MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220271

ODIASP: Clinically Contextualized Image Analysis Using the PREDIMED Clinical Data Warehouse, Towards a Better Diagnosis of Sarcopenia

Katia Charrière[a], Pierre-Ephrem Madiot[b], Svetlana Artemova[ac], Pungponhavoan Tep[a], Christian Lenne[a] Brigitte Cohard[b], Alban Caporossi[acf], Isabelle Boudry[a], Juliette Meyzenc[a], Gilbert Ferretti[e], Ivan Bricault[aef], Joris Giai[acf], Jean-Luc Bosson[acf], Eric Fontaine[dg], Cécile Bétry[df], Alexandre Moreau-Gaudry[acf]

[a] Clinical Investigation Center-Technological Innovation, Univ. Grenoble Alpes, INSERM CIC1406, CHU Grenoble Alpes, F-38000, Grenoble, France
[b] Digital Services Management, CHU Grenoble Alpes, F-38000, Grenoble, France
[c] Public Health Department, CHU Grenoble Alpes, F-38000, Grenoble, France
[d] Artificial Nutrition Clinic, CHU Grenoble Alpes, F-38000, Grenoble, France
[e] Department of Radiology and Medical Imaging, CHU Grenoble Alpes, F-38000, Grenoble, France
[f] TIMC, Univ. Grenoble Alpes, CNRS, VetAgro'Sup, CHU Grenoble Alpes, Grenoble INP, F-38000, Grenoble, France
[g] LBFA, Univ. Grenoble Alpes, F-38000, Grenoble, France

Abstract

Big Data and Deep Learning approaches offer new opportunities for medical data analysis. With these technologies, PREDIMED, the clinical data warehouse of Grenoble Alps University Hospital, sets up first clinical studies on retrospective data. In particular, ODIASP study, aims to develop and evaluate deep learning-based tools for automatic sarcopenia diagnosis, while using data collected via PREDIMED, in particular, medical images. Here we describe a methodology of data preparation for a clinical study via PREDIMED.

Keywords:

Clinical data warehouse, Deep Learning, Big Data, Sarcopenia, Artificial Intelligence, Nutrition

Introduction

Clinical data warehouses (CDW) allow the collection of clinical data by gathering complex, heterogeneous and generally unstructured medical information produced for care in a unique and powerful technical environment. The Grenoble Alps University Hospital (CHUGA) received regulatory approval to deploy its CDW called PREDIMED [1]. Thanks to the Big Data tools available on such platform, large quantities of medical data can be reused for various applications. Large cohorts of patients with contextualized data can be created, which offers new perspectives for researchers. For example, in the context of the COVID19 pandemic, usage of CDW [2] and contextualized image data [3] have demonstrated significant potential. In parallel, Deep Learning (DL) methods have shown the ability of algorithms to help diagnose pathologies using medical images with at least the same performance as medical experts [4].

Sarcopenia is defined as a combination of either or both low muscle mass and function. Although it is associated with an increased length of hospital stay and mortality [5], it is rarely diagnosed, as there is no gold standard for muscle mass assessment. Skeletal muscle index (SMI), derived from the cross-sectional muscle area (CSMA) at the mid-third lumbar vertebra (L3) level, using computed tomography (CT), can be used for sarcopenia diagnosis [6]. Nevertheless, its measurement is still difficult, if not impossible, in current daily practice due to time-consuming manual tasks. Thus, SMI is not used in clinical practice or for clinical studies on large cohorts of patients. Additionally, thresholds on SMI measurement for sarcopenia diagnosis might be questioned [7]. More work needs to be done to establish SMI sarcopenia cut-off in a large cohort of healthy subjects of all ages.

In this paper, we propose a new approach of data collection for a clinical study based on PREDIMED, combining two sets of data: 1) dataset of images, clinically contextualized for training and evaluation of a DL-based tool performing a fast and robust automatic CSMA measurement; 2) large cohort of patients with SMI score, cross-referenced with clinical data, for new sarcopenia markers estimation.

Methods

First, necessary clinical study data should be made available in the PREDIMED architecture (Fig 1) consisting of a virtualized platform hosting a CEPH distributed data storage. Four containers are deployed on the virtualized platform. Text and structured data tools are hosted on two dedicated virtual machines (predimed-text and predimed-sql). Another container hosts various shared tools (predimed-tools). For medical images, which represent a large volume of unstructured clinical data, a specific data management has been developed based on a virtual host (predimed-imagerie), which carries three containers. An Orthanc [8] server (Orthanc VM1) is responsible for the data import from the Clinical facility's PACS leveraging DICOM protocol. A dynamic network bandwidth control is put in place on this container in response to the dual objective 1) control and limit network and performance impact adjusted on the current care usage and 2) incremental buildup of historical data over time. A second Orthanc server (Orthanc VM2) provides for unlimited traffic access for PREDIMED users. Both Orthanc containers share the same PostgreSQL database deployed on CEPH, providing underlying needed resource scalability. Apache-Airflow orchestrator is leveraged for dynamic resource control for data import. One of the challenges of such approach was to not impact the picture archiving and communication system (PACS) used for current care while copying data to CDW.

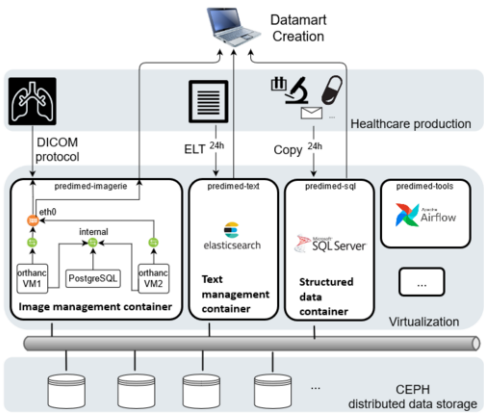

Figure 1. Architecture of the PREDIMED clinical data warehouse

Second, if the data necessary for the clinical study is available in PREDIMED, which is the case for ODIASP, data usage for a study should be validated by PREDIMED governance: the steering, scientific and ethics committees examine the project, give their approval, and inform medical division directors using the data produced in the corresponding medical units. The PREDIMED data processing team then identifies patients that can be included in the study and collect their administrative data for patients' individual information. When the selected patients are informed and agree to the usage of their data, patients' data required for the study (datamart) can be isolated and de-identified. De-identification of images is performed by removing all identifying DICOM tags and dose reports (often contain personal information). An exporting committee must then validate the datamart generated before its transmission for data annotation, DL algorithms development, and data analysis.

In ODIASP, the gathering of 680 CT scans with the third lumbar vertebra, with the corresponding patients' heights, is firstly required to develop and evaluate the DL-based tool. De-identified images are stored on a secure server, with access for medical experts to perform manual annotation (L3 position and muscular surface estimation) required for training and evaluation of DL models. Secondly, a cohort of about 3000 CT scans with automatically-computed SMI, combined with other medical information of interest (laboratory tests results, medical history, etc.), will be created to identify new markers of sarcopenia.

Results

For ODIASP clinical study, PREDIMED data allowed for the automatic identification of 3156 patients meeting the study criteria and automatic generation of information letters. As a result, 2907 patients were informed (249 were deceased). One hundred seventeen patients were finally excluded, and 111 were not willing to consent to the usage of their data. Data collection started for a subset of 680 patients required to develop and evaluate the DL-based tool. Patients' heights of 654 of them were automatically collected by PREDIMED automatic data exploration. Other data sources still need to be explored to find the 26 missing heights. CT scans are being collected on the flow, as massive retrieving of patient images could impact everyday care. At this time, 212 CT scans were collected, de-identified, and made available to clinicians for manual annotation.

Conclusions

The PREDIMED CDW is showing with ODIASP its ability to create a large dataset of clinically contextualized images, crucial for neural networks' training.

References

[1] S. Artemova et al., «PREDIMED: Clinical Data Warehouse of Grenoble Alpes University Hospital.,» *Studies in health technology and informatics,* vol. 264, p. 1421–1422, 2019.

[2] C. Apra et al., «Predictive usefulness of PCR testing in different patterns of Covid-19 symptomatology-Analysis of a French cohort of 12,810 outpatients,» *medRxiv,* 2020.

[3] W. Liang et al., «Early triage of critically ill COVID-19 patients using deep learning,» *Nature communications,* vol. 11, p. 1–7, 2020.

[4] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean et R. Socher, «Deep learning-enabled medical computer vision,» *npj Digital Medicine,* vol. 4, p. 1–9, 2021.

[5] A. S. Sousa, R. S. Guerra, I. Fonseca, F. Pichel et T. F. Amaral, «Sarcopenia and length of hospital stay,» *European journal of clinical nutrition,* vol. 70, p. 595–601, 2016.

[6] B. Amini, S. P. Boyle, R. D. Boutin et L. Lenchik, «Approaches to assessment of muscle mass and myosteatosis on computed tomography: a systematic review,» *The Journals of Gerontology: Series A,* vol. 74, p. 1671–1678, 2019.

[7] L. Caudron, A. Bussy, S. Artemova, K. Charrière, A. Moreau-Gaudry, J.-L. Bosson, G. Ferretti, E. Fontaine et C. Bétry, «Sarcopenia diagnosis: comparison of automated with manual computed tomography segmentation in clinical routine,» *Journal of Cachexia, Sarcopenia and Muscle Rapid Communications,* 2021.

[8] S. Jodogne, «The Orthanc ecosystem for medical imaging,» *Journal of digital imaging,* vol. 31, p. 341–352, 2018.

Address for correspondence

Katia Charrière, KCharriere@chu-grenoble.fr.