IPCPREDICT

Colonization prediction

Racha Gouareb

Content

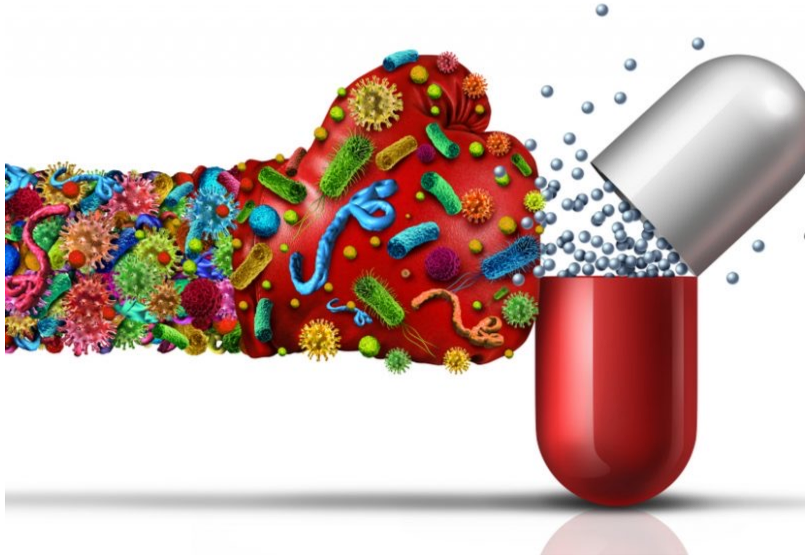
- Introduction
  - Background
  - Objective

- Methods
  - Dataset and features
  - Models

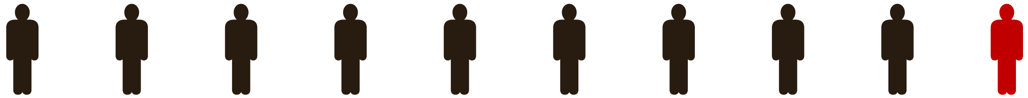
- Results

- Conclusion and future work
  - Environmental sample data set

Background - HAI



- 1 in every 10 inpatients develop a HAI*.
- Healthcare units are one of the biggest reservoirs of Carbapenem-resistant Enterobacteriaceae (CRE).
- 1.7 million infections and 99,000 associated deaths from AMR bacteria every year.
- HAIs result in unnecessary death and prolong hospital stays.



* World Health Organization (2016) Health care without avoidable infections - The critical role of infection prevention and control. World Health Organization, Geneva, Switzerland

Background - IPC

- The major goal of an IPC program is to decrease the incidence of healthcare associated infections (HAI), ideally to zero.

- World Health Organization (WHO) data demonstrates that
  - Effective IPC programs Reduce HAI rates by at least 30%
  - Surveillance contributes to a up to 57% reduction in HAIs

- WHO calls for
  - Better hand hygiene
  - Adequate environmental cleaning and disinfection
  - Adequate ventilation

- IPC strategies and measures are required to prevent or limit pathogens transmission

- High risk an inpatient faces of being colonized or infected by HCW, caregivers, other patients or visitors

Objective

- *Develop new systems as alternative solution to fight infection within the healthcare.*

- *Develop a machine learning model-based graph that predict inpatients colonization risks.*

MIMIC III

- Medical Information Mart for Intensive Care
- Data associated with 53,423 distinct hospital admissions between 2001 and 2012
- Consist of 22 tables
- Tables are linked by identifiers

The admissions table

Table source: Hospital database.

Table purpose: Define a patient's hospital admission, HADM_ID

The patients table

Table source: CareVue and Metavision ICU databases.

Table purpose: Defines each SUBJECT_ID in the database, i.e. defines a single patient.

The microbiologyevents table

Table source: Hospital database.

Table purpose: Contains microbiology information, including cultures acquired and associated sensitivities.

Number of rows: 631,726

MIMIC III - Deidentification

- Data removal :
  - Patient name
  - Telephone number
  - Address

HIPAA
Health Insurance Portability and Accountability Act

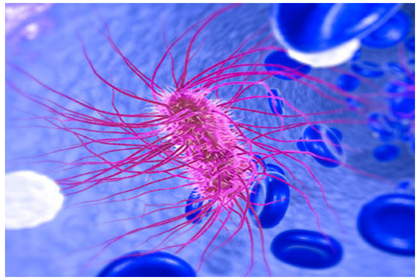
- Data shifting: From 2001-2012 to 2100 - 2200
  - Dates : shifted into the future preserving intervals
  - Time of day, day of the week, and approximate seasonality were conserved during date shifting.

State of art using MIMIC

- Prediction:
  - Mortality[1]
  - Length-of-Stay (LOS) [2]
  - Phenotyping (ICD code) [3]
  - Acute Respiratory Failure (ARF) [4]

[1] Huang B, Liang D, Zou R, Yu X, Dan G, Huang H, Liu H, Liu Y. Mortality prediction for patients with acute respiratory distress syndrome based on machine learning: a population-based study. Annals of Translational Medicine. 2021.
[2] T. Gentimis, A. J. Alnaser, A. Durante, K. Cook and R. Steele. Predicting Hospital Length of Stay Using Neural Networks on MIMIC III Data," 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing
[3] Gehrmann, Sebastian, et al. "Comparing rule-based and deep learning models for patient phenotyping." arXiv preprint arXiv:1703.08705 (2017)
[4] Wong, An-Kwok Ian, et al. "Machine learning methods to predict acute respiratory failure and acute respiratory distress syndrome." Frontiers in big Data 3 (2020): 39.

Enterobacteriaceae



Escherichia coli

Klebsiella pneumoniae

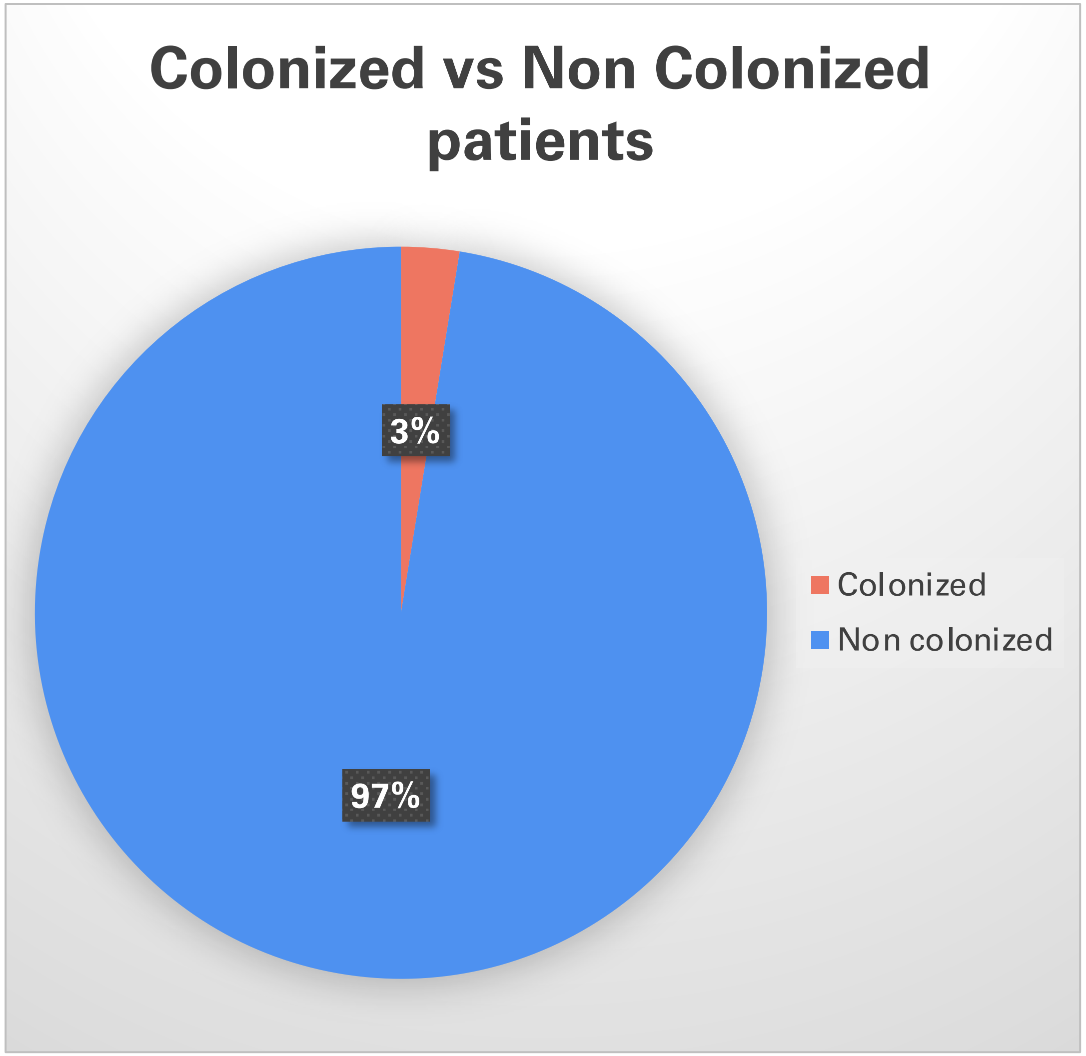
- Enterobacteriaceae are a large family of Gram-negative bacteria that includes E.coli and are a normal part of the gut flora.
- These pathogens can spread to the bloodstream resulting in life-threatening complications.
- Carbapenem-resistant *Enterobacteriaceae* (CRE) are *Enterobacteriaceae* that develop resistance to a group of antibiotics called carbapenems.

Example of specimens

Specimen	Flag	Category
SPUTUM	1	Respiratory
BLOOD CULTURE	**1**	**Blood**
URINE	**1**	**Urine**
BLOOD CULTURE	1	Blood
MRSA SCREEN	0	Other
SEROLOGY/BLOOD	1	Blood
EYE	1	Other
PLEURAL FLUID	1	Respiratory
STOOL	**0**	**Gastro**
SWAB	1	Other

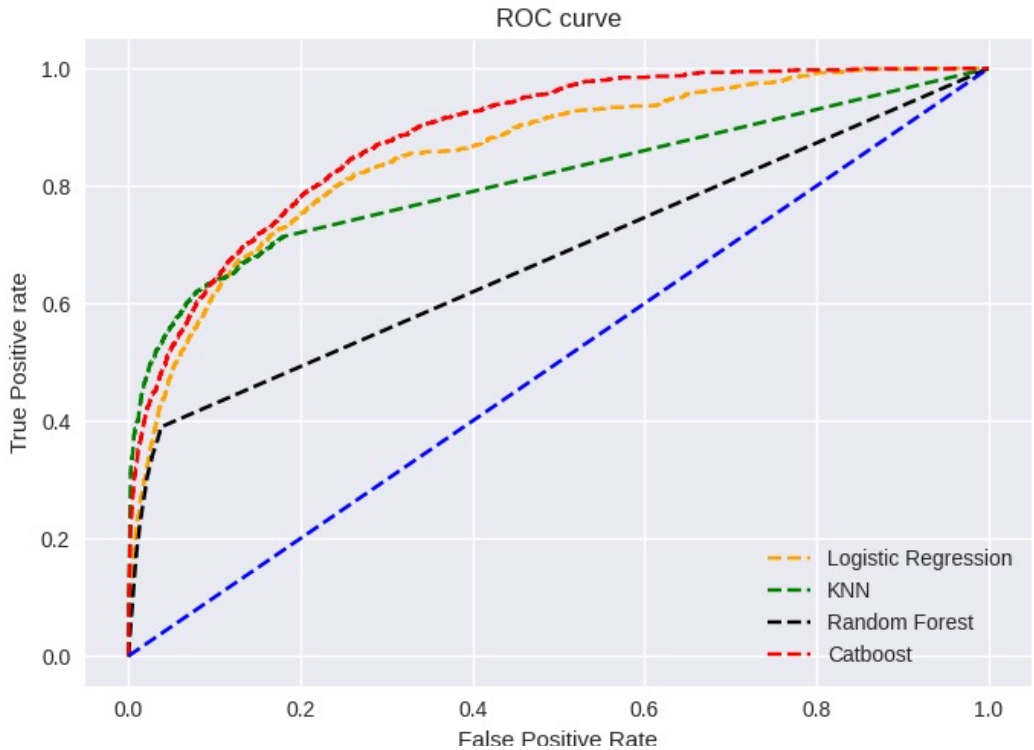
Data preprocessing

- Filter the data
- Link between tables
- Define relevant features
- Create new features



Results – Imbalanced (RWD)

	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**AUC (%)**
Logistic regression	82.2	10.2	**72.8**	17.9	85.4
KNN	**97.8**	**72.5**	33.5	**45.9**	80.9
Random forest	94.4	21.3	41	28	68.6
Catboost	**97.8**	69.1	31.5	43.2	**88.5**



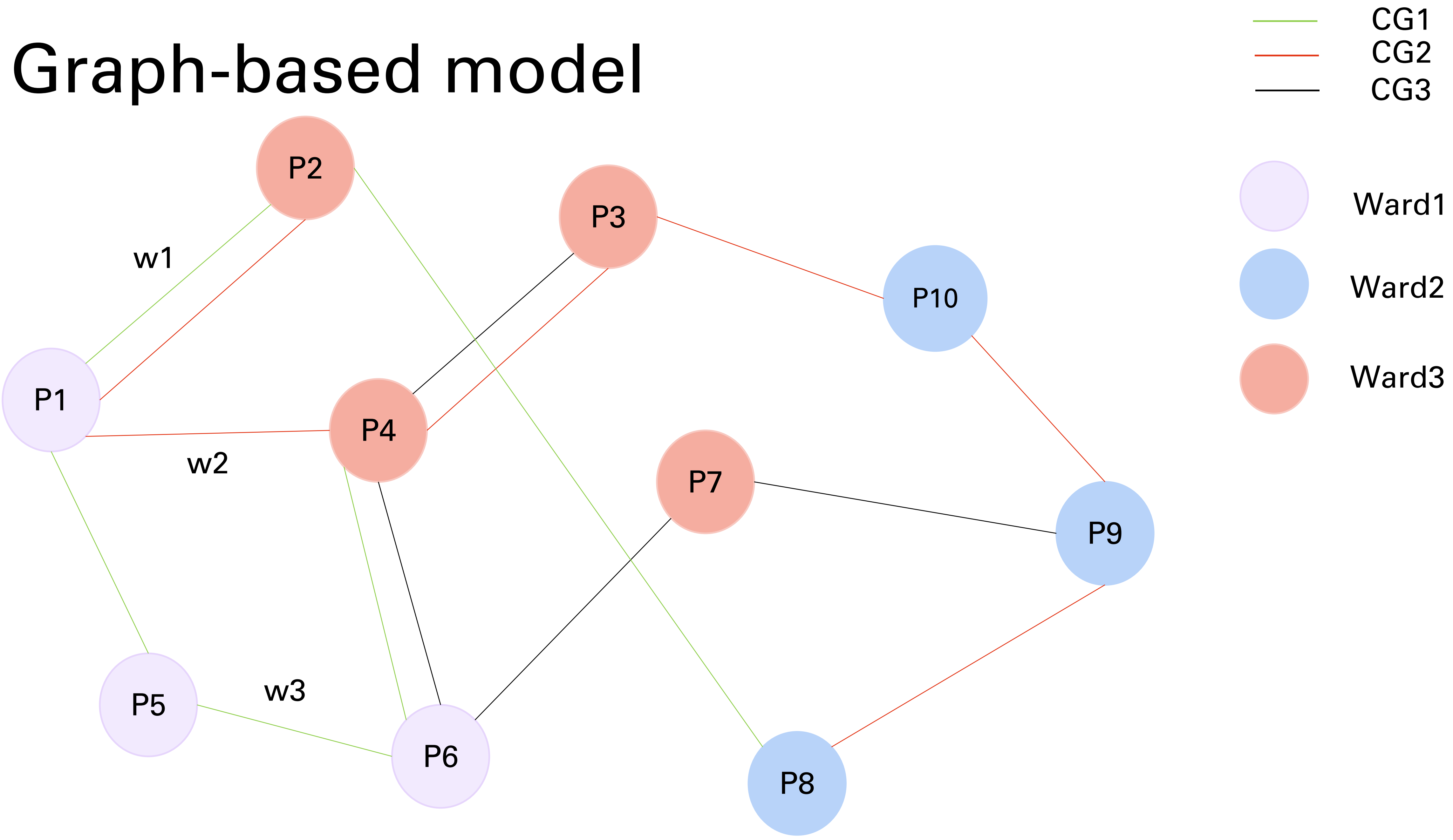
Results – Balanced (Optimal)

	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**AUC (%)**
Logistic regression	75.64	76.78	72.06	74.35	75.57
KNN	78.94	78.14	79.13	78.63	78.94
Random forest	73.08	72.25	73.12	72.68	73.08
Catboost	**80.67**	**79.15**	**82.17**	**80.63**	**80.70**

Next step and future work

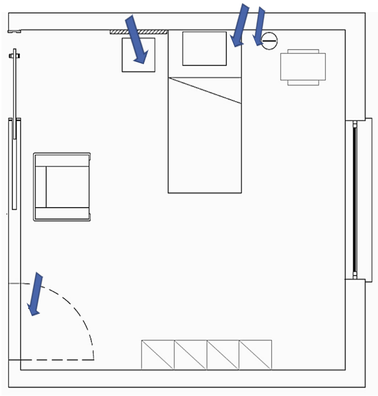
- Graph based-model
- Apply our model to long-term healthcare unit (LTHU) dataset
- Predict risk of infection in addition to the risk of colonization

Graph-based model

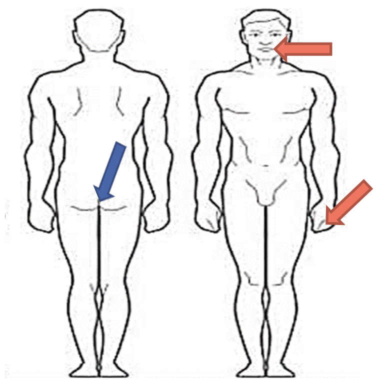


LTHU data

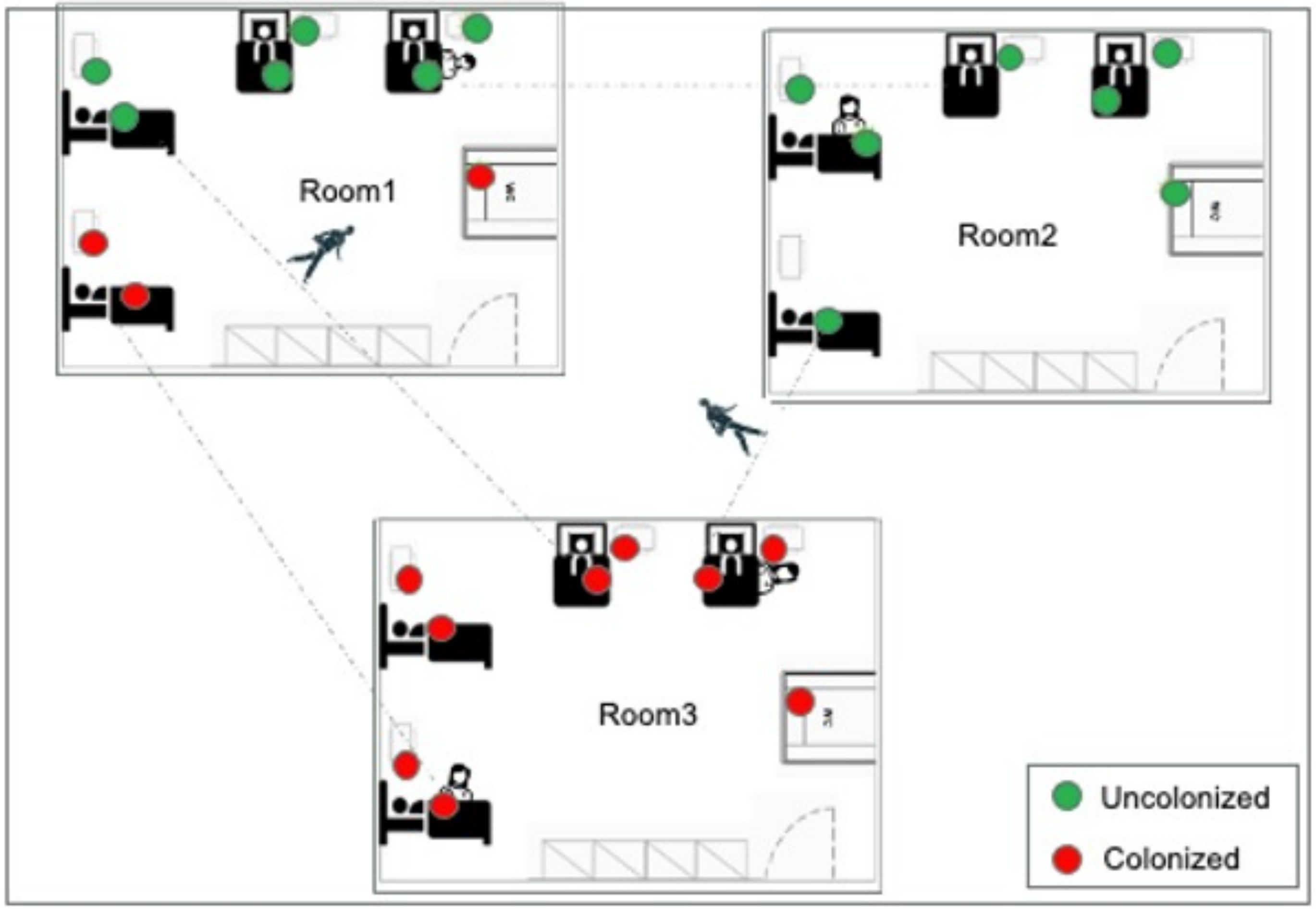
- Environmental samples
- Clinical samples
- Long-term healthcare unit (LTHU)



Environmental sampling



Clinical sampling



IPCpredict Team

University of Geneva

Douglas Teodoro

Racha Gouareb

Polytechnic of Leiria

Sónia G. Pereira

Catarina Marques
PhD student

Filipa Amorim
MSc student

THANK YOU