

HOW TO AUTOMATICALLY ANONYMISE MEDICAL IMAGES?

CURRENT SITUATION AND THE CHALLENGE

In Lithuania, the need for the re-use of health data is growing. The re-use of health data can contribute to progress in the areas of research and development, innovation, education, knowledge management, health policy-making, health planning and statistics. In order for health data to be processed, it must be anonymised (processed so that a person can no longer be directly or indirectly identified), thus ensuring the protection of personal data.

One type of health data is medical images. To prepare medical images for reuse, it is necessary to mechanically find and hide personal data in each image. Because this process is time-consuming, the practice of anonymising medical images is not widespread, and the potential for reuse of medical images is untapped.

With the development of an automated medical imaging anonymisation tool, a variety of medical images could be prepared for re-use without significant time and human resource costs.

WHAT IS THE MAIN PROBLEM

The process of anonymising medical images (one of the types of health data) is different from standard anonymisation processes for text or digital files. In addition, different medical devices capture different personal data in a medical image and often place it in different locations of the medical image. There is no solution on the market for automatic anonymisation of medical images. Solutions that are in the market today can only anonymise metadata, but many medical devices store personal data at the medical image level (DICOM format) rather than in a separate metadata file.

WHAT IS THE SCOPE OF THE PROBLEM

It is estimated that currently about 500 million medical images that could be adapted for re-use` are accumulated in Lithuania. Anonymisation of these images would contribute to the development of the Lithuanian life sciences industry, more accurate diagnosis and improvement of medical services. In the absence of an automated medical image anonymisation tool, the potential of this health data cannot be exploited.

DESIRED RESULT

Automation of the medical image anonymisation process is necessary when large volumes and medical images from different medical devices need to be processed, due to the high time required to prevent reuse of medical images at present. The use of anonymised medical images would contribute to improving the quality of medical services, digitizing services, reducing human error and saving resources.

SOLUTION MAIN FUNCTIONS

The solution should automatically anonymise and delete all information related to personal data contained in both the image itself and a separate metadata file, and be adaptable to different medical images generated by different devices, without requiring specific data science competencies. The requirements for the anonymisation solution are given in the table below.



Anonymisation tool requirements (the tool should have all the listed functionalities)

NR.	FUNCTION
1.	Development and implementation of an anonymisation tool; anonymisation of data only at the DICOM level, removal of structural reports from the study (DICOM SR, DICOM SC, DICOM PDF, DICOM ECG, and others).
2.	Development and implementation of an anonymisation tool; anonymisation of data at the image level (the dicom file is decomposed and the anonymised image is saved in *.jpg, *.mpeg format. This method is suitable when image analysis is performed by AI).
2A.	Creating a tool to ensure that information at the image level is 99.99% anonymised.
3.	Development and implementation of an anonymisation tool; anonymisation of data at the image level while maintaining the original DICOM format of the file (information is anonymised at the image level, leaving the original DICOM format and DICOM format structure. This method is suitable when anonymised images are analyzed by users with DICOM viewers, image reconstructions are performed).
3A.	Development of a tool to ensure that information at the image level is 99.99% anonymised and the DICOM structure is maintained.
4.	Anonymisation of data at the DICOM PDF, DICOM ECG, DICOM SR image, and DICOM level while maintaining the original DICOM format.
4A.	Creating a tool to ensure that information in DICOM PDF, DICOM ECG, DICOM SR is completely depersonalized.
5	Creating a database of anonymised images and storing anonymised images in *.jpg format.
6	Creating and storing a database of anonymised images in DICOM format in PACS system.
7.	Algorithm for anonymisation quality assessment.
7A.	Development and implementation of anonymisation quality assessment tool.

INDICATORS OF SUCCESS

The success of the solution will be measured by the depreciation reliability indicator. The reliability requirements for the medical image anonymisation tool are high (> 99.99%) to ensure compliance with the 2016 April 27 Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46 / EC (GDPR)

AVAILABLE DATA

Medical images collected by Santara clinics with fake personal data can be used to create the image anonymisation tool, so they are not subject to GDPR and can be used without legal restrictions. There are 2,405 such images from 7 different examination methodologies (computed tomography, electrocardiogram, X-ray radiology, and others). This data set is sufficient for the challenge activity and for determining the effectiveness and quality of anonymisation of the developed tool.

FUTURE OPPORTUNITIES

A successfully developed automated anonymisation tool for a variety of medical images could be used as needed by individual healthcare institutions to develop innovative solutions to optimize treatment processes, improve governance, or provide anonymised health data to third parties (researchers, startups, pharmaceuticals, etc.). development of commercial solutions.

ADDITIONAL SUPPORTING INFORMATION THAT MAY BE RELEVANT TO SOLUTION DEVELOPERS

- Improving the availability of health data: current situation analysis
- Availability of health data
- Anonymisation methods: recommendation
- Life Sciences Industry: Looking to the Future volume
- How we can place a value on health care data

