CPHM-Sunday-237

# Evaluation of Artificial Intelligence Software for Screening of Stool Cultures

Tristan R. Grams[1,2], Angela Ma[3], Natali Baker[2], Megan Hirschi[2], and Mark A. Fisher[1,2]

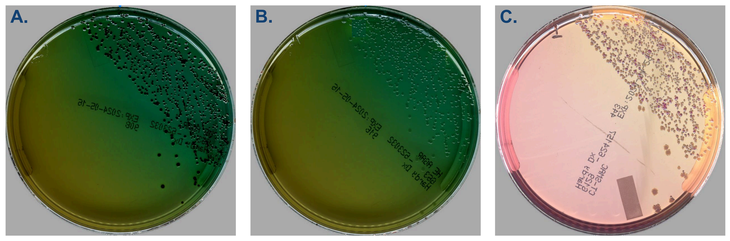
[1]Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT
[2]ARUP Laboratories, Salt Lake City, UT
[3]Public Health Ontario, Toronto, Ontario, Canada

## Abstract

**INTRODUCTION:** Over the past decade, the implementation of laboratory automation platforms in clinical microbiology has sought to streamline workflows and aid in the interpretation of often complex cultures. One way this has been achieved is through machine learning algorithms and artificial intelligence (AI) to analyze digital images of cultures to distinguish specific organisms based on growth and colony characteristics. PhenoMATRIX (PM) AI software developed by COPAN (Brescia, Italy) uses this approach, for example, to screen urine cultures and identify specific organisms on chromogenic agars, which has resulted in laboratory improvements. In this study, we evaluated the utility of a novel PM algorithm to screen stool cultures for *Salmonella*, *Shigella*, and *E. coli* O157. The PM stool culture algorithm interprets images of MacConkey, Hektoen Enteric (HE), and MacConkey with sorbitol (CT-SMAC) agar plates at 24 and 48-hours of incubation. **METHODS:** Over 117,000 images of clinical and spiked stool specimens from the WASPLab system (Copan) at ARUP Laboratories were used for algorithm training. The PM software separated culture plates based on growth and colony characteristics into 6 folders: 24H Negative for all 3 pathogens, 24H suspicious for *Salmonella/Shigella*, 24H suspicious for *E. coli* O157, 24H Positive-mixed (>1 pathogen) culture, 48H Positive for any of the pathogens, and 48H Negative. The algorithm accuracy was assessed by comparing manual to automated analysis of 2,034 cultures. **RESULTS:** The overall agreement of PM classification at 24- and 48-hours of incubation to manual reading was 95.4% and 82.9% respectively with most disagreements resulting from a false positive flag. Importantly, the negative predictive value (NPV) for 24- and 48-hour negative stool cultures was 99.0% and 99.4%. An additional prospective analysis of the algorithm was performed for 6 months prior to implementation in the laboratory. This revealed that the most common algorithm false positives included *Pseudomonas, Enterobacter, and Citrobacter* species. Additionally, the software identified several cultures that may otherwise have been missed for further workup, highlighting a potential patient-care benefit of AI-assisted culture reading. **CONCLUSIONS:** Initial pre-sorting of cultures by PM software can streamline the review of stool cultures. The high NPV allows technologists to focus on potential positives. Overall, PhenoMATRIX may improve sensitivity and consistency of stool culture interpretation in the clinical microbiology laboratory.

## Background

- Bacterial gastroenteritis in the United States
  - *Salmonella* – 1.35 million annual cases
  - *Shigella* – 450,000 annual cases
  - *E. coli* O157 – 74,000 annual cases
- Stool culture setup
  - Blood, MacConkey, Hektoen Enteric x2, MacConkey w/ sorbitol, Campy-CVA agars
- WASPLab Full Lab Automation
  - PhenoMATRIX artificial intelligence (AI) suite
    - Automatic reading, interpretation, and sorting of cultures
    - Currently employed for urine cultures and chromogenic agars
- Can PhenoMATRIX AI software be used to assist technologists with screening stool cultures?




**A.** Growth of *Salmonella* species on Hektoen Enteric (HE) agar. **B.** Growth of *Shigella* species on Hektoen Enteric (HE) agar. **C.** Growth of non-sorbitol fermenting *E. coli* O157 colonies on MacConkey with sorbitol (CT-SMAC).

## Methods

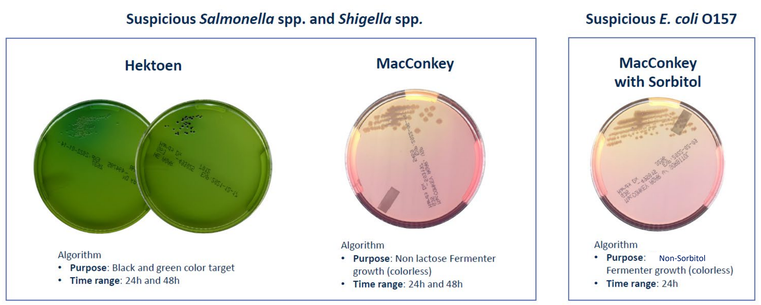
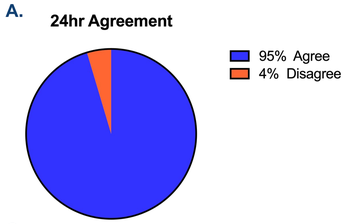
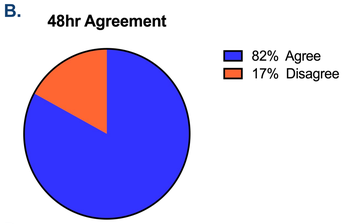



**Figure 1. Overview of Algorithm Training for the PhenoMATRIX Stool Culture AI.** The PM stool culture algorithm interprets images of MacConkey, Hektoen Enteric (HE), and MacConkey with sorbitol (CT-SMAC) agar plates at 24- and 48-hours of incubation. Additionally, a 6-hour broth-enriched HE plate is read at 14- and 38-hours post plating. Over 117,000 images of clinical and spiked stool specimens from the WASPLab system (Copan) at ARUP Laboratories were used for algorithm training. The PM software separated culture plates based on growth and colony characteristics into 6 folders: 24H Negative for all 3 pathogens, 24H suspicious for *Salmonella/Shigella*, 24H suspicious for *E. coli* O157, 24H Positive-mixed (>1 pathogen) culture, 48H Positive for any of the pathogens, and 48H Negative.

## Results

### Retrospective Pre-Plate Assessment of Software Performance




C.

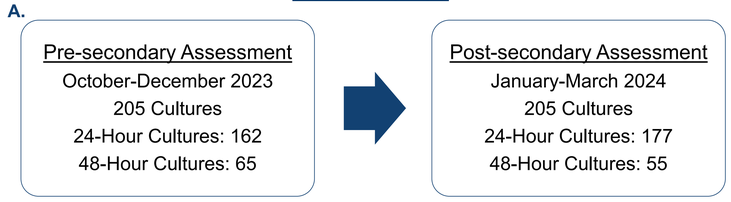
| PhenoMATRIX® Folder | Sample Count | Sample x Day | % |
|---|---|---|---|
| 24H Negative | 615 | 6 | 46.7 |
| 24H Suspicious Salmonella/Shigella | 435 | 4 | 33.0 |
| 24H Suspicious E. coli o157 | 87 | 1 | 6.5 |
| 24H Positive Mixed | 180 | 2 | 13.7 |

D.

| PhenoMATRIX® Folder | Sample Count | Sample x Day | % |
|---|---|---|---|
| 48H Negative | 316 | 3 | 46.7 |
| 48H Positive | 401 | 4 | 33.0 |

**Figure 2. Plate Based Assessment of PhenoMATRIX AI Software Performance.** The PhenoMATRIX AI Software was retrospectively evaluated using cultures collected over a 3-month period. Plate-level agreement at **A.** 24- and **B.** 48-hours. Agreement was determined by comparing the call of the PM software to that of the technologist on a per-plate basis. Breakdown of the number of samples evaluated and PhenoMATRIX folder ID that were used to determine precent agreement at **C.** 24- and **D.** 48-hours.

### Prospective Culture-Based Assessment of Software Performance



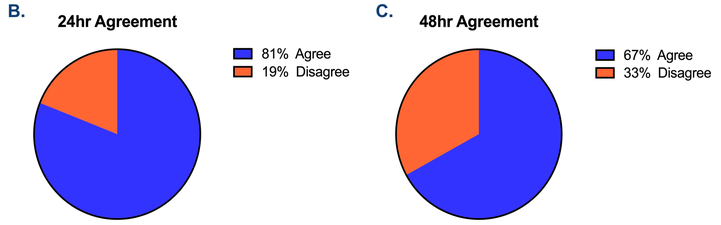







**Figure 3. Culture-Based Performance of PhenoMATRIX AI Stool Culture Software. A.** Outline for assessment of performance for the PhenoMATRIX (PM) AI software. A total of 410 cultures were evaluated over a 6-month period to assess software performance. Additionally, 32 negative samples were spiked with *Salmonella*, *Shigella*, or *E. coli* O157 due to the low incidence of positives during the evaluation period. The evaluation was split into two-phases; 1) without and 2) with secondary review by a microbiology fellow of cultures not in agreement between the technologists and PM AI software. **B-C.** Agreement at 24- and 48-hours. Agreement was defined as the final overall culture result from the PM software matched that of the technologist.

### PhenoMATRIX AI Software Increases the Consistency of Culture Reading

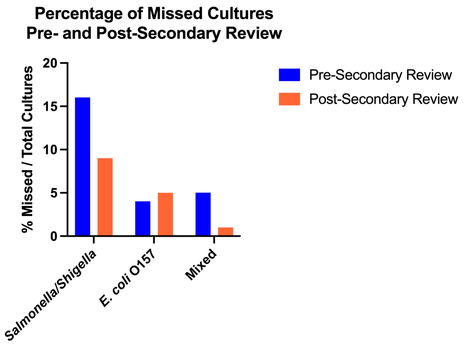


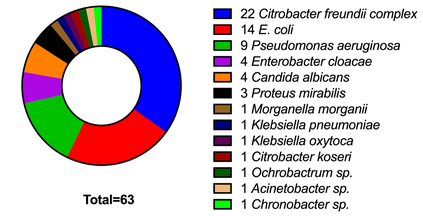

**Figure 4. PhenoMATRIX AI Software Decreases the Number of Potentially Positive Cultures Missed by Technologists.** Cultures were evaluated before and after the implementation of a secondary review of disagreements between the PM AI software and technologists as described in Fig 3A. After the post-review period was implemented, there was a 7% decrease in potential *Salmonella/Shigella* cultures and a 4% decrease in mixed cultures that were not initially worked up. There was a 1% increase in potential *E. coli* O157 cultures missed by technologists.

## Results

### PhenoMATRIX AI Software "False Positives"

A.

| PhenoMATRIX® Folder | False-Positive Culture Count |
|---|---|
| 24H Suspicious *Salmonella/Shigella* | 36 |
| 24H Suspicious *E. coli* O157 | 1 |
| 24H Positive Mixed | 5 |
| 48H Positive | 25 |



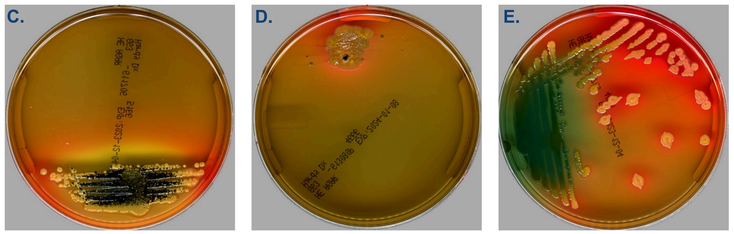




**Figure 5. "False Positive" PhenoMATRIX AI Software Calls are Driven by Media Limitations. A.** Number of cultures with "false positive" PM flag. **B.** MALDI identification of organisms that resulted in "false positive" flag by PM AI software shown as a percentage of total organisms (n=63). The number of times the organism was identified is listed before the name of the organism. **C-E.** Examples of "false positive" *Salmonella/Shigella* flag by software because of **C)** *Citrobacter* species **D)** *Pseudomonas* species and **E)** 48-hour color change in areas of high growth on HE agar.

### PhenoMATRIX AI Software Flags Cultures Otherwise Missed by Lab Staff

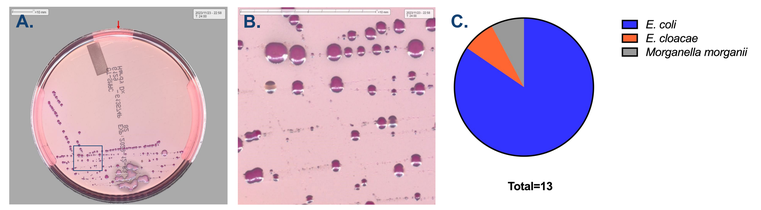



**Figure 6. PhenoMATRIX AI Software Identifies Cultures Otherwise Missed by Manual Reading. A.** Example of CT-SMAC plate for the evaluation of *E. coli* O157 presence that was missed by the technologist but flagged by the PM AI software. **B.** Enlarged image of area indicated in **6A. C.** MALDI identification of organisms from cultures that were flagged positive for *E. coli* O157 but were not worked up by lab technologists. All *E. coli* isolates were non-O157 as determined by *E. coli* O157 latex test.

## Conclusions and Future Directions

- Excellent performance of PhenoMATRIX software for screening stool cultures
  - 100% negative predictive value
  - Flagging of possible pathogens missed by technologists
- "False positives" are driven by limitations of media rather than software performance
  - Highest with "24H *Salmonella/Shigella*" and "48H Positive" reads
- Improved consistency for evaluation of stool cultures
  - Initial interpretation is performed by software with review by technologist
- Facilitates training and continuing education for technologists
  - Record of images and decisions made by software/technologist

## Acknowledgments

We acknowledge the tireless effort of the ARUP Bacteriology lab staff, algorithm development and logistical support from Copan Diagnostics, and evaluation media from Hardy Diagnostics.