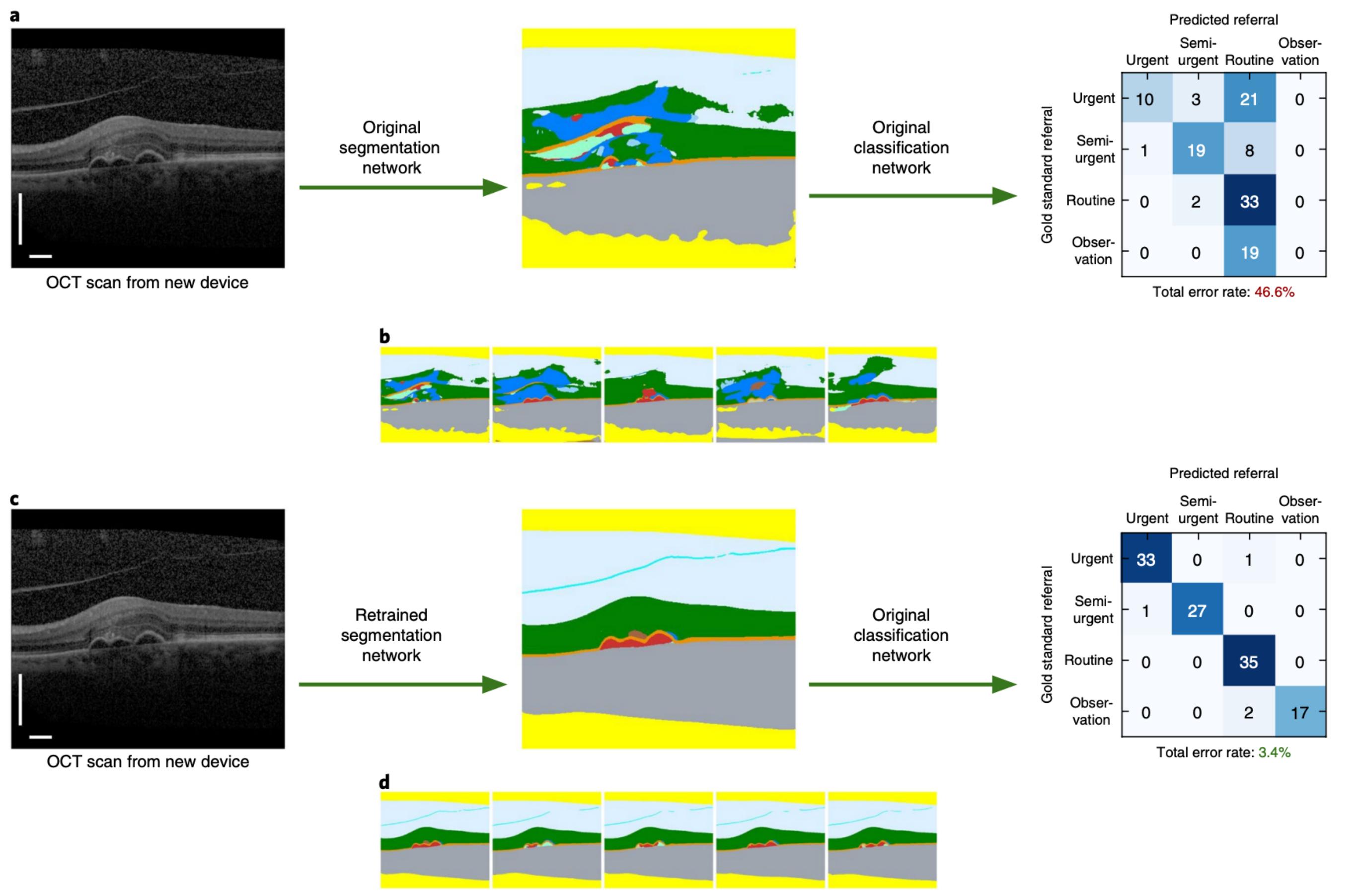


Fig. 4 | Generalization to a new scanning device type. a, Low performance of original network on OCT scans from the new device type 2. Left, the selected slice shows the different appearance of structures in device type 2. Middle, a poor quality segmentation map created with our original segmentation network (color legend in Supplementary Table 2). Right, resulting performance on a new test set of $n = 116$ patients. The confusion matrix shows patient numbers for the referral suggestion. **b**, All five segmentation hypotheses from our original network. The strong variations show the large uncertainty. **c**, High performance was attained on the device type 2 test set ($n = 116$) after retraining the segmentation network with OCT scans from device type 1 and device type 2. The classification network is unchanged. **d**, All five segmentation hypotheses from the retrained segmentation network. The network is confident in the interpretation of most structures, and just highlights the ambiguities in the sub-retinal pigment epithelium (RPE) space. Scale bars: 0.5 mm.

[r] Ronneberger et al., Aug 2018