<u>ScHARe</u>

Science Collaborative for Health disparities and Artificial intelligence bias REduction

Machine Learning Unveiled as a Bridge-building Trailblazer (*really a set of bridging paths falling under the AI* umbrella!*)

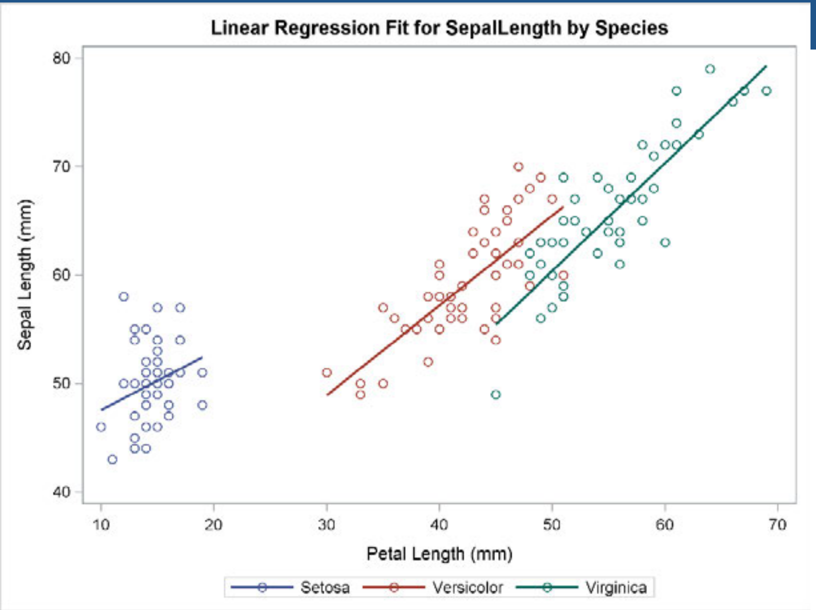
A. ML Essentials: roots in data analysis methods

B. Computational Strategies: varied forms of ‘learning’ and applying ‘learning’ algorithms…

* Recall *cynical* definition offered at recent NIH meeting: if it *actually* works in practice somehow, it’s ‘machine learning’ otherwise it may just be termed ‘artificial intelligence’ that still has more to learn…

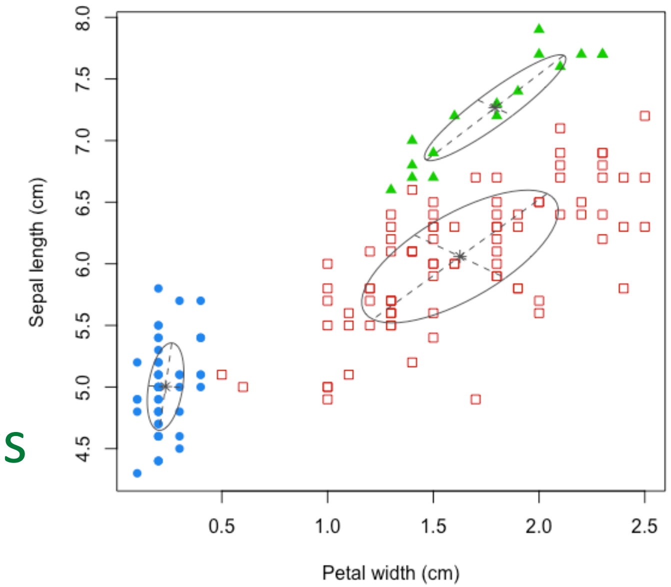
DEPARTMENT OF HEALTH & HUMAN SERVICES · USA
NIH National Institute of Diabetes and Digestive and Kidney Diseases

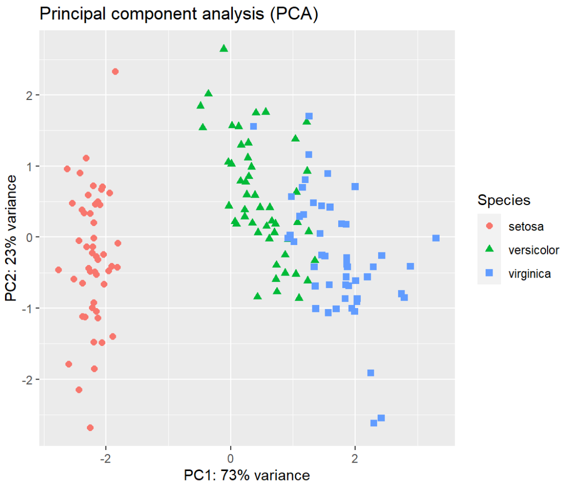
ML Essentials: roots in data analysis methods



- Data analysis methods to ‘learn’ how to predict patterns in data
  - Classic iris flower regression example

- Data analysis methods to ‘learn’ *novel* patterns in data: clustering & ‘mixture modeling’
  - Discover ‘clusters’ by length measures
  - Data reduction by principal components

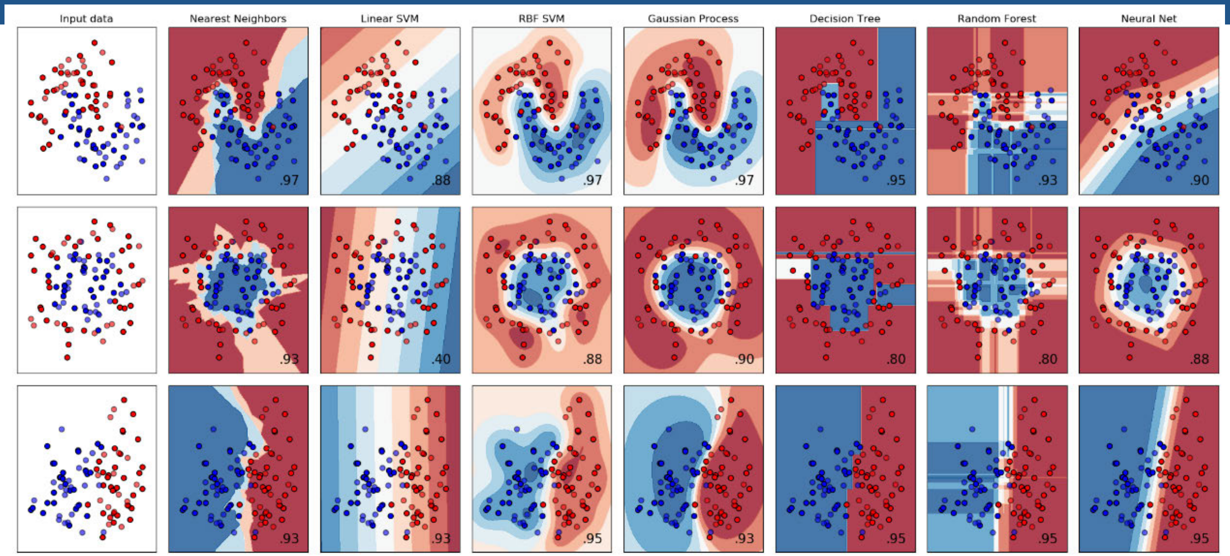




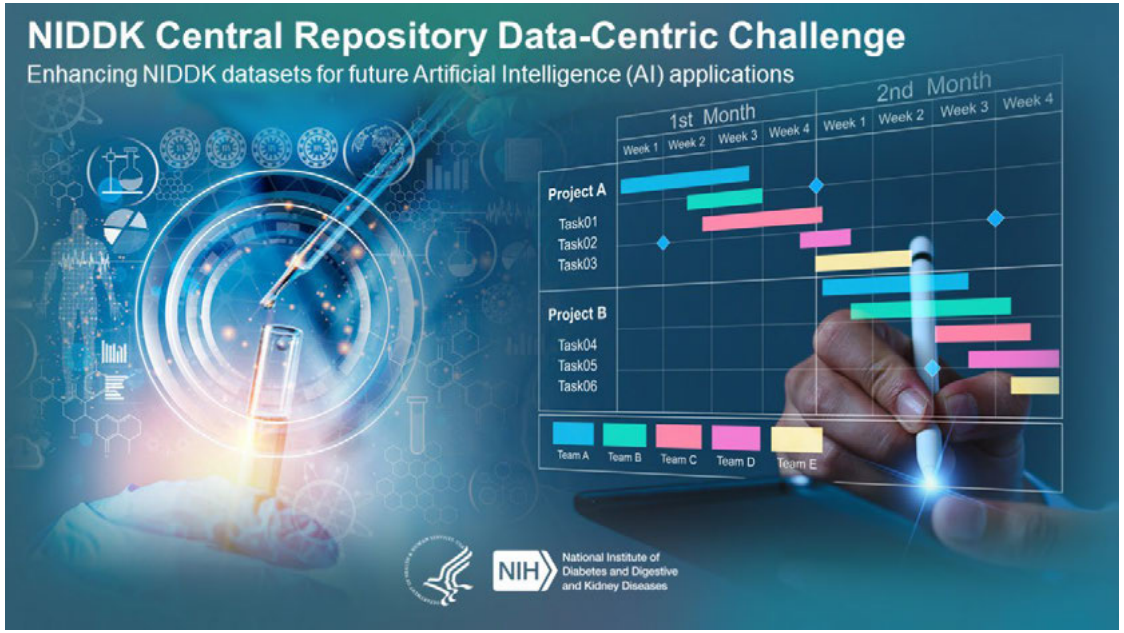
https://exeter-data-analytics.github.io/MachineLearning/_main_files/figure-html/unnamed-chunk-72-1.png | https://gexijin.github.io/learnR/book_files/figure-html/13-5-1.png

ML Essentials: roots in data analysis methods

- Data analysis methods to ‘learn’ how to predict patterns in data
- Data analysis methods to ‘learn’ *novel* patterns in data: clustering



- Relates to UN-supervised v. semi-supervised v. Supervised learning
  – Hearken back to prior ScHARe Think-a-thon
  – Underway: PHASE 2 of NIDDK CR Data-Centric Challenge (till Jan 22, 2024)

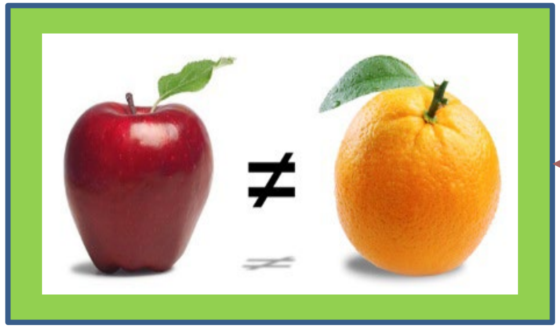


ML Essentials: supervised v. semi-supervised v. unsupervised learning

'Machine Learning' as a tool for Data Science (thus, for health equity research)

- *Does one term cover all approaches?* Types of ML, matching use cases & data
- *e.g. (**extent of 'supervision'**; goals of analysis)*

- *What does "extent of 'supervision'" mean in this context?*



Supervised Learning

<u>Supervision here</u>: *each instance is given exactly 1 'label' to distinguish*

https://www.gettyimages.ca/detail/news-photo/republican-presidential-hopeful-and-texas-governor-george-news-photo/51543471?adppopup=true

ML Essentials: supervised v. semi-supervised v. unsupervised learning

'Machine Learning' as a tool for Data Science (thus, for health equity research)

- *Does one term cover all approaches?* Types of ML, matching use cases & data
- *e.g. (**extent of 'supervision'**; goals of analysis)*

- *What does "extent of 'supervision'" mean in this context?*

Semi-supervised Learning

RED color shows up on the **RIGHT**
No picture lacks a 'mirror' image

<u>Supervision here</u>:
Only some instances given a 'label' to distinguish 'labeling' pattern overall…

https://www.gettyimages.ca/detail/news-photo/republician-presidential-hopeful-and-texas-governor-george-news-photo/51543471?adppopup=true

ML Essentials: supervised v. semi-supervised v. unsupervised learning

'Machine Learning' as a tool for Data Science (thus, for health equity research)

- *Does one term cover all approaches?* Types of ML, matching use cases & data
- *e.g. (**extent of 'supervision'**; goals of analysis)*

- *What does "extent of 'supervision'" mean in this context?*

Four-eyed Fox
Four-eyed Wolf
Red kidneys & butterflies around a 2 person drum circle

Unsupervised Learning

Supervision here: *Only intrinsic parts of instances used to 'label' them, elicit any pattern overall…*

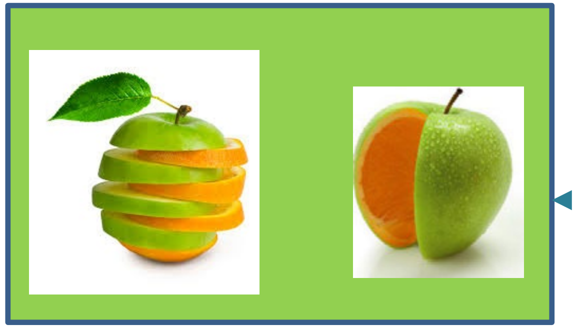
https://www.gettyimages.ca/detail/news-photo/republican-presidential-hopeful-and-texas-governor-george-news-photo/51543471?adppopup=true

ML Essentials: supervised v. semi-supervised v. unsupervised learning

'Machine Learning' as a tool for Data Science (thus, for health equity research)

- *Does one term cover all approaches?* Types of ML, matching use cases & data
- *e.g. (**extent of 'supervision'**; goals of analysis)*

- *What does "extent of 'supervision'" mean in this context?*



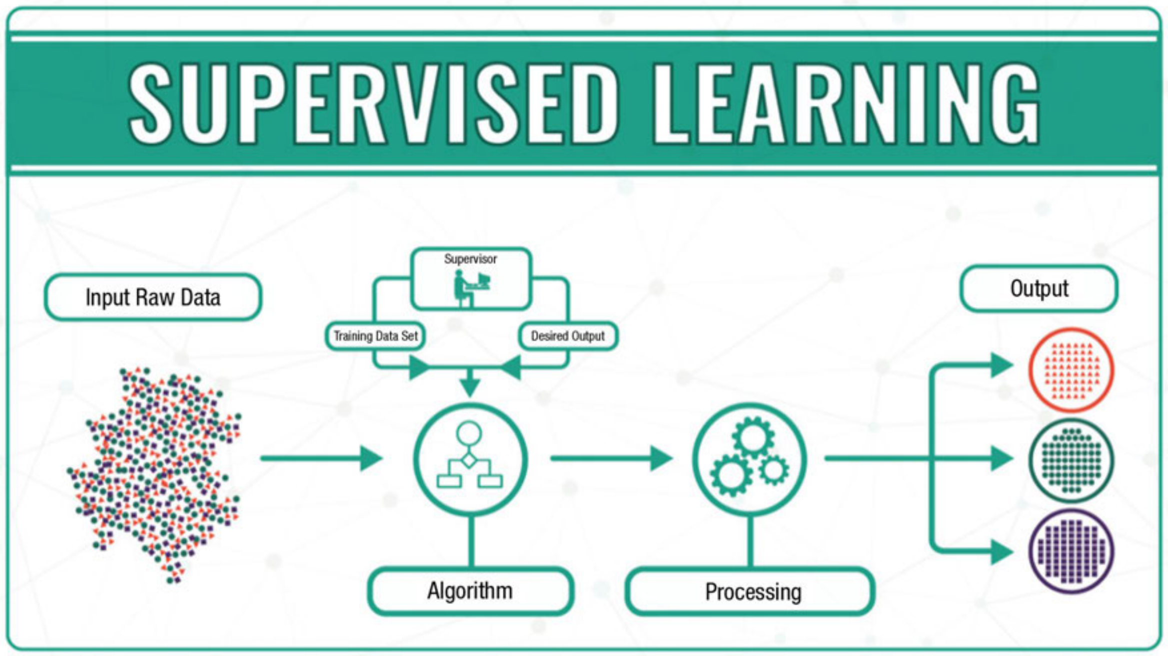
Unsupervised Learning

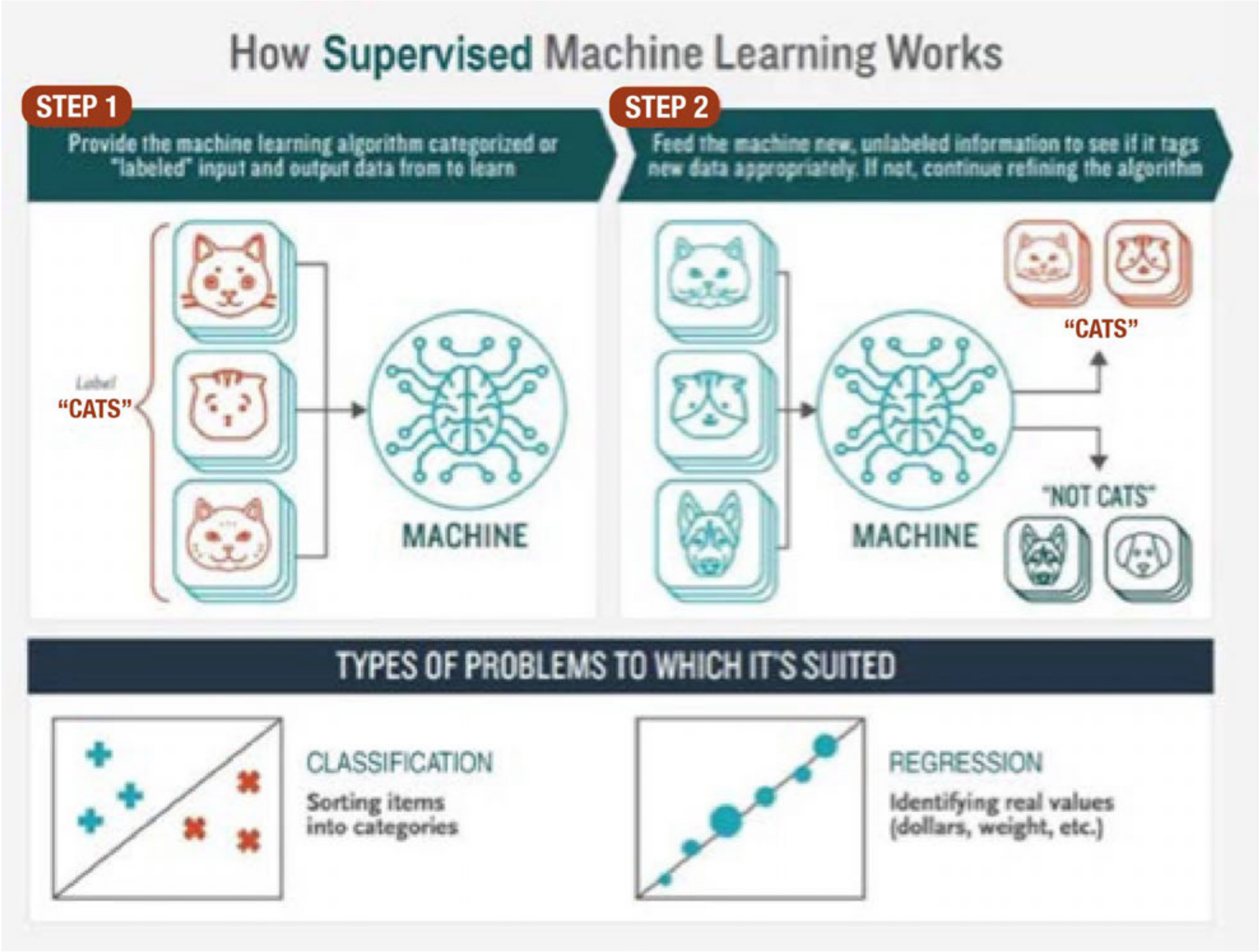
<u>Supervision here</u>: *LACK of such can elicit patterns NOT typically within human intuition*

https://www.gettyimages.ca/detail/news-photo/republician-presidential-hopeful-and-texas-governor-george-news-photo/51543471?adppopup=true

ML Essentials: supervised v. semi-supervised v. unsupervised learning

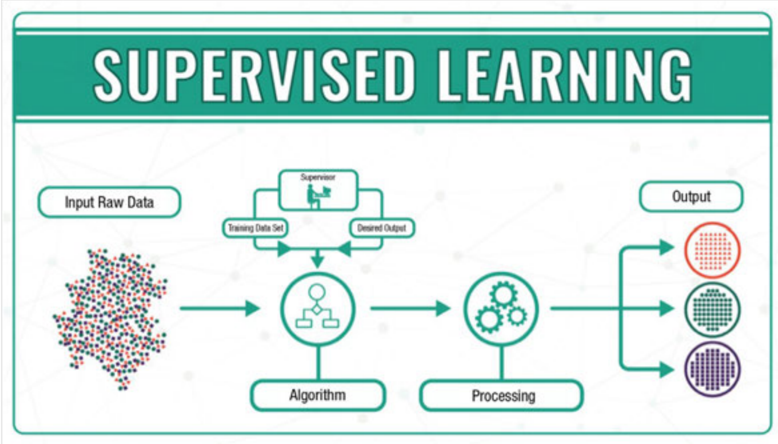
- From Booz Allen Team for CKD





ML Essentials: supervised v. semi-supervised v. unsupervised learning

- From non-NIH-funded team



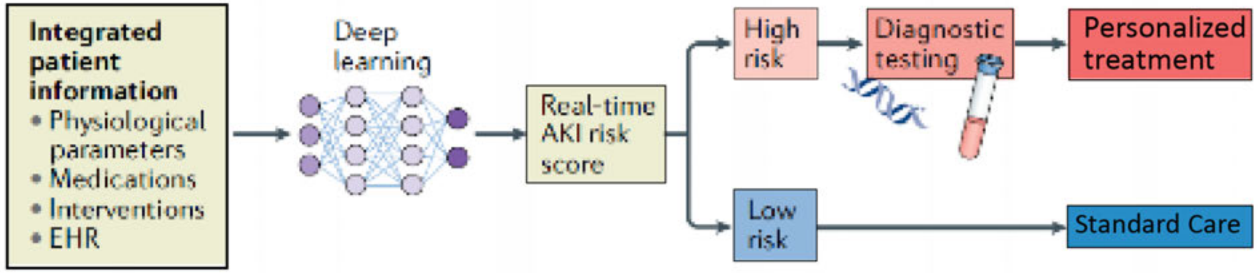
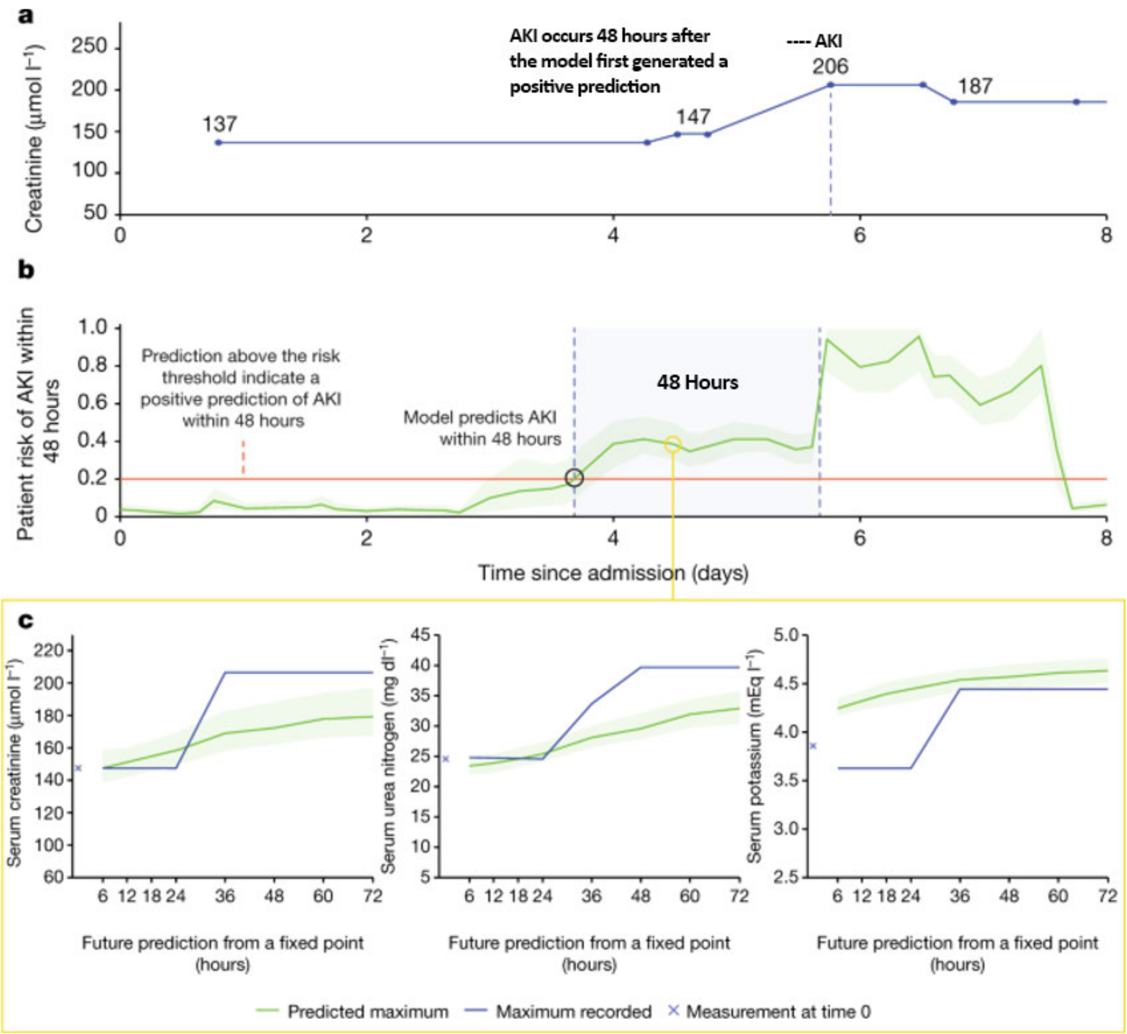


Fig. 1 | **Implementation of deep learning algorithms to identify patients at high risk of AKI.** Deep learning algorithms developed to support clinical decisions in real time should be based on integrated patient information, including electronic health records (EHRs) with detailed medical history (including ongoing problems and procedures), physiological parameters (such as vital signs and laboratory results) and medication details. Acute kidney injury (AKI) risk scores derived from such an algorithm would stratify patients and inform clinical decisions, including the use of additional diagnostics to enable personalized treatment.

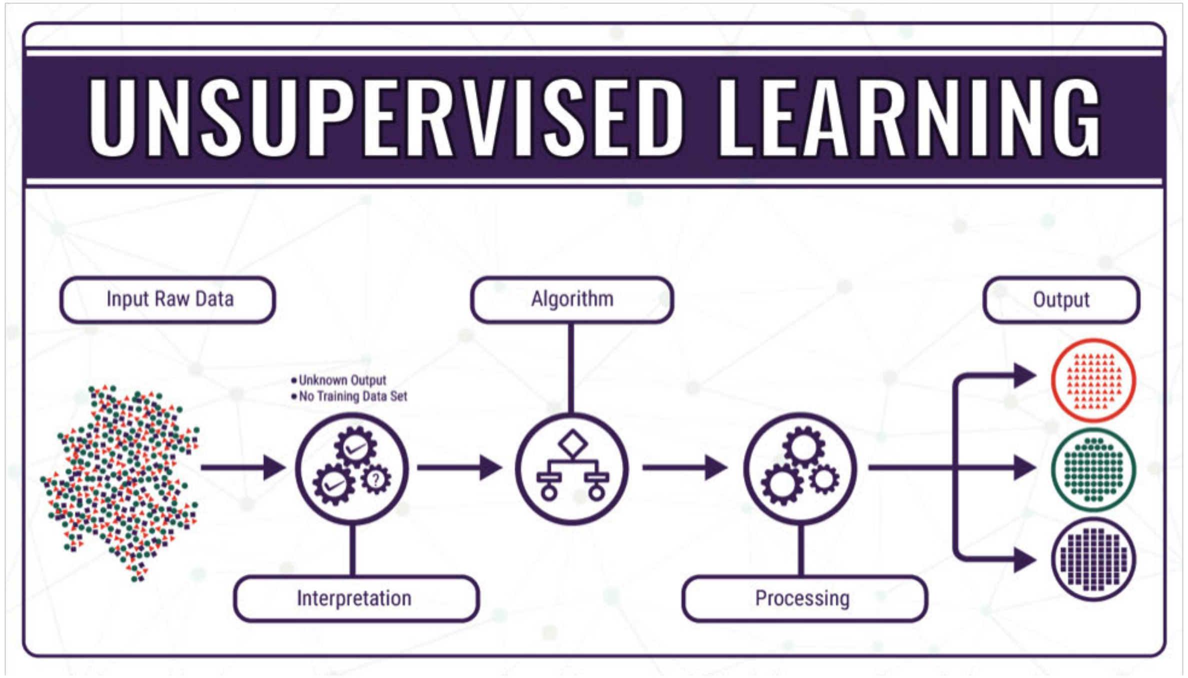


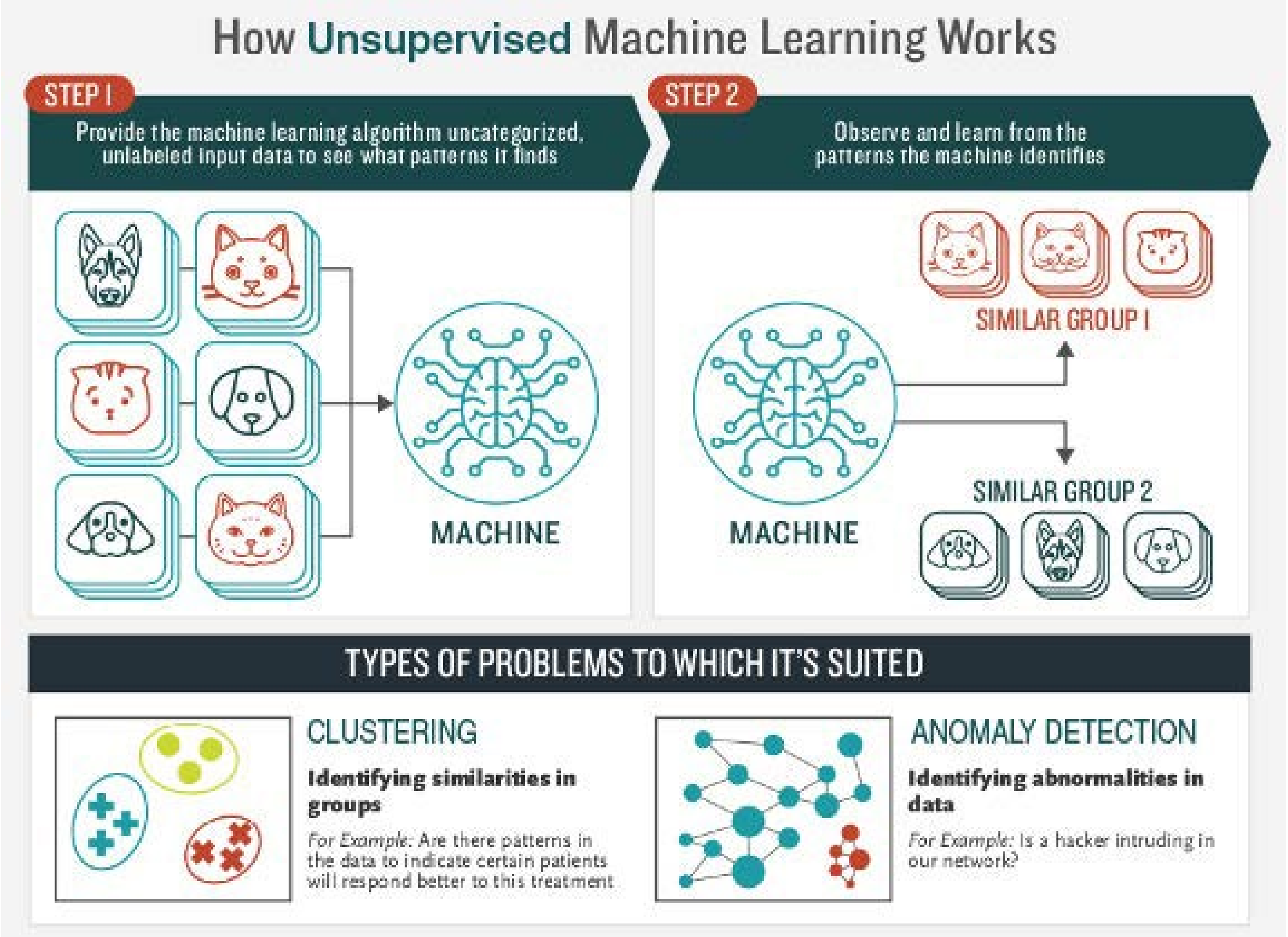
Figures 1 from editorial on and paper of DeepMind's AKI approach in Tomašev, N. et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* **572**, 116–119 (2019).

"We make use of several open-source libraries to conduct our experiments: the machine learning framework TensorFlow (https://github.com/tensorflow/tensorflow) along with the TensorFlow library Sonnet (https://github.com/deepmind/sonnet)"

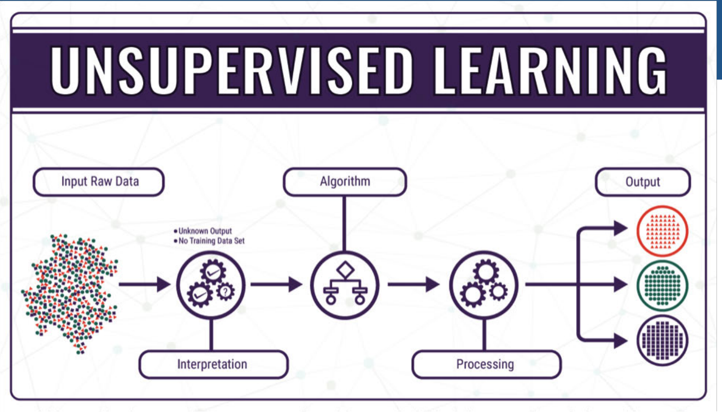
ML Essentials: supervised v. semi-supervised v. unsupervised learning

- From Booz Allen Team for CKD





ML Essentials: supervised v. semi-supervised v. unsupervised learning



- From NIDDK-funded team →
- From other NIH-funded team ↓
  - Mammograms
  - Role of density
  - Blend: un+sup

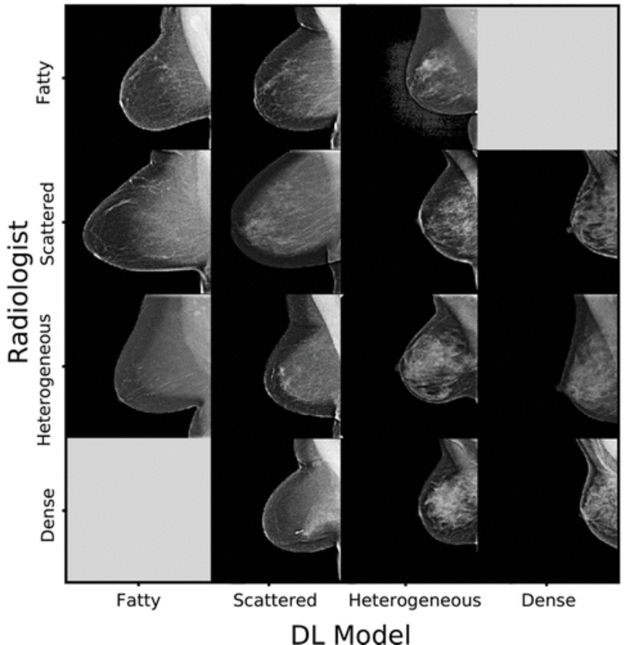
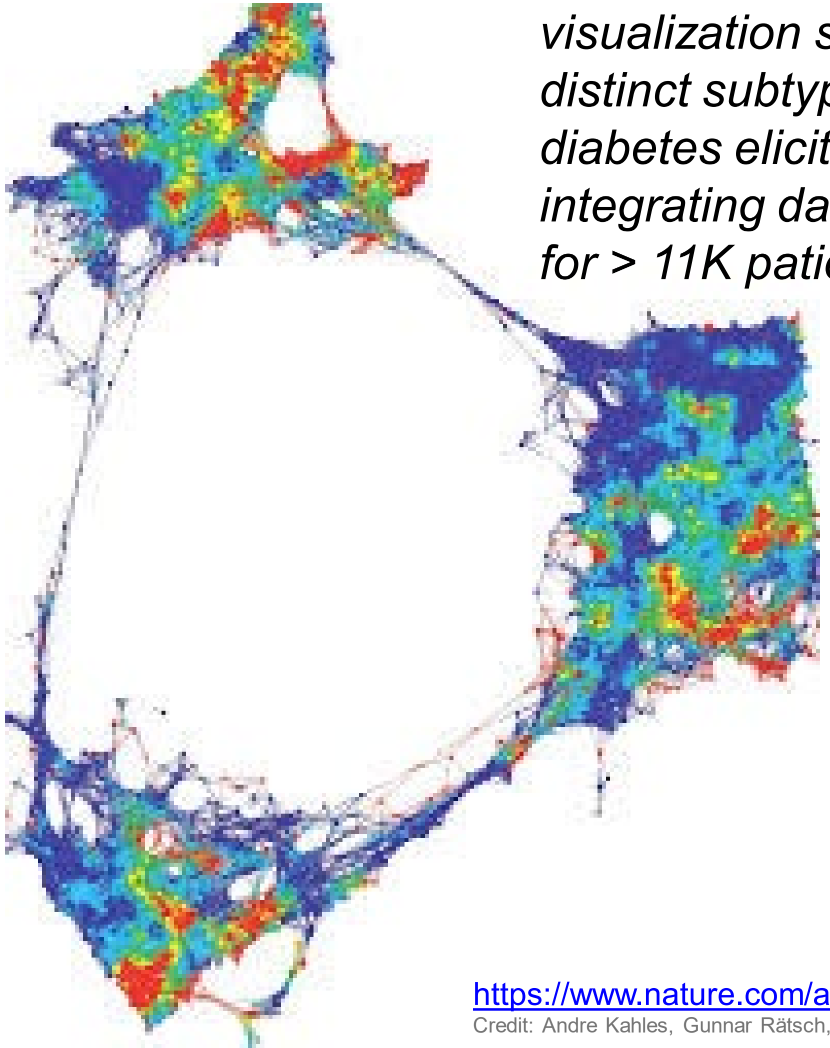


Figure 1d: Test set assessment. Comparison of the original interpreting radiologist assessment with the deep learning (DL) model assessment for **(a)** binary and **(c)** four-way mammographic breast density classification. **(b, d)** Corresponding examples of mammograms with concordant and discordant assessments by the radiologist and with the DL model.

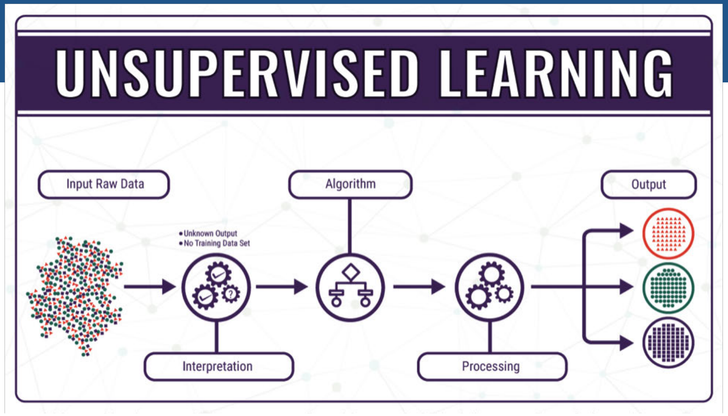


From NIDDK report: *network visualization showing 3 distinct subtypes of Type 2 diabetes elicited after integrating data from EHRs for > 11K patients*

https://www.nature.com/articles/d42473-019-00035-5

Credit: Andre Kahles, Gunnar Rätsch, Chris Sander

ML Essentials: supervised v. semi-supervised v. unsupervised learning



- From NIDDK-funded team →
- From other NIH-funded team ↓
  - Mammograms
  - Role of density
  - Blend: un+sup

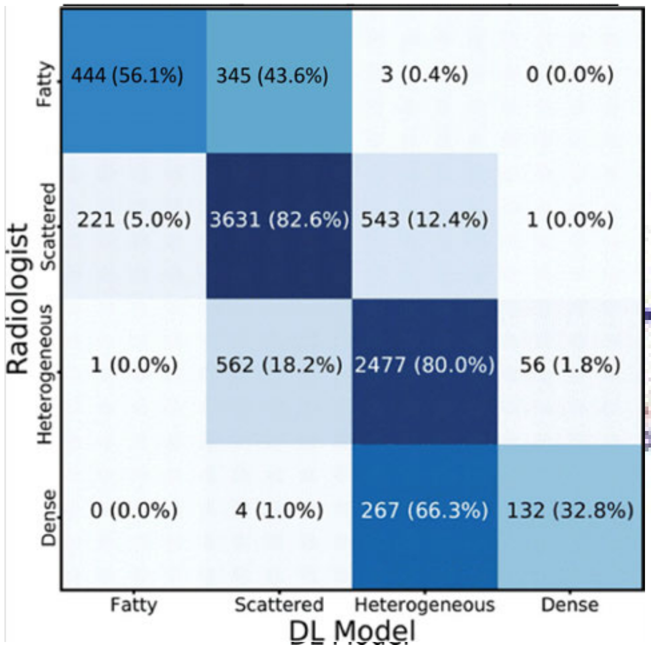
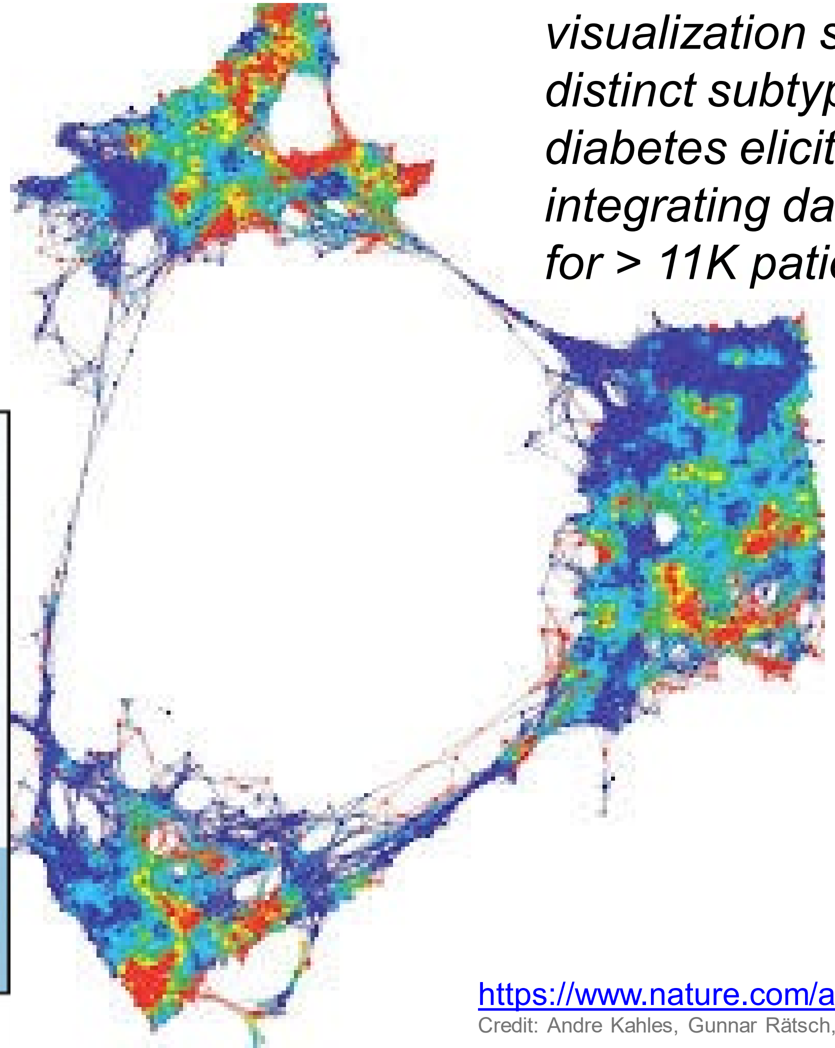


Figure 1d: Test set assessment. Comparison of the original interpreting radiologist assessment with the deep learning (DL) model assessment for **(a)** binary and **(c)** four-way mammographic breast density classification. **(b, d)** Corresponding examples of mammograms with concordant and discordant assessments by the radiologist and with the DL model.

From NIDDK report: *network visualization showing 3 distinct subtypes of Type 2 diabetes elicited after integrating data from EHRs for > 11K patients*



https://www.nature.com/articles/d42473-019-00035-5

Credit: Andre Kahles, Gunnar Rätsch, Chris Sander

Machine Learning Computational Strategies

- We now engage participants to check our mutual understanding