Nowcasting of fecal coliforms presence using an artificial neuron network in India

Asaf Pras

Advisor: Prof. Hadas Mamane

Water-tech lab

Research motivation - Global safe water shortage

- **2 billion**

  Lack safely managed drinking water

- **771 million**

  Without basic drinking water services

- **159 million**

  Drink water directly from surface water (streams or lakes)

- **800 children**

  Under 5 years of age die daily because of using unsafe drinking water

Research motivation - Global safe water shortage

Population expansion, urbanization and industrialization

Increased effluent production

Insufficient sewage infrastructure and water treatment plants

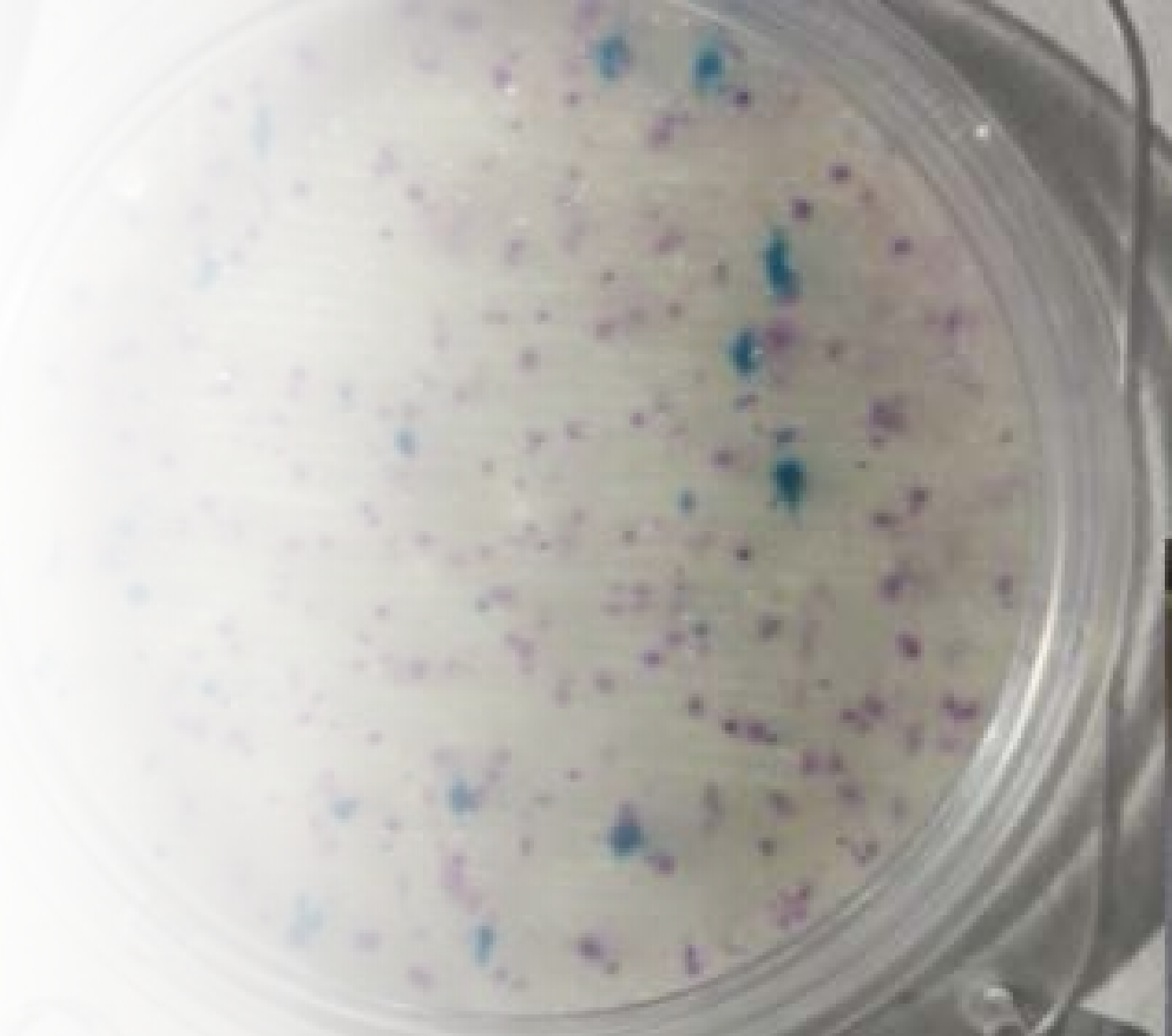
Contaminated water sources

Drinking water quality monitoring

- Water quality should be routinely analyzed
- Many kinds of parameters can be monitored in real-time
- Accurate information of biological contamination is an integral part of water systems management

Fecal coliforms monitoring

- Fecal coliforms (FC) or Escherichia coliform should be tested
- WHO standard is 0 MPN/100ml
- Traditional monitoring test takes at least 18-24h
- In rural communities the test routine is infrequent – up to 6 months
- Real-time monitoring is expansive

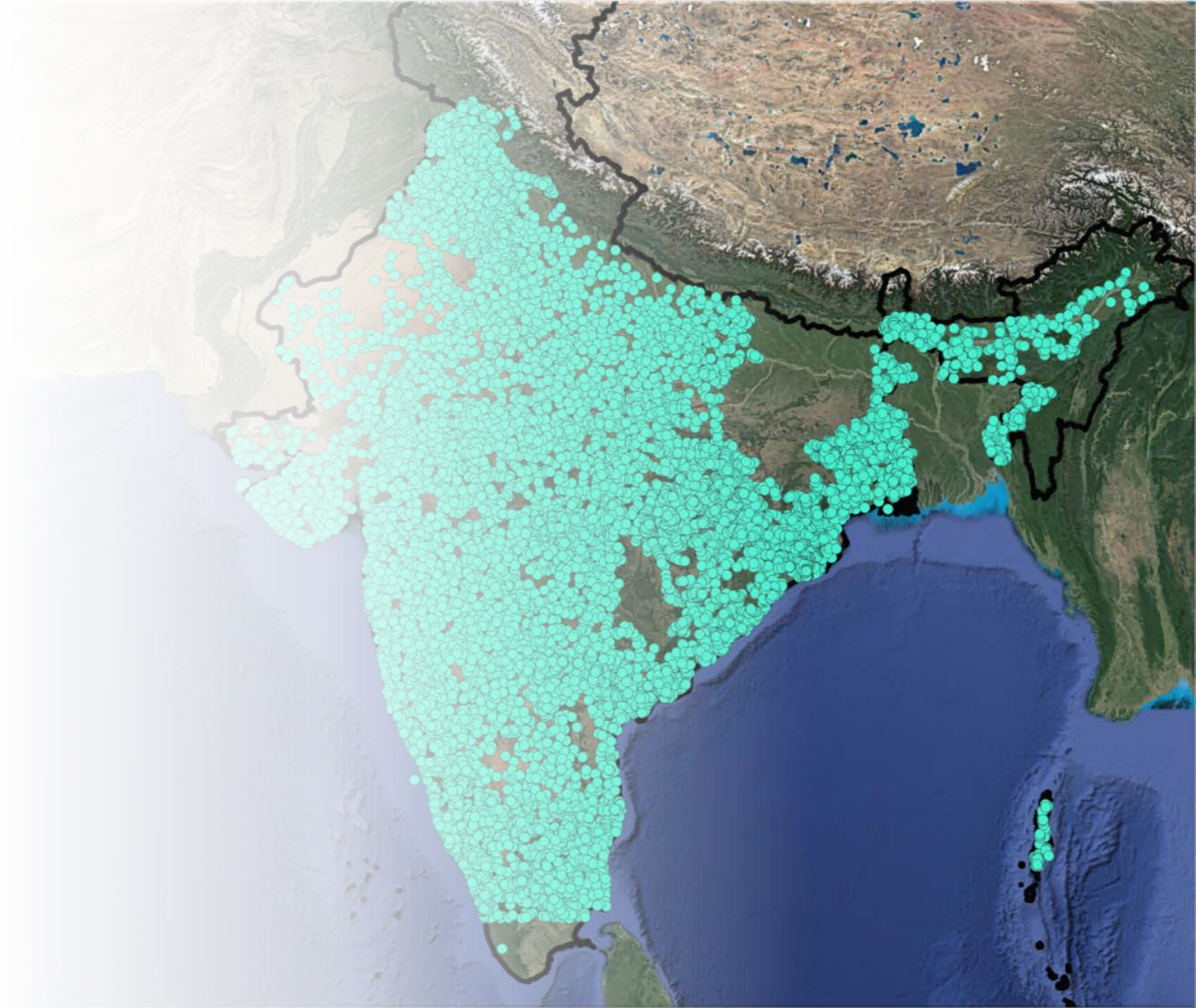


The Goal

- Find a cheap, real-time and in-situ solution for FC monitoring

The Data

- Was collected by the Central Pollution Control Board of India
- From throughout India
- Various kinds of sources
- A large number of parameters

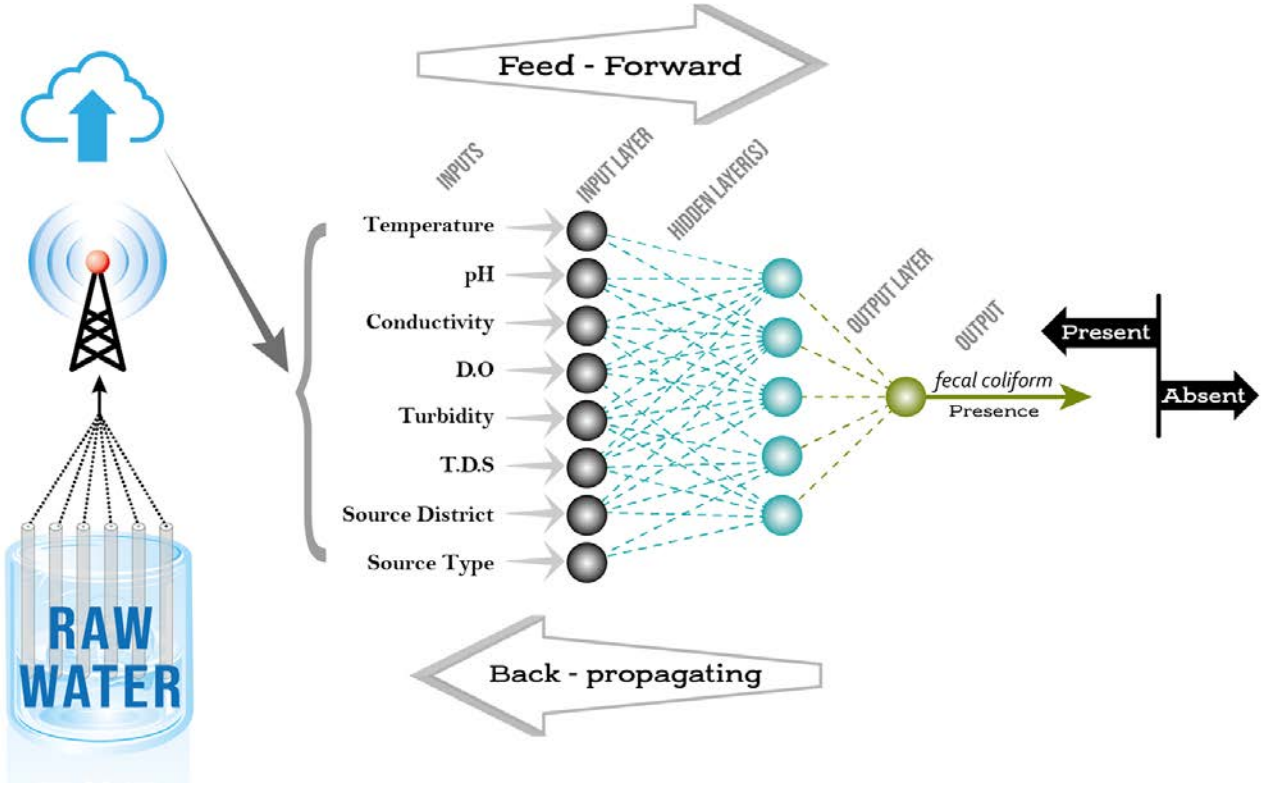


Our solution

- FC sensor based on machine learning or artificial intelligence, which uses easy-to-monitor, real-time and *in situ* parameters as inputs
- Gives a ‘nowcast’ of FC absence or presence in water sources
- Can be later upgraded for more parameters

Method

- Using a multilayer perceptron model
- The model is fed in with the following low-cost and easy to monitor parameters:
  - Temperature
  - pH
  - Electrical conductivity
  - Total dissolved solids
  - Dissolved oxygen
  - Turbidity
  - Geographical parameters (Type of source, and location)



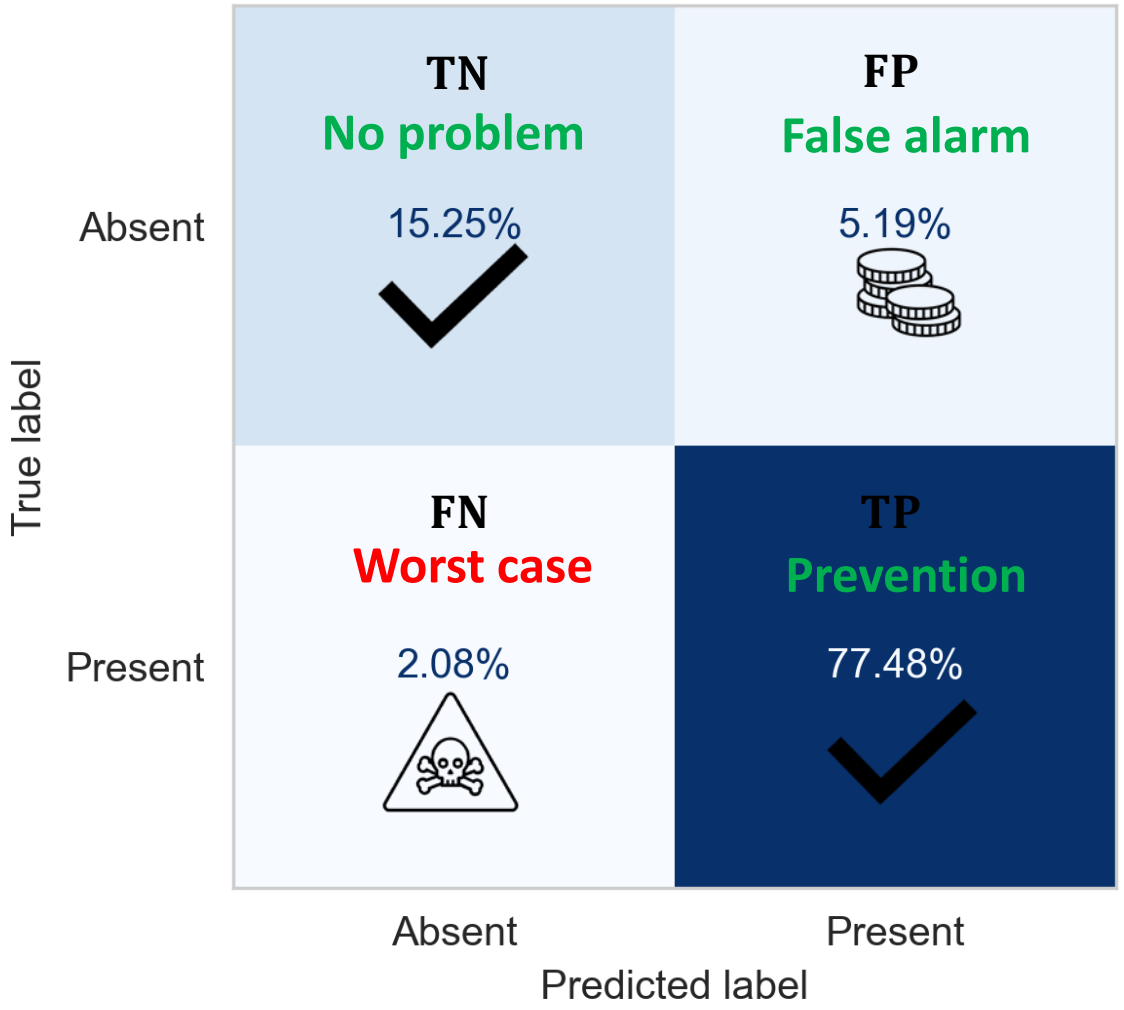
Results – Confusion matrix

$$\textbf{Accuracy} = \textbf{TP} + \textbf{TN} = 92\%$$

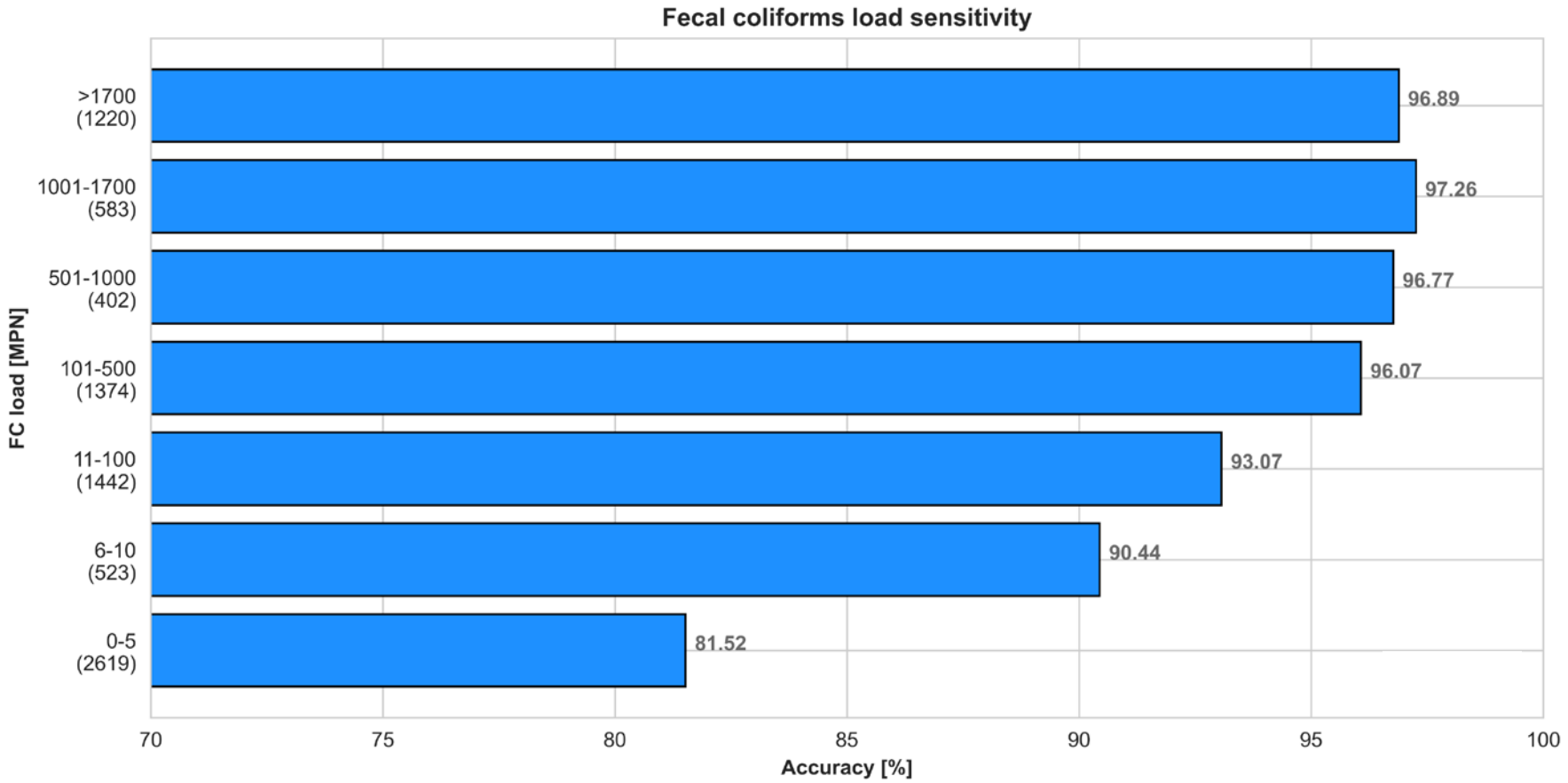
$$\textbf{Sensitivity} = \frac{\textbf{TP}}{\textbf{FN} + \textbf{TP}} = 98\%$$

$$\textbf{Precision} = \frac{\textbf{TP}}{\textbf{FP} + \textbf{TP}} = 93\%$$

$$\textbf{Specificity} = \frac{\textbf{TN}}{\textbf{FP} + \textbf{TN}} = 75\%$$

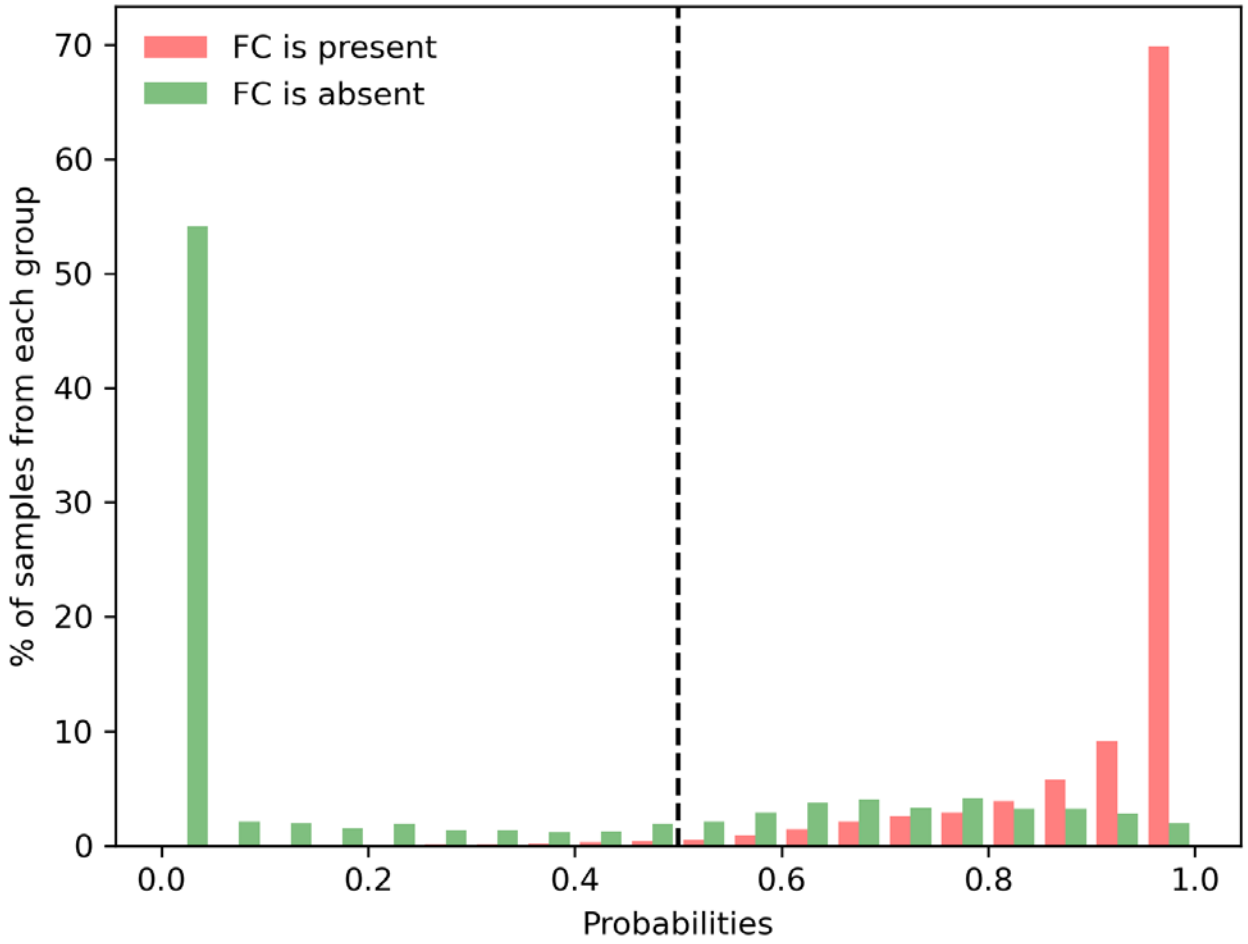


Results –
Biological load effect on sensitivity



Results – Probability histogram

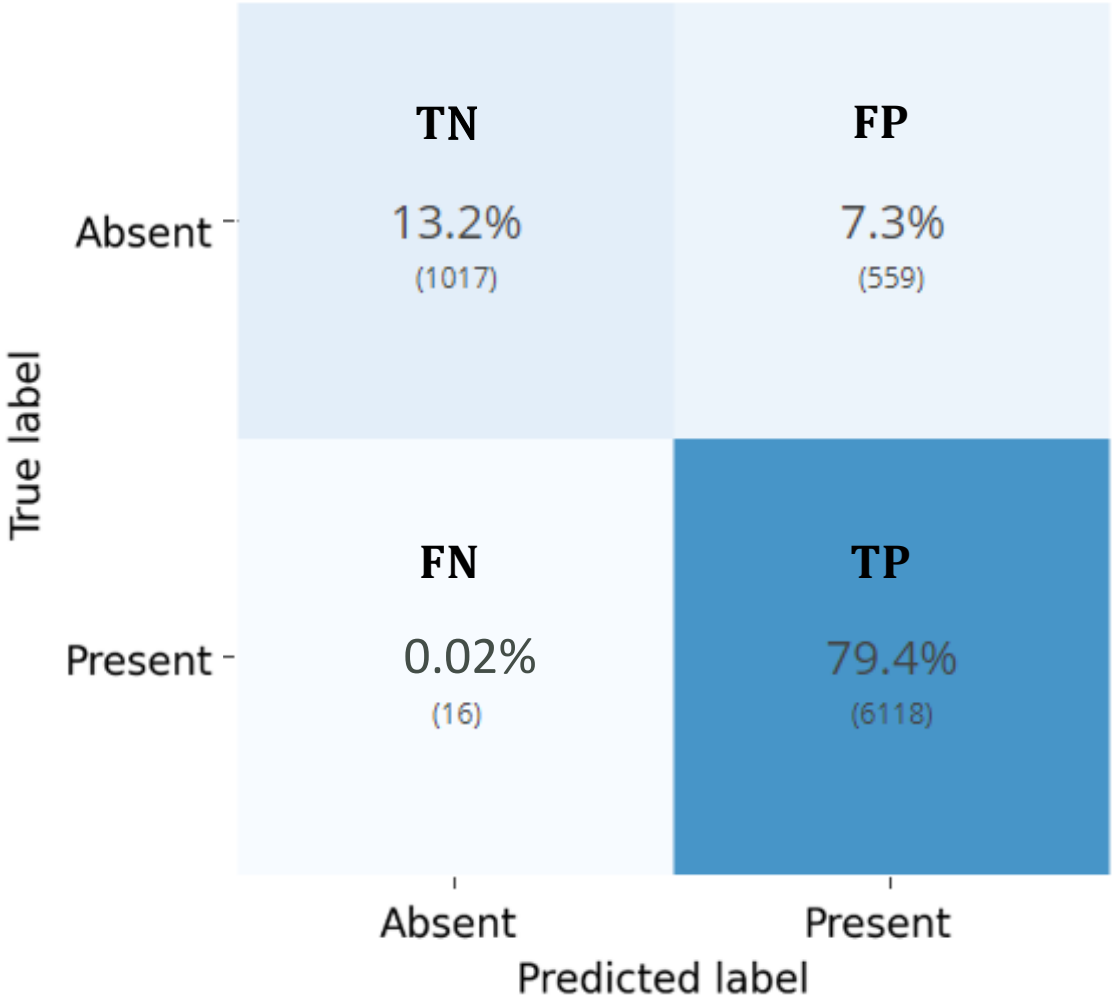
- The probability of a sample to be present



Results –
Sensitivity–specificity–precision tradeoffs

- when raising the cutoff to 0.75:
  - Accuracy: 90.56%
  - Sensitivity: 92.66%
  - Precision: 95.34%
  - Specificity: 82.36%

- when lowering the cutoff to 0.25:
  - Accuracy: 92.54%
  - Sensitivity: 99.74%
  - Precision: 91.63%
  - Specificity: 64.53%



Summary

- We showed the feasibility of nowcasting FC contamination:
  - About 93% of accuracy
  - 98% of sensitivity
  - 93 of precision
- By moving the threshold, we can “trade” sensitivity with precision

Thanks for listening
Drink water and keep safe