

This preprint has not undergone peer review or any post-submission improvements or corrections.

The Version of Record of this contribution is published in

Pechlivanis D, Didaskalou S, Kaldoudi E, Drosatos D. US-DICOMizer: An Open-Source Tool for Preparing Ultrasound DICOM Data for AI Applications. Biomedical Engineering Systems and Technologies, In: Fred, A., et al. Biomedical Engineering Systems and Technologies. BIOSTEC 2025. Communications in Computer and Information Science, vol 3115. Springer, Cham, 2016. doi: 10.1007/978-3-032-34459-5_24

and is available online at:

https://link.springer.com/chapter/10.1007/978-3-032-34459-5_24

US-DICOMizer: An Open-Source Tool for Preparing Ultrasound DICOM Data for AI Applications

Dimitrios Pechlivanis¹[0009–0001–3583–772X], Stylianos Didaskalou²[0000–0001–5932–9626], Eleni Kaldoudi^{1,2}[0000–0002–6054–4961], and George Drosatos²[0000–0002–8130–5775]

¹ School of Medicine, Democritus University of Thrace, 68100 Alexandroupoli, Greece
`{dimipech2,kaldoudi}@med.duth.gr`

² Institute for Language and Speech Processing, Athena Research Center, 67100 Xanthi, Greece
`{stelios.didaskalou,gdrosato}@athenarc.gr`

Abstract. Ultrasound imaging is a widely used diagnostic modality in diverse clinical contexts, but preparing ultrasound data for artificial intelligence (AI) applications presents notable challenges – including the need to ensure data privacy, extract diagnostically relevant regions, and associate contextual metadata. This paper presents US-DICOMizer, a standalone, open-source application designed to streamline the preparation of ultrasound DICOM files for AI workflows. The tool supports three core functions: (1) anonymisation, to remove sensitive metadata in compliance with privacy regulations; (2) cropping, to isolate relevant regions in images or multiframe videos; and (3) tagging, to enrich each file with metadata such as anatomical position and imaging purpose. Built with a user-friendly interface and a robust processing backend, US-DICOMizer incorporates semi-automated features, including a newly introduced autocropping function, to enhance workflow efficiency. Its effectiveness was evaluated using deep vein thrombosis (DVT) ultrasound datasets, showing fast processing times and favourable usability feedback from 15 healthcare professionals. These results demonstrate the tool’s suitability for scalable, standardised image preparation in AI development pipelines. The application offers a generalisable framework for ultrasound data preparation, while being adaptable to broader clinical use cases. Future work will focus on expanding support to other imaging modalities, integrating AI-assisted features, and conducting large-scale clinical validations.

Keywords: Ultrasound Imaging · DICOM · Anonymisation · Cropping · Tagging · Artificial Intelligence (AI).

1 Introduction

Machine learning (ML) is increasingly transforming medicine, particularly in medical imaging, where it supports automated diagnosis, optimised clinical workflows, and molecular structure analysis [11]. A fundamental requirement for

training reliable ML models in this domain is the availability of high-quality, labelled datasets. Preparing such datasets is complex due to strict requirements for patient privacy (anonymisation), the need to isolate diagnostically relevant image regions, and the association of contextual metadata such as anatomical site and image purpose.

Ultrasound imaging presents specific challenges for data preparation. It is widely used in clinical settings due to its safety, real-time imaging, and versatility. However, variability in operator technique and device configurations complicates standardisation. This is particularly true in conditions such as deep vein thrombosis (DVT) [27], where accurate and reproducible imaging is essential. As the primary diagnostic modality for DVT, ultrasound’s complexity and subjectivity make it a compelling use case for AI [2], which can assist in tasks like vessel segmentation and thrombus detection [19].

Medical imaging data is predominantly stored and exchanged in the DICOM (Digital Imaging and Communication in Medicine) format [14], which includes both metadata and pixel data. Preparing such data for AI applications involves anonymising sensitive metadata, addressing “burned-in” text within images (especially common in ultrasound), cropping to relevant regions, and enriching the data with task-specific annotations. Manual approaches to these tasks are time-consuming and prone to inconsistency, limiting scalability.

To address these issues, we introduce US-DICOMizer, an open-source tool designed to automate and streamline the preparation of ultrasound DICOM data for AI workflows. While suitable for general ultrasound use cases, the tool is evaluated using ultrasound exams from a DVT clinical study. US-DICOMizer integrates three core functionalities: (1) anonymisation of sensitive patient information; (2) cropping of relevant image regions; and (3) tagging with clinically meaningful metadata. These features support efficient and standardised preparation of datasets while ensuring regulatory compliance.

This work is part of the ThrombUS+ project, co-funded by the European Union (EU) under Grant Agreement No. 101137227 [15]. The ThrombUS+ initiative aims to develop an innovative, operator-independent wearable diagnostic device designed to facilitate the early detection and diagnosis of DVT using various diagnostic data modalities, including ultrasound imaging. As part of this initiative, the US-DICOMizer application addresses critical data preparation needs for AI-based workflows in DVT and other ultrasound imaging applications; a preliminary version of this tool was introduced in [23].

The remainder of this paper is organised as follows: Section 2 introduces the background concepts relevant to this work, including ultrasound imaging, the DICOM standard, and medical data anonymisation. Section 3 provides an overview of related work in DICOM data preparation and AI applications in ultrasound imaging. Section 4 details the methodology and design of the US-DICOMizer application, including its anonymisation, cropping, and tagging functionalities. Section 5 discusses the technical implementation details, highlighting the tools and frameworks employed. Section 6 presents results demonstrating the per-

formance and usability of the application. Finally, Section 7 summarises the contributions of this work and outlines future directions.

2 Background

To support the development and application of US-DICOMizer, this section outlines the foundational concepts that underpin its design. Artificial intelligence (AI) systems in medical imaging require structured, clean, and privacy-compliant data. We therefore provide an overview of ultrasound imaging as the target modality, the DICOM standard for storing and exchanging medical image data, and the principles of medical data anonymisation as mandated by international regulations.

2.1 Ultrasound Imaging

Ultrasound (US) [6] is a non-invasive medical imaging technique that uses high-frequency sound waves to visualize internal body structures. These waves are emitted by a transducer and transmitted into the body through a water-based gel to improve conductivity. Reflected echoes are captured and converted into images, offering real-time insights into anatomical and functional features without the use of ionizing radiation, making ultrasound a safe modality for a broad range of clinical applications, including obstetrics, cardiology, and vascular imaging [4].

Ultrasound imaging encompasses several display modes [7]:

- **A-mode (Amplitude Mode):** The earliest ultrasound technique, now largely obsolete in clinical practice.
- **B-mode (Brightness Mode):** The most widely used mode, generating high-resolution two-dimensional grayscale images.
- **M-mode (Motion Mode):** Records motion over time, particularly useful in cardiac imaging.
- **Doppler Modes:** Measure and visualize blood flow using the Doppler effect. Color Doppler overlays flow information on top of B-mode images, aiding in vascular assessments.

These characteristics make ultrasound particularly suited for dynamic and bedside evaluations, but also introduce unique challenges in data standardization and preprocessing for AI applications.

2.2 DICOM Standard

The Digital Imaging and Communications in Medicine (DICOM) standard [14] defines the format and exchange protocol for handling, storing, and transmitting medical images and related metadata. A DICOM file consists of a structured hierarchy that includes a 128-byte preamble, the “DICM” prefix, followed by

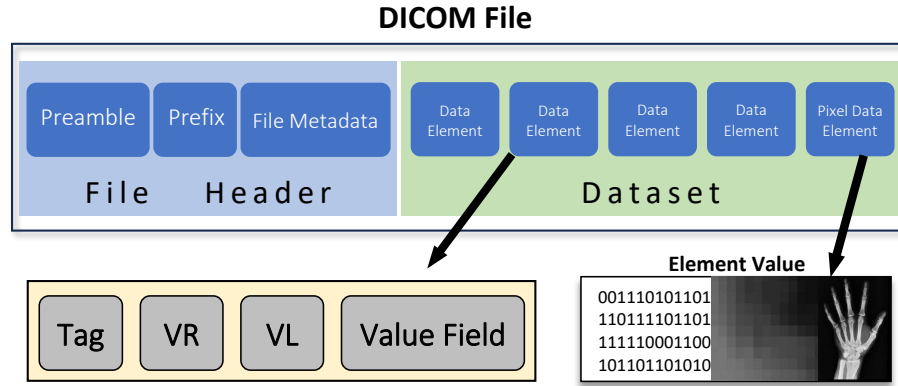


Fig. 1: Schematic representation of the DICOM file format.

File Metadata and the Dataset, which itself contains both headers (tags) and pixel data (see Figure 1).

This standardized architecture allows interoperability across imaging devices, PACS systems, and software applications. However, it also introduces complexities for anonymization, as patient information may reside in both standard and vendor-specific private tags, as well as be embedded (burned-in) within the image itself. For ultrasound data, this challenge is compounded by modality-specific variations in acquisition protocols and DICOM encoding, requiring tailored pre-processing pipelines for anonymization, cropping, and metadata tagging.

2.3 Data Anonymization

Medical imaging inherently contains sensitive personal health information (PHI) embedded in both metadata and image content. The protection of such data is governed by international regulations such as the General Data Protection Regulation (GDPR) [10] in the EU and the Health Insurance Portability and Accountability Act (HIPAA) [22] in the USA. Both frameworks mandate strict compliance regarding data privacy and impose penalties for violations.

Anonymization plays a pivotal role in enabling the ethical and legal reuse of medical data for research and clinical applications. In the context of biomedical research and healthcare analytics, anonymization ensures the confidentiality of patient data while supporting secondary uses such as AI training, epidemiological studies, and quality improvement initiatives. By removing or masking identifiers, anonymization mitigates the risk of re-identification and facilitates data sharing under data governance protocols aligned with GDPR and HIPAA standards.

3 Related Work

The preparation of ultrasound imaging data for artificial intelligence (AI) tasks requires specialised tools to address the unique challenges of this modality. Key

preprocessing steps include anonymisation, cropping, and tagging, which are essential to ensure patient privacy, enhance data utility, and support effective data sharing. While several academic tools have been proposed—mostly focusing on general medical imaging or specific modalities like neuroimaging—they often fail to meet the specific needs of ultrasound workflows, particularly for tasks such as vessel segmentation or thrombus detection. In parallel, a number of commercial and open-source solutions provide partial support for DICOM data anonymisation. However, these typically focus on metadata and lack capabilities for cropping, tagging, or removing burned-in personal data embedded in the image pixels. This section reviews both academic and practical tools relevant to anonymisation, cropping, and tagging in medical imaging, highlighting their applicability and limitations in the context of ultrasound data preparation.

3.1 Anonymisation Tools for Medical Imaging

Numerous anonymisation tools have been developed to safeguard patient privacy while ensuring compliance with DICOM (Digital Imaging and Communications in Medicine) standards. Monteiro et al. [21] introduced a de-identification pipeline leveraging convolutional neural networks (CNNs) for character recognition, achieving an anonymisation rate of 89.2%. Similarly, Rodríguez González et al. [24] developed an open-source toolkit for DICOM data de-identification tailored to multicenter trials, offering customisation based on specific privacy needs. However, these tools are primarily designed for neuroimaging and lacked features specific to ultrasound workflows. Haselgrove et al. [13] presented a system for anonymising DICOM data in neuroimaging research, ensuring compliance with quality control standards but leaving the unique needs of ultrasound imaging unaddressed.

In a comprehensive evaluation, Aryanto et al. [1] reviewed the functionality and safety of various free DICOM de-identification tools used in clinical research. The authors identified significant shortcomings, including inconsistent handling of private tags introduced by different vendors and inadequate detection of burned-in patient information. These limitations are particularly critical in ultrasound imaging, where personal data is often embedded directly into the image pixels. Such issues limit the effectiveness of existing de-identification solutions in real-world clinical workflows and highlight the need for more adaptable and domain-aware tools.

3.2 Cropping and Tagging in Ultrasound Workflows

Cropping and tagging techniques play a crucial role in isolating diagnostically relevant regions, anonymising burned-in annotations, and associating metadata with medical images. Interactive anonymisation tools, such as the Cornell Anonymisation Toolkit [28], allow users to manually crop and anonymise datasets but do not offer automation or optimisation for ultrasound imaging. Tools like DICOM Anonymiser [8] and dicom-anon [26] support tagging within the framework

of DICOM standards [12], but they are designed as general-purpose anonymisation utilities and do not offer features tailored to ultrasound workflows, such as cropping or removal of burned-in annotations. More recently, image cropping techniques based on grid-anchor formulations have emerged in general computer vision (e.g., CNN-based methods), but these approaches are designed for aesthetic image processing rather than clinical applications [29].

3.3 Commercial and Open Source Solutions

Several commercial and open-source tools offer anonymisation features, but none fully address the requirements of ultrasound workflows:

- **Rubo DICOM Anonymizer** provides metadata anonymisation, UID tracking, and pseudonymisation, but lacks image-level (burn-in text) support [25].
- **doRadiology DICOM Anonymizer** is a free macOS app with drag-and-drop input and re-identification features, but no burn-in removal [8].
- **MicroDicom DICOM Viewer** is a free Windows DICOM viewer with basic anonymisation options but no pixel-level processing [20].
- **Kitware DICOMAnonymizer** is an open-source Python tool that allows flexible header anonymisation but cannot detect image-embedded identifiers [16].
- **CARE DICOM Anonymizer** allows manual burn-in removal by drawing over image regions, but it is not scalable for large datasets [5].

A comparative summary of these tools is provided in Table 1, highlighting key platform characteristics, availability, and their ability to handle pixel-level anonymisation and ultrasound-specific workflows.

3.4 Summary of Limitations and Motivation for US-DICOMizer

Despite various advancements, significant gaps remain in addressing the specific requirements of ultrasound imaging. Existing tools are often modality-specific and fail to provide a comprehensive workflow that integrates anonymisation, cropping, and tagging in a streamlined and user-friendly manner. Moreover, inconsistent standards across imaging communities and lack of support for removing burned-in patient data pose challenges to the development of unified and scalable data preparation pipelines. While many tools offer basic metadata anonymisation, none combine both header and pixel anonymisation, structured tagging, and cropping with automation tailored to the characteristics of ultrasound imaging. US-DICOMizer aims to fill this gap by introducing a standalone, cross-platform tool that unifies all essential preprocessing steps. The tool is designed to be broadly applicable across ultrasound imaging use cases and is evaluated using deep vein thrombosis (DVT) examinations as a representative scenario to demonstrate its functionality and practical value in both research and clinical settings.

Table 1: Comparison of existing DICOM anonymisation tools.

Tool	Platform	Free/Open-Source	Burn-in Text Removal	Ultrasound-Specific
Rubo DICOM Anonymizer	Windows	Limited Free	No	No
doRadiology DICOM Anonymizer	macOS	Free	No	No
MicroDicom DICOM Viewer	Windows	Free	No	No
Kitware DICOMAnonymizer	Cross-platform (Python)	Open-Source	No	No
CARE DICOM Anonymizer	Windows & macOS	Free	Yes (manual)	No
US-DICOMizer (this work)	Cross-platform (Python)	Open-Source	Yes (auto & manual)	Yes

4 Methodology

The development of US-DICOMizer is guided by the specific needs of preparing ultrasound imaging data for artificial intelligence (AI) applications, incorporating insights from the ThrombUS+ project [15]. As part of this initiative, Clinical Study A [9], described in detail in the data collection manual available at <https://app.thrombus.eu/studies/a/>, serves as the primary context for validation. The methodology combines established imaging protocols with targeted tool development to support key preprocessing tasks—such as anonymisation, cropping, and tagging—without covering subsequent annotation or labelling, which are addressed in separate workflows.

4.1 Requirements and Design Principles

Clinical Study A specifies a structured procedure for acquiring ultrasound data, focusing on precise imaging of predefined anatomical sites, robust metadata tagging, and full privacy compliance. US-DICOMizer adopts a user-centred design to reduce manual effort and ensure consistent, high-quality data preparation. The application is designed based on the following principles:

1. **Data Privacy Compliance:** Full adherence to anonymisation protocols in line with HIPAA and GDPR standards.
2. **Standardisation and Consistency:** Integration of metadata tagging aligned with DICOM standards to support AI model training and evaluation.
3. **Workflow Efficiency:** Minimised manual intervention through user-friendly, interactive features.

4. **Automation and Scalability:** Automated execution of repetitive tasks (e.g., cropping, anonymisation) to support large-scale datasets.
5. **Clinical Relevance:** Support for metadata structures corresponding to anatomical sites and diagnostic contexts defined in Clinical Study A.

4.2 Data Collection Context

In Clinical Study A, clinicians capture ultrasound images and videos using a standardised protocol focused on assessing venous structures in the lower extremities. The dataset includes:

- *Compression Ultrasound:* Frames and videos assessing vein compressibility at specific anatomical landmarks (e.g., common femoral vein, popliteal vein).
- *Color Doppler Ultrasound:* Supplementary views that visualise blood flow in selected regions.
- *Optional Imaging:* Pathological views, such as thrombus visualisation, when clinically indicated.

US-DICOMizer supports these acquisition scenarios by providing a consistent pipeline for file import, cropping, tagging, anonymisation, and export, preserving the diagnostic and research utility of the data.

4.3 Workflow

US-DICOMizer follows the structured workflow shown in Figure 2, encompassing the following steps:

1. **File Loading and Validation:** Users can load individual DICOM files, folders, or compressed archives. The tool performs file validation and displays metadata and preview images. Multiframe DICOMs include a video slider for interactive review.
2. **Cropping and Tagging:**
 - **Cropping:** Regions of interest can be defined manually via the GUI, applied from configuration presets, or generated automatically using the autocropping feature, which identifies content-rich regions through image processing.
 - **Tagging:** Each file is annotated with structured metadata, including anatomical position (e.g., proximal/distal), laterality (left/right), and diagnostic purpose.
3. **Anonymisation:** Patient identifiers are removed or replaced based on user-defined settings. Unique identifiers (UIDs) for studies, series, and instances are regenerated to maintain DICOM compliance.
4. **Export:** All prepared files are exported as a ZIP archive with a consistent internal structure, ready for downstream AI processing.

By aligning with the Clinical Study A protocol, US-DICOMizer bridges the gap between clinical image acquisition and AI research workflows.

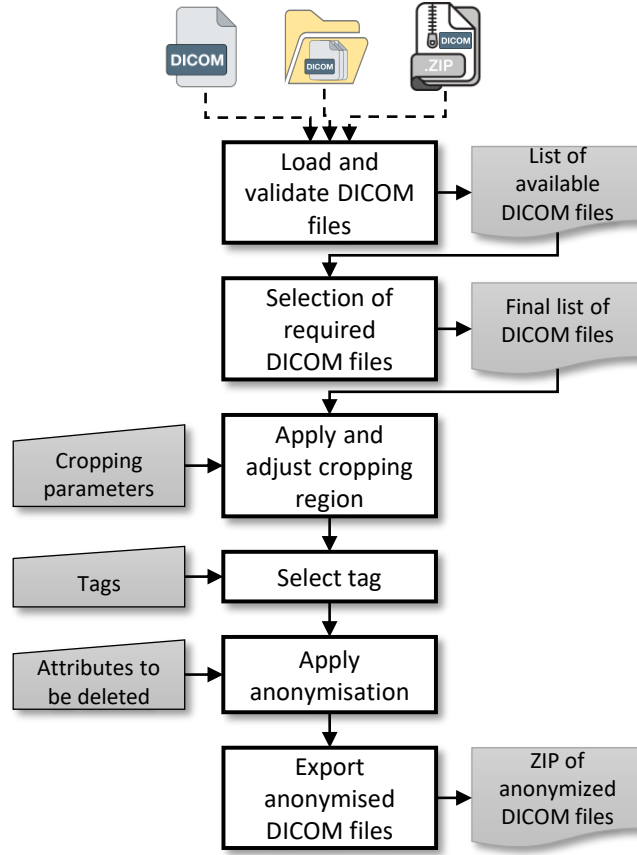


Fig. 2: US-DICOMizer workflow [23].

4.4 Usability Study Design

To evaluate user experience in real-world settings, a formal usability study is conducted with healthcare professionals from five clinical sites participating in the ThrombUS+ Study A [9]. The evaluation uses a survey-based methodology incorporating two validated instruments: the System Usability Scale (SUS) [18] and the Post-Study System Usability Questionnaire (PSSUQ) [17]. All participants had prior experience using US-DICOMizer in their clinical workflows.

The online questionnaire, administered via the EU Survey platform³, consisted of:

- 10 SUS items, rated on a 5-point Likert scale from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*);
- 16 PSSUQ items, rated on a 7-point Likert scale from 1 (*Strongly Agree*) to 7 (*Strongly Disagree*), where lower scores indicate better perceived usability.

³ <https://ec.europa.eu/eusurvey/runner/US-DICOMizer>

Demographic questions and open-ended prompts were also included to gather qualitative feedback. The study was approved by the ATHENA RC Ethics Committee and was conducted anonymously to encourage honest user responses.

5 Implementation Details

US-DICOMizer is an open-source application developed to support the preparation of ultrasound DICOM files for artificial intelligence (AI) tasks through an intuitive and semi-automated workflow. The code is publicly available on GitHub at <https://github.com/thrombusplus/US-DICOMizer>.

5.1 Development Environment and Frameworks

The application is implemented in Python and builds upon the following key libraries:

- **Tkinter**: For developing the graphical user interface (GUI).
- **Pydicom**: For reading, editing, and anonymising DICOM files.
- **SimpleITK**: For image processing and DICOM image visualisation.
- **Pillow (PIL)**: For basic image manipulation, such as cropping and saving.
- **NumPy**: For efficient numerical operations on image arrays.
- **Matplotlib**: For plotting and image visualisation during debugging.
- **Scikit-Image**: For advanced image analysis and enhancement.
- **OpenCV**: For implementing the autocropping function via edge detection and contour analysis.

5.2 Core Functionalities

Users can load individual DICOM files, entire folders, or ZIP archives. Loaded files appear in a tree view, allowing users to preview images and inspect DICOM metadata. Multiframe DICOMs are visualised using an interactive slider for frame navigation.

Cropping can be applied manually through the GUI or automatically based on predefined settings stored in a `settings.ini` file. These settings are applied uniformly across all frames in multiframe DICOMs. In addition, the application supports an autocropping feature that detects the primary region of interest using a processing pipeline comprising greyscale conversion, Gaussian blurring, Canny edge detection, and contour detection (via OpenCV). The detected contours are converted into bounding boxes used for automated cropping, with optional user review.

Metadata tagging enables users to associate files with clinically relevant labels such as anatomical location and diagnostic purpose. Cropping parameters and tags are saved to the configuration file to ensure reproducibility across sessions.

Anonymisation is achieved by removing or replacing sensitive patient identifiers. Tags specified under the `[attributes_to_del]` section of the configuration

file are deleted. Unique identifiers (UIDs) for studies, series, and instances are regenerated to maintain DICOM compliance.

Finally, processed files are exported as a compressed ZIP archive. The export structure and naming conventions are configurable to support integration with downstream pipelines.

5.3 Error Handling and Logging

Robust error handling ensures the application runs smoothly:

- Validation checks prevent non-DICOM files from being processed.
- Cropping inconsistencies or missing parameters generate clear error messages prompting user correction.
- A structured logging system captures events and errors for debugging and audit purposes.

5.4 Customisability

User preferences – including cropping areas, compression settings, and attribute removal – are configurable via an INI file. This file allows fine-grained control over application behaviour, supporting diverse use cases and device-specific settings. Users may also toggle the autocropping functionality, allowing a flexible mix of manual and automated processing.

5.5 Performance Optimisation

To maintain responsiveness, the application uses multithreading for preview rendering and batch operations. Image handling with NumPy arrays and PIL ensures lightweight and fast processing, even for large datasets.

6 Results

The proposed US-DICOMizer application is evaluated for its effectiveness in streamlining the preparation of ultrasound DICOM files for artificial intelligence (AI) tasks. The results are organised into three parts: a functional walkthrough of the workflow, a benchmark of processing performance, and a usability evaluation based on user feedback.

6.1 Workflow Evaluation

US-DICOMizer guides users through an intuitive sequence of steps: file loading and preview, cropping and tagging, anonymisation, and export. The following describes how each stage operates within the tool and presents the outcomes observed during practical testing.

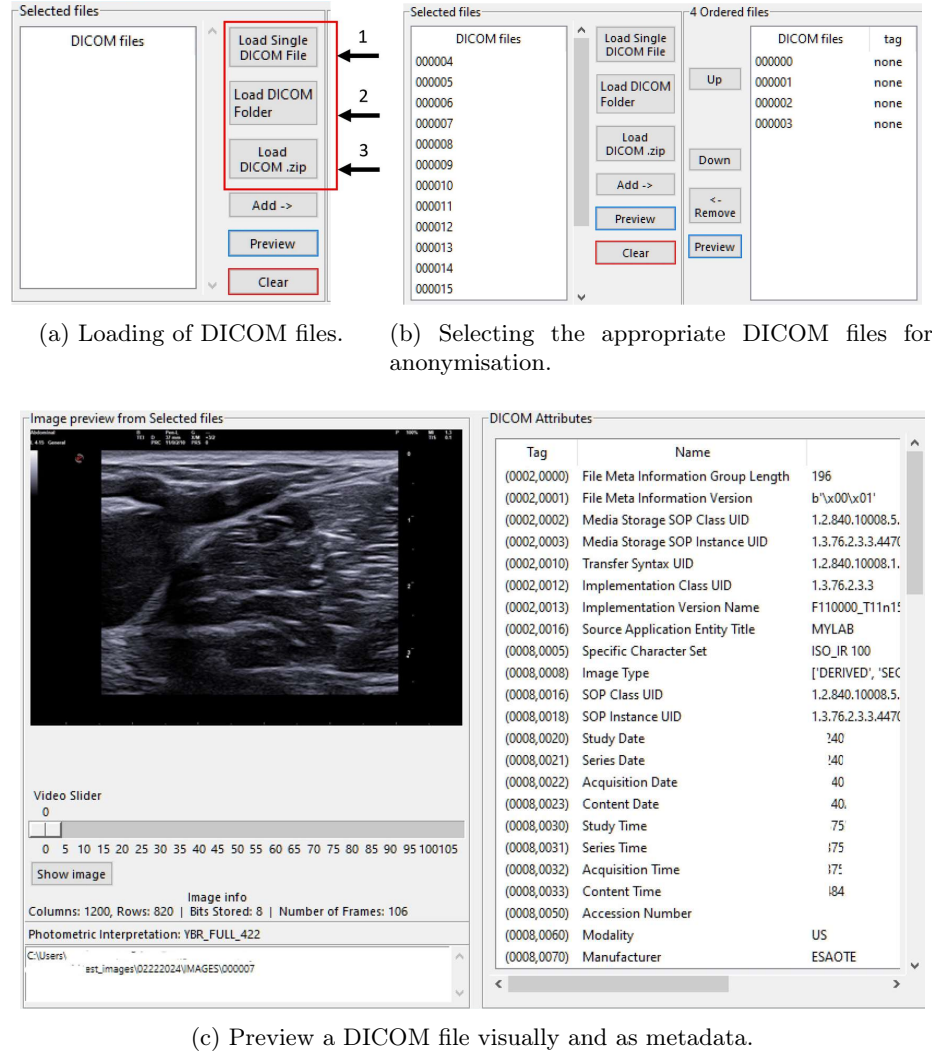


Fig. 3: Loading, selecting and previewing DICOM files [23].

Initially, users start the process by importing DICOM files through multiple options, including selecting individual files, importing entire folders, or uploading compressed archives (Figure 3.a). Imported files are displayed in the “Selected File” list, allowing users to preview images and inspect DICOM attributes, such as dimensions and metadata (Figure 3.c). Multiframe DICOM files include an interactive video slider for frame-by-frame inspection. The interface provides users with the flexibility to reorder, remove, or review files before further processing (Figure 3.b).

Once files are selected, users define cropping regions to isolate diagnostically relevant portions of the image or video frames. Predefined cropping settings are automatically applied where applicable, with the option for manual adjustment using an interactive preview tool. When no device-specific cropping configuration is available – such as for DICOM files from unknown ultrasound systems – the tool applies autocropping automatically. Additionally, users can manually invoke this functionality at any time using the **Auto** button, which performs region-of-interest detection based on image content (Figure 4.a). Cropping parameters could be applied uniformly across all frames or adjusted for individual files, ensuring consistency and relevance in data preparation. Additionally, cropping parameters are saved to a configuration file, allowing users to replicate consistent settings across multiple sessions (Figure 4.b). Concurrently, users add tags to each file to specify anatomical position (e.g., femoral vein (FV) or popliteal vein (PV)), side (e.g., left or right leg), and image purpose (e.g., diagnostic or non-diagnostic quality) (Figure 4.c). These tags ensure the association of critical metadata with the prepared files, enhancing their utility for downstream AI tasks.

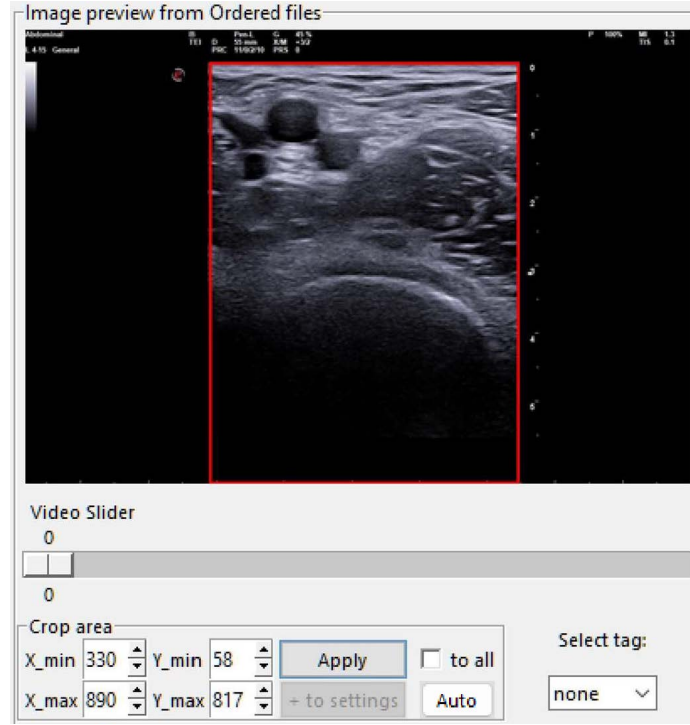
With cropping and tagging completed, the anonymisation process is initiated (Figure 5). The application removes sensitive patient identifiers while preserving essential metadata for AI processing. Files with correctly defined cropping areas and tags are visually marked to indicate readiness for anonymisation. Validation checks prevented the anonymisation of incomplete or improperly prepared files, maintaining data integrity.

Anonymised DICOM files are exported as a compressed ZIP archive for convenient sharing and storage (Figure 6). Users specify the destination folder for the exported archive, completing the preparation process. This final step ensures that the files are ready for integration into AI workflows with minimal manual intervention.

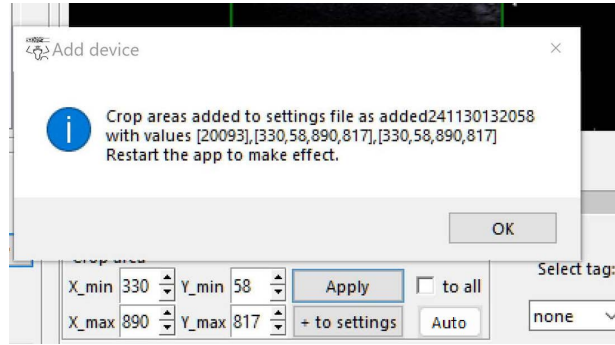
6.2 Performance Evaluation

The application demonstrates high efficiency and ease of use. On average, ultrasound exams comprising 15-30 DICOM files (including single-frame and multi-frame) are fully prepared (including loading, cropping, tagging, anonymisation, and export) within 3-6 minutes. Exporting to a ZIP archive required less than 1-2 seconds for exams of similar size.

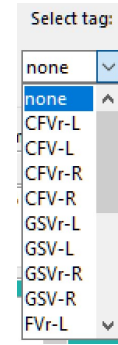
The anonymisation process was evaluated for its average execution times per frame across various photometric interpretations and cropping areas, as shown in Table 2. For single-frame ultrasound images with a resolution of 1200x800 pixels, Monochrome2 photometric interpretation required 19-23 milliseconds, depending on the cropping area. In contrast, the YBR_FULL_422 and RGB interpretations took longer, with times ranging from 25-30 milliseconds and 49-61 milliseconds, respectively. For ultrasound multiframe DICOMs, the processing times increased as expected: Monochrome2 required 102-108 milliseconds per frame, YBR_FULL_422 required 117-124 milliseconds, and RGB interpretation ranged from 241-250 milliseconds per frame.



(a) Process of image cropping.



(b) Save cropping area in settings.



(c) Select tag.

Fig. 4: Cropping and tagging process [23].

These results were produced using ultrasound exams from eight different ultrasound medical devices manufactured by four vendors, and calculated on a laptop computer running Windows 10, with an Intel Core i7-7700HQ CPU at 2.8 GHz and 16 GB of RAM. While DVT datasets were used for testing, the tool is designed for broader ultrasound imaging use cases. Conclusively, this

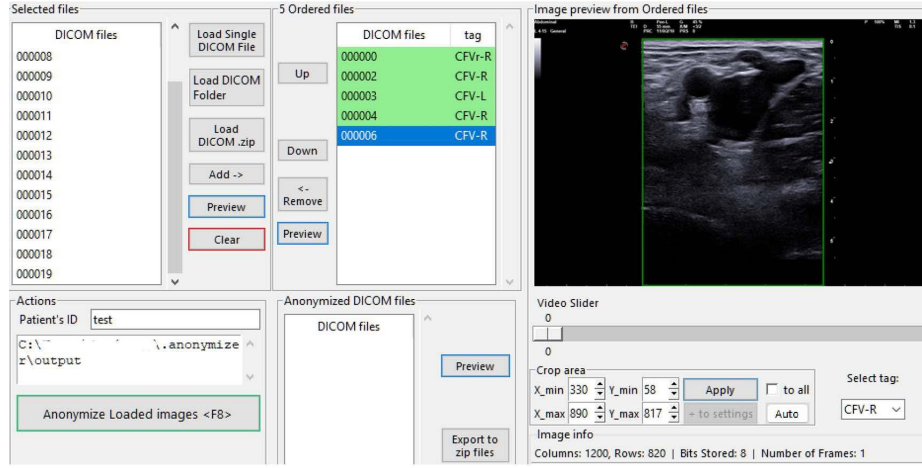


Fig. 5: Start the anonymisation process after properly cropping and tagging the required DICOM files [23].

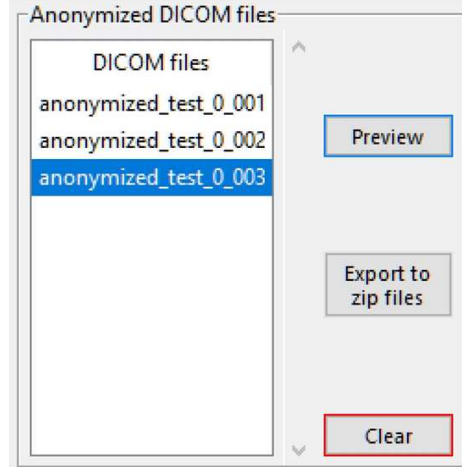


Fig. 6: Export anonymised DICOM files as a ZIP file [23].

experiment highlights that performance trade-offs are associated with both photometric complexity and cropping dimensions.

6.3 Usability Evaluation

A total of 15 healthcare professionals participated in the usability evaluation. Their descriptive characteristics, including gender, age, education level, and prior experience with US-DICOMizer, are summarised in Table 3.

Table 2: Average execution times (per frame) of anonymisation process for various photometric interpretations and cropping areas [23].

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

To evaluate different aspects of usability, both quantitative metrics (SUS and PSSUQ) and qualitative feedback were analysed.

The System Usability Scale (SUS) yielded a mean score of **66.33** out of 100, with a standard deviation (SD) of **11.57**, indicating moderate perceived usability of the tool. The distribution of SUS scores is presented in Figure 7. According to established SUS benchmarks [3], scores above 68 are considered average. Thus, the score of 66.33 suggests marginally below the industry average, but still within a range considered acceptable for early-stage tools.

For the Post-Study System Usability Questionnaire (PSSUQ), the average overall score was **2.35** (SD = 0.81), where lower values indicate higher usability. The category-wise scores were as follows:

- **System Usefulness:** Mean = **2.33**, SD = 0.88
- **Information Quality:** Mean = **2.44**, SD = 0.87
- **Interface Quality:** Mean = **2.31**, SD = 0.96

Since lower scores reflect more positive user experiences, the results indicate generally favorable perceptions across all usability dimensions. Figure 8 shows the distribution of responses across these PSSUQ categories, reinforcing the trend of generally positive feedback.

Table 3: Descriptive characteristics of the participants.

Evaluation Participants' Characteristics		N=15
Gender		
Female		6 (40%)
Male		9 (60%)
Age (years)		
26–45		11 (73%)
46–65		4 (27%)
Highest Education Level		
Bachelor's degree		2 (13%)
Master's degree		8 (53%)
PhD		5 (33%)
Usage Experience		
First time		4 (27%)
2–5 times		2 (13%)
6–15 times		2 (13%)
More than 15 times		7 (47%)

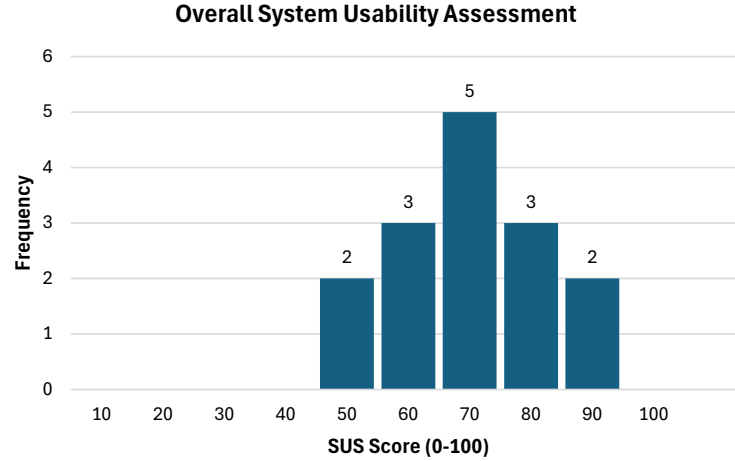


Fig. 7: Distribution of SUS scores from the usability study (N=15).

Participants were also asked to describe any challenges they encountered while using US-DICOMizer and to suggest potential improvements. Most respondents reported no significant difficulties. However, several comments pointed to minor usability issues, including the time-consuming nature of preparing large batches of data, occasional difficulty in defining cropping margins, and challenges understanding the tool when relying solely on the user guide.

Suggestions for improvement included:

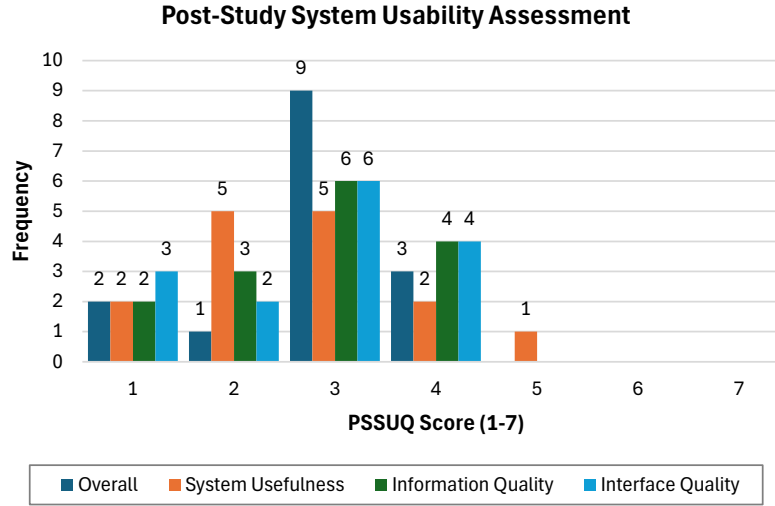


Fig. 8: Distribution of PSSUQ scores by category (N=15).

- More automated processing to improve efficiency for larger datasets;
- A smarter and more intuitive interface;
- UI enhancements such as clearer labelling (e.g., renaming “Actions” to “Anonymisation”), colour coding of interface sections, and drag-and-drop functionality for defining cropping boundaries.

One participant also noted that a short training video was significantly more effective than the written guide alone. Overall, the qualitative feedback supports the quantitative findings: while the tool is broadly usable and functional, enhancements to the interface and workflow automation could further improve the user experience. These results suggest that US-DICOMizer is a practical and usable tool with clear opportunities for refinement based on user-centered feedback.

7 Conclusions and Future Directions

This paper presented US-DICOMizer, an open-source tool designed to streamline the preparation of ultrasound DICOM files for artificial intelligence (AI) applications. By automating critical tasks – such as anonymisation, cropping, and metadata tagging – the tool addresses key challenges in ultrasound imaging workflows, including privacy compliance, region-of-interest extraction, and contextual annotation. Although initially validated using datasets from deep vein thrombosis (DVT) assessments, the tool is designed for broader applicability across clinical and research settings.

The evaluation results demonstrate that US-DICOMizer enhances both the efficiency and consistency of data preparation workflows. Performance bench-

marks show fast processing times even for multiframe DICOMs. Usability testing with 15 healthcare professionals indicated generally favourable perceptions of the interface and functionality, particularly highlighting features such as interactive previews, intuitive tagging, and the newly introduced autocropping functionality. This autocropping option was especially useful in reducing the time required to prepare large datasets. These findings support the tool’s suitability for integration into AI development pipelines, particularly in projects requiring standardised and scalable image preprocessing.

Despite these strengths, several opportunities for improvement remain. Future development will focus on:

- Extending support to additional imaging modalities, such as computed tomography (CT) and magnetic resonance imaging (MRI);
- Integrating AI-based modules for automated region-of-interest detection and semantic tagging;
- Conducting large-scale clinical validation across diverse sites, use cases, and user profiles;
- Enhancing interoperability with electronic case report form (eCRF) systems and research data platforms to support automated workflows and regulatory compliance.

In conclusion, US-DICOMizer provides a robust and extensible framework for preparing ultrasound imaging data for AI-driven applications. Its design balances automation with user control, making it a practical solution for both clinical research and translational AI. As development continues, the tool is poised to support more comprehensive imaging pipelines, accelerate dataset curation, and contribute to the advancement of precision diagnostics in healthcare.

Acknowledgments. This work is co-funded by the European Union, under the Horizon Europe Innovation Action ThrombUS+ (Grant Agreement No. 101137227). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or HADEA as the granting authority. Neither the European Union nor the granting authority HADEA can be held responsible for them. Also, this work was carried out in the context of the Inter-Institutional Master’s Program “Biomedical Informatics” with the support of the School of Medicine, Democritus University of Thrace and the Athena Research Center in Greece.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aryanto, K., Oudkerk, M., van Ooijen, P.: Free DICOM de-identification tools in clinical research: Functioning and safety of patient privacy. *European Radiology* **25**(12), 3685–3695 (2015). <https://doi.org/10.1007/s00330-015-3794-0>
2. Baker, M., Anjum, F., dela Cruz, J.: Deep venous thrombosis ultrasound evaluation. In: StatPearls. StatPearls Publishing, Treasure Island (FL) (2024)

3. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies* **4**(3), 114–123 (2009)
4. Brown, S.M., Sekiguchi, H., Pinsky, M.R.: A new era in critical care ultrasound: Professionalization. *Annals of the American Thoracic Society* **14**(12), 1747–1749 (Dec 2017). <https://doi.org/10.1513/AnnalsATS.201708-624PS>
5. CARE ARTHRITIS Ltd: CARE DICOM Anonymizer. <https://www.carearthritis.com/downloads/#anonymizer>, Accessed: July 2025
6. Carovac, A., Smajlovic, F., Junuzovic, D.: Application of ultrasound in medicine. *Acta Informatica Medica* **19**(3), 168 (2011). <https://doi.org/10.5455/aim.2011.19.168-171>
7. Conrad, S.A.: Focused echocardiography in the ICU. In: Frankel, H.L., deBoisblanc, B.P. (eds.) *Bedside Procedures for the Intensivist*, pp. 139–181. Springer New York, New York, NY (2010). https://doi.org/10.1007/978-0-387-79830-1_7
8. doRadiology: DICOM Anonymizer. <https://dicomanonymizer.com>, Accessed: July 2025
9. Drougka, I.: D7.1: Clinical studies initiation package – Study A (Jun 2024). <https://doi.org/10.5281/zenodo.11371924>, ThrombUS+ Project
10. European Parliament and Council: Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* pp. 1–88 (2016), <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
11. Habehh, H., Gohel, S.: Machine learning in healthcare. *Current Genomics* **22**(4), 291–300 (Dec 2021). <https://doi.org/10.2174/1389202922666210705124359>
12. Haselgrove, C., Poline, J.B., Kennedy, D.N.: Comment on “A simple tool for neuroimaging data sharing”. *Frontiers in Neuroinformatics* **8** (Oct 2014). <https://doi.org/10.3389/fninf.2014.00082>
13. Haselgrove, C., Poline, J.B., Kennedy, D.N.: A simple tool for neuroimaging data sharing. *Frontiers in Neuroinformatics* **8** (May 2014). <https://doi.org/10.3389/fninf.2014.00052>
14. ISO:12052:2017: Health informatics – Digital imaging and communication in medicine (DICOM) including workflow and data management. Standard, International Organization for Standardization (2017), <https://www.iso.org/standard/72941.html>
15. Kaldoudi, E., Marozas, V., Rytis, J., Pousset, N., Legros, M., Kircher, M., Novikov, D., Sakalauskas, A., Moustakidis, P., Ayinde, B., Moltani, L.A., Balling, S., Vehkaoja, A., Oksala, N., Macas, A., Balciuniene, N., Bigaki, M., Potoupnis, M., Papadopoulou, S.L., Grandone, E., Gautier, M., Bouda, S., Schloetelburg, C., Prinz, T., Dionisio, P., Anagnostopoulos, S., Drougka, I., Folkvord, F., Drosatos, G., Didaskalou, S., The ThrombUS+ Consortium: Towards wearable continuous point-of-care monitoring for deep vein thrombosis of the lower limb. In: Jarm, T., Šmerc, R., Mahnič-Kalamiza, S. (eds.) *9th European Medical and Biological Engineering Conference*. pp. 326–335. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-61628-0_36
16. Kitware, Inc.: DICOMAnonymizer. Available on GitHub: <https://github.com/KitwareMedical/dicom-anonymizer>, Accessed: July 2025
17. Lewis, J.R.: Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. *Proceedings of the Human Factors Society Annual Meeting* **36**(16), 1259–1260 (1992). <https://doi.org/10.1177/154193129203601617>

18. Lewis, J.R.: The system usability scale: Past, present, and future. *International Journal of Human-Computer Interaction* **34**(7), 577–590 (2018). <https://doi.org/10.1080/10447318.2018.1455307>
19. Maiti, D., Arunachalam, S.P.: Non-invasive diagnosis of deep vein thrombosis to expedite treatment and prevent pulmonary embolism during gestation. In: *Proceedings of the 2022 Design of Medical Devices Conference*. p. V001T01A003 (Apr 2022). <https://doi.org/10.1115/DMD2022-1054>
20. MicroDicom Ltd: MicroDicom DICOM Viewer. <https://www.microdicom.com>, Accessed: July 2025
21. Monteiro, E., Costa, C., Oliveira, J.L.: A de-Identification pipeline for ultrasound medical images in DICOM format. *Journal of Medical Systems* **41**(5), 89 (May 2017). <https://doi.org/10.1007/s10916-017-0736-1>
22. Nosowsky, R., Giordano, T.J.: The health insurance portability and accountability act of 1996 (hipaa) privacy rule: Implications for clinical research. *Annual Review of Medicine* **57**(Volume 57, 2006), 575–590 (2006). <https://doi.org/10.1146/annurev.med.57.121304.131257>
23. Pechlivanis, D., Didaskalou, S., Kaldoudi, E., Drosatos, G.: Preparing ultrasound imaging data for artificial intelligence tasks: Anonymisation, cropping, and tagging. In: *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, vol. 2, pp. 951–958. SciTePress (2025). <https://doi.org/10.5220/0013379400003911>
24. Rodríguez González, D., Carpenter, T., Van Hemert, J.I., Wardlaw, J.: An open source toolkit for medical imaging de-identification. *European Radiology* **20**(8), 1896–1904 (Aug 2010). <https://doi.org/10.1007/s00330-010-1745-3>
25. Rubo Medical Imaging BV: Rubo DICOM Anonymizer. https://rubomedical.com/dicom_anonymizer/, Accessed: July 2025
26. The Children’s Hospital of Philadelphia: Python DICOM Anonymizer. <https://github.com/chop-dbhi/dicom-anon>, Accessed: July 2025
27. Waheed, S.M., Kudaravalli, P., Hotwagner, D.T.: Deep vein thrombosis. In: *StatPearls*. StatPearls Publishing, Treasure Island (FL) (2024)
28. Xiao, X., Wang, G., Gehrke, J.: Interactive anonymization of sensitive data. In: *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. pp. 1051–1054. ACM, Providence Rhode Island USA (Jun 2009). <https://doi.org/10.1145/1559845.1559979>
29. Zeng, H., Li, L., Cao, Z., Zhang, L.: Grid anchor based image cropping: A new benchmark and an efficient model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(3), 1304–1319 (Mar 2022). <https://doi.org/10.1109/TPAMI.2020.3024207>