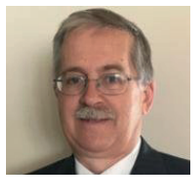
Solving the persistent shortage of clinical data available to medical algorithm developers

More and more, radiologists are turning to Artificial Intelligence as a possible means of handling the ever-increasing growth in the number of imaging exams that they have to read and report. The diagnostic accuracy and reliability of an AI-developed algorithm depends on the quality and applicability of the images used to train the algorithm. To meet this requirement a recently formed company, maiData Corporation has been set up with the express aim of solving the persistent shortage of clinical data available to medical algorithm developers. We spoke to Julian Marshall, CEO of maiData and Dr Robert Nishikawa, Professor of Radiology at the University of Pittsburgh.



Julian Marshall, CEO of maiData
email: julian.marshall@maidata.io

Prof R ."Bob" Nishikawa, Professor of Radiology, University of Pittsburgh, PA, USA
email: rmn29@pitt.edu

Q *Just why is it so important that AI developers get larger data sets?*

JM: There are in fact several reasons that developers need larger data sets. ***First,*** they need to avoid bias in training data, which basically means that the training data needs to be sufficiently broad to ensure that it includes all manifestations of subjects with both the disease — or features they're going after — as well as normal subjects. ***Second,*** and similarly, the data sets used for testing the algorithm have to represent the population at large to ensure generalizability. ***Third,*** the datasets need to be plentiful enough to allow AI companies to segregate data into separate Training, Validation, Testing and Regulatory Approval databases [Figure 1]because that's the only way to minimize the risk of overtraining and to get reasonable stand-alone performance results.

The process of bringing an AI algorithm to market requires the use of clinical data at several points. Ideally, the data used at each point will be different. But since the late 1940s, scientists have blurred the lines between these data sets using techniques such as jack-knife testing (1949), leave-one-out cross-validation, etc. But the cost of using these methods is that performance measured in the lab is not realized when algorithms are deployed clinically. The result can be products that don't work as expected on unknown populations, and performance expectations that are inaccurate. In addition, improper handling of data can put products into regulatory jeopardy.

AI companies have traditionally created their own relationships with individual clinical facilities, which is time- consuming and adds unpredictability in cost, effort and quality that can result in development delays. maiData can help AI companies streamline case collection efforts and provide seamless delivery of large volumes of medical images and metadata for developing robust AI algorithms.

BN: In terms of diminishing returns, as the dataset gets bigger, there will be smaller improvements in the performance of the AI algorithm. However, until you get very big datasets (and I admit I don't know exactly how big "very big" is) you can always gain in robustness if not accuracy.

However let me point out that not only do you want to enlarge your dataset, but you want to diversify it also. So collecting cases from multiple sites is important, not only for different patient demographics, but also for different radiologist interpretations. There is a lot of variability between radiologists. This has been well documented for breast and chest images. Cases selected from a site depends on what the radiologist calls normal and abnormal, and this will differ somewhat between radiologists. For example, in breast imaging, cases (cancerous, benign and normal) from a radiologist with

a low recall rate will have different characteristics than cases from a radiologist with a high recall rate. Since the radiologist is essentially the gatekeeper to what goes into training an AI algorithm, you want to have as many different gatekeepers as possible to remove any potential bias.

I am sure that companies have such datasets, as presumably each release of a new software version improves the algorithm performance. Some of the improvement is from collecting more cases over time. Given enough time, it is possible to collect a large and diverse set of images, but it is very difficult and cost prohibitive to do so in a short period of time. Time to market is a critical factor in the success of any company. As a result, datasets are continuously collected over time enabling more accurate and robust algorithms to be developed.

One big advantage of deep learning (DL) techniques is that they scale well with database size. In the pre-DL days, where multiple features were extracted from images and input into a statistical classifier, as the database got bigger it became more difficult to develop and train an algorithm. With DL, the training takes longer but it is relatively easy to manage. Obviously, the Googles of the world and large academics sites have sufficient computing power for most AI applications. Start-up companies are at a disadvantage there. However, relatively simple DL models are available and can run on a $15K computer. And while more complex models are being developed, depending on the application they may not be needed. In the pre-DL days, some statistical classifiers were better than others, but the increase in performance based on classifier choice was really quite small. Similarly, the data given to the model is more important than the DL model themselves.

Q *Following up on the point about the need to diversify the data set, don't some people maintain that the best population to train an algorithm is the local one from the area in which it will ultimately be applied.*

JM: Yes indeed, you often hear talks at conferences discuss training algorithms for a specific, local population, but that makes my alarm bells go off for several reasons. First, what if someone moves here from elsewhere? Would the algorithm work on them? Second, how is the quality of the algorithm assessed from a regulatory standpoint? If an algorithm is trained, or tuned, specifically for a local population, was FDA clearance or MDR approval granted afterwards? And how do you stop an algorithm training in one locale from being moved to another locale where the training may not be applicable? Third, patients deserve all AI algorithms to be designed and tested with sufficient rigor. If an AI algorithm is trained locally, how will the process be controlled, from a Quality Management System perspective? It reminds me of continuous training AI algorithms … a great idea as long as the on-going training doesn't send algorithm performance off-track.

An AI algorithm that really helps doctors must be generalizable, which means that it must work on all intended patients. This, in turn, means the algorithm must have been trained on a dataset that closely resembles the target population. For example, an AI algorithm for detecting breast cancer should be able to detect lesions irrespective of the ethnicity of the patient. But how does the developer know it will work if it hasn't been trained or tested on a broad population that represents all ethnicities?

AI in medicine is all about helping individual patients, but algorithm developers require vast data sets to ensure the quality and efficacy of each and every algorithm.

Q *Are the advantages/requirements for large and diverse training sets more important for certain body parts/pathologies than others? For example, it's known that the average breast density of Asian women is greater than that of Caucasian women, but are there similar differences between racial populations for other organs?*

JM: I think you could look at the needs of a breast algorithm here vs a lung nodule tool…all AI are not created equal; just because you can look for breast cancer doesn't mean it can also do lung; that's why it is critical to develop, train and test specific algorithm with large, specific databases. You could also support your point about local populations – e,g, if you train on Asian women, what happens when that tool is used in Detroit?

JM: In the way the question is framed, it is more for physicians and anthropologists to answer than for maiData. Environmental conditions, most specifically diet, are believed to influence body habitus in many ways. Referring to the example cited, it has been shown in the literature that the breast density of Asian women who move to western countries is reduced, while breast cancer risk is strongly associated with the origins of their ancestors, both in terms of genetics and urban habitat.

At a macroscopic level, people are

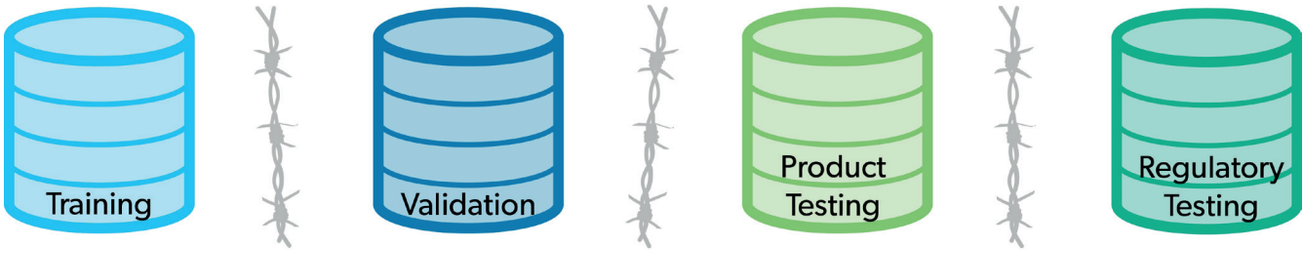


Figure 1: Database Segregation Best Practice

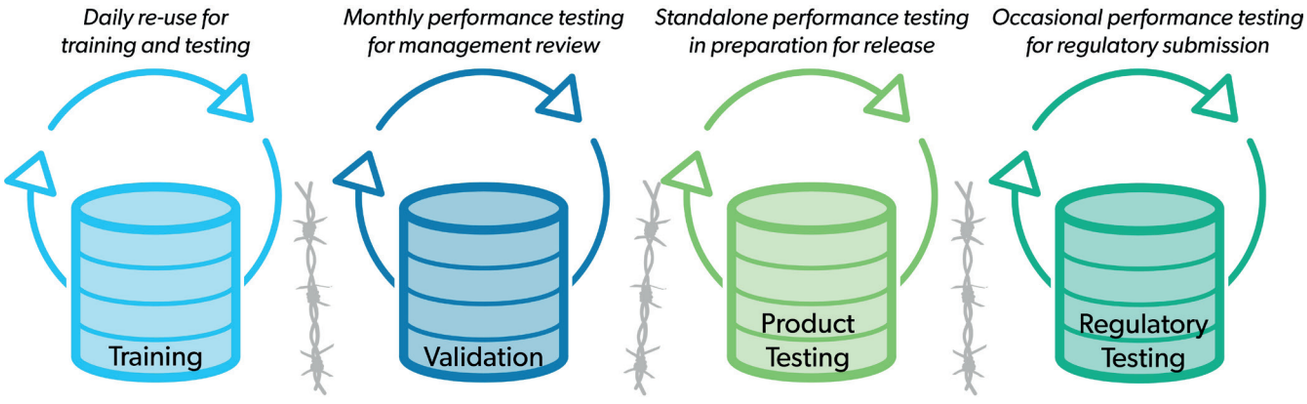


Figure 2: Ongoing Use of Segregated Data

the same. But as you zoom in, there are many distinct differences. A great example is seen in mammography. If you look at images of the left and right breast in the same patient, the breasts are grossly symmetrical, but there are lots of nuanced differences when you look in detail. I think that alone provides the rationale for large, diverse data sets. The data set has to be big enough to at least attempt to cover normal human variation.

Q *What currently is the radiological field that is most active in AI development?*

BN: From my perspective, all modalities and most body organs are active areas of AI development. For example, while COVID-19 is relatively new disease, the National Institutes of Health in the US has created and funded the Medical Imaging and Data Resource Center at the University of Chicago, led by Maryellen Giger. The Center will create an open-source database of medical images along with associated data from thousands of COVD-19 patients. The goal is to help understand better and treat the disease through the use of images and AI. This shows that many believe that medical images and AI can be used to tackle *important health problems.*

Q *There seem to be more and more prototype/development papers being published on AI but what about the number of algorithms in actual routine clinical use?*

BN: According to the ACR Data Science website (*https://www.acrdsi.org/DSI-Services/FDA-Cleared-AI-Algorithms*), the FDA has cleared 79 AI algorithms. The majority of these were approved in 2019 or 2020. While many of these are being used clinically, it still too early for any large studies of actual routine clinical use.

"...maiData's unique solution will reduce the cost and burden of data collection for AI companies and shorten AI algorithm time-to-market, meaning that clinicians get clinical decision support more quickly..."

However, proper integration of the algorithm is critical. That is one lesson that is evident from the clinical implementation of computer-aided detection (CADe) for screening mammography. Early studies showed that CADe could increase the cancer detection rate with a commensurate increase in the recall rate. However, long term studies showed a different result. The study by Lehman *et al.* showed that compared to a period when radiologists did not use CADe, it turned out that reading with CADe actually decreased sensitivity and increased specificity. This is exactly the opposite of what was expected based on the earlier studies. In fact, that result is impossible if radiologists used CADe as a second reader, as the system was labeled for use by the FDA. The most plausible explanation is that radiologists did not use CADe as a second reader, but probably used the system to read more quickly. That is, using CADe as a second reader was not viable and therefore the system was not properly integrated into the workflow. I believe that in parallel to algorithm development, an equal effort (in terms of money and time) should be devoted into integration issues; otherwise, the AI tools will have a diminished and possibly a negative impact on radiologists' performance, and ultimately, patient care.

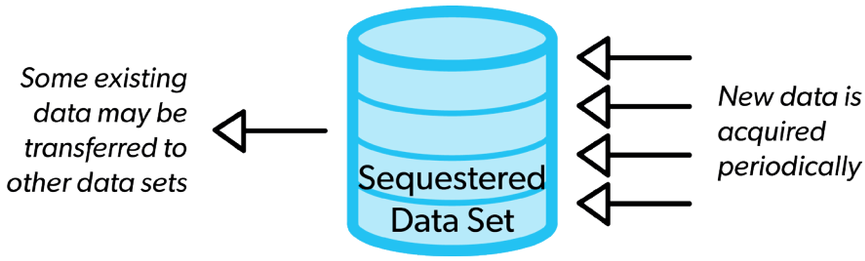


Figure 3: Simple Data Refreshing.

Q *Now, turning to maiData, what was the rationale for its creation, what precisely are the company's objectives and how in practice will the company attain these objectives?*
What is it that your company has which makes it worthwhile for developers to go through you rather than try to source images directly?

JM: maiData was founded to solve the persistent shortage of clinical data available to medical algorithm developers. maiData's unique solution will reduce the cost and burden of data collection for AI companies and shorten AI algorithm time-to-market, meaning that clinicians get clinical decision support more quickly, which can result in better patient care.

Over the last three decades, every company I've worked in has been starved for data for algorithm training, validation, testing and regulatory approvals. And, if you listen to developers speaking at AI companies, that is a shared experience. So, I thought "*let's solve this problem once and for all*". There are really two steps to what we are doing. First, gain access to a very large, very diverse data set. And second, start delivering meticulously pseudonymized and annotated data to the AI companies that need it. There are four reasons that AI companies will benefit from using maiData, all of which make data collection faster: we will already have the contracts in place; we will already have been through IT security approval and have our data collection software installed; and pricing will already be set. maiData removes many burdens from AI companies.

Q *If a radiologist reading this interview article wants to get involved, what practically does he/she have to do?*

JM: In my experience, most clinicians understand that their patients will benefit in the long-term if AI training datasets are much, much larger. I sense an altruism when I talk to clinicians and administrators … they want the datasets to play a part in the development of AI because they know that will help not only their patients, but all patients in the future. maiData is happy to talk to more facilities about collecting cases, including in Europe. maiData is starting in the United States because, especially with COVID-19, it is easier to start closer to home. But, we are mindful that European and Asian data needs be available to developers (to reduce bias and increase generalizability). We want to be extremely careful of regulations around patient data and privacy, such as GDPR, and mindful of the ethics of using patient data for commercial purposes.

" Artificial intelligence will not replace radiologists … but radiologists who use AI will replace radiologists who don't"

Curt Langlotz, Stanford University

Q *In the company's streamlined approach to data collection how are issues such as Intellectual Property handled? E.g. Images taken of a patient, even though anonymized presumably still belong to the patient; not to mention the annotation/report by the radiologist; Who's got the IP? What about liability issues ?*

JM: There is no single answer to this question because the ownership of images, reports and metadata varies between jurisdictions. The universe of data used to train an algorithm does help developers create IP. However, taken alone, the images, metadata and reports from a single individual have little influence in the result. There is a big societal good question here. If we all want the benefits of AI (earlier detection of disease, better outcomes, better workflow for clinicians), then we all have to accept that our data, properly pseudonymized, needs to be part of the solution.

Q *Finally broadening out from AI development, how are radiologists in general now reacting to the prospect of AI? Are we past the stage of them worrying about their job –are they now keen to get hold of AI to ease their work-load?*

BN: I think we are over the hump of radiologists fearing for their jobs. Radiologists do more than interpret images and once you understand the complexity and diversity of the challenges radiologists face daily you can envision how AI can be integrated to make radiologists more efficient and effective. I think it is a very exciting time for radiology. Our eyes can only extract a fraction of the information contained in a medical image and can only do so qualitatively. AI can unlock the hidden information in a quantitative manner making radiology an even more valuable component for optimal patient care. As for attracting young medics into radiology, I think it is at a low point right now because of the perceived threat of AI. As more and more AI tools are integrated into radiology, thefield will again be favored by those who are technology savvy. I can see a future, where images, image biomarkers and genetic/molecular biomarkers are the fundamental information used to decide on proper patient management.

JM: Radiologists are smart. They know there are certain tasks that are difficult for the human observer, and they do want the help. The American College of Radiology has set up a Data Sciences Institute just to deal with AI on behalf of ACR's members, who do worry about the influence of AI. But as Curt Langlotz (Stanford University) is famously quoted "Artificial intelligence will not replace radiologists … but radiologists who use AI will replace radiologists who don't." Medical schools need to do a better job of talking about AI with students to ensure a supply of radiologists going forward. A well-known radiologist from the UK once told me "*AI will probably take over screening, allowing radiologists to spend a lot more time on careful diagnostic work-ups*". I think that's a good thing both for radiologists and for patients.

More Information:
maiData
Palo Alto, CA, USA
www.maidata.io