WARWICK
THE UNIVERSITY OF WARWICK

The Philosophy of ML Algorithms

Fayyaz Minhas

The Philosophical foundations of Machine Learning

WORLD VIEW *A personal take on events*

Train PhD students to be thinkers not just specialists

Many doctoral curricula aim to produce narrowly focused researchers rather than critical thinkers. That can and must change, says **Gundula Bosch**.

Under pressure to turn out productive lab members quickly, many PhD programmes in the biomedical sciences have shortened their courses, squeezing out opportunities for putting research into its wider context. Consequently, most PhD curricula are unlikely to nurture the big thinkers and creative problem-solvers that society needs.

That means students are taught every detail of a microbe's life cycle but little about the life scientific. They need to be taught to recognize how errors can occur. Trainees should evaluate case studies derived from flawed real research, or use interdisciplinary detective games to find logical fallacies in the literature. Above all, students must be shown the scientific process as it is — with its limitations and potential pitfalls as well as its fun side, such as serendipitous discoveries and hilarious blunders.

This is exactly the gap that I am trying to fill at Johns Hopkins University in Baltimore, Maryland, where a new graduate science programme is entering its second year. Microbiologist Arturo Casadevall and I began pushing for reform in early 2015, citing the need to put the philosophy back into the doctorate of philosophy: that is, the 'Ph' back into the PhD. We call our programme R3, which means that our students learn to apply rigour to their design and conduct of experiments; view their work through the lens of social responsibility; and to think critically, communicate better, and thus improve reproducibility. Although we are aware of many innovative individual courses developed along these lines, we are striving for more-comprehensive reform.

PUT THE PHILOSOPHY BACK INTO THE DOCTORATE OF PHILOSOPHY.

Our offerings are different from others at the graduate level. We have critical-thinking assignments in which students analyse errors in reasoning in a *New York Times* opinion piece about 'big sugar', and the ethical implications of the arguments made in a *New Yorker* piece by surgeon Atul Gawande entitled 'The Mistrust of Science'. Our courses on rigorous research, scientific integrity, logic, and mathematical and programming skills are integrated into students' laboratory and fieldwork. Those studying the influenza virus, for example, work with real-life patient data sets and wrestle with the challenges of applied statistics.

A new curriculum starts by winning allies. Both students and faculty members must see value in moving off the standard track. We used informal interviews and focus groups to identify areas in which students and faculty members saw gaps in their training. Recurring themes included the inability to apply theoretical knowledge in statistical tests in the laboratory, frequent mistakes in choosing an appropriate set of experimental controls, and significant difficulty in explaining work to non-experts.

Introducing our programme to colleagues in the Johns Hopkins life-sciences departments was even more sensitive. I was startled by the oft-expressed opinion that scientific productivity depended more on rote knowledge than on competence in critical thinking. Several principal investigators were uneasy about students committing more time to less conventional forms of education. The best way to gain their support was coffee: we repeatedly met lab heads to understand their concerns.

With the pilot so new, we could not provide data on students' performance, but we could address faculty members' scepticism. Some colleagues were apprehensive that students would take fewer courses in specialized content to make room for interdisciplinary courses on ethics, epistemology and quantitative skills. In particular, they worried that the R3 programme could lengthen the time required for students to complete their degree, leave them insufficiently knowledgeable in their subject areas and make them less productive in the lab.

We made the case that better critical thinking and fewer mandatory discipline-specific classes might actually position students to be more productive. We convinced several professors to try the new system and participate in structured evaluations on whether R3 courses contributed to students' performance.

So far, we have built 5 new courses from scratch and have enrolled 85 students from nearly a dozen departments and divisions. The courses cover the anatomy of errors and misconduct in scientific practice and teach students how to dissect the scientific literature. An interdisciplinary discussion series encourages broad and critical thinking about science. Our students learn to consider societal consequences of research advances, such as the ability to genetically alter sperm and eggs.

Discussions about the bigger-picture problems of the scientific enterprise get students to reflect on the limits of science, and where science's ability to do something competes with what scientists should do from a moral point of view. In addition, we have seminars and workshops on professional skills, particularly leadership skills through effective communication, teaching and mentoring.

It is still early days for assessment. So far, however, trainees have repeatedly emphasized that gaining a broader perspective has been helpful. In future, we will collect information about the impact that the R3 approach has on graduates' career choices and achievements.

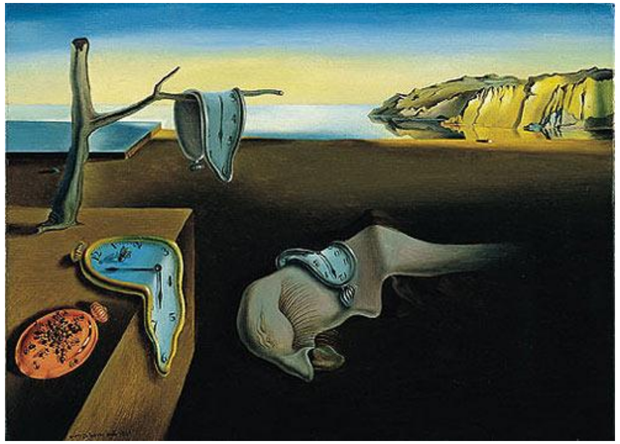
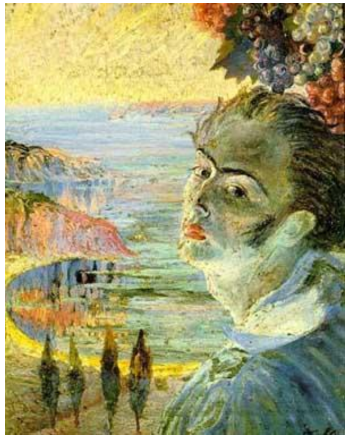
We believe that researchers who are educated more broadly will do science more thoughtfully, with the result that other scientists, and society at large, will be able to rely on this work for a better, more rational world. Science should strive to be self-improving, not just self-correcting. ■

Gundula Bosch *directs the R3 Graduate Science Initiative at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland.*
e-mail: gbosch@jhu.edu

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

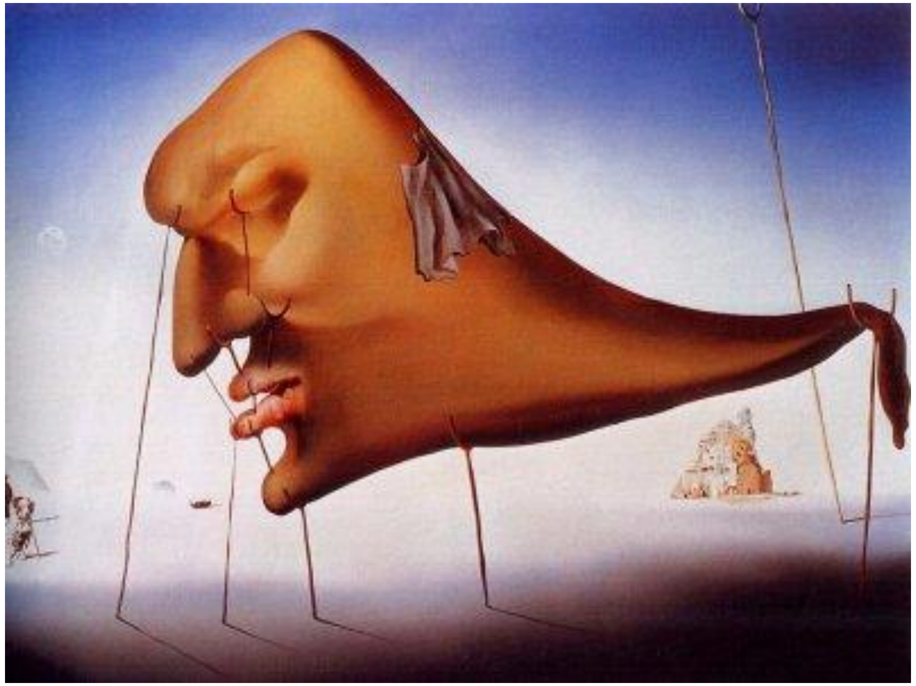
Paintings by two different painters



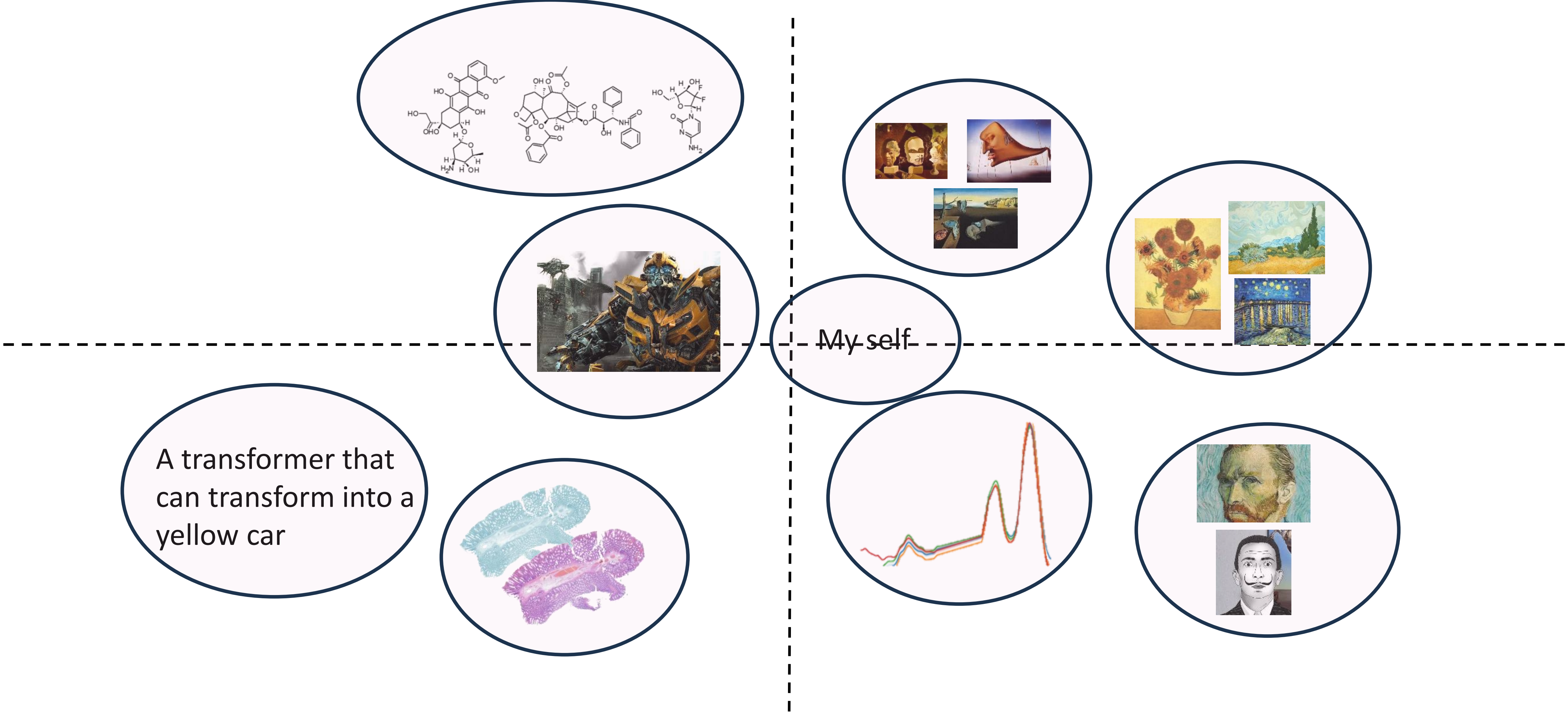

Who’s painting is this?

And this?



learning from data for generalization to unseen cases

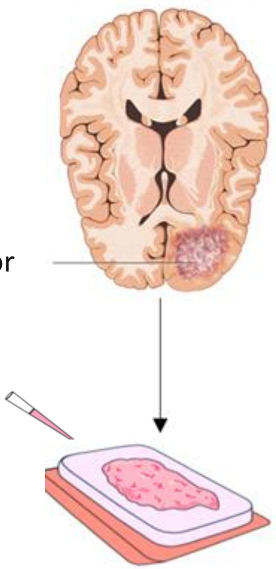
inductive inference



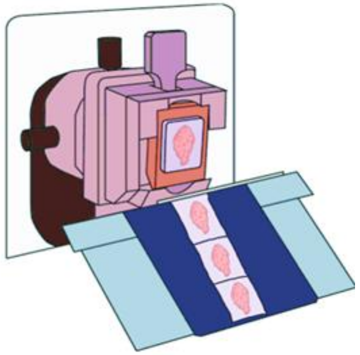
I. Entities have (explicit or implicit) representations

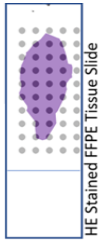
Application: Conventional Histopathology

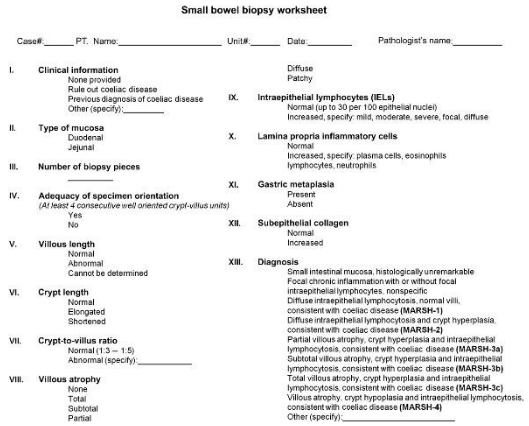
Tissue Acquisition → **Slide Preparation** → **Conventional Microscopy** → **Clinical Decision Making**



Block preparation
Addition of dyes







Small bowel biopsy worksheet

Case#:______ PT. Name:______________________ Unit#:______ Date:__________ Pathologist's name:__________

I. **Clinical information**
None provided
Rule out coeliac disease
Previous diagnosis of coeliac disease
Other (specify):_________

II. **Type of mucosa**
Duodenal
Jejunal

III. **Number of biopsy pieces**
____

IV. **Adequacy of specimen orientation**
(At least 4 consecutive well oriented crypt-villus units)
Yes
No

V. **Villous length**
Normal
Abnormal
Cannot be determined

VI. **Crypt length**
Normal
Elongated
Shortened

VII. **Crypt-to-villus ratio**
Normal (1:3 – 1:5)
Abnormal (specify):___________

VIII. **Villous atrophy**
None
Total
Subtotal
Partial
Diffuse
Patchy

IX. **Intraepithelial lymphocytes (IELs)**
Normal (up to 30 per 100 epithelial nuclei)
Increased, specify: mild, moderate, severe, focal, diffuse

X. **Lamina propria inflammatory cells**
Normal
Increased, specify: plasma cells, eosinophils
lymphocytes, neutrophils

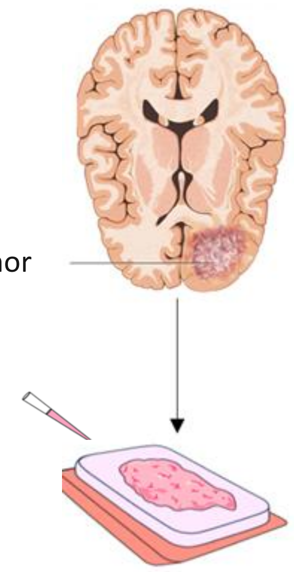
XI. **Gastric metaplasia**
Present
Absent

XII. **Subepithelial collagen**
Normal
Increased

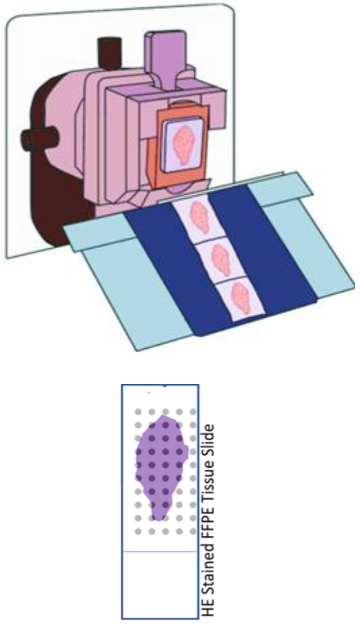
XIII. **Diagnosis**
Small intestinal mucosa, histologically unremarkable
Focal chronic inflammation with or without focal intraepithelial lymphocytes, nonspecific
Diffuse intraepithelial lymphocytosis, normal villi, consistent with coeliac disease **(MARSH-1)**
Diffuse intraepithelial lymphocytosis and crypt hyperplasia, consistent with coeliac disease **(MARSH-2)**
Partial villous atrophy, crypt hyperplasia and intraepithelial lymphocytosis, consistent with coeliac disease **(MARSH-3a)**
Subtotal villous atrophy, crypt hyperplasia and intraepithelial lymphocytosis, consistent with coeliac disease **(MARSH-3b)**
Total villous atrophy, crypt hyperplasia and intraepithelial lymphocytosis, consistent with coeliac disease **(MARSH-3c)**
Villous atrophy, crypt hypoplasia and intraepithelial lymphocytosis, consistent with coeliac disease **(MARSH-4)**
Other (specify):________________________

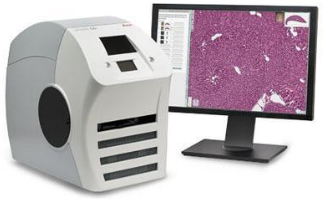
Application: Digital Pathology

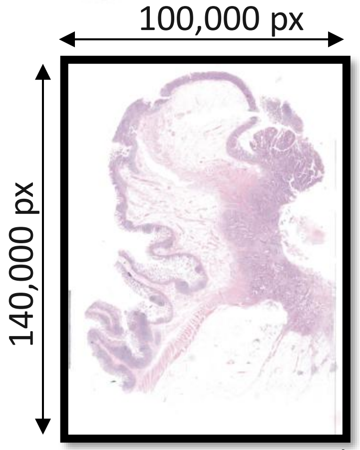
Tissue Acquisition → **Slide Preparation** → **Whole Slide Scanning** → **Clinical Decision Making**



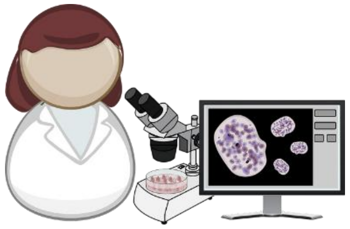
Block preparation
Addition of dyes







0.25 microns per pixel

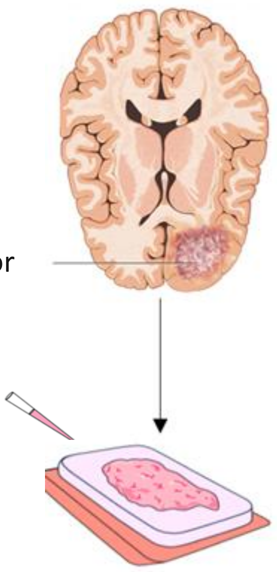


- Flexible working
- Good concordance with slide based clinical decision making based on equivalence studies
- However: *digitization of glass slides alone does not resolve the pressures of an increasing workload on a diminishing workforce of pathologists*

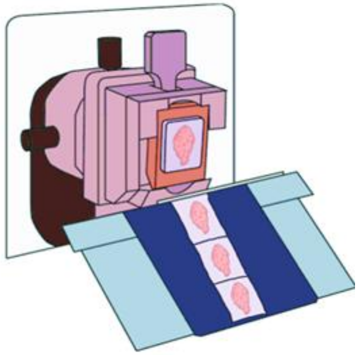
Example equivalence studies: Snead, David R. J., Yee-Wah Tsang, Aisha Meskiri, Peter K. Kimani, Richard Crossman, Nasir M. Rajpoot, Elaine Blessing, et al. "Validation of Digital Pathology Imaging for Primary Histopathological Diagnosis." *Histopathology* 68, no. 7 (June 2016): 1063–72. https://doi.org/10.1111/his.12879.
Hanna, Matthew G., Victor E. Reuter, Meera R. Hameed, Lee K. Tan, Sarah Chiang, Carlie Sigel, Travis Hollmann, et al. "Whole Slide Imaging Equivalency and Efficiency Study: Experience at a Large Academic Center." Modern Pathology 32, no. 7 (July 2019): 916–28. https://doi.org/10.1038/s41379-019-0205-0.

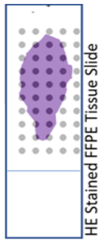
Application: Computational Pathology

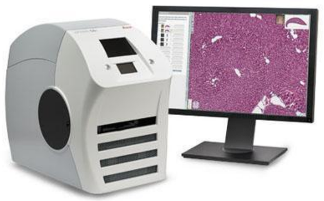
Tissue Acquisition → Slide Preparation → Whole Slide Scanning → **AI Assisted Clinical Decision Making**

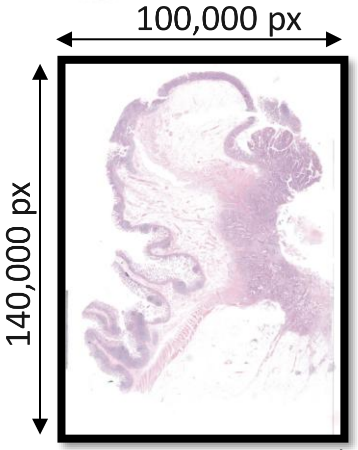


Block preparation
Addition of dyes

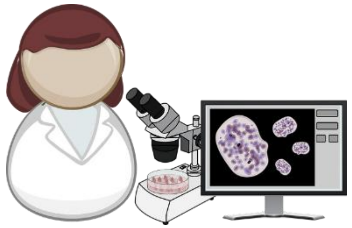




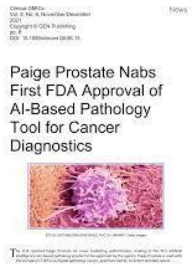




0.25 microns per pixel

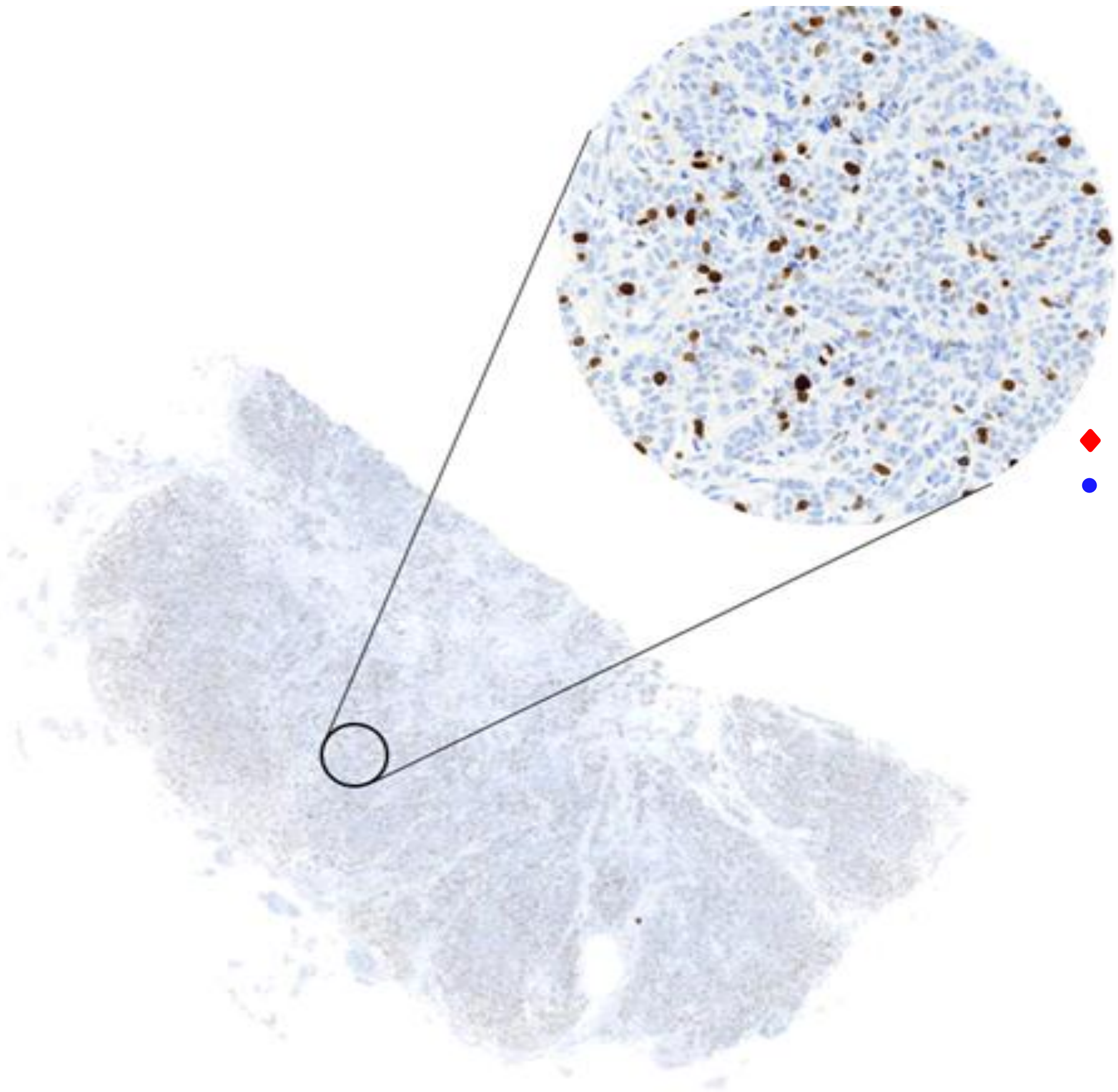


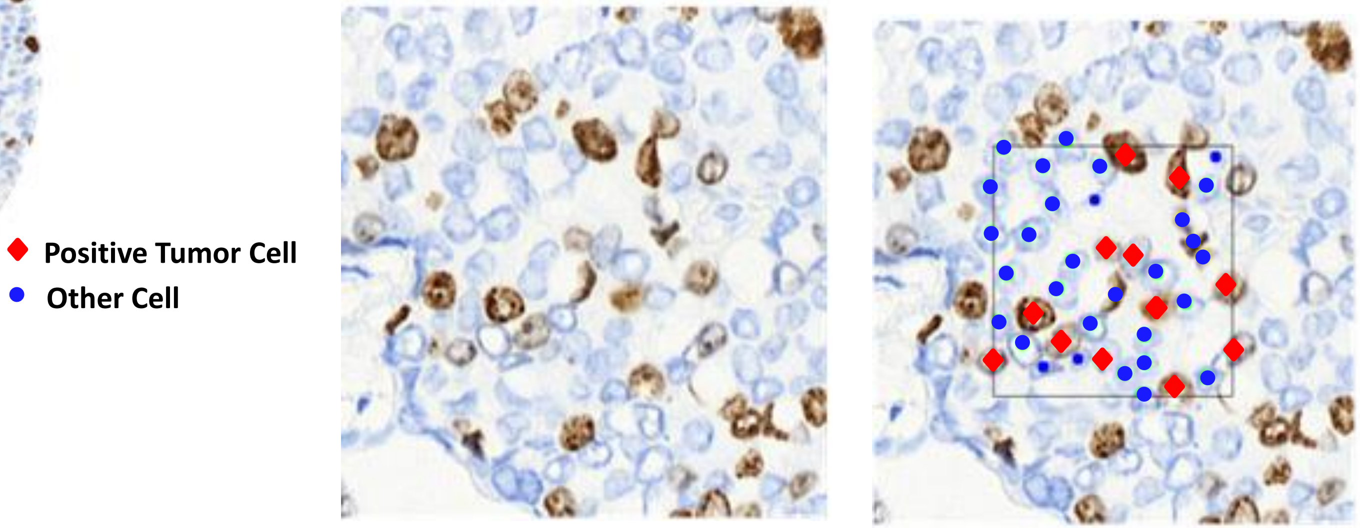
https://info.paige.ai/prostate

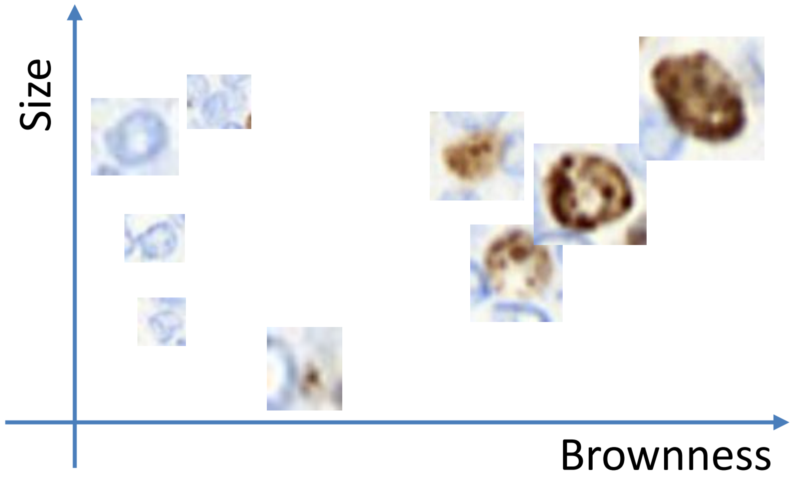


Example Independent validation study of PAIGE Prostate: Kanan, Christopher, Jillian Sue, Leo Grady, Thomas J. Fuchs, Sarat Chandarlapaty, Jorge S. Reis-Filho, Paulo G O Salles, Leonard Medeiros da Silva, Carlos Gil Ferreira, and Emilio Marcelo Pereira. "Independent Validation of Paige Prostate: Assessing Clinical Benefit of an Artificial Intelligence Tool within a Digital Diagnostic Pathology Laboratory Workflow." Journal of Clinical Oncology 38, no. 15_suppl (May 20, 2020): e14076–e14076. https://doi.org/10.1200/JCO.2020.38.15_suppl.e14076.

How does ML work?



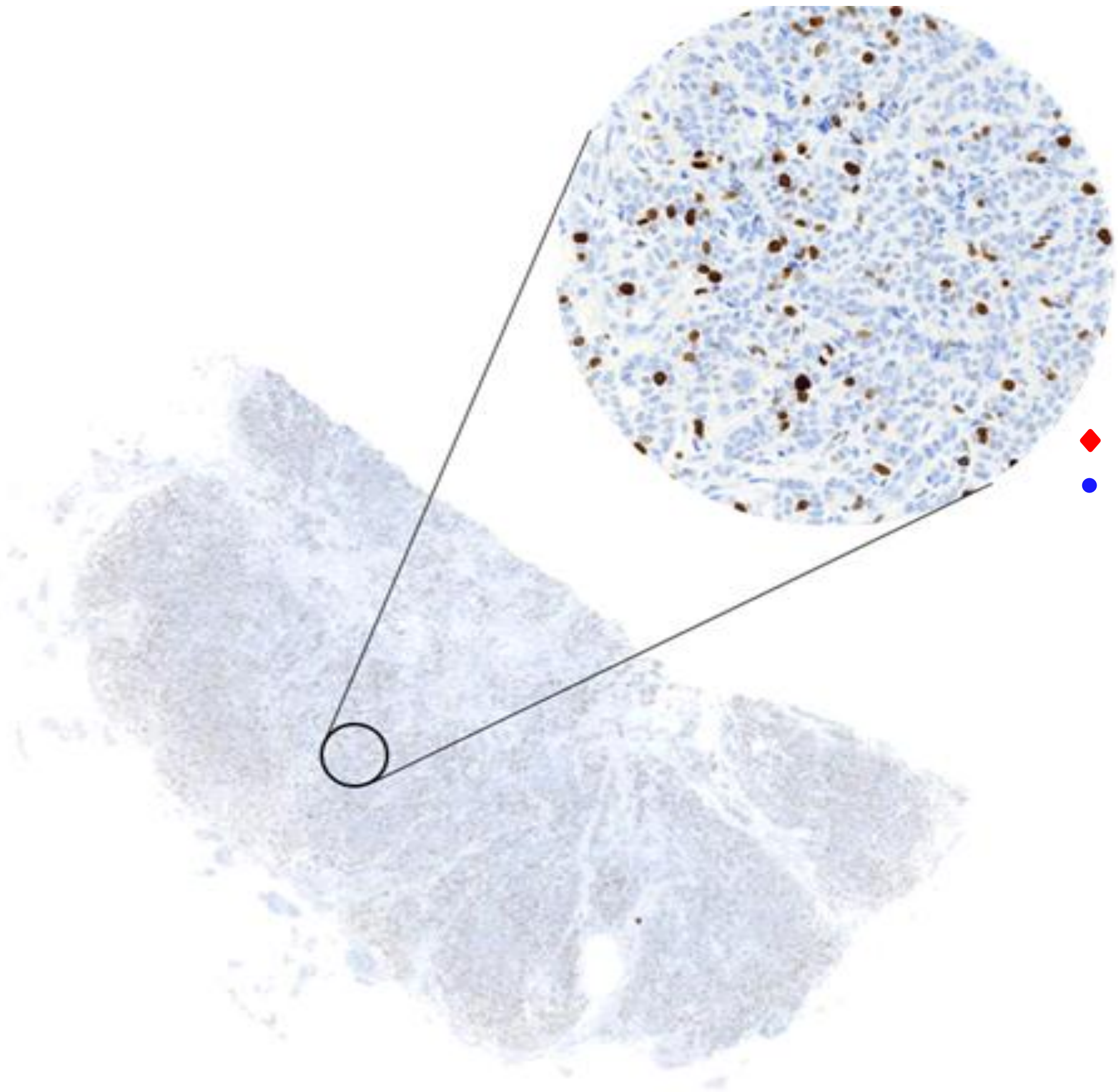


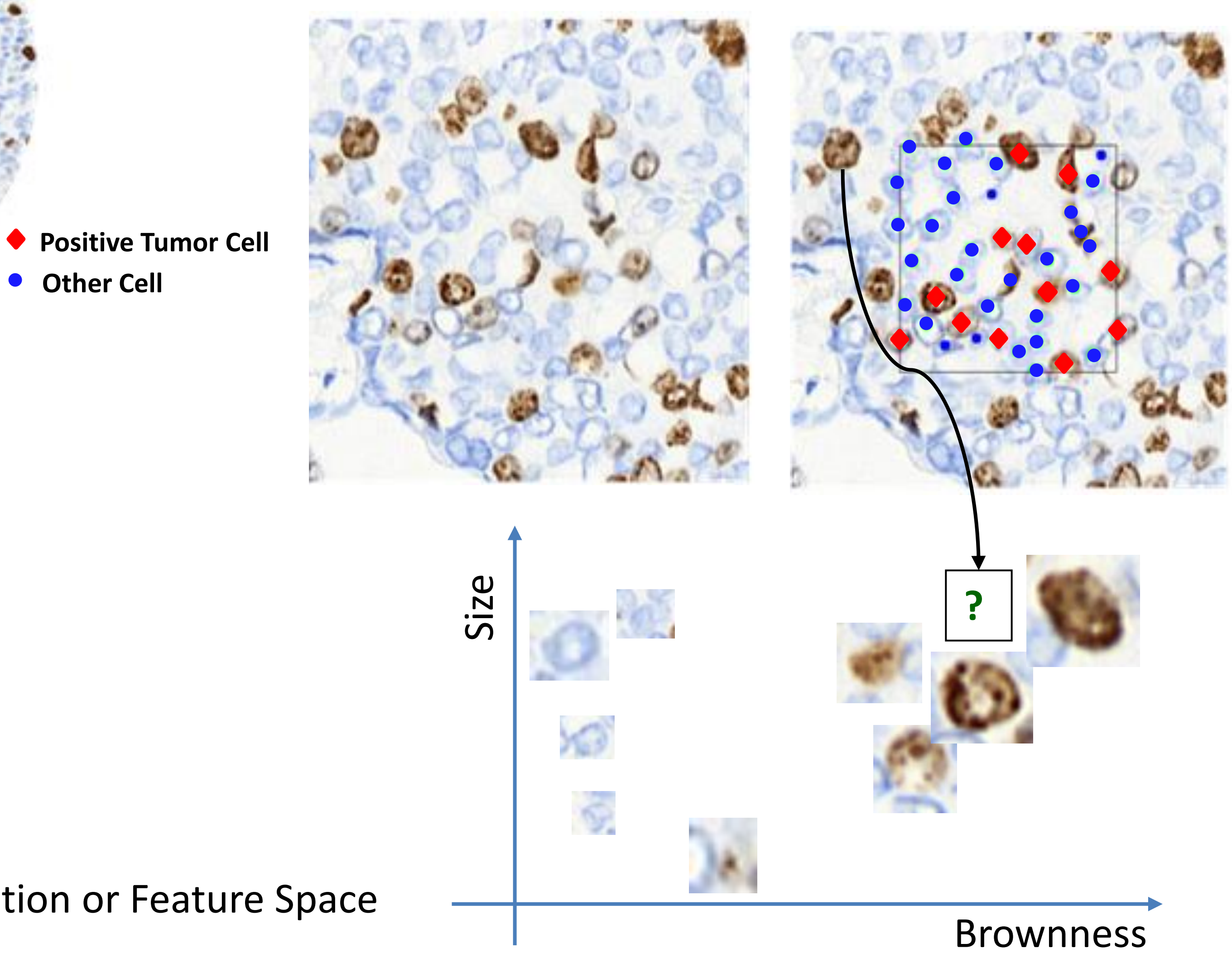


Representation or Feature Space

Wenqi Lu, Islam M Miligy, Fayyaz Minhas, Young Saeng Park, David R J Snead, Emad A Rakha, Clare Verrill, Nasir Rajpoot "Lessons from a Breast Cell Annotation Competition Series for School Pupils." Scientific Reports, 2022. https://ora.ox.ac.uk/objects/uuid:9e34d4e6-c677-4380-9403-759808b349aa.

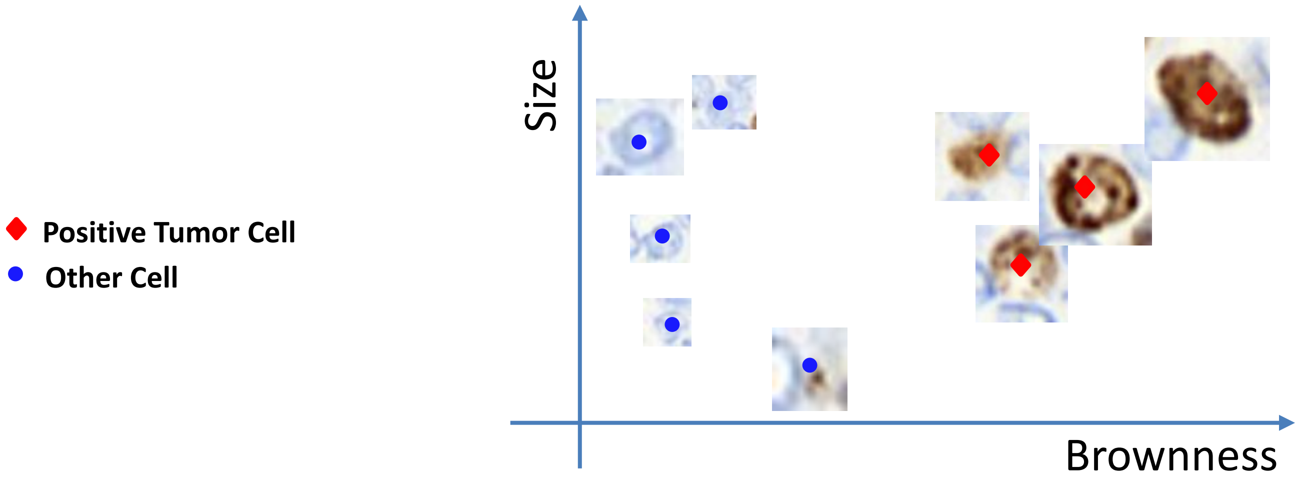
How does ML work?

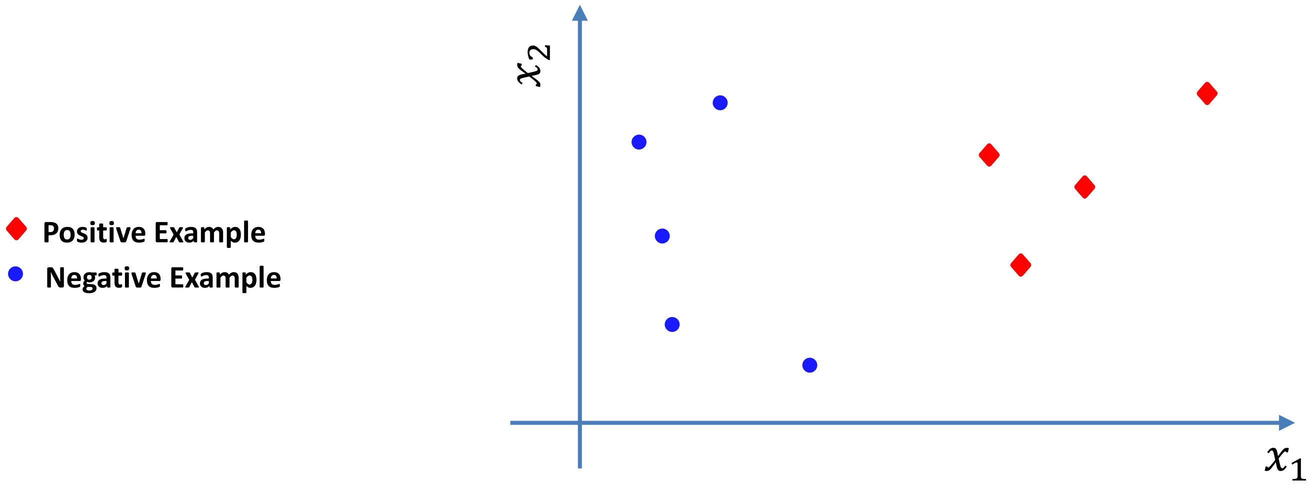




Representation or Feature Space

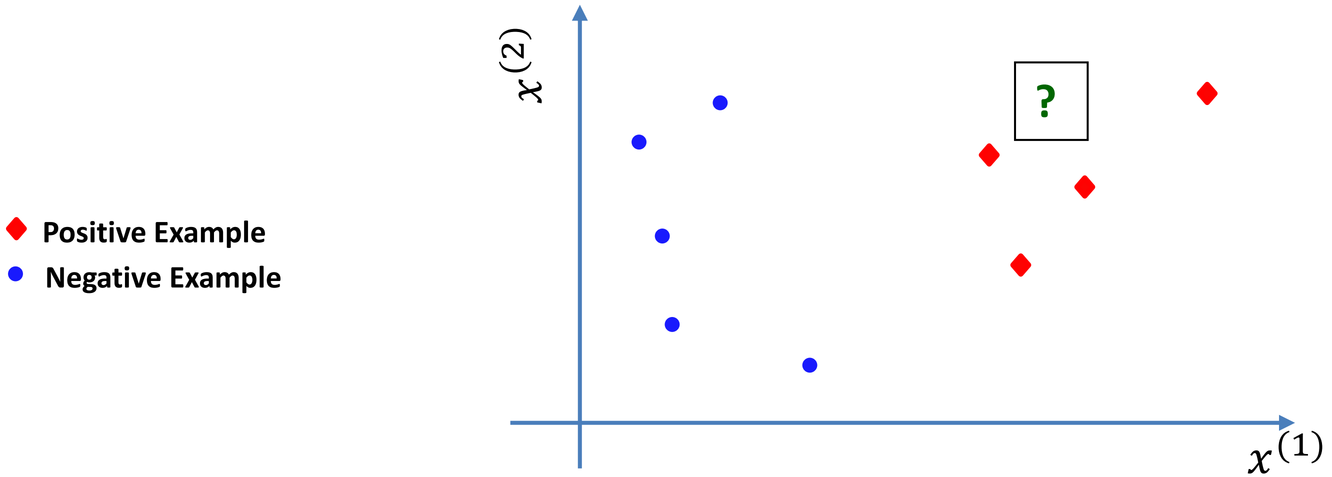
Wenqi Lu, Islam M Miligy, Fayyaz Minhas, Young Saeng Park, David R J Snead, Emad A Rakha, Clare Verrill, Nasir Rajpoot "Lessons from a Breast Cell Annotation Competition Series for School Pupils." Scientific Reports, 2022. https://ora.ox.ac.uk/objects/uuid:9e34d4e6-c677-4380-9403-759808b349aa.


Positive Tumor Cell
Other Cell
Size
Brownness


Positive Example
Negative Example
x_2
x_1

Classification Approaches: Nearest Neighbor and kNN

$$D(\boldsymbol{x_a}, \boldsymbol{x_b}) = \sqrt{\left(x_a^{(1)} - x_b^{(1)}\right)^2 + \left(x_a^{(2)} - x_b^{(2)}\right)^2}$$



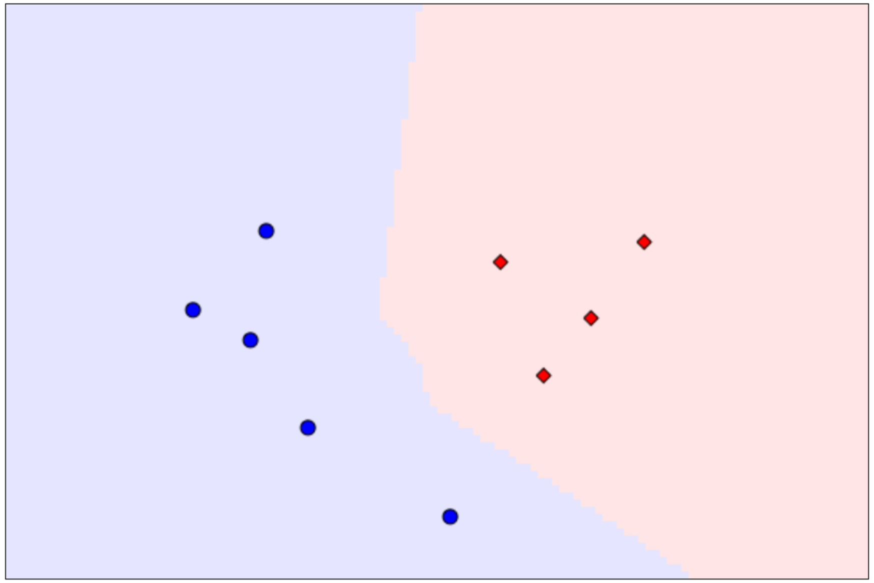
- Python Warm-up Lab Exercise
- https://github.com/foxtrotmike/CS909/blob/master/DM_1_kNN.ipynb

Example (k=1)-Nearest Neighbor Classification

K-Nearest Neighbors Demo

Instructions:

Select the type of point you want to place (Red or Blue) using the dropdown menu.
Click anywhere on the canvas to place the selected point.
Click on an existing point to select it. The `K` nearest neighbors of the selected point will be highlighted with lines.
Adjust the number of neighbors (K) using the input box to see how the neighbors and decision boundary change dynamically.
Select a distance metric from the dropdown to observe its effect on the decision boundaries.
Use the "Clear Points" button to reset the canvas and start over.



Point Type: Blue Num Neighbors (K): 1 Distance Metric: Euclidean Clear Points

(c) Fayyaz Minhas

Demo: https://foxtrotmike.github.io/CS909/knn.html

Discuss:

- Partitioning of the representation space
- Is k-NN a good rule?
- Can we separate points with a line?

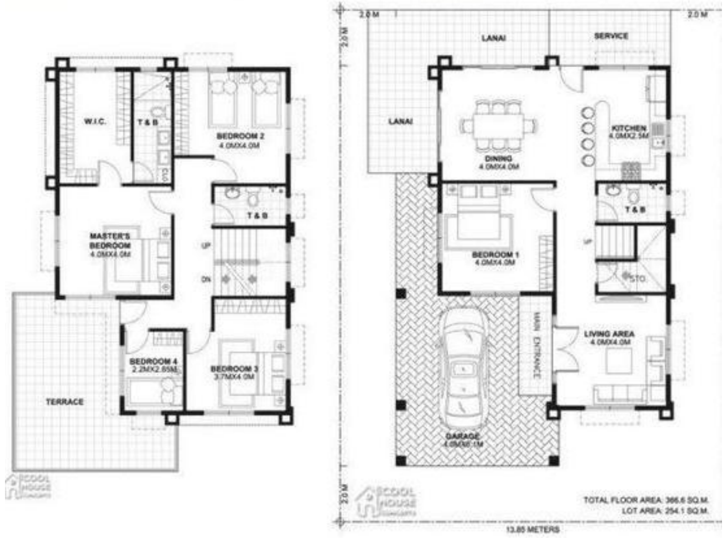
"Bank" in which statement is more semantically related to the picture?

- **A:** As he walked by the **bank**, he saw some tillers
- **B:** As he walked by the **bank**, he saw some tellers

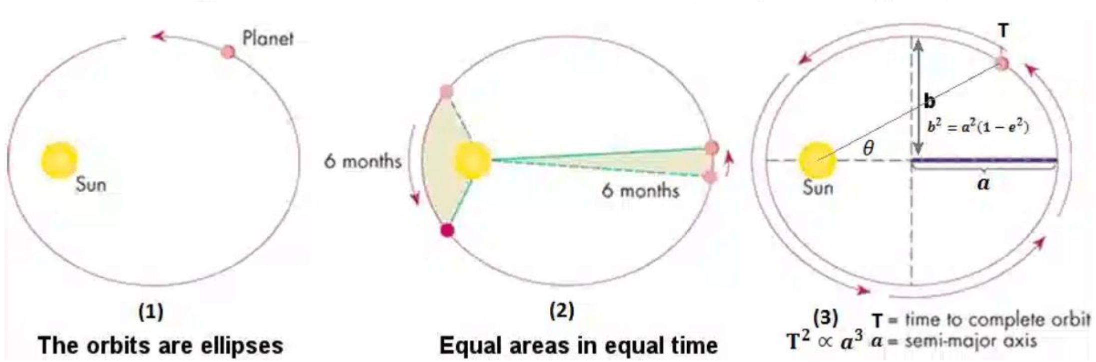
As he walked by the **bank**, he saw some tillers

As he walked by the **bank**, he saw some tellers

II. Semantic relatedness of entities is context dependent and thus their representations are contextual

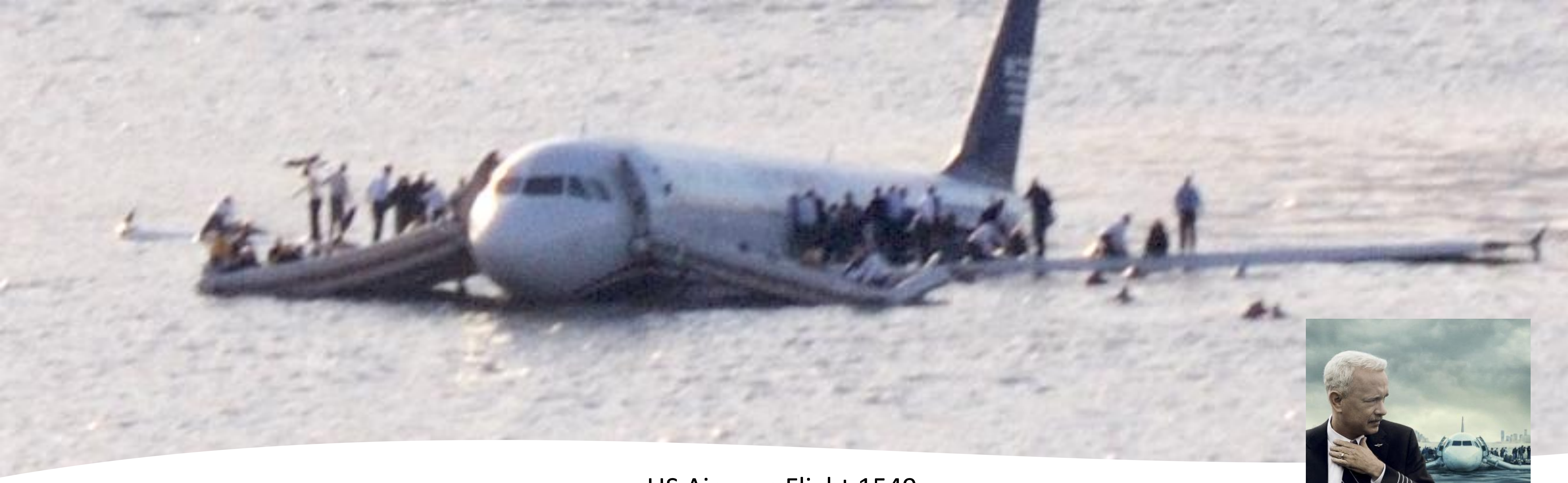


III. Representation of an entity can allow us to reconstruct or "generate" it



IV. It is possible to develop representations in an inductive manner (i.e., through empirical observations)

https://en.wikipedia.org/wiki/Johannes_Kepler
https://earthobservatory.nasa.gov/features/OrbitsHistory
https://plato.stanford.edu/entries/induction-problem/



US Airways Flight 1549

V. To act effectively and adaptively towards a goal, a being requires developing and using causal representations of entities at an appropriate level of complexity.

Only if we could have a mechanism that would enable developing such representations from empirical observations

Machine Learning and Deep Learning give mechanisms that allow us to use or develop effective representations from empirical observations that would generalize to unseen cases

Philosophical basis	Algorithms
I. Entities have (explicit or implicit) representations	Feature analysis / Representation learning
II. Semantic relatedness of entities is context dependent and thus their representations are contextual	Using Convolutions, Transformers or Graph Layers
III. Representation of any entity can allow us to reconstruct or “generate” it to a “sufficient”	Generative Machine Learning: GANs, Latent Diffusion Models
IV. It is possible to develop representations in an inductive manner (through empirical observations)	Learning Algorithm: Optimization of model parameters through gradient descent based on existing data Learning mechanisms: Self Supervised Learning, Next word prediction
V. To act effectively and adaptively towards a goal, a being requires developing and using causal representations of entities at an appropriate level of complexity.	Reinforcement Learning? Structural risk minimization (controls the model complexity and hence complexity of representations it learns)