# Balancing Predictive Performance and Explainability for Integrating Clinical, Biological, and Radiological Data in the Basque Country COVID-19 Cohort

**Lander RODRIGUEZ-IDIAZABAL**, Irantzu BARRIO, Jose Maria QUINTANA-LOPEZ

Univ Pais Vasco (SP)

Marta AVALOS-FERNANDEZ, Dylan RUSSON

Univ of Bordeaux (FR)

**65th ISI World Statistics Congress**

7 Oct 2025, The Hague, the Netherlands

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea
bcam basque center for applied mathematics
EXCELENCIA SEVERO OCHOA
ISI 5 - 9 OCTOBER 2025 The Hague
Inria
université de BORDEAUX
Inserm
BORDEAUX POPULATION HEALTH Research Center - U1219

# CONTEXT

- Modern **healthcare** routinely generates massive, heterogeneous data:

EHRs, Biological data, X-ray, Clinical notes, Treatments,…

Need for:

- **BALANCE** → Predictive power VS Explanaibility
  - **Accurate** → To usefully guide decisions
  - **Explainable** → Clinicians need transparent models to gain trust
- **INTEGRATION** of multiple data sources and multimodal data → Efficient

# The Basque Country COVID-19 Cohort

Adult patients diagnosed with COVID-19 and admitted to Galdakao-Usansolo Hospital's ED between March 2020 and January 2022 (N = 5,504)

- **Tabular data:**
  - **Sociodemographic (3 variables):** age, sex and vaccination status
  - **Comorbidities (13 variables):** Charlson index, and presence of comorbidities (heart failure, dementia, diabetes…)
  - **Laboratory (22 variables):** Partial $CO_2$ and $O_2$ pressure, glucose…
- **Image data: RAW & UNLABELED** Frontal X-ray image of lungs
- **Outcome** : mortality status

[Quintana-Lopez, Rodríguez-Idiazabal, 2024]

# OBJECTIVE

No longer in the urgency of developing predictive models for COVID-19 mortality

Rich data, valuable opportunity to develop/test methods and integration techniques

**- Challenges:**

- ***Missing*** chest X-ray → 965 / 5,504
- **Unlabeled** chest X-ray images → No reports, No class labels, No expert feedback
- ~40 mostly continuous tabular → Potentially **non-linear** relationships
- **Imbalanced** outcome classes → <10% mortality rate
- **Trade off** → Predictive performance vs. Explainability

**- Objective:** To investigate the **predictive performance** of COVID-19 death models

- **Single and multimodal** approaches using chest **X-ray images and EHR data**
- Models differ in their **explainability capacities**

# METHODS: Model architectures (1st attempt) - clustering

**Categorical** (Tabular) Variable Representing X-ray Information:

- Gaussian Mixture Variational Autoencoders with a Convolutional Neural Network architecture to classify frontal chest X-ray images into **clusters**
- Patients without X-ray images were assigned to a **“missing value” category**

Aim:

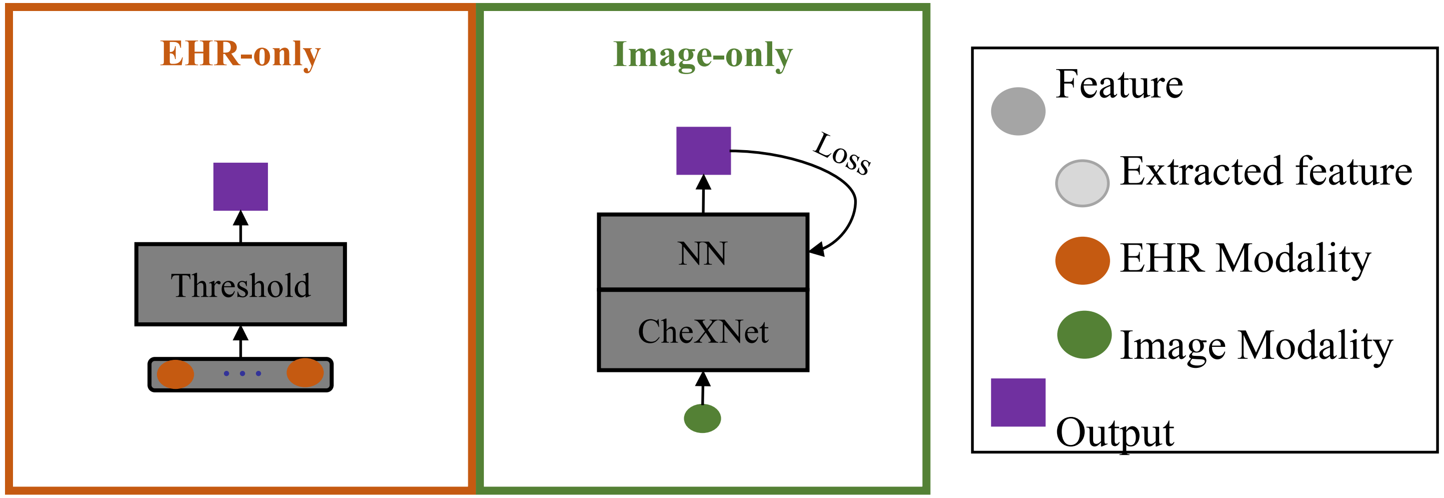
- To **facilitate data integration** across modalities
- To leverage the **full cohort** of >5,000 patients

*Naïve approach*: Clustering chest X-ray images of COVID-19 patients **without prior clinical knowledge** is neither predictive nor explainable

# METHODS: Model architectures – single modality

- **EHR-only** (Threshold)**: Cost-sensitive Lasso logistic regression** with **0-degree spline** to categorize variables for clinical interpretation

- **Image-only: Adapted CheXNet architecture** replacing its final classification layer with a task-specific set of fully connected (dense) layers



# METHODS: Model architectures – single modality

- **EHR-only** (Threshold)**: Cost-sensitive Lasso logistic regression** with **0-degree spline** to categorize variables for clinical interpretation

$$\max_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \omega_i \ln \frac{e^{y_i \sum_{j=1}^{p} \mathbf{x}_{ij} \boldsymbol{\beta}_j}}{1 + e^{\sum_{j=1}^{p} \mathbf{x}_{ij} \boldsymbol{\beta}_j}} - \lambda \|\boldsymbol{\beta}\|_1 \right\}$$

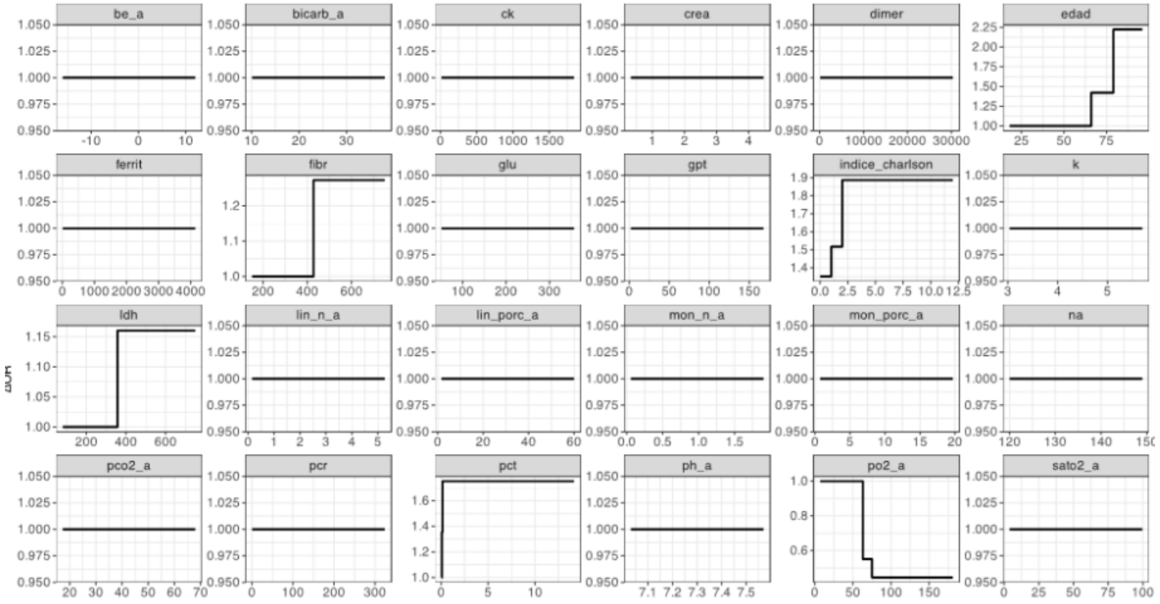
With $\mathbf{X} = \{x_{ij}\}, i = 1, \ldots, n; j = 1, \ldots, p$ and

$q_{kj}$ the $k$-th percentile of $X_j, k = 1, \ldots, k_j$

$$\chi_{ik}^{j} = \begin{cases} 1 & \text{if} \quad x_{ij} > q_{kj} \\ 0 & \text{otherwise.} \end{cases}$$

$\beta^j \in \mathbb{R}^{kj}$ and $\|\beta\|_1 = \sum_{j=1}^{p} \|\beta^j\|_1 = \sum_{j=1}^{p} \sum_{k=1}^{k_j} |\beta_k^j|$

$$\omega_i = \begin{cases} \omega & \text{if} \quad y_i = 1 \\ 1 & \text{otherwise.} \end{cases}$$



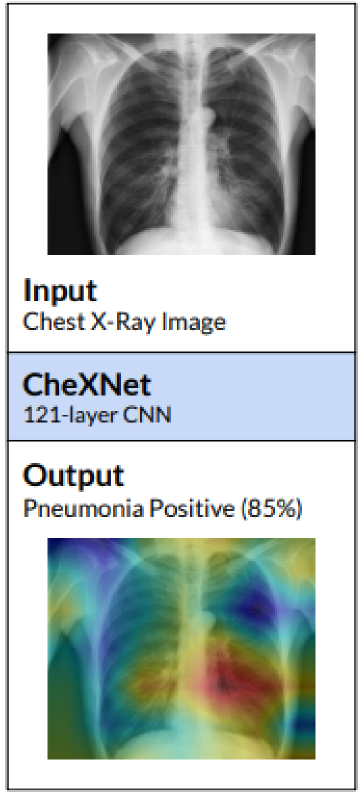

[Avalos, 2021]

# METHODS: Model architectures – single modality

- **Image-only: Adapted CheXNet architecture** replacing its final classification layer with a task-specific set of fully connected (dense) layers

## CheXNet Algorithm:

- Predicts the **probability of 14 disease categories** from chest X-rays
- Detects pneumonia at a level exceeding practicing radiologists
- 121-layer Convolutional Neural Network (**CNN**)
- Trained on >100,000 **frontal-view chest X-ray images** across 14 disease categories (ChestX-ray14 dataset)




- **Transfert learning**: pre-trained CheXNet adapted to predict
  - The probability of 14 disease categories or
  - The death

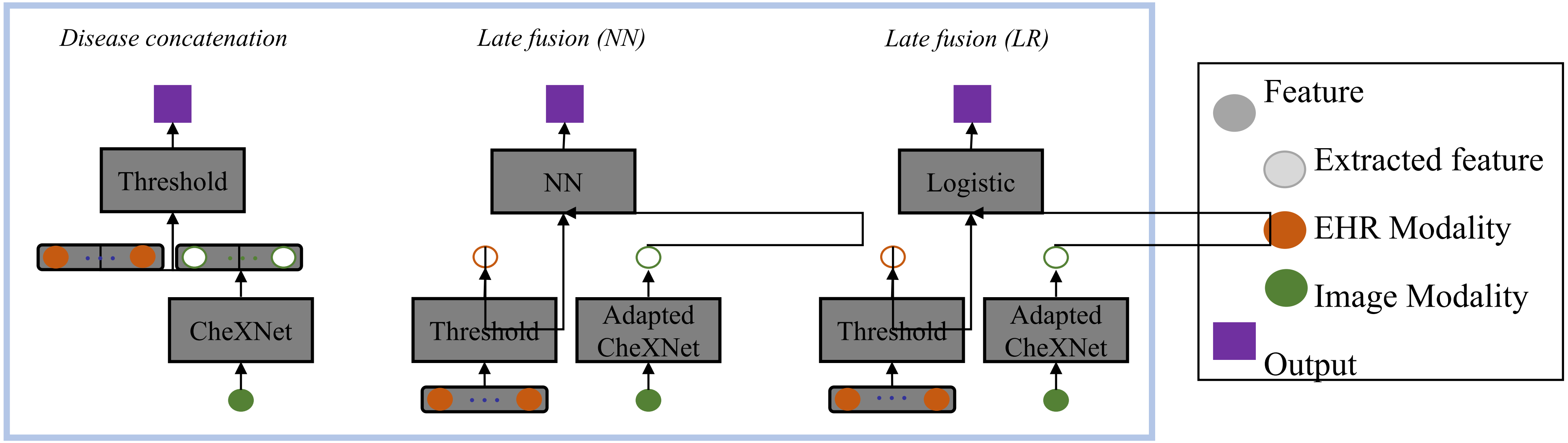
  in our COVID-19 data

[Rajpurkar, 2017]

# METHODS: Model architectures – multimodal

- **Disease concatenation:** Cost-sensitive Lasso with 0-splines applied to EHR data + 14 disease probability outputs from (unmodified) CheXNet

- **Late fusion:** Predicted probabilities from single modality models are inputs to:
  1. A set of fully connected networks (NN) or
  2. A logistic regression (LR)



# METHODS: Performance criteria

- **Train** (60%), **validation** (20%), and **test** (20%) **stratified** split of the data

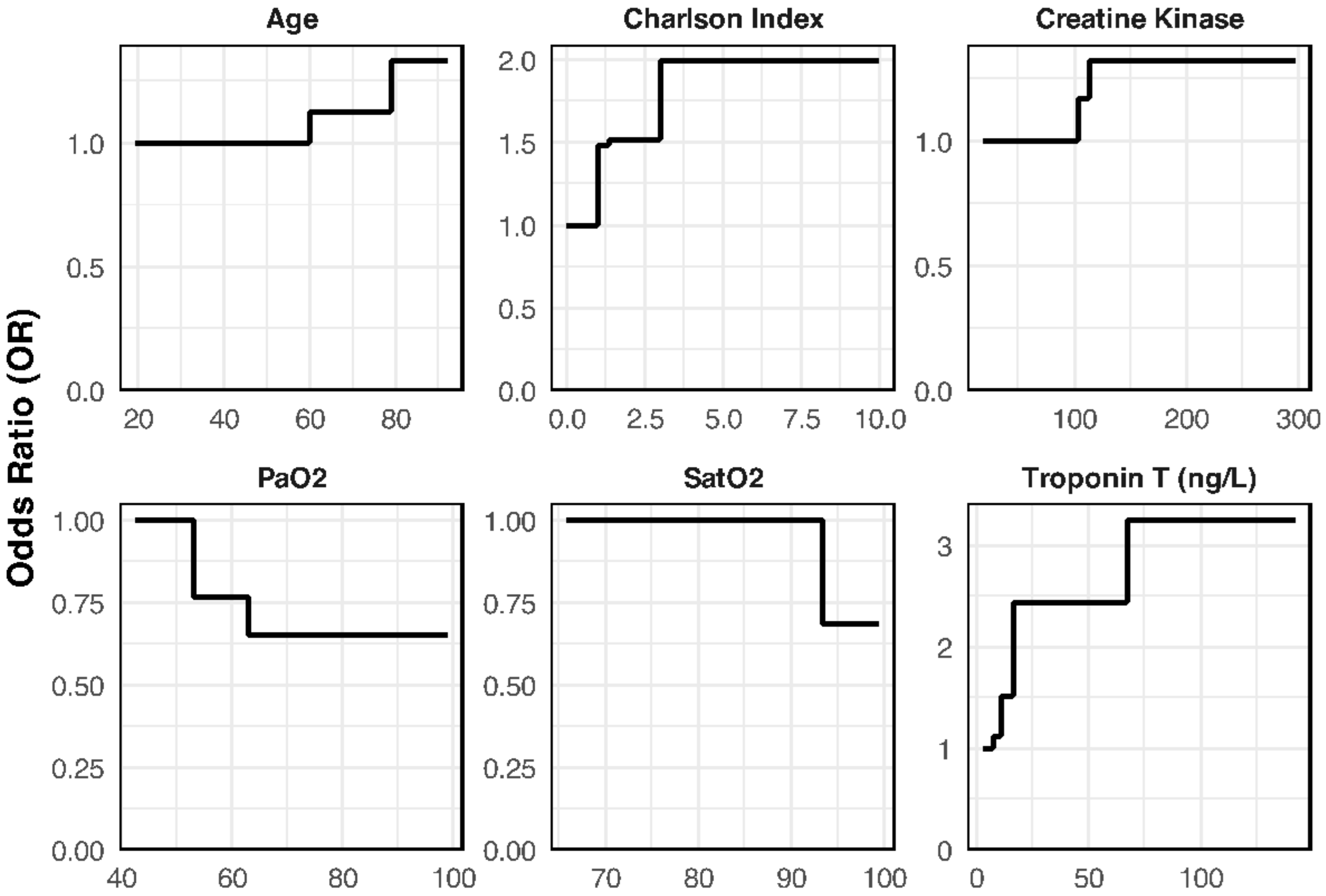
- **Predictive performance** assessed on the **test set**:
  - Area Under the Receiver Operating Characteristic Curve **(AUROC)**
  - Area Under the Precision-Recall Curve **(AUPRC)**

- **95% CIs**: bootstrap resampling with 1000 replicates

# RESULTS: Descriptive statistics

| Variable | Total sample | Non-deceased | Deceased | p-value |
|---|---|---|---|---|
| *N, (%)* | 965 (100%) | 901 (93.37%) | 64 (6.33%) | |
| *Sociodemographic variables* | | | | |
| **Age** | 63 [54-72] | 62 [53-71] | 78 [69-85] | <0.05 |
| **Charlson Index** | 1 [0-2] | 1 [0-2] | 3 [1-5] | <0.05 |
| **Sex at birth** | | | | 0.108 |
| Female | 322 (33.4%) | 307 (34.1%) | 15 (23.4%) | |
| Male | 643 (66.6%) | 594 (65.9%) | 49 (76.6%) | |
| *Comorbidities* | | | | |
| **Heart failure** | 91 (9.4%) | 69 (7.7%) | 22 (34.4%) | <0.05 |
| **Cerebrovascular disease** | 102 (10.6%) | 82 (9.1%) | 20 (31.2%) | <0.05 |
| **Arterial hypertension** | 465 (48.2%) | 422 (46.8%) | 43 (67.2%) | <0.05 |
| **Dementia** | 17 (1.8%) | 9 (1%) | 8 (12.5%) | <0.05 |
| **Diabetes** | | | | <0.05 |
| No | 764 (79.2%) | 724 (80.4%) | 40 (62.5%) | |
| Without chronic complications | 163 (16.9%) | 144 (16%) | 19 (29.7%) | |
| With chronic complications | 38 (3.9%) | 33 (3.7%) | 5 (7.8%) | |
| **Kidney disease** | 128 (13.3%) | 108 (12%) | 20 (31.2%) | <0.05 |
| **Cancer** | 54 (5.6%) | 48 (5.3%) | 6 (9.4%) | 0.28 |

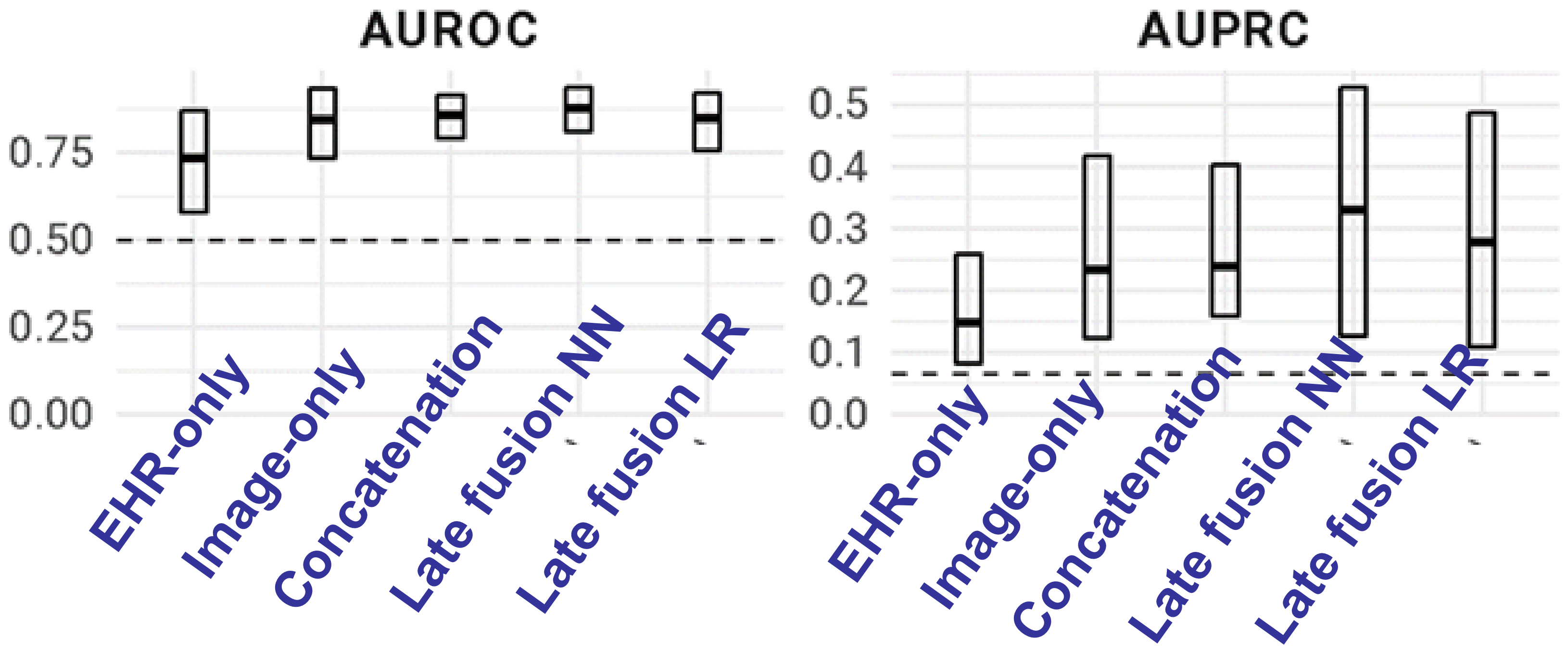
# RESULTS: EHR-only model



# RESULTS: Models performance

**Late Fusion Approaches:**

- Showed a **higher AUPRC** point estimate compared to single modality models
- However, variability was high (**wide 95% CIs**)



# CONCLUDING REMARKS

- **Integrating multimodal** data improves prediction - but also increases complexity
- New **open-source deep learning methods for medical imaging** offer exciting opportunities
- Simpler models (e.g., 0-degree splines) are well understood and mimic **clinical reasoning**
- Methods developed here can be **extended** beyond COVID-19 to other diseases
- **Balancing accuracy and transparency** remains the central challenge

# REFERENCES


**[Quintana-Lopez, Rodríguez-Idiazabal, 2024]**
Quintana-Lopez, Rodríguez-Idiazabal, Portuondo, García, Legarreta, Gascón, Larrea, Barrio & COVID-Health Basque Country Research Group (2024). Relevance of comorbidities for main outcomes during different periods of the COVID-19 pandemic. *Influenza and other respiratory viruses* https://doi.org/10.1111/irv.13240

**[Avalos, 2021]**
Avalos, Touchais, Henríquez-Henríquez (2021). Optimising Criteria for Manual Smear Review Following Automated Blood Count Analysis: A Machine Learning Approach. *In: Advances in Intelligent Systems and Computing, vol 1372. Springer*
https://doi.org/10.1007/978-3-030-73603-3_35

**[Rajpurkar, 2017]**
Rajpurkar, Irvin, Zhu, Yang, Mehta, Duan, Ding, Bagul, Langlotz, Shpanskaya, Lungren, & Ng (2017). CheXNet: *Radiologist-level pneumonia detection on chest X-rays with deep learning* [Preprint]. arXiv https://doi.org/10.48550/arXiv.1711.05225