Targeting microbiology lab efficiency with AI

Amy Carpenter Aquino

August 2020—Bringing an automated culture plate reading system into the Hennepin County Medical Center microbiology laboratory was never a question of if but when.

“We need artificial intelligence to help us with active decision-making processes in the lab,” says Glen Hansen, PhD, medical director of clinical microbiology and molecular diagnostics at Hennepin County Medical Center in Minneapolis. “The difference between robotics and AI is that we are now at a point where we want these instruments to help us with decision-making processes, not just robotic functions.”

His laboratory’s interest in automated plate reading is born of the trio of aims that labs everywhere are trying to achieve with automation, he says: “How can we get more testing out of the current staffing levels in our laboratory, how can we decrease turnaround times, and how can we introduce tools into the laboratory that allow the microbiology work to be done more efficiently, with fewer errors, and more time to protect the skill set of my laboratory techs for the areas we want them to look at?”

When Dr. Hansen’s team acquired the Automated Plate Assessment System (APAS) Independence (Clever Culture Systems) to process urine culture readings in December 2018, one of the selling points was the modular option that could be adapted easily to his laboratory’s existing layout, he says. Another was the pipeline of modules for methicillin-resistant *Staphylococcus aureus* (under FDA consideration), vancomycin-resistant enterococci, and zone size reading of antibiotic sensitivity discs. A study performed at Johns Hopkins Hospital found that the APAS demonstrated high accuracy compared with manual readings of MRSA cultures and detected a low level of positives missed by manual reading (Gammel N, et al. Poster presentation at ASM Microbe Online 2020).

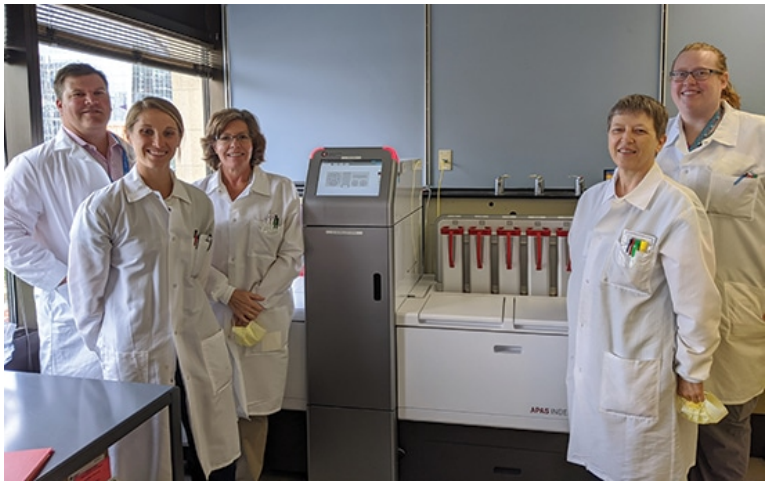
This year, Hennepin finalized its purchase of the APAS Independence, which uses AI to interrogate colonies for size, pigment, and granularity. “We believe systems like this are needed in the lab,” Dr. Hansen says.

The APAS Independence, which the FDA cleared in 2019, screens urine cultures for significant growth and removes negative cultures from the workflow. Dr. Hansen and colleagues in 2019 reported the results of the first U.S. use of the instrument to screen urine cultures. In the study, APAS categorized cases as significant growth, marked for review, no growth, or errors. Of 720 urine cultures evaluated, APAS identified 98 percent (370/374) of cases reported as no growth by the clinical lab. APAS identified as errors (caused by defects in the media) four cases that the lab reported as no growth. Thirty percent of the 720 cultures were marked for review by the lab. The combination of significant growth plus cases marked for review by APAS correctly identified 100 percent of all positive cultures (Hansen G, et al. Poster presentation at ASM Microbe 2019).

“We can target the work better with these tools,” and instruments make it possible for technologists to apply their skills where they count, “whether that’s decreasing turnaround times or focusing on cases that are growing,” Dr. Hansen says.
“I don’t want to spend 50 percent, let alone 80 percent, of talented tech time to read a negative plate. Lab automation allows us to focus the work of our techs on cases that matter.”

A shrinking workforce means maximizing resources, he says. “I’ve got eight talented people in the lab who work my day shift, whom I depend on to come to work Monday through Sunday and work from anywhere between five in the morning to five o’clock at night. How can I use their skills more effectively, so that if we’re asking them to do four things in a day, and I can provide them with more tools, they could do seven things in a day?”

The APAS Independence has had an impact, Dr. Hansen says. “The ability for us to put urines through the APAS has changed our entire workflow. Because of the way my benches are organized in the lab, the techs sit at a bench and see all the cases that come in. If I can remove the urine workflow from those benches, my benches become a lot more efficient, and as a result there are cases we can get to earlier.”



Dr. Glen Hansen, far left, at Hennepin County Medical Center with (from left) Jacqueline Salden, laboratory technologist/microbiology; Betsy Wesenberg, microbiology supervisor; Susan Nelson, microbiology technical lead; and Alexandra Nussbaum, laboratory technologist/microbiology/informatics. “This is a system that allows us to target their talents,” Dr. Hansen says of the staff and the lab’s APAS Independence (center).

For labs with dedicated urine benches, he says, an automated culture reading system could provide enough efficiency to allow for a 50 percent or greater reduction in the number of urine benches.

“With the pandemic, laboratories are under more financial constraints than they’ve ever been before, and staffing levels are going to be assessed like they’ve never been before,” Dr. Hansen says.

In Hennepin’s microbiology laboratory, the primary job of some staff is to look out for urine cultures that could be loaded onto the APAS after incubation. “Right now, the system does not do incubation,” Dr. Hansen says. “We do incubation in our normal incubators, and then we pull the plates over in the morning. The plates are time-stamped so the techs know how long they’ve been incubating, and then we put them into the system.”

APAS sorts the plates and autoclears negative plates from the workflow within 16 seconds. FDA clearance is for all urine cultures with detected positive growth to be sent back to the microbiologists for review. “So it doesn’t autocall systems,” Dr. Hansen says. “It identifies the growth patterns that are likely to be of clinical significance.”

He and his team have worked with Clever Culture to code the system to allow the technologists to understand what is a high-probability-positive urine culture versus a category they call “doubtful.” “So we have used the system to fine-tune it to provide a little more diagnostic decision-making in the lab, but those [positive] plates right now come back to the workflow benches for the techs to review.”

More than half of the urine culture workflow were negatives or was considered to be of limited clinical significance.

“Preliminary experience so far”—preliminary because the APAS isn’t yet connected to the reporting system—“suggests efficacy increases and a reduction in hands-on times with urine culture reporting of around 15 to 20 percent,” Dr. Hansen says. “For us, that equates to between a one- to two-hour savings per day in urine culture handling.”

“At the end of the day, it allows us to get through the workload more effectively and quicker than we’d been able to do without it. That in and of itself has endeared the system to the work staff.”

Time saved from the automation of urine culture screening allows other culture screenings to be completed more quickly. “If I’ve got a tech sitting at a bench that has urine, throat, and respiratory cultures, there are a series of cases that, without the APAS, we might not get to until after lunch because of how the benches are configured,”

Dr. Hansen says, citing as an example a respiratory sample from an ICU patient. “Because we can get to that case before lunchtime, if that case represents a significant case where the tech would have to do an antibiotic susceptibility on it, all of the work that flows from that case is now also adjusted. We believe that the susceptibility parameters of doing conventional culture in the lab will also shift because of the APAS.”

‘We have to learn how to set the expectations for AI in microbiology and culture reading.’ Karissa Culbreath, PhD, D(ABMM)

It didn’t take long to see the benefits, he says, “and the pandemic has only strengthened our understanding of how tools are needed in microbiology to get through the day.” The Hennepin microbiology laboratory, like other labs in the U.S., has had to shift staff to the molecular laboratory to help with the COVID-19 testing workload. “Tools like the APAS have unquestioningly allowed us to get through the day more effectively with lower burnout, lower stress, and, quite frankly, higher accuracy as you start to work people longer hours.”

“This is a system that allows us to target their talents,” he says of the staff.

APAS’ high level of accuracy in ruling out cases of limited significance is not what appealed most to Dr. Hansen. “I expect it to work. I’m not going to bring anything into my lab that doesn’t work,” he says. “What I want to know is how does this change the workflow? How does it change turnaround time? How does it help with staff engagement?”

So far the instrument is delivering on the expectation of lowered turnaround times and making it possible for staff to be redeployed to other areas, Dr. Hansen says. “We think these advantages are generalizable across all different laboratory spectrums.”

TriCore Reference Laboratories has seen similar efficiencies with its use of WASPLab. “The workflow impact is huge, especially in such a strained environment that we’re experiencing now,” says Karissa Culbreath, PhD, D(ABMM), medical director of the Infectious Disease Department at TriCore. She shared her laboratory’s experience with WASPLab in an AI session last year at the AMP annual meeting and in a July 9 interview with CAP TODAY.

“What we’ve been able to see in our lab is that about 30 percent of urine cultures are generally negative and another 30 percent are normal urogenital cultures. Which leaves a small proportion that are positive and would require workup by the technologists,” Dr. Culbreath says.

Even before the screening algorithms were implemented to screen out the negatives, “just by having the digital cultures and being able to have the manual process to move those into different buckets allowed us to redistribute our workforce,” says Dr. Culbreath, who is also associate professor in the Department of Pathology at the University of New Mexico School of Medicine. “It doesn’t take a highly trained microbiologist to read a negative

culture."

The redistribution meant the entry-level and newer staff looked at the negative cultures and those with normal flora and "even some of the cultures that were pure growth of one pathogen," she says. "The mixed cultures from nephrostomy patients or from other immunocompromised patients that could contain a lot of different pathogens go to the highly trained staff for their expertise."

At a time when staff from the microbiology laboratory, already understaffed because of the workforce shortage, has been reallocated to the molecular lab for COVID-19 testing, "having digital imaging up top helps us to be more efficient and optimize our process," Dr. Culbreath says. The PhenoMatrix culture reading algorithms used in WASPLab add another level of efficiency, she says, "because it's an algorithm that you can confidently say can screen everything as negative, and then have someone look at the cultures visually—in our case, we are able to view a batch of 30 cultures that have been flagged by the system as potentially negative and confirm those results with a single action."

Technologists still interact with the cultures "but on a larger scale, and with one foot on the button they can say all of those are negative. It's so much faster than going through every culture one by one."

In the AMP session, where Dr. Culbreath reported on the use of machine learning algorithms to support microbiology culture interpretation, she said her "MALDI eyes" tell her when she's looking at *E. coli*, for example, and the question is, "Can I teach a system and machine learning software to do the same type of thing that my MALDI eyes can do?"

Certain morphologies are a challenge to any system. "But they are the same challenges we would see in our own laboratories. How good are your MALDI eyes at differentiating *Enterococcus* from *Lactobacillus*?" she asks.

WASPLab uses convolutional neural networks to identify organisms "and not just based on morphology and large morphologic groupings but into species and genus levels with fairly high accuracy of 91 to 99 percent," she said. Accuracy was identified as the likelihood that the predicted assignment is correct, while specificity was based on rejecting the bacteria when it does not apply to the classification, such as the ability to reject *Klebsiella pneumoniae* from the *Staphylococcus* aureus group. "That's important—you need to know that it's not *Staph aureus* if it looks like *Klebsiella*."

The sensitivity—correctly classifying bacterial colonies into their species—wasn't as high, nor was precision. "The precision is the correct assignment of many bacterial colonies from the same species," Dr. Culbreath said. "We know that bacteria don't always read the book, so they may not always look the same in the laboratory and even in the same culture."

Intra-colony variability, in which bacterial colonies from the same species have multiple different morphologies, can also affect their classification (Huang L, et al. *Theor Biol Med Model*. 2018;15[1]:22).

"This has been directly related to the number of trained images that are used in building the system to say that all of these morphologies would fall into the *Streptococcus agalactiae* class." Because of the many possible appearances, "you have to continue to build out the system."

Morphologic relatedness, or the ability of the system to distinguish between morphologies that may look similar, also presents a challenge, she said. For example, *Klebsiella pneumoniae, Klebsiella oxytoca*, and *Enterobacter* species could all be detected within the lactose fermenting group.

"I know in the laboratory we may have some techs who say, 'I know this is not a *Klebsiella* because of how it smells,'" Dr. Culbreath said. "But since our artificial intelligence doesn't yet have Smell-O-Vision, we can't use the smell test to differentiate it. It's really based on morphology. So can we differentiate this because of the relatedness to the organisms?"

"This confusion matrix in convolution neural networking is what we use to determine the ability of the neural

networking system to differentiate the organism."

She and colleagues have been working on grouping and clustering based on morphologic and organism classification and have been able to correctly classify with a fairly high level of accuracy. "But it's also important to make sure we don't have too many that are misclassified."

In a colony recognition study, they challenged 247 isolates against the image analysis library. Organisms were considered correct if they were within the appropriate classification. Her team found it was unable to classify between eight and 13 percent of organisms in the staphylococci, streptococci, *Enterobacteriaceae*, and *Pseudomonas* groups. "That's the important part of a machine learning algorithm: If it doesn't know what it is, we need it to not call it something that it's not, and to say, 'I don't know the answer.'"

"You want to build into your system that it doesn't give you any answer at all. That's how we've been working to refine our system."

The technologists' interpretation of urine culture results presented an unexpected challenge and shed light on the limitations of automated culture reading. "When we started to employ the algorithm on some data sets, we realized that our baseline measure of the technologist providing the answer was part of the problem," Dr. Culbreath said. "Technologists are not machines and they result cultures in different ways. You can give different technologists the same culture and get different answers."

Dr. Culbreath and her colleagues set out to determine the cause of the discrepancies by running urine culture samples through the technologists and asking them to provide an answer. Next, the senior technologists followed the procedure precisely for another set of culture results. Agreements between what the procedure said was the correct answer and the technologist result ranged from 96 percent to 64 percent.

When the technologists were asked why they had not followed the procedure precisely, they found the technologists had added many questions of their own to the usual basic steps to determining a result. In addition to asking, "Is there any bacteria present?" the technologists wanted to know if the bacteria were from the specimen or from contaminant on the plate. In addition to how much is present, they wanted to know if it was on other plates too. In addition to "what does it look like?" they asked how long the culture had been incubating and what other plates look like. And in addition to "Is there more than one thing, and what do the other things look like?" they wanted to know: Are there different morphologies of one thing? Are there the same morphologies of different things?

Says Dr. Culbreath: "We have to learn how to set the expectations for AI in microbiology and culture reading. We are not at the point where we can put in a plate and we are going to get a perfect answer." The baseline expectations are accurate assessment of presence or absence of growth and quantitation of organisms present, the ability to detect an organism on multiple types of media along different time points, the ability to discriminate different morphologies of different organisms, and detection of different morphologies of the same organism. "And we're still at a point where the technologist will ultimately perform the final culture analysis," she said.

When they used the software with the rules they applied, there was near perfect agreement with what the technologists resulted by the reference result determined in the lab. "We had agreement that reached up to 99 percent for the detection of no growth and 92 percent with a differentiation of what would get ID and AST performed in the laboratory," she said.

Even more steps were used to determine a positive culture: Has the patient been on antibiotics? Any other cultures ordered? Patient diagnosis? Is there a rare scenario the protocol didn't address?

"Resulting clinical microbiology is not a yes or no answer," Dr. Culbreath said. "It is both an art and a science. We can program the algorithm to determine the science, but there is an art to the final interpretation."

What does fully integrated AI look like? "It's the culture result being processed through the neural network, but when it works best, and in the best future of it, we would have integration of patient data, the local antibiogram,

and hospital information," such as the diagnosis and patient experience, she said.

Whether AI in the microbiology laboratory is friend or foe is still unknown, in Dr. Culbreath's view. "And I think we still have a long way to go," she said.

Amy Carpenter Aquino is CAP TODAY senior editor.