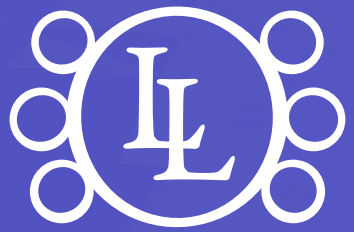
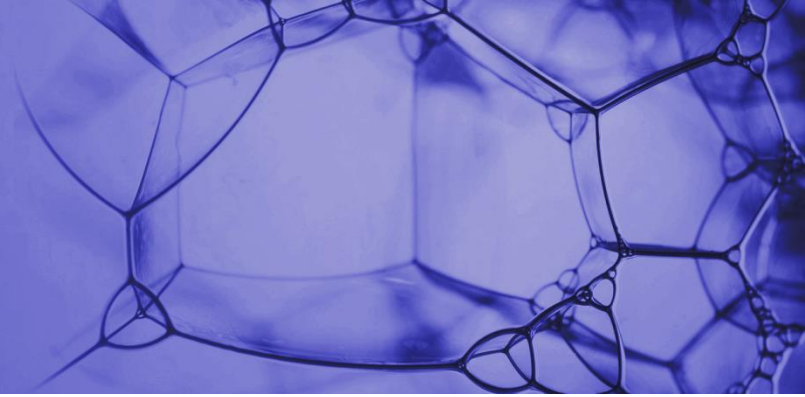


LOSCHMIDT
LABORATORIES



Artificial Intelligence

in Microbiology

by Stanislav Mazurenko, PhD

mazurenko@mail.muni.cz

Outline

- Motivation
- Introduction to AI and ML
- Recent applications in Microbiology

Motivation

Motivation: sequences and chemicals

- Large volumes of digital data
- Affordable computing power and storage
- Complex study objects

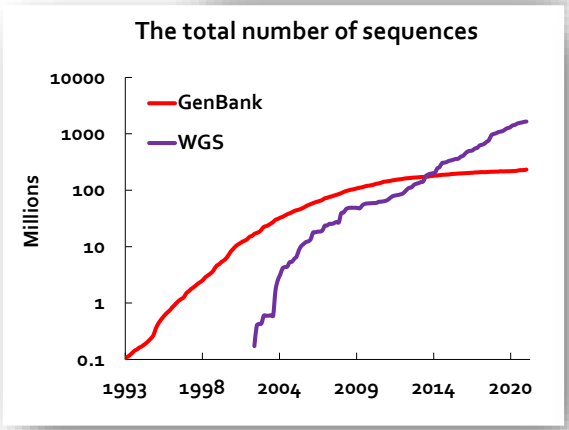
GenBank FREE
Eric W Sayers ✉, Mark Cavanaugh, Karen Clark, James Ostell, Kim D Pruitt, Ilene Karsch-Mizrachi

JGI GOLD biobank uk
Improving the health of future generations

RCSB PDB PROTEIN DATA BANK
UniProt
BRENDA
FIREPROT DB

PubChem
ChEMBL
JOURNAL OF CHEMICAL INFORMATION AND MODELING

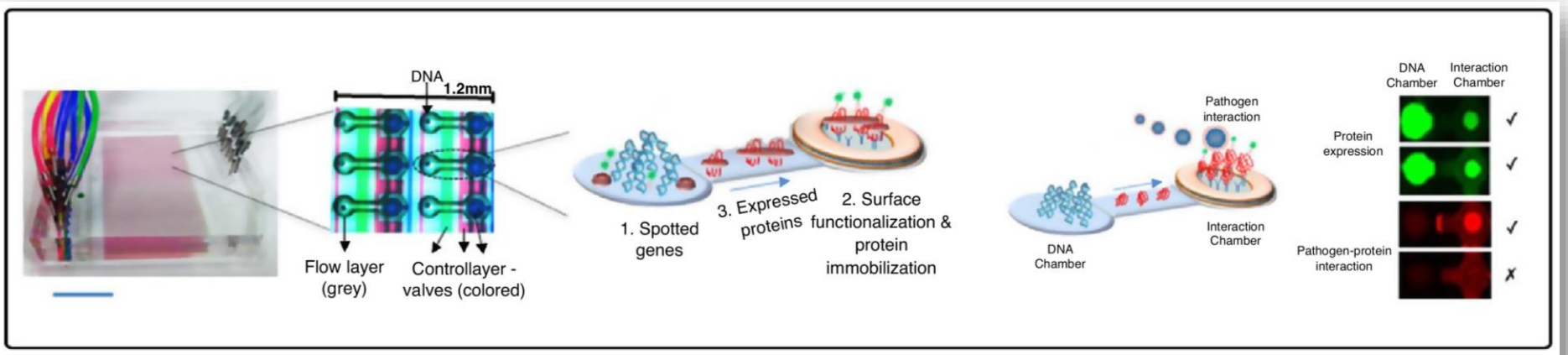
ZINC 15 – Ligand Discovery for Everyone
Teague Sterling and John J. Irwin*



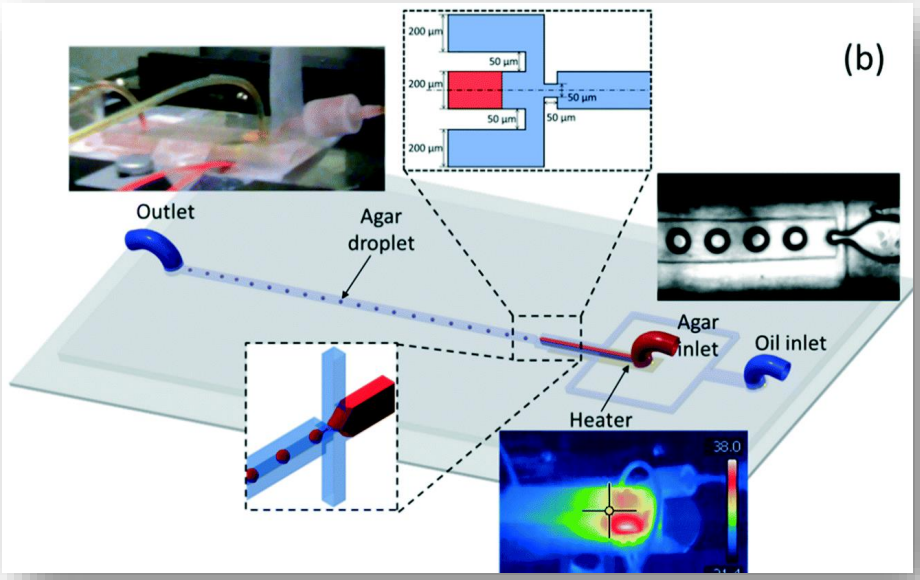
14 Recommended Categories | 253 Origin Categories | 12 Core Categories | Categories

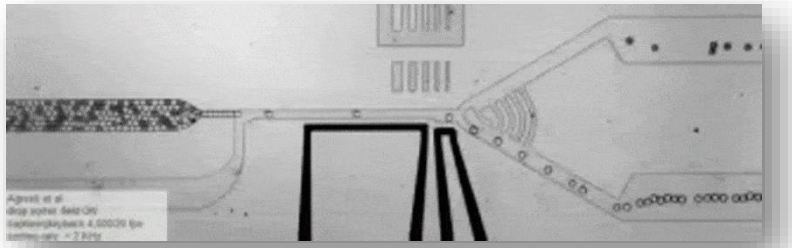
Category ID	Description	Items
1014	Brain MRI	922
1005	Cognitive function summary	5
1004	Diet and alcohol summary	321
1002	Early life	13
1007	Education and employment	16
1017	Genomics	30
100113	Geographical and location	13
1015	Heart MRI	39
1019	Linked health outcomes	75
1016	Main abdominal MRI fields likely to be of interest to researchers.	22
1018	Mental health	186
1006	Physical measure summary	66
1001	Primary demographics	8
1003	Self-reported medical conditions	118

Motivation: big experimental data



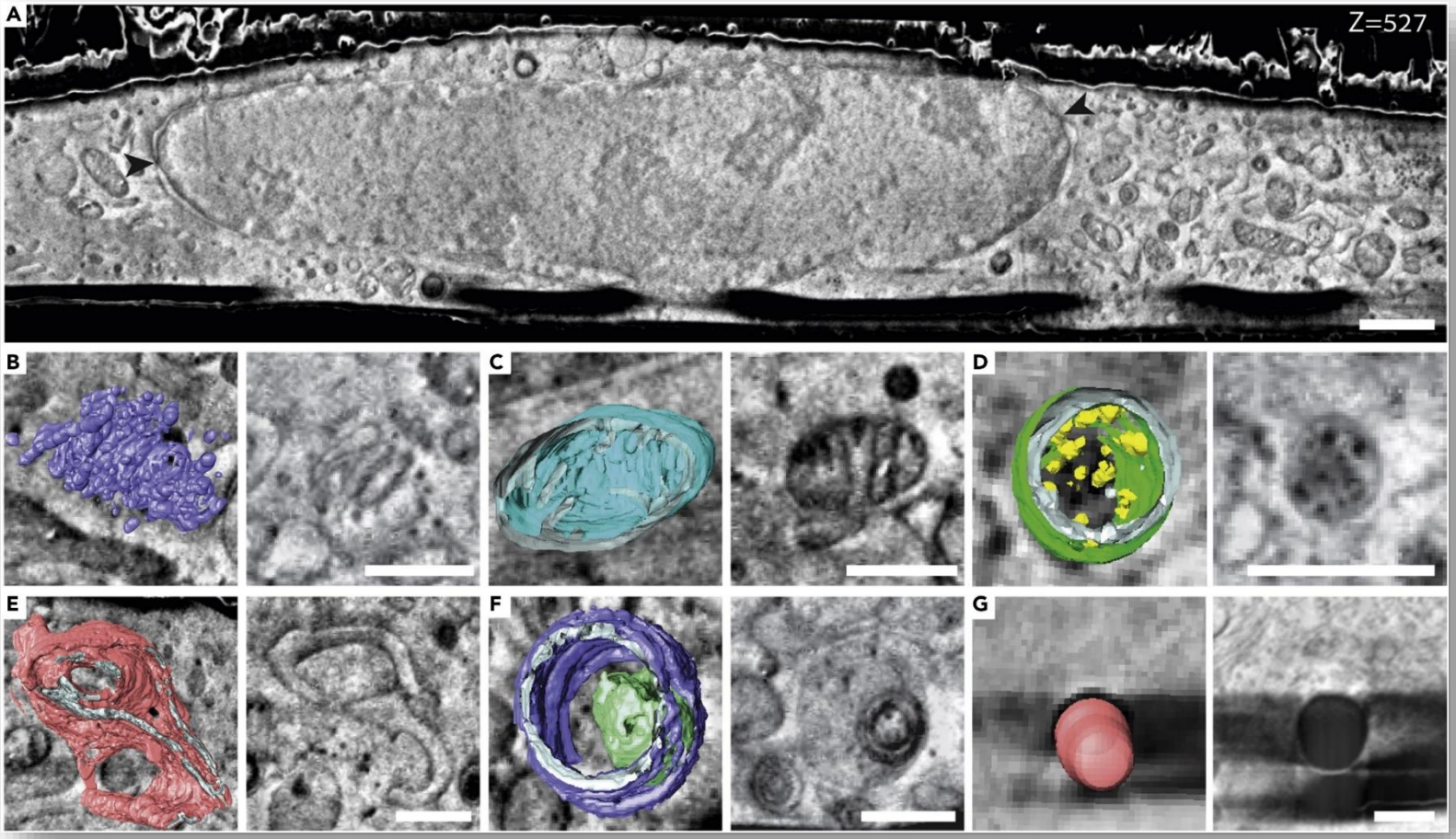
Current Opinion in Biotechnology

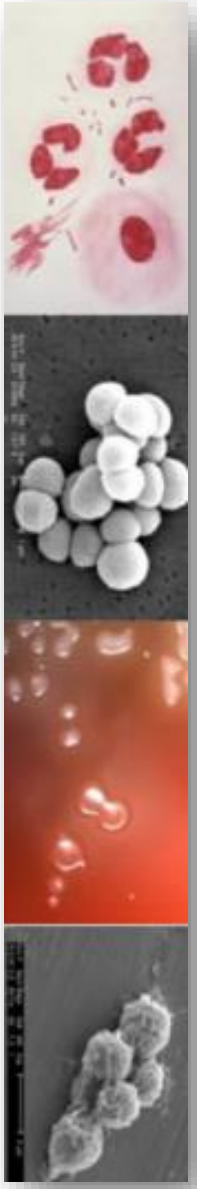


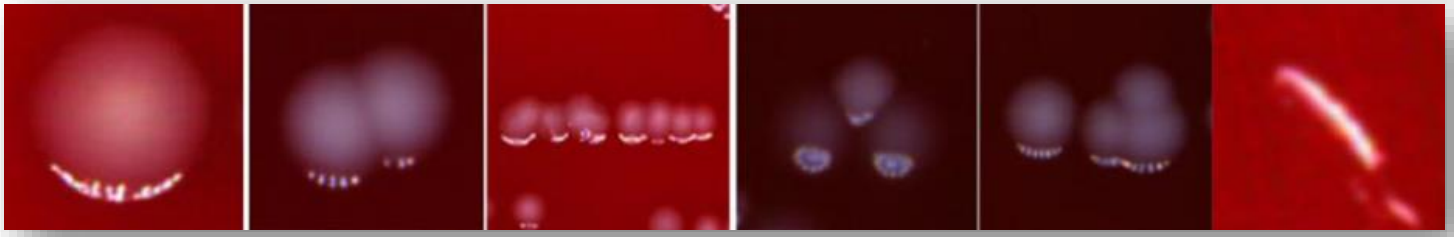


Source: Scheler et al. "Recent developments of microfluidics as a tool for biotechnology and microbiology." Current opinion in biotechnology 2019 Khater et al. "Picoliter agar droplet breakup in microfluidics meets microbiology application: numerical and experimental approaches." Lab on a Chip 2020.

Motivation: cell imaging





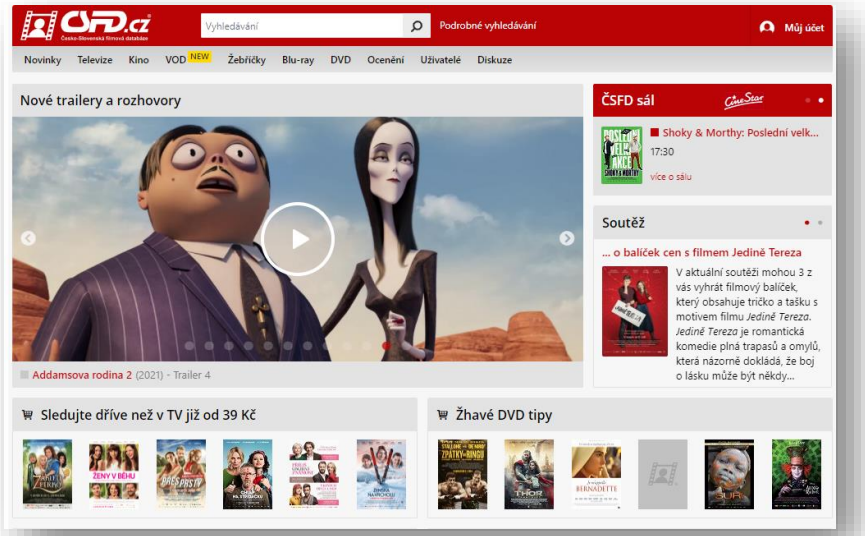


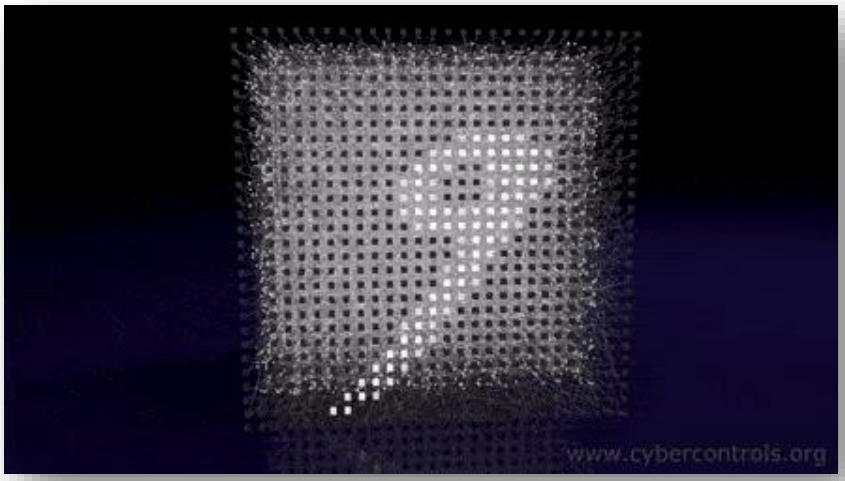
Source: Scher et al. "In-situ fiducial markers for 3D correlative cryo-fluorescence and FIB-SEM imaging." iScience (2021): 102714.

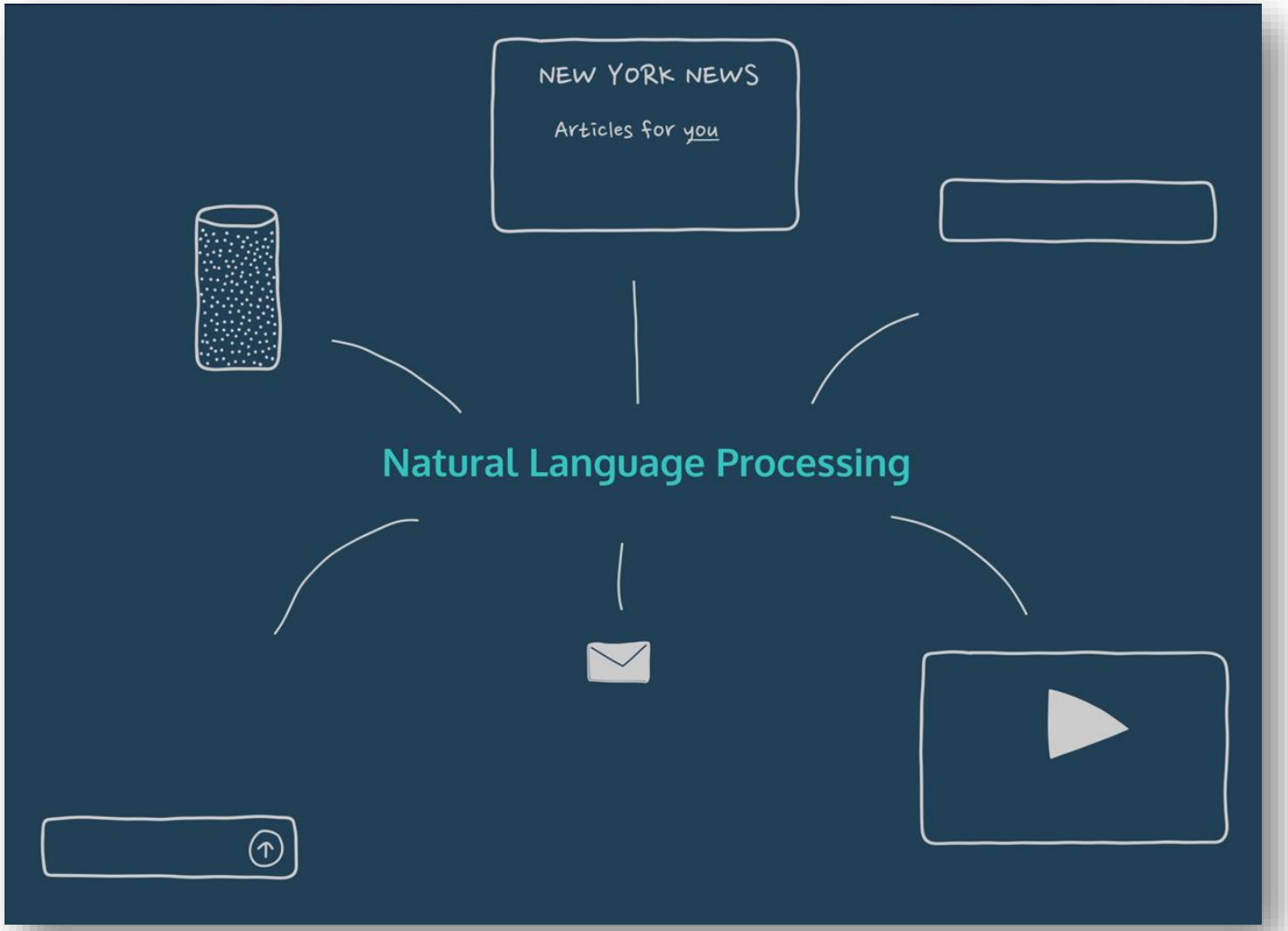
Introduction to AI and ML

Introduction to AI and ML

- Recommendation engines
- Image & speech recognition
- Anomaly detection
- Natural language processing
- Data mining…







Source: towardsdatascience.com

Introduction to ML

Faces

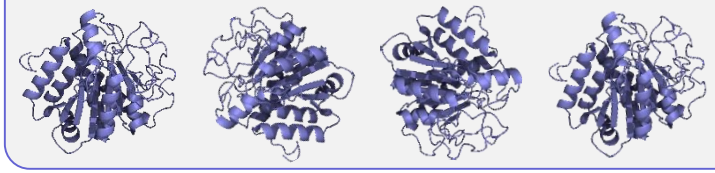
Not faces

- **Historically, people tried to find rules themselves, e.g. detection of particular shapes or color contrasts;**
- **Often such manual rules are too simplistic to give good results;**
- **Machine Learning gives the means to generates those rules automatically!**

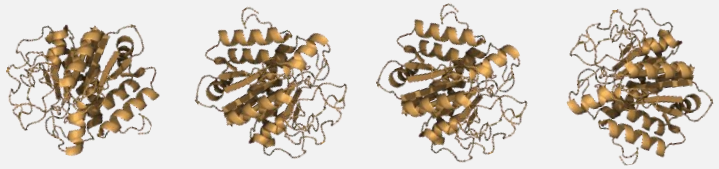
F( **) = + 1**

F( **) = - 1**

Non-vaccine candidates:



Vaccine candidates:

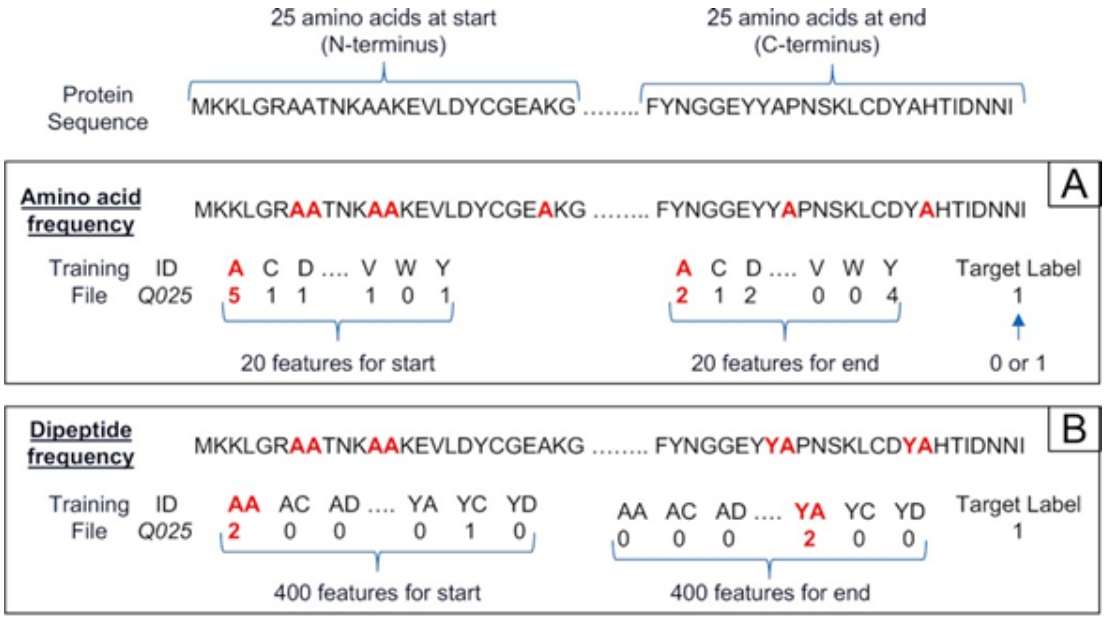


Basics of ML: data representation

MKKLGRAATNKAAKEVLDYCGEAKG… → **Feature vector: (5, 1, 1, -5.67, 0.69, …)**

Examples:

- **AA frequency**
- **AA sequence**
- **Conservation scores**
- **Structural elements**
- **…**

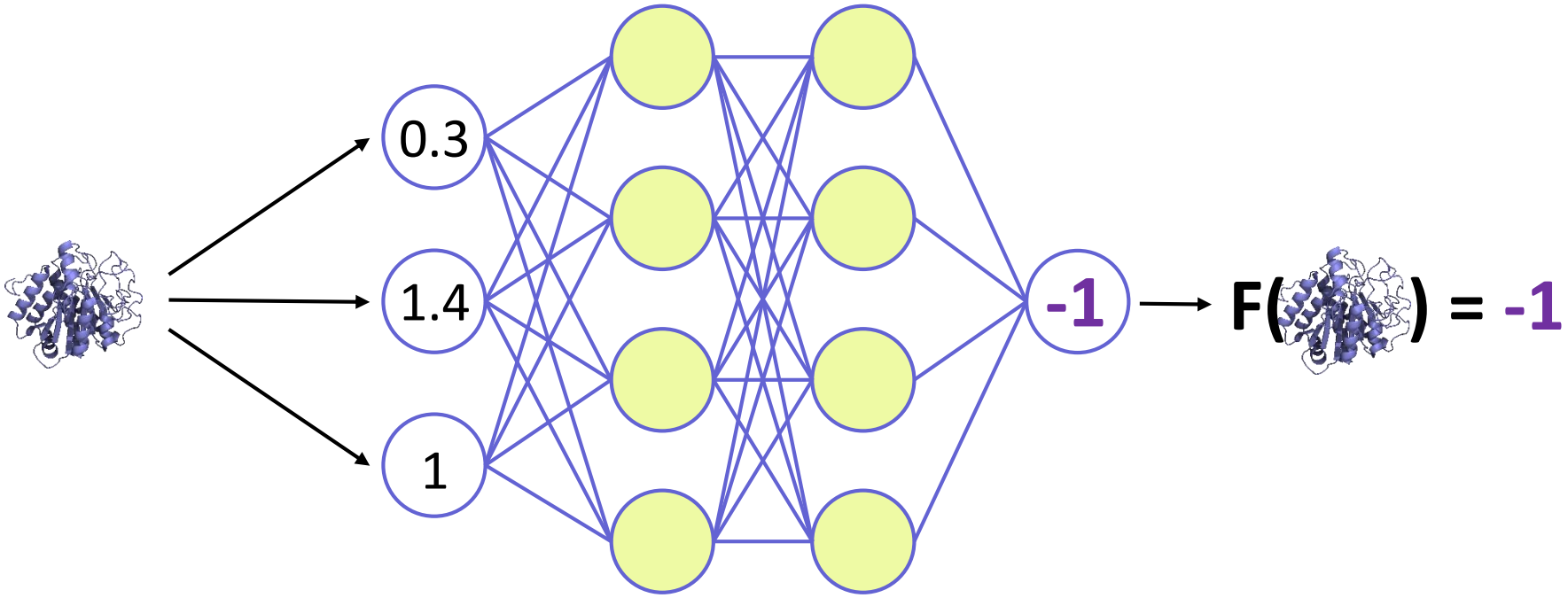


One-hot encoding:

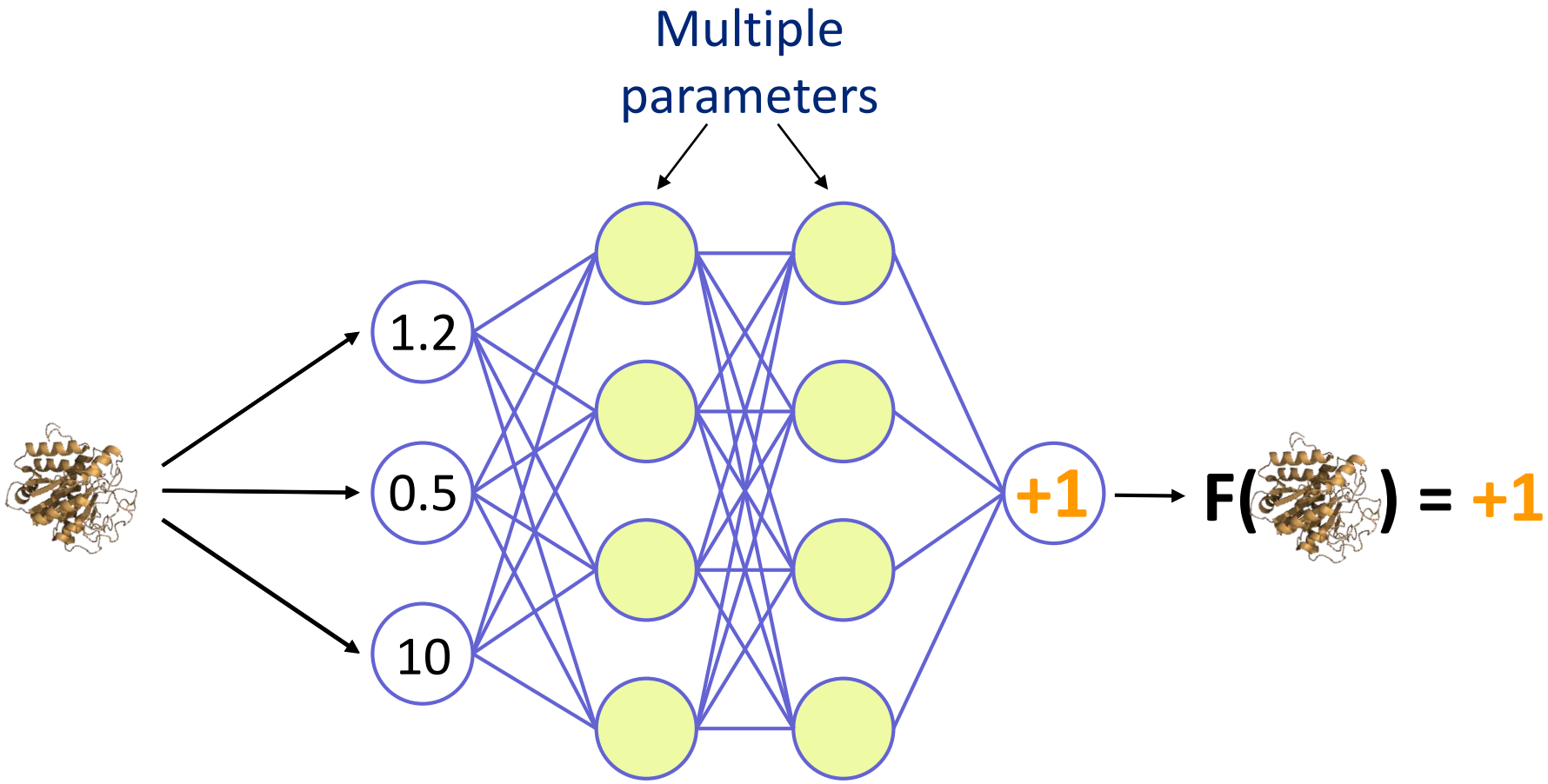
	K	K	L	G	R	A	A	T	…
A	0	0	0	0	0	1	1	0	…
K	1	1	0	0	0	0	0	0	…
L	0	0	1	0	0	0	0	0	…
G	0	0	0	1	0	0	0	0	…
R	0	0	0	0	1	0	0	0	…
T	0	0	0	0	0	0	0	1	…
…									

Source: Goodswen et al. "Machine learning and applications in microbiology." FEMS Microbiology Reviews (2021).

Basics of ML: training

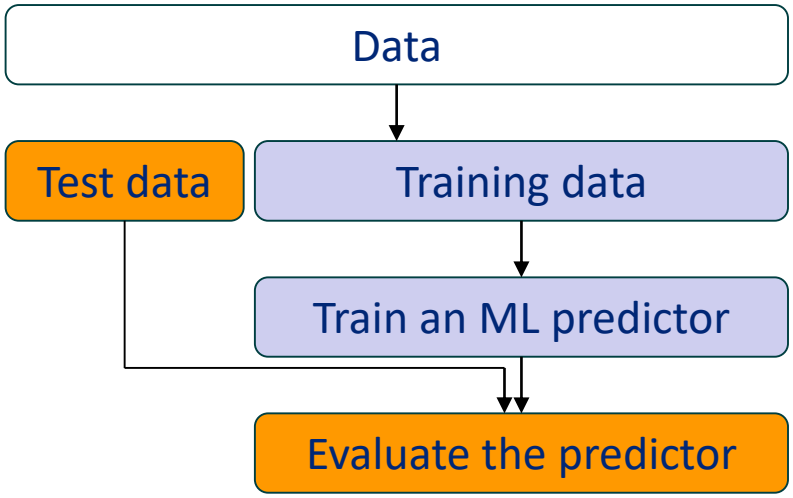


Basics of ML: training

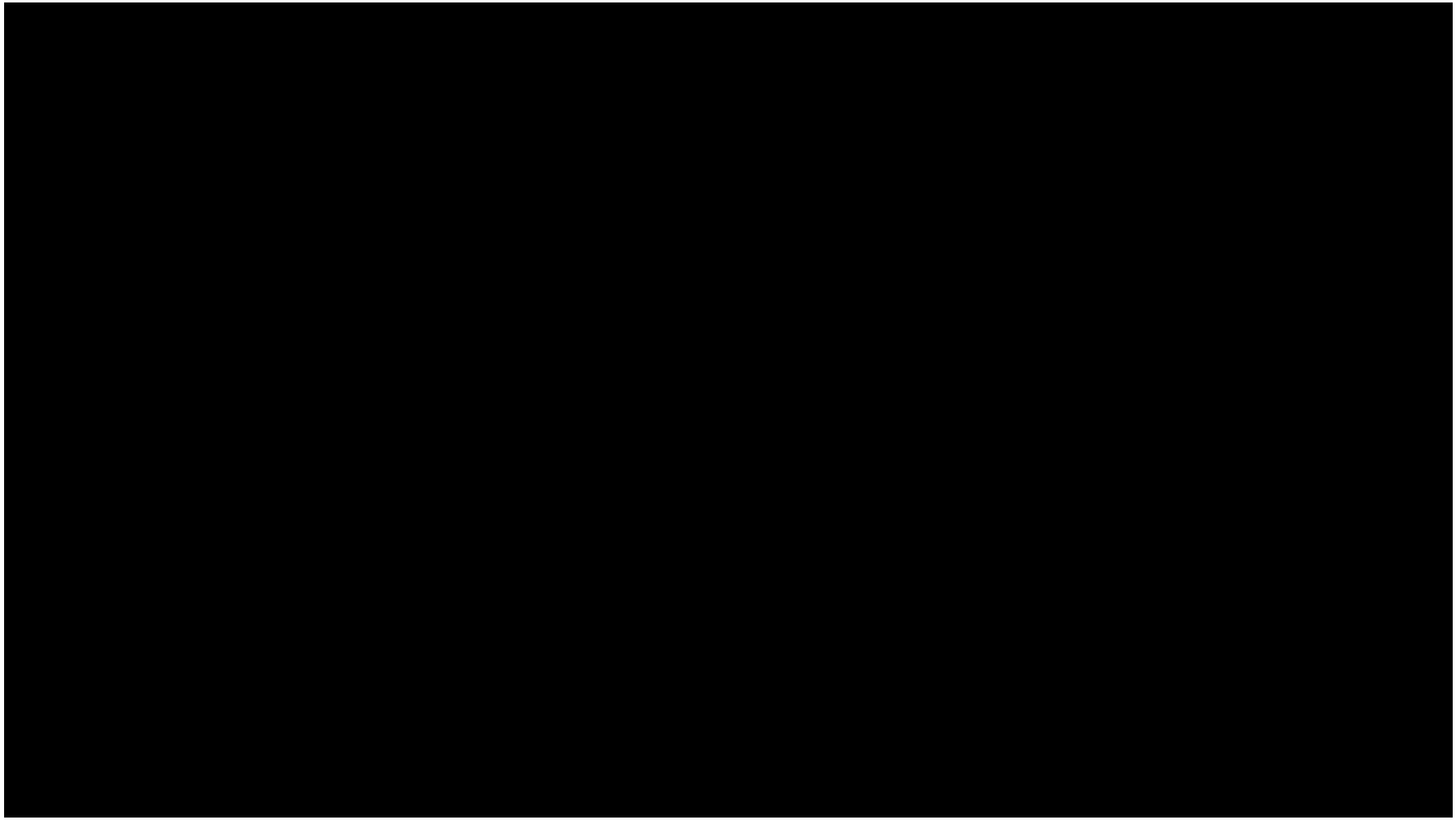


Basics of ML: validation

- **The goal of ML is to identify generalizable patterns in your training data.**
- **These patterns must be valid for future data!**
- **Therefore, the core of ML protocol is to evaluate the predictor on the test data, hidden from the predictor:**



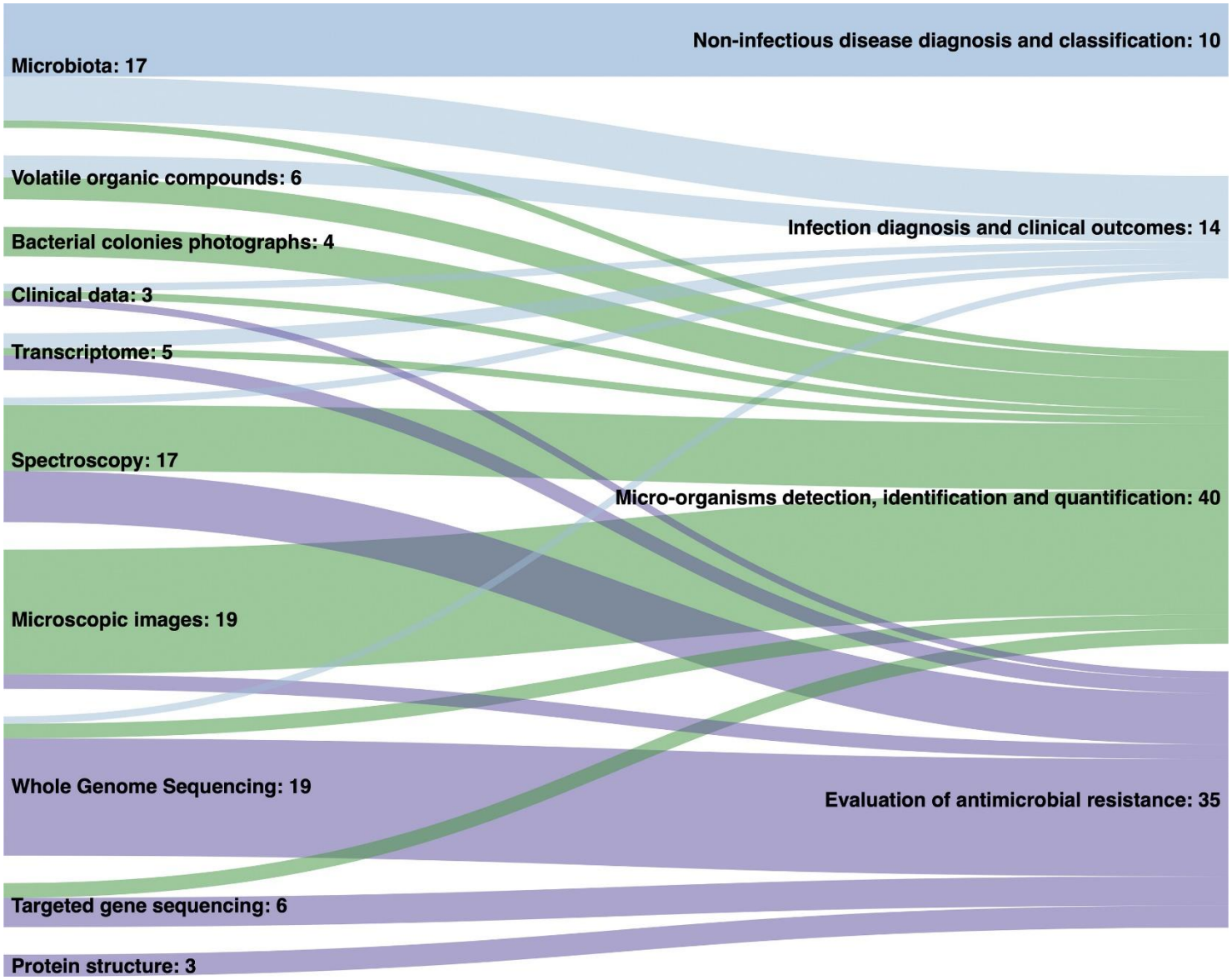
Artificial Neural Networks



Link: https://vimeo.com/154085950

Recent applications

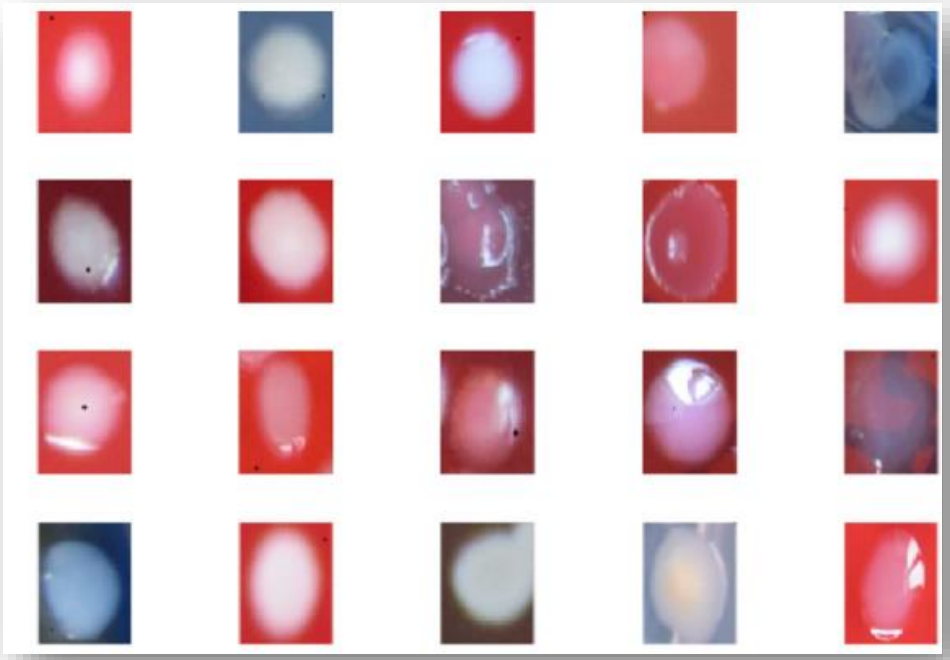
Overview



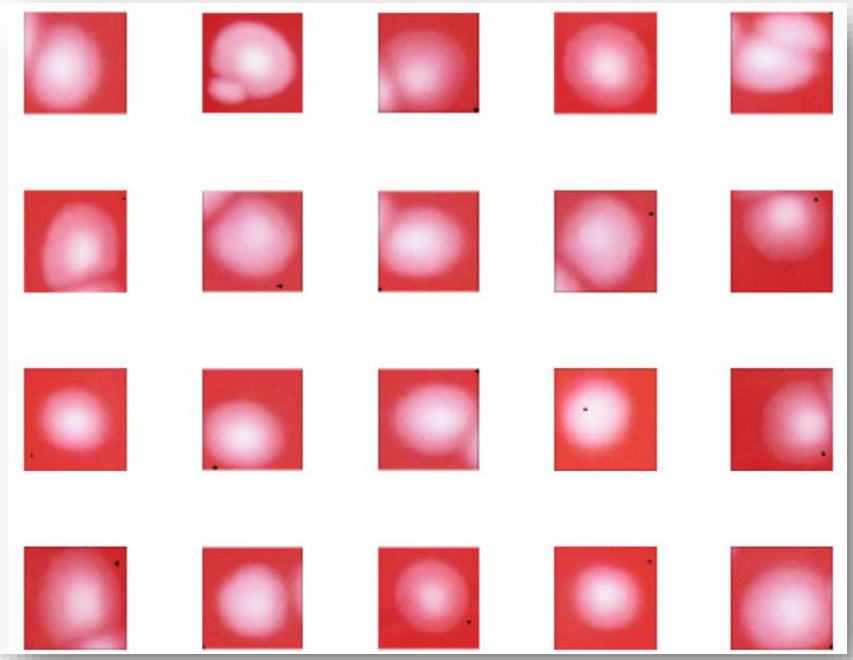
Source: Peiffer-Smadja et al. "Machine learning in the clinical microbiology laboratory: has the time come for routine practice?" Clinical Microbiology and Infection 2020

Bacterial colony morphology

Interclass variations



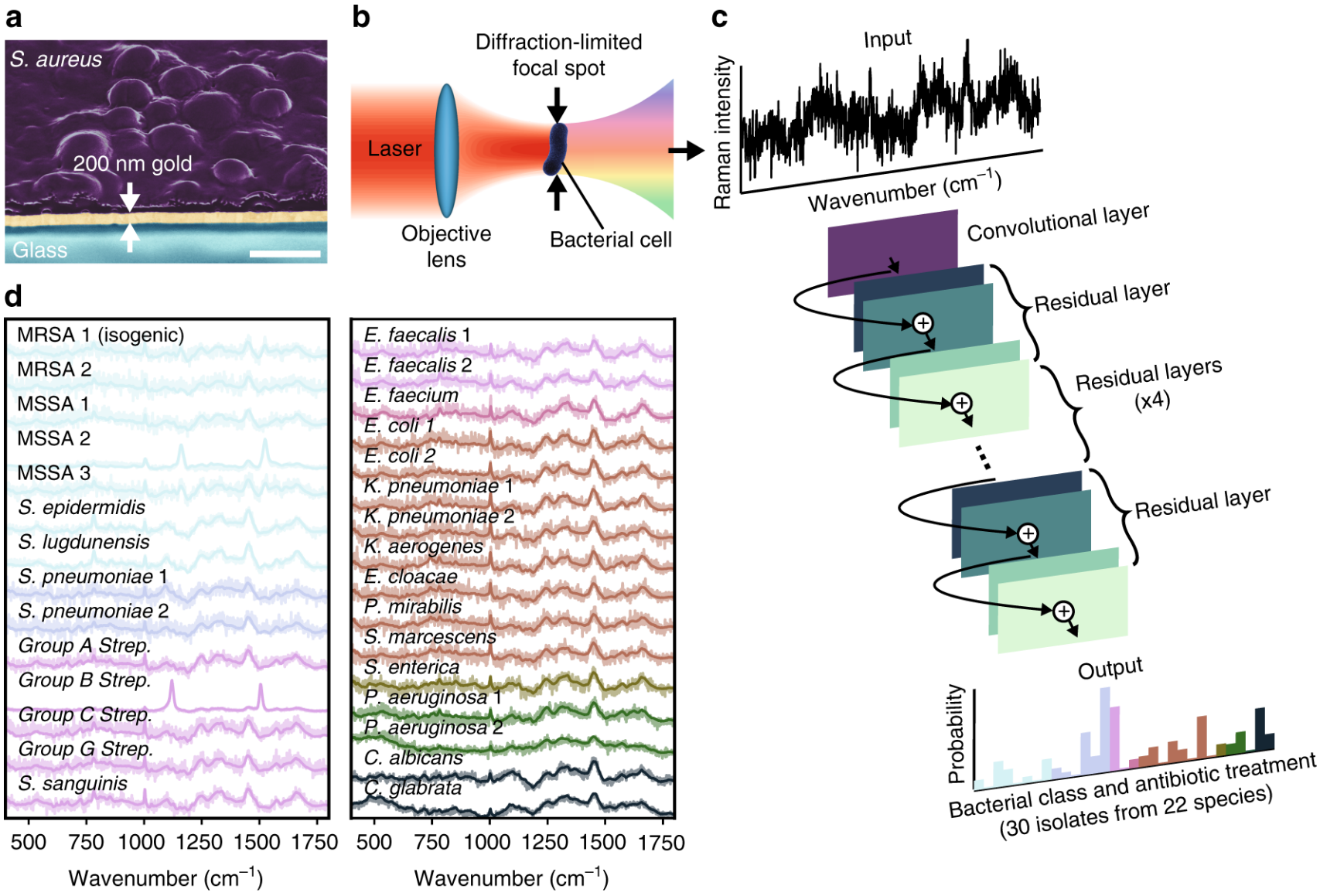
Intraclass variations (*Streptococcus agalactiae*)



A convolutional neural network was able to discriminate between 18 classes of bacterial colonies.

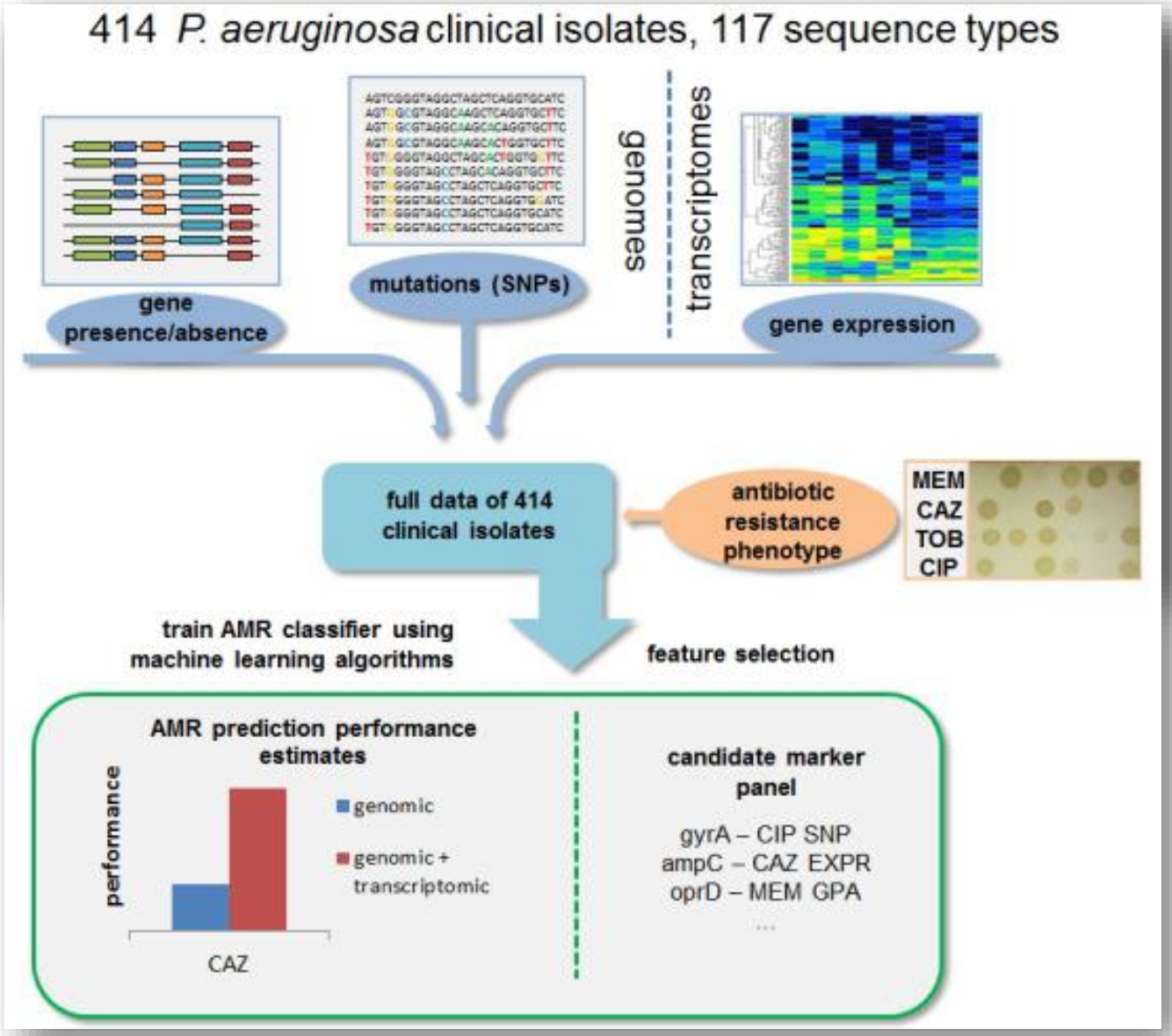
Source: Huang, Lei, and Tong Wu. "Novel neural network application for bacterial colony classification." Theoretical Biology and Medical Modelling 15.1 (2018): 1-16.

Identification of pathogens



Source: Ho Chi-Sing et al. "Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning." Nature communications 10.1 (2019): 1-8.

Antimicrobial resistance

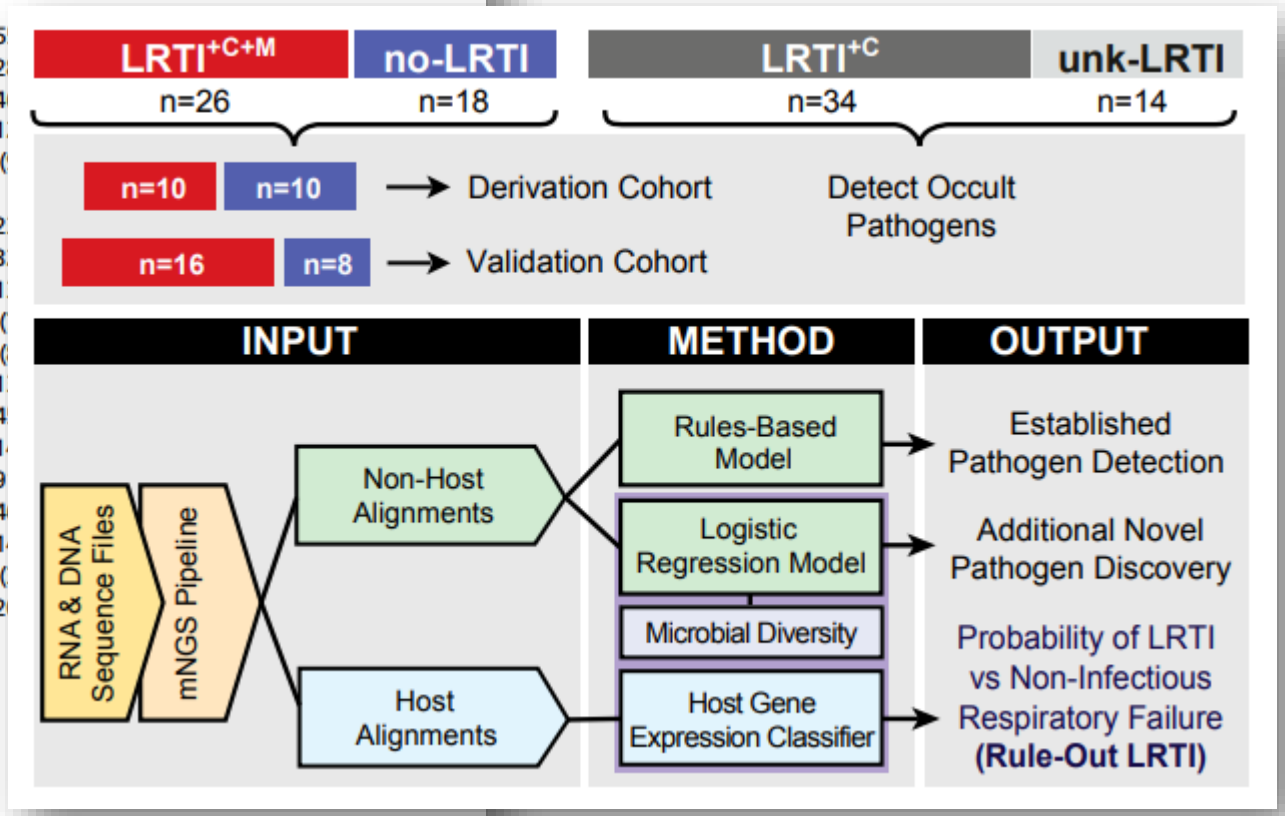


Source: Khaledi Ariane et al. "Predicting antimicrobial resistance in Pseudomonas aeruginosa with machine learning-enabled molecular diagnostics." EMBO molecular medicine 12.3 (2020): e10264.

Clinical outcomes

Table 1. Demographics and clinical characteristics of study cohort

Cohort characteristics	Cohort overall	LRTI^{+C+M}	no-LRTI
Patient characteristics			
Total enrolled	92	26	18
Age, y	62[†]	61	63
Female gender	31 (34%)	6	9
Race			
African American	5		
Asian	26 (2		
Caucasian	50 (4		
Other	11 (1		
Hispanic ethnicity	8 (		
Comorbidities and outcomes			
Bacteremia	21 (2		
Nonpulmonary infections	29 (3		
COPD	12 (1		
Diabetes mellitus	6 (		
Congestive heart failure	7 (		
Current smoker	12 (1		
Immune suppression	41 (4		
Solid-organ transplantation	13 (1		
Prior antibiotic use	84 (9		
Community acquired pneumonia	42 (4		
Hospital acquired pneumonia	13 (1		
Ventilator associated pneumonia	3 (		
30-d mortality	18 (2		
Clinical metrics			
Max temperature, °C			
Max WBC count, 10^6 cells/μL			
Max heart rate, bpm			
Max respiratory rate, breaths/min			
SIRS criteria, mean			
APACHE III score, mean	97	101	94
Pneumonia severity index, mean	151	148	137



Source: C4X Discovery, Fernández-Torras et al. "Connecting chemistry and biology through molecular descriptors." Current Opinion in Chemical Biology 66 (2022): 102090.

Summary

- Machine Learning method is a powerful data-driven alternative to traditional modelling;
- One turns data into numbers (features) and trains a generic algorithm to discriminate between labels in the feature space;
- It is essential to have a separate test set for evaluation of the resulting predictor;
- In Microbiology, a wide range of tasks is already solved by Machine Learning.

Bi9680En: AI in Biology, Chemistry, and Bioengineering

- **Období: podzim**
- **Rozsah: přednáška 2 hodiny/týden**
- **Vyučující: Dr. Stanislav Mazurenko**
- **Osnova:**
  - modern bio-challenges: drug design, DNA interpretation, protein engineering
  - types of AI algorithms and workflow for designing predictors
  - clustering algorithms, random forests, artificial neural networks
  - features, databases, and predictors used in applications