Is Your De-identified Radiograph Truly De-identified?

Pouria Rouzrokh, Bardia Khosravi, Shahriar Faghani, Austin Grove, Elizabeth Sun Kaji, Michael J Taunton[1], Bradley James Erickson[1], Cody Wyles[1]
[1]Mayo Clinic

INTRODUCTION: De-identifying medical imaging data is crucial for ensuring patient privacy during research endeavors, data transfers, and the development of artificial intelligence (AI) models. For radiographs, the two most common strategies for de-identification are: 1) the removal or substitution of tags containing protected health information (PHI) in the DICOM metadata, and 2) altering pixel values in areas containing PHI on the image itself (e.g., by covering the area with a black box). Considering these strategies, we hypothesized that specialized AI models might still detect PHI "fingerprints" from imaging features, potentially revealing patient identities.

METHODS: After obtaining institutional review board approval, we collected 661,124 anterior-posterior (AP), lateral, and oblique hip and pelvis radiographs from our institutional total joint arthroplasty registry. We de-identified them using the aforementioned strategies. The data were then split into training, validation, and test folds at the patient level. We trained a ConvNeXt (v2) model to encode each input radiograph into a one-dimensional vector of size 768. The model was designed to increase the cosine similarity between encoded vectors of imaging pairs belonging to the same patient, regardless of view and time distance, and to decrease it for other pairs. After training, we identified a cosine similarity cutoff threshold for distinguishing the output vectors for a pair of radiographs as belonging to similar or different patients by plotting the Receiver Operating Characteristic (ROC) curve of that prediction on our validation set. We then applied this threshold to our test set and also leveraged t-SNE plots and integrated gradient maps (IGMs) to elucidate the model's performance.

RESULTS: The comparison of cosine similarity between encoded vectors for radiographic pairs against the calibrated threshold successfully detected true identity relationships in 94% of all possible pairs in the test set. The pipeline had reliable performance even across pairs that contained radiographs of different views and with years of time distance between them. IGMs highlighted the contour of the bones as the most prominent features learned by the model (Figure).

DISCUSSION AND CONCLUSION:

Despite de-identifying radiographs by both of the two most common methods to protect PHI, a self-supervised AI model demonstrated the capability to detect radiographs belonging to the same patients but taken from different views or times with 94% accuracy.

Clinical Relevance/Application:

Relying solely on DICOM metadata alteration and pixel value changes in radiographs may not sufficiently de-identify them. Medical systems need more robust de-identification strategies for highly sensitive data access situations.