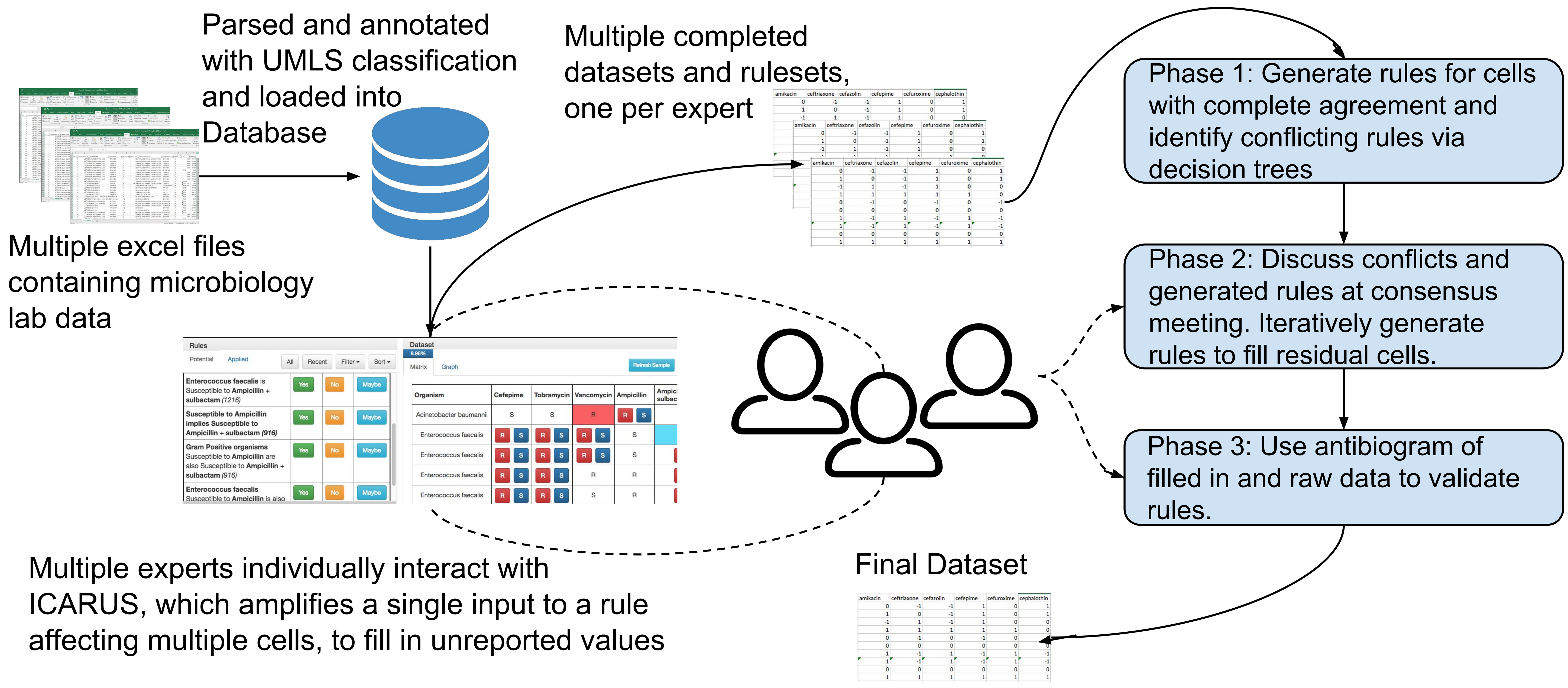
Derivation of Expert Consensus Rules for Missing Antimicrobial Susceptibility Data

The Ohio State University
Wexner Medical Center

Protiva Rahman, BS[1,4]; Erinn M. Hade, PhD[1]; Arnab Nandi, PhD[4]; Preeti Pancholi, PhD[2]; Mark Lustberg, MD, PhD[3]; Kurt Stevenson, MD, MPH[3]; Courtney Hebert, MD, MS[1,3]; [1]Department of Biomedical Informatics, [2]Department of Pathology, [3]Internal Medicine, College of Medicine, [4]Department of Computer Science, The Ohio State University, Columbus, Ohio, USA



Motivation

- Preprocessing and cleaning data prior to analysis is iterative and time-consuming.
- Engaging domain experts at this stage can be challenging, due to difficulties in data interaction and expression.
- We discuss our methods for getting expert consensus in filling in 75,000 cells in an antibiotic susceptibility dataset.

Background

- Antibiotic susceptibility data from microbiology labs are parsed and annotated with Unified Medical Language System (UMLS)[1] classification.
- The parsed data are loaded into ICARUS[2] (our data completion tool built for amplifying domain expertise).
- Experts individually complete datasets which need to be consolidated.

Table 1: Description of Data Challenges and Proposed Solutions

Theme	Challenge	Solution
Size of dataset	Laborious to annotate manually	ICARUS provides small snapshots of data to the expert and suggests rule generalizations to amplify input.
	Difficult to make more generalized and less organism-specific rules	Ontology classifications underlying the dataset allow for rules that generalize
Reproducibility and Transparency	Difficult to apply rules to a new dataset manually	Because of the ontologies behind the dataset, once a new dataset has been annotated with needed information, rules can be automatically used to fill in unreported data.
	Difficult to easily document all decisions transparently	All accepted rules are encoded and stored.
Consensus	Experts differ in their interpretation, and expertise provided by only one expert may introduce bias	Allow multiple experts to use ICARUS to provide expertise and then synthesize their input in the form of rules.

Acknowledgement: We acknowledge the support of NSF under awards IIS-1422977, IIS-1527779, and CAREER IIS-1453582 and NIAID of NIH under award R01AI116975.

NSF NIH

Methods

Rule Generation Via Decision Tree

- Given multiple datasets which were filled in by experts, we use custom decision trees to extract rules that fill in consensus cells (cells that did not have conflicts between experts).
- Features include other antibiotic susceptibilities of the culture, and UMLS annotations.
- The decision tree splits were made along semantic features (antibiotics in the same class and organism hierarchy).

Consensus Meeting and Validation

- Conflicting rules were extracted and resolved at a consensus meeting with 4 experts.
- Rules were again generated to fill in residual unfilled cells.
- Final set of rules were validated through manual inspection of an antibiogram (a table that summarizes resistances by organism and antibiotics).

Results

Validation results (Table 2)

- 171 rules were automatically generated, 105 of which were accepted and 22 were modified.
- 68 conflicting rules were discussed at the consensus meeting
- After final rule generation, 94% of the database was filled.
- Experts reviewed the antibiogram and select visualization (Figure 1) for final validation.

Figure 1: Antibiogram Trends for Validation

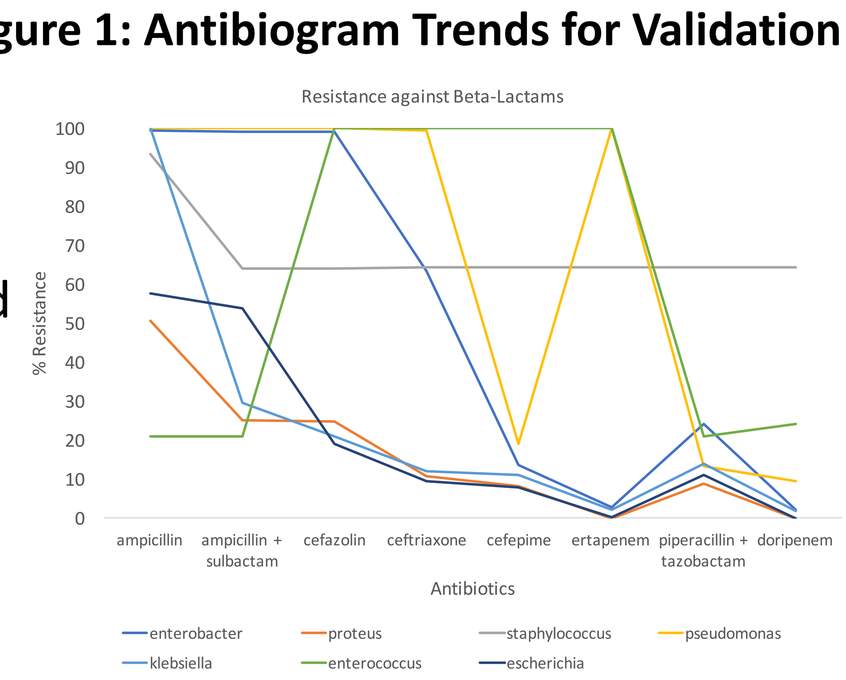


Table 2: Results of Rule Validation

	Total Rules	Accepted	Split	Removed Constraint	Precision*	Added Constraint	Rejected	Cell Impact (% of total missing)
Initial Rule Gen.	171	105	0	22	74%	0	44	52,569 (70%)
Consensus Meeting	68	40	0	0	n/a	1	27	16,647 (22%)
Final Rule Gen.	55	16	3	5	43%	36	0	4,519 (6%)

*Precision = (accepted + split + removed constraint)/total

Future Directions

- Create effective interactive visualization of rules to accelerate consensus meeting [3].
- Validate our methods at a different institution.

References

1. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research. 2004 Jan 1;32(suppl_1):D267-70.
2. Rahman P, Hebert C, Nandi A. ICARUS: Minimizing Human Effort in Iterative Data Completion. Proceedings of the VLDB Endowment. 2018 Sep 1;11(13):2263-76.
3. Rahman P, Chen J, Hebert C, Pancholi P, Lustberg M, Stevenson K, Nandi A. Exploratory Visualizations of Rules for Validation of Expert Decisions. In DSIA Workshop, IEEE VIS 2018.