13th INTERNATIONAL CONFERENCE

TYPHOID & OTHER INVASIVE SALMONELLOSES

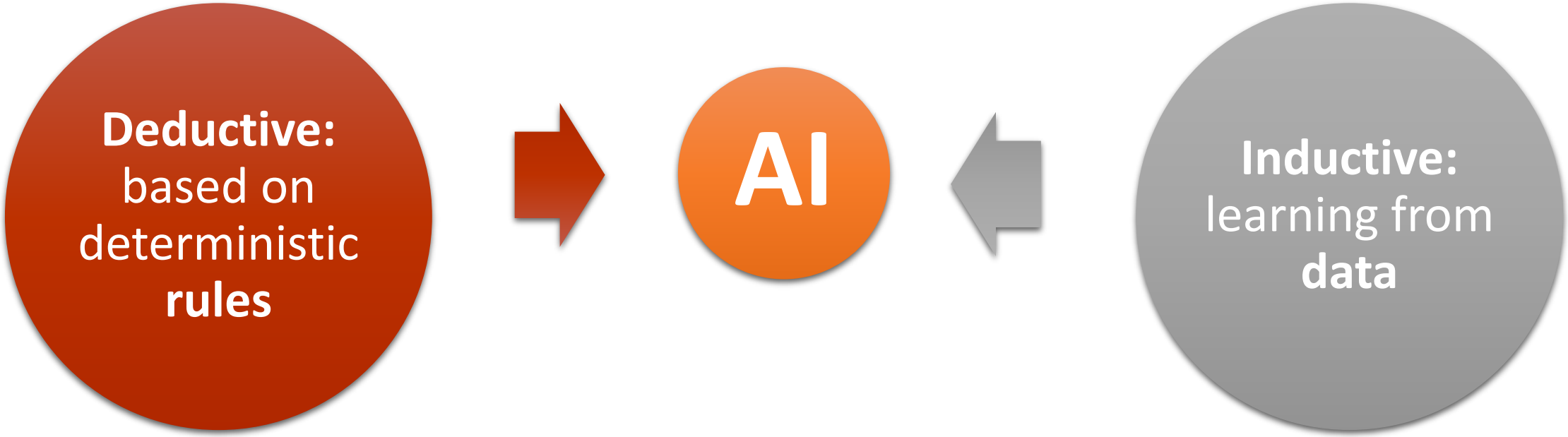
December 5-7, 2023 | Kigali, Rwanda

Integrating artificial intelligence to analyze typhoid & other invasive salmonelloses data: from zero to hero!

Bráulio RGM Couto

BIOBYTE
TECNOLOGIA EM EPIDEMIOLOGIA

Artificial Intelligence (AI)



- Rule-Based Systems
- Knowledge Graphs
- Programming in Logic - Prolog
- Ontology-Based Systems

- Linear and Logistic Regression
- Decision Trees
- Neural Networks, SVM, LSI, NLP
- Probabilistic models

Ganesh, V., Seshia, S.A. & Jha, S. **Machine learning and logic: a new frontier in artificial intelligence**. Form Methods Syst Des 60, 426–451 (2022). https://doi.org/10.1007/s10703-023-00430-1

Artificial Intelligence (AI)

- If you analyze a database employing advanced or basic methods and document the findings, including models, in a paper or report **it doesn't qualify as AI**.
- However, if you integrate rules and/or analysis results, even from a basic assessment, into an application for decision-making, **that constitutes AI.**

Inductive AI applied to typhoid & other invasive salmonelloses

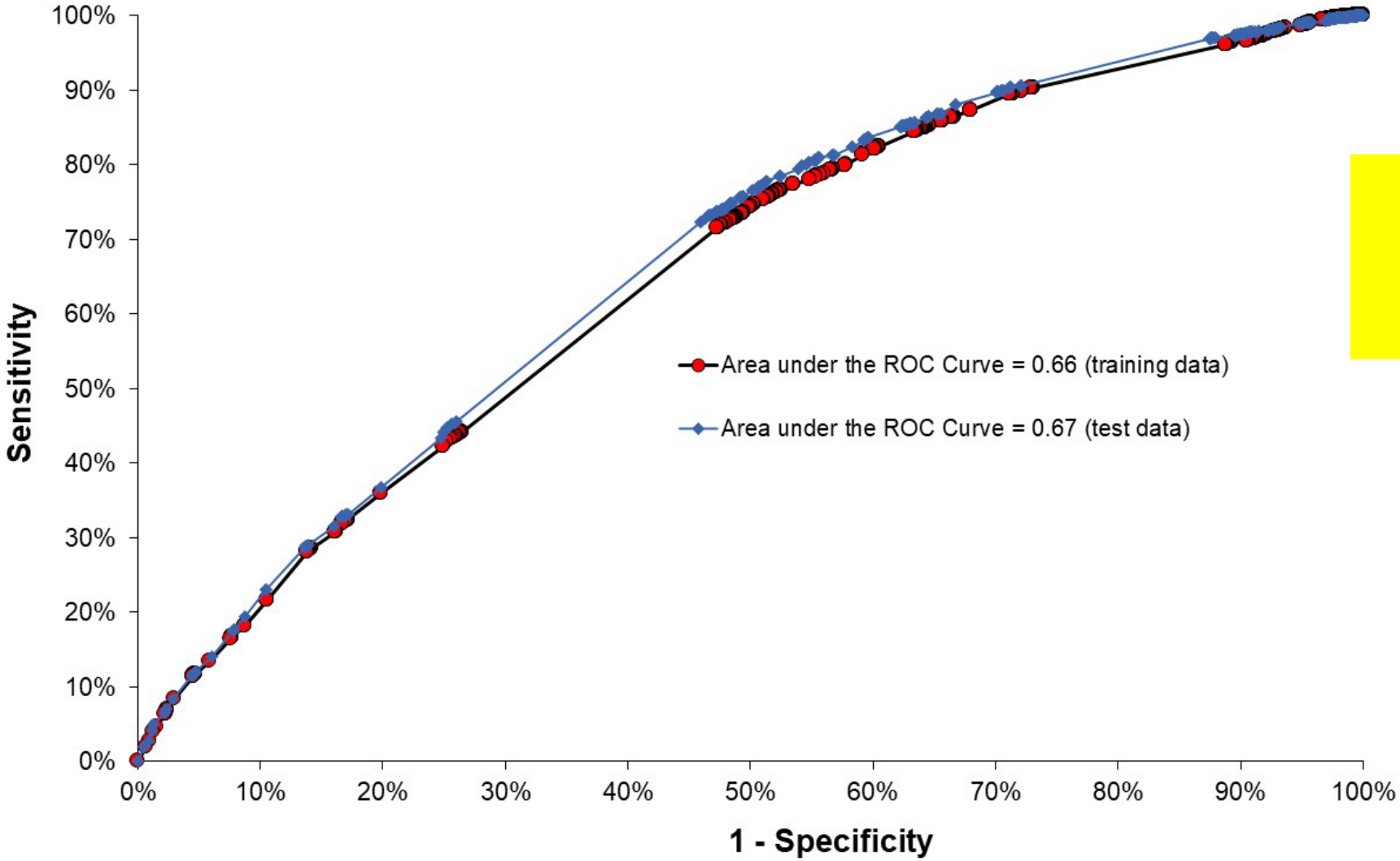
Using logistic regression model

1. How to identify typhoid/paratyphoid enteric fever without culture?

Variable	Logistic coefficient	S.E.	Odds Ratio	[95% C.I.]	p-value
Patient had diarrhea ((≥3 watery stools/day)	0.515	0.041	1.67	[1.5; 2.2]	< 0.001
Patient had vomiting	0.077	0.034	1.08	[1.5; 2.2]	0.025
Patient had headache	0.369	0.036	1.45	[1.1; 1.6]	< 0.001
Patient reports fever	0.514	0.168	1.67	[1.1; 1.6]	0.002
Patient had abdominal pain	0.087	0.038	1.09	[1.5; 2.2]	0.022
Patient had cough	-0.904	0.035	0.40	[1.0; 3.0]	< 0.001
Patient had shortness of breath / difficulty breathing	-0.416	0.060	0.66	[1.1; 1.6]	< 0.001
Patient had constipation	-0.485	0.075	0.62	[1.1; 1.6]	< 0.001
Patient had rash	-0.426	0.097	0.65	[1.5; 2.2]	< 0.001
Patient had chills	-0.406	0.126	0.67	[1.1; 1.6]	0.001
Patient had bodyaches	-0.650	0.170	0.52	[1.5; 2.2]	< 0.001
Patient had sore throat	-1.759	0.392	0.17	[1.1; 1.6]	< 0.001
Patient had arthralgia / joint pain	-1.076	0.278	0.34	[1.5; 2.2]	< 0.001
Patient had myalgia / muscle pain	-1.314	0.605	0.27	[1.1; 1.6]	0.030
Patient had burning micturition / urination	-0.574	0.277	0.56	[1.5; 2.2]	0.038
Patient had chest pains	-0.985	0.404	0.37	[1.1; 1.6]	0.015
Constant	-1.428				

Database: Garrett et al., **Incidence of typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan: results of the Surveillance for Enteric Fever in Asia Project**. www.thelancet.com/lancetgh Vol 10 July 2022

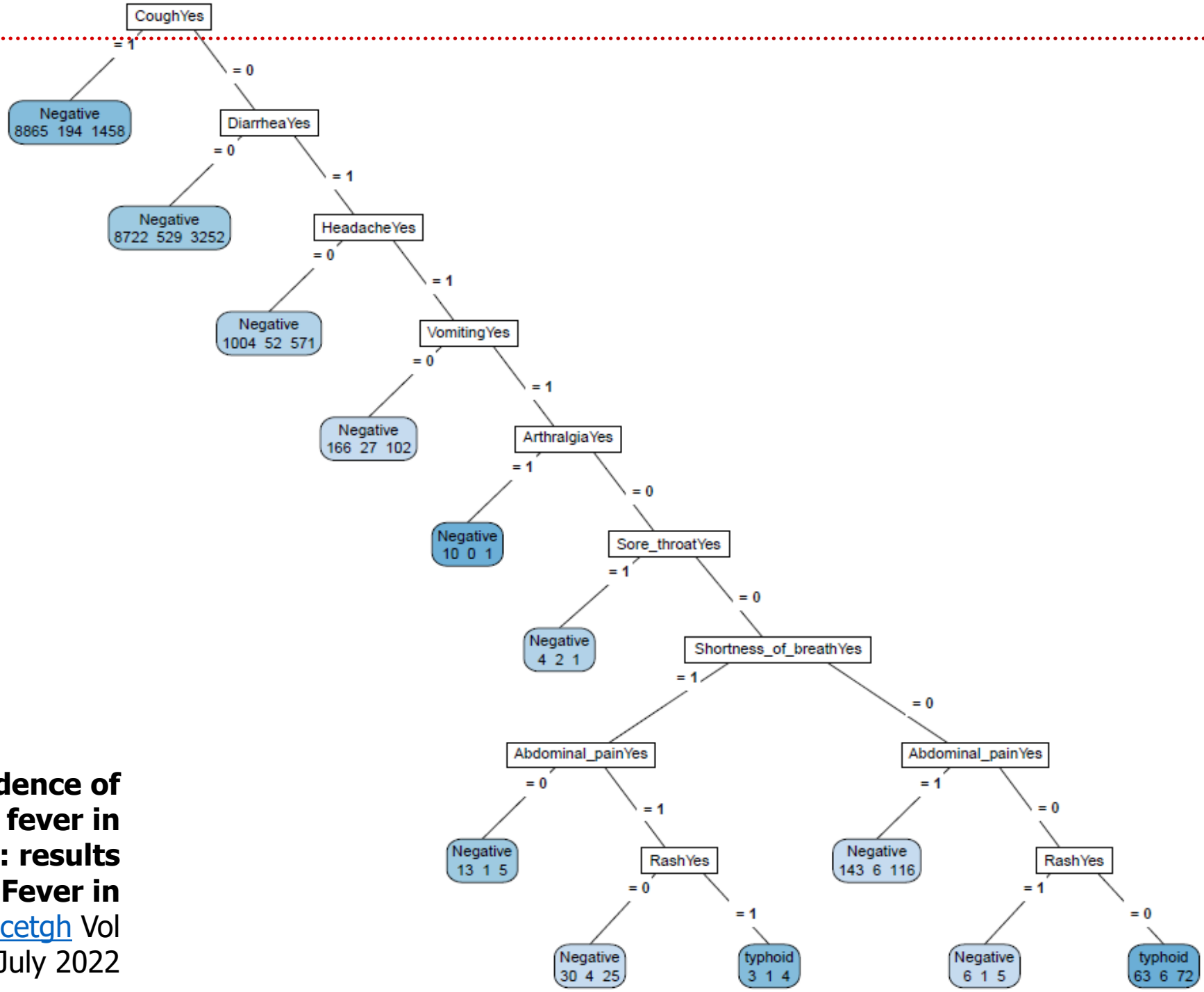
Performance of the logistic regression model in identifying typhoid/paratyphoid cases without culture?



EnteriRisk Model

Using decision tree

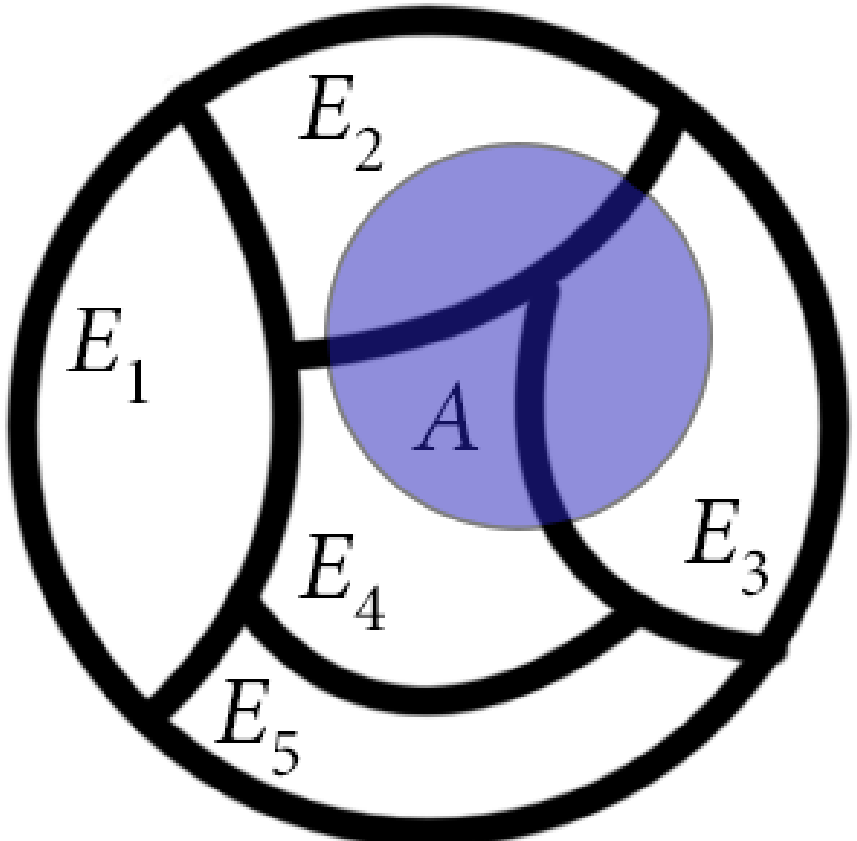
1. How to identify typhoid/paratyphoid enteric fever without culture?



Database: Garrett et al., **Incidence of typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan: results of the Surveillance for Enteric Fever in Asia Project**. www.thelancet.com/lancetgh Vol 10 July 2022

Empiric Antibiotic Therapy: a Probabilistic Approach

2. How to start optimal antibiotic therapy before the results of cultures and antimicrobial susceptibility tests are available?



Law of total probability: success rate of a monotherapy.

$$P(A) = \sum_{i=1}^{n} P(E_i \cap A) = \sum_{i=1}^{n} P(E_i) \times P(A/E_i)$$

COUTO et al., **Optimizing Empiric Antibiotic Therapy: a Probabilistic Approach**. Open Forum Infectious Diseases, 2022https://doi.org/10.1093/ofid/ofac492.267

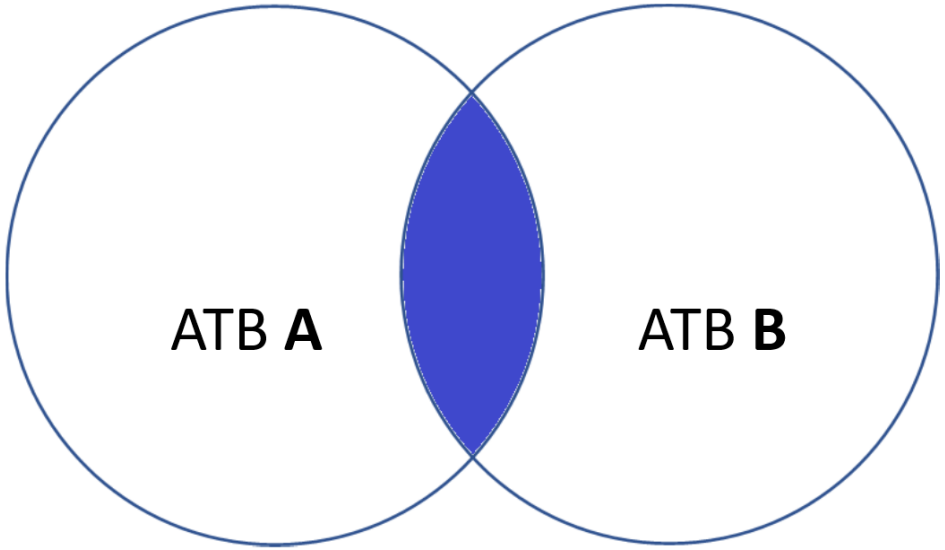
SONG et al., **Clinical profiles and antimicrobial resistance patterns of invasive Salmonella infections in children in China**. European Journal of Clinical Microbiology & Infectious Diseases (2022) 41:1215–1225. https://doi.org/10.1007/s10096-022-04476-7

		Salmonella enterica serotype			
		Salmonella Typhi	*Salmonella Paratyphi*	*Nontyphoidal Salmonella*	Total
	Frequency	**11**	5	114	130
Probability of a specific serotype =		8%	4%	88%	100%
Ampicillin	n	10	5	110	**44%**
	S	50%	60%	43%	
Piperacillin	n	8	0	64	**52%**
	S	50%	0%	55%	
Ampicillin/ sulbactam	n	5	0	106	**47%**
	S	80%	0%	46%	
Piperacillin/ tazobactam	n	11	0	106	**94%**
	S	100%	0%	97%	
Cefotaxime	n	11	5	111	**90%**
	S	100%	80%	90%	
Ceftriaxone	n	11	5	111	**90%**
	S	100%	80%	90%	
Ceftazidime	n	10	5	111	**93%**
	S	100%	80%	93%	
Cefepime	n	11	5	110	**94%**
	S	91%	100%	94%	
Aztreonam	n	11	0	109	**90%**
	S	100%	0%	93%	
Gentamicin	n	5	0	108	**79%**
	S	100%	0%	81%	

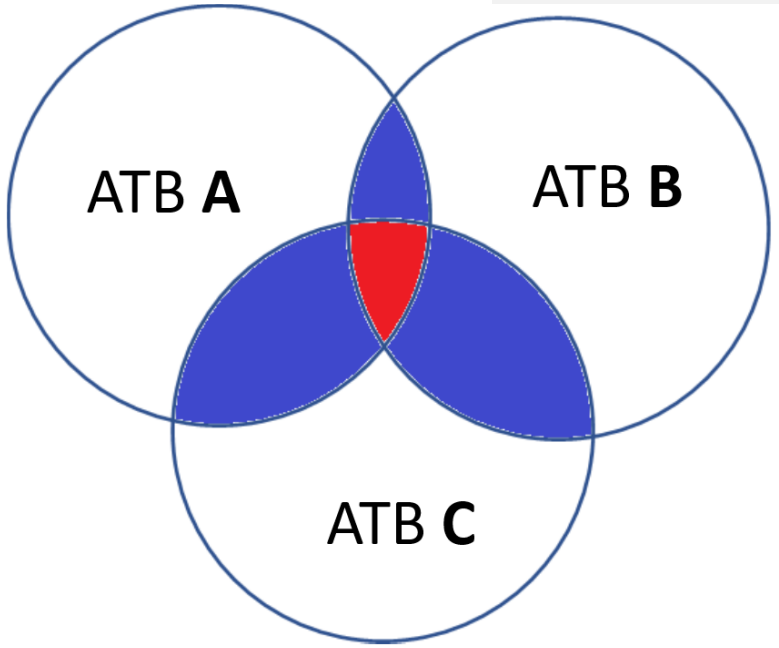
		Salmonella enterica serotype			
		Salmonella Typhi	*Salmonella Paratyphi*	*Nontyphoidal Salmonella*	Total
	Frequency	**11**	5	114	130
Probability of a specific serotype =		8%	4%	88%	100%
Tobramycin	n	5	0	108	**79%**
	S	100%	0%	81%	
Amikacin	n	5	0	105	**87%**
	S	100%	0%	89%	
Imipenem	n	11	0	108	**96%**
	S	100%	0%	100%	
Meropenem	n	8	0	88	**96%**
	S	100%	0%	100%	
Tetracycline	n	2	0	75	**68%**
	S	100%	0%	68%	
Ciprofloxacin	n	9	3	98	**88%**
	S	100%	33%	89%	
Levofloxacin	n	9	3	98	**88%**
	S	100%	33%	89%	
Sulfamethoxazole/ trimethoprim	n	11	5	109	**77%**
	S	91%	80%	75%	
Chloramphenicol	n	5	0	75	**80%**
	S	80%	0%	84%	

Empiric Antibiotic Therapy: a Probabilistic Approach

Success probability of two or three antibiotics: the sensitivity to an antimicrobial is considered independent of sensitivity to any other



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

COUTO et al., **Optimizing Empiric Antibiotic Therapy: a Probabilistic Approach**. Open Forum Infectious Diseases, 2022https://doi.org/10.1093/ofid/ofac492.267

SONG et al., **Clinical profiles and antimicrobial resistance patterns of invasive Salmonella infections in children in China**. European Journal of Clinical Microbiology & Infectious Diseases (2022) 41:1215–1225. https://doi.org/10.1007/s10096-022-04476-7

Antibiotic regimen	Probability of the antimicrobial regimen being successful in treating invasive Salmonella infections
Ciprofloxacin + Cefotaxime	99%
Ciprofloxacin + Ceftriaxone	99%
Levofloxacin + Cefotaxime	99%
Levofloxacin + Ceftriaxone	99%
Imipenem	96%
Meropenem	96%
Cefepime	94%
Piperacillin/ tazobactam	94%
Ceftazidime	93%
Cefotaxime	90%
Ceftriaxone	90%
Aztreonam	90%

Antibiotic regimen	Probability of the antimicrobial regimen being successful in treating invasive Salmonella infections
Ciprofloxacin	88%
Levofloxacin	88%
Amikacin	87%
Chloramphenicol	80%
Gentamicin	79%
Tobramycin	79%
Sulfamethoxazole/ trimethoprim	77%
Tetracycline	68%
Piperacillin	52%
Ampicillin/ sulbactam	47%
Ampicillin	44%

Success probability of one or two antibiotics to treat invasive *Salmonella* infections before the results of cultures and antimicrobial susceptibility tests are available.

		Nepal				Bangladesh				Pakistan			
		Salmonella Typhi	*Salmonella Paratyphi*	Total			*Salmonella Typhi*	*Salmonella Paratyphi*	Total			*Salmonella Typhi*	*Salmonella Paratyphi*	Total
	Frequency	1366	235	1601		Frequency	4065	735	4800		Frequency	2093	137	2230
Probability of a specific serotype =		85%	15%	100%		**Probability of a specific serotype =**	85%	15%	300%	**Probability of a specific serotype =**		94%	6%	139%
ampicillin	n	1345	229	**97%**		ampicillin	n	4065	735	**77%**	ampicillin	n	2091	137	**29%**
	S	97%	99%				S	73%	100%			S	18%	94%	
ceftriaxone	n	948	173	**99%**		ceftriaxone	n	4065	735	**100%**	ceftriaxone	n	2089	136	**44%**
	S	100%	98%				S	100%	100%			S	35%	99%	
chloramphenicol	n	1345	230	**98%**		chloramphenicol	n	4065	735	**84%**	chloramphenicol	n	2087	136	**30%**
	S	98%	98%				S	81%	100%			S	18%	97%	
ciprofloxacin	n	1345	232	**13%**		ciprofloxacin	n	4064	735	**2%**	ciprofloxacin	n	2085	135	**4%**
	S	13%	10%				S	2%	0%			S	5%	1%	
cotrimoxazole	n	1347	230	**98%**		cotrimoxazole	n	4065	735	**85%**	cotrimoxazole	n	2087	135	**29%**
	S	98%	100%				S	82%	100%			S	18%	96%	
imipenem	n	1259	200	**100%**		imipenem	n	0	0	**na**	imipenem	n	1507	5	**100%**
	S	100%	100%				S	na	na			S	100%	100%	
azithromycin	n	1348	232	**99%**		azithromycin	n	4065	735	**97%**	azithromycin	n	1530	5	**99%**
	S	99%	97%				S	98%	90%			S	99%	100%	

GARRETT et al., **Incidence of typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan: results of the Surveillance for Enteric Fever in Asia Project**. www.thelancet.com/lancetgh Vol 10 July 2022

Using dynamic programming to analyze pseudoDNA

3. How to identify clusters of related microorganisms, even without traditional DNA fingerprinting methods?

a) Create a **pseudoDNA** by concatenating the antimicrobials test results as either sensitive (S), resistant (R), or intermediate (I) for each strain.

b) Use dynamic programming to calculate the minimum **edit distance** between each pair of pseudoDNA sequences, creating a matrix of pairwise similarities.

c) Identify cluster of similar strains based on a cutoff in the edit distance matrix, such as a distance of less than 1.

COUTO et al., **Integrating artificial intelligence and dynamic programming to identify microbial clusters through antimicrobial susceptibility sequence analysis**. Open Forum Infectious Diseases, 2023 Open Forum https://doi.org/10.1093/ofid/ofad500.1263

Francisco et al., **First report on antimicrobial resistance and molecular characterization of Salmonella enterica serotype Typhi isolated from human specimens in Luanda, Angola (2010).**

https://doi.org/10.1016/j.jgar.2018.02.001

Strain ID	Ampicillin	Ampicillin-sulbactam	Cefuroxime	Piperacillin	Sulphamethoxazole-trimethoprim	Chloramphenicol	Ciprofloxacin	PseudoDNA
S27	R	R	R	R	R	S	S	RRRRRSS
S32	R	R	S	R	R	S	S	RRSRRSS
S2	S	S	S	S	S	S	S	SSSSSSS
S18	R	R	R	R	R	S	S	RRRRRSS
S26	R	R	S	R	R	S	S	RRSRRSS
S29	S	S	S	S	S	S	I	SSSSSSI
S3	S	S	S	S	R	S	S	SSSSRSS
S28	R	R	S	R	R	S	S	RRSRRSS
S34	S	S	S	S	S	S	I	SSSSSSI
S19	S	S	S	S	S	S	S	SSSSSSS

	S27	S32	S2	S18	S26	S29	S3	S28	S34	S19
S27		1	5	0	1	6	4	1	6	5
S32	1		4	1	0	5	3	0	5	4
S2	5	4		5	4	1	1	4	1	0
S18	0	1	5		1	6	4	1	6	5
S26	1	0	4	1		5	3	0	5	4
S29	6	5	1	6	5		2	5	0	1
S3	4	3	1	4	3	2		3	2	1
S28	1	0	4	1	0	5	3		5	4
S34	6	5	1	6	5	0	2	5		1
S19	5	4	0	5	4	1	1	4	1	

Cutoff in the edit distance matrix: 1.00

Number of clusters identified = 2

- ✓ Cluster 1 = {S27 = S32 = S18 = S26 = S28}
- ✓ Cluster 2 = {S2 = S29 = S3 = S34 = S19}

GARRETT et al., **Incidence of typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan: results of the Surveillance for Enteric Fever in Asia Project**. www.thelancet.com/lancetgh Vol 10 July 2022

Total sample size (n=6,552):

ID	pseudoDNA	...	ID	pseudoDNA	...	ID	pseudoDNA	...	ID	pseudoDNA	...	ID	pseudoDNA
1	SSSISSS	...	435	SSSISSS	...	607	SSSISSS	...	4614	SSSISSS	...	34734	RRRRRRS
2	RSRIRSS	...	438	SSSISSS	...	615	SSSISSS	...	4618	SSSISSS	...	34735	RRRRRRS
15	RSSISSS	...	439	SSSISSS	...	624	SSSISSS	...	4631	SSSISSS	...	34736	SSSRSSS
21	SSSISSS	...	443	SSSISSS	...	628	SSSISSS	...	4633	SSSISSS	...	34737	RRRRRRS
27	SSSISSS	...	448	SSSISSS	...	636	SSSISSS	...	4636	SSSISSS	...	34738	RRSRSRS
40	SSSISSS	...	461	RSRIRSS	...	638	SSSISSS	...	4651	SSSISSS	...	34740	RRRRRRS
48	SSSISSS	...	465	SSSISSS	...	640	SSSISSS	...	4672	RSSRSSS	...	34741	RRRRRRS
56	SSSSSSS	...	466	SSSISSS	...	651	RSRIRSS	...	4675	SSSISSS	...	34742	RRRRRRS
60	SSSISSS	...	470	RSSRSSS	...	652	SSSISSS	...	4676	SSSISSS	...	34743	RRRRRRS
64	SSSISSS	...	490	SSSISSS	...	656	SSSISSS	...	4683	SSSISSS	...	34744	RRRRRRS
78	SSSISSS	...	496	SSSISSS	...	662	RSSIRSS	...	4686	SSSISSR	...	34745	RSRRRSS
85	SSSISSS	...	501	SSSISSS	...	663	SSSISSS	...	4699	SSSISSS	...	34746	RRRRRRS
89	SSSISSS	...	508	SSSISSS	...	671	SSSISSS	...	4707	SSSISSS	...	34747	RRRRRRS
90	SSSISSS	...	513	SSSSSSS	...	677	SSSISSS	...	4715	SSSISSS	...	34748	RRRRRRS

Cutoff in the edit distance matrix: 2.00 Number of clusters identified = 5

- ✓ Cluster 1 = (1=3=4=5=6=7=8=9=10=11=12=13=14=15=16=17=18=19=20=22…
- ✓ Cluster 2 = (21=25=32=45=64=67=77=79=83=86=100=103=122=124…
- ✓ Cluster 3 = (34=251=261=303=331=358=387=396=449=523=542=553=558=564…
- ✓ Cluster 4 = (1667=1996=4917=4969=5496)
- ✓ Cluster 5 = (6204=6265=6340=6387=6398=6405=6413=6415=6417=6423=6459=6470=6511=6515)

Visualization of high-dimensional space object

4. Is it possible to visualize objects that exist in a multidimensional space as a fourth dimension and beyond?

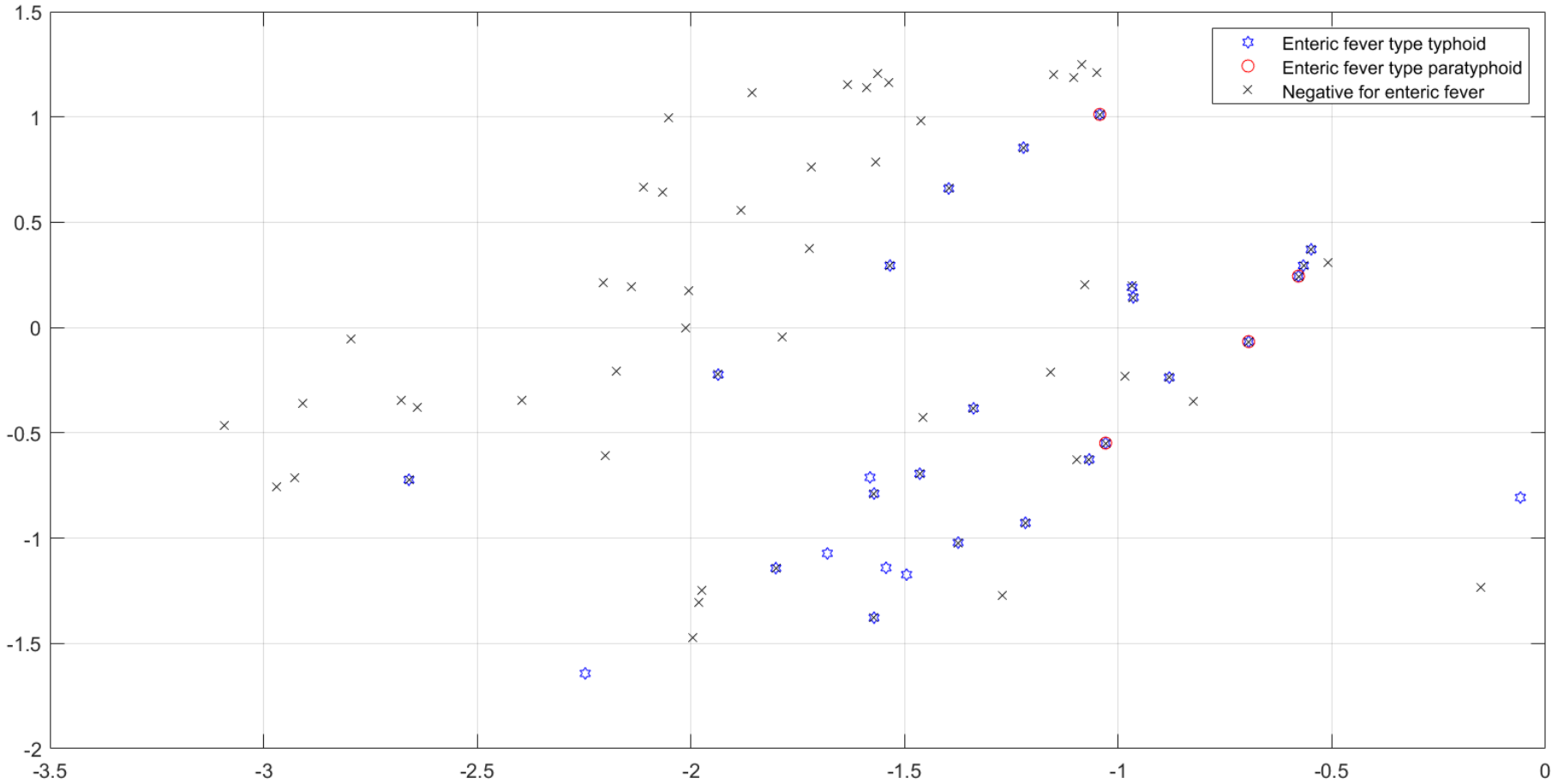
- The high-dimensional visualization problem in R^m can be formulated as a distance-geometry problem:

$$\Re^m \approx \Re^3 \quad \text{or} \quad \Re^m \approx \Re^2$$

- to find n points in low space (2D or 3D) so that their interpoint distances match the corresponding values from R^m as closely as possible.

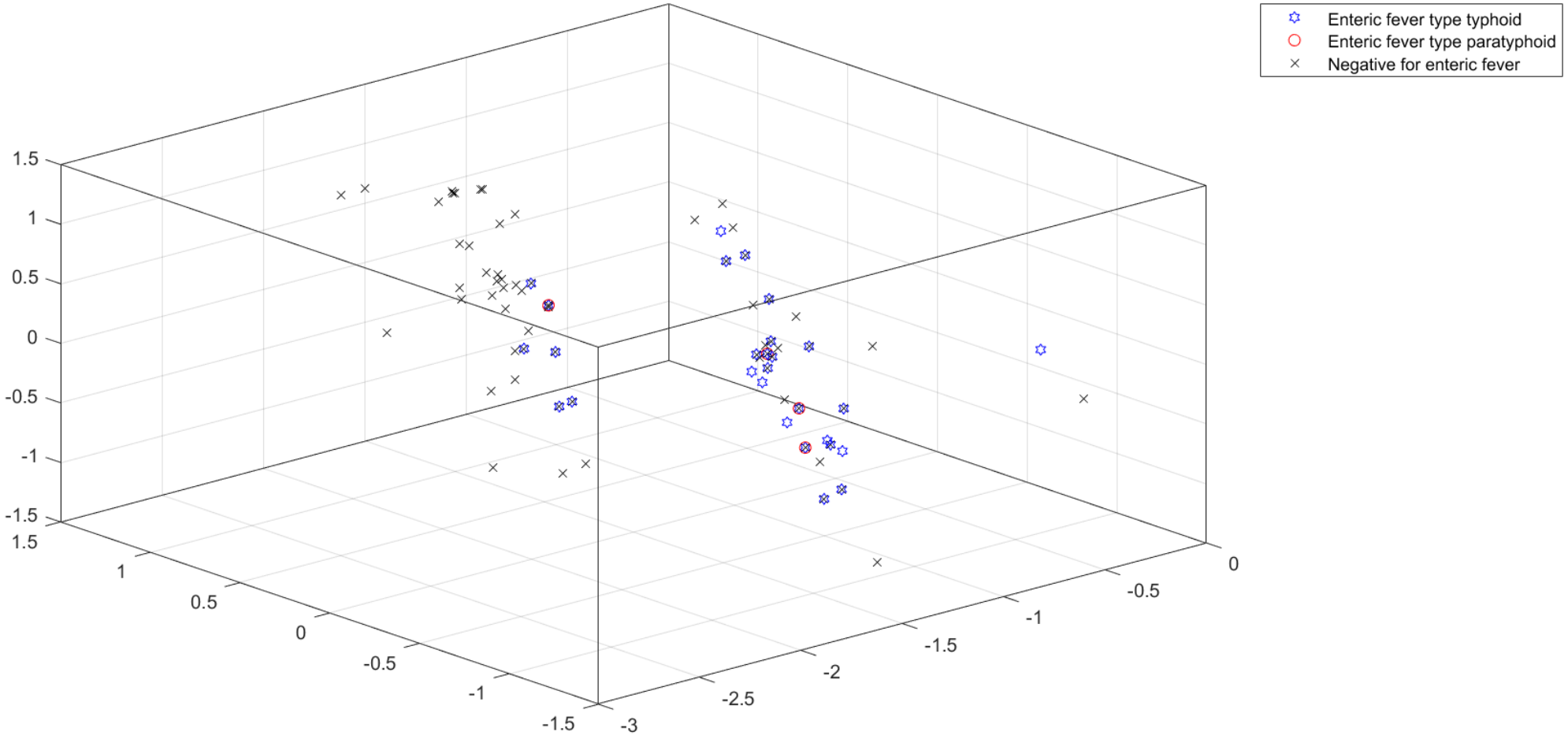
MARCOLINO LS, COUTO BRGM, SANTOS MA. **Genome Visualization in Space.** Advances in Soft Computing, v. 74, p. 225-232, 2010.

GARRETT et al., **Incidence of typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan: results of the Surveillance for Enteric Fever in Asia Project**. www.thelancet.com/lancetgh Vol 10 July 2022



Visualization in a reduced space (2D) of 34,015 cases of suspected and confirmed typhoid fever, codified as 16 frequency vectors of signs and symptoms (16th dimension space).

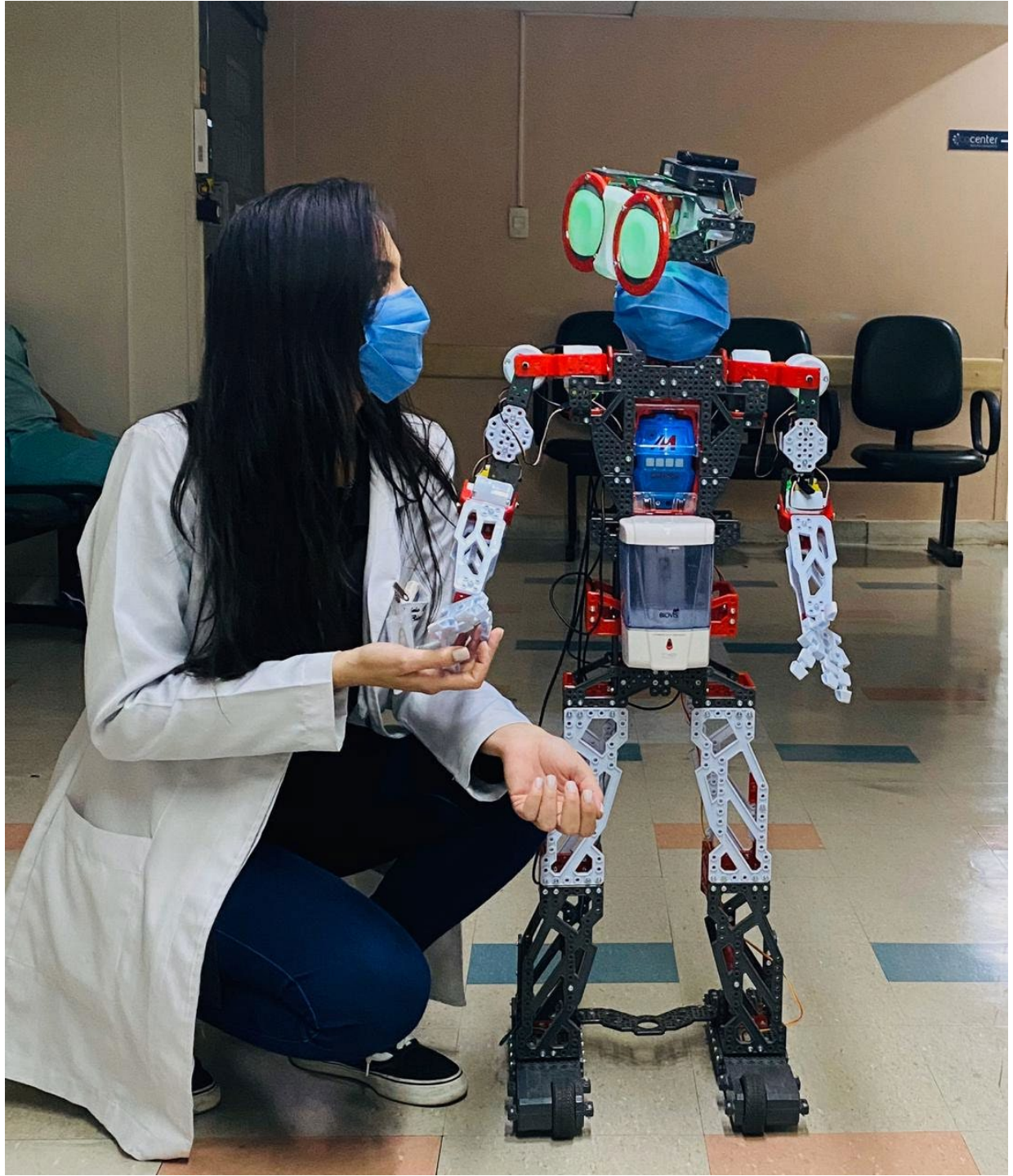
GARRETT et al., **Incidence of typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan: results of the Surveillance for Enteric Fever in Asia Project**. www.thelancet.com/lancetgh Vol 10 July 2022



Visualization in a reduced space (3D) of 34,015 cases of suspected and confirmed typhoid fever, codified as 16 frequency vectors of signs and symptoms (16th dimension space).

Conclusion: bringing it all together in a cloud app (AWS)

1. How to identify typhoid/paratyphoid enteric fever without culture?
   - Logistic regression (EnteriRisk), decision tree, and…?!
2. How to start optimal empiric antibiotic therapy?
   - Law of total probability.
3. How to identify clusters in an outbreak without genotyping?
   - pseudoDNA with edit distance clusterization.
4. How to visualize objects in a multidimensional space?
   - Vector representation of cases with minimization methods.



Source code

EnteriRisk Model

– Thank you very much!

brauliocouto@biobytebrasil.com

+55 31 98896-6098