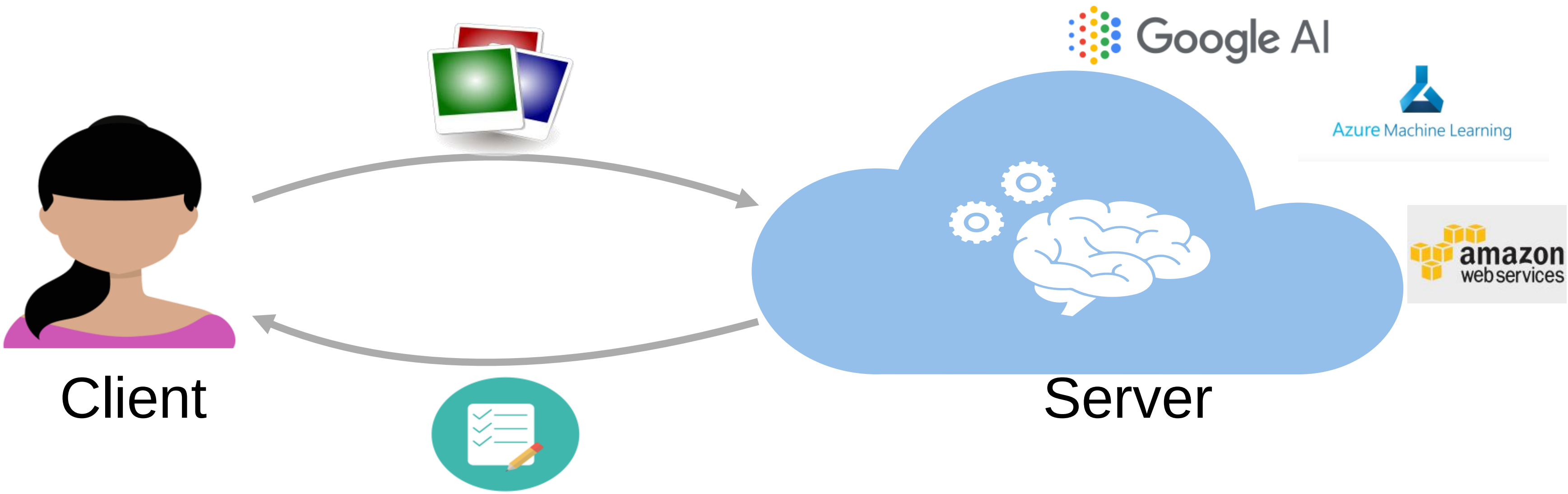
Privacy-preserving federated learning for healthcare

Melek Önen

January 2022, AI4Health

Machine Learning as a Service

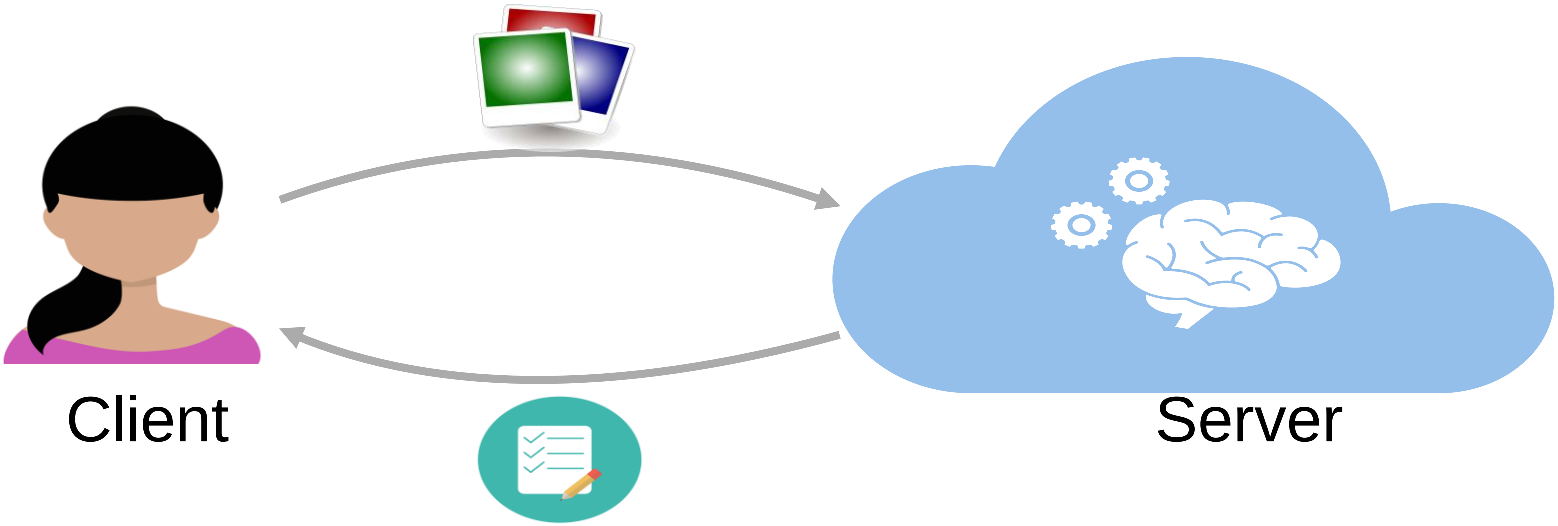


Performance

No need for ML knowledge

Cost reduction

Sensitive and confidential data



- Sensitive personal data
- Corporate data (IP)

- Intellectual property

- Legal restrictions

Data Protection - Tools

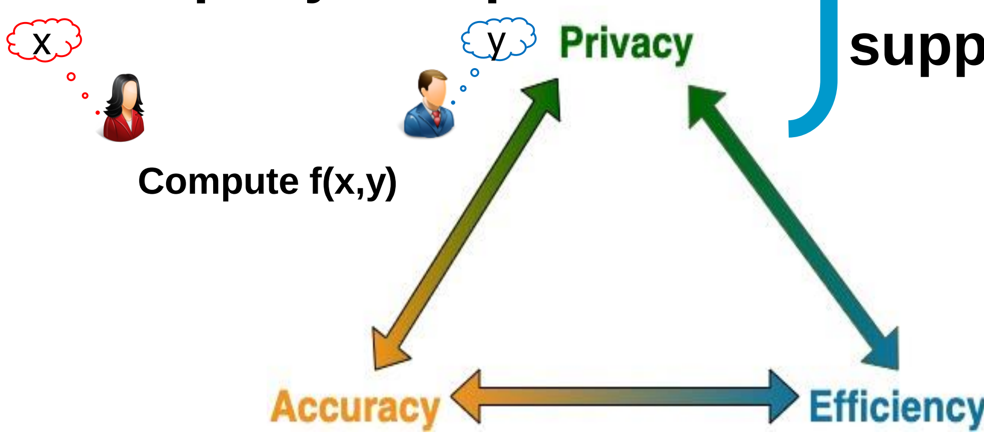
Traditional crypto → **not adapted**

Advanced crypto

Homomorphic Encryption

$$Enc(m_1) \blacklozenge Enc(m_2) = Enc(m_1 \spadesuit m_2)$$

Secure Multiparty Computation



- Additional overhead

- Only, some operations are supported

Homomorphic encryption

$$Encrypt(m_1)\ op1\ Encrypt(m_2) = Encrypt(m_1 op2\ m_2)$$

Partially HE

Support one operation only

Somewhat HE

Support arbitrary + and limited number of x

Fully HE

Support any function

El Gamal Cryptosystem

- $KeyGen(1^n)$
  - $SK: x$
  - $PK: \{p, g, g^x mod\ p\}$
- $Encryption(g^x, m)$
  - Choose $r \leftarrow Z_N^*$
  - Output $c = (c_1, c_2) = (g^r, m(g^x)^r)$
- $Decryption(x, c)$
  - $Compute\ c_1^x$
  - Output $m = \frac{c_2}{c_1^x}$

El Gamal Homomorphism

- $c = (c_1, c_2) = (g^r, m(g^x)^r),\ c' = (c_1', c_2') = (g^{r\prime}, m'(g^x)^{r\prime})$

- $(c_1 c_1', c_2 c_2{}') = (g^{r+r'}, m.m' g^{(r+r')x}) = Encrption(g^x, mm')$

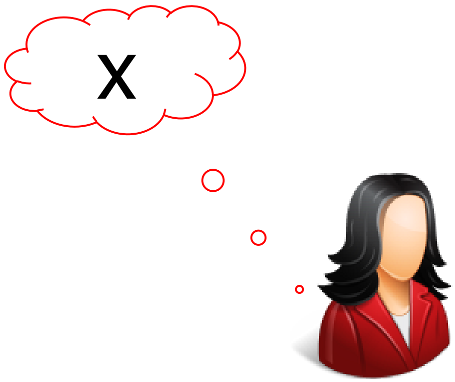
EURECOM

Sophia Antipolis

Homomorphic encryption

- **Partially homomorphic encryption**
  - Practical
  - Used for electronic voting, simple statistical operations
- **Somewhat homomorphic encryption**
  - Useful for low-degree polynomials
- **Fully Homomorphic Encryption**
  - Still not very practical
  - Ongoing research

Secure Two-party computation



Compute f(x,y)

leak no other information than what Ideal model leaks

Yao's GC

Arithmetic sharing

Boolean sharing

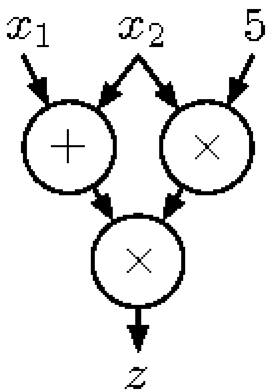
Yao's Secure Two-Party Computation

- **First secure two-party computation**

- **Problem**
  - Inputs: A: x; B: y;
  - Output: f(x,y)

- **Idea**
  - Represent f as a Boolean Circuit
  - Transform this circuit to a garbled circuit that hides all information but the output

2PC with Arithmetic circuits



$$z = (x_1 + x_2) \times (x_2 \times 5)$$

- **Arithmetic Circuit**
  - Addition gates
  - Multiplication gates

Secret sharing for addition gates [Ben-or et al.'88]

- **Goal: compute** $\mathrm{z} = x_A + x_B$
  - Secret share x_A
    - $x_A = x_{A1} + x_{A2}$
  - Secret share x_B
    - $x_B = x_{B1} + x_{B2}$
  - Send shares to two parties
    - P1:$x_{A1},\ x_{B1}$; P2:$x_{A2},\ x_{B2}$;
  - Compute sum
    - P1: $res1 = x_{A1} + x_{B1}$; P2: res2 = $x_{A2} + x_{B2}$;

  $\Rightarrow x_A + x_B = \mathrm{res1} + \mathrm{res2}$

$\Rightarrow$ **No communication needed!**

How to multiply?

- **Goal compute** $z = x_A . x_B$
  - Secret share x_A
    - $x_A = x_{A1} + x_{A2}$
  - Secret share x_B
    - $x_B = x_{B1} + x_{B2}$
  - Send shares to two parties
    - P1: $x_{A1},\ x_{B1}$; P2: $x_{A2},\ x_{B2}$;
  - Compute multiplication
    - P1: $res1 = x_{A1}.x_{B1}$; P2: res2 = $x_{A2}.x_{B2}$;

$\Rightarrow x_A . x_B \neq \mathrm{res1}.\mathrm{res2}$

$\Rightarrow x_A . x_B = (x_{A1} + x_{A2})(x_{B1} + x_{B2})$

How to compute this?

$$= x_{A1}x_{B1} + x_{A2}x_{B2} + x_{A1}x_{B2} + x_{A2}x_{B1}$$

Beaver triplets

- **Let $c = ab$ and let secret share and obtain** $[a], [b], [c]$
- **Secret share x_A and x_B**
- **Compute**
  - Construct $e = x_A + a;\ d = x_B + b$; from [a], [b], $[x_A]$, $[x_B]$
  - $[z] = [x_A . x_B] = [c] + e[x_B] + d[x_A] - ed$

$\Rightarrow$ **Pre-processing and storage needed**

$\Rightarrow$ **1-round communication**

2 Party Computation - Summary

- **2 different methods**
  - Yao's GC
  - 2-PC through arithmetic sharing
- **Assumption**
  - Semi-honest parties
- **Open questions**
  - From 2PC to multi-party computation
  - From semi-honest to malicious adversaries

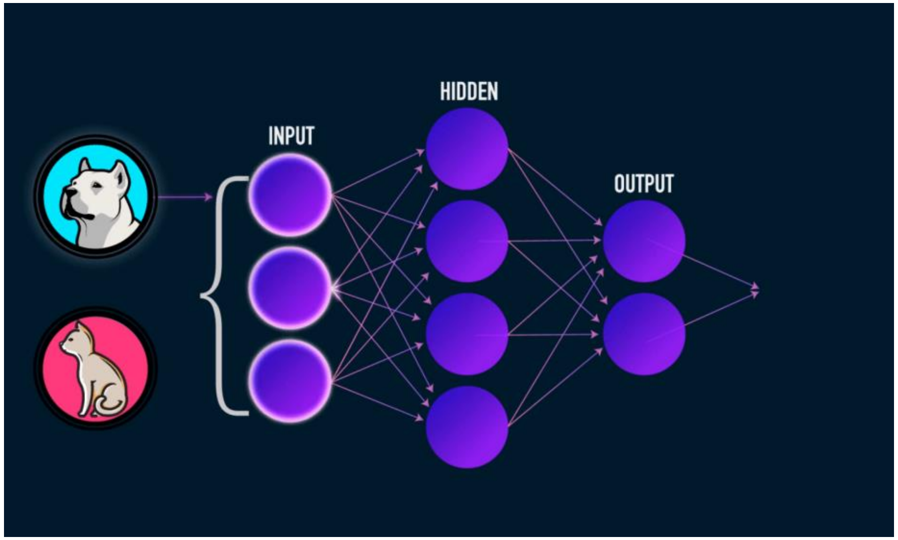
HE vs. 2PC

HE
Non-interactive
Only linear operations
Expensive in computation cost
No communication cost

2PC
Interactive - Client is involved
Linear and nonlinear operations
Efficient in computation cost
Expensive in communication cost

Artificial Neural Networks

Supervised machine learning technique



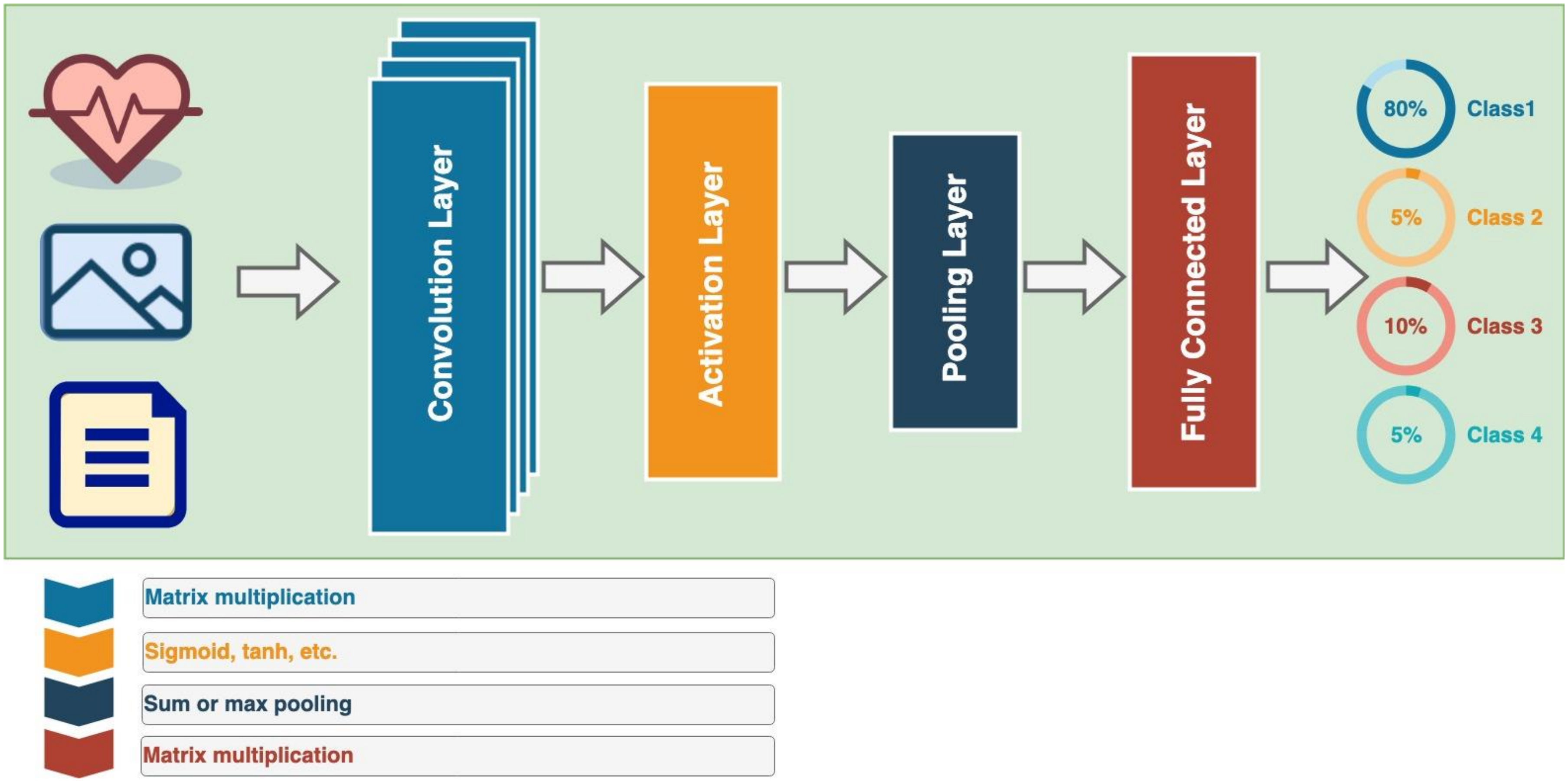
Two phases:

- Training
- Classification

NN layers:

- Activation layer
- Pooling layer
- Fully-connected layer
- Convolution layer (optional)

Neural Networks - Architecture



Privacy preserving NN Classification

Use Advanced cryptographic techniques

- Homomorphic encryption, Secure 2PC

Challenge : Privacy vs. Performance

- Additional overhead (Computation, memory & bandwidth)
- Complex operations (sigmoid, tanh, etc.)
- Real numbers (vs. integers with PETs)

Goal

- Reduce NN complexity
- Approximate complex operations $\Rightarrow$ Use low degree polynomials
- Approximate real numbers $\Rightarrow$ Use integers

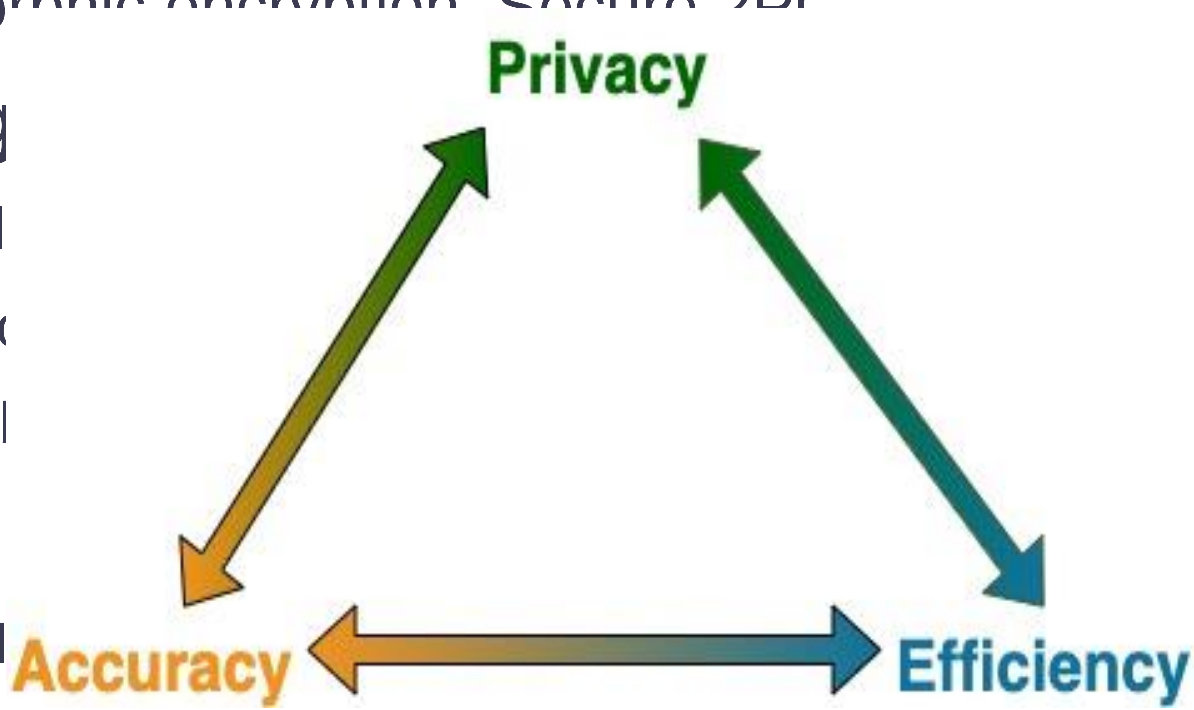
Privacy preserving NN Classification

Use Advanced cryptographic techniques

- Homomorphic encryption, Secure 2PC

Challeng

- Additional ... n)
- Complex ...
- Real numl



Goal

- Reduce N
- Approximate complex operations $\Rightarrow$ Use low degree polynomials
- Approximate real numbers $\Rightarrow$ Use integers

Approximation of NN layers

- Convolution layer
  - Matrix multiplications $\Rightarrow$ No need for approximation
- Activation layer
  - Most common approach: $\boldsymbol{x^2}$ and ReLU
- Pooling layer
  - Sum or average
- Fully Connected layer
  - Matrix multiplications $\Rightarrow$ No need for approximation
- Real numbers
  - Most common approach: Multiplying with 10^n

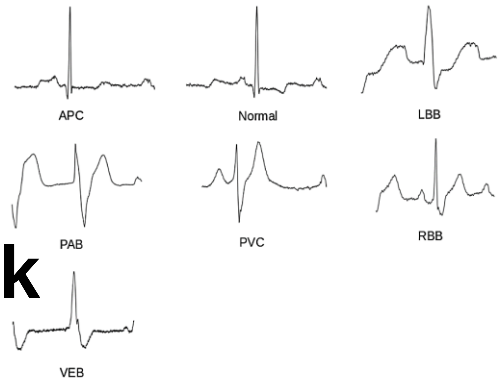
PAC: Pp Arrhythmia Classification

[FPS 2019]

NN based ECG analysis

2PC based NN classifier

- Low degree polynomials for activation functions
- Approximation of real numbers
- PCA for size reduction



Performance results with PhysioBank

- 96.34% accuracy
- 1 sec prediction time in real environment

PAC in batches

- Efficient solution for real scenarios

What about federated learning for healthcare?

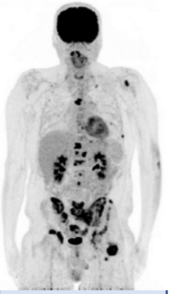
EPIONE Team– Marco Lorenzi
- Development of the **FedBioMed platform**
- Development of the interface, adaptation of the central infrastructure to the project algorithms, installation of the "client" software infrastructure

https://fedbiomed.gitlabpages.inria.fr/

Antoine Lacassagne
CENTRE DE LUTTE CONTRE LE CANCER
unicancer NICE

unicancer

10 Comprehensive Cancer Centers
- Definition of a multi-center AI clinical project
- Structuring and storage of the database in a local client application (harmonization of formats and structuring)
- Hospitals transmit only the parameters of the learning model



EURECOM
Sophia Antipolis

Melek Önen

Security and privacy for emerging technologies such as Big Data or IoT

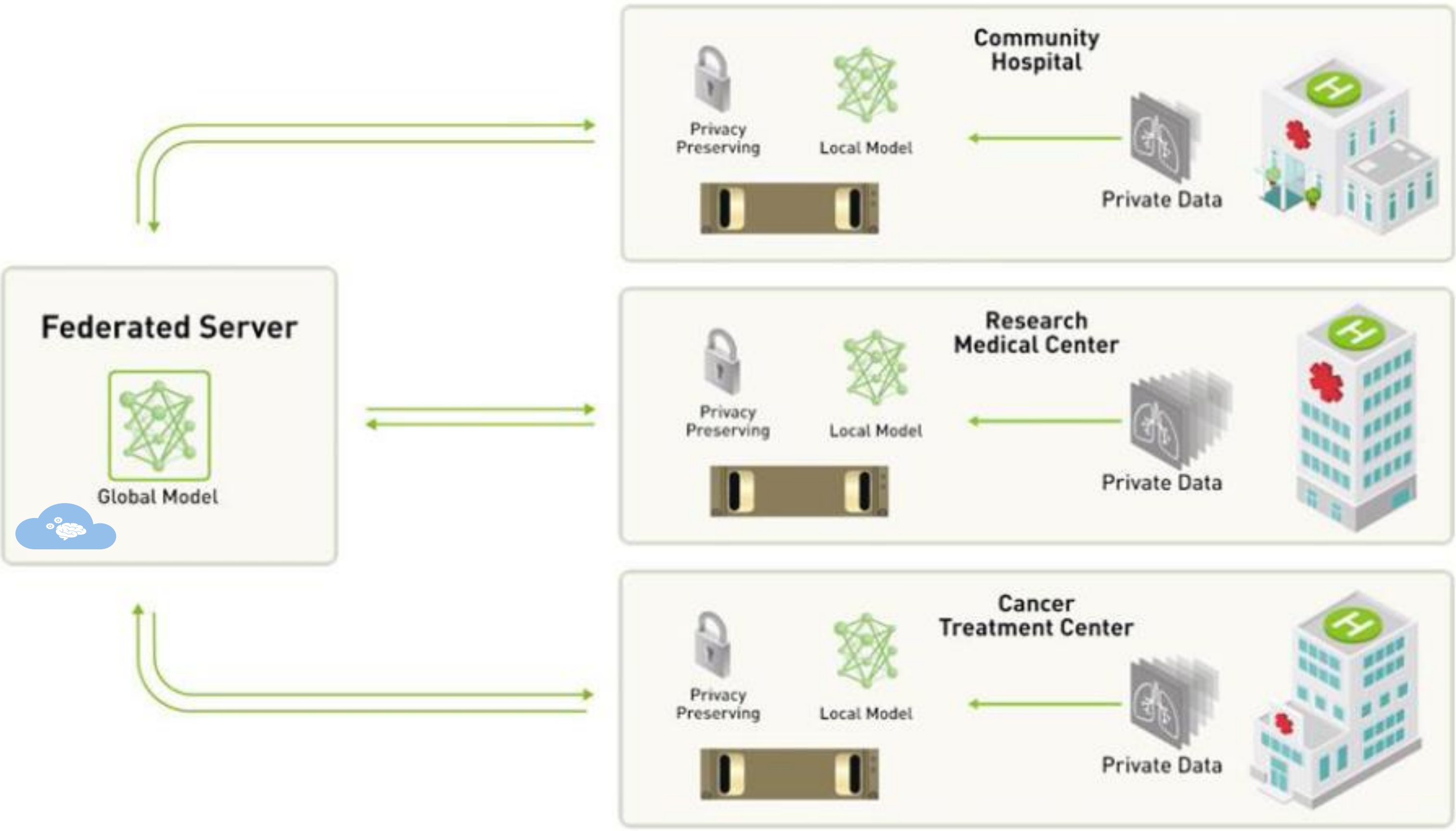
3iA Côte d'Azur
Interdisciplinary Institute for Artificial Intelligence

Objective at the end of the project
a functional platform, open to the project's partner hospitals, versatile

Picture credit: Olivier Humbert

EURECOM
Sophia Antipolis

What about federated learning for healthcare?



Picture credit: Olivier Humbert

Open questions

- **Multiple parties**
  - Need for a dedicated key management scheme?
  - Need for MPC with multiple parties?
  - Need for a new adversarial model (the more the risky?)?
- **Healthcare data**
  - Usually large datasets $\Rightarrow$ adds even more complexity
  - What about operations
  - More bandwitdh

Thank you!

melek.onen@eurecom.fr

Joint work with: Olivier Humbert, Marco Lorenzi, Riccardo Taiello