

АНОНИМИЗАЦИЯ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ В СОВРЕМЕННЫХ ТЕХНОЛОГИЯХ

Бондаренко Владимир Александрович, студент 2 курса бакалавриата института Экономики и управления АПК, ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, atombaytovich@gmail.com

Научный руководитель – Красовская Людмила Владимировна, к.т.н., доцент, доцент кафедры прикладной информатики, ФГБОУ ВО РГАУ – МСХА имени К. А. Тимирязева, kraslud@yandex.ru

Аннотация. Рассмотрены возможности и подходы к обезличиванию цифровых медицинских изображений, таких как рентгенограммы и флюорографические снимки. Представлены алгоритмы и методы удаления персональных данных из файлов форматов DICOM и IOD [2, 4]. Показаны преимущества обезличивания для защиты конфиденциальности пациентов и интеграции с информационными системами.

Ключевые слова: рентгенография, обезличивание, цифровые изображения, DICOM, информационная безопасность

ANONYMIZATION OF DIGITAL IMAGES IN MODERN TECHNOLOGIES

Bondarenko Vladimir Alexandrovich, second-year undergraduate student at the Institute of Economics and Management of the Agro-Industrial Complex, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, atombaytovich@gmail.com

Scientific supervisor – Krasovskaya Lyudmila Vladimirovna, Ph.D in Technical Science, Associate Professor of the Department of Applied Informatics, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy, kraslud@yandex.ru

Annotation. The possibilities and approaches to de-identification of digital medical images, such as X-rays and fluorographic images, are considered. Algorithms and methods for removing personal data from DICOM and IOD file formats are presented. The advantages of de-identification for protecting patient confidentiality and integration with information systems are demonstrated.

Key words: radiography, depersonalization, digital images, DICOM, information security

Обезличивание медицинских данных, включая цифровые рентгенограммы и флюорографические снимки, является важным этапом обеспечения

информационной безопасности пациентов в современной медицине. В связи с ростом цифровизации и передачи медицинских изображений между учреждениями, возникает необходимость в удалении личной информации, которая может нарушить конфиденциальность пациента. Статья направлена на исследование алгоритмов и технологий, используемых для обезличивания цифровых рентгенограмм и флюорографических изображений, а также их эффективности [1, 3].

Основная цель обезличивания цифровых рентгенографических и флюорографических изображений заключается в исключении любой идентифицирующей информации из файлов, что обеспечивает их безопасное хранение и использование в исследованиях и клинической практике. Для достижения этой цели разработана программа, способная обезличивать файлы форматов DICOM и IOD, в которых содержатся персональные данные.

Файлы DICOM (Digital Imaging and Communications in Medicine) представляют собой стандарт для хранения и передачи медицинских изображений и соответствующей информации. Они содержат как графическую информацию, так и метаданные, которые включают идентифицирующие данные пациента: имя, дату рождения, идентификаторы исследований и другие чувствительные данные. Обезличивание данных DICOM и IOD (Information Object Definitions) требует тщательного подхода к обработке и удалению метаданных [9].

Для реализации обезличивания был разработан комплексный подход, включающий следующие этапы:

1. Удаление персональных метаданных. В первую очередь удаляются элементы данных, которые могут идентифицировать пациента, такие как PatientName, PatientID, PatientBirthDate и другие.

2. Замена значений на нейтральные: в некоторых случаях вместо удаления данных целесообразно заменять их на нейтральные значения, чтобы сохранить структуру данных для анализа, но исключить любую возможность идентификации. Например, поля с информацией о дате могут быть заменены на "01-01-2000", что сохраняет формат данных.

3. Проверка корректности обезличивания: Важным этапом является тестирование алгоритмов обезличивания на различных типах изображений. Были проведены тесты на рентгенографических и флюорографических изображениях с целью проверки правильности удаления идентифицирующих данных и сохранения диагностической ценности изображений. После обезличивания с помощью данной программы изображения всегда открываются не только в системах PACS, но и в RadiAnt и других программах уже в обезличенном виде, что позволяет сохранять диагностическую ценность и обеспечивать полную совместимость.

Для обработки данных поступает поток необезличенных файлов. Программа рекурсивно проходит по директориям, чтобы найти все файлы форматов DICOM и IOD, и применяет алгоритм обезличивания к каждому из

них. Все обезличенные файлы сохраняются в директории, указанной пользователем [6, 8].

Формат DICOM включает в себя не только графическую часть, но и значительное количество метаданных, которые содержат информацию о пациенте, оборудовании и параметрах исследования. Эти данные представлены в виде тэгов (например, Tag (0010,0010) – PatientName). Для обезличивания используются алгоритмы, которые проходят по всем тэгам файла и удаляют или модифицируют те из них, которые содержат персональные данные [4, 5].

Особое внимание уделяется IOD (Information Object Definitions), которые представляют собой стандартизированные структуры данных в DICOM [9]. Обезличивание IOD необходимо для обеспечения того, чтобы все элементы, потенциально содержащие идентифицирующую информацию, были либо удалены, либо модифицированы.

Анализ распределения идентифицирующих тэгов в DICOM-файлах показал, что около 60% тэгов могут содержать персональные данные, которые необходимо удалить или изменить (рисунок 1). Это было подтверждено в ходе исследования, проведенного на нескольких наборах данных DICOM, и результаты показали необходимость тщательного подхода к удалению или модификации этих данных.

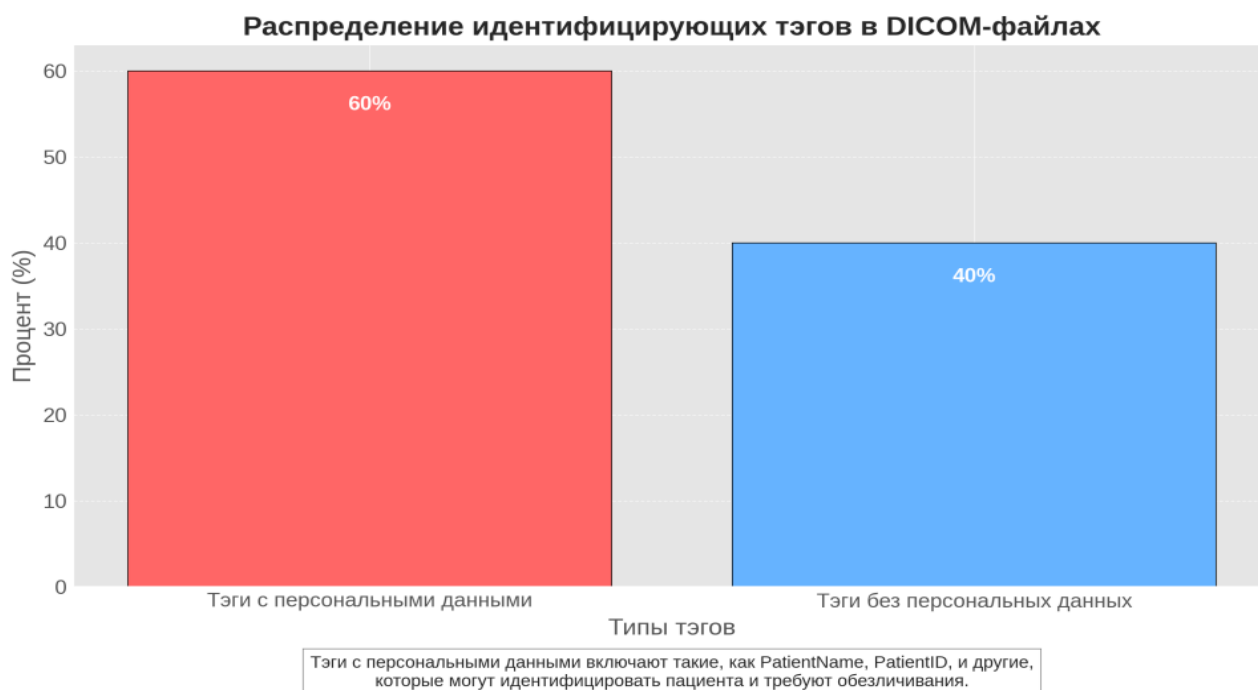


Рисунок 6 – Диаграмма анализа распределения идентифицирующих тэгов в DICOM-файлах

Цифровизация медицинских изображений является одним из наиболее значительных достижений в медицине последних десятилетий. По данным ВОЗ, ежегодно проводится около 3,6 миллиарда диагностических медицинских обследований, таких как рентген, и это число продолжает расти по мере того, как больше людей получают доступ к медицинскому обслуживанию. Примерно 350

миллионов из них проводятся детям в возрасте до 15 лет. Эти данные основаны на официальных отчетах ВОЗ и отражают текущий объем медицинских обследований в мире. В среднем, около 80% больниц и медицинских центров используют системы PACS для хранения и передачи изображений, что делает обезличивание ключевым аспектом защиты данных пациентов.

Однако утечек данных при реализации предложенного подхода не произойдет, так как процесс обезличивания происходит локально на устройстве, которое изначально защищено и соответствует требованиям информационной безопасности для медицинских организаций. Интернет не используется на этапе обезличивания, что минимизирует риски утечки.

Разработанная программа успешно справляется с задачей обезличивания файлов форматов DICOM и IOD на всех типах цифровых медицинских изображений, таких как рентгенограммы и флюорографические снимки. Программа поддерживает работу как на платформах Windows, так и на Linux. В ходе тестирования было установлено, что все персональные данные были удалены или заменены на нейтральные значения, а изображения сохранили свою диагностическую ценность. Алгоритмы продемонстрировали высокую эффективность и производительность, что позволяет интегрировать их в существующие медицинские информационные системы.

Библиографический список

1. Luo SZhang JXiao NQiang YLi KZhao JMeng LSong P(2022)DAS-Net: A lung nodule segmentation method based on adaptive dual-branch attention and shadow mapping Applied Intelligence 10.1007/s10489-021-03038-252:13(15617-15631) Online publication date: 1-Oct-2022 <https://dl.acm.org/doi/10.1007/s10489-021-03038-2>
2. Cinaglia PCannataro M(2024)Quantum Computing for BioinformaticsReference Module in Life Sciences 10.1016/B978-0-323-95502-7.00182-2 Online publication date: 2024 <https://doi.org/10.1016/B978-0-323-95502-7.00182-2>
3. Cinaglia P(2024)Time Series Analysis and Forecasting for Epidemiology and Pandemic SurveillanceReference Module in Life Sciences 10.1016/B978-0-323-95502-7.00133 <https://doi.org/10.1016/B978-0-323-95502-7.00133-0>
4. Marco Eichelberg, Klaus Kleber and Mark Kammerer. 2020. Cybersecurity issues for PACS and medical imaging. Academic Radiology 27, 8 (2020), 14. <https://doi.org/10.1016/j.acra.2020.03.026>
5. Yi T, Pan I, Collins S, Chen F, Cueto R, Hsieh B, Hsieh C, Smith JL, Yang L, Liao WH, Merck LH, Bai H, Merck D. DICOM Image Analysis and Archive (DIANA): an Open-Source System for Clinical AI Applications. Digit Imaging. 2021 Dec;34(6):1405-1413. <https://doi.org/10.1007/s10278-021-00488-5>
6. Gorman C., Punzo D., Octaviano I., Pieper S., Longabaugh W.J.R., Clunie D.A., Kikinis R., Fedorov A.Y., Herrmann M.D. Interoperable slide microscopy viewer and annotation tool for imaging data science and computational pathology. Nat. Commun. 2023;14:1572. <https://doi.org/10.1038/s41467-023-37224-2>

7. Kolekar S, Chen H, Kim K. 2023. Design of Precision Medicine Web-service Platform Towards Health Care Digital Twin. 2023 Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN). <https://doi.org/10.1109/ICUFN57995.2023.10199942>
8. Larobina M. Tomography. Thirty Years of the DICOM Standard. 2023 Oct 6;9(5):1829-1838. <https://doi.org/10.3390/tomography9050145>
9. "Dicom iod", Jun. 2020, [online] Available: [http://dicom.nema.org/medical/dicom/current/output/chtml/part04/chapter_6.html#:~:text.](http://dicom.nema.org/medical/dicom/current/output/chtml/part04/chapter_6.html#:~:text=)