# CS502: Deep Learning in Biomedicine

Maria Brbić

Fall 2023

École
polytechnique
fédérale
de Lausanne

# This Lecture

- Logistics
- Why this course?
- Deep learning basics

# Logistics: Teaching Staff

## Instructor

Maria Brbic

## Teaching assistants

El Mahdi Chayti
TA

Akash Dhasade
TA

Artyom Gadetsky
TA

Bettina Messmer
TA

Nicolas Talabot
TA

Shuo Wen
TA

Omar El Malki
AE

Jeremy Goumaz
AE

Anna-Rose Gryspeert
AE

Colin Hofmann
AE

Somesh Mehra
AE

Sander Miesen
AE

Daniel Molinuevo
AE

# Logistics: Communication

- **Moodle**
  - Course website: Moodle course link
  - We will post all course materials on Moodle: homeworks, practical sessions, project information
  - All announcements will be on Moodle in the Announcement channel

- **To reach teaching staff, always use:**
  - Ed discussion forum
  - If you have a very specific question, contact TAs/AEs

# Logistics: Communication

- **Forum: Ed Discussion**
  - Link: https://edstem.org/eu/courses/754/discussion/50022
  - Main way to ask your questions
  - Dedicated channels for lectures/exercises/homeworks/project related questions
  - The channels/threads are structured, please check if your question has already been answered before asking !
  - For general/administrative questions, use the "general" channel
  - You can use the "social" channel to discuss freely (no answer will be provided in this channel)

# Logistics: Website & Readings

- **Slides**
  - Final version available on Moodle after the lecture
  - Feedback form for each lecture
- **Readings:**
  - Deep Learning book by Goodfellow, Bengio, Courville
  - Deep Learning in Bioinformatics
  - Research papers
- **Optional readings:**
  - Papers and pointers to additional literature in slides
  - Especially useful for course projects

# Grading

- **Final grade will be composed of:**
  - **Homework: 50%**
    - Homework 1: 10% grade
    - Homework 2: 20% grade
    - Homework 3: 20% grade
  - **Course project: 50%**
    - Proposal: 5%
    - Code: 10%
    - Final report: 20%
    - Presentation: 15%
- **The course assignments and project will be based on Pytorch**

# Course Projects

- **Performed in groups of up to 3 students:**
  - Fine to have groups of 1 or 2 but 3 person teams can be more efficient. The team size will be taken under consideration when evaluating the scope of the project in breadth and depth
- **Teams:**
  - Form teams before October 15th
  - To find team members (if needed), use Finding team members channel on Moodle
  - To register a team, use Team registration channel on Moodle

# Course Projects

- **Course project:**
  - **Options 1&2:**
    - We will provide few-shot learning benchmark with dataset and algorithms
    - **Option 1:** Create a new benchmark dataset and test existing algorithms
    - **Option 2:** Implement a new algorithm and test on existing benchmarks
  - **Option 3:**
    - Choose biomedical problem of your choice and provide solution
  - More details to follow.

# Assignments Schedule

| Week | Assignment | Due on (23:59 CET) |
|---|---|---|
| 1 | Homework 1 release | Sun, Oct 8th |
| 3 | Homework 2 release | Sun, Oct 29th |
| 6 | Homework 3 release | Sun, Nov 19th |
| 12 | Homeworks debriefing | Fri, Dec 8th |
| 7 | Project release | Tue, Dec 12th |
| | Team forming | Sun, Oct 15th |
| | Project pitch and discussion | Fri, Nov 17th |
| | Project proposal deadline | Sun, Nov 19th |
| | Project presentations | Weeks 13 & 14 |

**Deadline policy:**

- Late periods are not acceptable, **no exceptions!**
- Submissions won't be graded in such case

# Assignments Schedule

This week!

| Week | Assignment | Due on (11:59pm PT) |
|---|---|---|
| 1 | Homework 1 release | Sun, Oct 8th |
| 3 | Homework 2 release | Sun, Oct 29th |
| 6 | Homework 3 release | Sun, Nov 19th |
| 12 | Homeworks debriefing | Fri, Dec 8th |
| 7 | Project release | Tue, Dec 12th |
| | Team forming | Sun, Oct 15th |
| | Project pitch and discussion | Fri, Nov 17th |
| | Project proposal deadline | Sun, Nov 19th |
| | Project presentations | Weeks 13 & 14 |

**Deadline policy:**

- Late periods are not acceptable, **no exceptions!**
- Submissions won't be graded in such case

# Prerequisites

- **Background in:**
  - Machine Learning
  - Mathematical concepts: linear algebra, probability and statistics
- **Programming:**
  - You should be able to write non-trivial programs in Python
- **No single topic is too hard by itself, but we will cover and touch upon many topics**

# Lectures

- 2h lecture/week & 2h practical session/week
- **Lectures:** Thursdays 10am@CM12 [2h]
- **Practical sessions:** Fridays 9am@INF1,2 [2h]
  - Implementation of topics covered in course
  - Additional materials covered
  - Walkthrough of homework assignments
  - **Attend practical sessions!**
- Once projects are released, they can be discussed on Fridays 8am@INF1,2 [1h]
  - Walkthrough of project expectations

# Course Goals

- Cover recent deep learning methods and demonstrate how can they be used for different problems in biomedical domain
- Understand fundamentals of covered methods, their advantages and failure modes
- Understand basics of different biomedical problems to which these methods are applicable
- Utilize learned concepts in practice by implementing common deep learning methods including different neural network architectures and methods for different problem settings

# Course Lectures Schedule

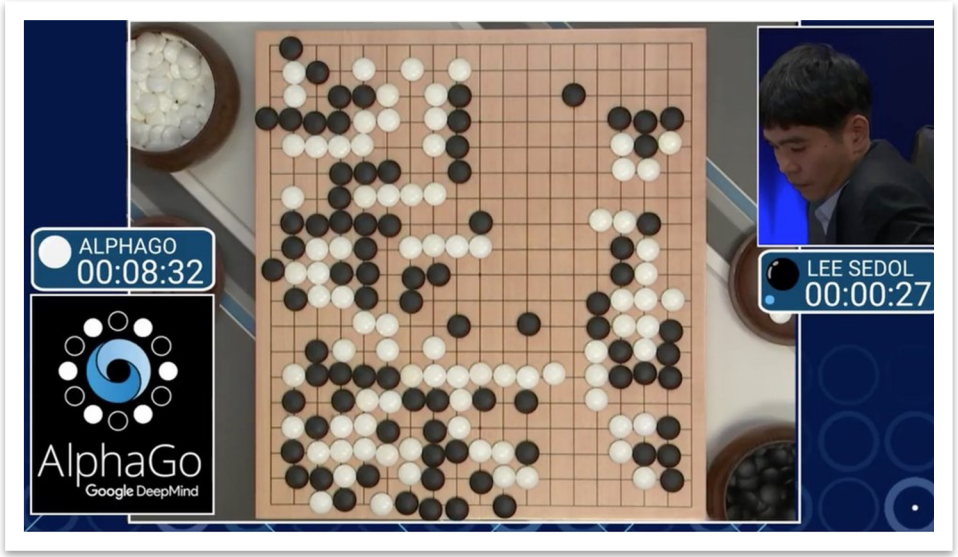
| Date | Topic |
|---|---|
| **Thu, Sep 21st** | Course intro + deep learning basics |
| **Thu, Sep 28th** | Convolutional neural networks |
| **Tue, Oct 5th** | Graph neural networks |
| **Thu, Oct 12th** | Guest lecture: Vikas Garg (Aalto University) |
| **Thu, Oct 19th** | Transformers, DNABert |
| **Thu, Oct 26th** | Transformers, single-cell |
| **Tue, Nov 2nd** | Transfer learning, few-shot learning |
| **Thu, Nov 9th** | Few-shot learning, meta-learning |
| **Tue, Nov 16th** | Domain adaptation |
| **Thu, Nov 23rd** | Unsupervised learning |
| **Tue, Nov 30th** | Self-supervised learning |
| **Thu, Dec 7th** | Project discussions |
| **Thu & Fri, Dec 14th and 15th** | Project presentations |
| **Thu & Fri, Dec 21st and 22nd** | Project presentations |

# Course Exercises Schedule

| Date | Topic | |
|---|---|---|
| **Fri, Sep 22nd** | Pytorch tutorial | |
| **Fri, Sep 29th** | Convolutional neural networks | |
| **Fri, Oct 6th** | Graph neural networks | |
| **Fri, Oct 13th** | Graph neural networks | |
| **Fri, Oct 20th** | Transformers, DNABert | |
| **Fri, Oct 27th** | Transformers, single-cell | |
| **Fri, Nov 3rd** | Transfer learning, few-shot learning | |
| **Fri, Nov 10th** | Few-shot learning, meta-learning | |
| **Fri, Nov 17th** | Project pitch | **Everyone needs to attend!** |
| **Fri, Nov 24th** | Domain adaptation, Unsupervised learning | |
| **Fri, Nov 31st** | Self-supervised learning | |
| **Fri, Dec 8th** | Homeworks debriefing | **Everyone needs to attend!** |
| **Thu & Fri, Dec 14th and 15th** | Project presentations | |
| **Thu & Fri, Dec 21st and 22nd** | Project presentations | |

# Why this course?

# Revolution of AI




# Revolution of AI

"A photo of the astronaut riding a horse"

"Teddy bears mixing sparkling chemicals as mad scientists"

# Revolution of AI

"An expressive oil painting of a basketball player dunking, depicted as an explosion of a nebula."

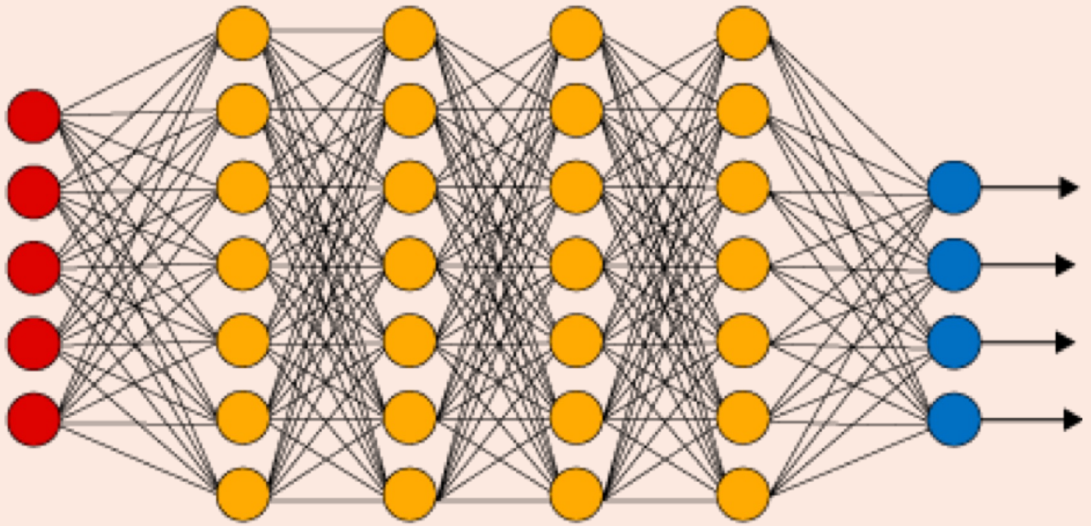
DALL·E 2

DALL·E 3

AI

Machine learning

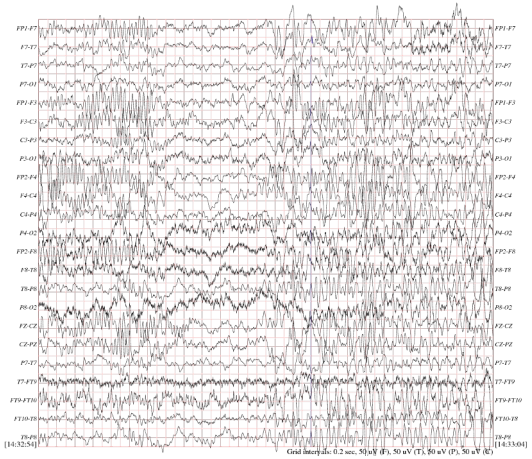
Deep learning

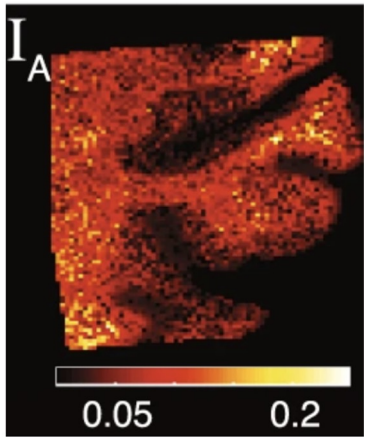


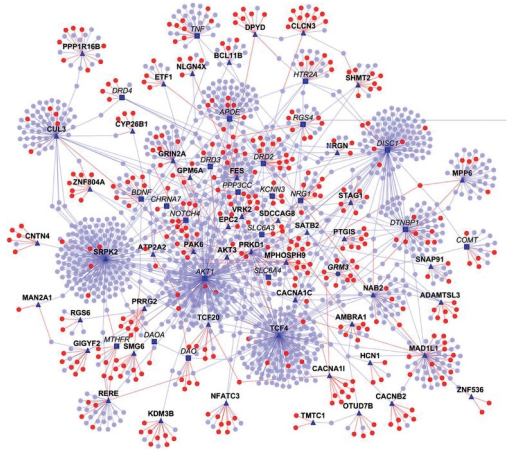
# What makes biomedicine different?

# What makes biomedicine different?

- We often need methods that can make inferences and deliver insights that **humans can not**
- Biomedical data is **often uninterpretable** by humans
  - Genome data
  - EEG data
  - Multidimensional MRI data
  - Complex graph-data (e.g., protein interactions)



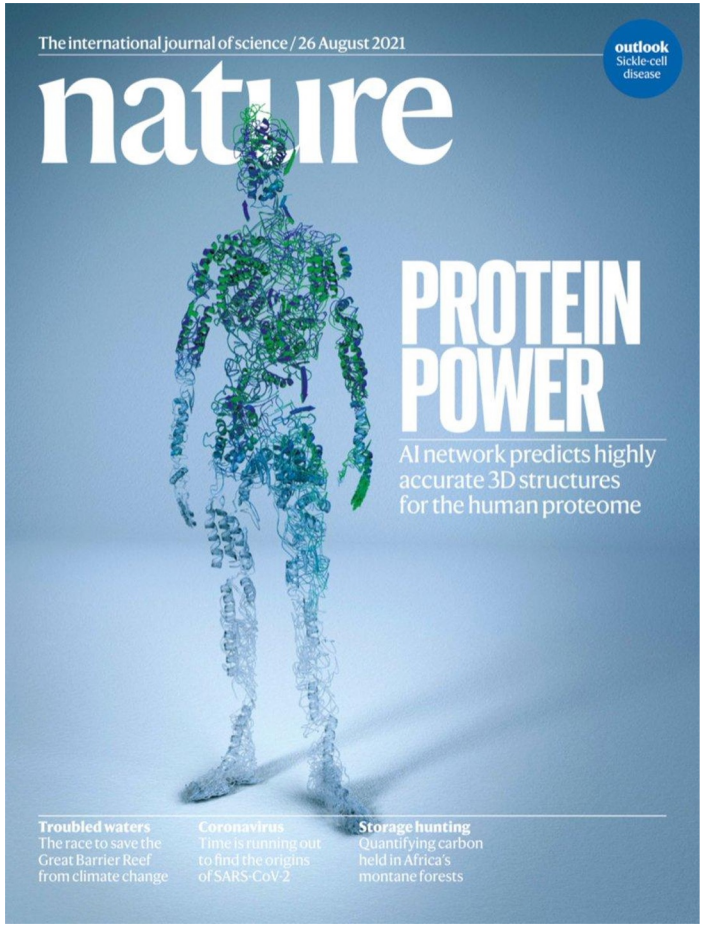


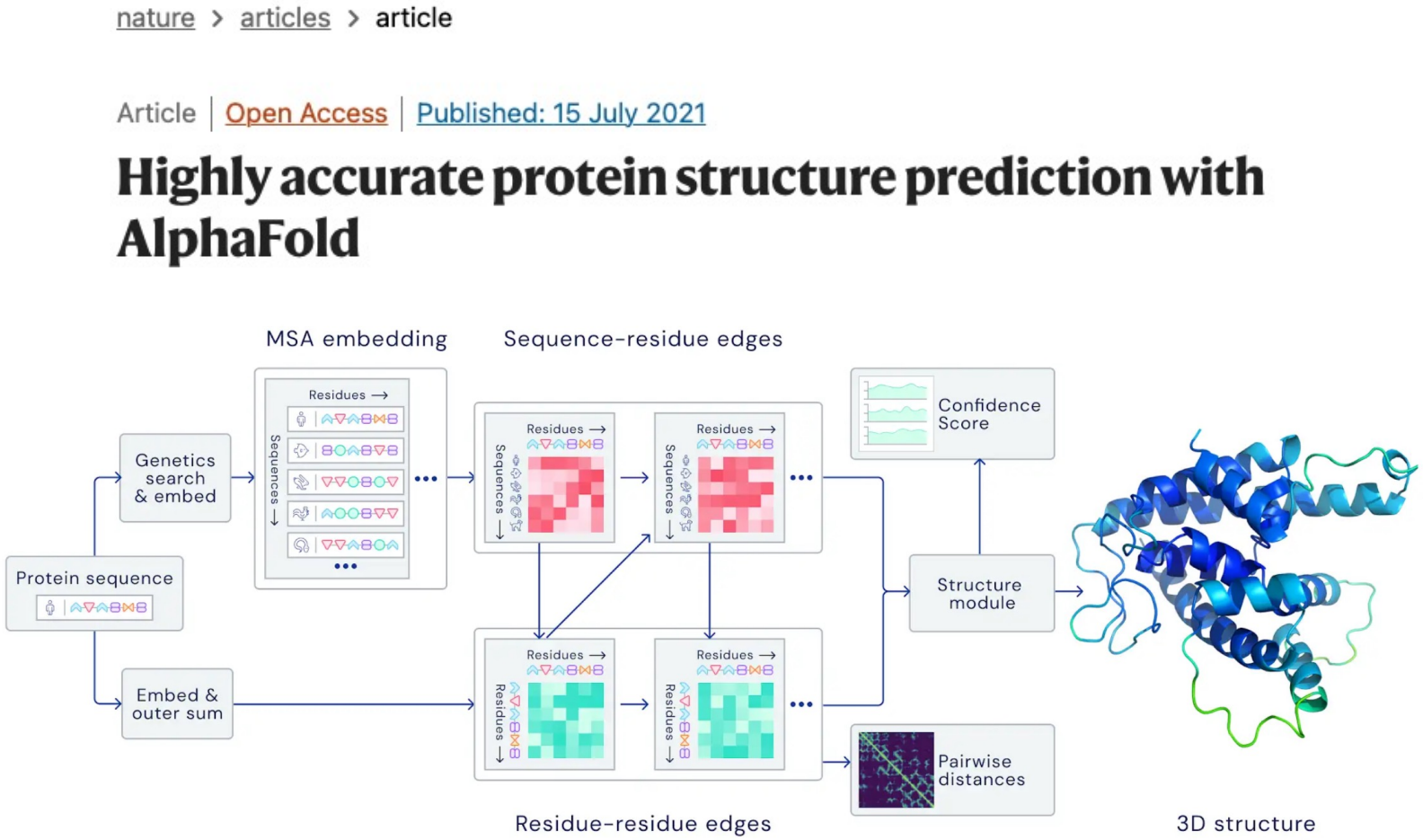



# AI in Biomedicine: Success stories



# AI in Biomedicine: Success stories



# What are challenges of applying AI methods in biomedicine?

# Biomedicine: Challenges

Deep learning success: Availability of large-scale labeled training data

**Biomedicine: Labeled datasets are very difficult to obtain**

- Expert knowledge required to annotate datasets
- Incomplete understanding of complex biological mechanisms
- Not enough samples for rare cases

# Biomedicine: Challenges

**Biomedicine: Labeled datasets are very difficult to obtain**

**Example:** How to label sufficient number of patients for **each possible** disease state?

# Biomedicine: Challenges

Deep learning success: Black-box models

**Biomedicine: We often need explainable (white-box) models**

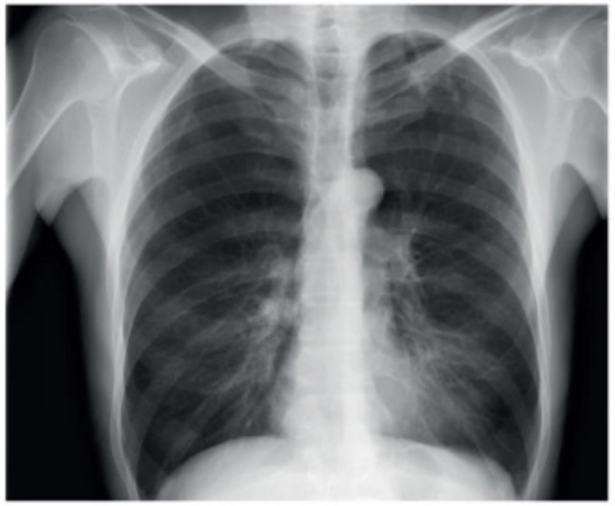
- Why are predictions made? By what evidence?
- Model's predictions understandable to humans
- Trust in the model's predictions is of crucial importance in biomedical applications

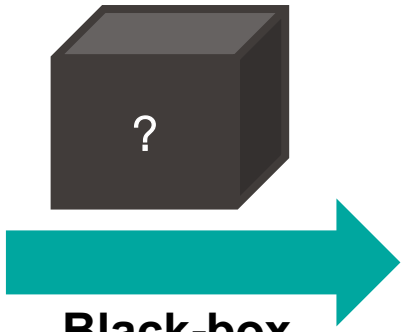
# Biomedicine: Challenges

**Biomedicine: We often need explainable (white-box) models**

## Example:

**Input**
Chest x-ray image






**Black-box ML model**

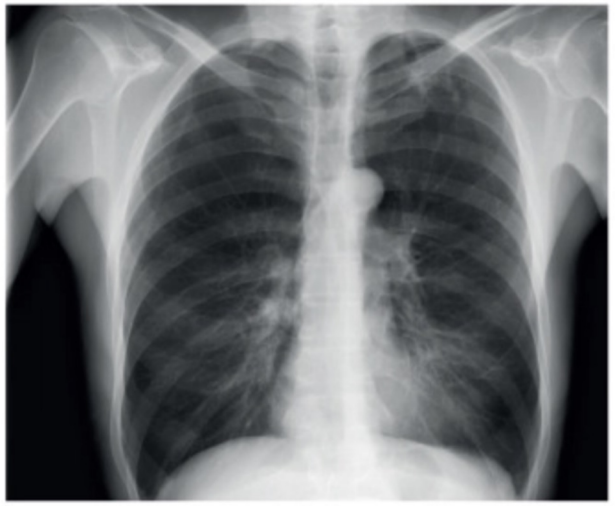
**Output**
Pneumonia positive (85%)

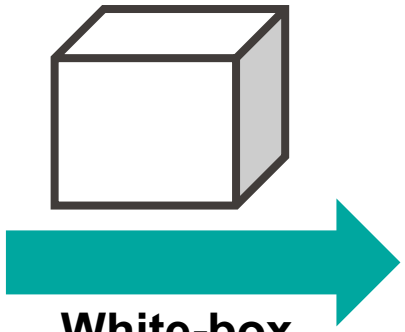
# Biomedicine: Challenges

**Biomedicine: We often need explainable (white-box) models**

## Example:

**Input**
Chest x-ray image





**White-box ML model**

**Output**
Pneumonia positive (85%)

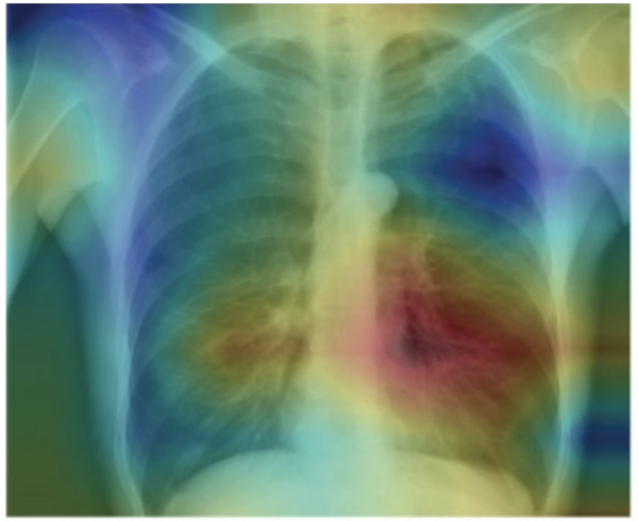


Figure from Ghassemi et al. *Lancet* '21

# Biomedicine: Challenges

Deep learning success: Train and test dataset are generated from the same source

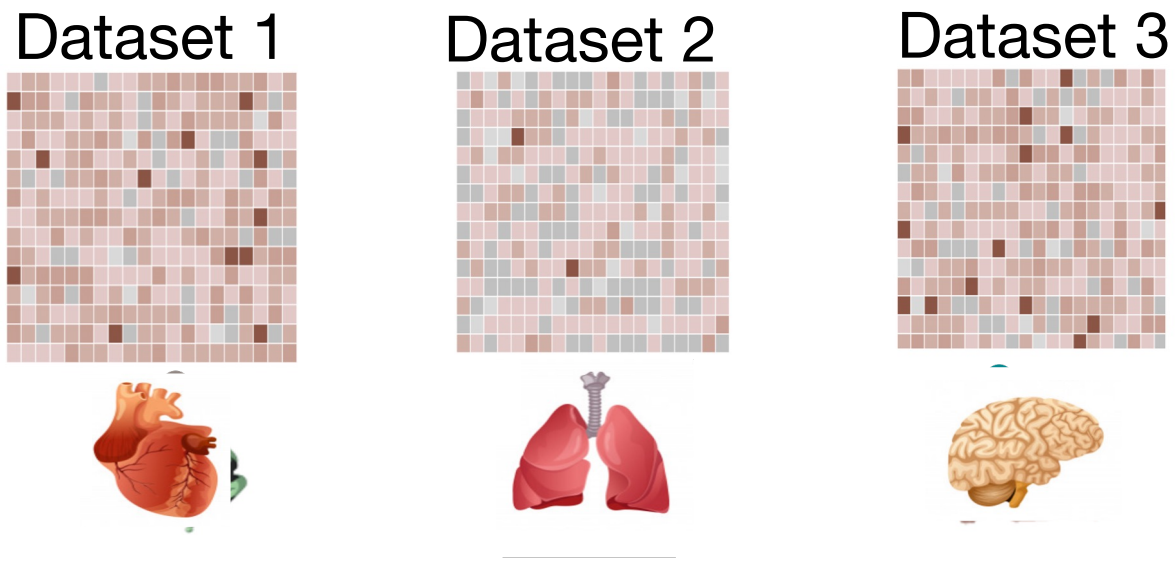
**Biomedicine: Train and test dataset are often generated from heterogenous experimental conditions**

- We need models that can generalize across different experimental conditions

# Biomedicine: Challenges

**Biomedicine:** Train and test dataset are often generated from heterogenous experimental conditions

**Example:** How to transfer annotations **across tissues, donors, species**?



# Biomedicine: Many Other Challenges ...

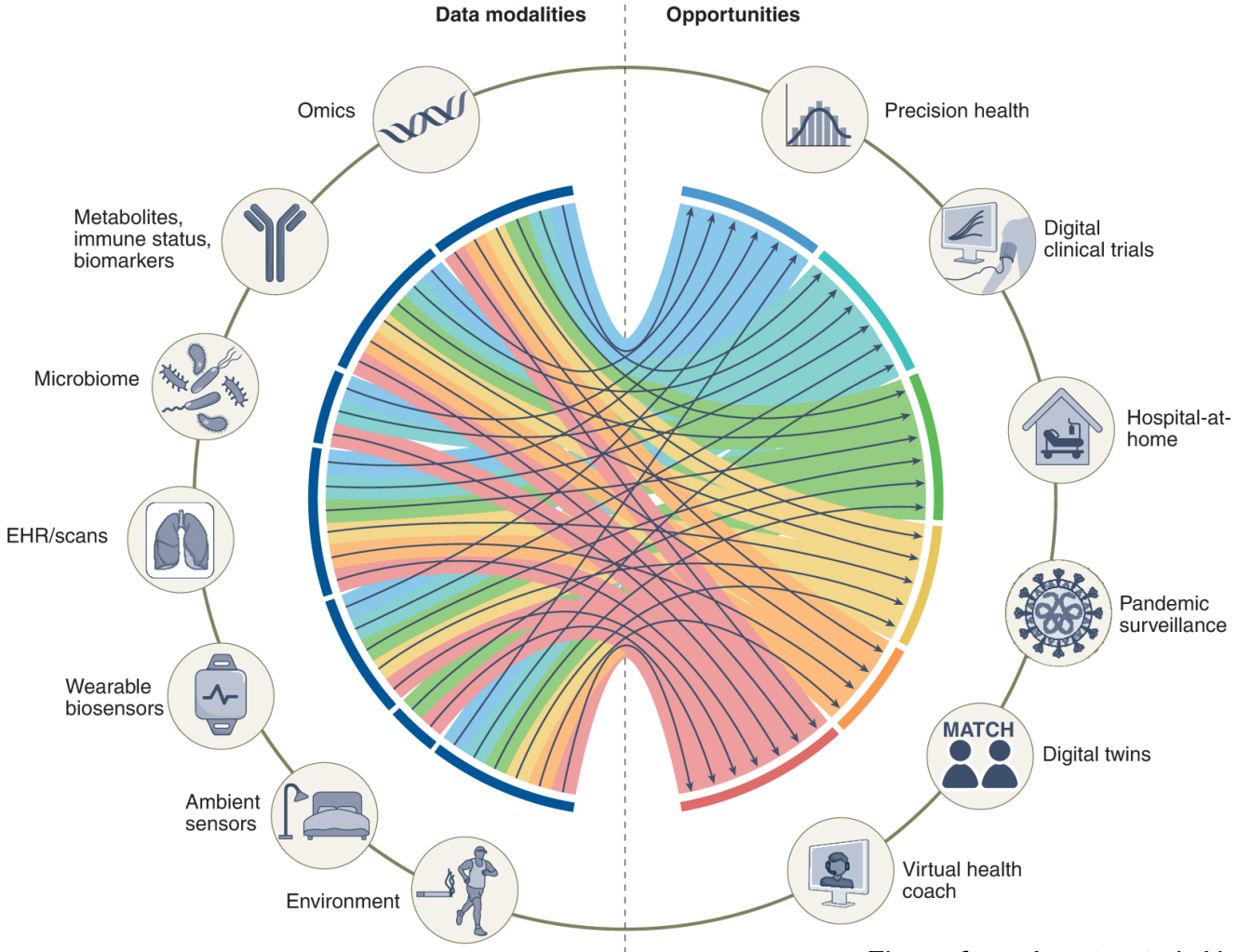



Figure from Acosta et al. *Nature* '22

# This course: Deep learning approaches for addressing biomedical challenges

Graph neural networks

Transformers

Convolutional neural networks

Few-shot learning

Domain adaptation

Meta-learning

Self-supervised learning

# This course: Deep learning approaches for addressing biomedical challenges

Drug response prediction

Drug side effect prediction

Cell type annotation

Medical images classification

Datasets integration