



First generation saliency maps of OSIC AI models

New Consortium Takes Aim at IPF

Idiopathic pulmonary fibrosis (IPF) is a progressive disease that affects more than 125,000 people in the United States, resulting in more deaths each year than breast cancer. By the time it's identified, most people with IPF are well within the three- to five-year window of life expectancy that's typical from the time of diagnosis. This is due in part because they've been misdiagnosed multiple times, in part because it takes a year to confirm the diagnosis before proceeding with treatments and, mostly, because IPF is rare. At least, that has been the prevailing belief.

"In reality, fibrotic lung disease is more common than we thought. We just weren't picking it up," notes Simon Walsh, MD, a radiologist at the National Heart and Lung Institute at Imperial College London, and lead radiologist of the Open Source Imaging Consortium (OSIC). "Because of that, we have shown, fairly conclusively, that there is a lack of imaging expertise when it comes to evaluating patients with suspected interstitial lung disease on high resolution CT, which is the cornerstone of diagnostic imaging. [That's] the weak point in the chain."

OSIC began with a simple idea: make high-resolution CT chest scans widely and openly available and use AI to assist with identifying biomarkers and other patterns associated with IPF. What evolved, says OSIC executive director Elizabeth Estes, "is unequivocally the most complex project on the planet because not only are you [working] with a pretty siloed industry, but you're also dealing with issues like cross-border rules, patient privacy, [general data protection regulation], and all of the complexities of every university or academic medical center."

Despite the magnitude of the challenge, the OSIC data repository, which launched last September, is well on the

way to meeting its 2022 goal of acquiring 15,000 anonymized scans, with accompanying clinical data from an international community of partnering academic institutions and pharmaceutical companies. Personal identifiable information is removed before the scans, and

clinical data are released and uploaded to the repository, where they're available to computational scientists, radiologists, and clinicians who are encouraged to build algorithms from their observations. Estes predicts OSIC will have 25,000 scans committed by the end of the year, "which will be by far the largest database in the world for interstitial lung disease," she says. "It already is one of the largest and most diverse [databases] because the [uploads] are coming from everywhere."

The data analysis process is managed by OSIC's Science Leadership Team, consisting of Walsh; pulmonologist Kevin Brown, MD, of National Jewish Health in Denver; and computational science lead David Barber, PhD, a professor of machine learning at University College London. The leadership team oversees working groups focused on data architecture and privacy; publication, research, ethics, and oversight; biomarkers; and OSIC's challenge competitions and governance.

Walsh oversees the data group, which meets weekly to provide quality control on all of the information coming into the database. "It's the pairing of the clinical data and imaging data that is so critical," he says.

Networking

The OSIC network includes organizations based in industry, education, and patient advocacy. Some have criticized OSIC, an open-source project, for making membership free to academic institutions while charging for-profit companies to join. In an effort to make participation as inclusive and equitable as possible, membership fees are determined on a sliding scale based on revenue, according to Estes.

Membership fees support the cost of developing and maintaining the database, which uses Microsoft Azure and an interface designed by PricewaterhouseCoopers. Everything that evolves from the data created through OSIC-sponsored activities—for example, digital imaging biomarkers

and treatment algorithms related to IPF and other interstitial lung diseases—will be openly available to the medical and scientific community.

Will Perry, lead engineer on the OSIC project for PricewaterhouseCoopers, says, "It is the open-source model that we hope becomes the future in solving other complex diagnostic problems. We want to open up health care for the benefit of patients by bringing divergent groups together into a coalition, turning competitors into willing collaborators, and leveraging their collective knowledge to help accelerate the fight against other complex problems."

Membership also supports OSIC's open competitions on Kaggle, a data science community platform. The first Kaggle challenge in 2020 invited participants to develop algorithms to predict lung function decline in patients with pulmonary fibrosis based on images in the database. A prize of \$55,000 was divided among three winning investigators out of 2,500 global competitors. OSIC will announce its second Kaggle challenge later this year.

According to Barber, who manages algorithms for OSIC and organizes the Kaggle competitions, the medical community is only beginning to embrace the idea of openly sharing data in the way that the AI community has for so many years. As the database grows, so does the potential of developing an algorithm to predict the progression of IPF in individual patients.

In the weeks and months ahead, OSIC will continue its outreach to governmental agencies, patient advocacy groups, and patients themselves in an effort to raise awareness about its mission and broaden participation in the database.

Walsh hopes that patients with IPF, who typically feel disenfranchised, will be inspired by the advancements underway at OSIC. He also encourages radiologists and computer scientists to embrace this unique and important opportunity to collaborate. "These diseases are difficult to diagnose, and they require radiologists with decades of skill to interpret [the scans]. Computer scientists are looking for interesting challenges to engage their skills, and we would welcome them to come and look at our data sets and get involved." For more information on OSIC, visit osicild.org. ■

Claudia Stahl is a freelance writer based in Ambler, Pennsylvania. She specializes in writing about the health of people and the planet.