Private Multi-Winner Voting for Machine Learning

Adam Dziedzic, Christopher A Choquette Choo,
Natalie Dullerud, Vinith Menon Suriyakumar,
Ali Shahin Shamsabadi, Muhammad Ahmad Kaleem,
Somesh Jha, Nicolas Papernot, Xiao Wang

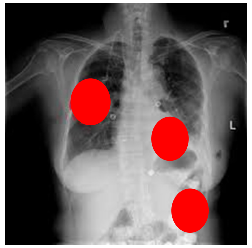
VECTOR INSTITUTE · UNIVERSITY OF TORONTO · Northwestern University · Google · THE UNIVERSITY of WISCONSIN MADISON

Single-Label vs Multi-Label Classification

Single-Label:
7

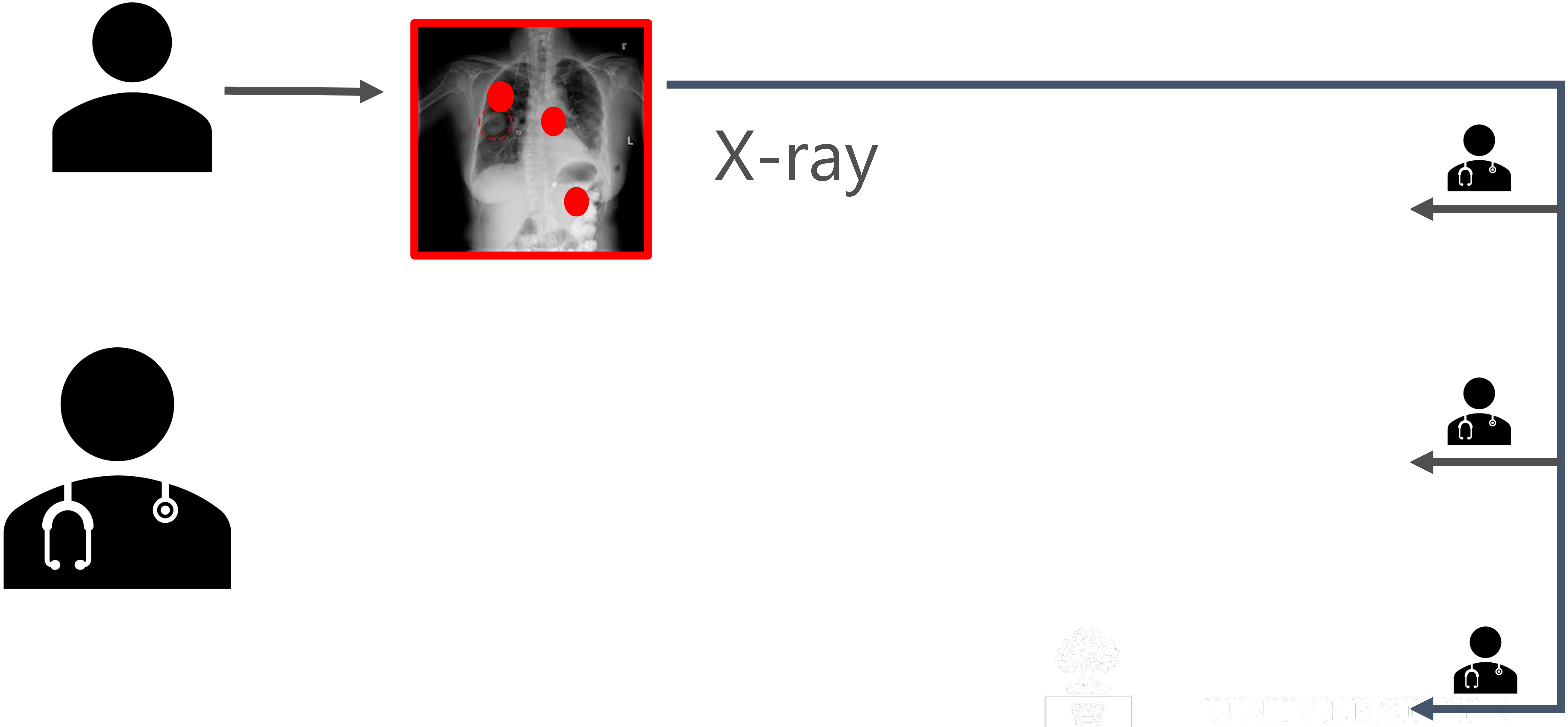
Single-Label vs Multi-Label Classification

Single-Label:
7

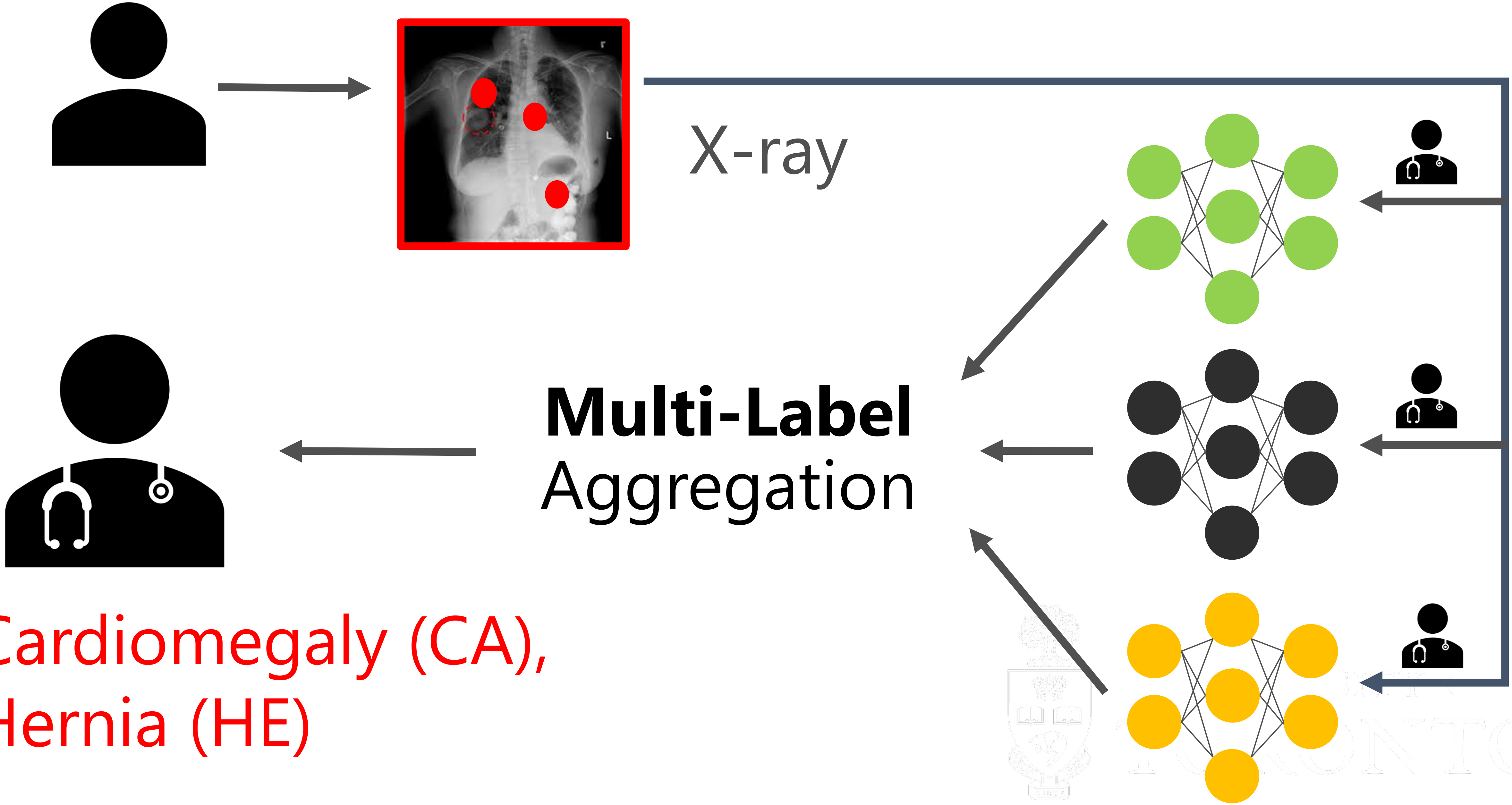


Multi-Label:
Cardiomegaly (CA) - enlarged heart
Edema (ED) - fluid trapped in a tissue
Hernia (HE) - organ bulges out

Multi-Label Classification in Medicine

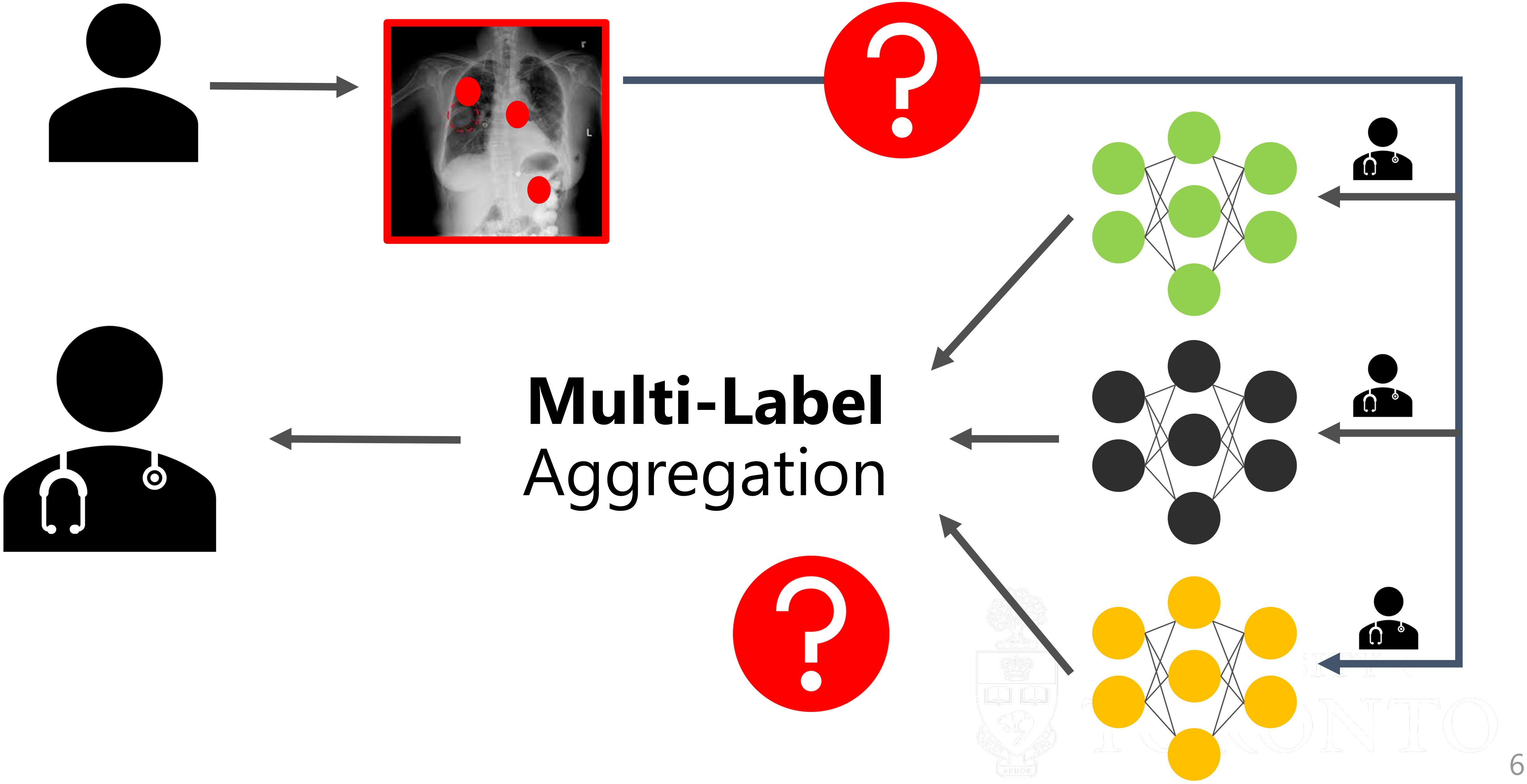


Multi-Label Classification in Medicine

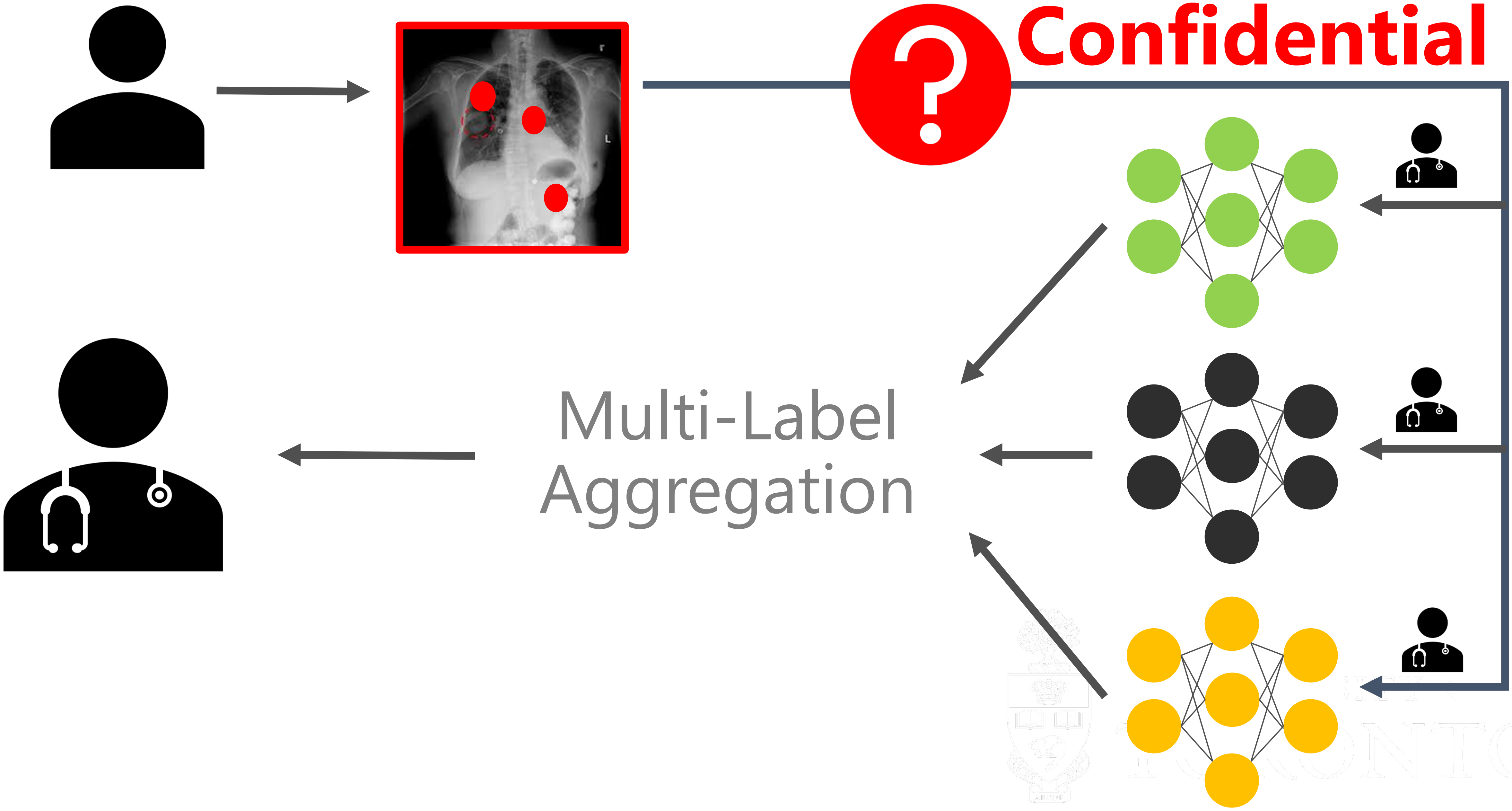


Cardiomegaly (CA),
Hernia (HE)

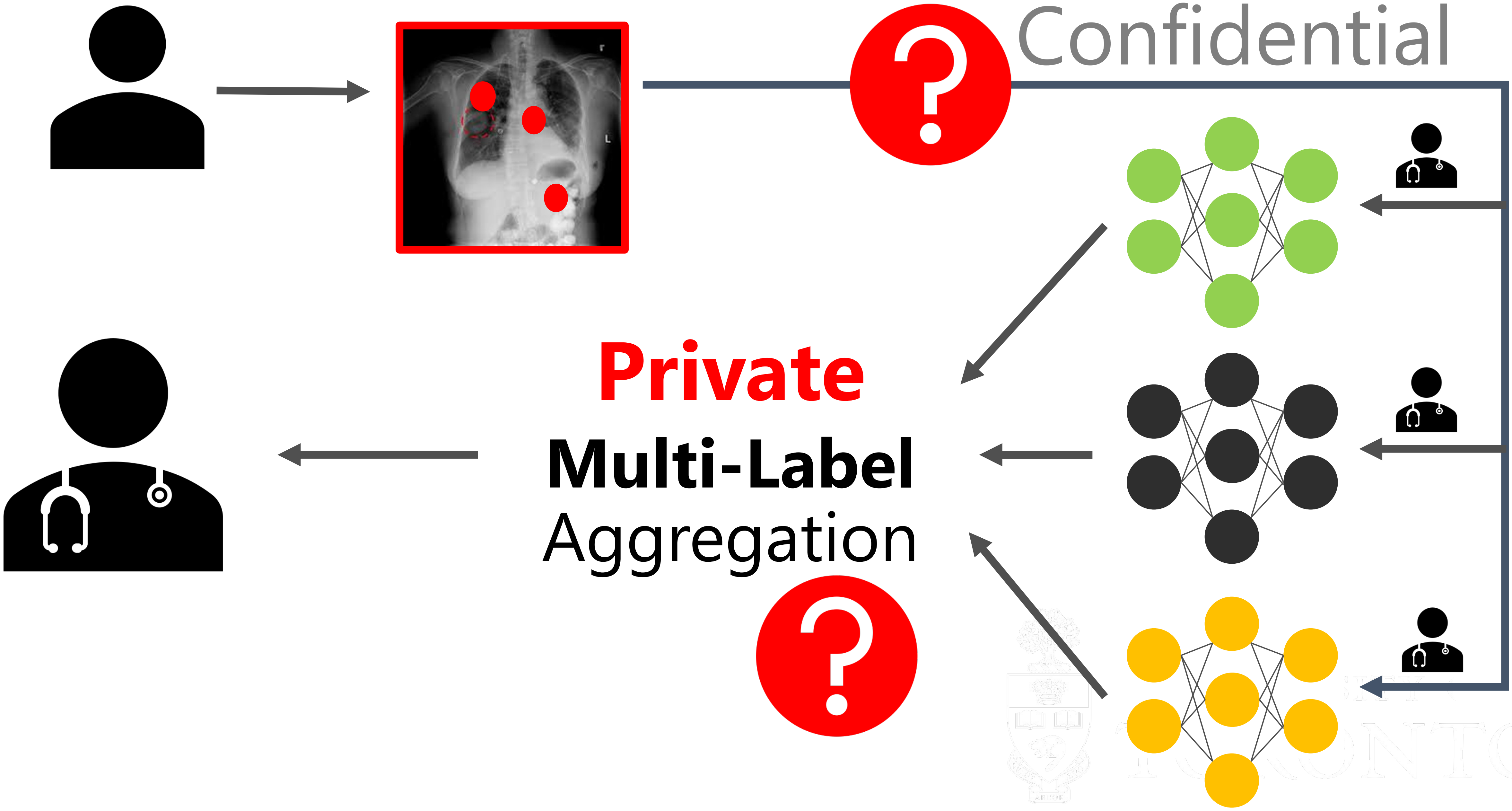
Multi-Label Classification in Medicine



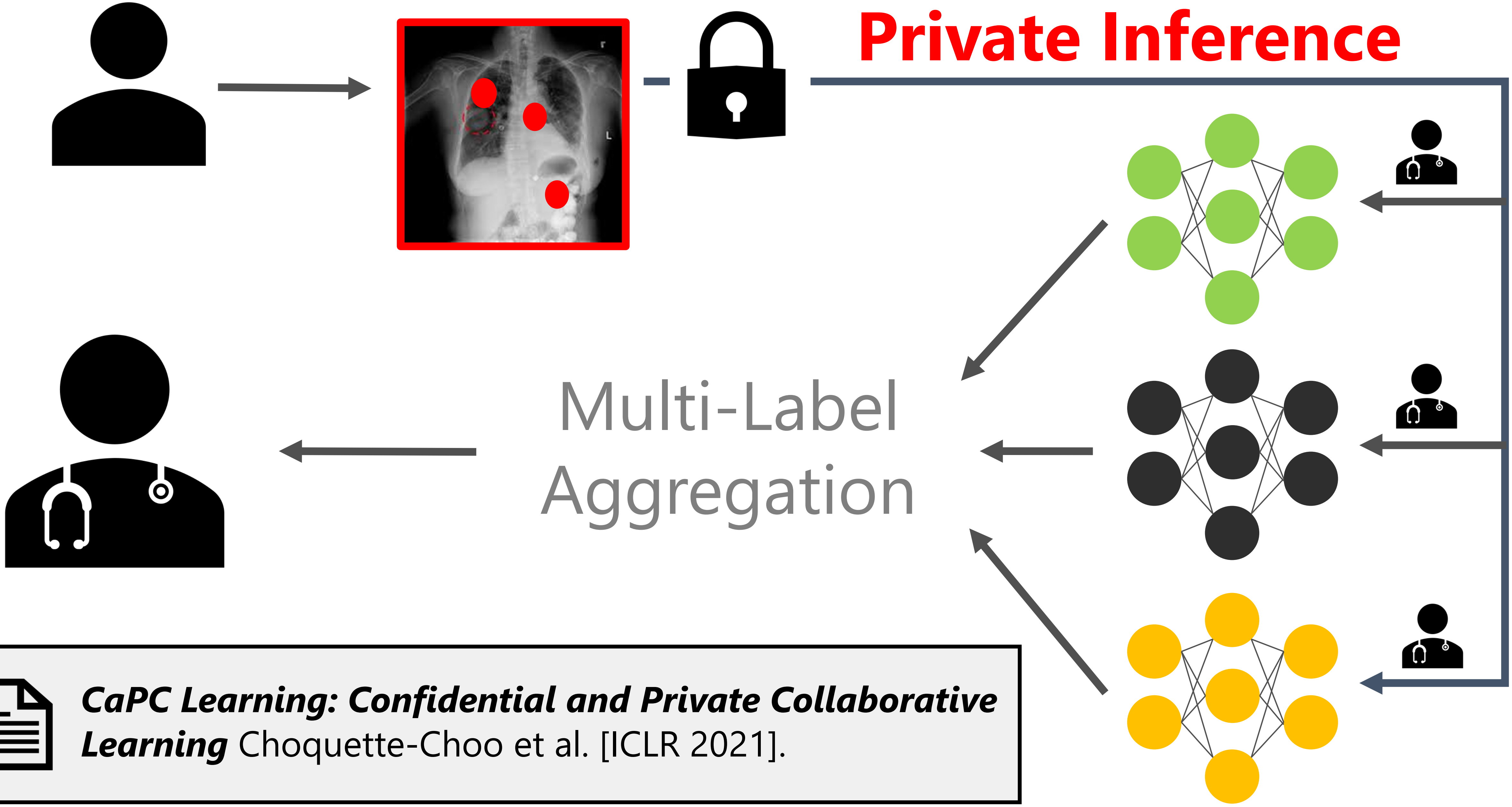
Multi-Label Classification in Medicine



Multi-Label Classification in Medicine

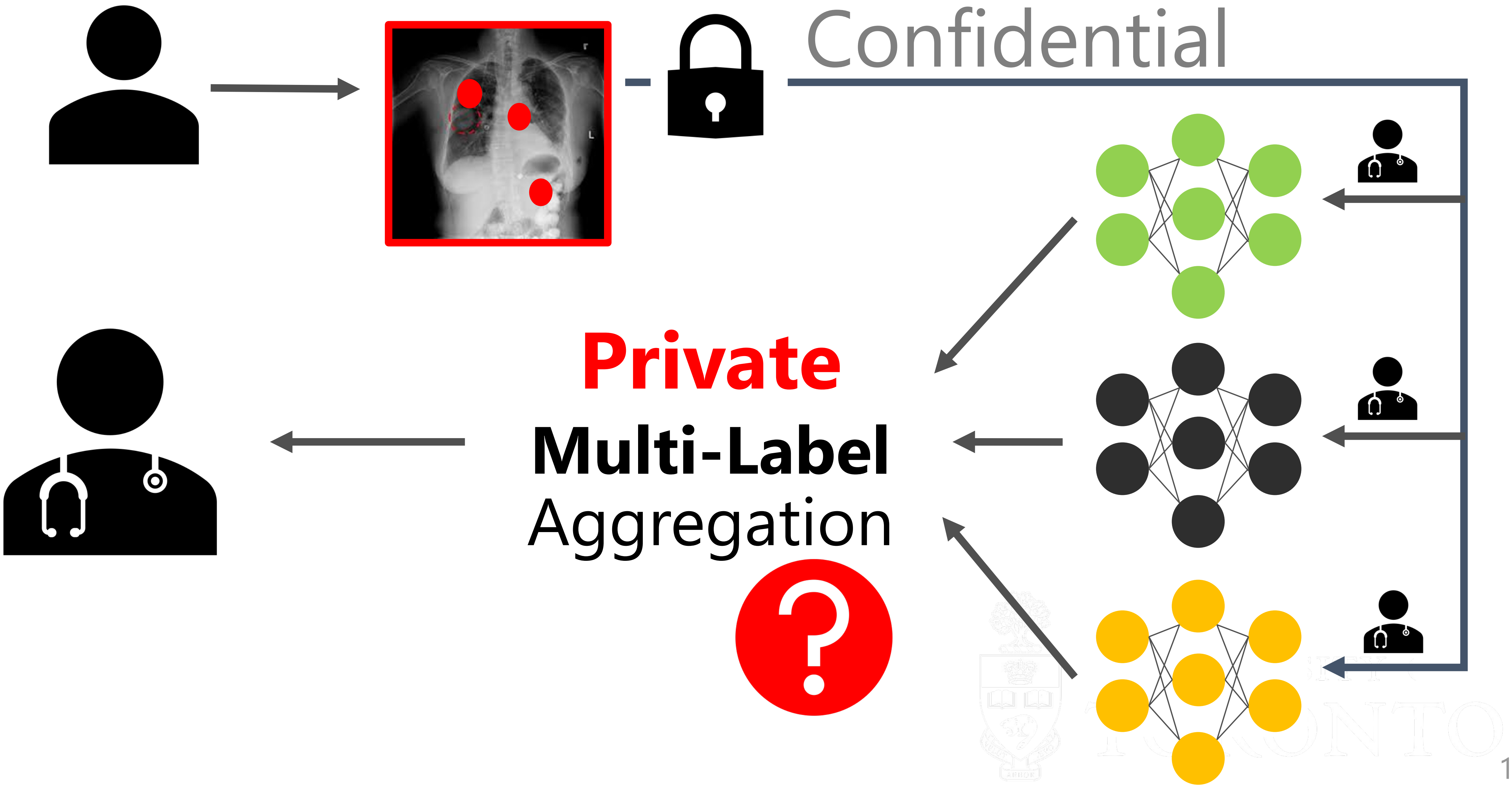


Confidential Collaborative Learning via CaPC

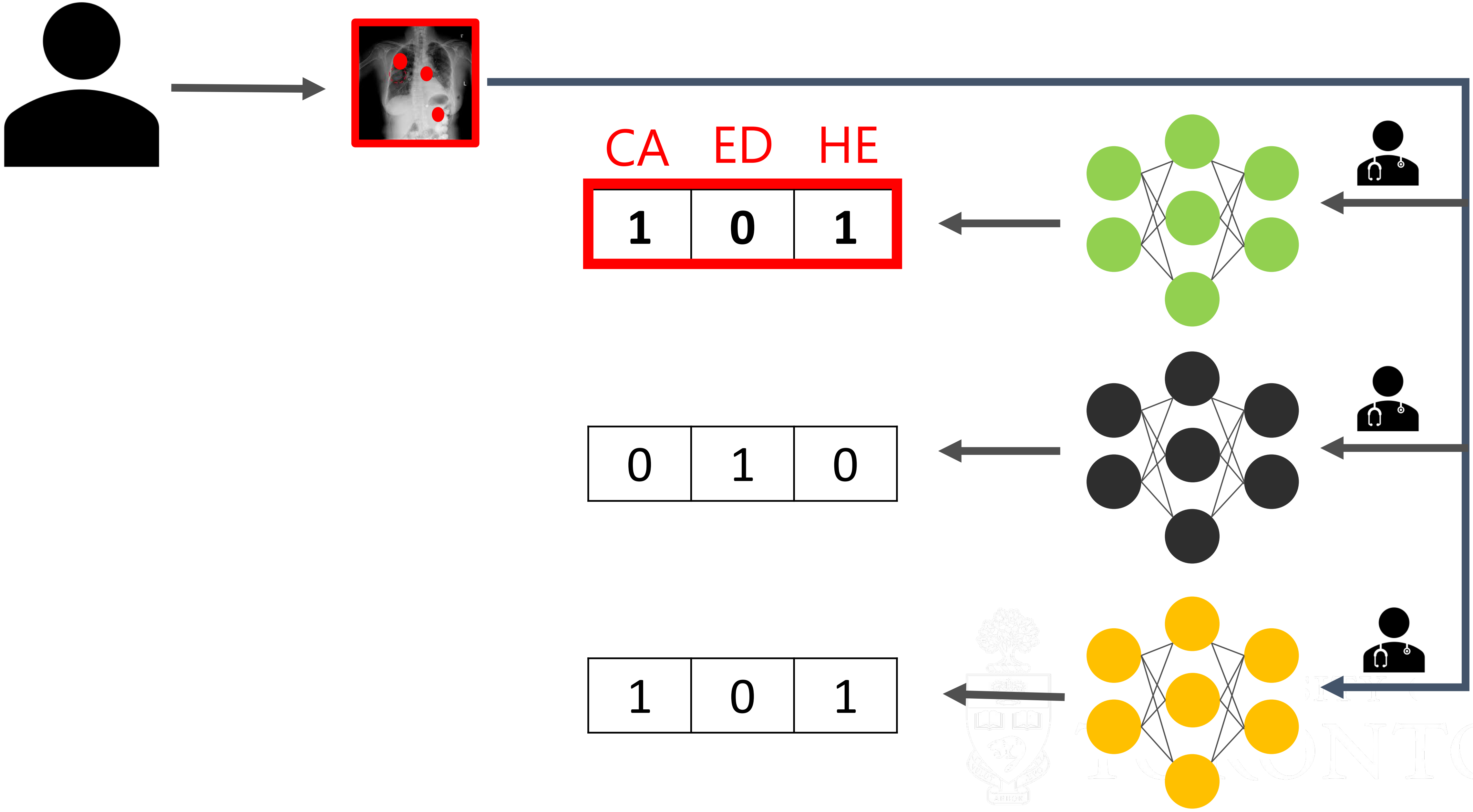


CaPC Learning: Confidential and Private Collaborative Learning Choquette-Choo et al. [ICLR 2021].

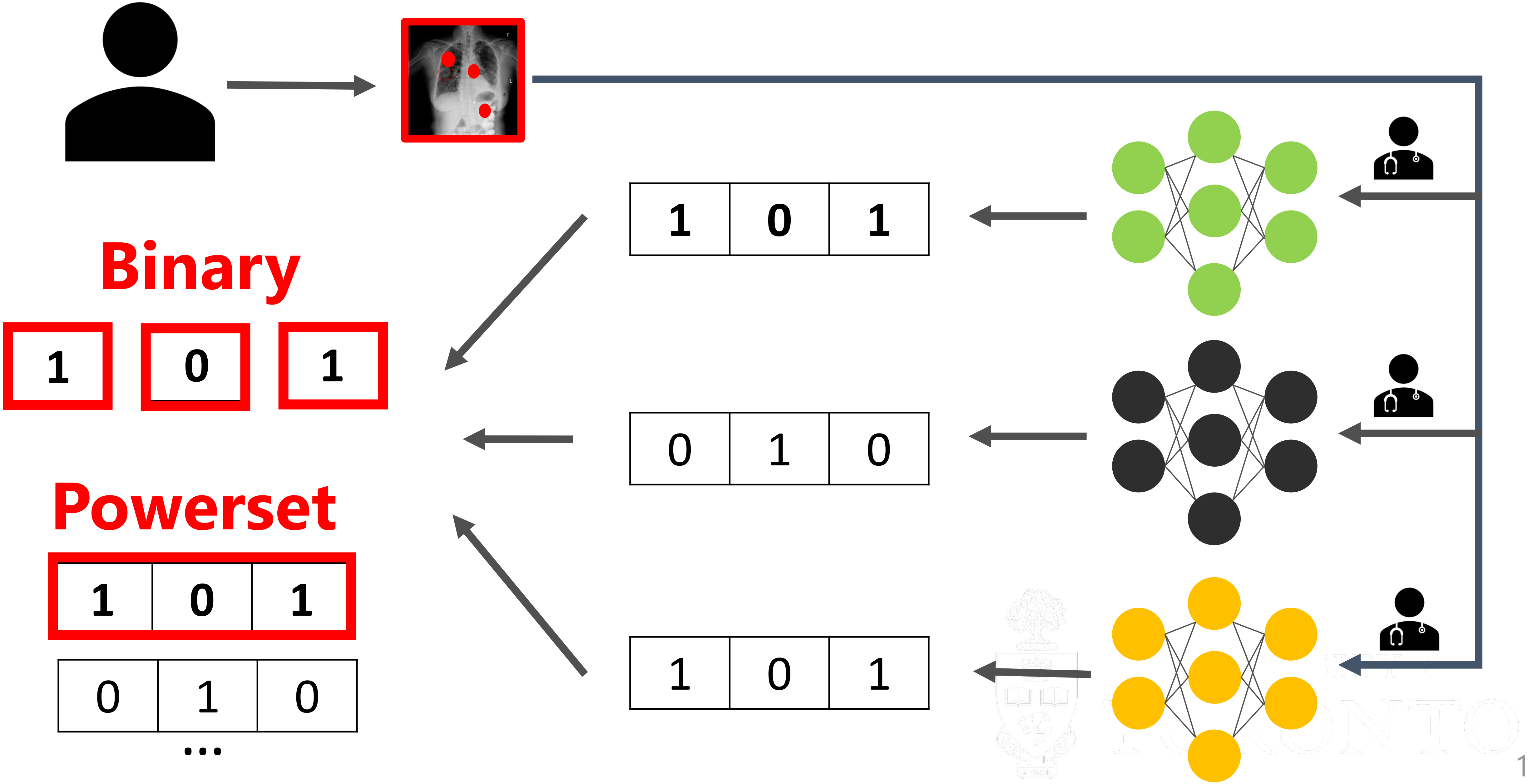
Private Multi-Label Classification in Medicine



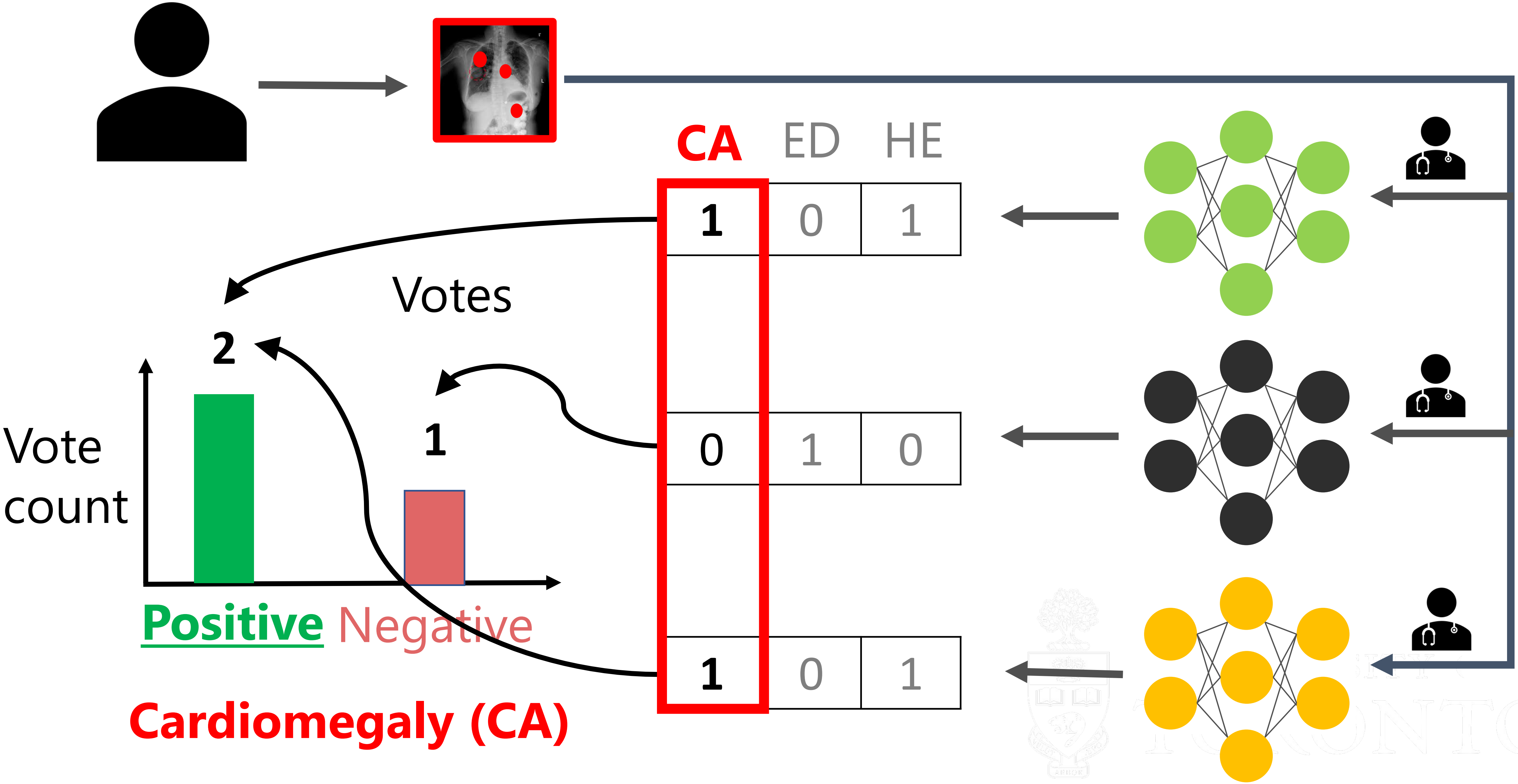
Aggregation for Multi-Label Classification



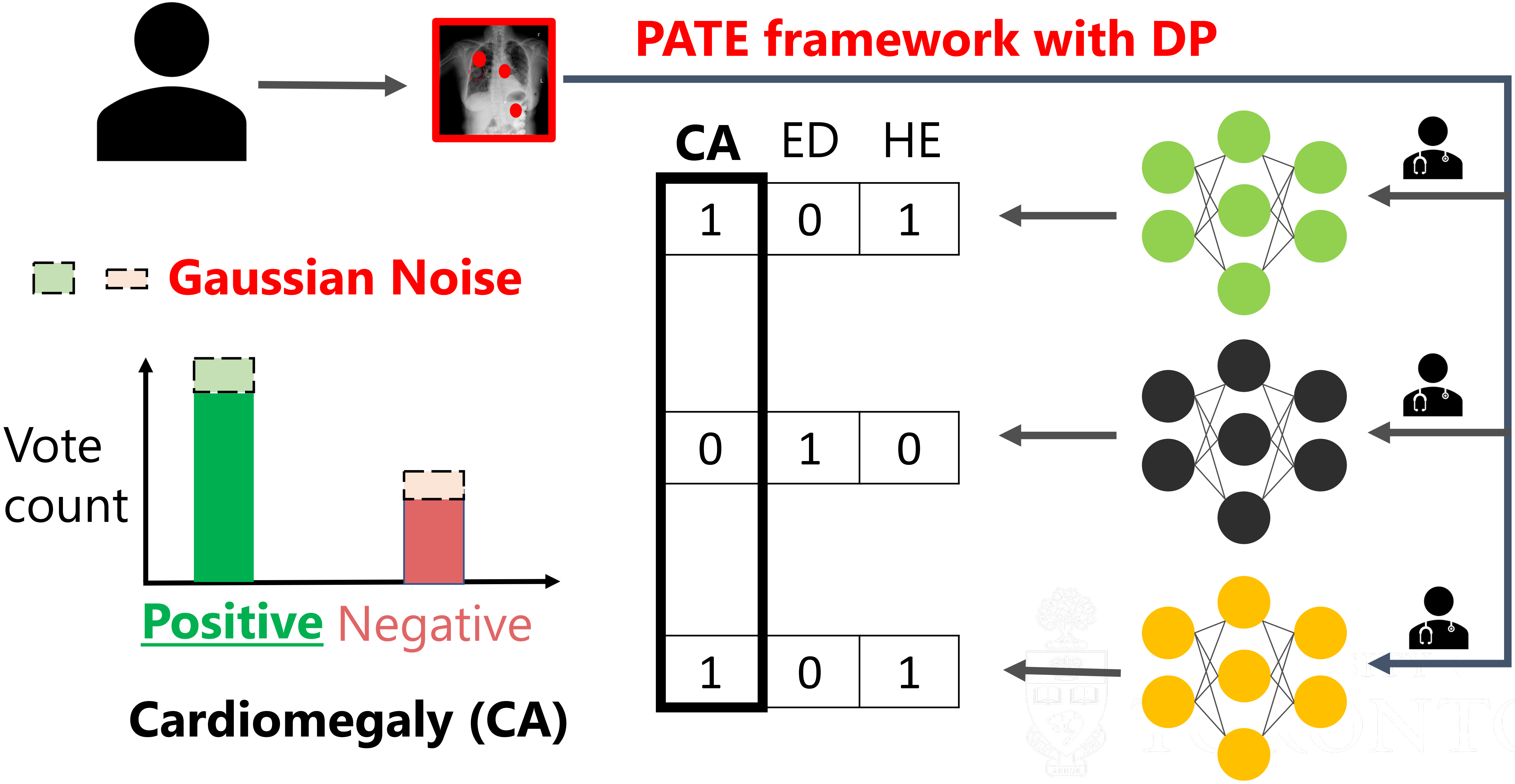
Two Methods for Multi-Label Classification



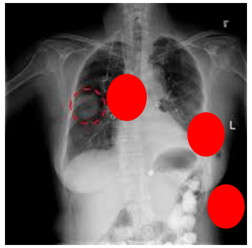
Binary Multi-Label Classification



Private Binary Multi-Label via Noisy ArgMax

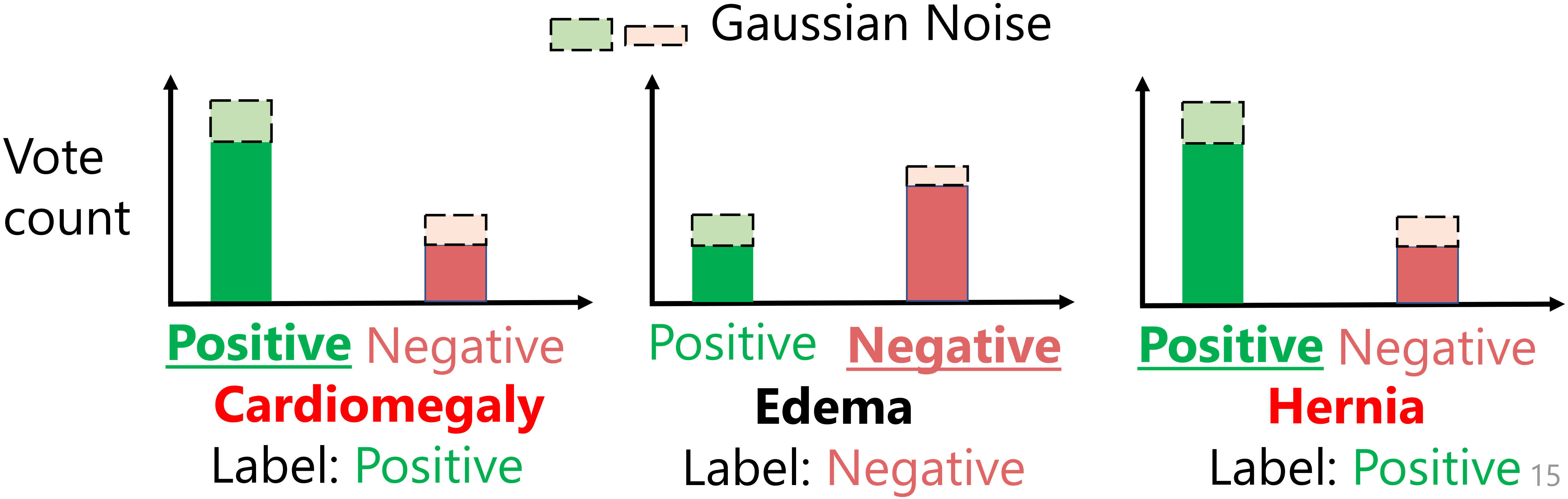


Private Binary Multi-Label Classification

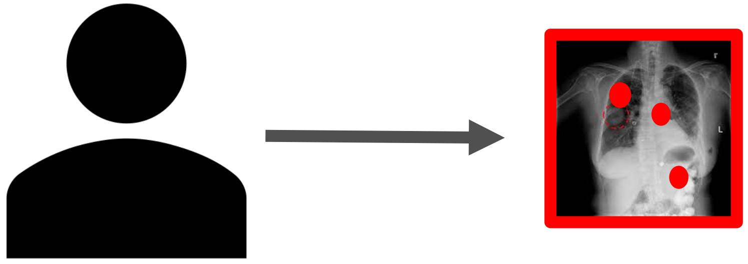


Multi-Label: Cardiomegaly and Hernia, no Edema

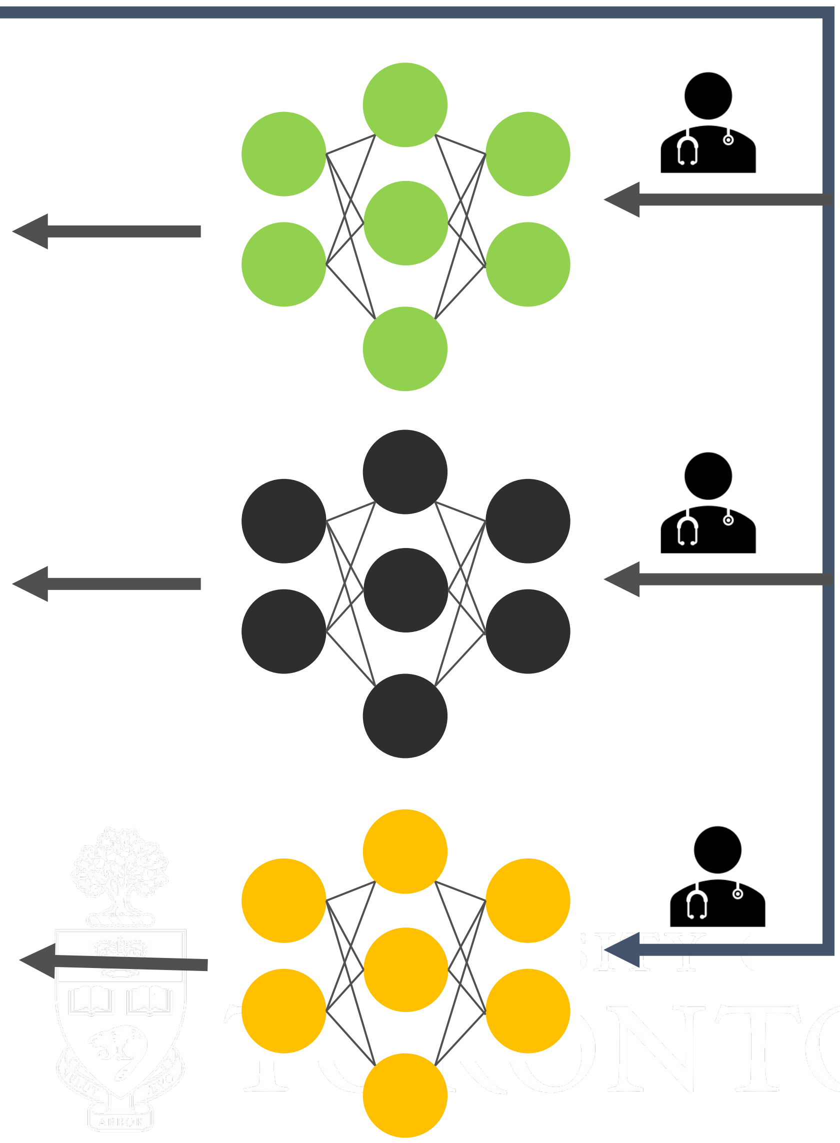
Multi-winner election for a set of voters each with a vote per label



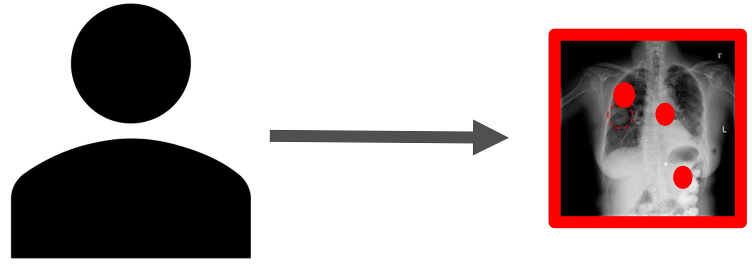
Alternative Powerset of All Possible Votes



0	0	0	0	4	1	0	0
1	0	0	1	5	1	0	1
2	0	1	0	6	1	1	0
3	0	1	1	7	1	1	1

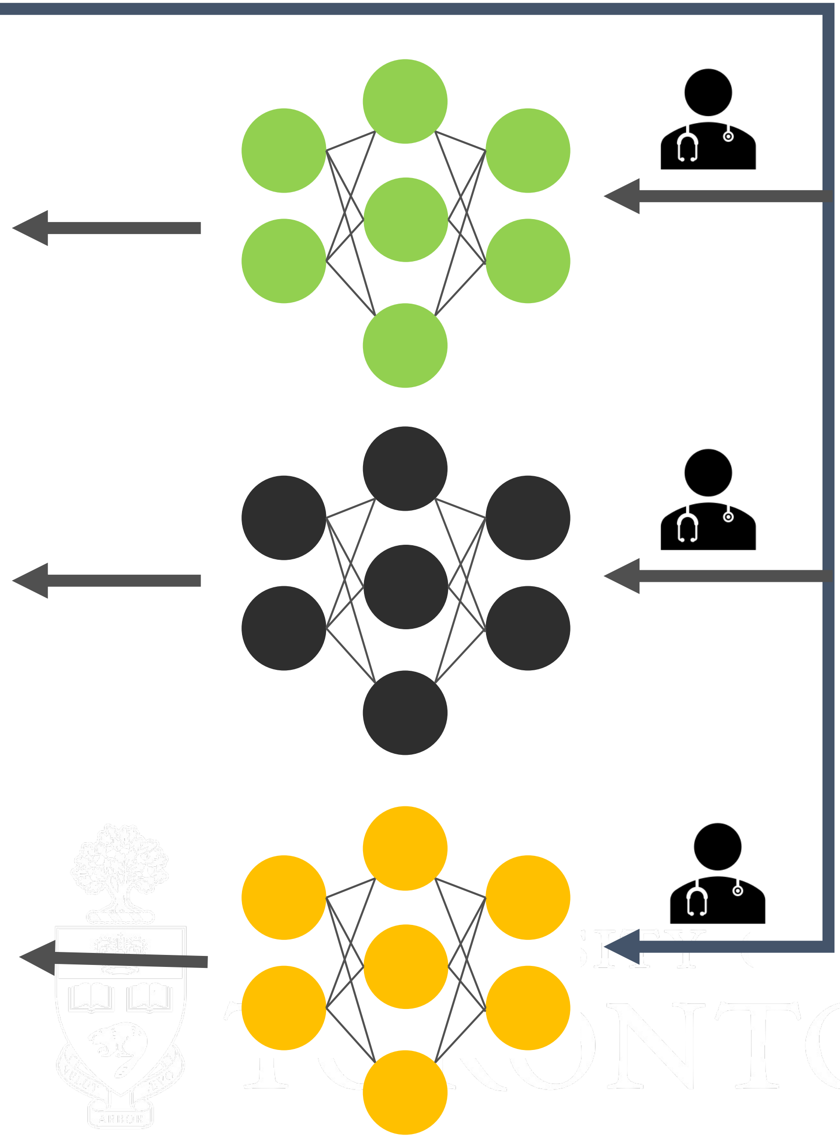


Alternative Powerset Multi-Label Classification

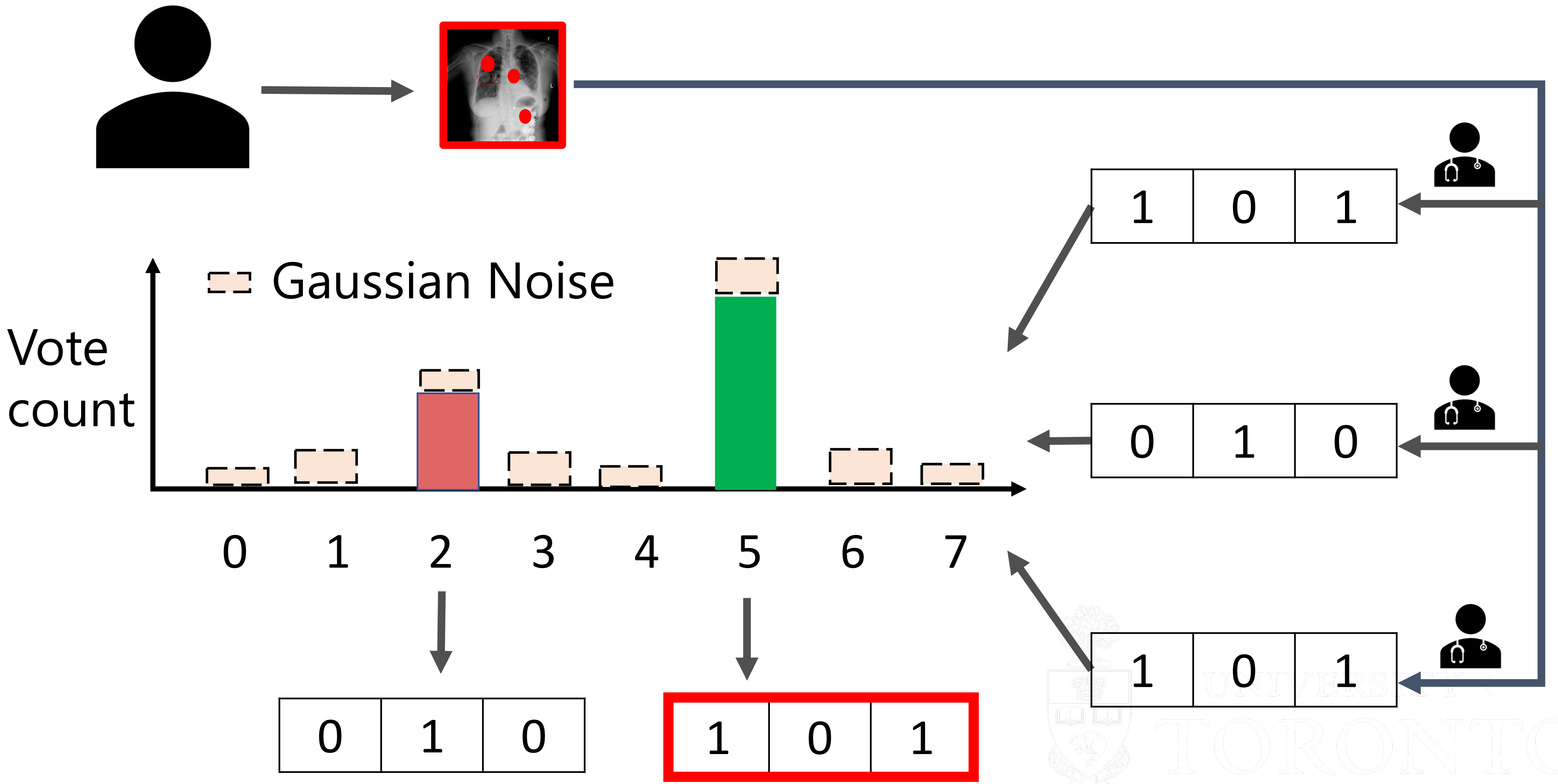


# of classes grows exponentially: 2^N

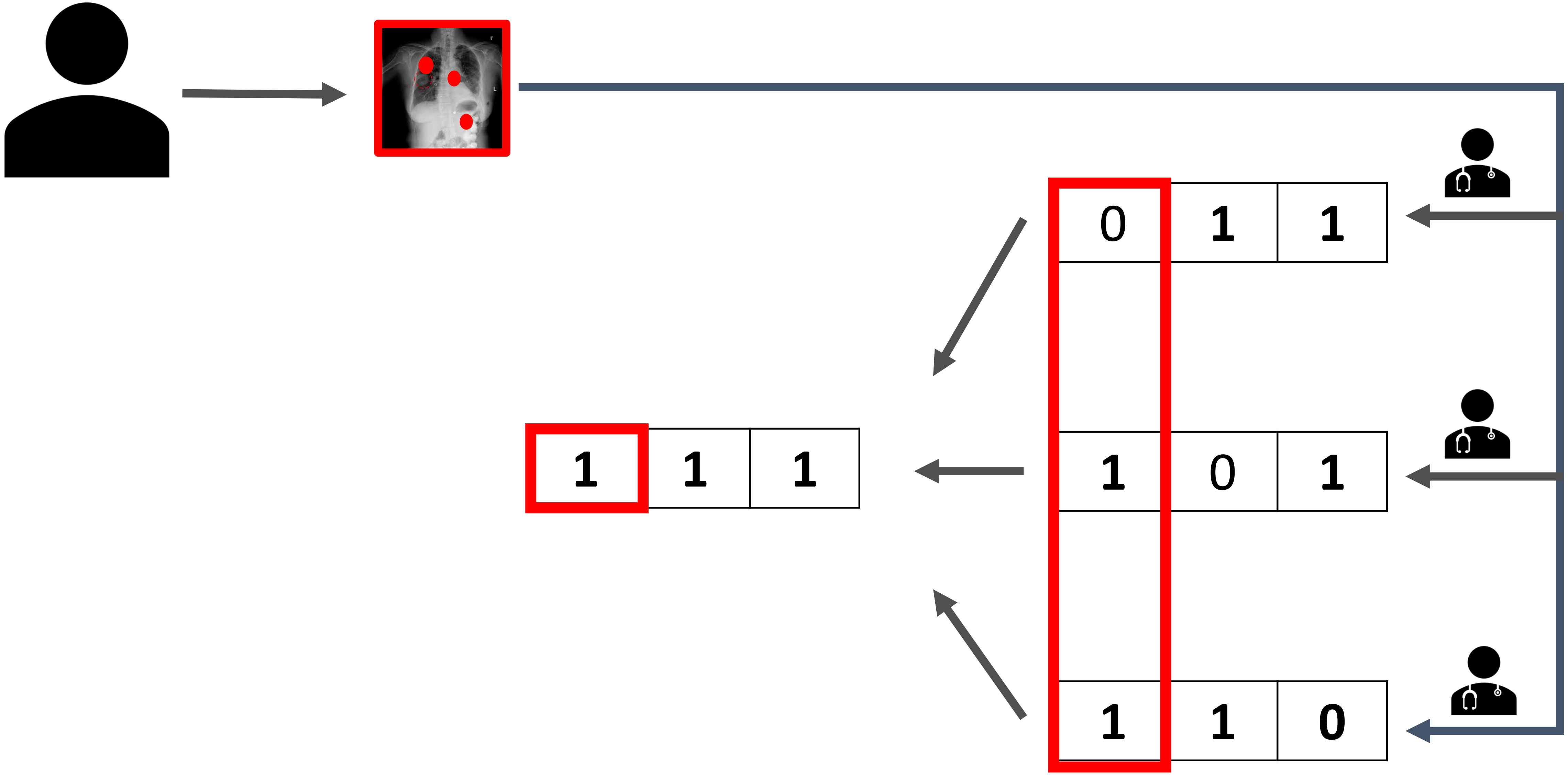
0	0	0	0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1



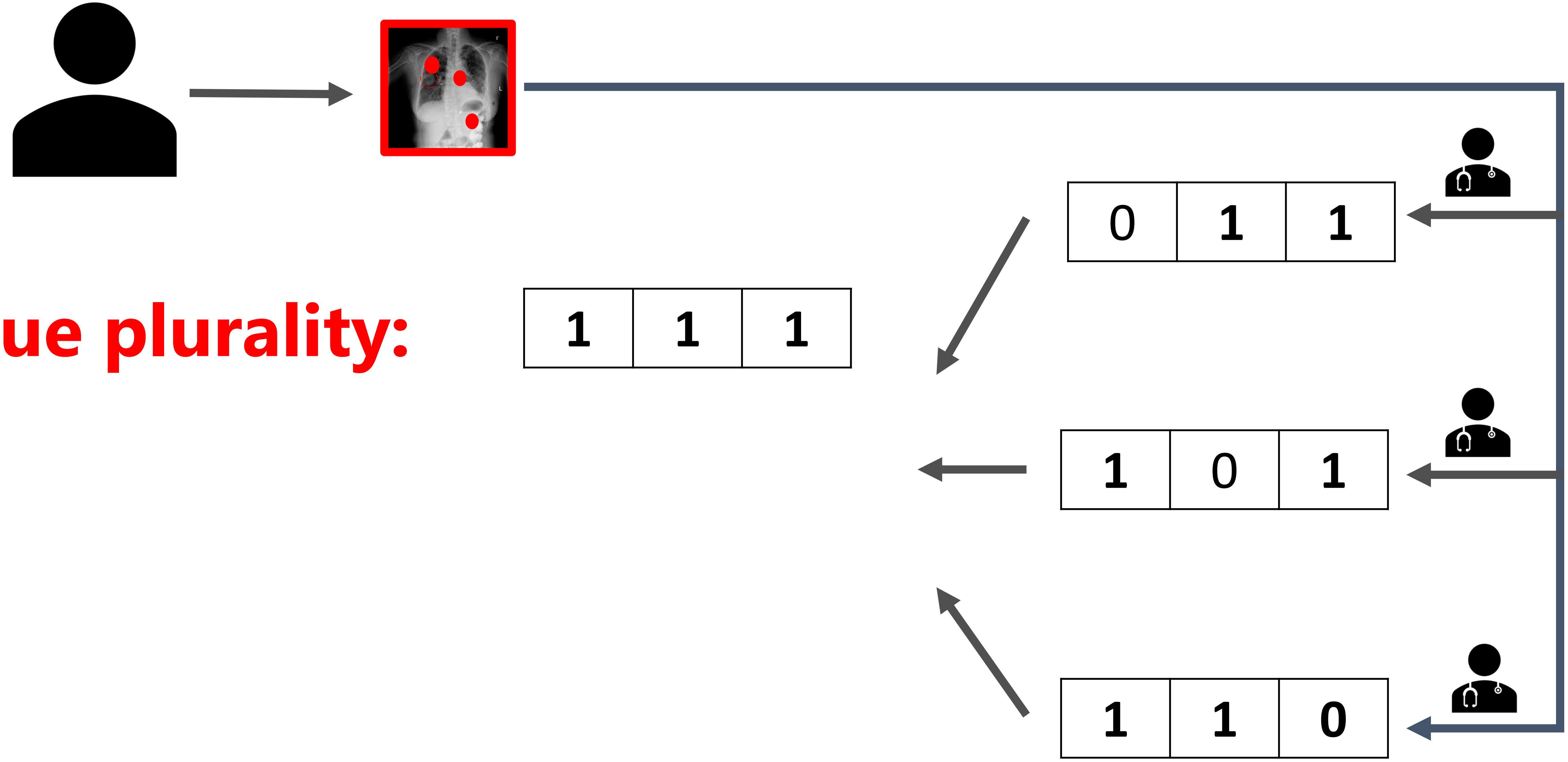
Powerset Multi-Label Classification



Binary vs Powerset

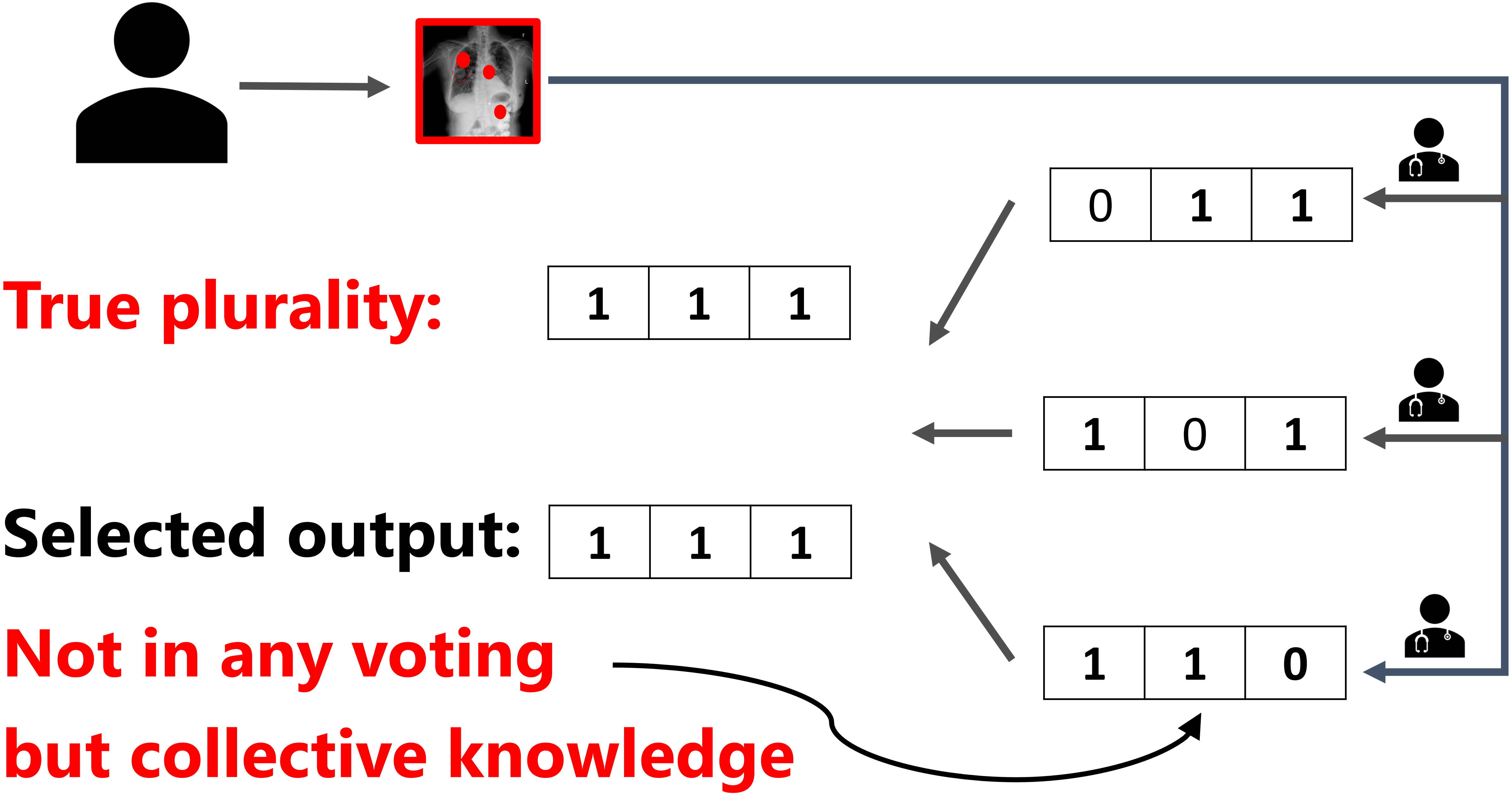


Binary Method Returns **True Plurality**

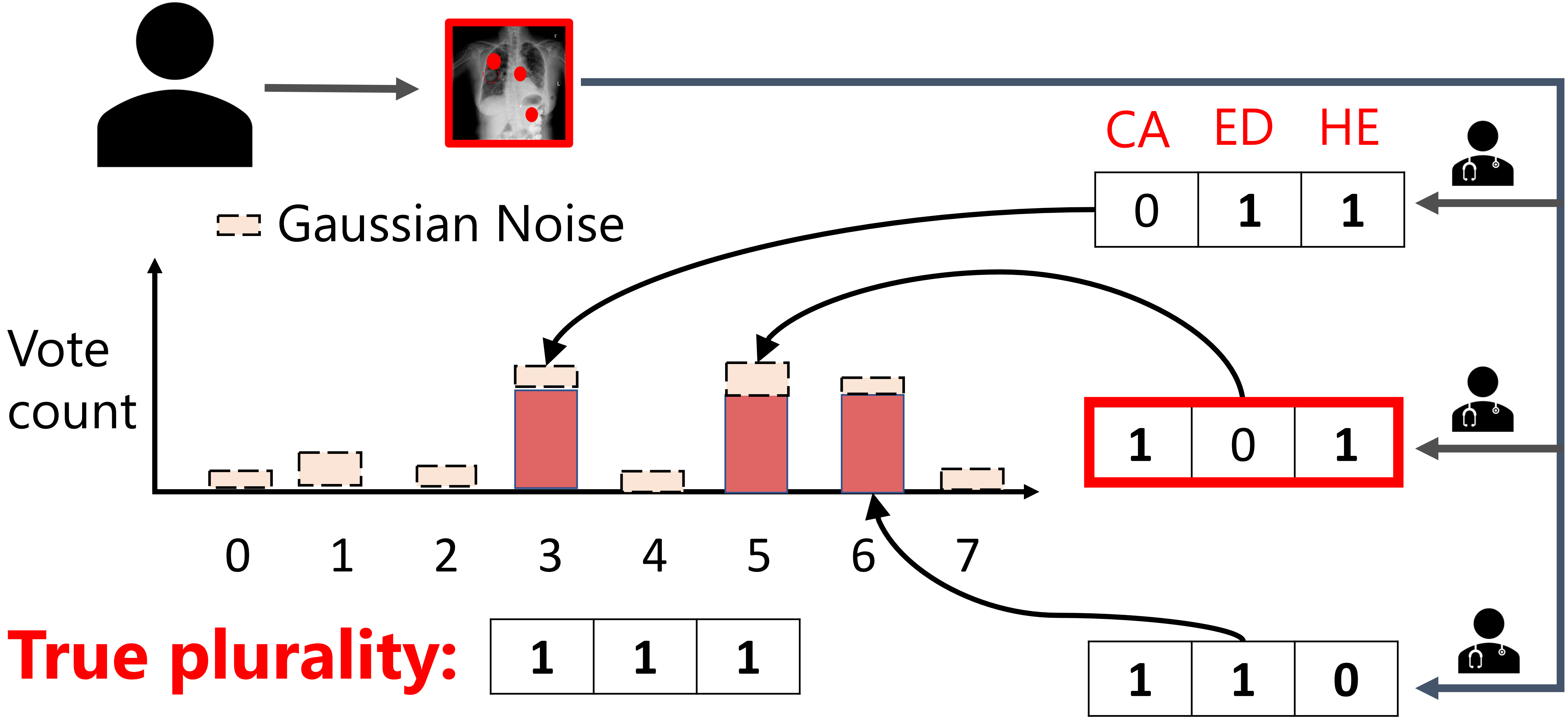


True plurality:

Binary Method Returns True Plurality



Powerset Method Approximates True Plurality



True plurality: 1 1 1

Selected 1 0 1 is one of the votings

Label Correlation: Binary vs Powerset

Uncorrelated Labels

CA	ED	HE
1	0	0
0	**1**	0
0	0	**1**
0	**1**	0

Binary is optimal

Label Correlation: Binary vs Powerset

Uncorrelated Labels

CA	ED	HE
1	0	0
0	**1**	0
0	0	**1**
0	**1**	0

Binary is optimal

Noisy or Strongly Correlated Labels

CA	ED	HE
0	**1**	**1**
0	**1**	**1**
0	**1**	**1**
1	0	0

ED $\Leftrightarrow$ HE

Powerset performs better

Clipping the Number of Positive Votes

0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Pascal VOC Dataset
Avg. 2 out of 20 labels

Clipping Positive Votes per Answering Party

$b_j = \boxed{1} \boxed{1} \boxed{1}$

Max 2 votes

$$b_j = \min\left(1, \frac{\tau}{\|b_j\|_2}\right) b_j$$

L_2 clipping for Binary

Clipping Positive Votes per Answering Party

$$b_j = \begin{array}{|c|c|c|}\hline 1 & 1 & 1 \\ \hline\end{array}$$

Max 2 votes $\downarrow$

$$b_j = \min\left(1, \frac{\tau}{\|b_j\|_2}\right) b_j$$

$$\|b_j\|_2^2 = 2 \downarrow$$

$$b_j = \begin{array}{|c|c|c|}\hline \sqrt{2}/\sqrt{3} & \sqrt{2}/\sqrt{3} & \sqrt{2}/\sqrt{3} \\ \hline\end{array}$$

L_2 clipping for Binary

Clipping Positive Votes per Answering Party

$b_j =$

1	1	1

Max 2 votes

$$b_j = \min\left(1, \frac{\tau}{\|b_j\|_2}\right) b_j$$

$$\|b_j\|_2^2 = 2$$

$b_j =$

$\sqrt{2}/\sqrt{3}$	$\sqrt{2}/\sqrt{3}$	$\sqrt{2}/\sqrt{3}$

L_2 clipping for Binary

Max 3 votes

0	0	0	0
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1

All Powerset classes

Clipping Positive Votes per Answering Party

$b_j =$ | 1 | 1 | 1 |

Max 2 votes

$$b_j = \min\left(1, \frac{\tau}{\|b_j\|_2}\right) b_j$$

$$\|b_j\|_2^2 = 2$$

$b_j =$ | $\sqrt{2}/\sqrt{3}$ | $\sqrt{2}/\sqrt{3}$ | $\sqrt{2}/\sqrt{3}$ |

Max 1 vote

0	**0**	**0**	**0**	**4**	**1**	**0**	**0**
1	**0**	**0**	**1**	5	1	0	1
2	**0**	**1**	**0**	6	1	1	0
3	0	1	1	7	1	1	1

L_2 clipping for Binary

Fewer classes for Powerset

Metrics for Multi-Label Classification

Accuracy (ACC) = (TP + TN) / (P + N)

Balanced Accuracy (BAC) = $\frac{1}{2}$(TPR + TNR)

Area-Under-the-Curve (AUC) = $\int_0^1 t(f)df$, t(f) = TPR / FPR

Mean-Average-Precision (MAP) = TP / (TP + FP)

Compare Binary vs Powerset Mechanisms

Pascal VOC, 20 labels, ResNet 50, $\varepsilon = 20$, $\delta = 1e^{-5}$

Method	**ACC**	**BAC**	**AUC**	**MAP**
Non-private	*0.97*	*0.85*	*0.97*	*0.85*
DPSGD	0.92	0.50	0.68	0.40
Powerset	**0.94**	0.58	0.70	0.29
Binary	**0.94**	**0.62**	**0.85**	**0.57**

Confidential & Private Collaborative Learning

Pascal VOC, 20 labels, ResNet 50, $\varepsilon = 20$, $\delta = 1e^{-5}$

Method	ACC	BAC	AUC	MAP
Before	0.93	0.59	0.88	0.54
After	**0.94**	**0.64**	**0.89**	**0.55**

Models Improve with Multi-Label CaPC

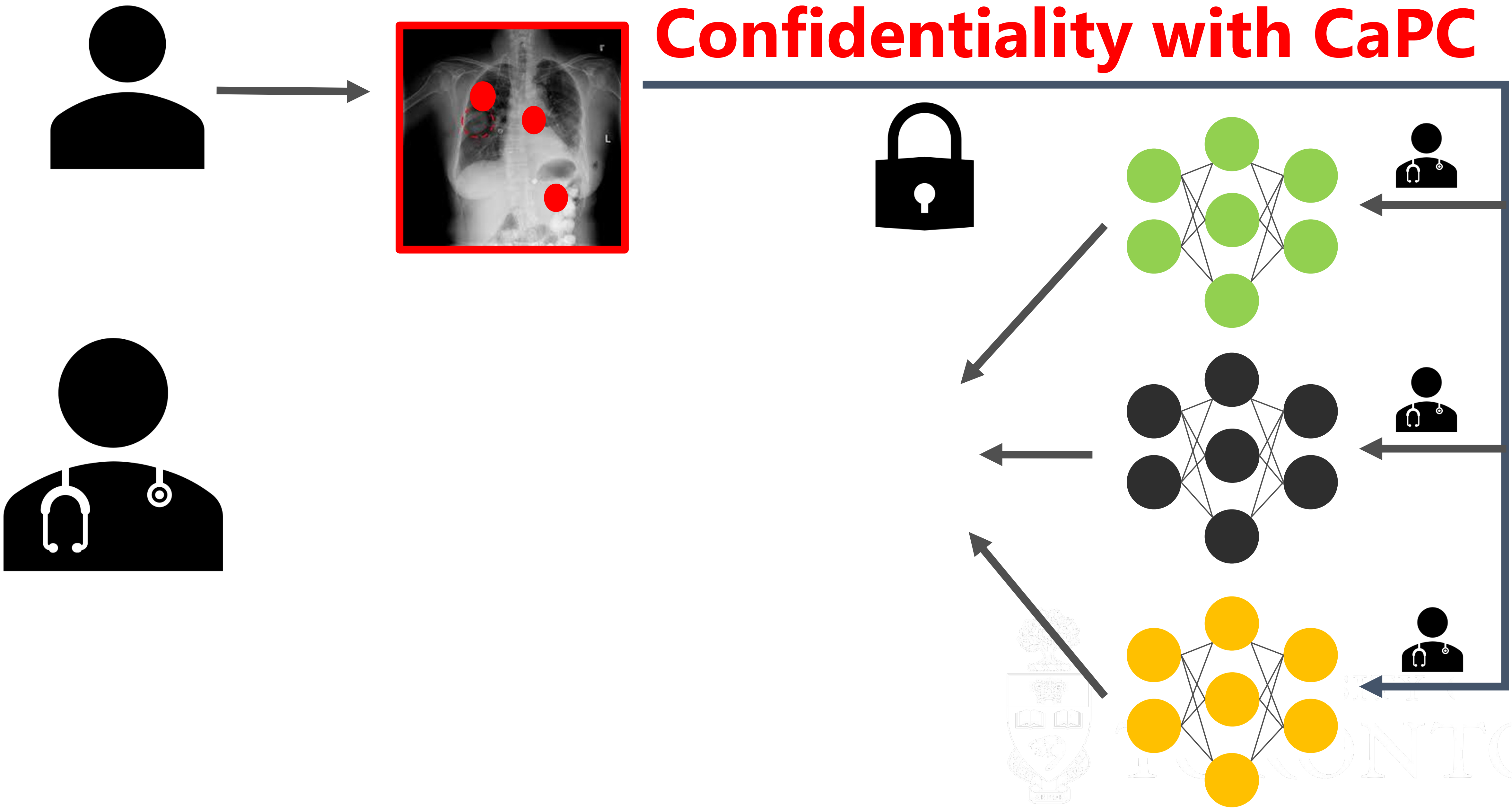
Pascal VOC, 20 labels, ResNet 50, $\varepsilon = 20$, $\delta = 1e^{-5}$

Method	ACC	BAC	AUC	MAP
Before	0.93	0.59	0.88	0.54
After	**0.94**	**0.64**	**0.89**	**0.55**

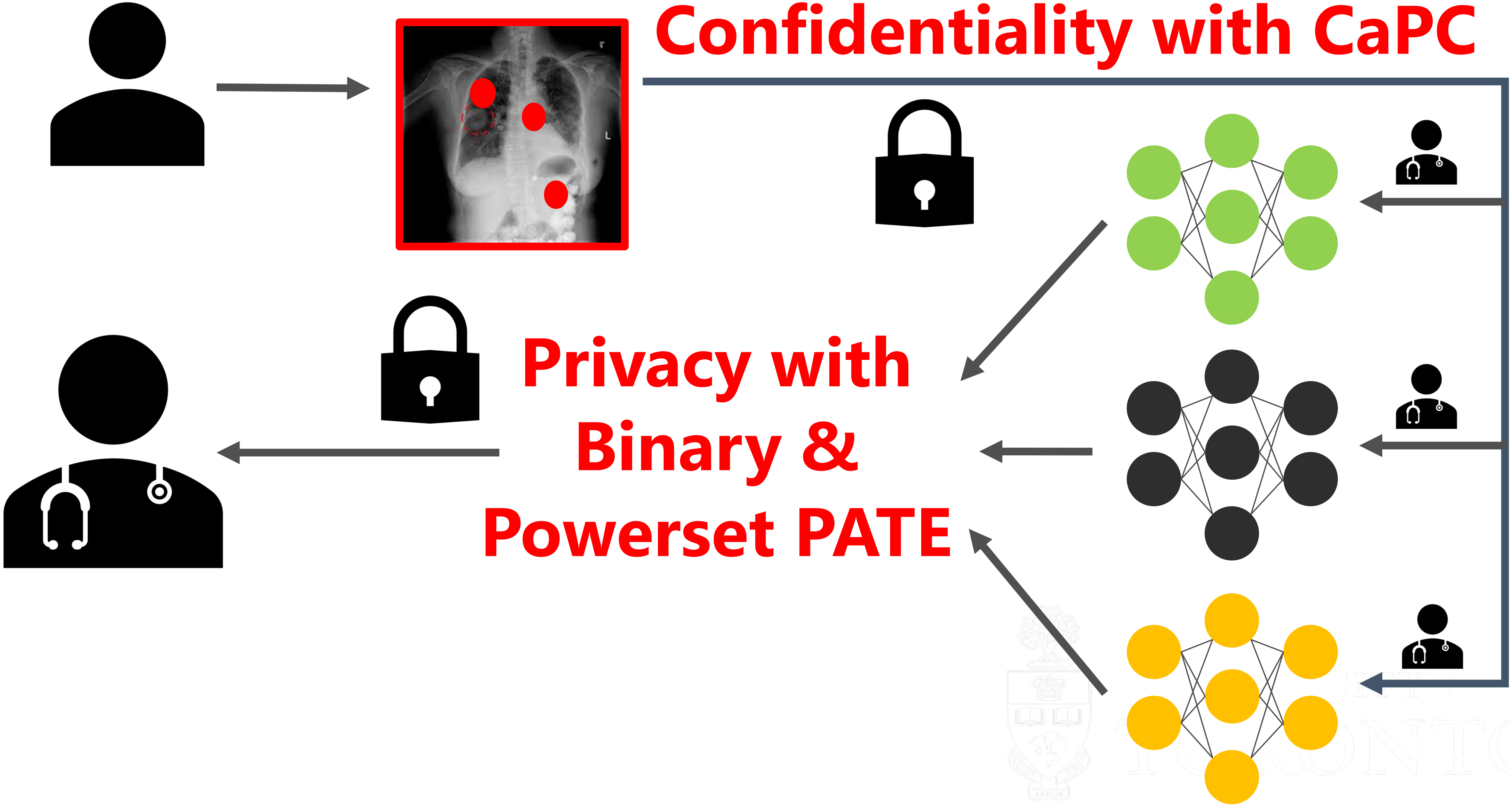
MIMIC, 11 labels, DenseNet 121, $\varepsilon = 20$, $\delta = 1e^{-6}$

Method	ACC	BAC	AUC	MAP
Before	0.84	0.63	0.78	0.43
After	**0.85**	**0.64**	**0.79**	**0.45**

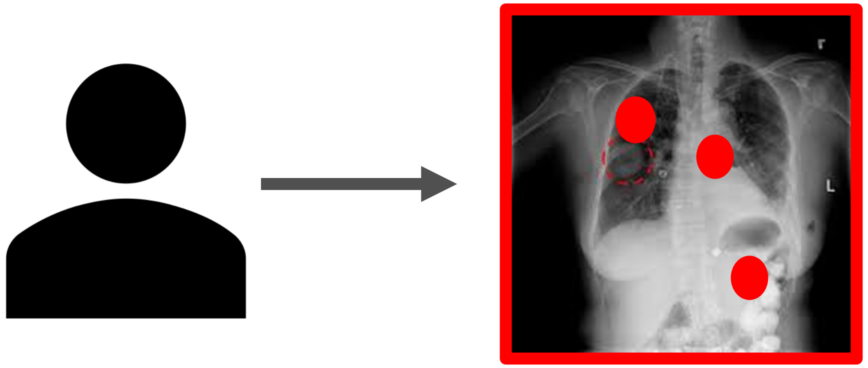
Confidential Multi-Label Classification

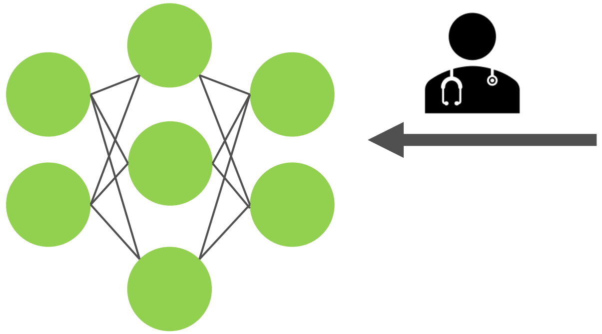


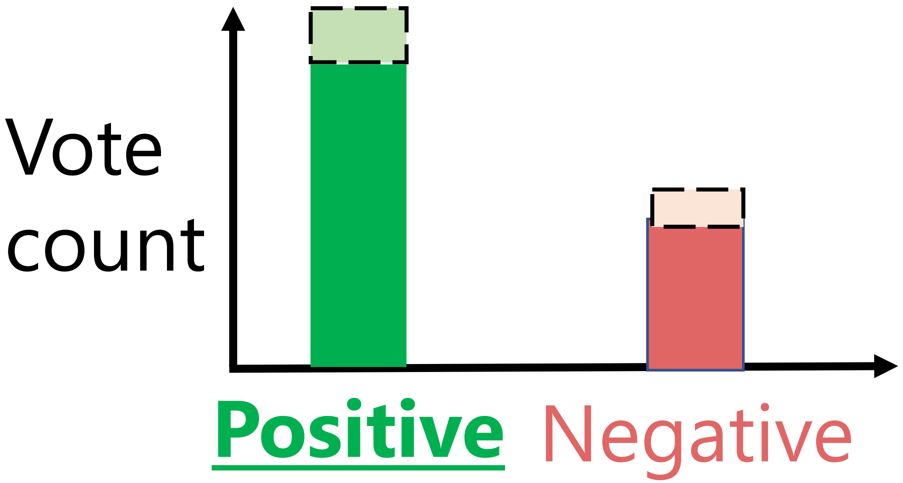
Private Multi-Label Classification with PATE



Thank you







Multi-Label PATE & CaPC

Join our SprintML Lab at CISPA!

We are **hiring Ph.D. students, Postdocs, and Research Interns** with a research focus in one or multiple of the following areas:

- Privacy-Preserving Machine Learning
- Secure and Robust Machine Learning
- Distributed and Federated Learning
- Machine Learning Model Confidentiality
- Trustworthy Language Processing

Backup

Retraining with Multi-Label CaPC ($\boldsymbol{\varepsilon = 10}$)

Dataset	# of Models	State	PB (ε)	ACC	BAC	AUC	mAP
Pascal VOC	1	Initial	-	.97	.85	.97	.85
	50	Before CaPC	-	.93±.02	.59±.01	.88±.01	.54±.01
	50	After CaPC	10	**.94±.01**	**.62±.01**	.88±.01	.54±.01
	50	After CaPC	20	**.94±.01**	**.64±.01**	**.89±.01**	**.55±.01**
CheXpert	1	Initial	-	.79	.78	.86	.72
	50	Before CaPC	-	.77±.06	.66±.02	.75±.02	.58±.02
	50	After CaPC	20	.76±.07	**.69±.01**	**.77±.01**	**.59±.01**
MIMIC	1	Initial	-	.90	.74	.84	.51
	50	Before CaPC	-	.84±.07	.63±.03	.78±.03	.43±.02
	50	After CaPC	20	**.85±.05**	**.64±.01**	**.79±.01**	**.45±.03**

Performance of Binary for different ε values

PB (ε)	Queries answered	ACC	BAC	AUC	MAP
1	**0**	-	-	-	-
2	**6**	.86	.62	.62	.44
3	**13**	.93	.67	.67	.53
4	**22**	.93	.64	.64	.44
5	**31**	.95	.63	.63	.39
6	**40**	.95	.67	.67	.45
7	**64**	.95	.64	.64	.35
8	**81**	.95	.66	.66	.40
9	**101**	.95	.60	.60	.28
10	**113**	.96	.63	.63	.30
11	**135**	.96	.64	.64	.33
12	**165**	.96	.65	.65	.35
13	**199**	.96	.63	.63	.32
14	**217**	.96	.64	.64	.35
15	**239**	.96	.63	.63	.32
16	**272**	.96	.63	.63	.31
17	**306**	.96	.63	.63	.30
18	**332**	.96	.63	.63	.31
19	**362**	.96	.63	.63	.30
20	**403**	.96	.63	.63	.30

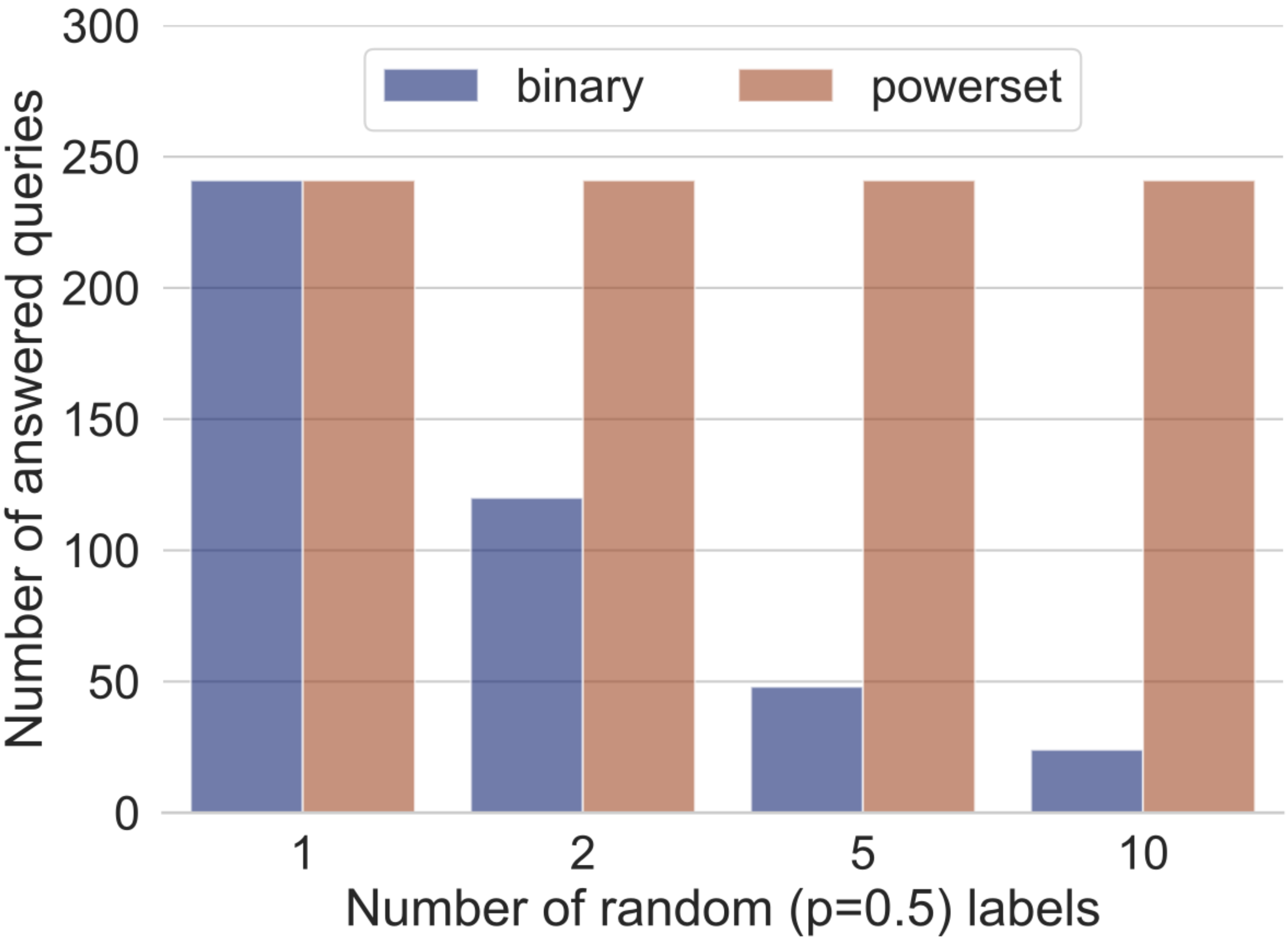
DPSGD vs PATE on the CheXpert Dataset

We compute the Area-Under-the-Curve (AUC) metric per label.
Adaptive denotes the Adaptive DPSGD for multi-label classification.

$$\varepsilon = 8, \delta = 10^{-4}$$

Method	AT	CA	CO	ED	EF	Average
Non-private	*0.84*	*0.80*	*0.87*	*0.90*	*0.91*	*0.87*
DPSGD	0.56	0.53	0.66	0.56	0.62	0.58
Adaptive	0.75	0.73	**0.84**	**0.79**	0.79	0.78
Binary PATE	**0.78**	**0.75**	**0.84**	0.76	**0.81**	**0.79**

Randomly Generated Labels



Compare Binary vs Powerset Mechanisms

Accuracy (ACC) = (TP + TN) / (P + N)

Balanced Accuracy (BAC) = $\frac{1}{2}$(TPR + TNR)

Area-Under-the-Curve (AUC) = $\int_0^1 t(f)df$, t(f) = TPR / FPR

Mean-Average-Precision (MAP) = TP / (TP + FP)

Method	ACC	BAC	AUC	MAP
Non-private	*0.97*	*0.85*	*0.97*	*0.85*
DPSGD	0.92	0.50	0.68	0.40
Powerset	**0.94**	0.58	0.70	0.29
Binary	**0.94**	**0.62**	**0.85**	**0.57**

Pascal VOC, 20 labels, ResNet 50, $\varepsilon = 20$, $\delta = 1e^{-5}$