

Artificial intelligence and high processing systems expand education and research capacity of Heart Institute (InCor)

Data anonymization project conducted in partnership with Intel expands research possibilities by using the results of imaging tests performed and stored by InCor



Created in 1977, InCor is today one of the largest cardiology centers in the world in terms of number of visits and cardiology and pulmonology subspecialties, and a reference institution in academic research, supported by the results of hundreds of thousands of imaging exams performed annually by the institution.

Since 2012, InCor's research team has worked on a project for patient data anonymization to ensure data confidentiality, which will allow them to use large volumes of exams in their research. In 2017, Intel started to support this project, inserting artificial intelligence and high processing capacity in the method that will process imaging exams and make them available to different research teams.



Challenges

- Automate the process of patient data anonymization from imaging exams;
- Increase the imaging volume available to research groups;
- Allow remote search without sharing critical data of patients.

Solution

- Use of artificial intelligence and high processing capacity to identify and anonymize critical data, and automate the process.

Impact

- At the end of the first stage of the project, the data recognition rates vary from 60% to 100%, depending on the type of imaging exam.



“Everything should be validated by the end of this year, resulting in process that will scientifically prove the method of identifying information in exams.”

Marco Antonio Gutierrez,
IT Chief at InCor.

The importance of imaging exam results in medical research

Every year, the Heart Institute of the University of São Paulo (InCor) performs about 200,000 imaging studies, and each of them may contain one to thousands of images. In addition to serving the hundreds of thousands of patients who visit the institution annually, the institution has built an essential database to conduct medical research.

In addition to being a reference center in cardiology, InCor is also an important center for medical research, and many of these studies are based on the institution's collection of exams. However, medical ethics define that sensitive data of patients cannot be disclosed by researchers and must be kept confidential.

Sensitive data refer to information that allows patient identification, such as name, age, birth date, among others; and these data cannot be shared by hospitals with third parties without prior consent of the patient. Until now, data anonymization was extremely difficult due to the lack of a standard implemented in the place where they are stored in imaging exams and due to the display format of each exam (which varies significantly).

Today, data privacy is ensured by confidentiality agreements signed by the researchers, who agree not to disclose critical data about the cases studied and only use them responsibly. But this model does not allow a massive use of exams, which need to be evaluated and have their data anonymized manually.

Because of this, since 2012, the InCor Technology team has worked on a project to eliminate this information in an automated way. According to Marco Antonio Gutierrez, IT Chief at InCor, medical equipment already records patient data as part of the image itself. “And this is done without standardization. Information such as weight, height, age and birth date can be found in different places in the image,” he says.

Gutierrez also says the initial idea was to develop a robot to do the job. “Then, we created a prototype to identify and blur sections with text information. That was

our first effort to solve this problem,” says Gutierrez, highlighting that the model had to be improved, as it blurred all text information.

“A lot of information, such as exam date and measurements, for example, needed to be kept. We had the challenge of teaching our robot how to distinguish sensitive data from regular data,” he says. According to Gutierrez, in this first stage of the project in 2017, the InCor Technology team divided the process into three stages: text identification in imaging exams, text conversion into a sequence of characters, and information blurring.

Increased processing capacity and use of artificial intelligence (AI) to identify sensitive information

In 2017, the robot development project gained a new momentum. InCor had built a partnership with Intel, which identified AI as one of the areas of interest for the development of new solutions. According to Gutierrez, the joint effort with Intel added one more stage to the process: deploy a classifier through artificial intelligence.

“AI is used in a stage before we blur information, it identifies whether the information is sensitive or not. For this reason, we've trained a neural network based on annotations done by one person. After this training, the neural network identifies what is sensitive in the text and blurs it, making it confidential,” he explains.

In this project, Intel assigned a team of software experts who, since 2018, have provided specialized resources to assist customers in their first AI-related developments. “We analyzed how to use this type of resource to fulfill InCor's demands and identified the need for data anonymization from imaging exams performed by the hospital, allowing their use in large batches in research,” explains André Ribeiro, New Business Director at Intel Brazil.

The first step was a pilot conducted in the last quarter of 2018, when Intel assigned technicians with experience in AI solution development and InCor assigned its own technical team. “Together, we created this prototype for imaging exam anonymization. We're not

only talking about X-rays, but also magnetic resonance, tomography and other exams, each of them performed in a different machine, from different manufacturers,” adds André.

It means that each exam has different information placed on different parts of the image. In practice, the AI system takes a number of different exams generated in different platforms and submits them to a machine learning system, teaching the computer how to recognize what data from that image can identify a person. Once these data are identified, this information has to be blurred so that researchers cannot identify the patient.

In total, 680 images were used to train the algorithm, and later other 320 images were used to test it, including different types of exams – some containing electrocardiographic tracing. This group of 1,000 images comprised the first stage of algorithm training, and the results were considered positive by the InCor team.

“We achieved 100% accuracy in cineangiography exams, about 60% in cardiovascular ultrasound exams, and more than 50% in electrocardiograms,” reveals Gutierrez, highlighting that, with humans, the margin of accuracy in identifying sensitive data is, in general, 85%.

In addition to identifying sensitive data of patients, the system records all images in DICOM (Digital Imaging and Communications in Medicine) standard, ensuring information compatibility between different devices.

With these results, InCor is preparing to conduct in 2020 a new stage of neural network training, now with a batch of 10,000 images. This new stage will also be supported by Intel. “It’s a medium- and long-term work to be performed by InCor with our support. In this second stage, our role is to provide technical support and see how to make the algorithms more efficient to run on the InCor platform,” explains André Ribeiro.

“Everything should be validated by the end of this year, resulting in process that will scientifically prove the method of

identifying information in exams,” says Gutierrez, emphasizing that the effort made so far is focused on research and development.

Now, the institution is focused on the scientific field, which would already produce impressive results. Currently, InCor conducts around 200,000 imaging exams every year, with each exam containing one to thousands of images. “We have over one thousand research projects running at the house. The subsets of these studies will certainly be used in these projects”, predicts Gutierrez.

In addition, the institution has 2 million images of all types stored in its system, with information from 1.4 million individuals. “By anonymizing sensitive data, we have the possibility of conducting thousands of predictive analyses,” he says.

In addition, the system should further increase the efficiency and quality of InCor's service. According to Guilherme Rabello, Commercial and Market Intelligence Manager at InCor, the system will become more robust in this new stage. “Our focus is to develop something to impact patient care, allowing improved patient care and more accurate studies,” he says.

“Our focus is to develop something to impact patient care, allowing improved patient care and more accurate studies.”

Guilherme Rabello,
Commercial and Market
Intelligence Manager at InCor.

“Together, we created this prototype for imaging exam anonymization. We’re not only talking about X-rays, but also magnetic resonance, tomography and other exams, each of them performed in a different machine, from different manufacturers.”

André Ribeiro,
New Business Director
at Intel Brazil.

