Kraków Applied Physics and Computer Science

Summer School'20

15th of Septeber 2020

AGH

Getting to know the Nvidia Clara environment and analysis of federated learning opportunities with the help of a local instance.

Presenter: Maciej Krzywda

Project supervisor: dr hab. inż Tomasz Szumlak

Project goals

- Searching for a comprehensive solution for the analysis of medical images (CT, MRI, x-ray, ultrasound etc.)
- Speed and optimized alghoritms
- Working with multiply databases

From project goals to research & development

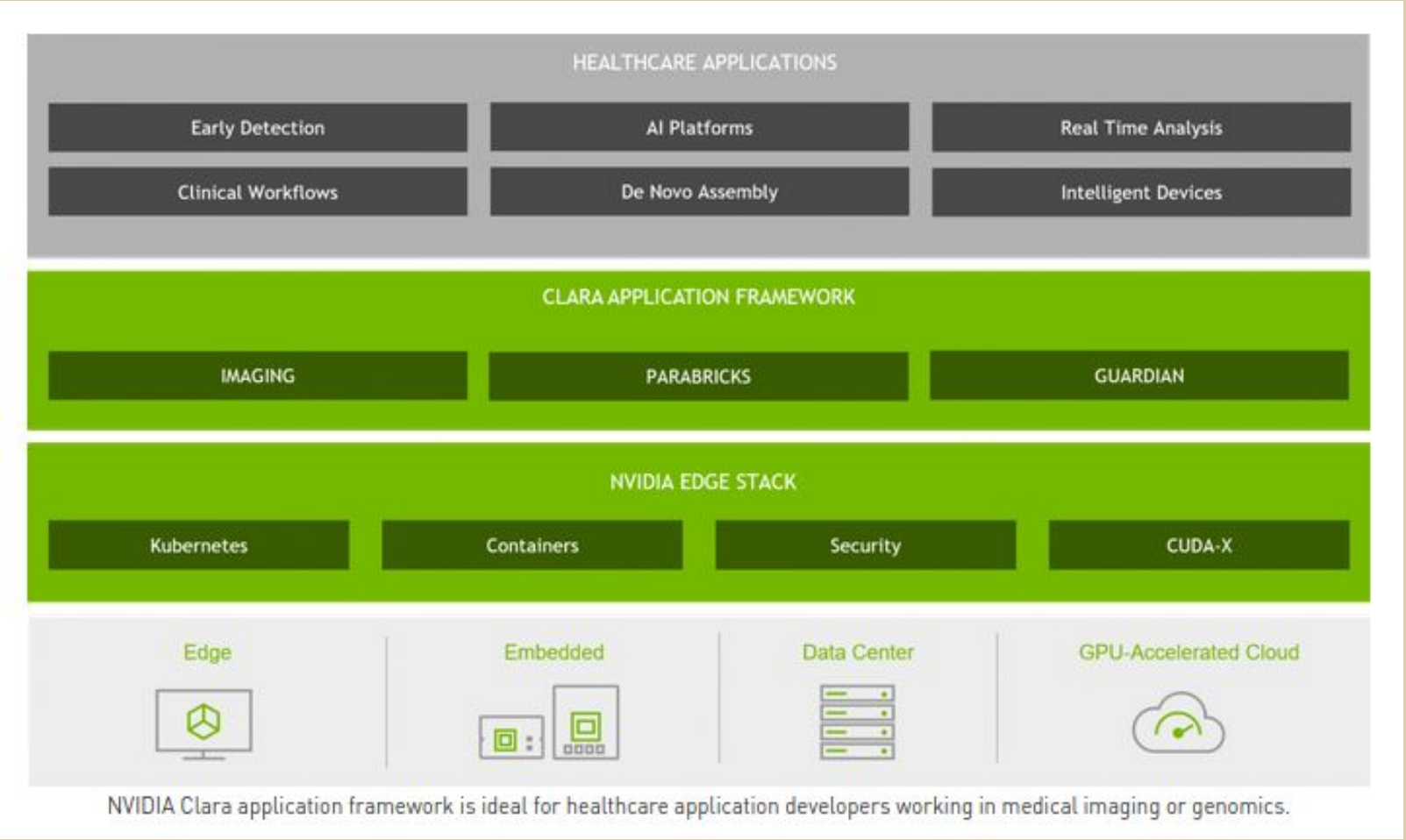
- Searching for a comprehensive solution for the analysis of medical images (CT, MRI, x-ray, ultrasound etc.)
- Speed and optimized alghoritms
- Working with multiply databases

NVIDIA® CLARA™

NVIDIA CLARA

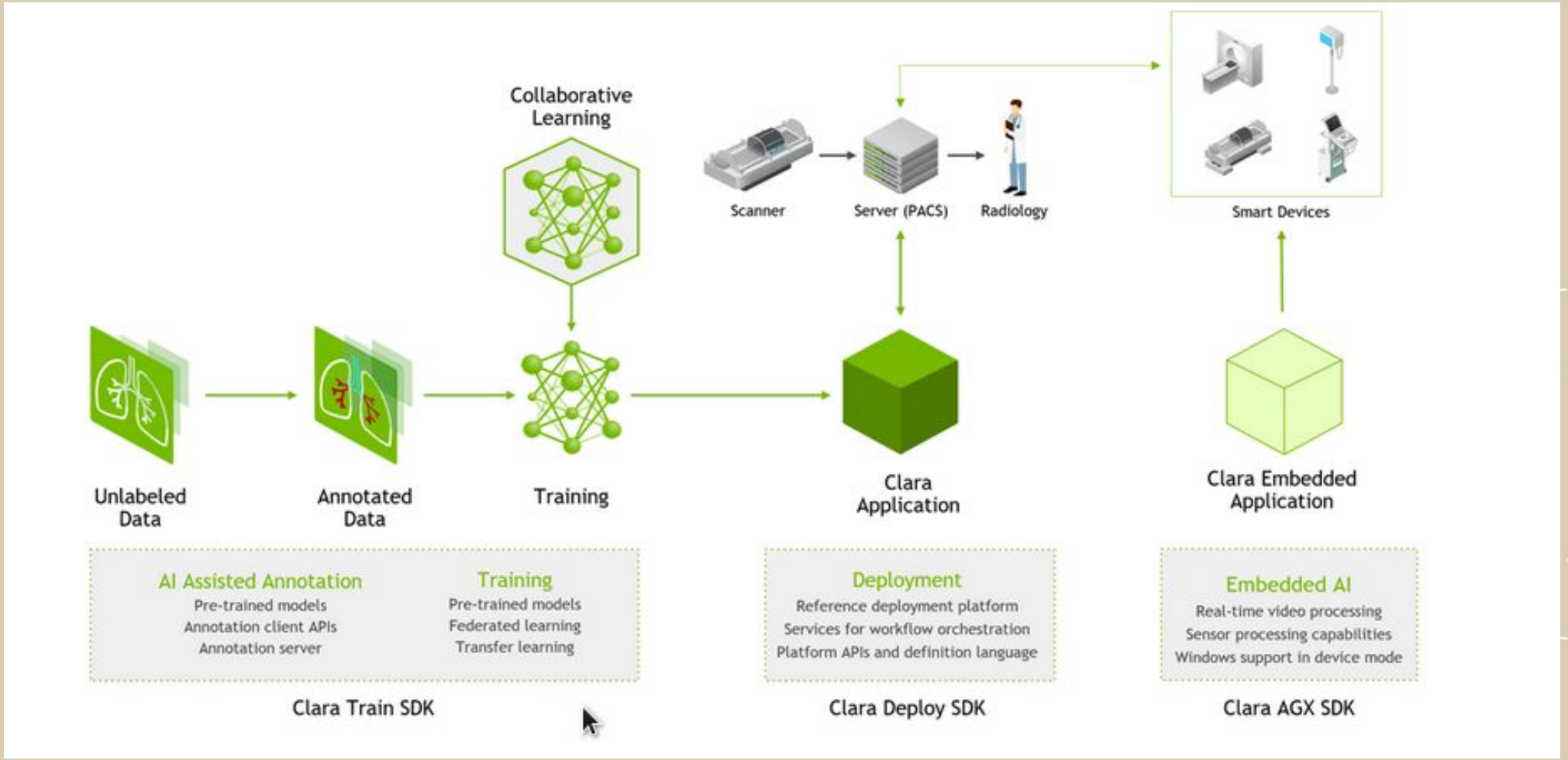
NVIDIA Clara is a healthcare application framework for AI-powered imaging, genomics, and for the development and deployment of smart sensors. It includes full-stack GPU-accelerated libraries, SDKs and reference applications for developers, data scientists and researchers to create real-time, secure and scalable solutions.

NVIDIA CLARA



NVIDIA Clara application framework is ideal for healthcare application developers working in medical imaging or genomics.

NVIDIA CLARA



Data Formats in Healthcare

- **DICOM** - Digital Imaging and Communications in Medicine (DICOM) is the standard for the communication and management of medical imaging information and related data

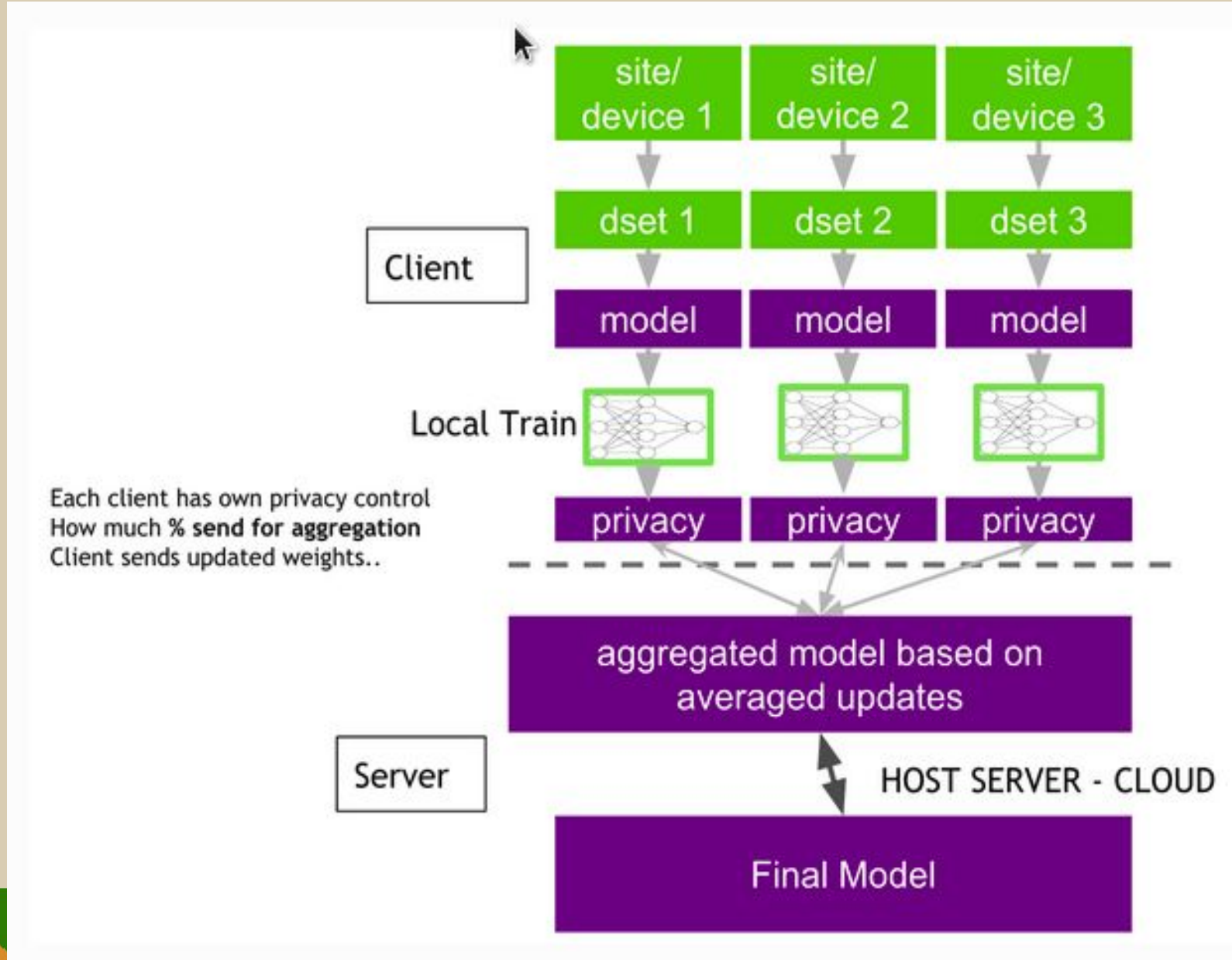
- **DICOM RTSTRUCT** - DICOM RT Structure Set ("**RT**" stands for radiotherapy

- **NIFTI** - The Neuroimaging Informatics Technology Initiative (NIfTI) is an open file format commonly used to store brain imaging data obtained using Magnetic Resonance Imaging methods.

NIFI (default data format fo NVIDIA CLARA)

If your native data format is different from **NIfTI** or if you want to convert the image and label mask to isotropic resolution, you can use the provided data converter or some other software of your choice, such as ITK-SNAP or directly in Python.

Federated learning




Federated Server
Global Model
Community Hospital
Privacy Preserving
Local Model
Private Data
Research Medical Center
Privacy Preserving
Local Model
Private Data
Cancer Treatment Center
Privacy Preserving
Local Model
Private Data

12 YEARS OF HEALTHCARE

2007 — NVIDIA CUDA

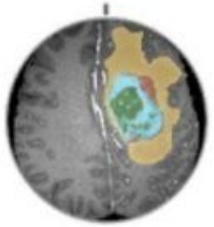
2009 — Siemens Ultrasound Amnioscopic Rendering

2011 — GE Revolution CT Iterative Recon & Dose Reductions

2012 — Thermo Fisher Ion Desktop Sequencing

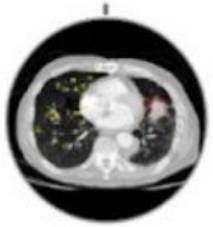
2013 — Molecular Dynamics HIV-1 Capsid, NAMD

2014 — Brain Tumor Segmentation with Deep learning

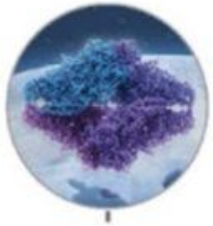


2015 — GE Ultrasound 4D Blood Flow

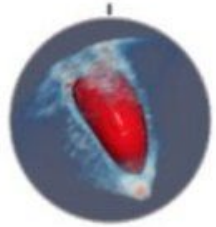
2016 — Lung Cancer Detection w/ Deep Learning



2017 — Nobel Prize Cryo-EM 3D Reconstruction



2018 — NVIDIA Project Clara



2018 — 1500 Deep Learning classes
270 Scientific Sessions
80 AI Startups

RSNA 2018
TOMORROW'S RADIOLOGY TODAY
NOVEMBER 25-30

2019 — Clara Federated Learning
Clara AGX Developer kit
Clara AI Development

Development - Data set

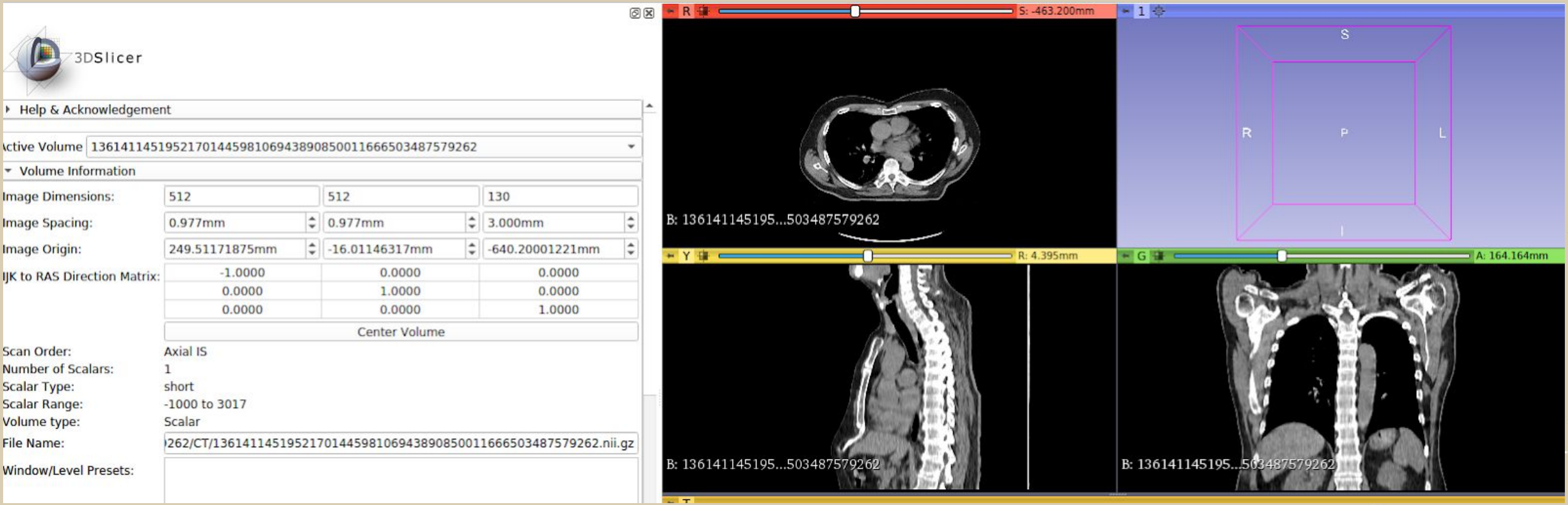
Lung CT Segmentation Challenge 2017

Each training dataset includes a set of DICOM CT image files and one DICOM RTSTRUCT file. Each training dataset is labeled as LCTSC-Train-Sx-yyy, with Sx (x=1,2,3) identifying the institution and yyy identifying the dataset ID in one institution. You may take advantage of this information to optimize your algorithm for testing data acquired from different institutions.

Detailed Description

Collection Statistics	Updated 2019/05/08
Modalities	CT, RT, RTSTRUCT
Number of Participants	60
Number of Studies	60
Number of Series	120
Number of Images	9,593
Image Size (GB)	4.8

Development - Data set



Sample form data Lung CT Segmentation Challenge 2017

Development - data preprocessing

- Convert Dicom to NIFTI using software, packages/modules like **dicom2nifti** (python) or **data converter** for Clara Train SDK.

In Data Converter input file can be format, can be .dcm, .nii, .nii.gz, .mha, .mhd.

nvmidl-dataconvert -d your/data/directory -r 1 -s .dcm -e .nii.gz -o your/output/directory

- All input images and labels must be in NIfTI format. Each input image and its corresponding label mask must have the same image dimension. To visualize or save NIfTI images, you can use free viewers such as ITK-SNAP or MITK.

Development - data preprocessing

- If the data format is DICOM or the resolution is not isotropic, one can use the provided data converter tool to convert the data to isotropic NIfTI format. Furthermore, many pre-trained models were trained on 1x1x1mm resolution images, and to use those pre-trained models as a starting point, convert the data to 1x1x1mm NIfTI format

Development - preprocessing

- Dataset (in this case small dataset) - 10 files (each with a lots of sample) from **Lung CT Segmentation Challenge 2017 (circa 300Mb)**
- Small dataset -> 70% train, 30% tests
- DICOM-structure for each catalog (filename 1-1.dcm)

more about converting:

Development - preprocessing

- The layout of data files can be arbitrary, but the JSON file describing the data list must contain the relative paths to all data files. For example, the datalist.json:

```
{
  "train": [
    {
      "image" : "train/im1.nii.gz",
      "label" : "train/lb1.nii.gz"
    },
    {
      "image" : "train/im2.nii.gz",
      "label" : "train/lb2.nii.gz"
    },
    {
      "image" : "train/im3.nii.gz",
      "label" : "train/lb3.nii.gz"
    },
    {
      "image" : "train/im4.nii.gz",
      "label" : "train/lb4.nii.gz"
    },
    {
      "image" : "train/im5.nii.gz",
      "label" : "train/lb5.nii.gz"
    },
    {
      "image" : "train/im6.nii.gz",
      "label" : "train/lb6.nii.gz"
    },
    {
      "image" : "train/im7.nii.gz",
      "label" : "train/lb7.nii.gz"
    },
  ],
  "test": [
    {
      "image" : "val/im8.nii.gz",
      "label" : "val/lb8.nii.gz"
    },
    {
      "image" : "val/im9.nii.gz",
      "label" : "val/lb9.nii.gz"
    },
    {
      "image" : "val/im10.nii.gz",
      "label" : "val/lb10.nii.gz"
    },
  ]
}
```

Development - perspective

- running learning process (with preprocessed data, bigger than small dataset)
- prepare separated environment (like docker container) to run federated learning
- doing benchmarks with other solutions for this same data

Thank you!

Presenter: Maciej Krzywda
Project supervisor: dr hab. inż Tomasz Szumlak