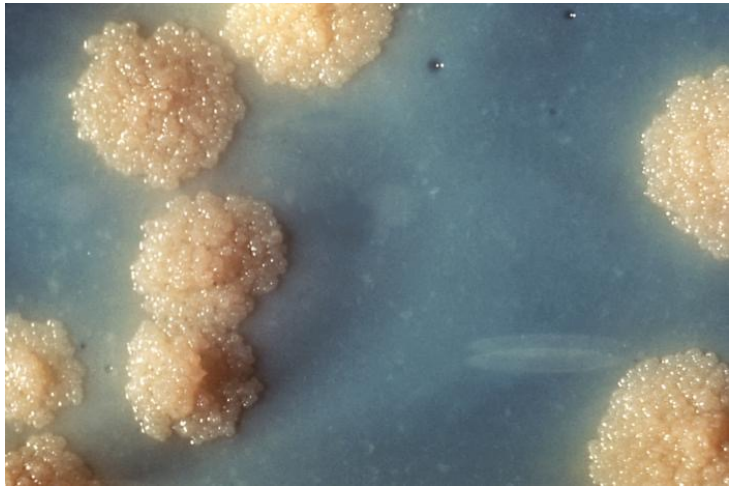
Explainability vs. interpretability in genotype-phenotype predictions: the case of antimicrobial resistance

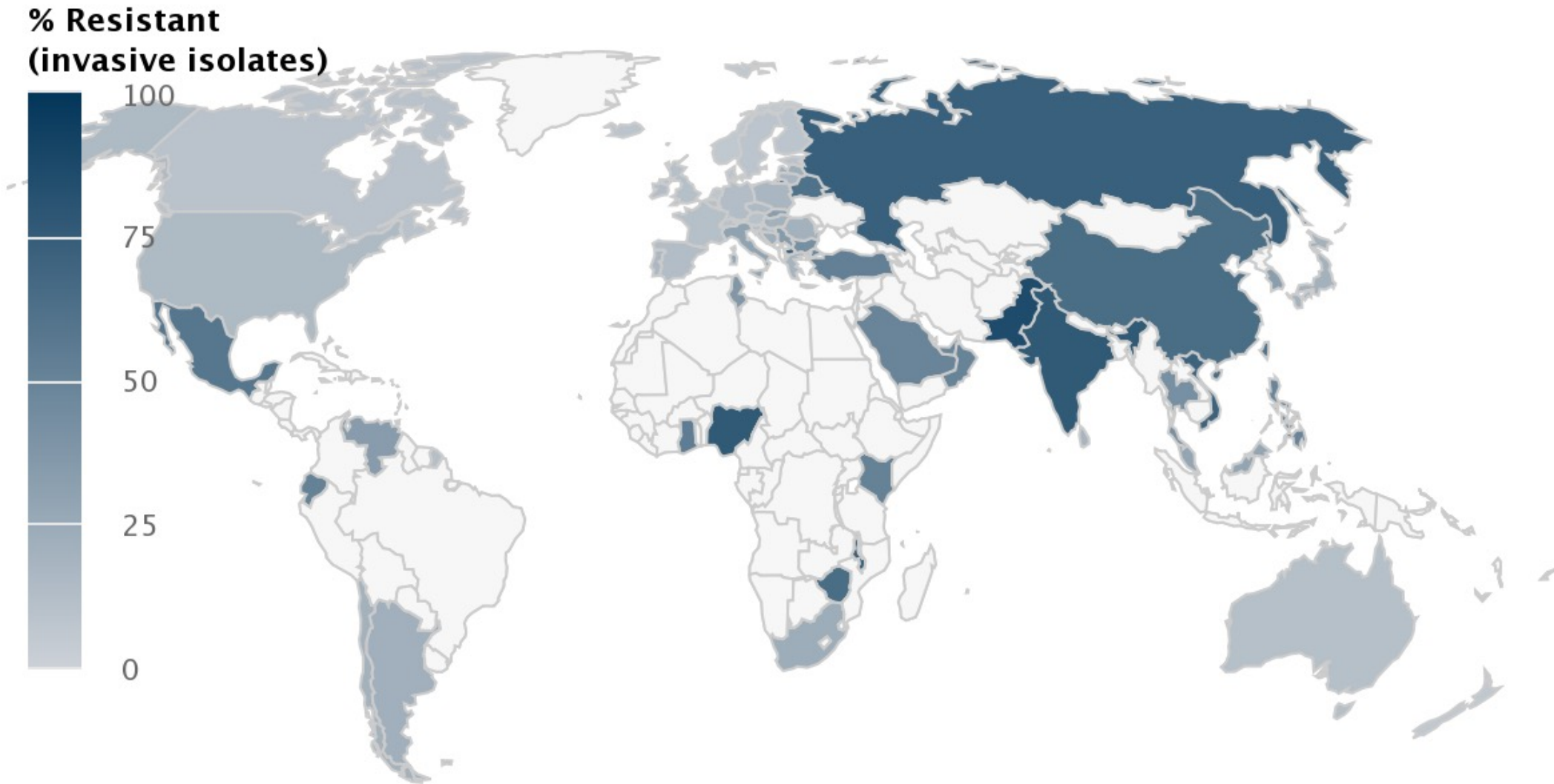


Leonid Chindelevitch
Infectious Disease Epidemiology
Imperial College London

Overview of antimicrobial resistance (AMR)

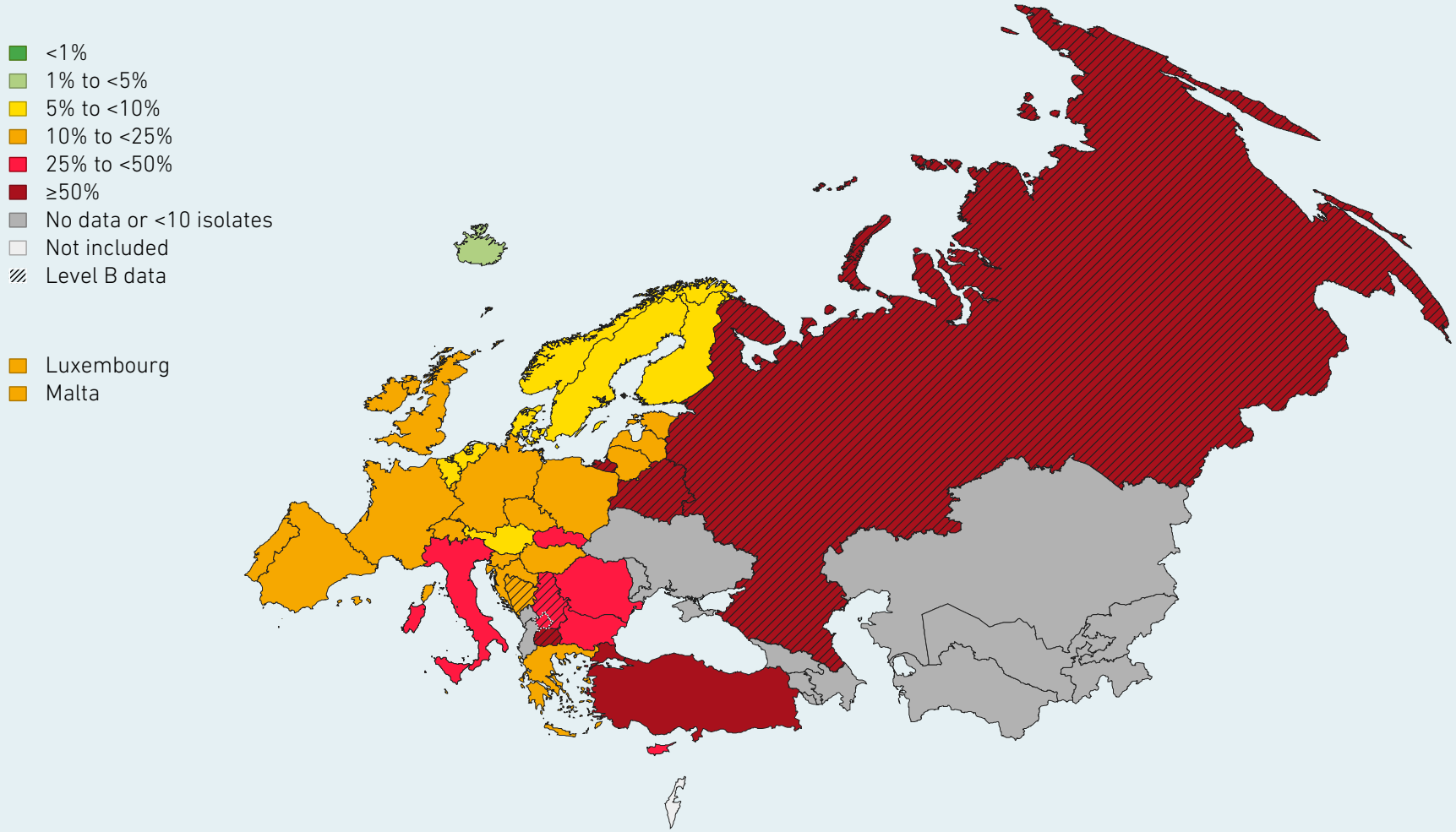
- Occurs when pathogenic microbes (bacteria, viruses, fungi, or protozoa) develop mechanisms to bypass the action of an antimicrobial drug
- Can be acquired, e.g. due to poor treatment adherence, or transmitted
- Many hospital-acquired infections are resistant, e.g. MRSA (methycillin-resistant *S. aureus*) and CRAB (carbapenem-resistant *A. baumannii*)
- AMR is projected to cause > 10M deaths annually by 2050 (O'Neill, '16)
- Its determinants are genetic, so can be studied via genome sequencing

Resistance of *Escherichia coli* to Cephalosporins (3rd gen)



Center for Disease Dynamics, Economics & Policy (cddep.org) © Natural Earth

Third-generation cephalosporin-resistant *E. coli* in the European Region (EARS-Net and CAESAR), 2015



Level B data: the data provide an indication of the resistance patterns present in clinical settings in the country, but the proportion of resistance should be interpreted with care. Improvements are needed to attain a more valid assessment of the magnitude and trends of AMR in the country. Levels of evidence are only provided for CAESAR countries and areas.

EARS-Net countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom.
CAESAR countries and areas: Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Georgia, Kazakhstan, Kyrgyzstan, Montenegro, Republic of Moldova, Russian Federation, Serbia, Switzerland, Tajikistan, The former Yugoslav Republic of Macedonia, Turkey, Turkmenistan, Ukraine, Uzbekistan and Kosovo (in accordance with United Nations Security Council resolution 1244(1999))

Data sources: 2015 data from the Central Asian and Eastern European Surveillance of Antimicrobial Resistance (CAESAR, ©WHO 2016) and 2015 data (data extracted from TESSy August, 2016 and not final) from the European Antimicrobial Resistance Surveillance Network (EARS-Net, ©ECDC 2016).

Potential goals of predicting AMR from WGS data

- **Ruling out** the use of particular drugs by predicting drug resistance
- **Ruling in** the use of particular drugs by predicting drug sensitivity
- **Monitoring** of novel or emerging resistance-associated mutations
- **Identifying** shortcomings in existing laboratory-based methods

Design criteria of a machine learning method

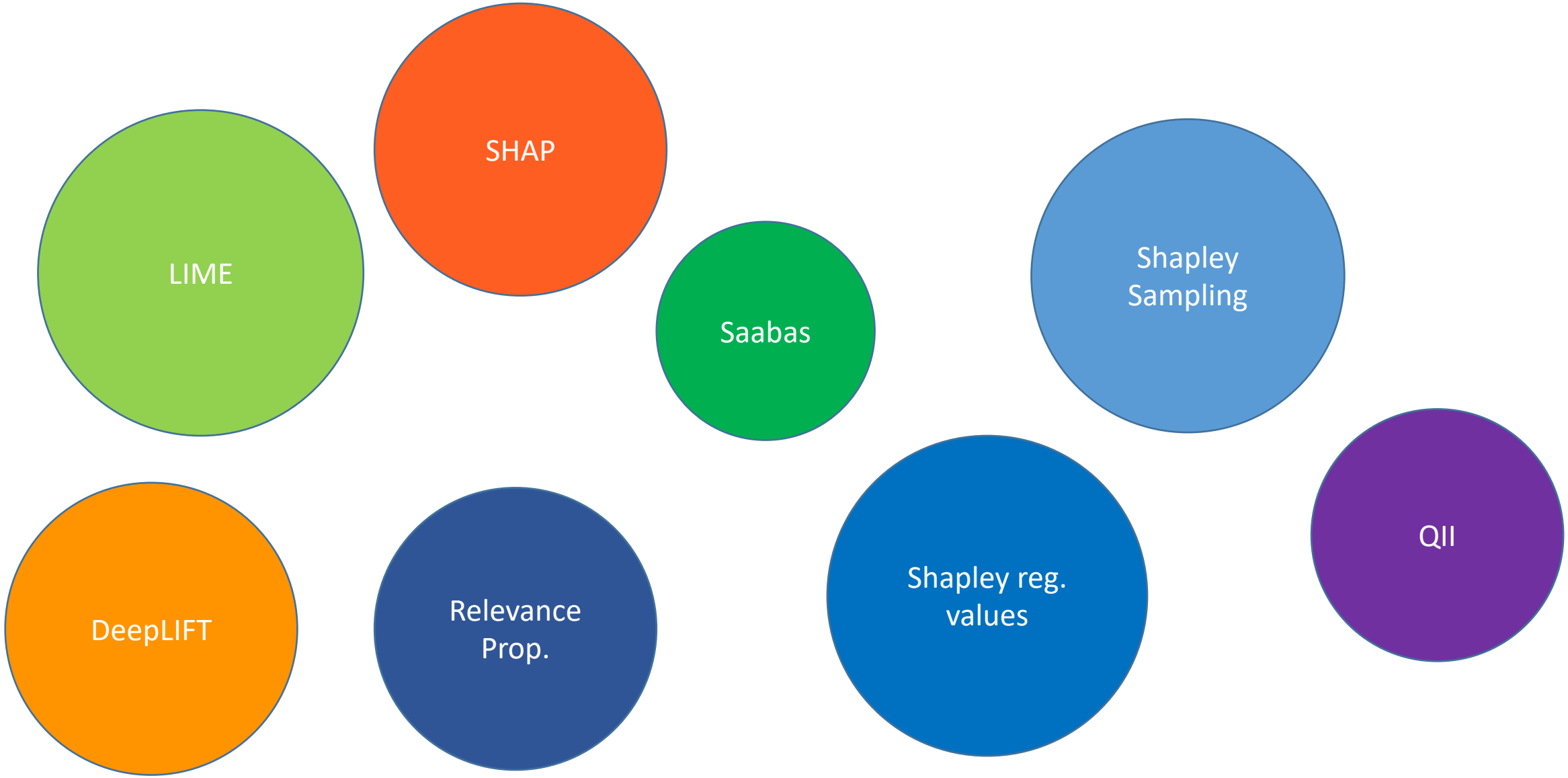
- **Accuracy**: we would like to get the most accurate prediction possible
- **Interpretability**: we would like to understand why this is the answer

  [past experiences have left biomedical practitioners mistrustful of ML]
- **Relevance**: we would like to identify only relevant (causal?) features
- **Reliability**: we would like to know how confident the predictions are

Black-box models can produce high accuracy…

- DeepAMR (Yang et al, 2019; Bioinformatics)
- Wide-and-deep (Chen et al, 2019; EBioMedicine)
- Multi-species (Aytan-Aktug et al, 2020; mSystems)
- WeightWatcher (Nguyen et al, 2021; SMU Data Science Review)
- LRCN (Safari et al, 2021; Proceedings of ACM-BCB – forthcoming)
- ResNet (Sedaghat et al, 2021; PhD thesis)

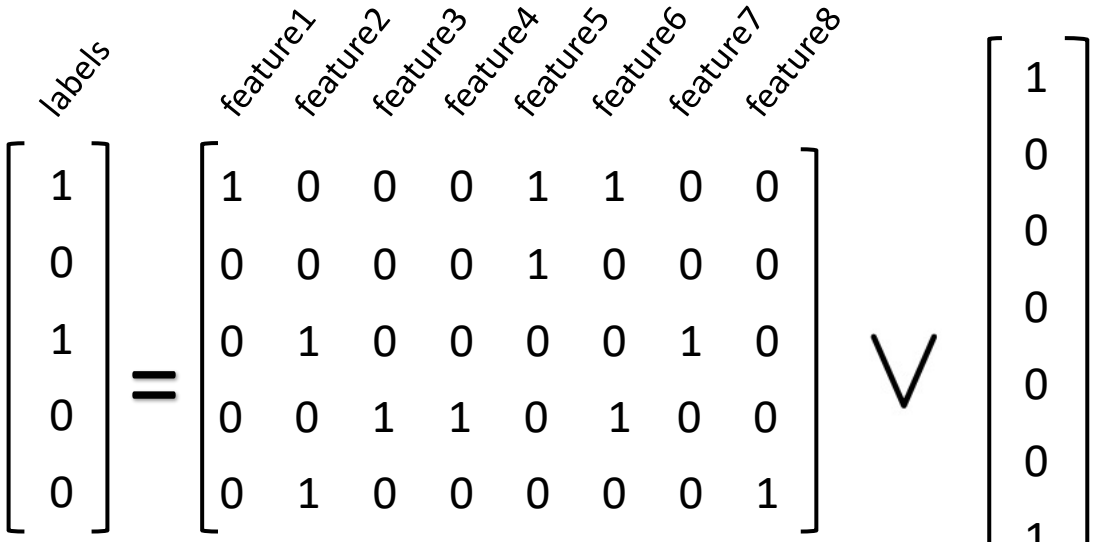
But which interpretation approach should we use?



The right classification for the wrong reason?

Ribeiro et al (2016). “Why Should I Trust You? Explaining the Predictions of Any Classifier”. NAACL Proceedings.

INGOT-DR: our interpretable rule-based classifier



```
If feature1 is active OR feature7 is active
Then predict the label is positive (here R)
```

$$\min \|w\|_0 \ \mathbf{s.t.}\ y = A \vee w,\ w \in \{0,1\}^n$$

Zabeti et al. **INGOT-DR: an interpretable classifier for predicting drug resistance in *Mycobacterium tuberculosis*** (2020). Available at https://www.biorxiv.org/content/10.1101/2020.05.31.115741v3 (journal version to appear in AlMoB shortly)

INGOT-DR produces a state-of-the-art performance

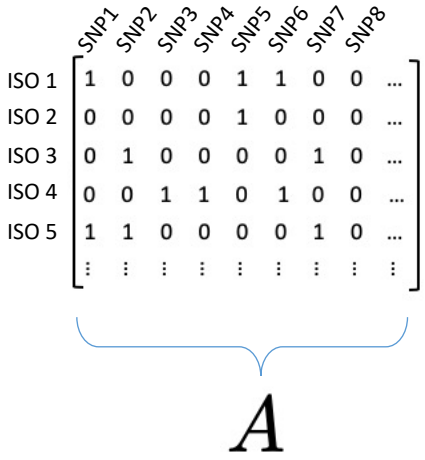
8000 Isolates

ReSeqTB
RELATIONAL SEQUENCING TB DATA PLATFORM

PATRIC
PATHOSYSTEMS RESOURCE INTEGRATION CENTER

Feature Matrix

Single-Nucleotide Polymorphisms (SNPs)



Group testing

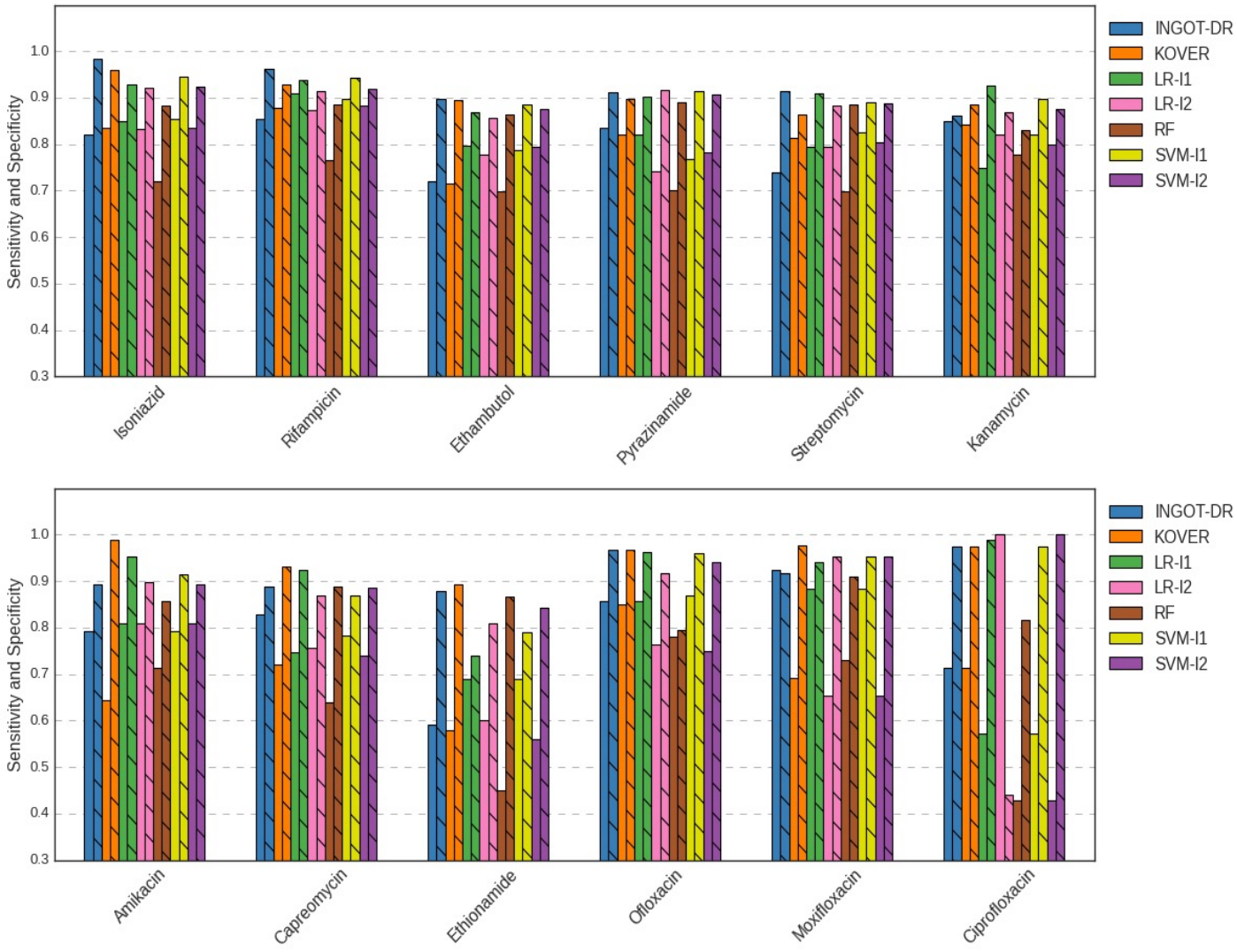
The goal is to select a **small** number of SNPs that yield a good sensitivity & specificity

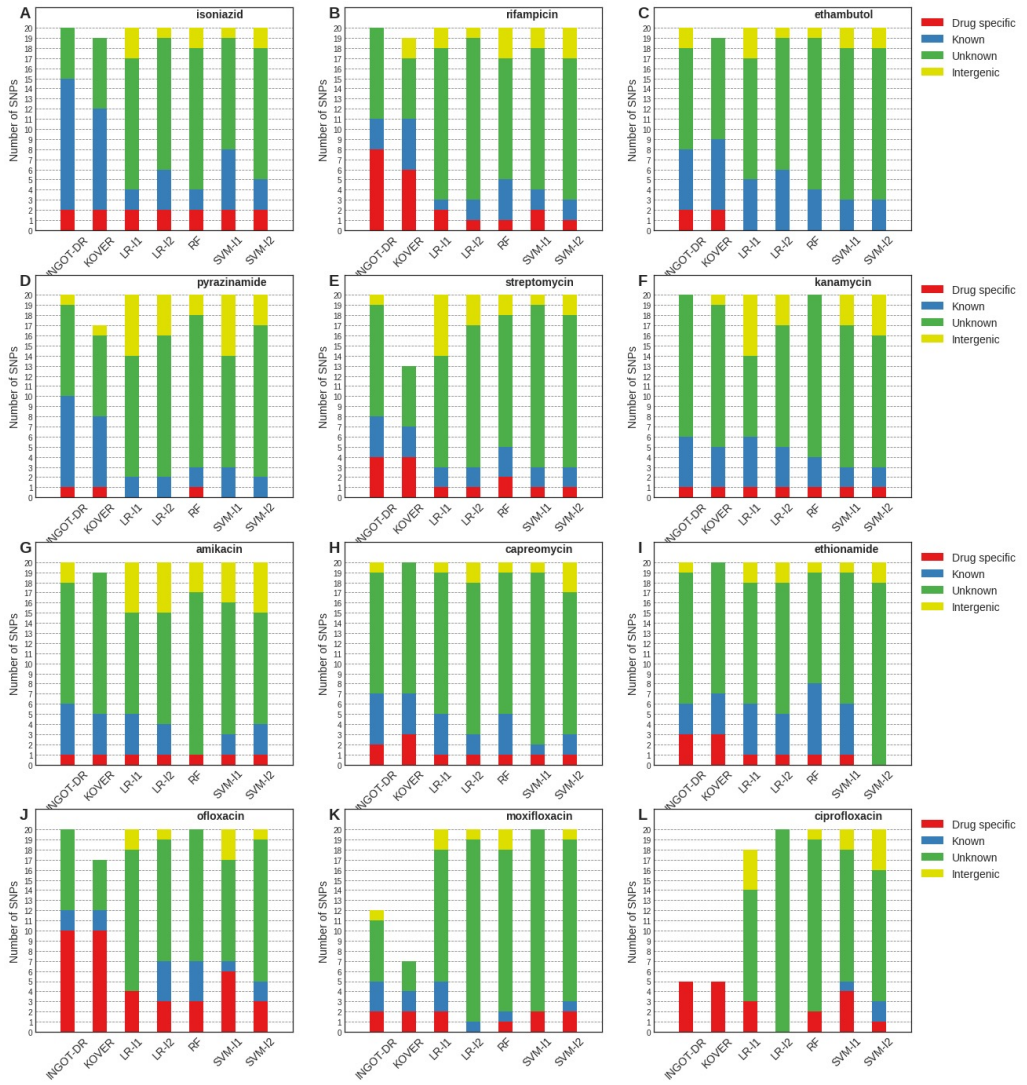
$$y = A \vee w \ , \quad w \in \{0,1\}^n$$

Example rule: **If** gyrA_A90V **OR** gyrA_S91P **OR** gyrA_D94A **OR** gyrA_D94G **OR** gyrA_D94Y **Then** resistant to ciprofloxacin

Zabeti et al. **INGOT-DR: an interpretable classifier for predicting drug resistance in *Mycobacterium tuberculosis*** (2020). Available at https://www.biorxiv.org/content/10.1101/2020.05.31.115741v3 (journal version to appear in AlMoB shortly)

INGOT-DR: better balanced accuracy **and** relevance

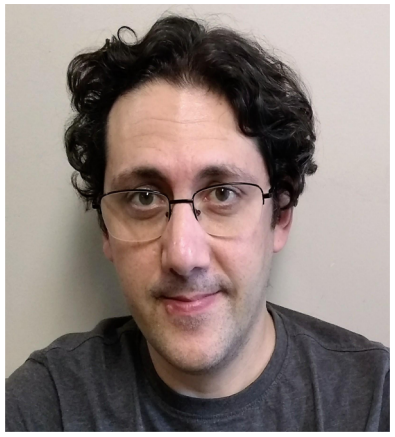




Summary of our contributions

- Developed INGOT-DR, an interpretable, flexible, and fast resistance prediction algorithm that designs **optimal rules** subject to constraints
- Applied it to 8,000 *M. tuberculosis* samples x 100,000 SNPs x 12 drugs obtaining **state-of-the-art accuracy** and identifying **relevant variants**
- Outperformed 6 other commonly used machine learning approaches
- The right tradeoff between prediction accuracy and interpretability: “how much accuracy are we willing to sacrifice for interpretability”?






Cedric Chauve

Maxwell Libbrecht

Pedro Feijao

Stanley Liang




Matthew Nguyen

Amir-Hosein Safari

Nafiseh Sedaghat

Hooman Zabeti




Nicholas Dexter

Niklas Stotzem

Fernando Guntoro

John Lees

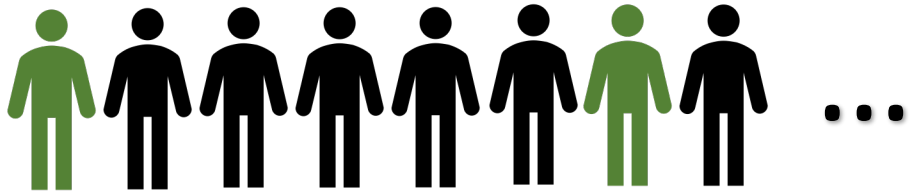
ALFRED P. SLOAN FOUNDATION
1934

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

GenomeCanada

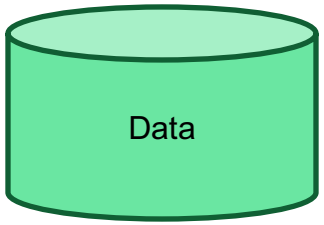
Group Testing

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & \dots \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \vee w$$



$$\min \|w\|_0 \ \textbf{s.t.}\ y = A \vee w,\ w \in \{0,1\}^n$$

Boolean Compressed Sensing



$\mathcal{D} = \{(x_1, y_1), \ldots, (x_m, y_m)\}$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \ldots \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & \ldots \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

A

$$y = A \vee w \ , \quad w \in \{0,1\}^n$$

$$\min \|w\|_0 \ \textbf{s.t.} \ y = A \vee w$$

$$w \in \{0,1\}^n$$

$$\begin{aligned} \min \quad & \sum_{j=1}^{n} w_j + \lambda \sum_{i=1}^{m} \epsilon_i \\ \text{subject to:} \quad & A_{\mathcal{P}} w + \epsilon_{\mathcal{P}} \geq 1 \\ & A_{\mathcal{Z}} w - \epsilon_{\mathcal{Z}} = 0 \\ & 0 \leq w_j \leq 1, \quad j = 1, \ldots, n \\ & 0 \leq \epsilon_i \leq 1, \quad i \in \mathcal{P} \quad \mathcal{P} = \{i | y_i = 1\} \\ & 0 \leq \epsilon_i, \quad i \in \mathcal{Z} \quad \mathcal{Z} = \{i | y_i = 0\} \end{aligned}$$

Our approach

$$\min \quad \sum_{j=1}^{n} w_j + \lambda \sum_{i=1}^{m} \xi_i$$

$$\text{s.t.} \quad w \in \{0,1\}^n$$

$$0 \le \xi_i \le 1, \quad i \in \mathcal{P}$$

$$0 \le \xi_i, \quad i \in \mathcal{Z}$$

$$A_{\mathcal{P}} w + \xi_{\mathcal{P}} \ge 1$$

$$A_{\mathcal{Z}} w - \xi_{\mathcal{Z}} = 0$$

$$\min \quad \sum_{j=1}^{n} w_j + \lambda_{\mathcal{P}} \sum_{i \in \mathcal{P}} \xi_i + \lambda_{\mathcal{Z}} \sum_{k \in \mathcal{Z}} \xi_k$$

$$\text{s.t.} \quad w \in \{0,1\}^n$$

$$0 \le \xi_i \le 1, \quad i \in \mathcal{P}$$

$$\xi_i \in \{0,1\}, \quad i \in \mathcal{Z}$$

$$A_{\mathcal{P}} w + \xi_{\mathcal{P}} \ge 1$$

$$A_{\mathcal{Z}} w - \xi_{\mathcal{Z}} \ge 0$$

$$m\xi_{\mathcal{Z}} - A_{\mathcal{Z}} w \ge 0$$

Our approach

$$\min \quad \sum_{j=1}^{n} w_j + \lambda \sum_{i=1}^{m} \xi_i$$

$$\text{s.t.} \quad w \in \{0,1\}^n$$

$$0 \le \xi_i \le 1, \quad i \in \mathcal{P}$$

$$0 \le \xi_i, \quad i \in \mathcal{Z}$$

$$A_{\mathcal{P}} w + \xi_{\mathcal{P}} \ge 1$$

$$A_{\mathcal{Z}} w - \xi_{\mathcal{Z}} = 0$$

$$\min \quad \sum_{j=1}^{n} w_j + \lambda_{\mathcal{P}} \sum_{i \in \mathcal{P}} \xi_i + \lambda_{\mathcal{Z}} \sum_{k \in \mathcal{Z}} \xi_k$$

$$\text{s.t.} \quad w \in \{0,1\}^n$$

$$0 \le \xi_i \le 1, \quad i \in \mathcal{P}$$

$$\xi_i \in \{0,1\}, \quad i \in \mathcal{Z}$$

$$A_{\mathcal{P}} w + \xi_{\mathcal{P}} \ge 1$$

$$A_{\mathcal{Z}} w - \xi_{\mathcal{Z}} \ge 0$$

$$m\xi_{\mathcal{Z}} - A_{\mathcal{Z}} w \ge 0$$

Our approach

FN

FP

$$\min \quad \sum_{j=1}^{n} w_j + \lambda_{\mathcal{P}} \sum_{i \in \mathcal{P}} \xi_i + \lambda_{\mathcal{Z}} \sum_{k \in \mathcal{Z}} \xi_k$$

$$\text{s.t.} \quad w \in \{0,1\}^n$$

$$0 \le \xi_i \le 1, \quad i \in \mathcal{P}$$

$$\xi_i \in \{0,1\}, \quad i \in \mathcal{Z}$$

$$A_{\mathcal{P}} w + \xi_{\mathcal{P}} \ge 1$$

$$A_{\mathcal{Z}} w - \xi_{\mathcal{Z}} \ge 0$$

$$m\xi_{\mathcal{Z}} - A_{\mathcal{Z}} w \ge 0$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} = 1 - \frac{\text{FP}}{\text{N}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} = 1 - \frac{\text{FN}}{\text{P}}$$

$$\sum_{i \in \mathcal{P}} \xi_i \le (1 - \text{Sensitivity})|\mathcal{P}|$$

$$\sum_{k \in \mathcal{Z}} \xi_k \le (1 - \text{Specificity})|\mathcal{Z}|$$

Our approach

$$\min \quad \sum_{j=1}^{n} w_j + \lambda_{\mathcal{P}} \sum_{i \in \mathcal{P}} \xi_i + \lambda_{\mathcal{Z}} \sum_{k \in \mathcal{Z}} \xi_k$$

$$\begin{aligned}
\text{s.t.} \quad & w \in \{0,1\}^n \\
& 0 \leq \xi_i \leq 1, \quad i \in \mathcal{P} \\
& \xi_i \in \{0,1\}, \quad i \in \mathcal{Z} \\
& A_{\mathcal{P}} w + \xi_{\mathcal{P}} \geq 1 \\
& A_{\mathcal{Z}} w - \xi_{\mathcal{Z}} \geq 0 \\
& m\xi_{\mathcal{Z}} - A_{\mathcal{Z}} w \geq 0
\end{aligned}$$

$$\min \quad \sum_{j=1}^{n} w_j + \lambda_{\mathcal{P}} \sum_{i \in \mathcal{P}} \xi_i$$

$$\begin{aligned}
\text{s.t.} \quad & w \in \{0,1\}^n \\
& 0 \leq \xi_i \leq 1, \quad i \in \mathcal{P} \\
& \xi_i \in \{0,1\}, \quad i \in \mathcal{Z} \\
& A_{\mathcal{P}} w + \xi_{\mathcal{P}} \geq 1 \\
& A_{\mathcal{Z}} w - \xi_{\mathcal{Z}} \geq 0 \\
& m\xi_{\mathcal{Z}} - A_{\mathcal{Z}} w \geq 0 \\
& 1^T \xi_{\mathcal{Z}} \leq (1 - \bar{t})|\mathcal{Z}|
\end{aligned}$$