GAZİ ÜNİVERSİTESİ

MÜHENDİSLİK FAKÜLTESİ BİLGİSAYAR MÜHENDİSLİĞİ

CENG482 / Introduction to Computer Security

Project of Security for Medical Data

Prof. Dr. Şeref SAĞIROĞLU

Aylin AYGÜL 201180060

Atakan KAPLAN 191180768

Aycan KAYNAKCI 191180052

2024

Contents

Table of Figures

Introduction

Magnetic Resonance Imaging (MRI) is a critical diagnostic tool in medical practice, providing detailed images of the internal structures of the body. While MRI scans are invaluable for diagnosing and monitoring various conditions, the storage and handling of MRI data raise significant privacy concerns. Patient information embedded within these images, if not properly anonymized, can lead to privacy breaches. This report addresses these concerns by implementing a Python-based solution to anonymize MRI DICOM files, ensuring patient privacy is maintained.

Objective

The objective of this project is to develop a Python program to load, display, and anonymize MRI DICOM files. The program aims to protect sensitive patient information while maintaining the integrity of the medical images. The key tasks include:

- Loading and displaying MRI DICOM images.
- Extracting and printing patient information from DICOM files.
- Anonymizing patient information in the DICOM files.
- Saving the anonymized DICOM files for further use.

Privacy Issues in MRI Data

MRI data typically contains sensitive personal information within the DICOM files. If these files are shared without proper anonymization, there is a risk of exposing patient identities and health information, which can lead to various privacy violations. The Health Insurance Portability and Accountability Act (HIPAA) in the United States and similar regulations in other countries mandate the protection of patient information, making anonymization a critical task.

Patient Identifiable Information (PII) in MRI Data

MRI DICOM files not only contain image data but also include a wealth of metadata that often carries personally identifiable information (PII). This metadata typically includes:

- Patient's Name: Identifies the patient by name.
- Patient ID: A unique identifier assigned by the healthcare provider.
- Birth Date: Indicates the patient's date of birth.
- Sex: Records the patient's gender.
- Medical History: Includes details about the patient's medical conditions and history.

The presence of such detailed information poses a significant risk if the data is accessed by unauthorized individuals. It can lead to identity theft, discrimination, and other privacy violations.

Legal and Ethical Considerations

The handling of MRI data is governed by various laws and regulations designed to protect patient privacy. Notable among these are:

- HIPAA (Health Insurance Portability and Accountability Act) in the United States: Requires the protection of patient information and sets standards for its electronic transmission and storage.
- GDPR (General Data Protection Regulation) in the European Union: Imposes strict guidelines on the handling of personal data, including health information.

Healthcare providers are ethically obligated to ensure the confidentiality of patient data. Failure to anonymize MRI DICOM files can result in severe legal repercussions, financial penalties, and loss of trust.

Risks of Data Breaches

Data breaches in the healthcare sector can have devastating consequences. When MRI data is not properly anonymized, the risks include:

- Unauthorized Access: Hackers or malicious entities gaining access to sensitive information.
- Data Leakage: Unintended exposure of patient data during transfers or sharing between entities.
- Misuse of Data: Use of personal health information for malicious purposes, such as fraud or discrimination.

Techniques for Ensuring Confidentiality

De-identification

De-identification involves removing or modifying personal information from medical images. Techniques such as face masking, skull stripping, and defacing are used to prevent re-identification of individuals from medical images.

Skull Stripping with FSL and BET

A common de-identification technique used in MRI data is skull stripping, which involves removing non-brain tissue from brain MRI images. This is often done using the Brain Extraction Tool (BET) from the FMRIB Software Library (FSL). BET efficiently removes the skull and other non-brain tissues, which helps protect patient identity by removing potentially identifiable facial features.

FSL is a comprehensive library of analysis tools for functional, structural, and diffusion MRI brain imaging data. BET is a tool within FSL that performs brain extraction (or skull stripping) and is widely used in neuroimaging research for preprocessing MRI data.

Pseudonymization

Pseudonymization replaces private identifiers with fake identifiers or pseudonyms. In DICOM files, required attributes are replaced with pseudo-values, while the original values are encrypted and stored securely.

Differential Privacy

Differential privacy adds random noise to the data to mask individual contributions while allowing aggregate data analysis. This makes it challenging to infer information about any specific individual.

Synthetic Data Generation

Synthetic data generation involves creating artificial datasets that mimic the statistical properties of real data. Techniques such as generative adversarial networks (GANs) are employed to generate synthetic medical images that do not correspond to real patients.

Data Masking

Data masking involves hiding or altering specific data within an image. Techniques include blurring, pixelation, and the application of masks to obscure sensitive regions.

Encryption

Encryption secures data by converting it into a coded format that can only be decoded with a specific key. Algorithms like RSA, AES, and Triple-DES are used to encrypt DICOM attributes, ensuring that sensitive data within medical images is protected from unauthorized access during storage and transmission.

Formats for Anonymized Medical Images

DICOM (Digital Imaging and Communications in Medicine)

The DICOM standard includes specific guidelines for de-identification and pseudonymization of medical images. The Basic Application Level Confidentiality Profile is commonly used, which mandates the anonymization of specific attributes, such as patient names and identifiers.

NIfTI (Neuroimaging Informatics Technology Initiative)

NIfTI is often used for storing neuroimaging data, especially in research settings. Anonymization in NIfTI involves removing or replacing identifying information embedded in the metadata.

Implementation and Compliance

Advanced tools and software solutions leverage deep learning algorithms to automate the anonymization process, ensuring scalability and consistency. These solutions can detect and obscure sensitive information in real-time while preserving the diagnostic quality of the images.

Anonymization practices must comply with regulations like HIPAA and GDPR. Tools typically include features for audit and logging to track anonymization processes, ensuring transparency and compliance.

Our Project

This Flask application serves as a robust backend for medical image processing and anonymization. It offers a range of functionalities to handle DICOM and NIfTI files securely while ensuring patient privacy and confidentiality.

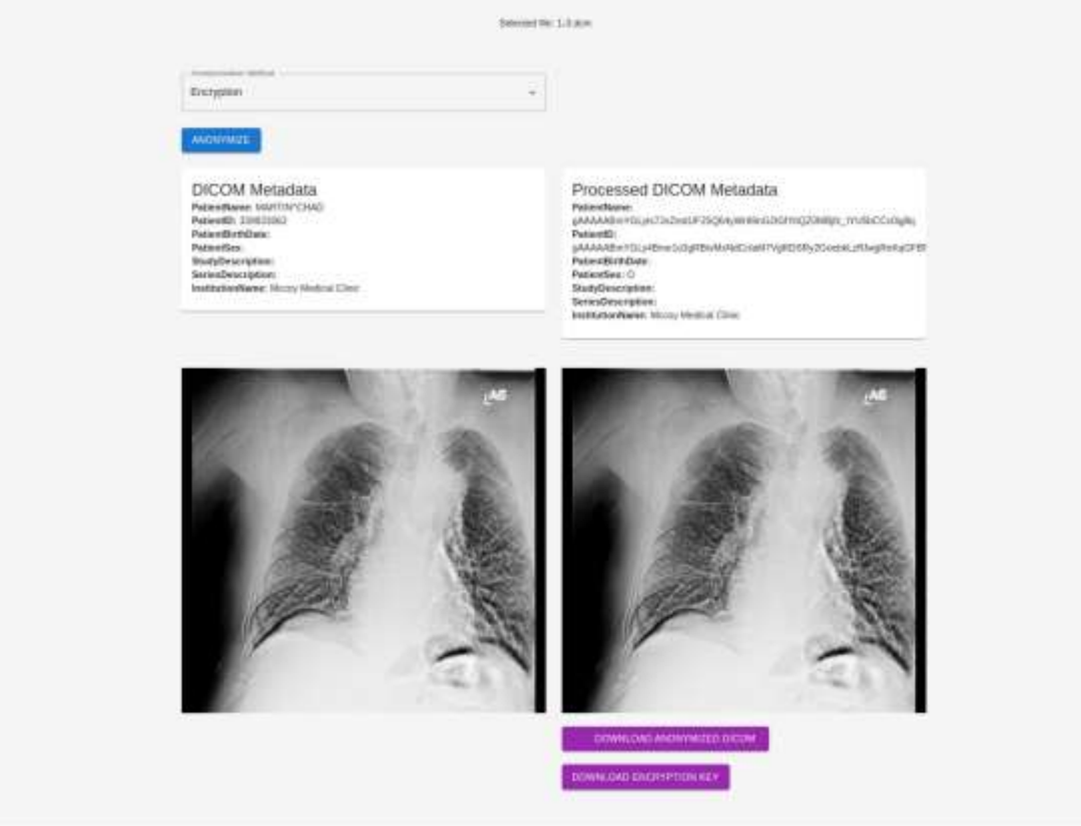
One of the core features is DICOM image processing, facilitated through the `/skull_strip` endpoint. This endpoint accepts DICOM files, performs skull stripping using the Brain Extraction Tool (BET), and provides the processed DICOM file as output. Additionally, the `/process_nifti` endpoint enables visualization of axial, coronal, and sagittal slices from NIfTI files, enhancing diagnostic capabilities.

Anonymization of DICOM files is a critical aspect addressed by the application. The `/anonymize` endpoint offers various anonymization methods, including suppression, randomization, pseudonymization, bucketization, slicing, and encryption. Users can also manipulate image text, such as blurring, removing, or pseudonymizing it, to further protect sensitive information.

For encryption and decryption operations, the application employs the Fernet symmetric encryption scheme. It ensures that patient names and IDs are securely encrypted and decrypted using a generated encryption key. This approach adds an extra layer of security, safeguarding patient data from unauthorized access.

To support these functionalities, the application relies on a set of dependencies, including Flask, Flask-CORS, pydicom, PIL (Python Imaging Library), numpy, nibabel, matplotlib, and base64. Error handling mechanisms are also in place to manage file uploads, processing errors, and encryption/decryption failures gracefully.

In essence, this Flask application serves as a comprehensive solution for medical image management and anonymization. It can be deployed in healthcare settings to facilitate secure handling of medical images, ensuring compliance with privacy regulations and maintaining patient trust and confidentiality.



Selected file: 1-3.dcm

Encryption

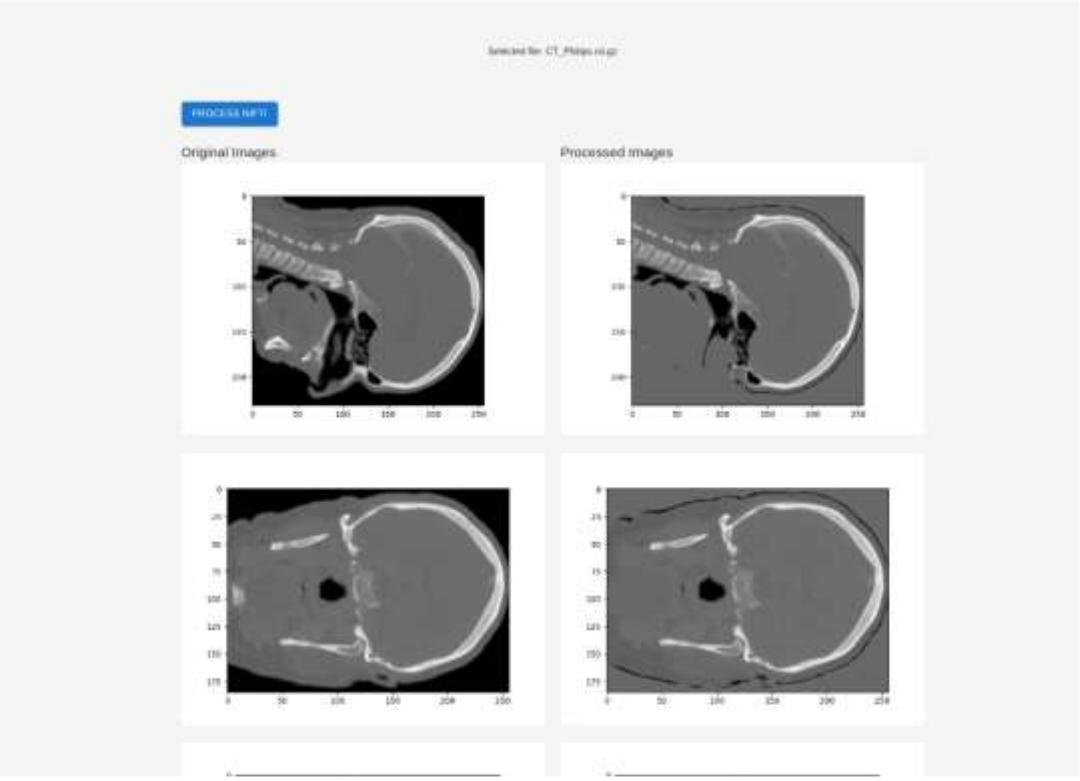
DICOM Metadata

PatientName: MARTIN^CHAD
PatientID: [illegible]
PatientBirthDate:
PatientSex:
StudyDescription:
SeriesDescription:
InstitutionName: Mccoy Medical Clinic

Processed DICOM Metadata

PatientName:
gAAAAAB[illegible]
PatientID:
gAAAAAB[illegible]
PatientBirthDate:
PatientSex: O
StudyDescription:
SeriesDescription:
InstitutionName: Mccoy Medical Clinic

Figure 1 Hide in the DICOM file



Selected file: CT_Philips.nii.gz

Figure 2 Hide in the NIFTI

Conclusion

Medical image anonymization is a multi-faceted approach involving various techniques to protect patient privacy. Methods such as de-identification, pseudonymization, differential privacy, and synthetic data generation play critical roles. Ensuring compliance with regulations like GDPR and HIPAA is paramount, and advanced tools are available to automate and streamline the anonymization process.

This project demonstrates the importance of anonymizing MRI DICOM files to protect patient privacy. By using Python and relevant libraries such as pydicom and matplotlib, we successfully loaded, displayed, anonymized, and saved DICOM files. This approach is crucial for complying with privacy regulations and ensuring that patient information remains confidential when sharing medical data for research or other purposes.

References

1. Health Insurance Portability and Accountability Act (HIPAA). (n.d.). HHS.gov. Retrieved from https://www.hhs.gov/hipaa/index.html
2. Pydicom Documentation. (n.d.). Retrieved from https://pydicom.github.io/
3. Matplotlib Documentation. (n.d.). Retrieved from https://matplotlib.org/
4. HIPAA Privacy Rule - Health Insurance Portability and Accountability Act, U.S. Department of Health & Human Services. (n.d.). Retrieved from https://www.hhs.gov/hipaa/for-professionals/privacy/index.html
5. General Data Protection Regulation (GDPR) - Regulation (EU) 2016/679 of the European Parliament and of the Council. (n.d.). Retrieved from https://eur-lex.europa.eu/eli/reg/2016/679/oj
6. Kahn, C. E., Jr., & Thao, C. (2015). The RSNA Image Share Validation Program: A Model for Implementing Best Practices in Imaging Informatics. *Journal of the American College of Radiology, 12*(2), 151-153. https://doi.org/10.1016/j.jacr.2014.10.007
7. Yasaka, K., & Abe, O. (2018). Deep learning and artificial intelligence in radiology: Current applications and future directions. *PLoS Medicine, 15*(11), e1002707. https://doi.org/10.1371/journal.pmed.1002707
8. Doyle-Lindrud, S. (2019). The Health Insurance Portability and Accountability Act (HIPAA): Privacy and Security Issues in the Digital Age. *Clinical Journal of Oncology Nursing*, *23*(1), 66-70. https://doi.org/10.1188/19.CJON.66-70