CHAVID3S: CHAVI DICOM DATA DE IDENTIFICATION SYSTEM

The user manual for the step-by-step installation and use of CHAVID3S.

This is a collaborative project between Indian Institute of Technology Kharagpur and Tata Medical centre Kolkata

Lead Investigators

Prof. Jayanta Mukhopadhyay
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Email:jay@cse.iitkgp.ac.in
and
Dr. Sanjoy Chatterjee
Department of Radiation Oncology
Tata Medical Center, Kolkata
Email:sanjoy.chatterjee@tmckolkata.com

Teams
Surajit Kundu[1](surajit.113125@gmail.com), Santam
Chakraborty[2](santam.chakraborty@tmckolkata.com), Jayanta
Mukhopadhyay[1](jay@cse.iitkgp.ac.in), Sanjoy Chatterjee[2](sanjoy.chatterjee@tmckolkata.com),
Rimpa Basu Achari[2](rimpa.achari@tmckolkata.com), Indranil
Mallick[2](indranil.mallick@tmckolkata.com), Partha Pratim Das[1](ppd@cse.iitkgp.ac.in), Moses
Arunsingh[2](moses.arunsingh@tmckolkata.com), Tapesh
Bhattacharyyaa[2](tapesh.bhattacharyya@tmckolkata.com), Soumendranath
Ray[2](soumen.ray@tmckolkata.com)

March 2023

---

[1]Indian Institute of Techonology Kharagpur
[2]Tata Medical Center Kolkata

Contents

1 An Overview

CompreHensive ArchiVe of Imaging (CHAVI) DICOM Data De-identification System (**CHAVID3S**) version 1.0 [1] has been developed for the de-identification of radiological DICOM images and associated radiation therapy (RT) planning (RTPLAN), structure (RTSTRUCT), and dose (RT-DOSE) data. It is a standalone application that is built using Java. The graphical user interface (GUI) is designed using JavaFX. The CHAVID3S uses a MySQL relational database management system (RDBMS) to store the data, which keeps tracking the references of the data in encrypted form for re-identification.

This system is also capable of de-identifying the clinical data where the de-identification is defined in the context of research goals and objectives [2].

1.1 Prerequisite terms and definition

1.1.1 DICOM

Digital Imaging and Communications in Medicine (DICOM) is an international medical standard that specifies a data interchange protocol for medical images and their associated information across all fields of medicine [3].

1.1.2 RTSTRUCT

A radiotherapy structure set (RTSTRUCT) object of the DICOM standard is used to transmit patient structures and related data across the treatment planning system (TPS). It primarily contains the regions of interest (ROIs) and points of interest (POIs) associated with one/multiple radiological studies.

1.1.3 RTPLAN

In DICOM, a radiotherapy plan (RTPLAN) object contains geometric and dosimetric data specifying a course of the external beam or brachytherapy treatment. An RTPLAN object can be generated by manually entering data from the TPS. It usually keeps the references of an RTSTRUCT to define a coordinate system and set of patient structures.

1.1.4 RTDOSE

The radiotherapy dose (RTDOSE) object of the DICOM standard is used for transferring the dose distributions calculated by radiation therapy TPS. The dose distributions in an RTDOSE may be presented as 2D or 3D grids.

1.1.5 De-identification

De-identification is a process of detecting the patients' personal identifiers and removing or modifying those identifiers from the data.

2 Scope and Utilization

Analysis of radiological data, particularly in radiation oncology, has many challenges in the imaging field, clinical research, technology, and computation. **"Radiomics"** involves the high-

throughput extraction of quantitative imaging features with the intent of creating mineable databases from radiological images [4]. CHAVI - CompreHensive Digital ArchiVe of Cancer Imaging is the first imaging data bank focussed on providing an open access system to allow researchers globally to access radiological imaging data of patients with cancer following FAIR principles. The CHAVID3S system ensures that any data uploaded into the CHAVI system is appropriately de-identified.

3   Installation

The CHAVID3S can be installed and run on both Windows and Linux operating systems.

3.1   Requirements

- Jdk 1.8 or higher

- MySQL 5.7 or higher

- dcm4che library

- Netbeans Integrated development environment (IDE) 8.2 or higher

3.2   Installation Process

Step 1.   Clone the CHAVID3S from GitHub repository.

Step 2.   Create a database in MySQL and import the **"chavid3s.sql"** file.

Step 3.   Configure the **"config.json"** file. Kindly refer to Table 1 for more details.

  Note: Please generate the encryption string for your MySQL database password. The default encryption password is generated for string "1234".

Step 4.   Open the project using Netbeans IDE.

Step 5.   Apply **"Clean and build"** project.

Step 6.   Click on **"Run"** button to execute the application.
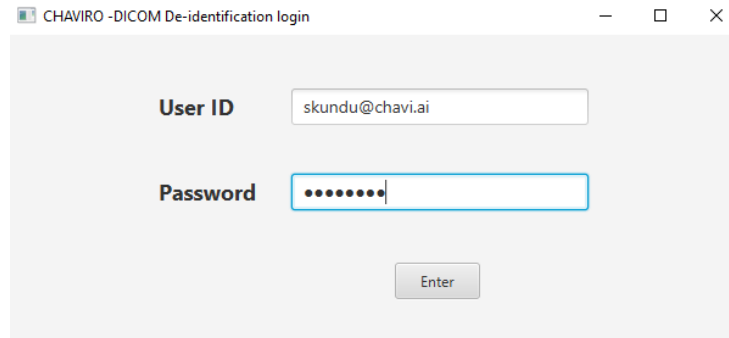
4   User Roles

In order to access the CHAVID3S, a user needs to have a login credential. The CHAVID3S has role-based control to restrict the access limit of different modules. There are two types of roles **users** and **administrators**. The user can log in to the system using the User ID and Password as shown in Figure 1.

4.1   Administrator

An Administrator has regulatory control over the system to manage the users' accounts and project creation. In this context, a project may be defined according to the type of cancer found at a certain anatomical location.

Configuration File Details (config.json)		
Key Name	**Definition**	**Default Value**
dbName	Database name	*chaviro*
dbURL	Database URL with port number	*jdbc:mysql://localhost:3306/*
dbUsername	Database user name	*root*
dbPassword	Database password	*xixjm4WhSMFQVBkaVl59oA==* *Encrypted form of "1234"*
dbDriver	Using JDBC database driver	*com.mysql.jdbc.Driver*
deidentifiedDCMSource	De-identified file directory	*/deidentifiedDICOM/*
globalPath	Full path where the CHAVID3S is located	*C:/Users/"username"/Documents/*
tempDirectory	Set a directory for storing the temporary files	*C:/Users/"username"/Documents/ NetBeansProjects/DICOMDIS/temp/*
globalID	It is unique identification for a CAHVID3S	*2013031112*
encryptionKey	This encryption key is used across the system for encryption/decryption.	*CHAVI-RO@S.KUNDU*
patientIDFormat	It is used if there is a specific format of original patient id	*MR/00/000000*

Table 1: CHAVID3S configuration file details

Figure 1: Login Page

4.1.1 User Management

As shown in Figure 2, this interface is used to create the users' accounts. The administrator will provide the required information following Name, Email-ID, Contact Number, Password, and user role. The email ID will work as a login id to access the CHAVID3S. The administrator has the control to deactivate and activate the users' accounts by clicking the corresponding button as shown in Figure 2.

4.2 Users

A user is primarily assigned for the de-identification of both DICOM images and clinical datasets. The GUI of a user is shown in Figure 4. The user has access to the following modules

- De-identify the DICOM images by clicking on the **Start De-identification Process** button.

3

Figure 2: User account management Interface

Figure 3: Project creation interface

- Clinical data can be de-identified by selecting the project name followed by choosing the clinical data (CSV format) from the **De-identify clinical data** section at right-bottom. Then the user has to click on the **De-identify EHR data** button.

- The de-identified files can be moved to any external/internal media drive. The user has to select the de-identified folder from the **De-identified file list**. Then the target drive needs to be selected from the **select drive** dropdown followed by clicking the **Move** button from

the **Move de-identified DICOM** section at the top-right.

- Users are advised to change the password after the first login. The password-management module is available at the top as shown in Figure 4.
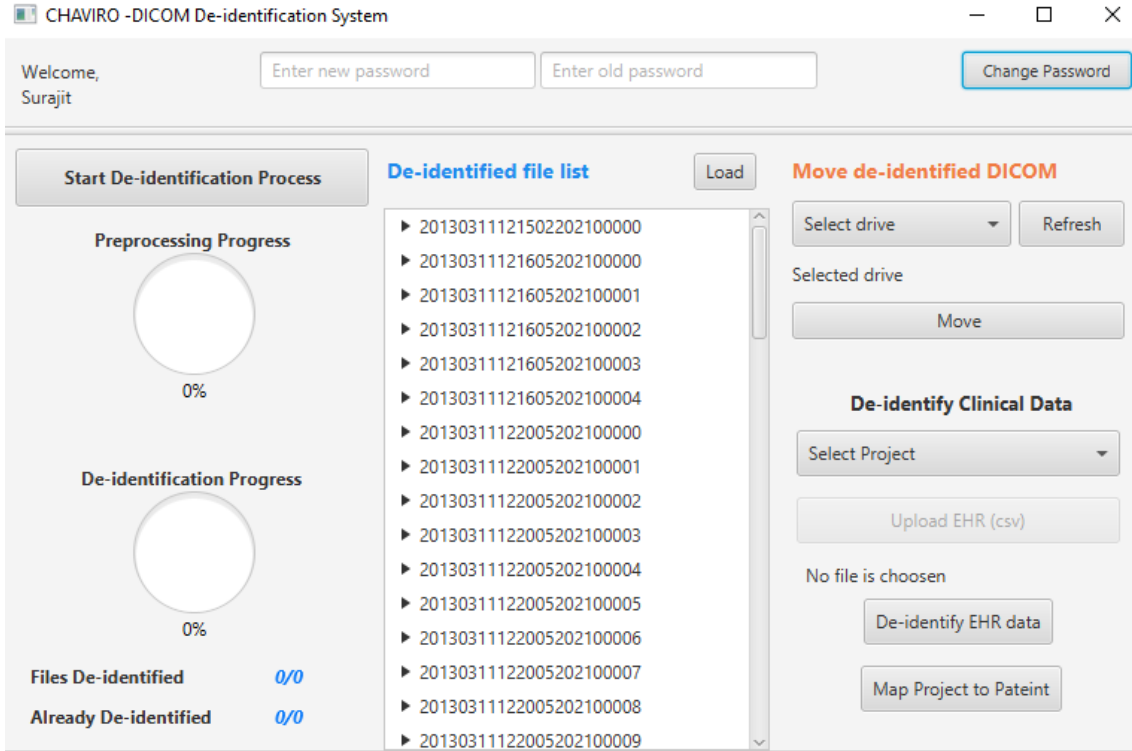

Figure 4: User interface for de-identification

5   DICOM De-identification Process

The CHAVID3S supports the de-identification of single DICOM or multiple files in compressed zip format. The DICOM de-identification interface is shown in Figure 5. Please refer to Figure 5 for the step-by-step process of DICOM de-identification.

- **Step 1.**   Select a single DICOM or multiple in **zip** format by clicking the **"Choose File (ZIP format)"** button.

- **Step 2.**   Choose the anatomic site from the **"Anatomic Site"** dropdown list.

- **Step 3.**   Pick the laterality of the anatomical site from the **"Side"** dropdown list.

- **Step 4.**   Select the radiological study type from the **"Choose Image Type"** dropdown list.

- **Step 5.**   Click on the **"SET"** button to check whether the valid data is provided or not. If everything is okay, **"De-identify"** button will be activated, otherwise, the system will guide the user by displaying the proper message.

- **Step 6.**   Click on the **"De-identify"** button to de-identify the DICOM dataset.

Once the de-identification is started, the progress status can be viewed from the same interface as shown in Figure 6.
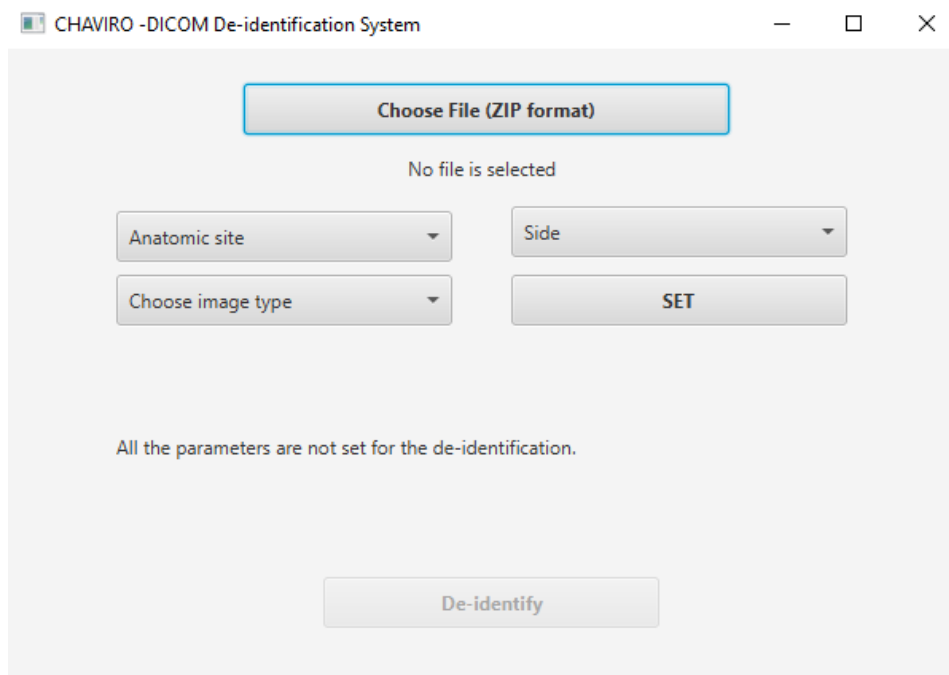
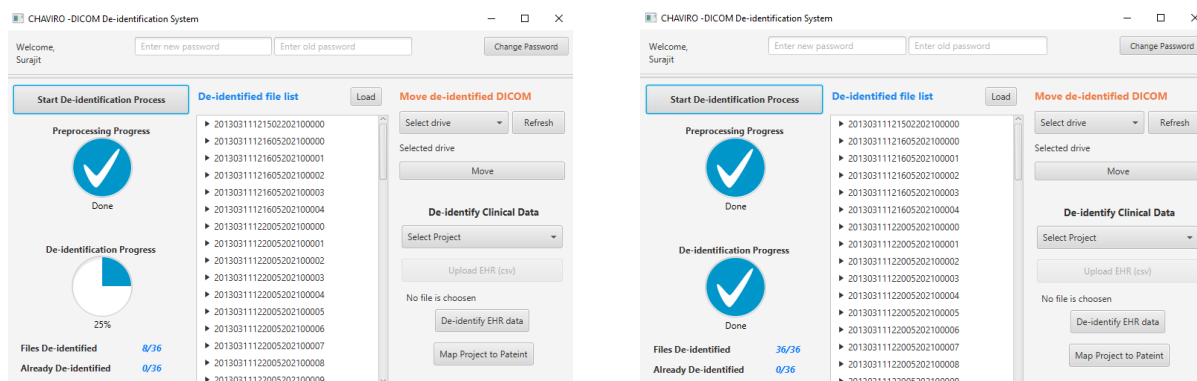**Figure 5:** DICOM de-identification interface

Figure 6: De-identification progress status

6   Clinical Data de-identification Process

We follow a specific format while acquiring the original clinical data. The template is provided in **"xls"** format for four projects. The original clinical data needs to be stored in a CSV file. Please refer to Figure 4 for the step-by-step process of clinical data de-identification.

    Step 1.   Select the project from **"De-identify Clinical Data"** section at bottom-right.

    Step 2.   Choose the original clinical data file by clicking on **"Upload EHR (csv)"** button.

    Step 3.   Click on the **"De-identify EHR Data"** button to execute the clinical data de-identification process.

Disclaimer

Neither the CHAVI community nor any of its members shall be liable for any direct, indirect, incidental, special, exemplary, or consequential damages.

References

[1] S. Kundu, S. Chakraborty, S. Chatterjee, S. Das, R. B. Achari, J. Mukhopadhyay, P. P. Das, I. Mallick, M. Arunsingh, T. Bhattacharyyaa *et al.*, "De-identification of radiomics data retaining longitudinal temporal information," *Journal of medical systems*, vol. 44, no. 5, pp. 1–15, 2020.

[2] S. Kundu, S. Chakraborty, J. Mukhopadhyay, S. Das, S. Chatterjee, R. Basu Achari, I. Mallick, P. Pratim Das, M. Arunsingh, T. Bhattacharyya *et al.*, "Research goal-driven data model and harmonization for de-identifying patient data in radiomics," *Journal of Digital Imaging*, vol. 34, no. 4, pp. 986–1004, 2021.

[3] P. Mildenberger, M. Eichelberg, and E. Martin, "Introduction to the dicom standard," *European radiology*, vol. 12, no. 4, pp. 920–927, 2002.

[4] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. Van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker *et al.*, "Radiomics: extracting more information from medical images using advanced feature analysis," *European journal of cancer*, vol. 48, no. 4, pp. 441–446, 2012.