



Abstract: Is Medical Chest X-ray Data Anonymous?

Deep Learning-based Patient Re-identification is Able to Exploit the Biometric Nature of Medical Chest X-ray Data

Kai Packhäuser, Sebastian Gündel, Nicolas Münster, Christopher Syben,
Vincent Christlein, Andreas Maier

Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
kai.packhaeuser@fau.de

With the rise and ever-increasing potential of deep learning techniques in recent years, publicly available medical datasets became a key factor to enable reproducible development of diagnostic algorithms in the medical domain. Medical data contains sensitive patient-related information and is therefore usually anonymized by removing patient identifiers, e. g., patient names before publication. To the best of our knowledge, we are the first to show that a well-trained deep learning system is able to recover the patient identity from chest X-ray data. We demonstrate this using the publicly available large-scale ChestX-ray14 dataset, a collection of 112,120 frontal-view chest X-ray images from 30,805 unique patients. Our verification system is able to identify whether two frontal chest X-ray images are from the same person with an AUC of 0.9940 and a classification accuracy of 95.55 %. We further highlight that the proposed system is able to reveal the same person even ten and more years after the initial scan. When pursuing a retrieval approach, we observe an mAP@R of 0.9748 and a precision@1 of 0.9963. Furthermore, we achieve an AUC of up to 0.9870 and a precision@1 of up to 0.9444 when evaluating our trained networks on external datasets such as CheXpert and the COVID-19 Image Data Collection. Based on this high identification rate, a potential attacker may leak patient-related information and additionally cross-reference images to obtain more information. Thus, there is a great risk of sensitive content falling into unauthorized hands or being disseminated against the will of the concerned patients. Especially during the COVID-19 pandemic, numerous chest X-ray datasets have been published to advance research. Therefore, such data may be vulnerable to potential attacks by deep learning-based re-identification algorithms. This work has previously been published in Scientific Reports [1].

References

1. Packhäuser K, Gündel S, Münster N, Syben C, Christlein V, Maier A. Deep learning-based patient re-identification is able to exploit the biometric nature of medical chest X-ray data. *Sci Rep.* 2022;12(1):1–13.