# *SpeedyAnnotate*: An Intuitive and Open-Source Tool for Efficient Image Annotation and Quality Comparison

Ian Selby[1,2], Anna Breger[3,4], Michael Roberts[4,5], Lorena Escudero Sánchez[1], Judith Babar[1,2], James Rudd[1,5], Fiona Gilbert[1,2], Carola-Bibiane Schönlieb[4], Evis Sala[6], Jonathan Weir-McCall[1,2]

Department of Radiology, University of Cambridge[1]; Cambridge University Hospitals NHS Trust, Cambridge[2]; Center of Medical Physics and Biomedical Engineering, Medical University of Vienna[3]; Department of Applied Mathematics and Theoretical Physics, University of Cambridge[4]; School of Clinical Medicine, University of Cambridge[5]; Advanced Radiodiagnostics Centre, Fondazione Policlinico Universitario Agostino Gemelli and Università Cattolica del Sacro Cuore[6]

## 1. Introduction

- We present *SpeedyAnnotate,* a freely available image annotation software designed to streamline labelling and image quality assessment.
- To demonstrate its utility, we investigated the reliability of DICOM headers in chest radiographs from five publicly available COVID-19 datasets covering 39 centres across seven countries.

## 2. Methods

**Development and Features:**

- Built using PyQt6 in Python.
- Fully customizable, offering flexibility for various annotation tasks:
  - **Individual Image Annotation**: Allows users to annotate diagnostic or quality control findings using checkboxes, radiobuttons, and free text (Figure 1).
  - **Bounding Box Creation**: Facilitates marking findings and regions of interest for segmentation model development.
  - **Comparative Image Quality Assessment (IQA)**: Enables side-by-side comparison of image quality against a reference using structured layouts with checkboxes, radiobuttons, dropdowns, and free text (Figure 2).

**Ease of Use:**

- Features a **wizard-based setup** for intuitive configuration.
- Requires **minimal technical expertise**.
- Cross-platform compatibility: Works seamlessly on both Mac and Windows.

**Accessibility:**

- **Freely downloadable from GitLab**: see the QR code below or https://gitlab.developers.cam.ac.uk/maths/cia/covid-19-projects/speedyannotate.

**Application in Research:**

- We investigated the reliability of DICOM headers for chest radiographs in five publicly available COVID-19 datasets, namely the NCCID, BIMCV-COVID-19, BriXia, RICORD-1C and CUH.

LinkedIn

SpeedyAnnotate

Thanks to: NCITA National Cancer Imaging Translational Accelerator; CANCER RESEARCH UK; intel; DRAGON; cmih; NIHR | Cambridge Biomedical Research Centre

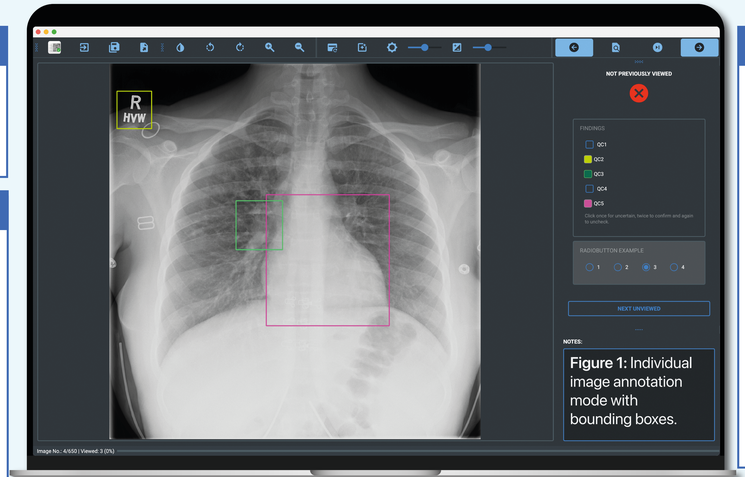



**Figure 1:** Individual image annotation mode with bounding boxes.

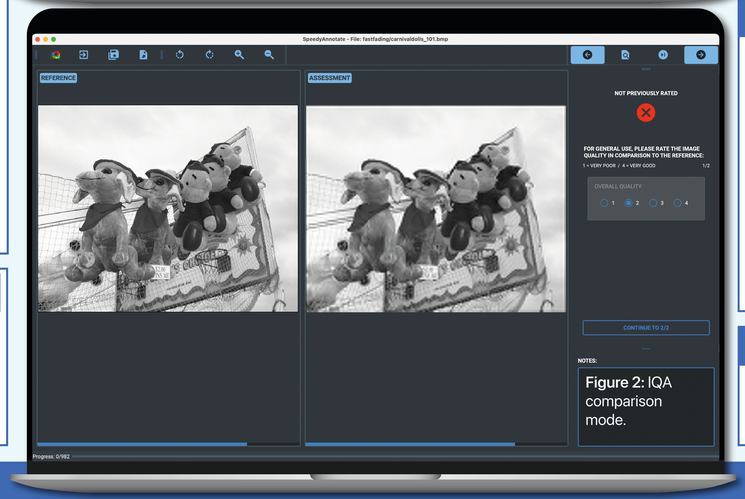



**Figure 2:** IQA comparison mode.

## 3. Results

22,274 chest radiographs were included in our analysis*:*

**Non-Chest Radiographs:**

- 229 radiographs (1.03%) were identified as non-chest images.

**Grayscale Inversion Discrepancies:**

- In one dataset, 83 images (0.95%) had a Photometric Interpretation header mismatch with grayscale inversion.
- All affected images belonged to the COVID-19 negative cohort.

**Series Description Header Analysis:**

- Found in 56.5% (12,580/22,274) of the DICOM metadata files analyzed.
- Incorrectly identified lateral projections in 2.9% (364/12,580) of cases.
- Missing, incomplete, or ambiguous in 43.5% (9,694/22,274) of cases.

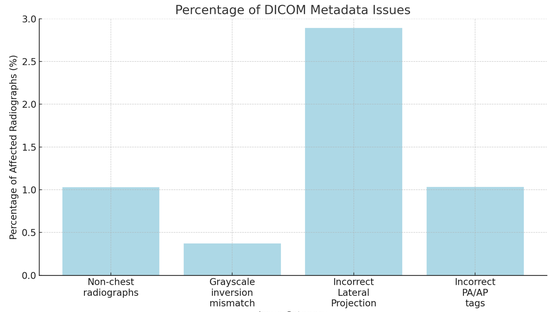



**Figure 3:** Percentage of radiographs affected by DICOM metadata issues across different tags.

## 4. Discussion

**DICOM Metadata Challenges:**

- DICOM metadata shows inconsistencies, including errors in identifying projections, grayscale inversion mismatches, and incomplete or ambiguous headers.
- These issues can lead to biases, reduced generalizability, and inaccuracies in AI models.

**Need for EDA:**

- Exploratory data analysis (EDA) is essential before AI development to identify and address metadata and imaging inconsistencies.
- Rigorous EDA helps ensure dataset reliability and prevents spurious patterns from being learned.

**Role of *SpeedyAnnotate*:**

- *SpeedyAnnotate* facilitates detailed EDA, quality control, and annotation, addressing dataset inconsistencies efficiently.
- Its customizable features support diverse tasks, such as metadata verification and labelling for segmentation models.

## 5. Conclusions

- **Tools like *SpeedyAnnotate* enhance dataset quality and AI robustness.**
- **Incorporating such tools into workflows ensures more accurate and clinically relevant AI systems.**