Newsbytes

Twitter tool connects pathologists with similar cases

October 2020—Many prolific Twitter users describe the social media site as a time sink, but Andrew Schaumberg, PhD, begs to differ. After observing pathologists turn to Twitter to seek advice about difficult patient cases, he developed Pathobot, a free, artificial intelligence-driven search tool on Twitter that is designed to help pathologists connect with colleagues faster.

“One reason social media helps is there is often another pathologist somewhere in the world with their smartphone on who is immediately available to give an opinion,” says Dr. Schaumberg, research fellow in pathology at Harvard Medical School and a postdoctoral fellow in computational pathology at Brigham and Women’s Hospital. Recognizing the wealth of histopathology data that could be exchanged in such interactions, Dr. Schaumberg began collecting tweets in 2018. He used the data to develop Pathobot, which uses machine-learning and deep-learning techniques in real time to help pathologists find cases similar to theirs on Twitter. The bot is unique as a search tool, he says, because it’s guided by prior pathology discussion on social media and helps reach pathologists “in an automatic and targeted way, based on the case at hand.”

Although the bot’s underlying architecture is complicated, the tool is user-friendly. Users need only mention “@pathobot” in their Twitter posts that include photomicrographs of H&E-stained tissue and pathology-related hashtags, such as “#dermpath,” or text, such as “pancreas” or “infection.” “In a minute or so,” says Dr. Schaumberg, “Pathobot will reply to the post with a ranked list of eight of the top most similar patient cases on Twitter, and Twitter will notify the pathologists who shared those cases” to help foster discussion. Pathobot’s artificial intelligence is trained on photomicrographs of H&E-stained tissue only, he notes, because that’s what’s shared the vast majority of the time.

The bot, which went live last year, also links to relevant PubMed articles and provides a disease state prediction (normal/nontumor/infection, or benign/low-grade malignant potential, or malignant) for each histopathology image in the Twitter post. Those predictions are displayed as boxplots to provide a quantitative assessment of confidence in them. “If Pathobot predicts the chance of malignancy is 50 to 70 percent, that’s different than 30 to 90 percent, and you see that spread in the boxplot,” Dr. Schaumberg explains. “These predictions are simple sanity checks. [For example,] if a pathologist determines Pathobot inaccurately predicted malignancy for a Leishmania case, then the pathologist also knows to disregard Pathobot’s search results as inaccurate because the same Pathobot AI drives predictions and search.”

Dr. Schaumberg

Dr. Schaumberg is quick to note that the bot is not FDA approved to diagnose or treat disease. “Pathobot does simple search-related grunt work to connect pathologists who have similar cases,” he says. “[Its] primary purpose isn’t really to find the top most similar cases—it’s to find the pathologists who shared those cases.”

In one example, Sofopoulos Michail, MD, shared photomicrographs of H&E-stained tissue and IHC results for a lower mediastinal mass pressing on a patient’s esophagus. Pathobot found a similar case posted by Sanjay Mukhopadhyay, MD, who, after being notified by Twitter that his case was being used for comparison, replied, “Need history of malignancy, imaging and more #ihcpath. My differential would include RCC and lung cancer.

Open to other things too." Others who responded to the Twitter post suggested metastatic chromophobe renal cell carcinoma and chordoma of the posterior mediastinum.

Pathobot may be particularly beneficial for pathologists in low-resource settings, Dr. Schaumberg says. For instance, a pathologist in a developing country who is attempting to diagnose a case of Kaposi's sarcoma may not have access to an HHV8 IHC stain. But if the pathologist takes the case to Pathobot, another pathologist might suggest checking for a medical history of HIV, which would support a diagnosis of Kaposi's sarcoma.

In areas with more resources, pathologists might seek advice about IHC antibodies or other IHC methods for subtyping a malignancy, he says, or a favored diagnosis on a morphological basis. Celina Stayerman, MD, posted such a case using "#GUPath prostate adenocarcinoma. Got quite a few patterns here. Please share your Gleason score and ISUP. Pattern 3 is the predominant. @pathobot requires 'prostatic adenocarcinoma' and 'grading.'" After several replies, those who joined the discussion reached a consensus on the score. "Dr. Stayerman demonstrates advanced usage of Pathobot here, using the 'requires' keyword to focus Pathobot's search results on prostatic adenocarcinoma, for instance," Dr. Schaumberg explains.

But Pathobot is more than just a useful Twitter tool, Dr. Schaumberg says. It has become an indispensable component of his research. Initially, he used data from Twitter to develop an algorithm that could learn simple tasks, such as predicting whether an image showed H&E-stained tissue. For more difficult tasks, the algorithm's performance was poor and "expecting an AI to distinguish tissue types or disease state from these data seemed ambitious," he says.

It took a rallying telephone call from Dr. Mukhopadhyay, director of pulmonary pathology at Cleveland Clinic, to spur Dr. Schaumberg to rise to the challenge of such an algorithm project. "For him, an algorithm that could distinguish benign from malignant disease was the holy grail because it's such a common question confronting pathologists," says Dr. Schaumberg. "He was very strong about convincing me to pursue that."

But the pursuit would require obtaining vast amounts of data. "We needed to think outside the box to grow the project," he says. "And that's how Pathobot started."

Dr. Schaumberg and colleagues took the supervised machine-learning algorithm they were training to make disease predictions and repurposed it to search Twitter for relevant patient cases. They then developed the Twitter "user" @pathobot to employ their algorithm in real time to help pathologists find cases similar to their own. "Pathologists who saw Pathobot in action sometimes reached out to us and collaborated," says Dr. Schaumberg, "and we gained some training data for the AI to make it better," which began a virtuous cycle of improvements to the tool.

Eventually the team improved the algorithm's performance enough to publish their findings as a study (Schaumberg AJ, et al. *Mod Pathol*. doi.org/10.1038/s41379-020-0540-1). Building the data set discussed in the study required manually annotating 13,626 images from Twitter. The algorithm used Twitter hashtags, text from associated tweets, and replies to determine disease state ground truth—that is, nontumor, benign, or malignant—for the images. The Twitter data were augmented with 113,161 images of H&E-stained tissue from PubMed annotated with the associated article titles, abstracts, and figure captions.

The team also combed through all cases to find the most debated diagnoses. Many of these were easy to identify, Dr. Schaumberg says, because there tends to be a disproportionate amount of discussion on Twitter when there is disagreement. For such cases, the algorithm took a majority vote, and if this approach did not automatically infer the correct diagnoses, the team annotated the cases themselves rather than use the text from the associated tweets.

Another, perhaps surprising, step in creating the data set involved collecting social media content that had nothing to do with pathology, such as pathologists' vacation pictures posted on Twitter. To develop a machine-learning classifier to identify histopathology stains, Dr. Schaumberg and colleagues had to include in the data set histopathology and nonpathology images, so they obtained consent to download images of ski slopes and birthday

parties from some of those who contributed relevant data and manually separated these from the pathology-related tweets. The team plans to make their annotated data set of photomicrographs used in the project, sans personal photos, available under an open-source Creative Commons license by the end of the year.

Using this expanded data set, Dr. Schaumberg and his collaborators developed the supervised machine-learning and deep-learning models that can identify histopathology stains, discriminate between tissue types, and classify images by disease state. The project employs "the first pan-tissue pan-disease (i.e., from infection to malignancy) method for prediction and search on social media, and [is] the first pathology study prospectively tested in public on social media," according to the *Modern Pathology* article.

Yet Twitter, like all social media platforms, has its drawbacks—among them, the difficulty of finding comparison cases for rare disease entities. It was this problem that inspired Dr. Schaumberg to incorporate data from PubMed into Pathobot's data set.

Dr. Schaumberg recognized the need to extend Pathobot's reach when Daliah Hafeez, MD, shared a liver subcapsular case on Twitter in which it was unclear whether the rounded structures shown in the photomicrographs were parasite eggs, nematodes, or vegetable matter mimicking disease. Dr. Stayerman responded "Perforation? Fistula?" to indicate how lentils may have reached the liver. "Pathbot found a similar lentils case from a colleague, and our internal tools for searching PubMed also found similar lentil cases," Dr. Schaumberg says. Dr. Hafeez concluded that the patient had a duodenal fistula that had healed. "Vegetable matter it is," she posted to Twitter.

It was after using internal tools to search PubMed for cases that might provide additional context for the lentil finding and posting those cases to the Twitter thread that Dr. Schaumberg realized Pathobot could perform an automated PubMed search for pathologists. To find relevant articles for the PubMed data set, the team used its machine-learning system trained to differentiate H&E images from all other images. The system filtered more than one million PubMed articles for figures that included images of H&E-stained tissue.

The number of cases in Pathobot's social media database can also be a limiting factor as a result of too few similar cases being returned for comparison. To address this shortcoming, Dr. Schaumberg encourages pathologists to contribute data by sending a direct message to @pathobot on Twitter or going to his website, www.pathobotology.org/contact. He also encourages users to repeatedly use @pathobot on the same case as more photomicrographs or diagnoses emerge to help refine Pathobot's search results.

As Dr. Schaumberg delves deeper into AI and labors over Pathobot, he's looking to the future. His plans include incorporating new sources of data, such as The Cancer Genome Atlas, into the project and shoring up gaps in the current data. "Extending the data set to whole slide images may make our methods more applicable to hospitals since hospitals often have the resources to invest in whole slide imaging," he says.

"It's exciting that AI can be applied at the clinical level rather than just the basic science level," he concludes. "To me, that's very motivating." □□—*Charna Albert*

X-Lab making inroads in laboratory network outreach

X-Lab has announced that it is focusing on bringing its product for digitizing and automating the laboratory test referral and results data-reporting process to the United States. The system, which will be marketed under the name Labgnostic, is designed to connect diagnostics laboratories worldwide through a single connection to an interoperable hub.

In March, the United Kingdom's National Health Service mandated that all NHS trusts transfer externally referred SARS-CoV-2 test requests and results through the system, which is marketed under the name NPEx in the United Kingdom. The NHS selected the technology solution for its ability to share workload capacity information and provide faster turnaround times and safer data exchange.

“With a hub connecting diagnostic systems across the UK, Ireland, and into Europe, X-Lab [is] now exploring use cases for this technology across other continents,” according to a press statement from the company.

X-Lab, 424-367-8081

Dr. Aller practices clinical informatics in Southern California. He can be reached at raller@usc.edu.