



Applications and Framework

Dr. Fayyaz Minhas

Department of Computer Science

University of Warwick

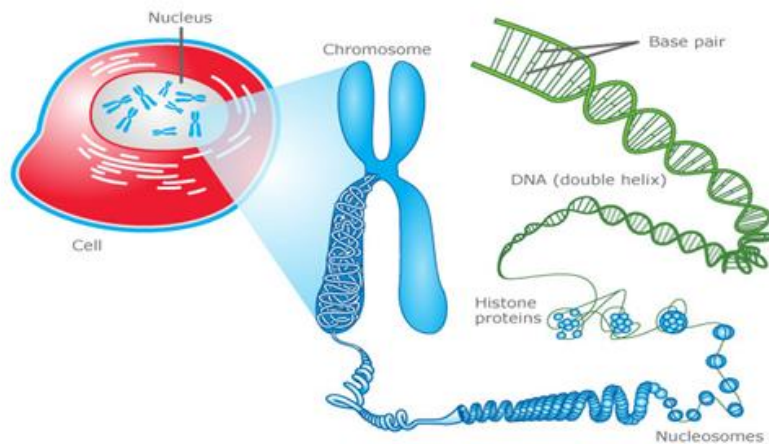
<https://warwick.ac.uk/fac/sci/dcs/teaching/material/cs909/>

Applications of Machine Learning

- An ability that I would like you to learn is to identify how to use machine learning in different domains.
- Machine learning can be applied in a wide array of real-world applications

Data, Big Data and Data Science

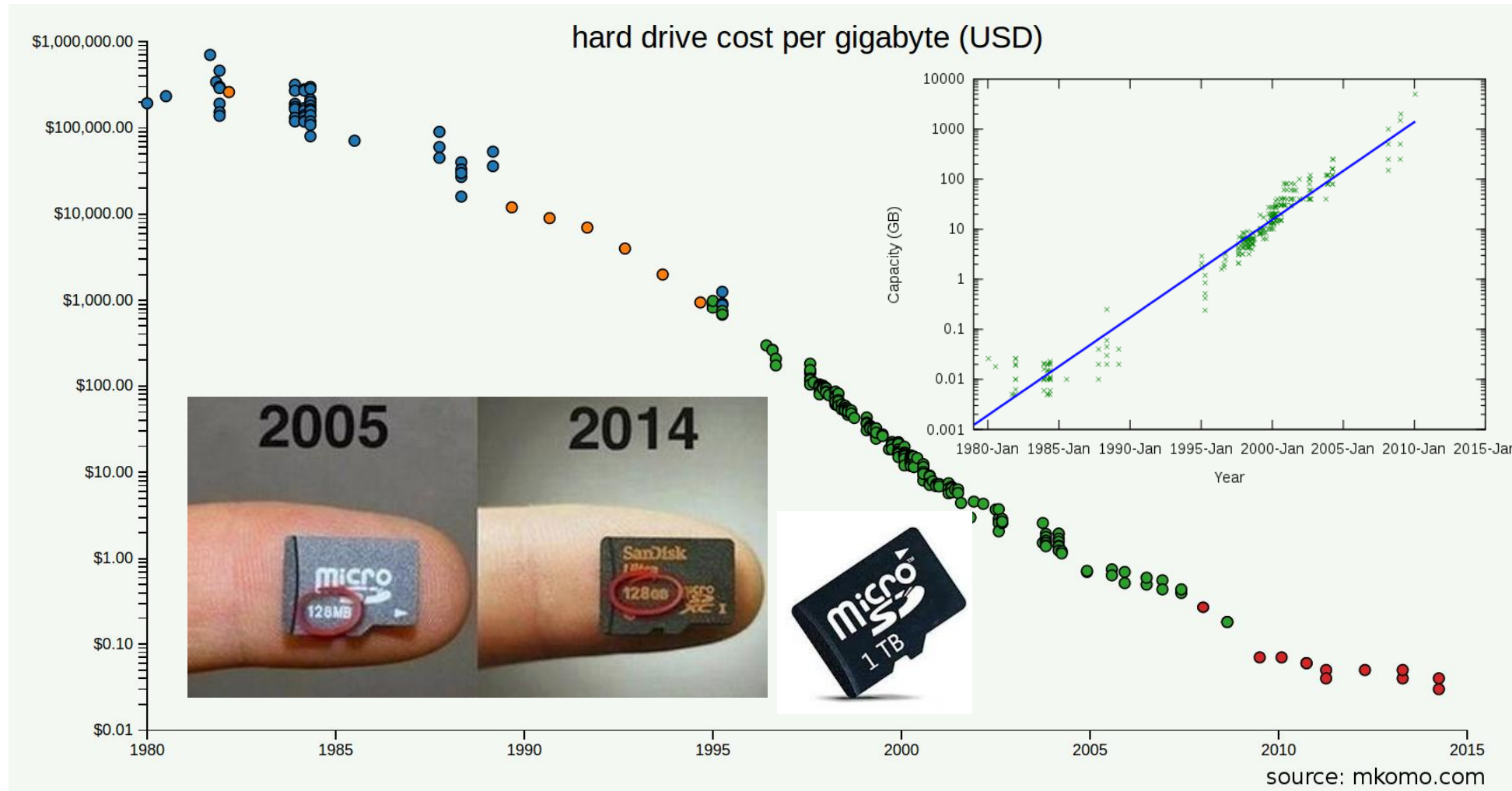
- We are going through an age of “Big Data”
 - Humans are the only biological entity that can store more data outside its body than inside it



If the Digital Universe were represented by the memory in a stack of tablets, in 2013 it would have stretched two-thirds the way to the Moon*

By 2020, there would be 6.6 stacks from the Earth to the Moon*

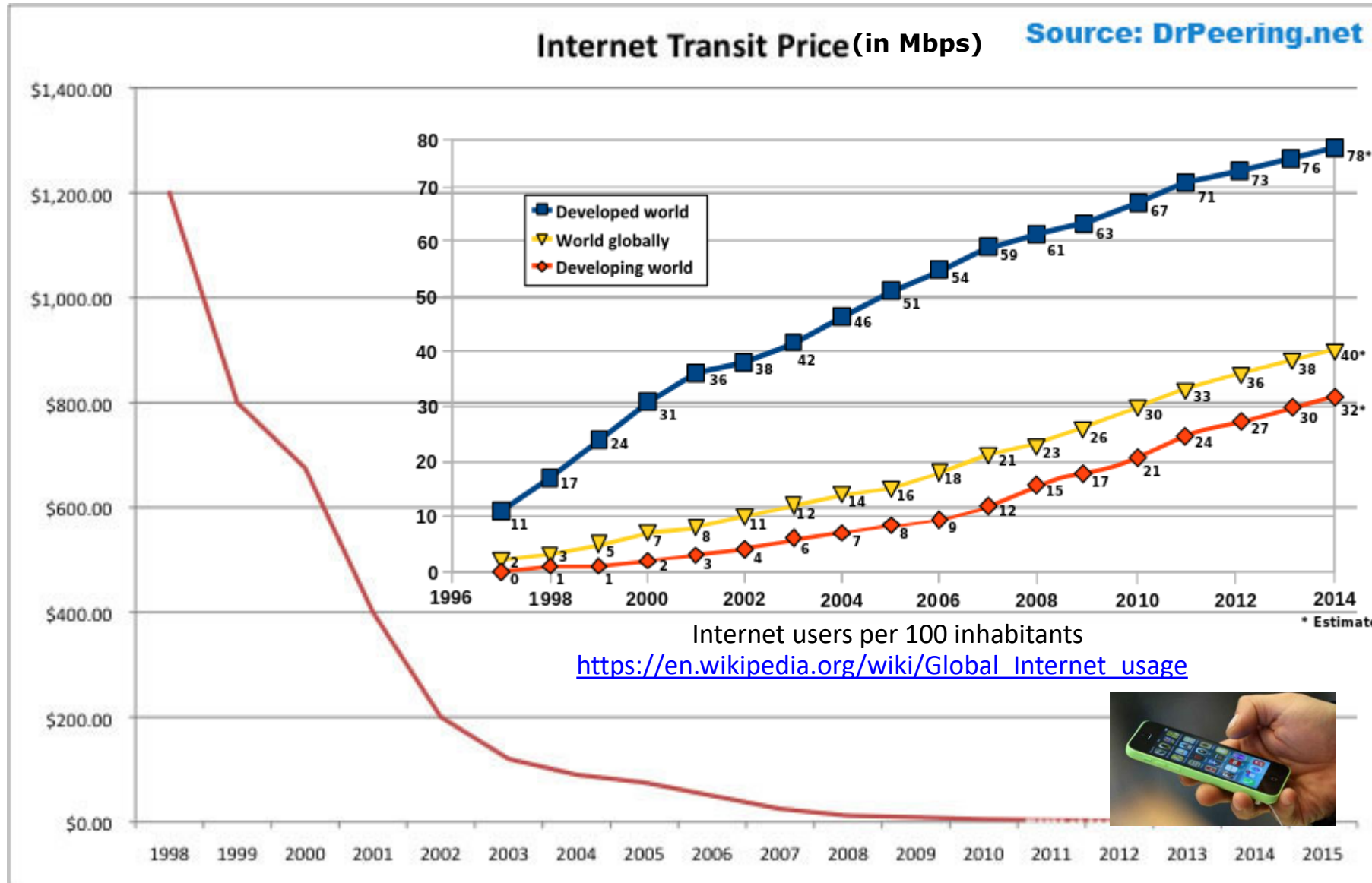
Age of Big Data



Our Emergent Digital Future

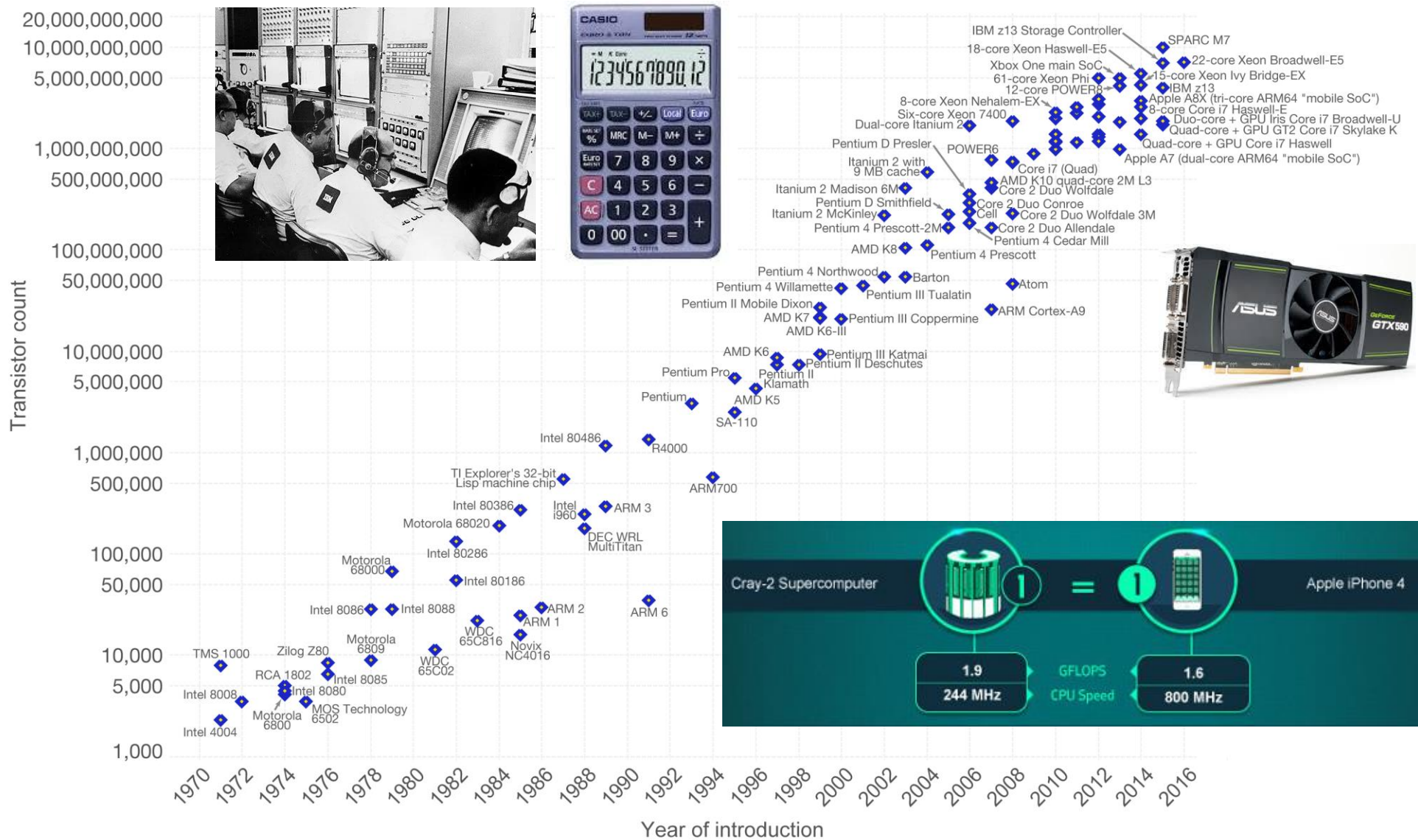
<http://www.digitaltonto.com/2011/our-emergent-digital-future/>

Age of Big Data



Moore's Law – The number of transistors on integrated circuit chips (1971-2016)

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are strongly linked to Moore's law.

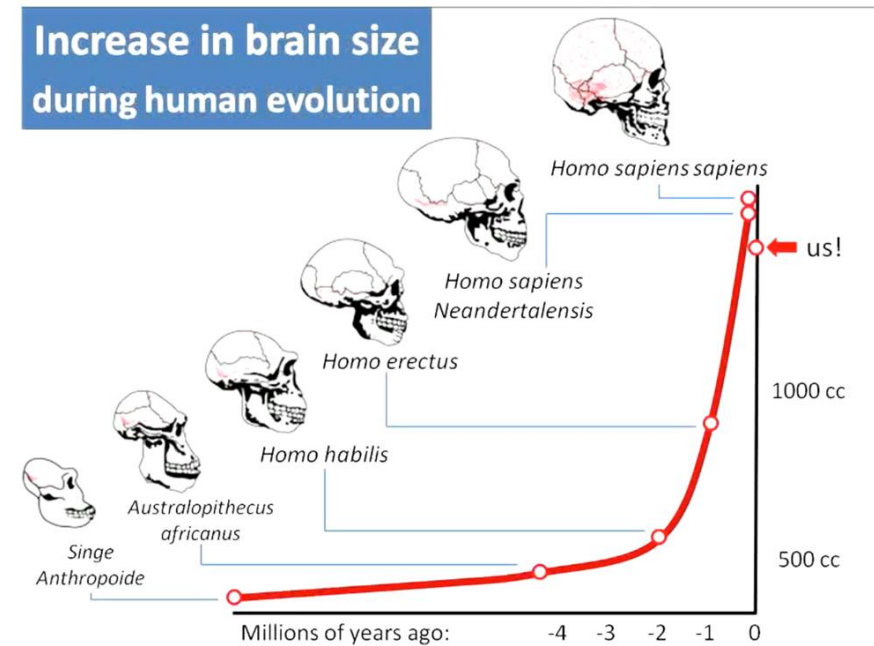


Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)
The data visualization is available at [OurWorldinData.org](https://www.ourworldindata.org). There you find more visualizations and research on this topic.

Licensed under CC-BY-SA by the author Max Roser.

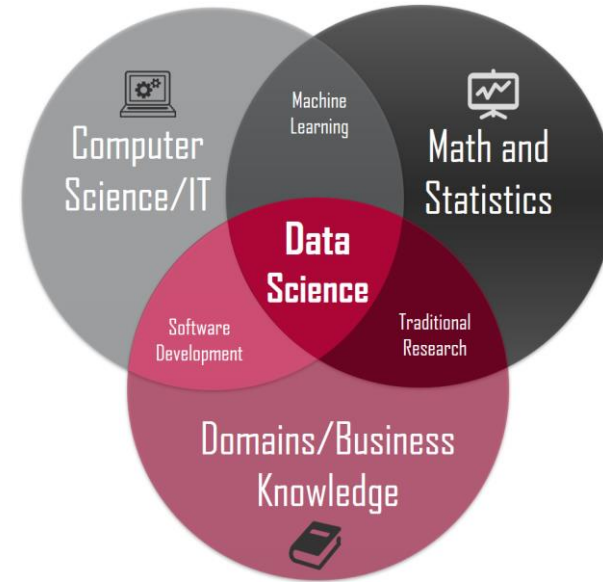
Question

- Where does the bottleneck lie?
 - Data Analysis by Humans
 - Have humans become smarter?
 - How about we automate that?



Data Science

- Use of the scientific method in the development of:
 - Processes
 - Algorithms
 - Systems
- For extraction of knowledge from structured or unstructured data
- Examples



LinkedIn's Fastest-Growing Jobs Today Are In Data Science And Machine Learning



Louis Columbus, CONTRIBUTOR

FULL BIO

Opinions expressed by Forbes Contributors are their own.

YOUR READING LIST

The Economist

Topics Current edition More

Regulating the internet giants

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules



TECH • SCIENCE

'Sexiest Job' Ignites Talent Wars as Demand for Data Scientists Soars



Growth in data scientist searches per million searches



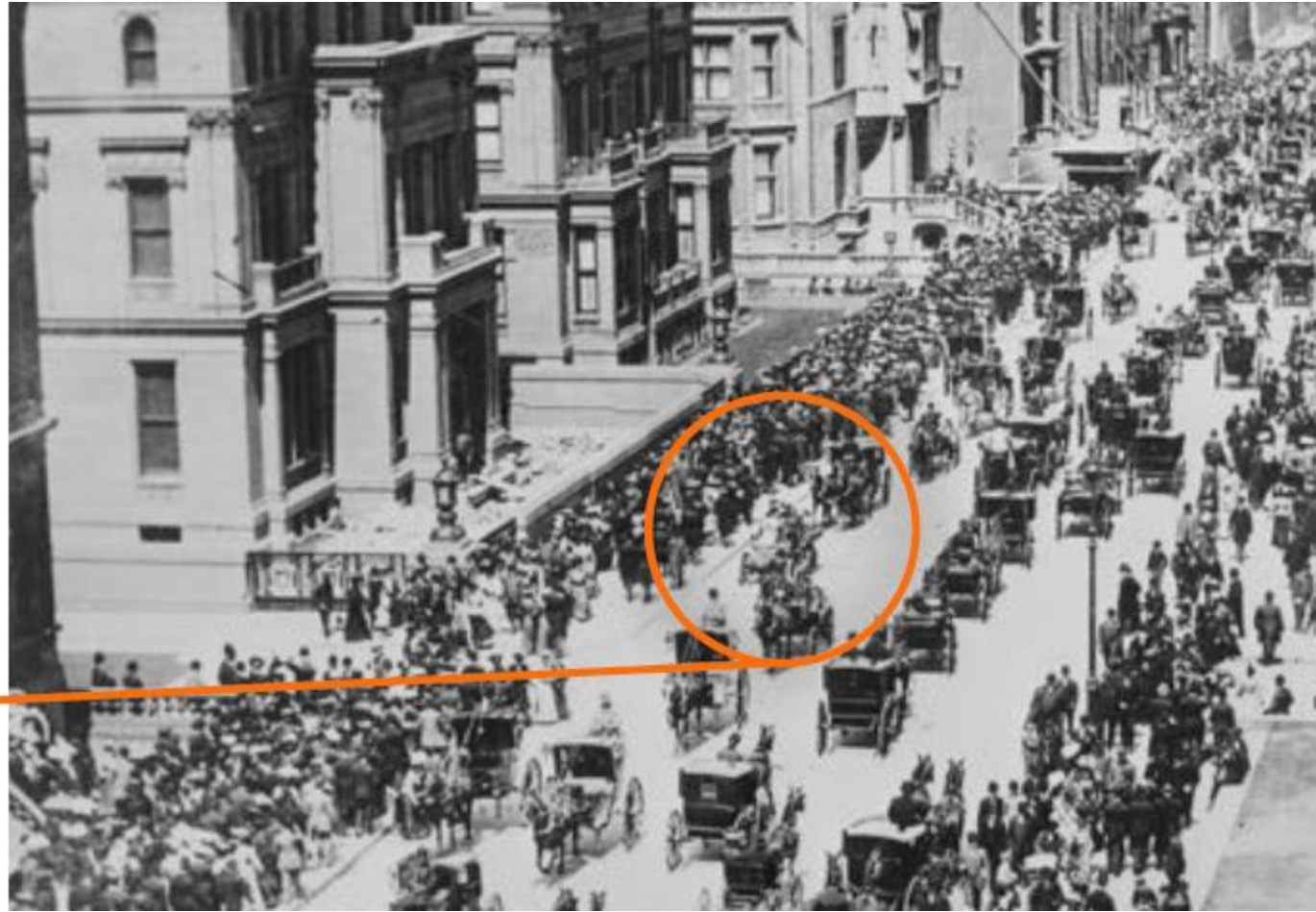
indeed

5th AVE NYC

1900

Where is

the
car?



5th AVE NYC
1913

Where is
the
horse?



Applications

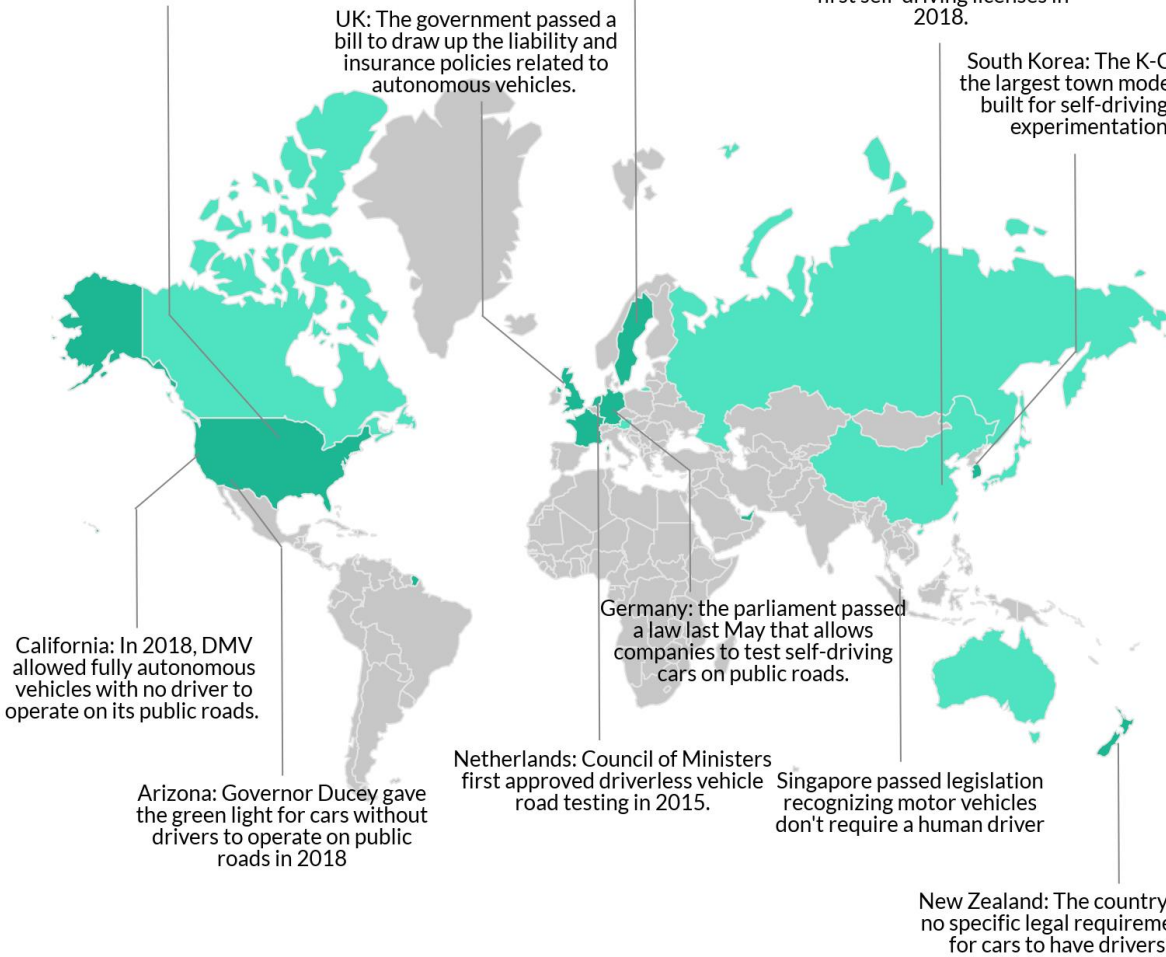
US: 33 states accommodate self-driving vehicles on public roads.

Sweden: Last December, Volvo launched its Drive Me project, which provided self-driving cars to a number of people.

China: Shanghai issued its first self-driving licenses in 2018.

UK: The government passed a bill to draw up the liability and insurance policies related to autonomous vehicles.

South Korea: The K-City is the largest town model ever built for self-driving car experimentation.



Autonomous Vehicle Access to Public Roads:

None or unknown Some access High access

Graphic: Tony Peng | Synced

Handwriting Recognition / OCR

From
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

Nov 10, 1999

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim

From
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

Nov 10, 1999

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

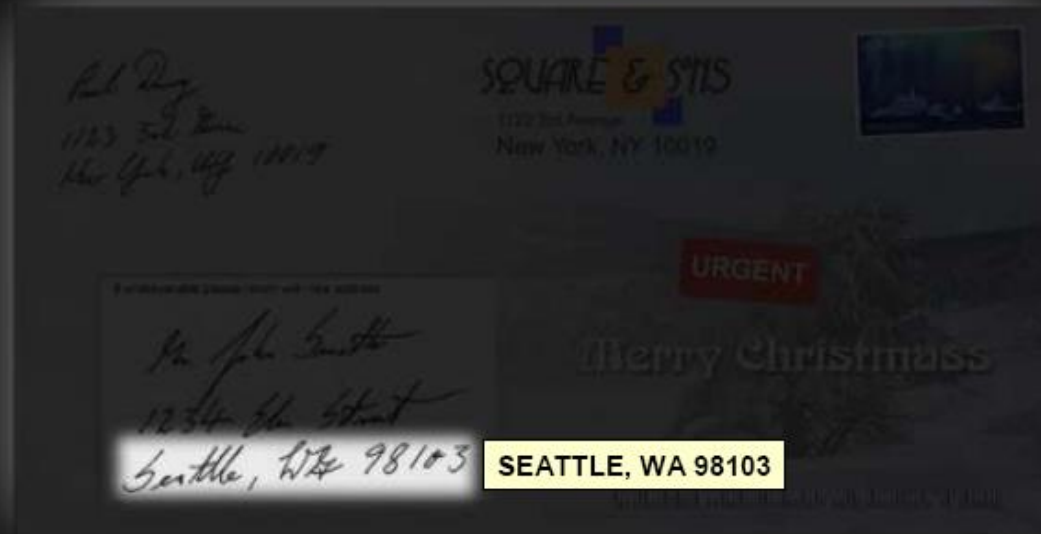
Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim

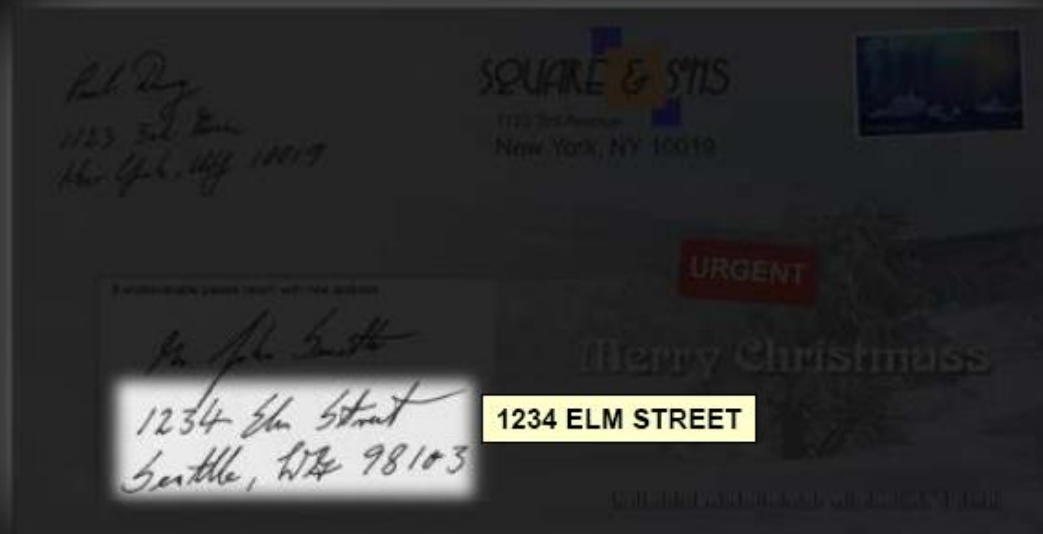
The Letter



1978: First Postal Code Reader Worldwide



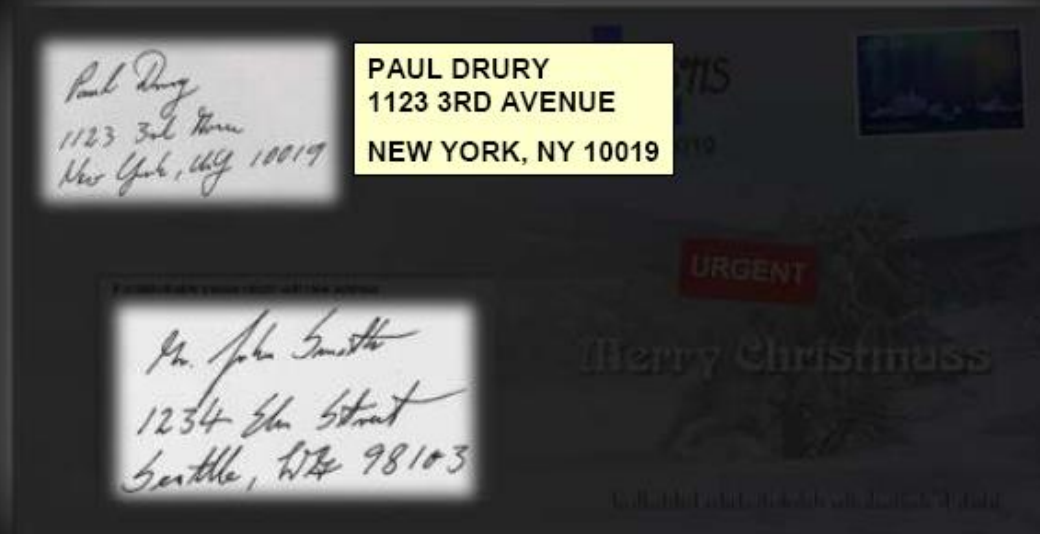
1982: First Address Reader Worldwide



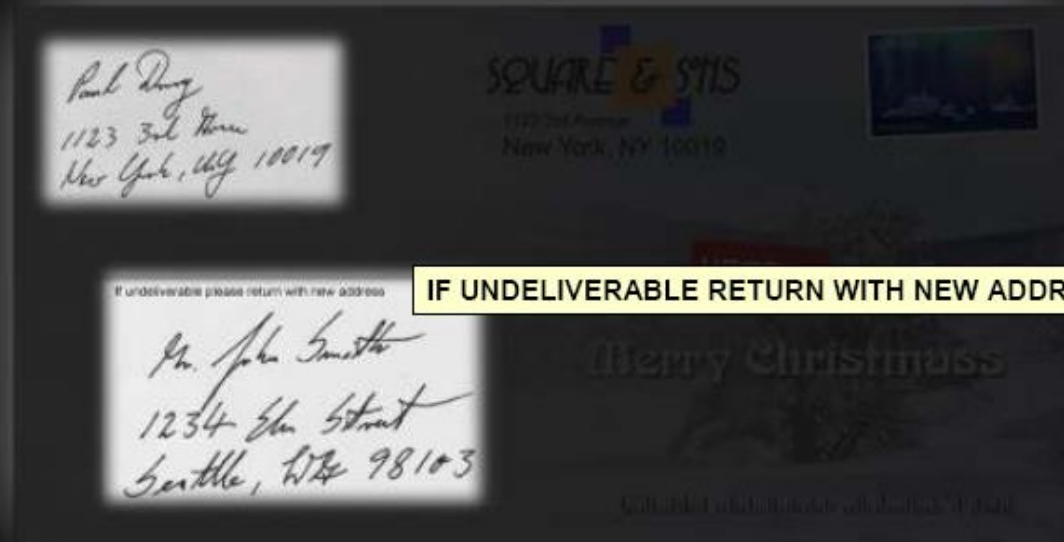
1984: First Multi Line Reader



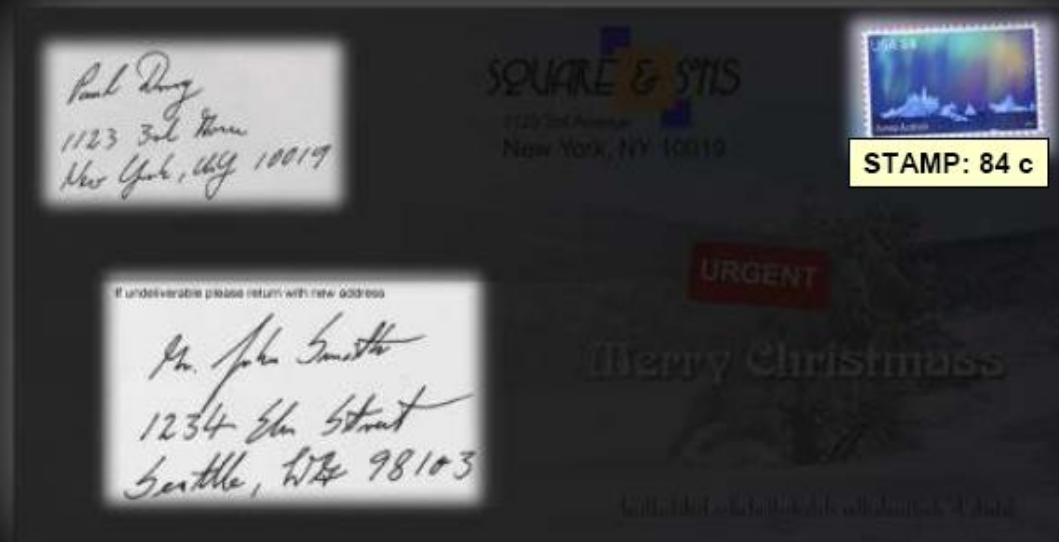
1996: First Sender's Address Reader



1998: First Full Text Reading



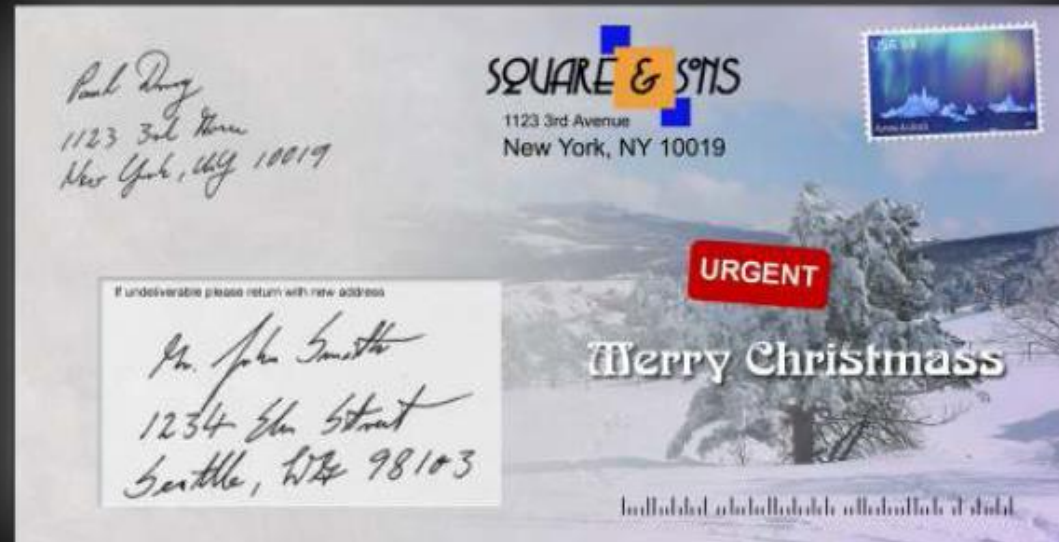
2000: First Graphics Recognition



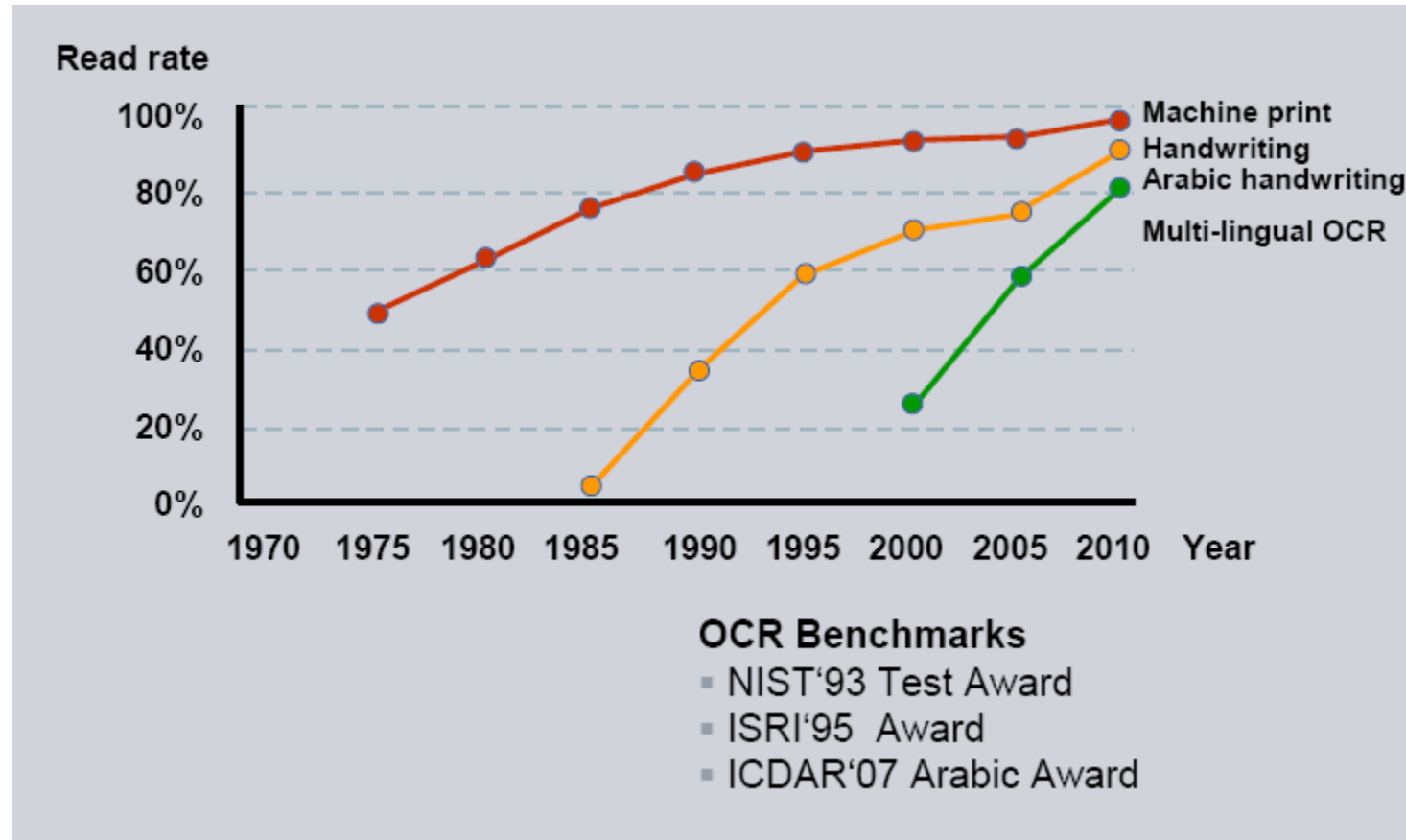
2004: First Full Recognition



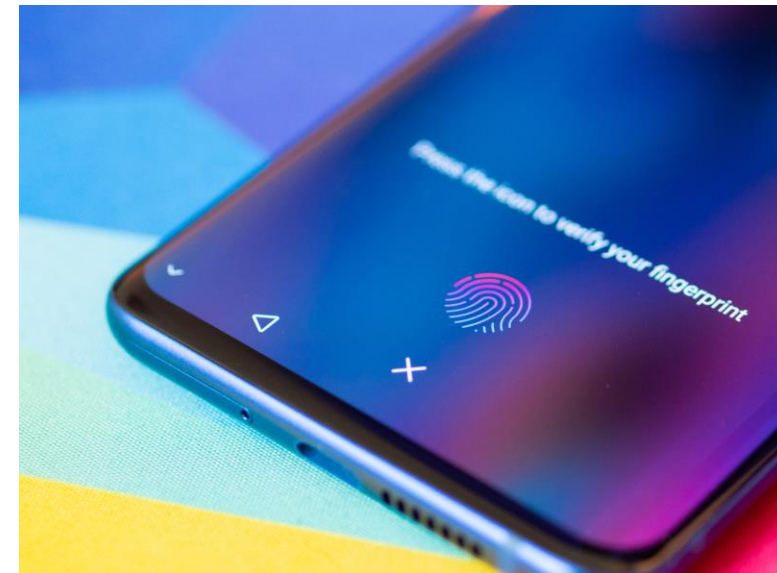
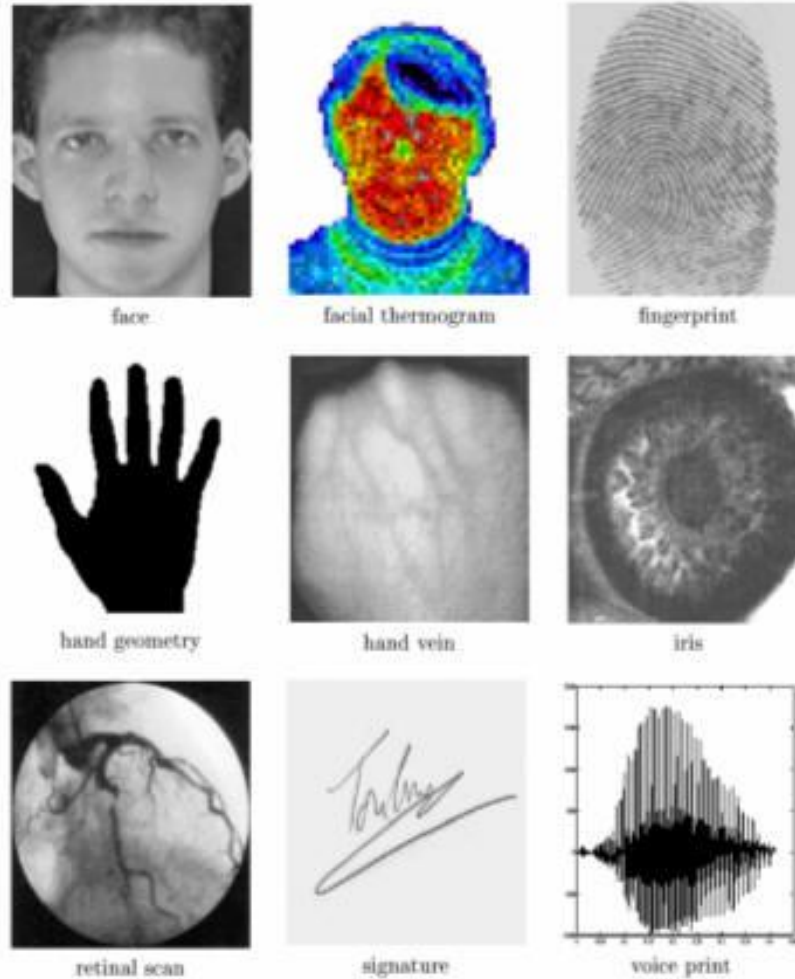
2008: Recognition on Both Sides of Envelope



OCR Accuracy



Applications: Biometrics



Facebook Friends Tagging

We've Suggested Tags for Your Photos

We've automatically grouped together similar pictures and suggested the names of friends who might appear in them. This lets you quickly label your photos and notify friends who are in this album.

Tag Your Friends

This will quickly label your photos and notify the friends you tag. [Learn more](#)



Who is this?



Who is this?



Who is this?



Who is this?



Who is this?



Who is this?

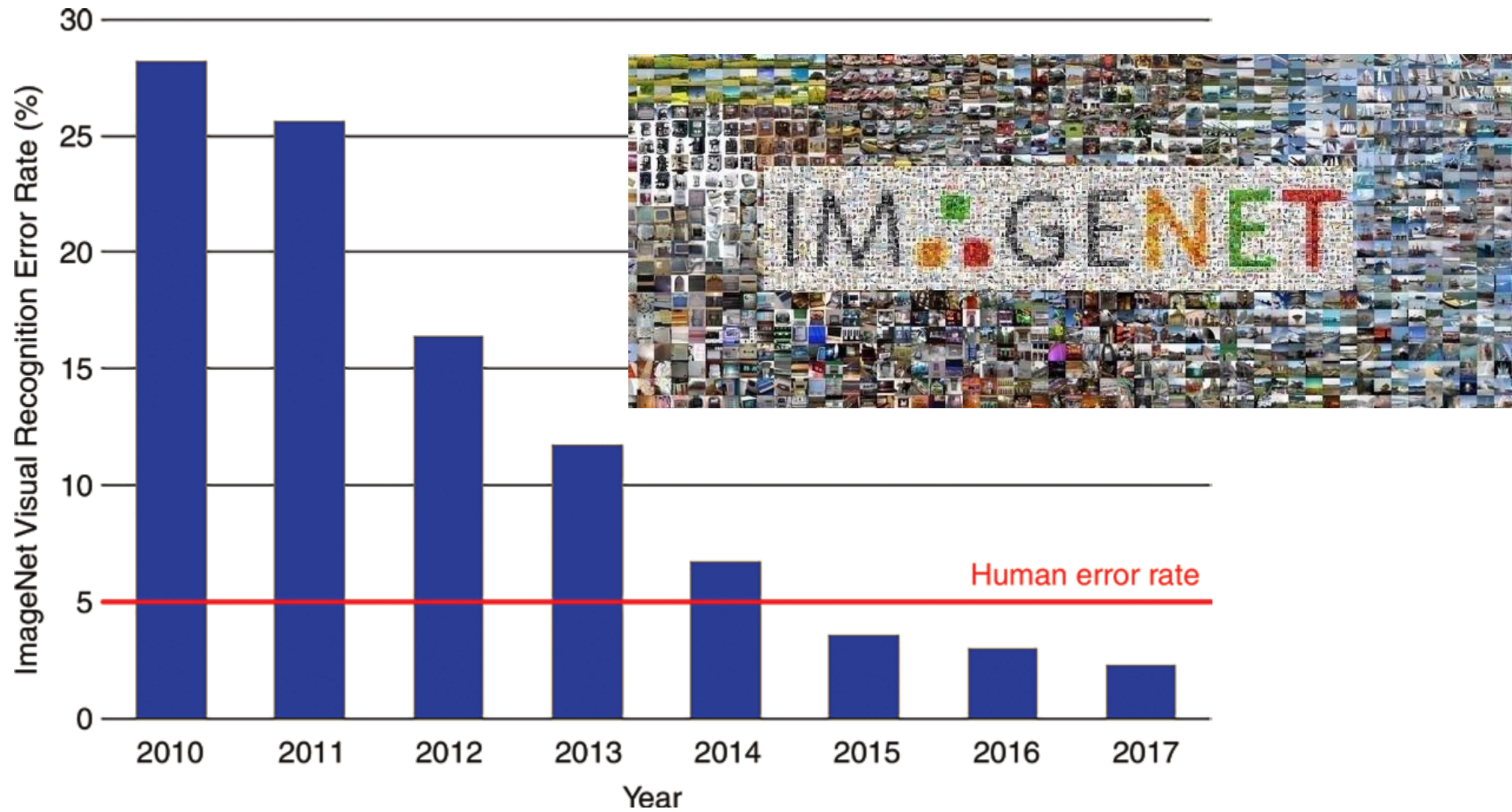


Francis Luu



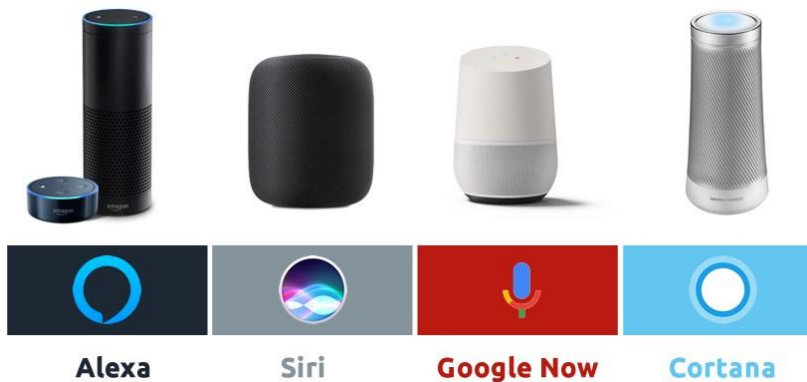
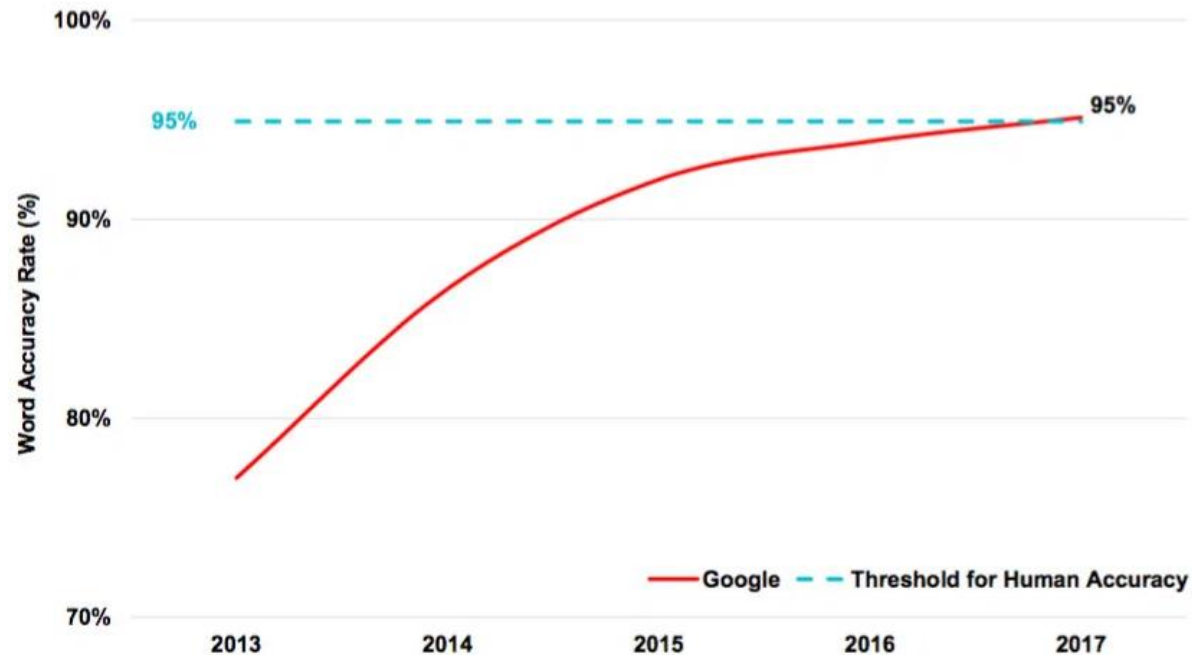
Skip Tagging Friends

Save Tags



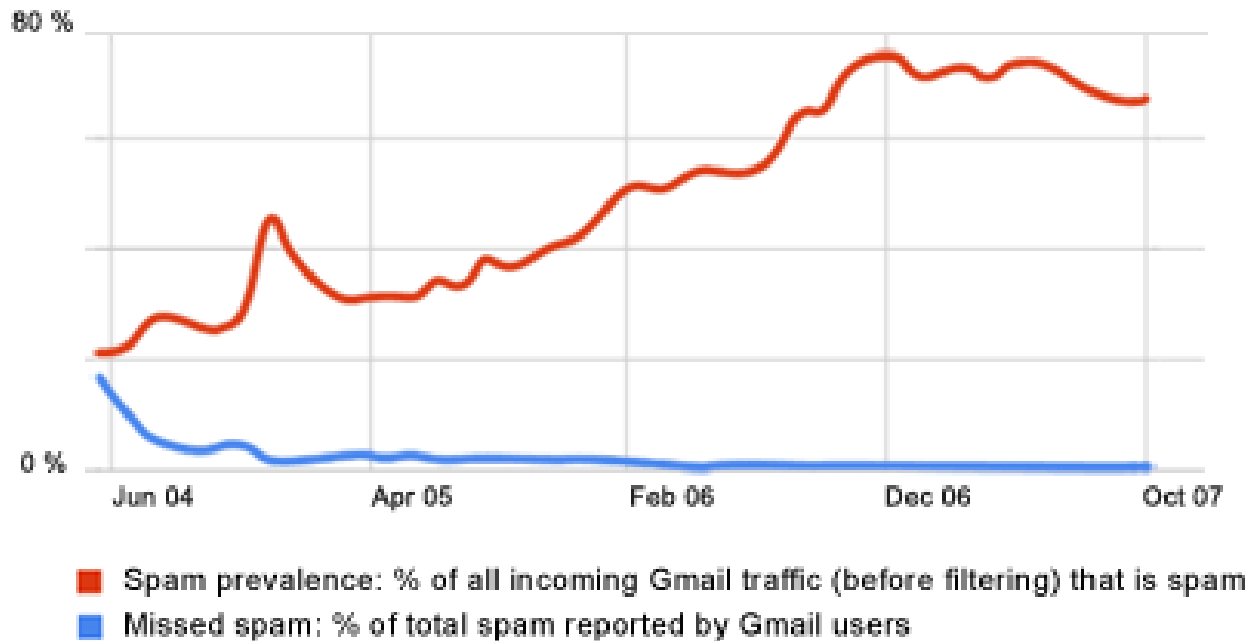
Voice Assistants

Google Machine Learning
Achieving Higher Word Accuracy, 2013-2017



Gmail: ML in NLP

More spam, but not in Gmail Inboxes



As the amount of spam has increased, Gmail users have received less of it in their inboxes, reporting a rate less than 1%.

William Shakespeare ¶

Transformer AI poetry: Poetry classics as reimaged and rewritten by an artificial intelligence.

“Sonnet 18”

William Shakespeare

Shall I compare thee to a summer's day?

Thou art more lovely and more temperate:

Rough winds do shake the darling buds of May,

And summer's lease hath all too short a date;

A winter's day, when beams of sun are few,

And skies are grey with clouds intemperate,

Is to thy beauty somehow sordid grown.

Her lovely-lovely eye is quite shut up,

Like windows pale before a sunny shower.

The earth that's cold and moist, like to my mistress,

When once that she unbrac'd, her lovely jewel,

Grows as both ugly and uncourtly.

Poor soul, the center of my motley host,

See, now they request thy sovereign presence;

The din of merry hearts hath brought me thus

To greet thee, thou unsoil'd and saint-like beauty.

Recommender Systems



- Recommend movies based on user preferences, interests and likes
- Similar ideas for facebook...
 - Find friends that share your interests

Keyboard acoustics

Text recognized by the HMM classifier, with cepstrum features (underlined words are wrong),

the big money fight has drawn the shoporo od dosens
of companies in the entertainment industry as well
as attorneys gnnerals on states, who fear the
file shading softwate will encourage illegal acyivitt,
srem the grosth of small arrists and lead to lost cobs and
dimished sales tas revenue.

Text after spelling correction using trigram decoding,

the big money fight has drawn the support of dozens
of companies in the entertainment industry as well as
attorneys generals in states, who fear the film sharing
software will encourage illegal activity, stem the growth
of small artists and lead to lost jobs and finished sales
tax revenue.

Original text. Notice that it actually contains two typographical errors, one of which is fixed by our spelling corrector.

the big money fight has drawn the support of dozens
of companies in the entertainment industry as well as
attorneys gnnerals in states, who fear the file sharing
software will encourage illegal activity, stem the growth
of small artists and lead to lost jobs and dimished sales
tax revenue.

PCR in HCI/CV

- Gesture Recognition



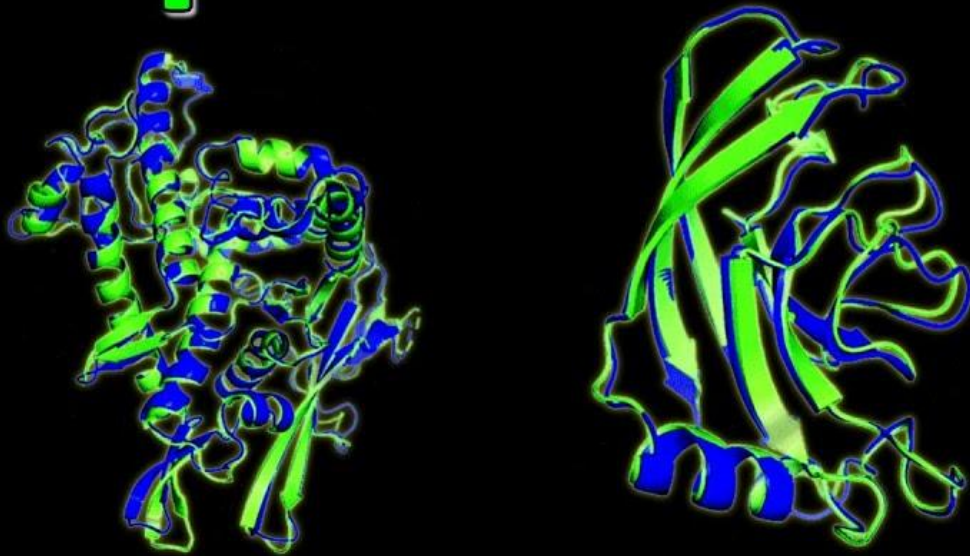
Synthesis





<https://deepmind.com/blog/article/Agent57-Outperforming-the-human-Atari-benchmark>

Google DeepMind's AlphaFold 2

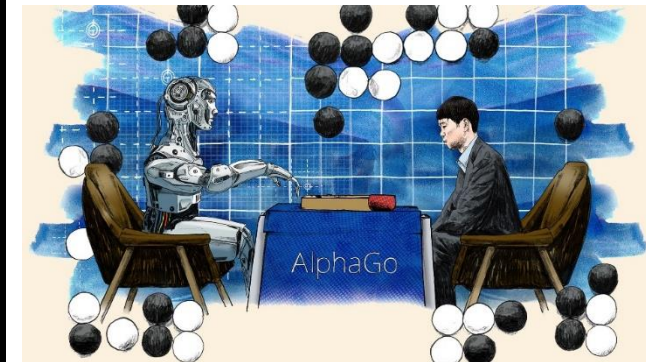


AI Breakthrough in Biology

NEWS · 30 NOVEMBER 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.



Machine Learning



What society thinks I do



What my friends think I do



What other computer scientists think I do



What my wife thinks I do

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \zeta_i \quad \max \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i x_j$$

$s.t. \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 - \zeta_i \quad \rightarrow \quad s.t. \quad 0 \leq \alpha_i \leq C$
 $\zeta_i \geq 0 \quad \sum_{i=1}^m \alpha_i y_i = 0$

$$h(\beta) = \frac{1}{2} \|\beta\|^2 + C \left[\frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i(\beta x_i)) \right]$$

↓

$$h(\beta) = \frac{1}{n} \sum_{i=1}^n \left[\frac{1}{2} \|\beta\|^2 + C * \max(0, 1 - y_i(\beta x_i)) \right]$$

What I think I do



```
Python 5.1.0 — An enhanced Interactive Python.
-> Introduction and overview of IPython's features.
quickref -> Quick reference.
elp -> Python's own help system.
bject? -> Details about 'object', use 'object??' for extra details.

n [1]: import PyTorch
ImportError Traceback (most recent call
ipython-input-1-f0a82a930083> in <module>()
--> 1 import PyTorch
ImportError: No module named 'PyTorch'

n [2]:
```

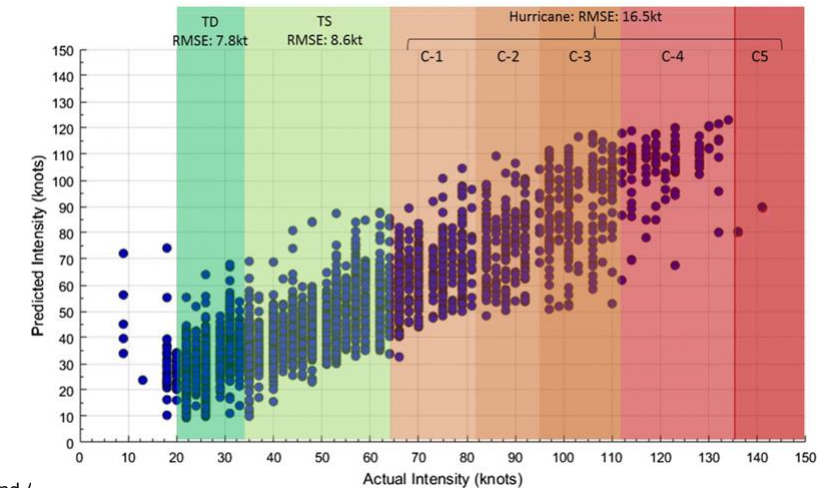
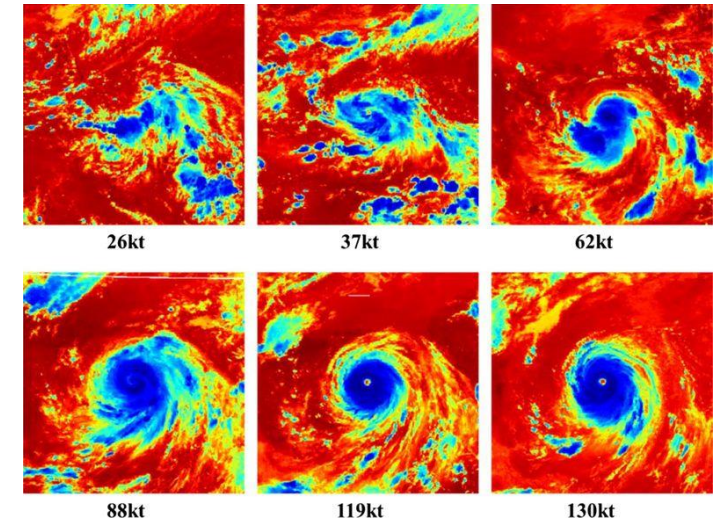
What I actually do

Applied Machine Learning

- Examples from my research

Data Science: Hurricane Intensity Prediction

- Hurricane Intensity Prediction
 - In collaboration with National Hurricane Center, USA
 - Deep-PHURIE




- Largest recorded hurricane: Typhoon Tip, Northwest Pacific, 12 October 1979
- Smallest recorded hurricane: Tropical Cyclone Tracy, near Australia, 24 December 1974

Deep-PHURIE: Deep Learning based Hurricane Intensity Estimation from Infrared Satellite Imagery, M. Dawood, A. Asif and Fayyaz Minhas, in *Neural Computing and Applications*, pp. DOI: 10.1007/s00521-019-04410-7, July 2019.

PHURIE: Hurricane Intensity Estimation from Infrared Satellite Imagery using Machine Learning, Amina Asif, Muhammad Dawood, Bismillah Jan, Javaid Khurshid, Mark DeMaria, and Fayyaz ul Amir Afsar Minhas, in *Neural Computing and Applications*, DOI: <http://dx.doi.org/10.1007/s00521-018-3874-6>, 2018

Data Science: Journal Recommendation System



PRISM: Journal Paper Recommendation System

PRISM helps you find journals that could be best suited for publishing your research papers. Please also consult the journal's Aims and Scope for further guidance. Ultimately, the Editor will decide on how well your article matches the journal.
Simply insert your title and abstract and select the appropriate field-of-research for the best results.

Hover-Net: Simultaneous segmentation and classification c ✓

Nuclear segmentation and classification within Haematoxylin & Eosin stained histology images is a fundamental prerequisite in the digital pathology work-flow. The development of automated methods for nuclear segmentation and classification enables the quantitative analysis of tens of thousands of nuclei within a whole-slide pathology image, opening up possibilities of further analysis of large-scale nuclear morphometry. However, automated nuclear segmentation and classification is faced with a major

Clear

Find Journal

Our Recommendations

- [medical image analysis](#)
- [ieee transactions on medical imaging](#)
- [medical image computing and computer assisted intervention – miccai 2018](#)
- [ieee transactions on geoscience and remote sensing](#)
- [ieee transactions on image processing](#)

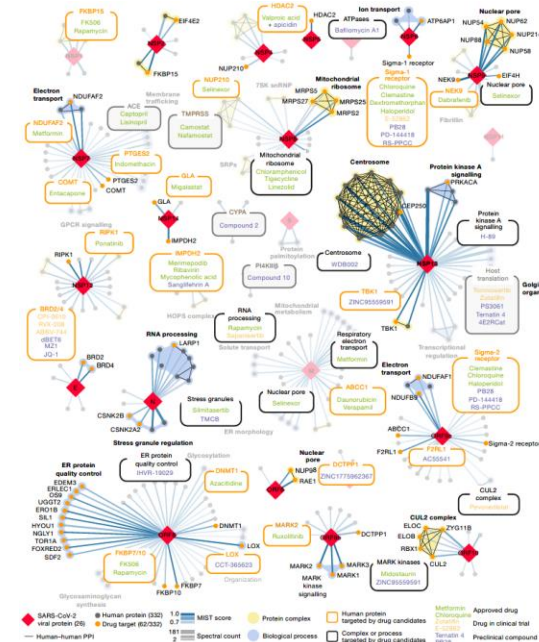
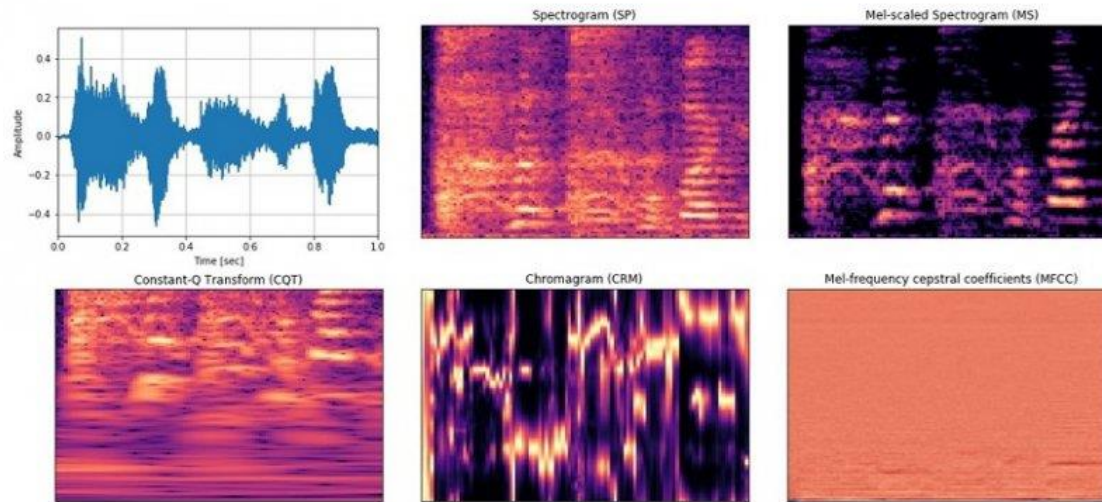
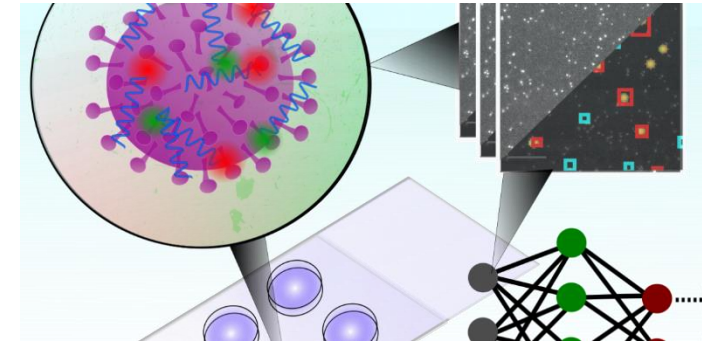
- Using classical NLP
- Using BERT



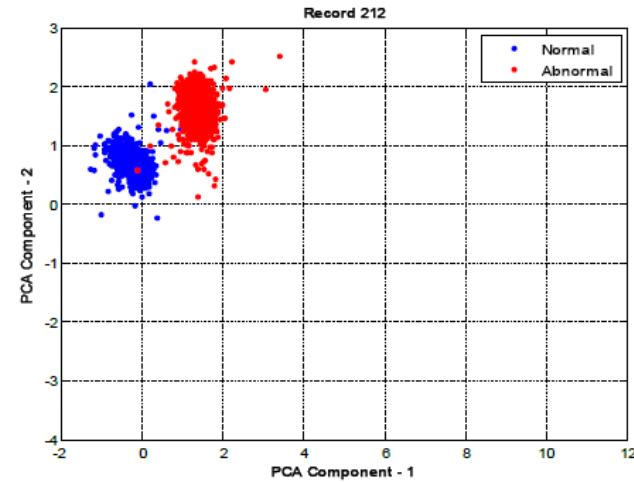
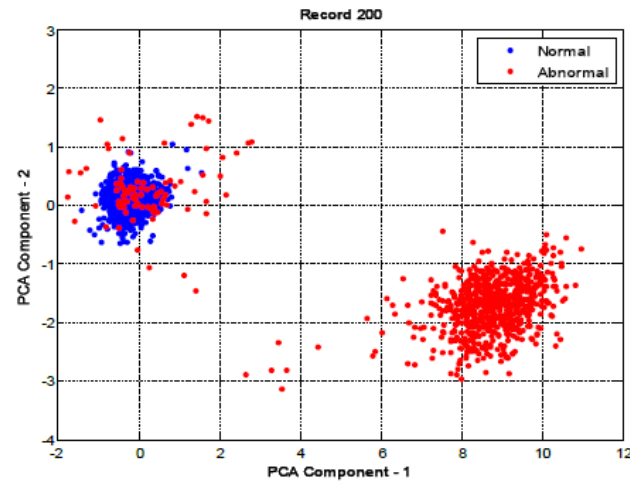
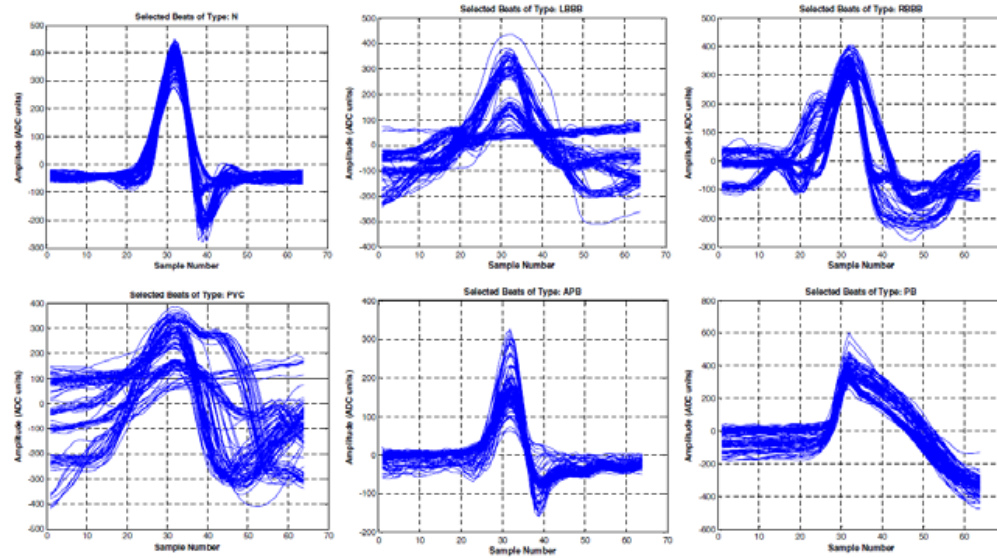
I can't be as confident about computer science as I can about biology. Biology easily has 500 years of exciting problems to work on. It's at that level.

— *Donald Knuth* —

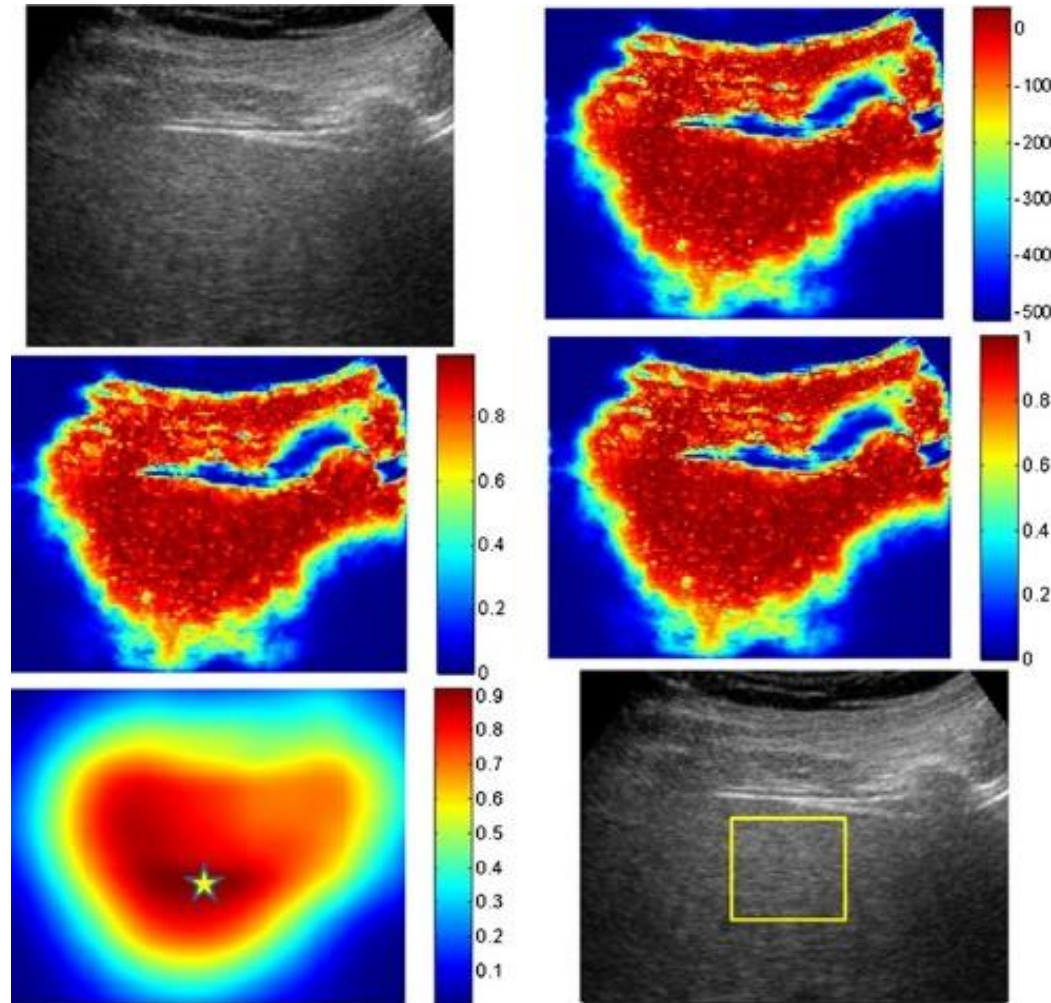
COVID19 Meets Machine Learning



ECG Classification

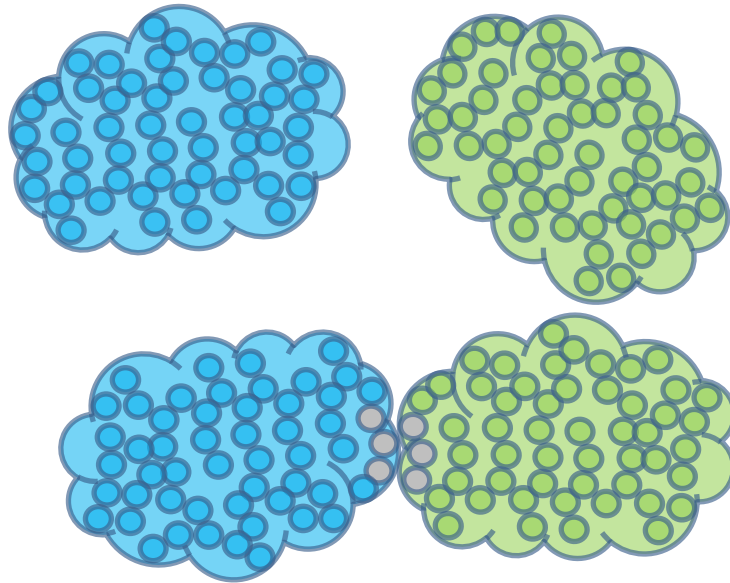


Medical Data Classification



Predicting Protein interactions, interfaces and affinity

- Input: Two protein structures or sequences
- Output: What residue pairs interact



PAIRpred: PARTner specific Interacting Residue PREDictor

What is PAIRpred?

We present a novel partner-specific protein-protein interaction site prediction method called PAIRpred. PAIRpred uses information from both proteins in a protein complex to predict pairs of interacting residues from the two proteins using both sequence and structure information. PAIRpred presents a more detailed model of protein binding, and offers state of the art accuracy in predicting binding sites at the protein level as well as inter-protein residue contacts at the complex level. PAIRpred is available as a cloud based webserver. You can use it to make predictions for proteins of your interest.

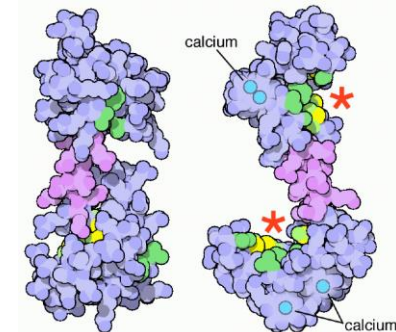
Apply PAIRpred to your proteins here

Input File Type:

File containing sequence or structure of ligand: No file chosen

File containing sequence or structure of receptor: No file chosen

Email:



[1] Fayyaz Minhas and Asa Ben-Hur, Multiple instance learning of Calmodulin binding sites. **Bioinformatics** 28, i416, 2012.

[2] Wajid Abbasi, Amina Asif, Saiqa Andleeb, and Fayyaz Minhas, CaMELS: In silico Prediction of Calmodulin Binding Proteins and their Binding Sites, in **Proteins: Structure, Function and Bioinformatics**, 2017.

[3] Issues In Performance Evaluation for Host-Pathogen Protein Interaction Prediction, Wajid A. Abbasi and Fayyaz Minhas, in **Journal of Bioinformatics and Computational Biology**, vol. 14, no. 3, 1650011, January 2016.

[4] Training host-pathogen protein-protein interaction predictors , Abdul Hannan Basit, Wajid Arshad Abbasi, Amina Asif, Sadaf Gull and **Fayyaz Minhas**, in **Journal of Bioinformatics and Computational Biology**, Vol. 16, No. 04, 1850014 (2018).

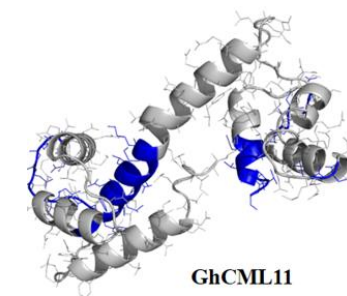
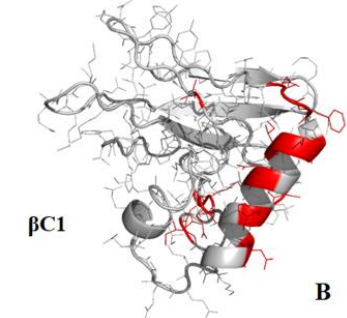
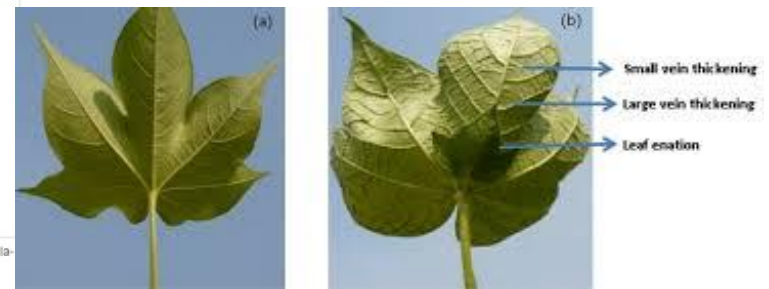
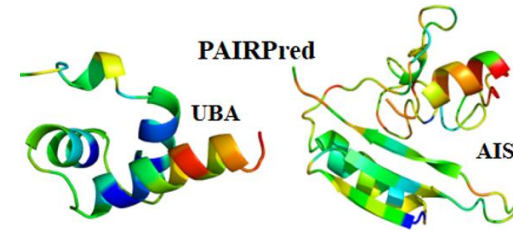
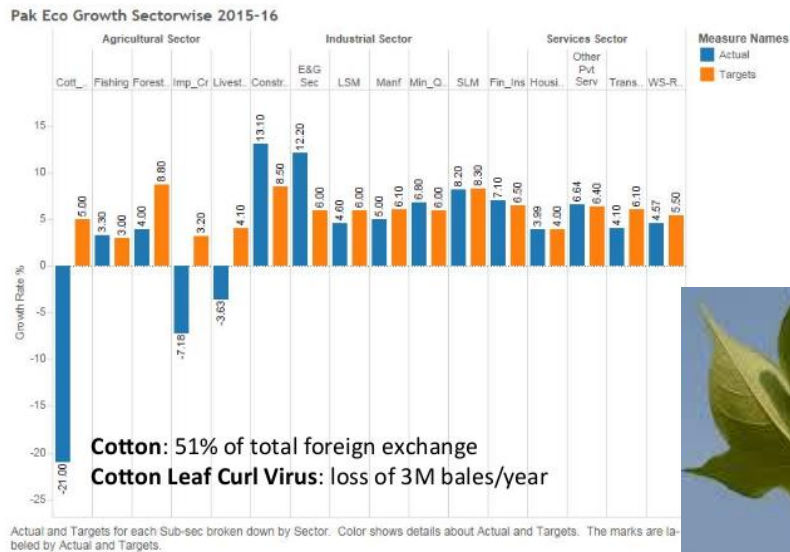
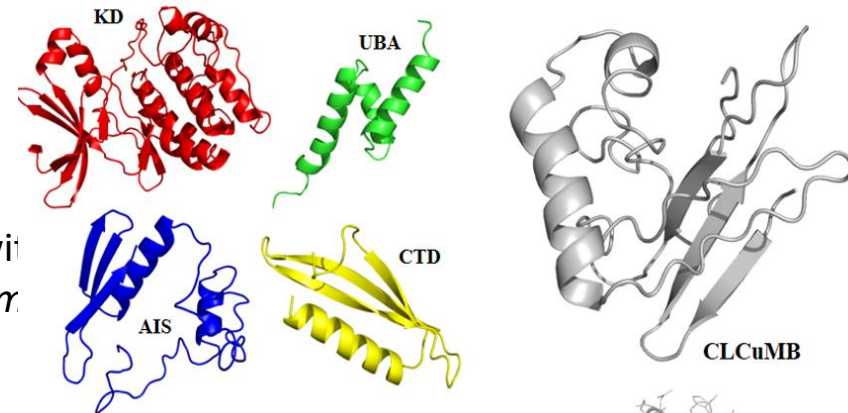
[5] Issues In Performance Evaluation for Host-Pathogen Protein Interaction Prediction, Wajid A. Abbasi and Fayyaz Minhas, in **Journal of Bioinformatics and Computational Biology**, vol. 14, no. 3, 1650011, January 2016.

[6] Training host-pathogen protein-protein interaction predictors , Abdul Hannan Basit, Wajid Arshad Abbasi, Amina Asif, Sadaf Gull and **Fayyaz Minhas**, in **Journal of Bioinformatics and Computational Biology**, Vol. 16, No. 04, 1850014 (2018).

Experimental Validation: SnRK1- β C1

- In Silico Prediction and Validations of Domains Involved in *Gossypium hirsutum* SnRK1 Protein Interaction with Cotton leaf curl Multan betasatellite encoded β C1
- β C1, pathogenicity determinant encoded by Cotton leaf curl Multan betasatellite interacts with calmodulin-like protein 11 (CML11) in *Gossypium hirsutum*

GhSnRK1 506aa



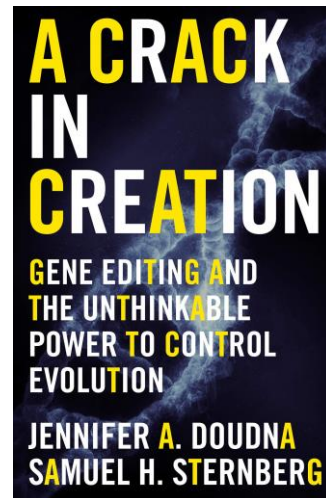
<http://www.dawn.com/news/1201323>

In Silico Prediction and Validations of Domains Involved in *Gossypium hirsutum* SnRK1 Protein Interaction with Cotton leaf curl Multan betasatellite encoded β C1, Kamal, Hira, Fayyaz ul Amir Afsar Minhas, Hanu Pappu, Imran Amin et al., in *Frontiers in Plant Science* 10 (2019): 656.

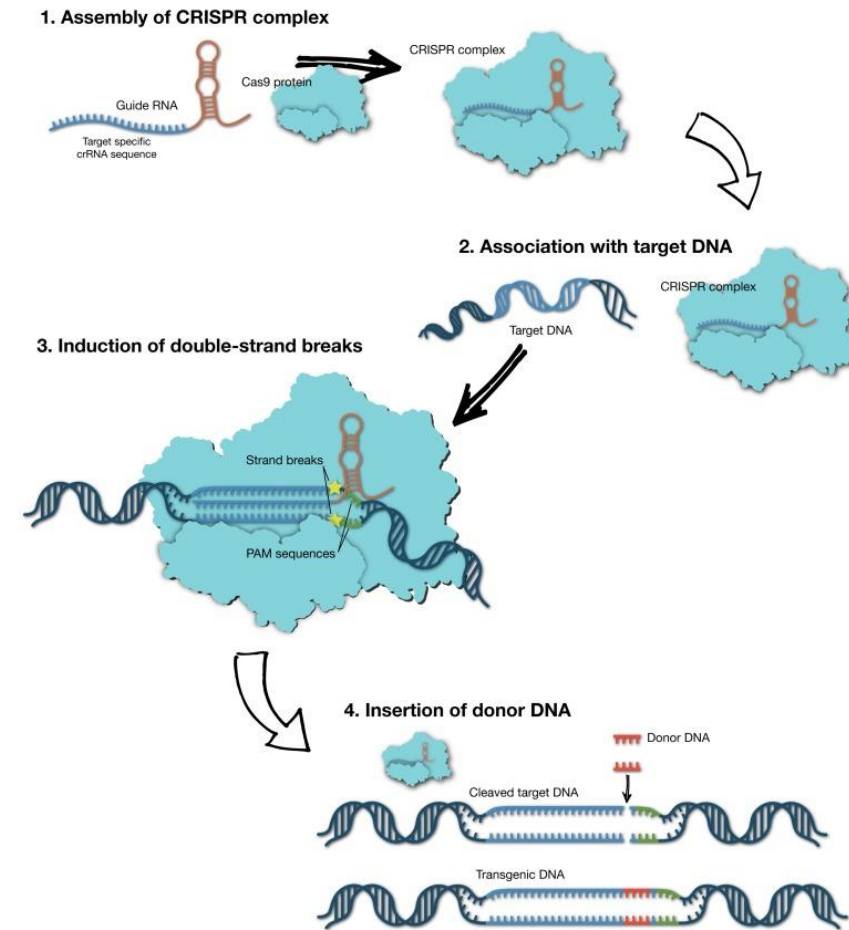
Bioinformatics and molecular analysis of *Gossypium hirsutum* calmodulin-like protein (CML11) interaction with begomovirus-transcription activator protein C2. Hira Kamal, Fayyaz Minhas, et al., in *PLoS One* (In press).

Predicting anti-CRISPR proteins

- 20 proteins for training
- Identified 3 new anti-CRISPR proteins using a ranking ML model



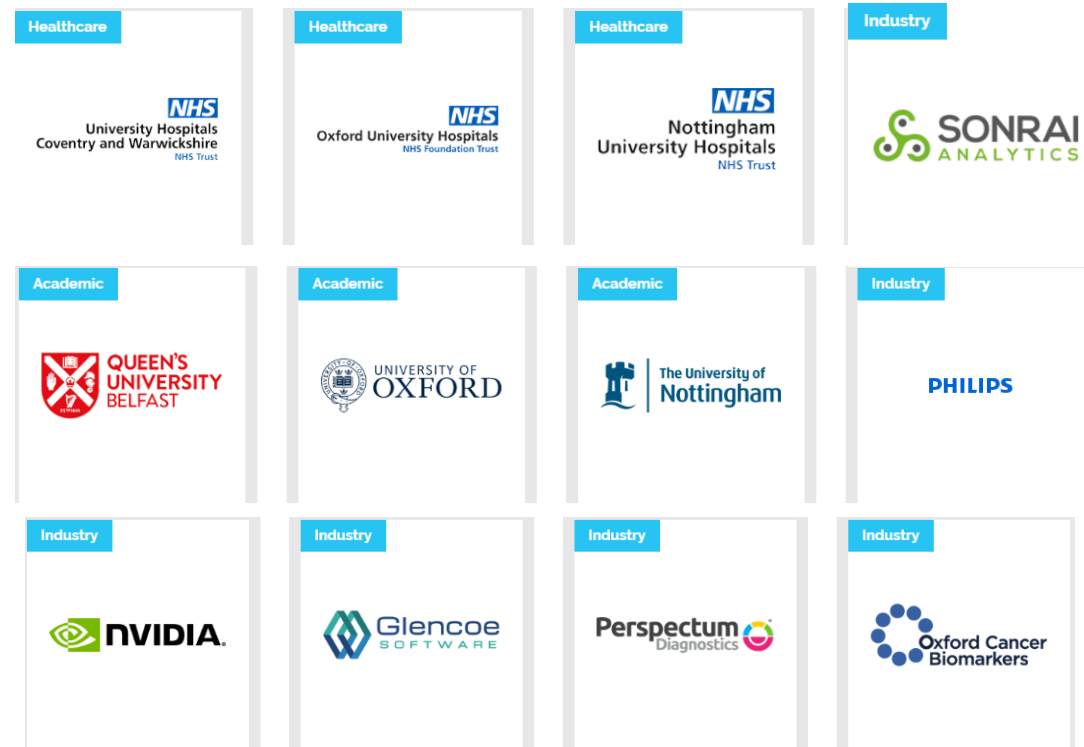
Jennifer Doudna



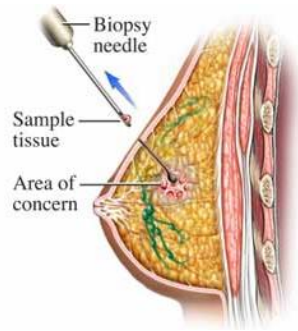
Eitzinger, Simon, Amina Asif, Kyle E. Watters, Anthony T. Iavarone, Gavin J. Knott, Jennifer A. Doudna, and Fayyaz ul Amir Afsar Minhas. "Machine Learning Predicts New Anti-CRISPR Proteins." *Nucleic Acids Research* 48, no. 9 (May 21, 2020): 4698–4708. <https://doi.org/10.1093/nar/gkaa219>.

Current Focus: PATHLake

- PATHology data Lake, Analytics, Knowledge and Education
- UK Research and Innovation
- £15.7 million
- Objective:
 - Improve speed and accuracy of cancer diagnosis



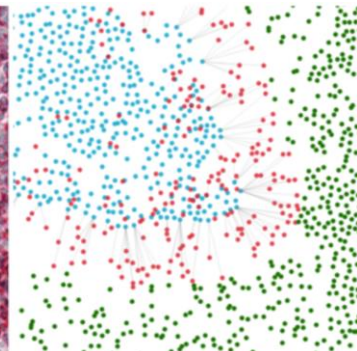
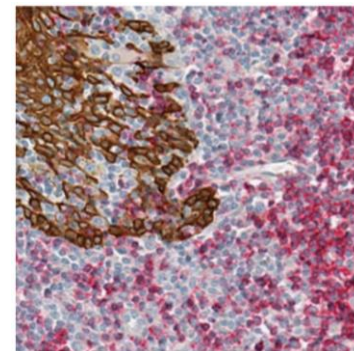
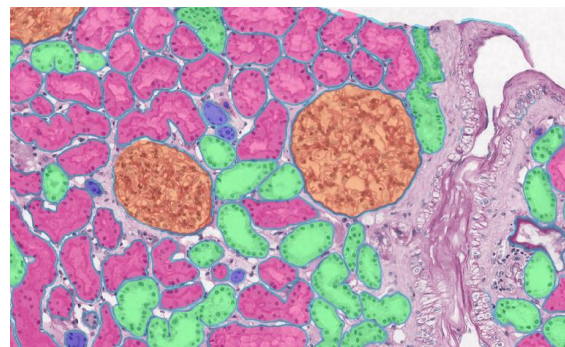
The Revolution in Pathology



Conventional Microscope Pathology



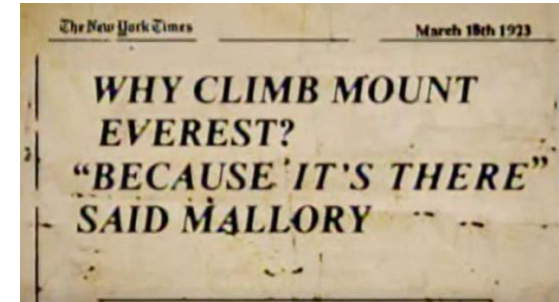
Digital Pathology



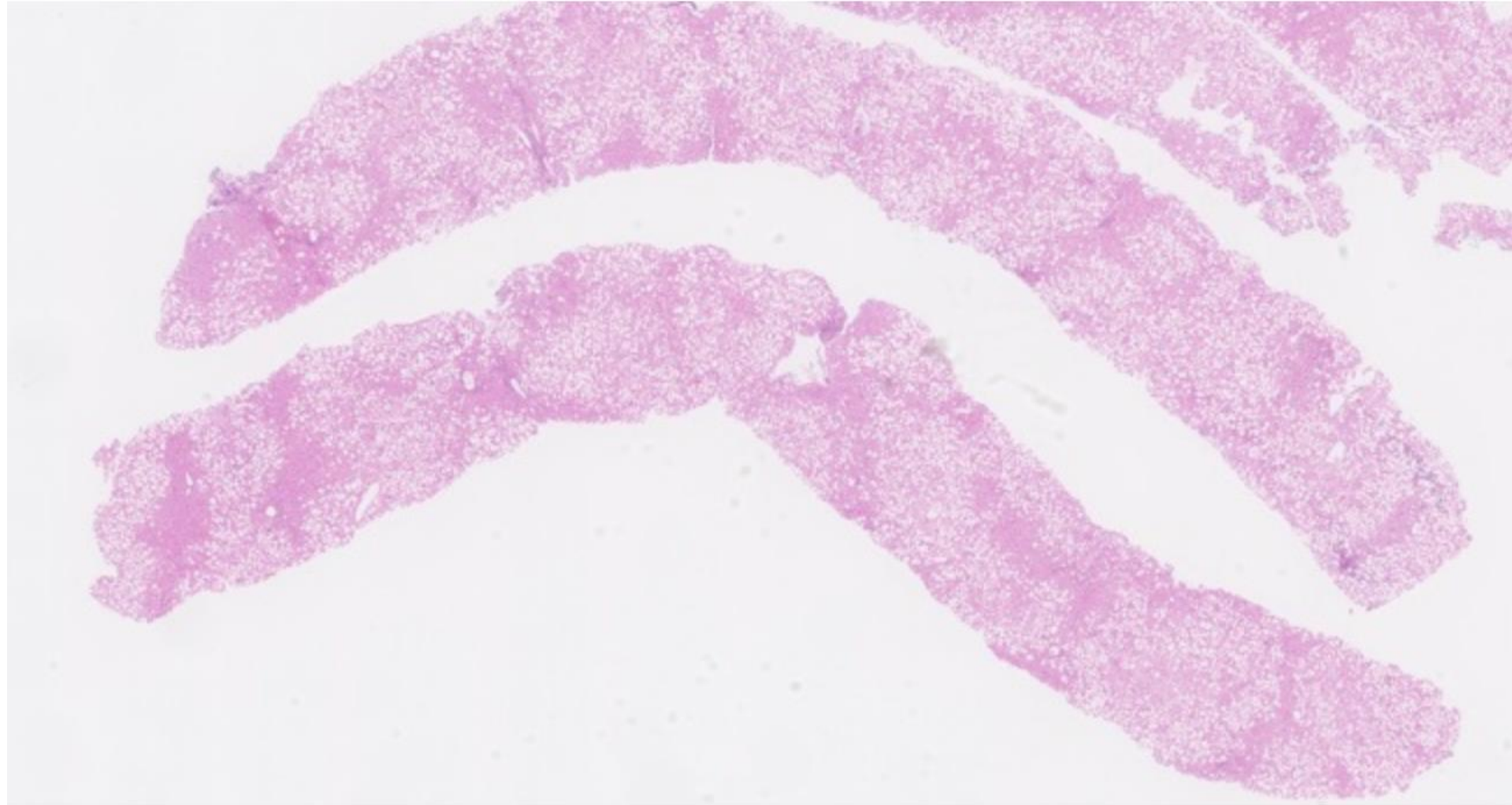
Computational Pathology

Why?

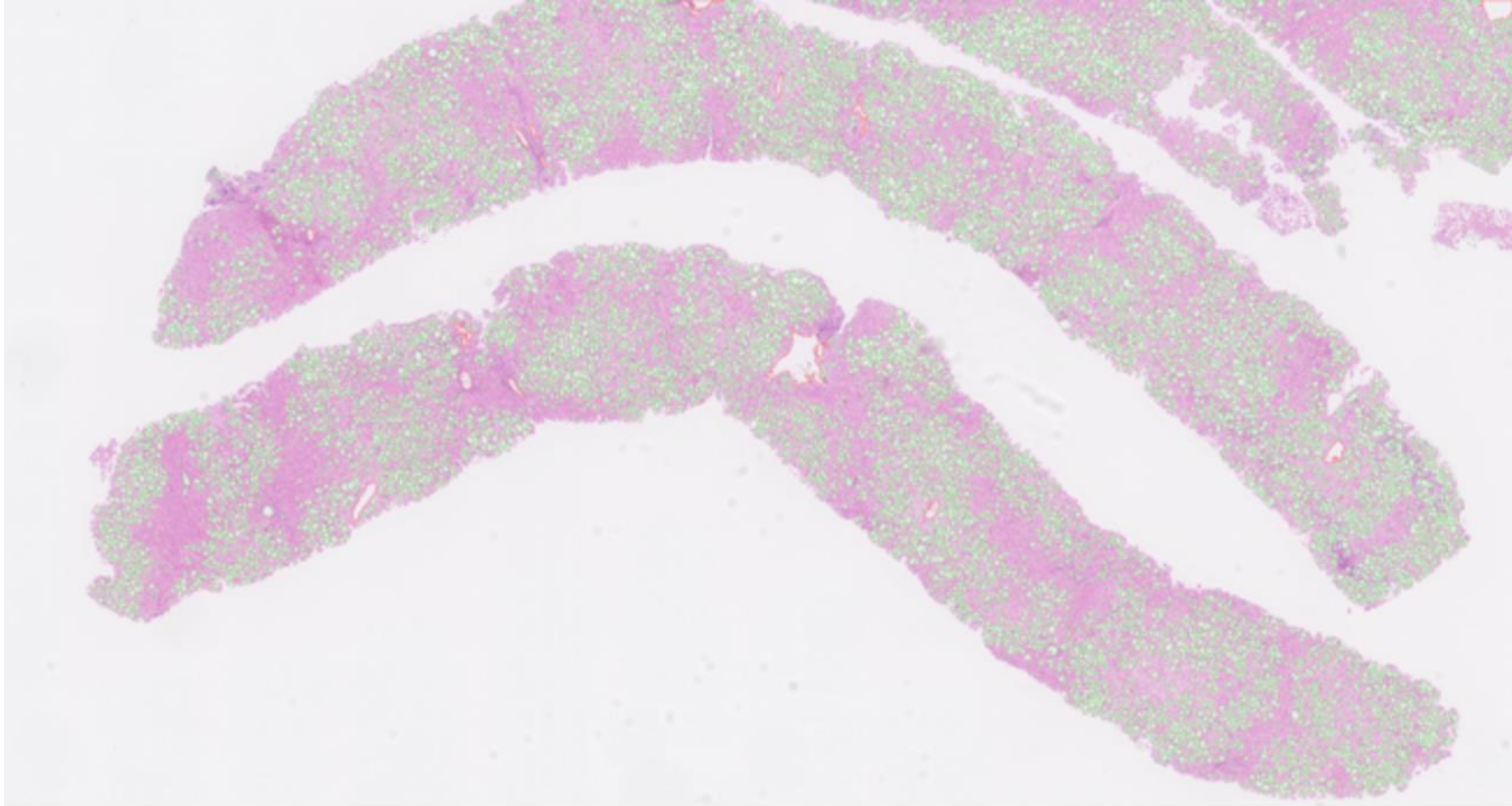
- Shortage of Pathologists
- Quantification is difficult
- Subjectivity
- Inter-observer variability



How much fat?

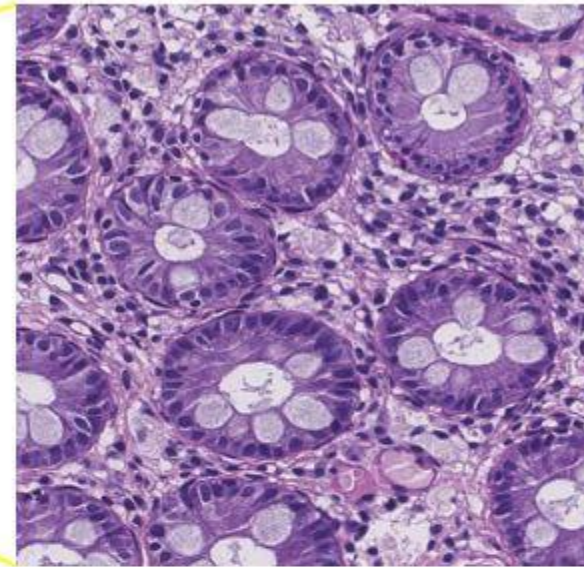
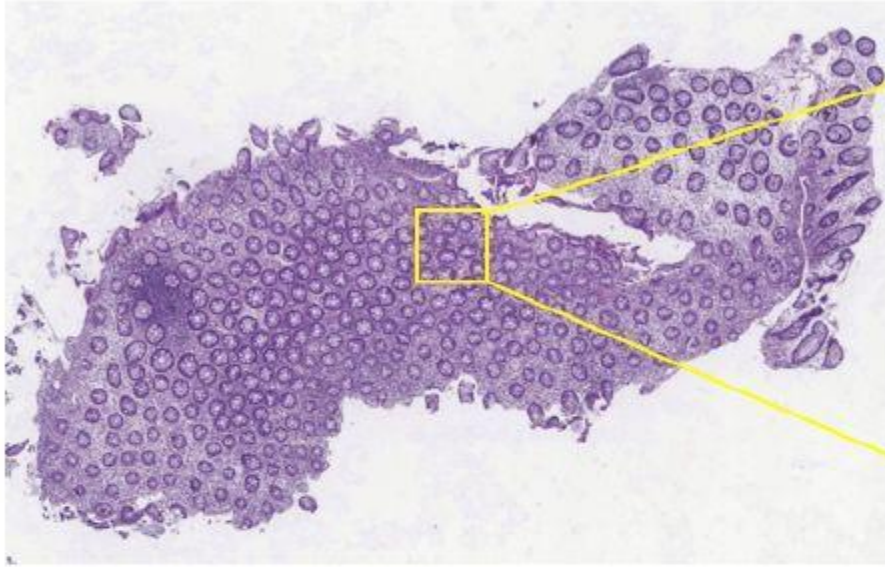


45.9%

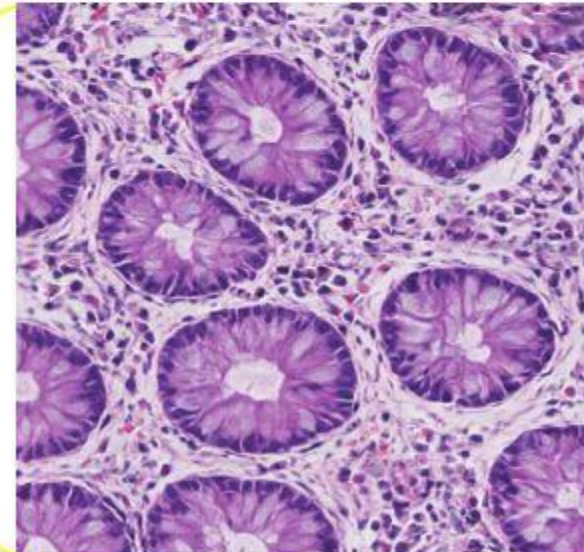
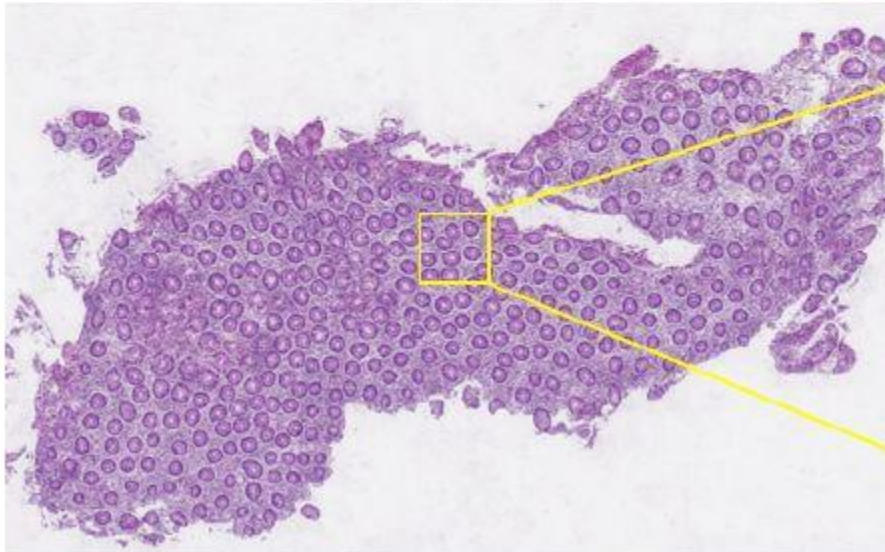


R. Glodin, "The Clinical Challenge of Image Analysis and AI", CRUK Early Detection Sandpit, Royal College of Pathologists, London, UK, 20 Nov. 2019

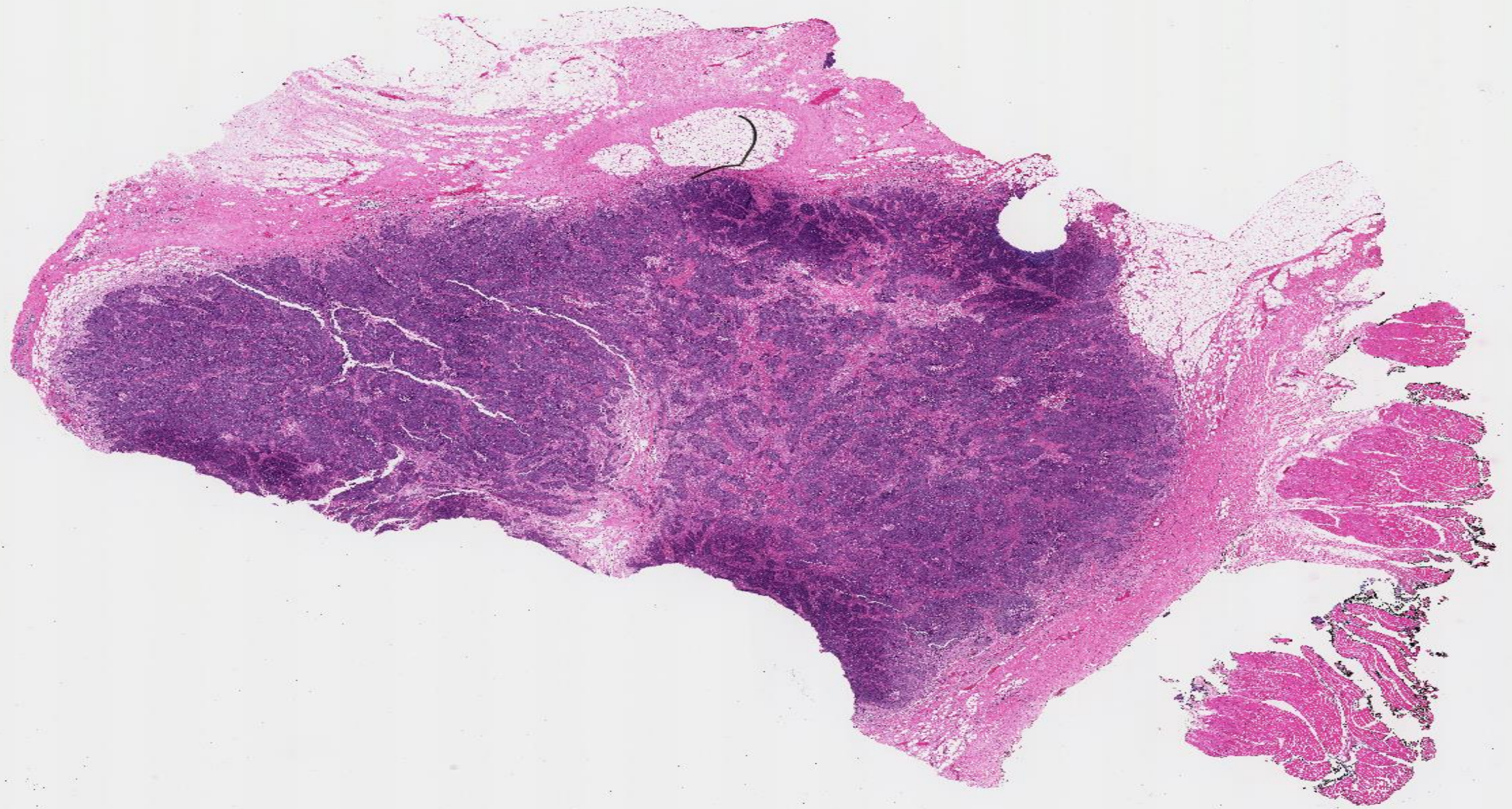
Real Tile

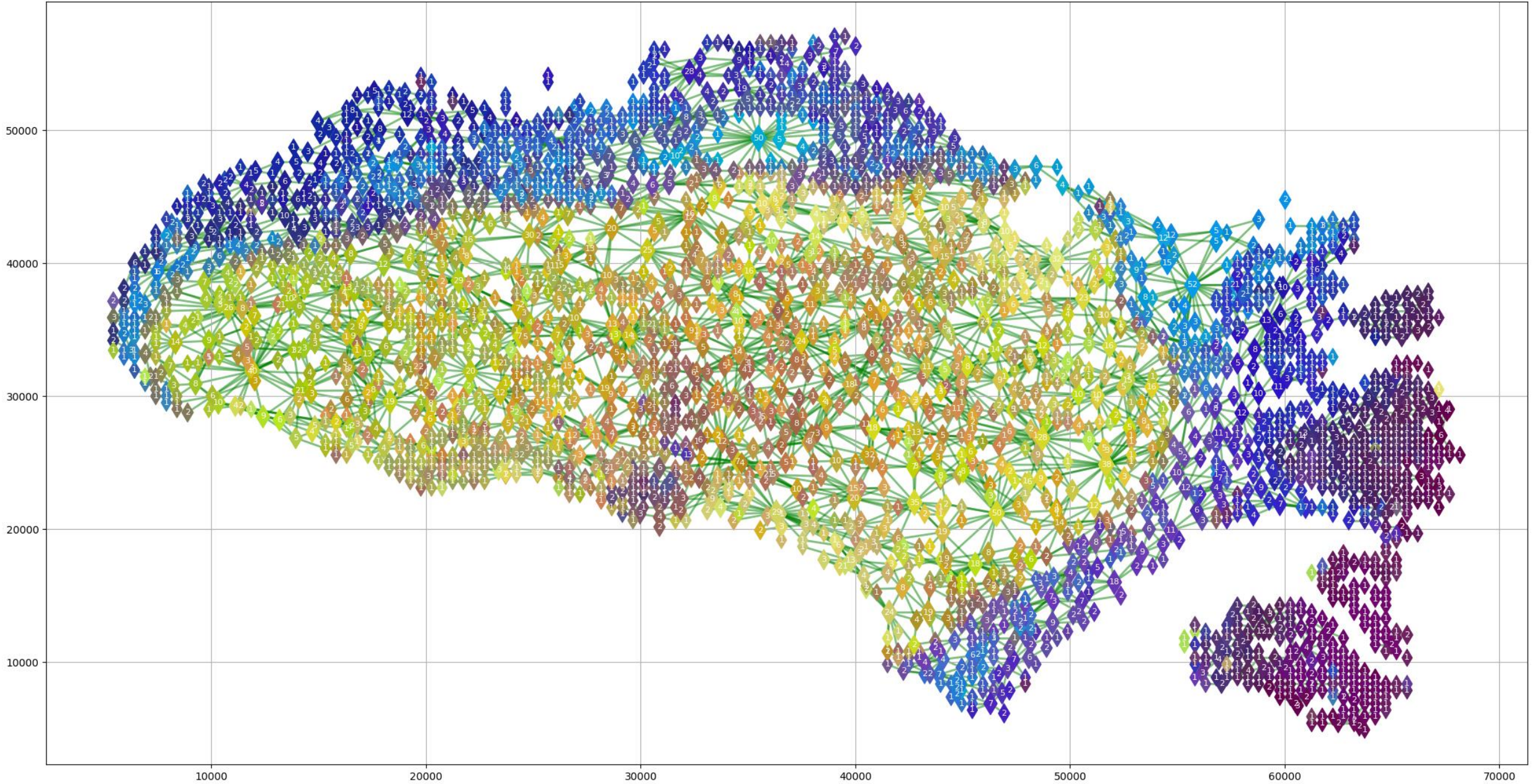


Generated Tile

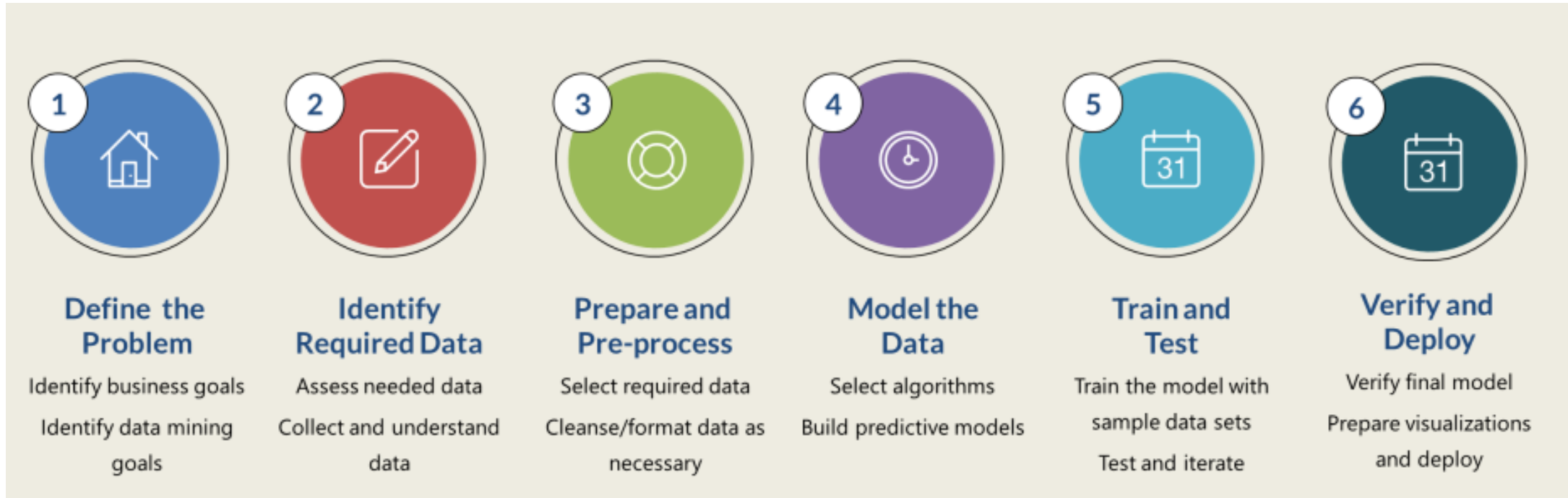


Deshpande, Srijay, Fayyaz Minhas, Simon Graham, and Nasir Rajpoot. "SAFRON: Stitching Across the Frontier for Generating Colorectal Cancer Histology Images." *ArXiv:2008.04526 [Cs, Eess]*, August 11, 2020. <http://arxiv.org/abs/2008.04526>.



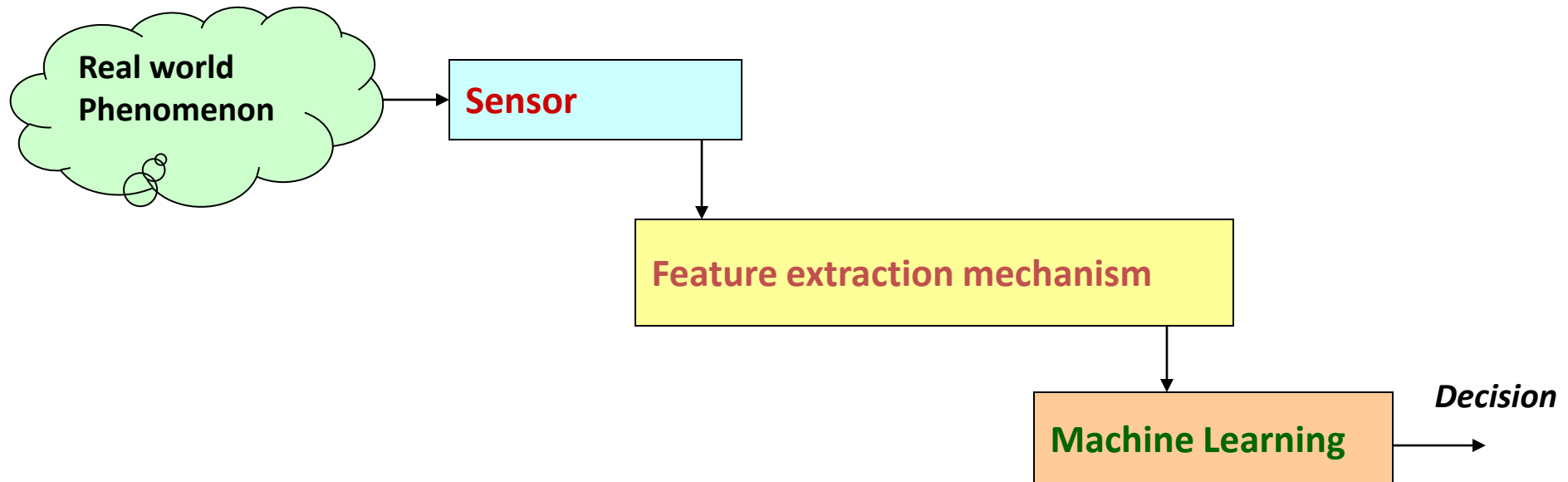


Steps in the development of a data science model



Constructs of a Data Mining System for Prediction

- Identify the objective
 - Identify the unit of classification (example)
 - Image block, protein sequence,



Learning from Data

- Example Case
 - Pathologists vs. Computer Scientists
 - Hypothetical!
 - Classify a person in their "native" environment

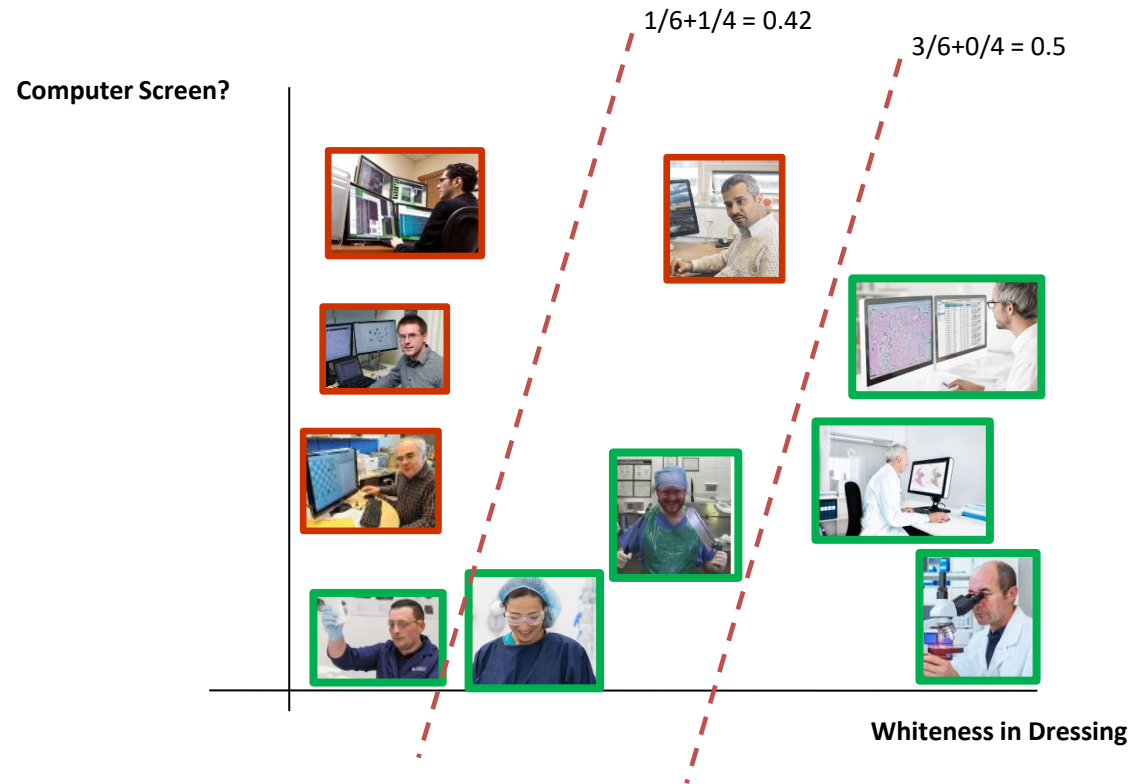
Constructs

- Sensor(s)
 - Camera

- Feature Extraction
 - White coats or lap aprons?
 - Computer Screen or microscopes?
 - Income?

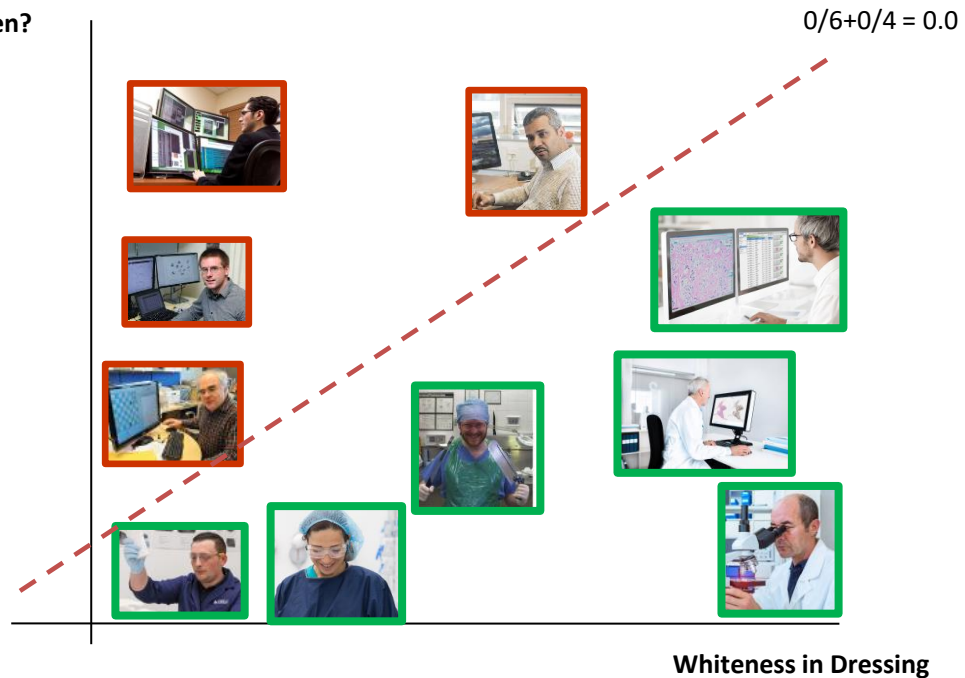
- Machine Learning

Feature Space



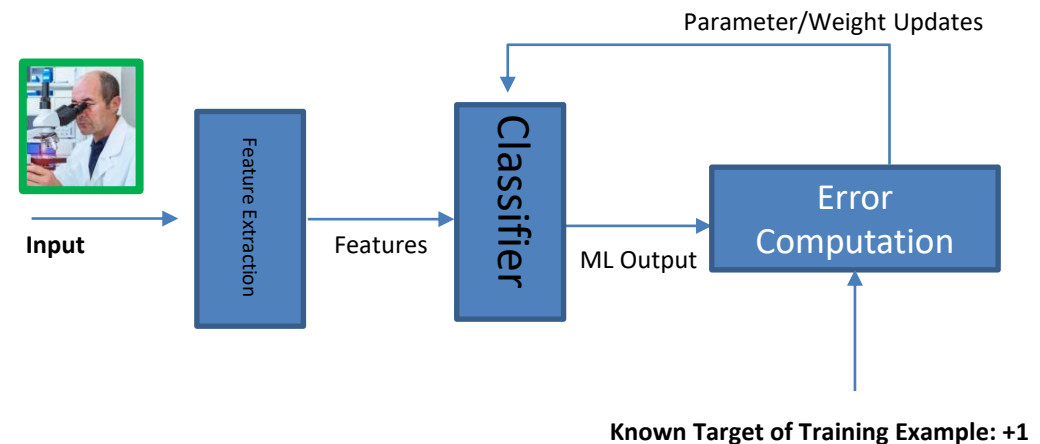
Feature Space Classification

Computer Screen?



So far?

- **Representation**
 - Represent examples in a feature space
 - Define a classification function
 - Line: $f(\mathbf{x};\mathbf{w}) = w_1x^{(1)}+w_2x^{(2)}+b = 0$
- **Evaluation**
 - Define an error function
 - Misclassifications
- **Optimize**
 - Reduce error
- **Real Test (Generalization)**
 - How does it perform on unseen data?



End of Lecture-2

We want to make a machine that will be proud of us.

- Danny Hillis