

Preparing Ultrasound Imaging Data for Artificial Intelligence Tasks: Anonymization, Cropping, and Tagging

D. Pechlivanis, S. Didaskalou, E. Kaldoudi, and G. Drosatos



presented by

Eleni Kaldoudi

Athena Research Center, Greece
Democritus University of Thrace, Greece
ThrombUS+ Horizon Project

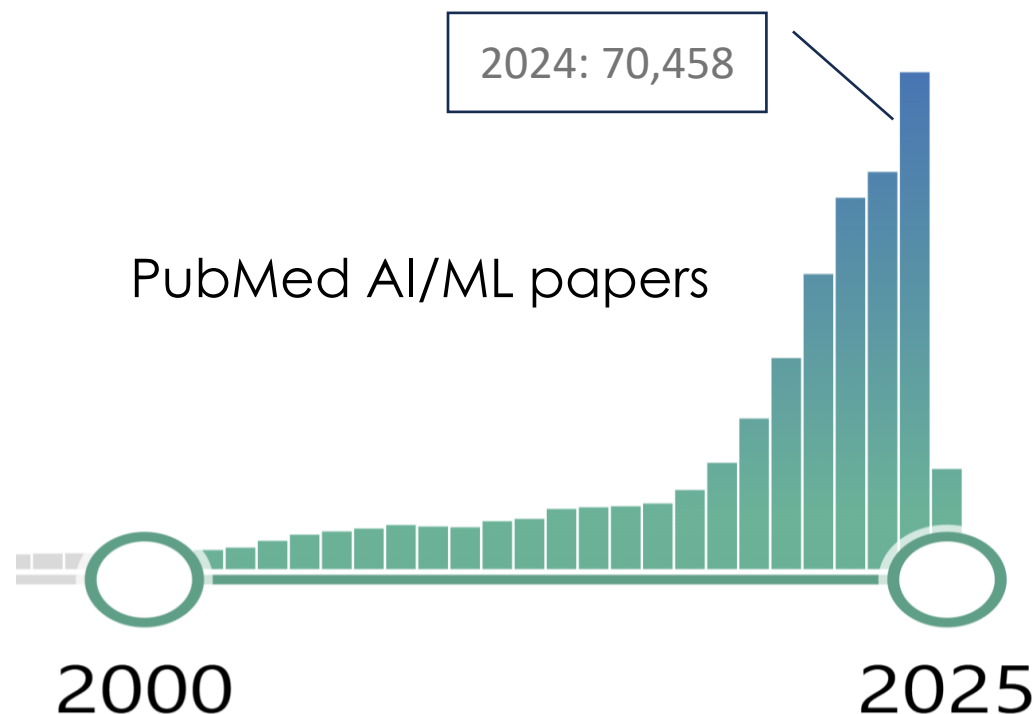




more than **350,000 papers** related to **AI/ML** in biomedical applications since 2000

11% of AI/ML papers in biomedicine relate to **diagnostic imaging**

20% of AI/ML papers in diagnostic imaging relate to **ultrasonography**



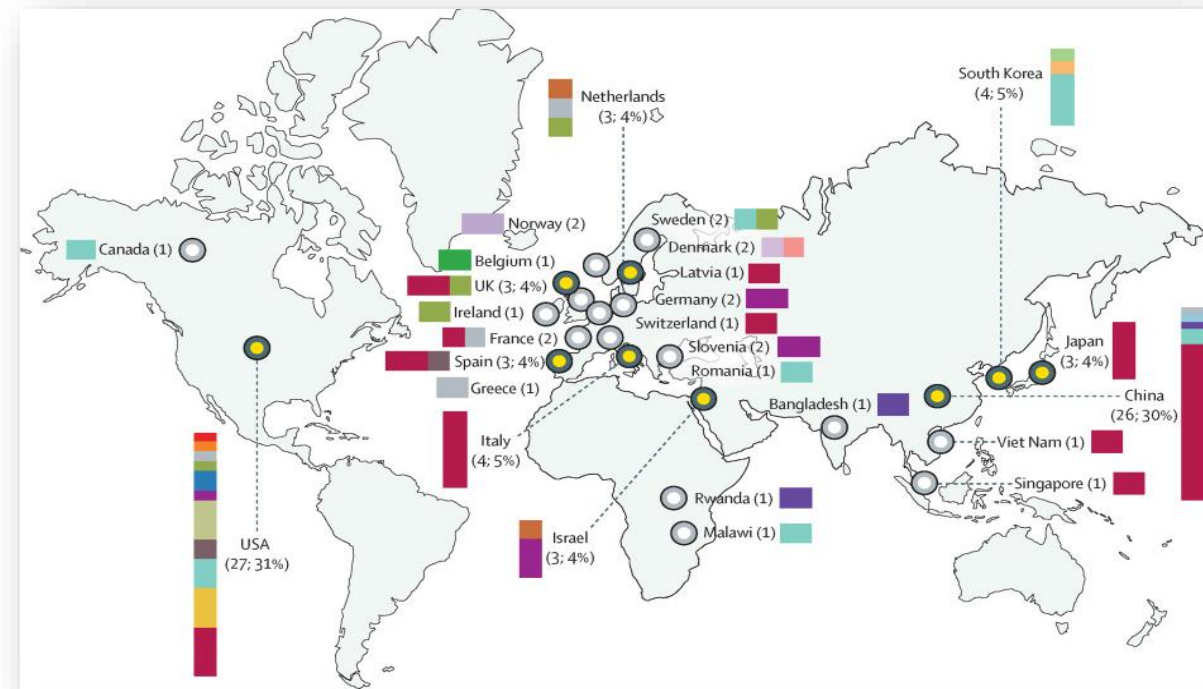
- 1) "artificial intelligence"[MeSH Terms] OR ("artificial"[All Fields] AND "intelligence"[All Fields]) OR "artificial intelligence"[All Fields] OR ("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields])
- 2) "diagnostic imaging"[MeSH Subheading] OR ("diagnostic"[All Fields] AND "imaging"[All Fields]) OR "diagnostic imaging"[All Fields]
- 3) "ultrasound"[All Fields] OR "ultrasonography"[MeSH Terms] OR "ultrasonography"[All Fields] OR "ultrasonics"[MeSH Terms] OR "ultrasonics"[All Fields] OR "ultrasounds"[All Fields] OR "ultrasound s"[All Fields]

#1 | #1 AND #2 | #1 AND #3

a 2024 scoping review identified
86 randomized controlled trials
that evaluate artificial intelligence
applications in clinical practice

finds that the majority

- involve AI applications
in diagnostic imaging
- report statistically significant
healthcare improvement
when AI is employed



THE LANCET
Digital Health

Log in

REVIEW · Volume 6, Issue 5, E367-E373, May 2024

Open Access

Download Full Issue

Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review

Ryan Han, MS^{a,b,c} · Julián N Acosta, MD^{d,e} · Zahra Shakeri, PhD^f ·

Prof John P A Ioannidis, MD DSc^{g,h} · Prof Eric J Topol, MD^{i,*} · Pranav Rajpurkar, PhD^{a,*}

Affiliations & Notes

Article Info

most **important factor** for successful application of AI/ML in diagnostic imaging is the **training data set**

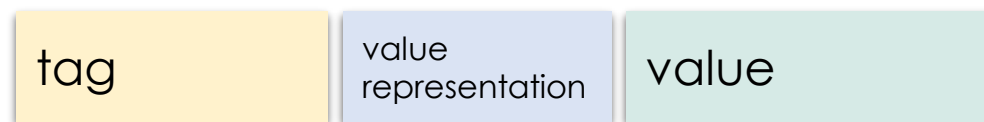
facts:

- universal use of **DICOM** standard
- rigorous **privacy** regulations (GDPR, HIPAA)

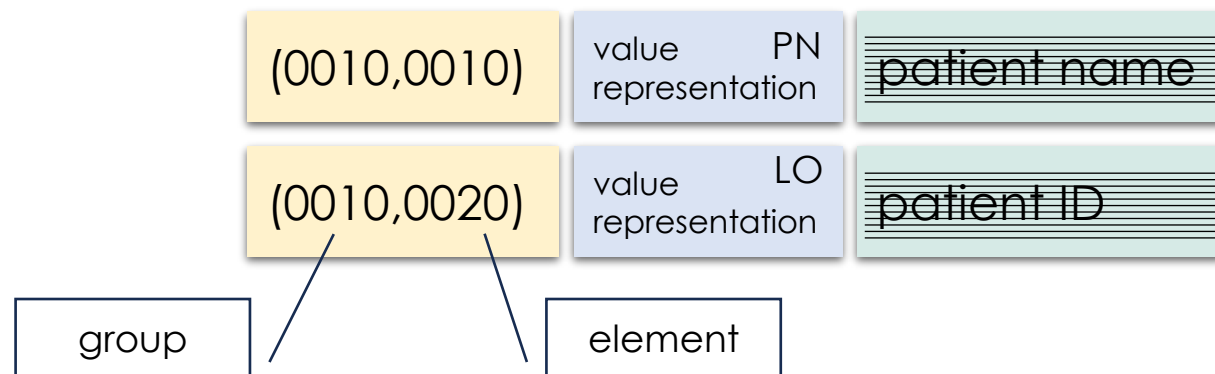
key challenges:

- ensure patient **anonymity** in DICOM data
- allow for **tagging, labeling**

DICOM metadata element



examples



many free solutions are available
for DICOM data de-identification

however,

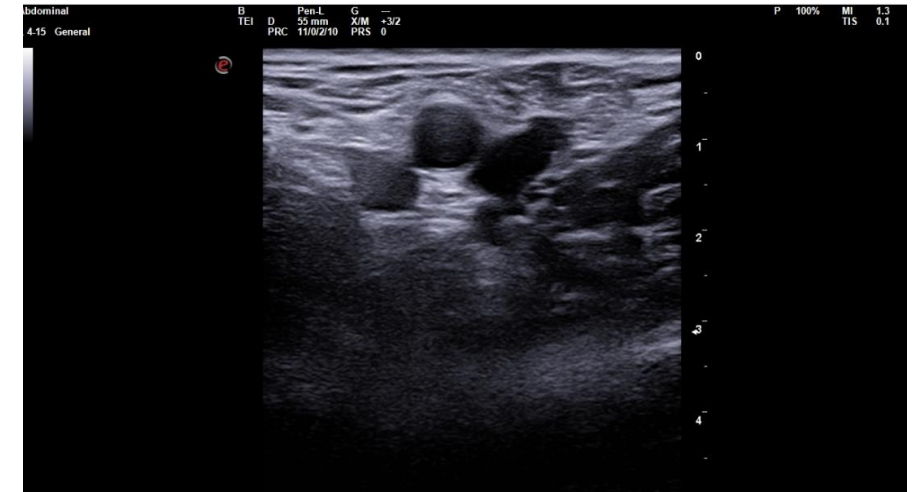
de-identification is not always efficient, since:

- DICOM implementations vary
by vendor and imaging modality
(e.g. vendors may introduce their own tags)
- very often, personal data is also
burned in the image

► Eur Radiol. 2015 Jun 3;25(12):3685–3695. doi: [10.1007/s00330-015-3794-0](https://doi.org/10.1007/s00330-015-3794-0)

Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy

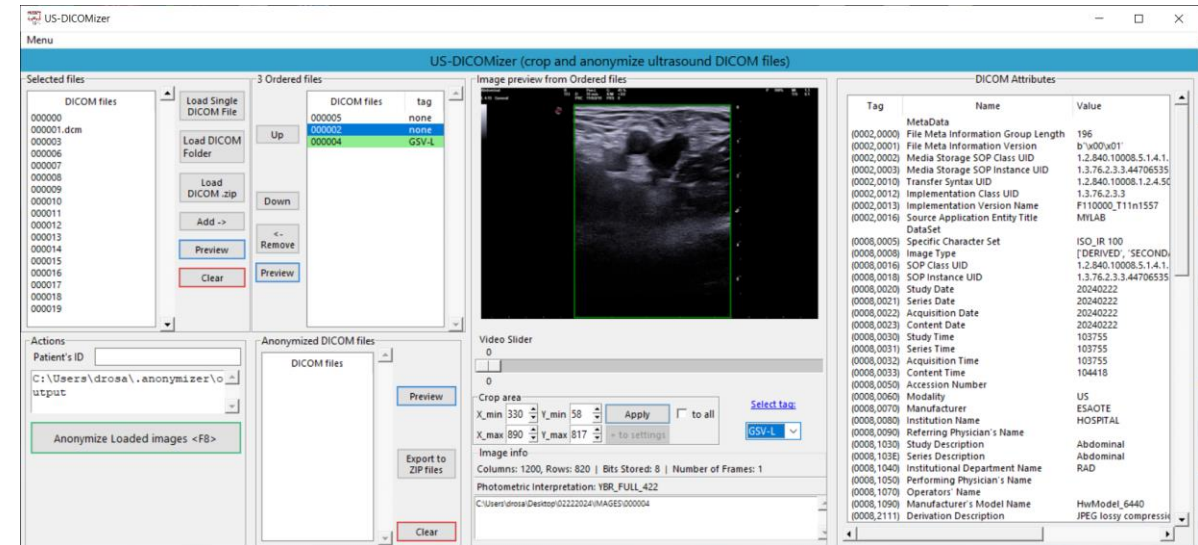
[K Y E Aryanto](#)^{1,✉}, [M Oudkerk](#)¹, [P M A van Ooijen](#)¹



US-DICOMizer

an integrated approach to ultrasound data preparation for AI/ML training sets

- DICOM metadata
de-identification:
customizable, can be inspected
- burned-in image cropping:
customizable
per imaging device
- generation of image tags related to training set
- batch image processing: customizable workflows



Horizon Europe Project

Wearable Continuous Point-of-Care Monitoring, Risk Estimation and Prevention for Deep Vein Thrombosis

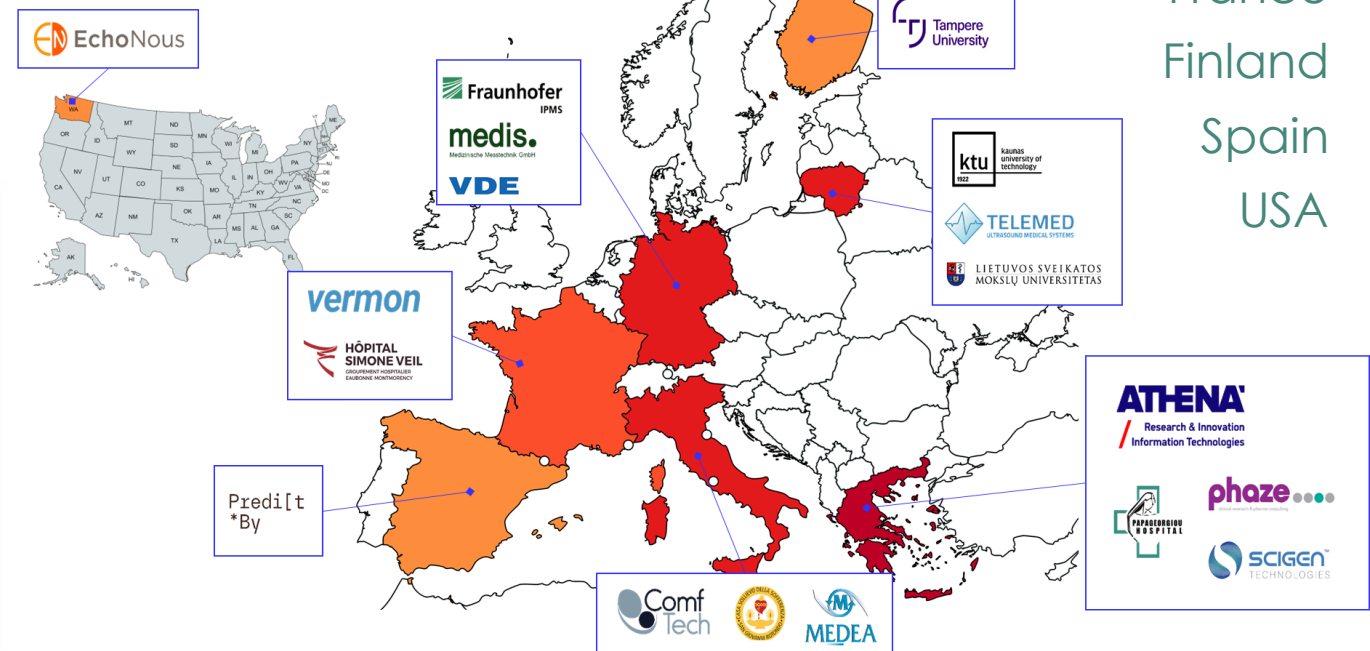
9.5 M € | 2024-2027

Coordinator: E. Kaldoudi, ATHENA RC, Greece

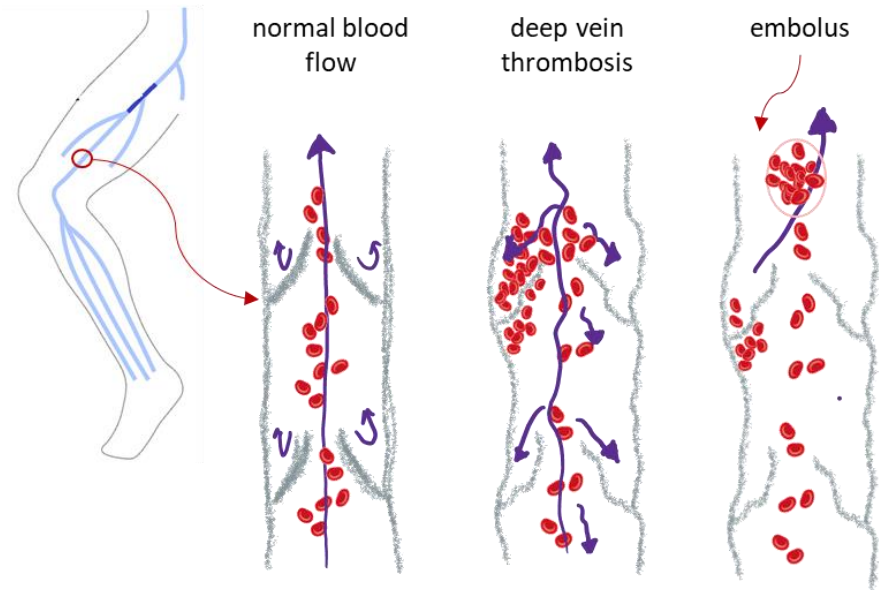
ThrombUS⁺

18 partners from 8 countries

Greece
Lithuania
Germany
Italy
France
Finland
Spain
USA



Deep Vein Thrombosis (DVT) is the clotting of blood in a deep vein of the pelvis or an extremity (usually calf or thigh)

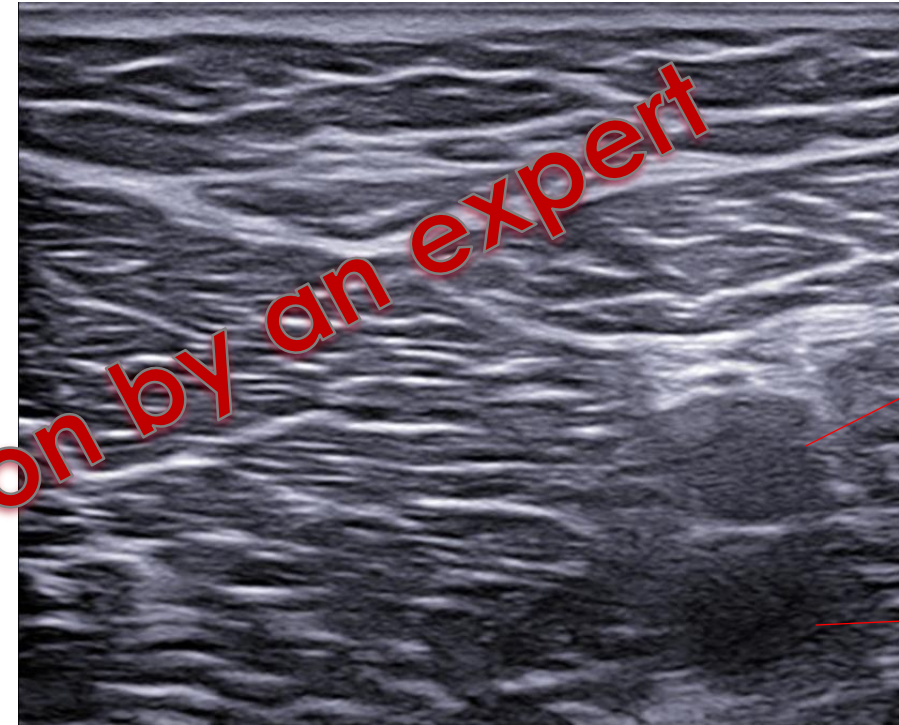
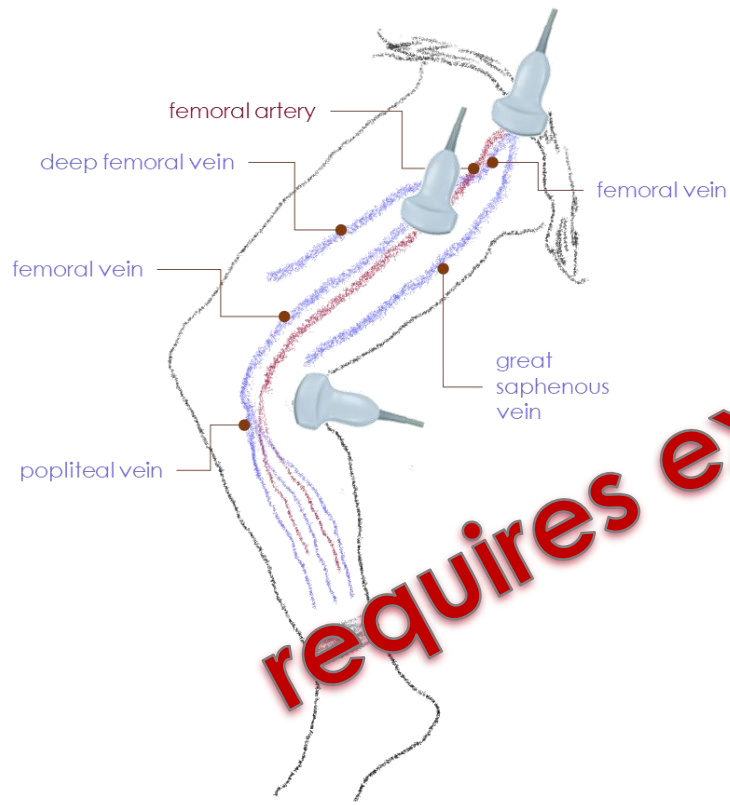


- ⇒ in immobile patients
- ⇒ after major surgery
- ⇒ in cancer patient
- ⇒ in pregnancy
- ⇒ during flights
- ⇒

- affects more than
1,000,000 Americans per year
700,000 Europeans per year
- 1/2 of people with DVT experience a sudden pulmonary embolism
- about 1/4 of those who have a pulmonary embolism die from it
- annual health expenditure related to DVT is €8.5 billion in EU

Olaf M et al. Deep Venous Thrombosis. Emerg Med Clin North Am. 2017
Cohen AT et al. Thromb Haemost. 2007 Oct;98(4):756-64
Barco S et al. Thromb Haemost. 2016 Apr;115(4):800-8

compression ultrasonography
is the method of choice for
DVT diagnosis



requires examination by an expert

when DVT is present,
vein does NOT fully collapse

>90% sensitive

>95% specific

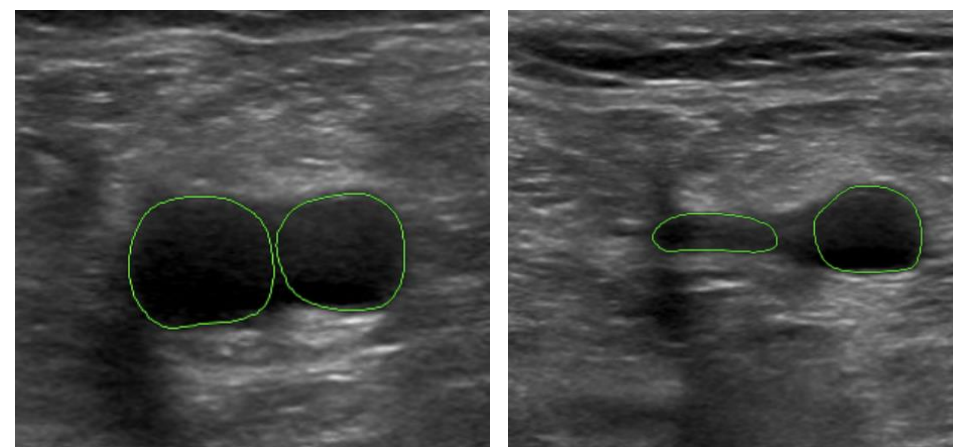


the 1st wearable,
autonomous
ultrasound

machine learning to

ThrombUS⁺

- ⇒ evaluate **diagnostic quality** of the image
- ⇒ **guide** the wearable positioning and compression
- ⇒ identify and **segment** vessels
- ⇒ **evaluate** vein compressibility
- ⇒ **alert** on DVT suspicion and **explain**



ThrombUS+ raining data set

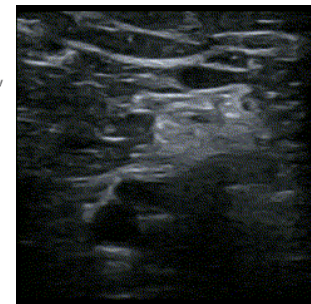
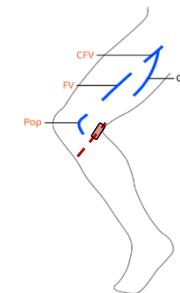
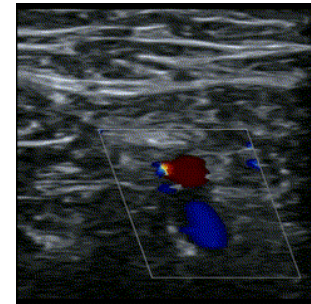
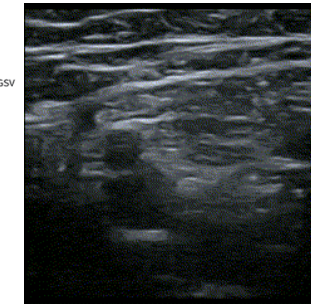
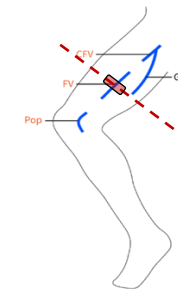
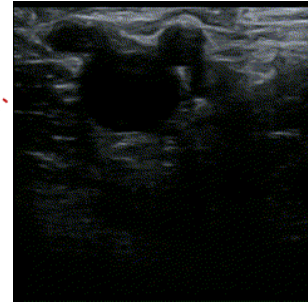
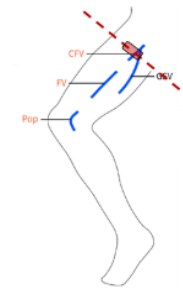
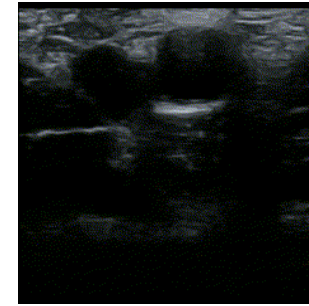
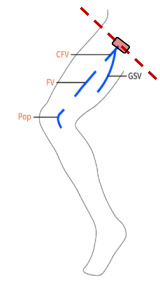
conventional ultrasound images and videos
from 5 hospitals in different countries

from >3,000 patients referred for DVT ultrasound

per patient:

- 4 imaging sites
- 5 compression ultrasound video clips
- additional images of no diagnostic quality

overall, 500-750 frames per patient X 3,000 patients



- **tags for de-identification**

- pre-define tags for de-identification per vendor or imaging device
- manually select tags for de-identification

- **cropping region**

- pre-define region per imaging device
- manually define region

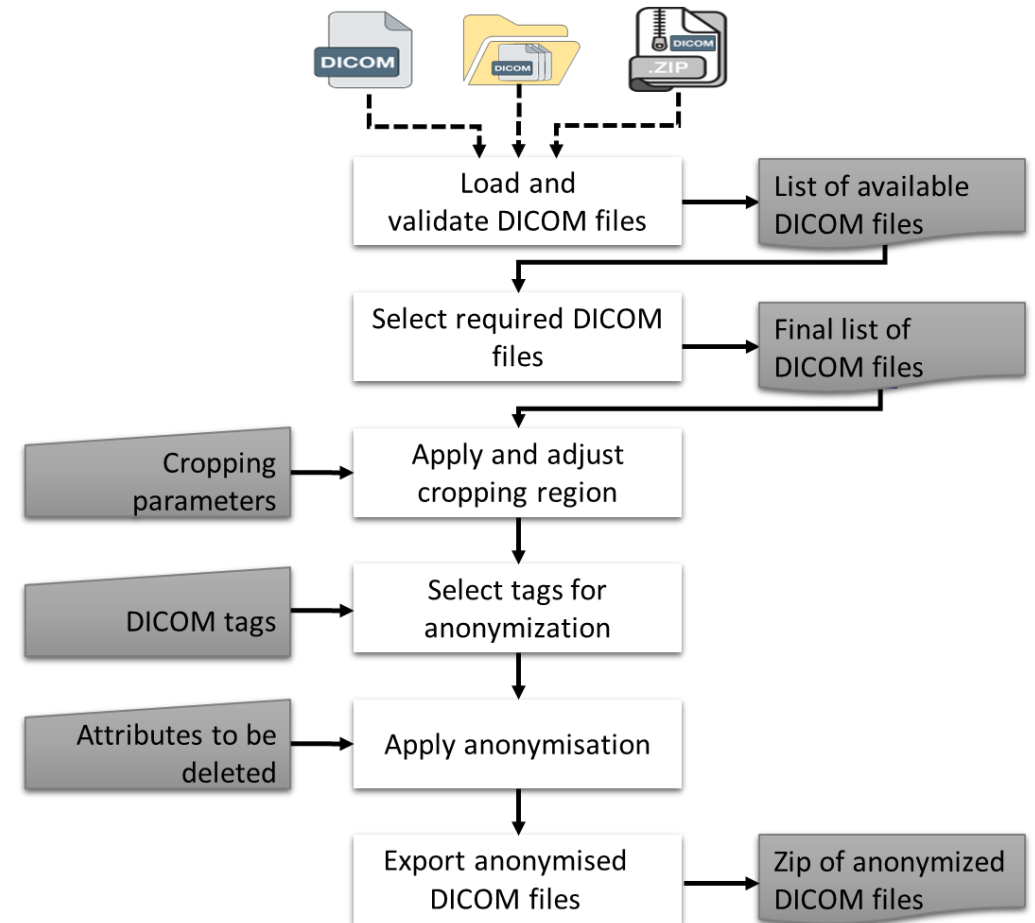
- **apply metadata tagging for training set**

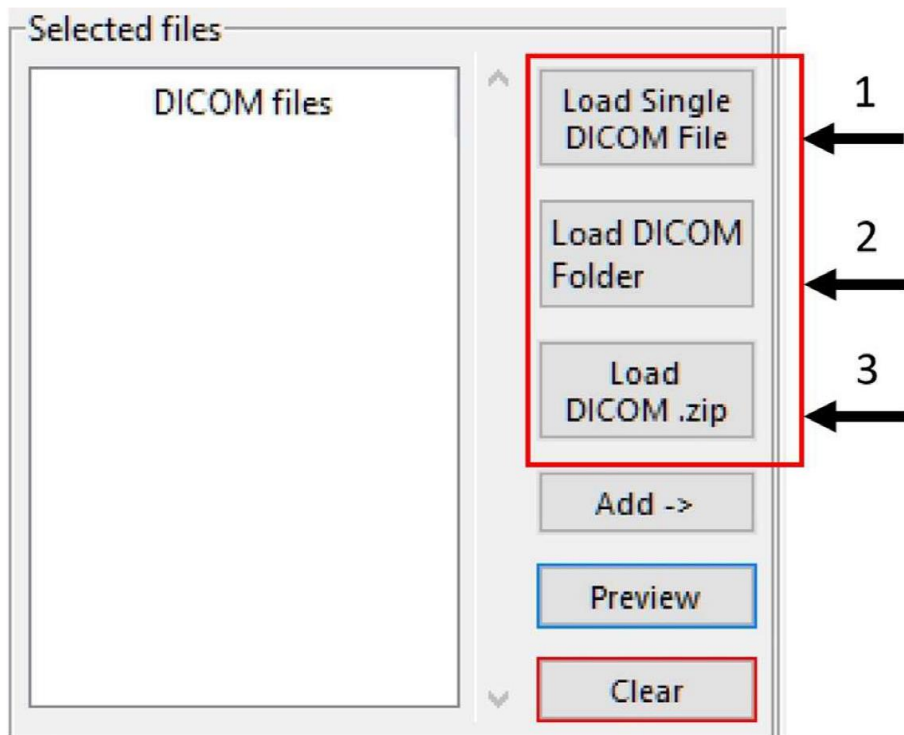
- add patient pseudo-id, anatomical site, diagnosis, etc.

- **preview original and de-identified data**

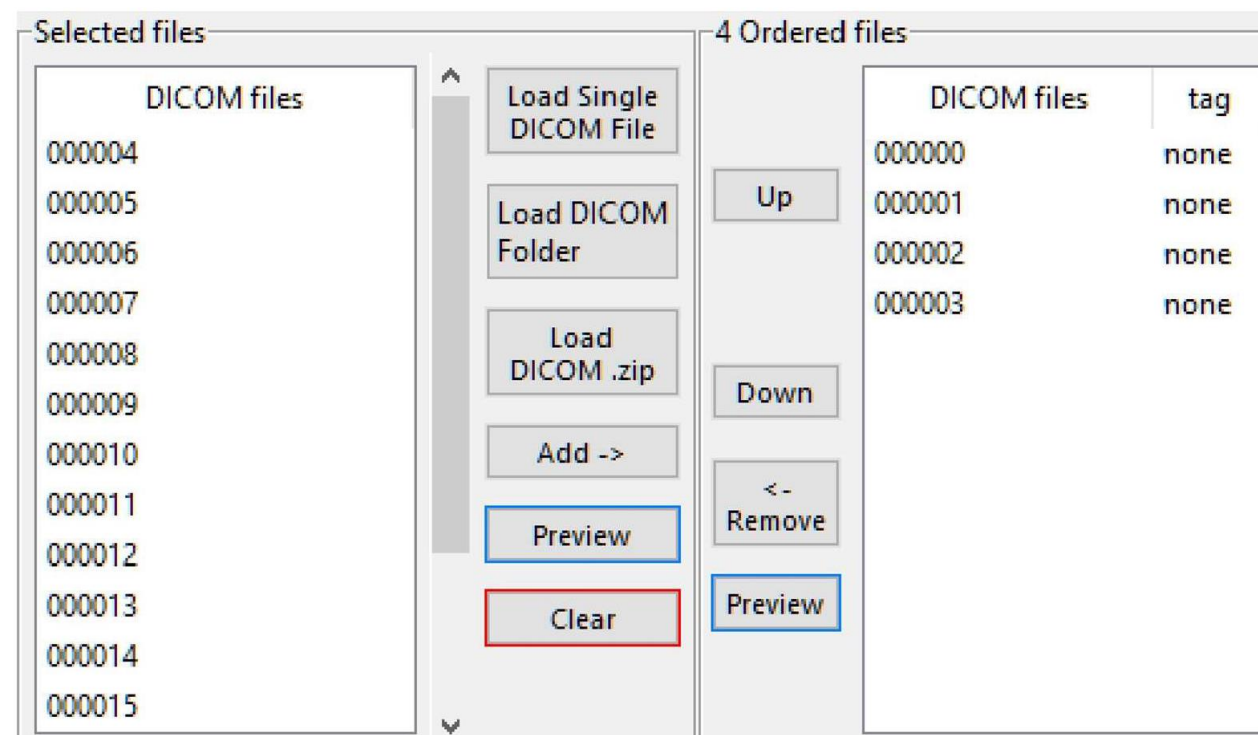
- image/video
- DICOM tags

US-DICOMizer workflow



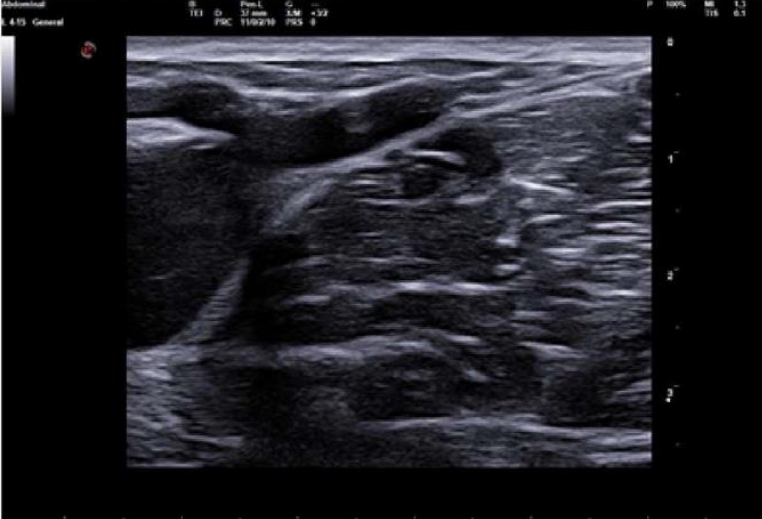


load DICOM files



select the appropriate DICOM files for anonymization

Image preview from Selected files



Video Slider
0

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105

Show image

Image info
Columns: 1200, Rows: 820 | Bits Stored: 8 | Number of Frames: 106

Photometric Interpretation: YBR_FULL_422

C:\Users\...est_images\02222024\IMAGES\000007

DICOM Attributes

Tag	Name	
(0002,0000)	File Meta Information Group Length	196
(0002,0001)	File Meta Information Version	b'\x00\x01'
(0002,0002)	Media Storage SOP Class UID	1.2.840.10008.5.
(0002,0003)	Media Storage SOP Instance UID	1.3.76.2.3.3.4470
(0002,0010)	Transfer Syntax UID	1.2.840.10008.1.
(0002,0012)	Implementation Class UID	1.3.76.2.3.3
(0002,0013)	Implementation Version Name	F110000_T11n15
(0002,0016)	Source Application Entity Title	MYLAB
(0008,0005)	Specific Character Set	ISO_IR 100
(0008,0008)	Image Type	['DERIVED', 'SEC
(0008,0016)	SOP Class UID	1.2.840.10008.5.
(0008,0018)	SOP Instance UID	1.3.76.2.3.3.4470
(0008,0020)	Study Date	240
(0008,0021)	Series Date	240
(0008,0022)	Acquisition Date	40
(0008,0023)	Content Date	40
(0008,0030)	Study Time	75
(0008,0031)	Series Time	175
(0008,0032)	Acquisition Time	175
(0008,0033)	Content Time	184
(0008,0050)	Accession Number	
(0008,0060)	Modality	US
(0008,0070)	Manufacturer	ESAOTE

Select tag:

none

none

CFVr-L

CFV-L

CFVr-R

CFV-R

GSVr-L

GSV-L

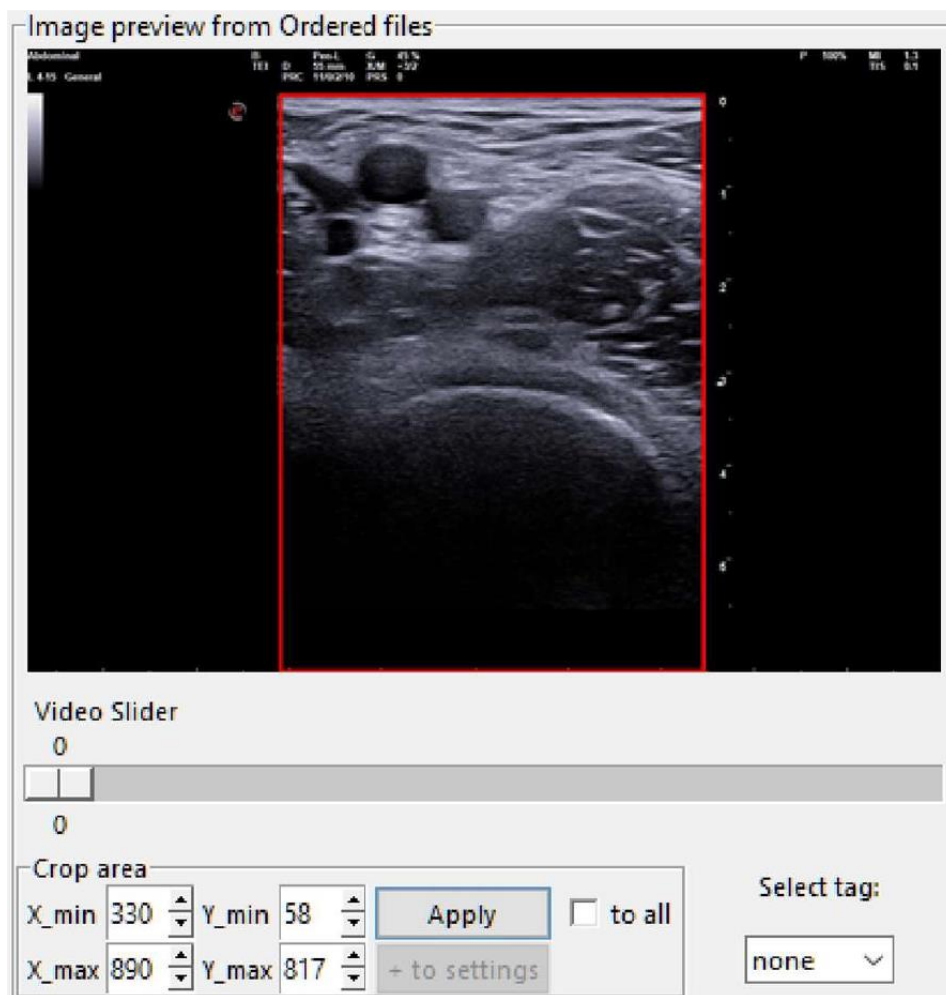
GSVr-R

GSV-R

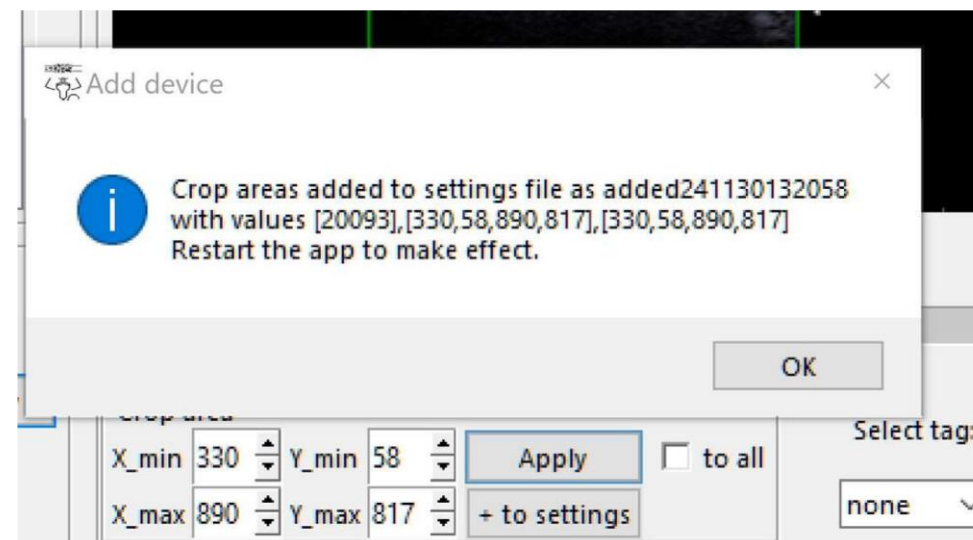
FVr-L

select
training set
predefined tags
for labelling
purposes

preview images and DICOM tags



define cropping region



save cropping region in settings

evaluation using DVT ultrasound datasets by 4 different vendors

- processing time per frame:
 - Monochrome2: 19-23 ms
 - YBR FULL 422: 25-30 ms
 - RGB: 49-61 ms
- full dataset preparation time:
3-6 minutes per ultrasound exam
- preliminary user feedback:
 - high usability and efficiency for medical imaging professionals
- extensive usability assessment
is currently underway in 5 hospitals
(via SUS: System Usability Scale questionnaire)

Media Type (resolution)	Photometric Interpretation	Cropping Area (pixels)	AVG Time per Frame (msec)
Ultrasound Image (1200x800) [100 runs]	Monochrome2	576x432	19
		768x576	21
		1024x768	23
	YBR FULL 422	576x432	25
		768x576	27
		1024x768	30
	RGB	576x432	49
		768x576	53
		1024x768	61
Ultrasound Multi-frame (1200x800) [10 runs]	Monochrome2	576x432	102
		768x576	105
		1024x768	108
	YBR FULL 422	576x432	117
		768x576	119
		1024x768	124
	RGB	576x432	241
		768x576	244
		1024x768	250

Windows 10 desk-top computer
Intel Core i7-7700HQ CPU at 2.8 GHz, 16 GB RAM

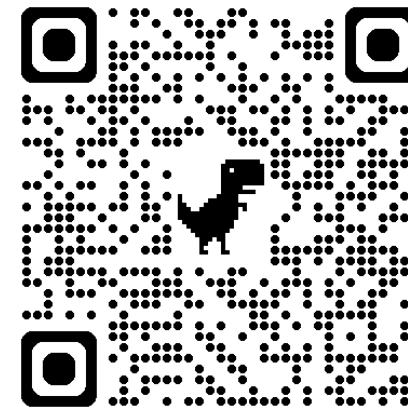
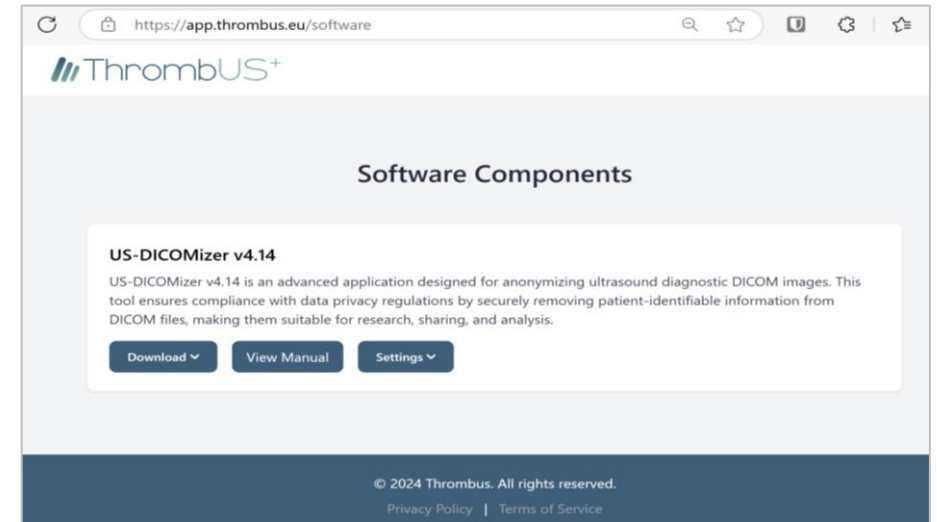
US-DICOMizer

open source, available under MIT License

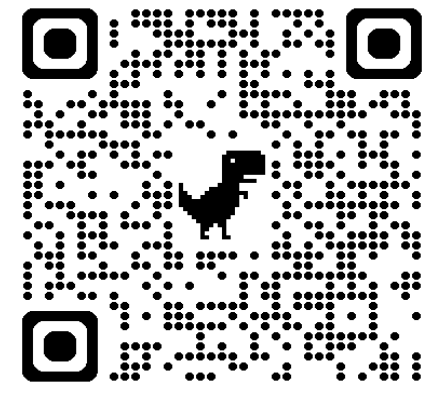
programming Language: Python

libraries used:

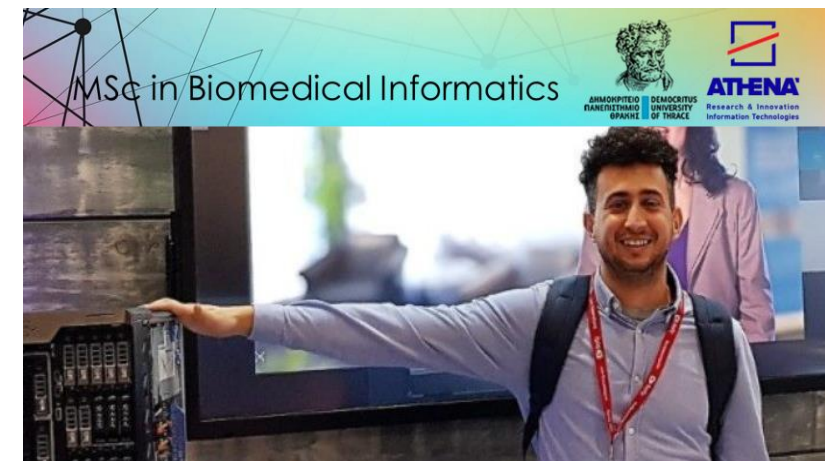
- Pydicom for handling DICOM files
- Tkinter for GUI development
- SimpleITK, NumPy, Pillow for image processing
- Matplotlib for visualization



app.thrombus.eu/software



 **GitHub**



 thrombus-project@athenarc.eu

 <https://thrombus.eu/>

 <https://www.linkedin.com/company/thrombus-eu-project/>

work conducted within the

ThrombUS+ Horizon project

in collaboration with the

MSc in Biomedical Informatics,
organized by
Democritus University of Thrace
and Athena Research Center,
Greece

MSc in Biomedical Informatics

 bmi@med.duth.gr

bmi@athenarc.gr

 <https://bmi.med.duth.gr/>



Co-funded by the
European Union



ΔΗΜΟΚΡΙΤΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΡΑΚΗΣ
DEMOCRITUS
UNIVERSITY
OF THRACE

