# Prediction of Intensive Care Unit (ICU) admission in COVID-19 patients in Brazil: a multicentric machine learning analysis

Alexandre Chiavegatto Filho

*School of Public Health, University of São Paulo, São Paulo, Brazil*

*Authors: Carine Savalli, Roberta Moreira Wichmann, Fabiano Barcellos Filho, Fernando Timoteo Fernandes.*

LABDAPS
LABORATÓRIO DE BIG DATA E ANÁLISE PREDITIVA EM SAÚDE

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE SAÚDE PÚBLICA

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

# IACOV-BR

- A multicenter cohort study involving 16,236 adult patients who tested positive for COVID-19
- March-August 2020
- 18 hospitals across all the five regions of Brazil
- Large diversity in demographics, resources, and clinical practices
- 22 predictors: laboratory, clinical, and demographic data

NORTH

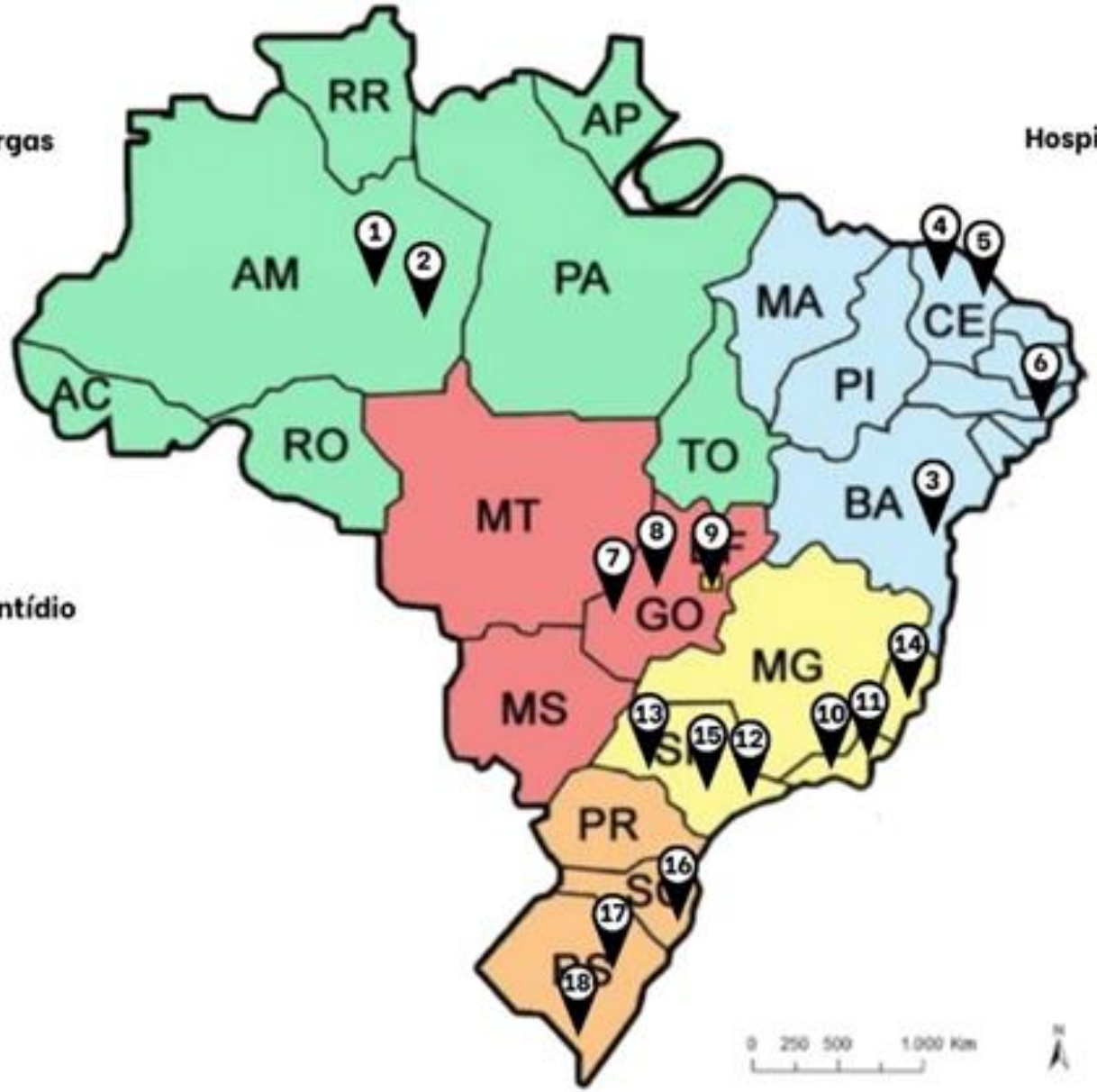
1. Hospital Universitário Getúlio Vargas
2. Hospital Santa Julia

NORTHEAST

3. Hospital Português da Bahia
4. Hospital Unimed Fortaleza
5. Hospital Universitário Walter Cantídio
6. Hospital Universitário HC

MIDWEST

7. Hospital Estadual de Luiziânia
8. Hospital Estadual de Trindade
9. Hospital Santa Lúcia




SOUTH EAST

10. Hospital Universitário Clementino Fraga Filho
11. Hospital Unimed-Rio
12. Hospital Santa Casa de São Paulo
13. Hospital São Francisco Mogi Guaçú
14. Hospital Evangélico de Vila Velha
15. Hospital das Clínicas da Faculdade de Medicina da USP

SOUTH

16. Hospital Santa Catarina Blumenau
17. Hospital Moinhos de Vento
18. Hospital Escola UFPEL

# Current study

- Aim: to predict Intensive Care Unit (ICU) admission in patients with COVID-19
- IACOV-BR
- Adult patients (> 18 years)
- 14 hospitals included with sample sizes ranging from 47 to 1500 patients

# Data preprocessing

For each hospital

- Box plots to identify extreme values
- Multinomial variables converted into dummy variables
- Continuous variables normalized with z-score transformation
- Variables with correlation exceeding 0.90 removed
- Variables with over 90% missing values were excluded
- Multiple imputation using chained equations (MICE)

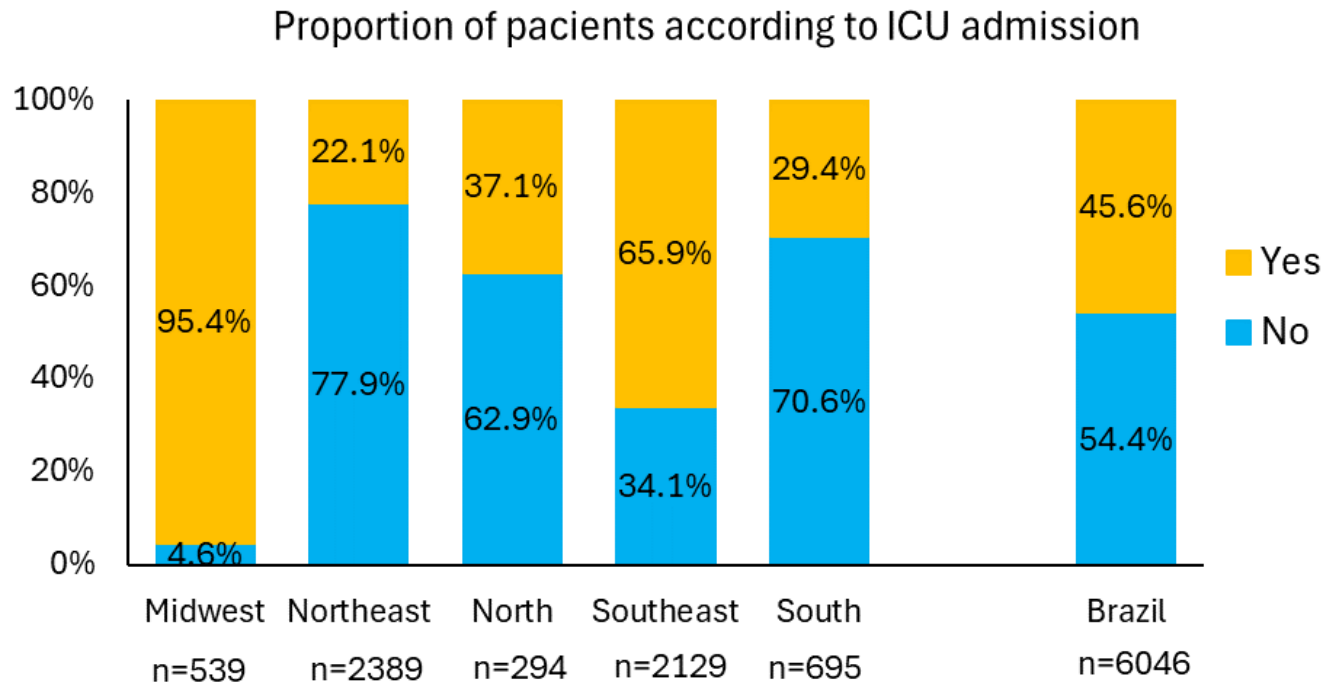
# Machine Learning methods

## For each hospital

- Three popular machine learning algorithms for tabular data: XGBoost, LightGBM, and CatBoost
- Hyperparameters optimization: HyperOpt 10-fold cross-validation
- Training/test (70% / 30%)
- To address the class imbalance: random oversampling in the training dataset
- Performance metric: area under the ROC curve (AUROC)

# Results

- Males: more prevalent (53.8%)
- Age: 57.5 years (SD=17.9)
- Among patients who provided information about self-declared race: White (65.4%)



# Results

| Region | Hospital | n | % of patients admitted to ICU | AUC | Best algorithm |
|---|---|---|---|---|---|
| SouthEast | SE2 | 1500 | 69.9% | 0.799 | XGBoost |
| | SE3 | 449 | 67.5% | 0.764 | LightGBM |
| | SE5 | 124 | 21.8% | 0.667 | LightGBM |
| | SE6 | 56 | 42.9% | 0.814 | Catboost |
| NorthEast | NE1 | 1359 | 18.0% | 0.940 | LightGBM |
| | NE2 | 845 | 21.9% | 0.666 | Catboost |
| | NE3 | 112 | 58.9% | 0.664 | LightGBM |
| | NE4 | 73 | 42.5% | 0.709 | Catboost |
| MidWest | MW1 | 539 | 95.4% | 0.673 | LightGBM |
| South | S1 | 456 | 30.0% | 0.748 | LightGBM |
| | S2 | 148 | 14.2% | 0.987 | Catboost |
| | S3 | 91 | 50.5% | 0.755 | LightGBM |
| North | N1 | 247 | 30.4% | 0.817 | XGBoost |
| | N2 | 47 | 72.3% | 0.679 | LightGBM |

# Discussion

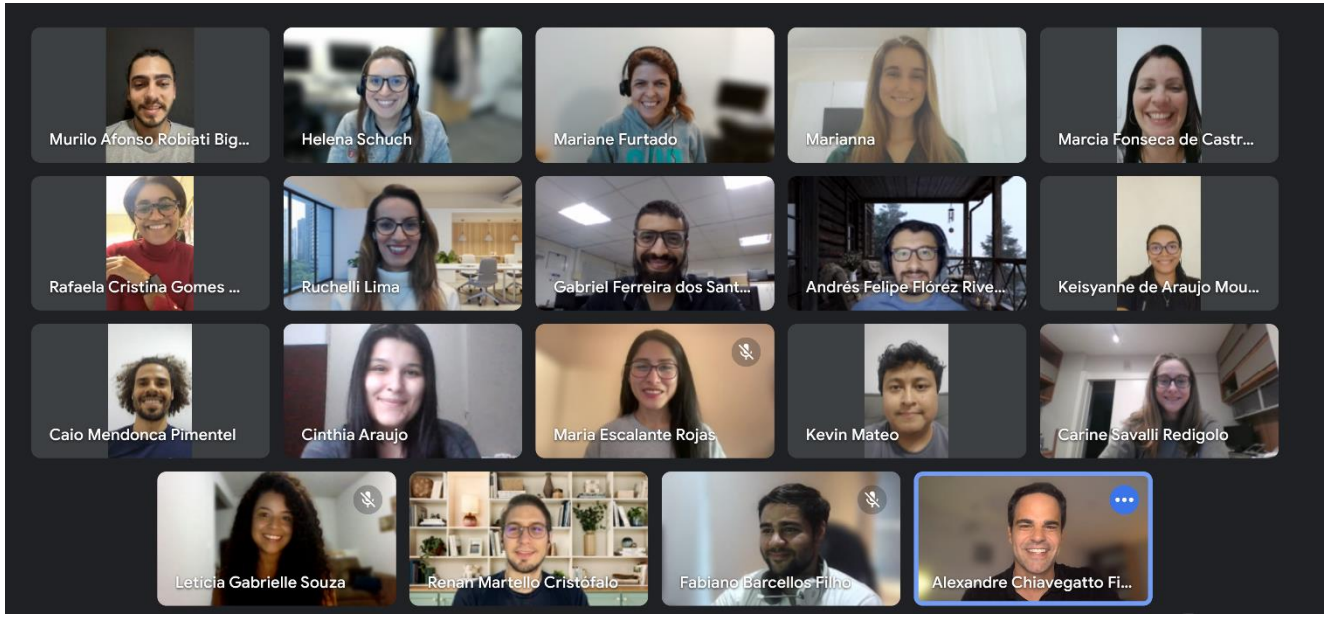
- The prediction of ICU admission with machine learning algorithms was good or excellent (AUC>0.7) for 9 hospitals
- Large variation among hospitals regarding the prevalence of the outcome
- Wider variety of predictive performances and best algorithms
- To those hospitals with poor predictive performance, alternative strategies should be evaluated (e.g. data aggregation or transfer learning)

# Acknowledgments

*Funding for this research was provided by the National Council for Scientific and Technological Development – CNPq and the Department of Science and Technology of Secretariat of Science, Technology, Innovation and Health Complex of Ministry of Health of Brazil – MoH, under grant No. 445020/2023-7*

LABDAPS

LABORATÓRIO DE BIG DATA E ANÁLISE PREDITIVA EM SAÚDE


*Thank you!*

Alexandre Chiavegatto Filho

http://labdaps.fsp.usp.br

@SaudenoBR

@labdaps

alexdiasporto@usp.br