Development of a HIPAA-compliant synchronizer to Enable Automatic Reporting and Inter-institutional Collaboration in Healthcare Systems Worldwide.

Fernando Yepes-Calderon[1,*], J. Gordon McComb[2]

[1]Science Based Platforms LLC, 604 Beach CT, Florida, 34950
[2]Children Hospital Los Angeles, Los Angeles, United States

Abstract
Having analytical capabilities inside clinics is an obeyed step forward since most diagnosing data is currently created in a format suitable to computer processing. The advantages of having the machines working on the extraction of numbers benefit all parts of the healthcare pipeline, fastening verdict generation, broadening the use of stored data, cross-correlating the multiple sources of information, and profiting from theoretically unlimited quantifications, which are the core for creating evidence-based verdicts, fast population analysis, more accurate diagnosis, enforcement of patients' adherence to treatment and interinstitutional cooperation envisaging the use of big data and artificial intelligence. Nevertheless, such a platform requires integrating several subsystems that developers still need to conceive to interact with each other. Additionally, the automation must comply with confidentiality regulations; therefore, security breaches are open while processing the data. The presented work involves the development of an architecture capable of allocating an unlimited number of algorithms inside healthcare facilities, which provides quantifications without perturbing the operation of the transport utilities that are already working. Moreover, the presented design can return numbers to a centralizer that keeps a simple yet robust protocol to assert inter-institutional cooperation around big data and artificial intelligence implementations.

Keywords
Artificial intelligence, PACS, Medical Imaging, HIPAA, Big Data

1. Introduction

With the advent of artificial intelligence, the data stored in healthcare facilities suddenly gained unexpected relevance. Hospitals, clinics, and healthcare administrators initially obeyed to save patient records by law [1], now recognize the value of the experience inherently attached to the data repositories since those records are precursors for machine-operated services. The new technological platform can speed up diagnosis, enhance accuracy, and improve patient adherence to treatment [2, 3]. Mid-term benefits encompass data centralization to constantly monitor infectious agents, population analysis, and better budgetary programming even within

ICAIW 2023: Workshops at the 6th International Conference on Applied Informatics 2023, October 26–28, 2023, Guayaquil, Ecuador
fernando.yepes@strategicbp.net (F. Yepes-Calderon); GMcComb@chla.usc.edu (J. G. McComb)
https://www.fernandoyepesc.com (F. Yepes-Calderon)
0000-0001-9184-787X (F. Yepes-Calderon)

universal health care systems. The utter purpose aims to exploit the experience attained from stored data to replace the medical practice's costly curative and preventive methods with prediction-based strategies [4, 5].

Recent Artificial Intelligence (AI) implementations in medicine and other high-impact fields have created an optimistic atmosphere around many possibilities that might emerge when machines execute tasks [6, 7]. Nevertheless, integrating cognitively capable tools into existing medical networks is challenging [8]. From a telecommunications perspective, medical data is protected by laws and strict regulations like the Health Insurance Portability and Accountability Act (HIPAA); therefore, technological utilities for data transport and administration – although running over the unsecured TCP/IP protocols – have been patched to provide the required security as in the Picture Archiving and Communication System (PACS) [9]. Up to here, having access to databases remains a technical issue that one can solve with simple actions such as opening a port to automation - a routine task within TCP/IP networks -. However, regulations impose a real difficulty. The data, raw or derived from original records, should not lead to patient identification. Recall that PACS or any other transport-specialized utility in hospitals and clinics move DICOM formats in which each package carries confidential information in the header [10]. Thus, any automation must deal with inherent TCP/IP security flaws, avoid human interaction in the pipeline's intermediate steps, and anonymize the data right after being gathered [11]. From an operational perspective, the data must flow in virtually closed channels, equivalent to saying that only assigned specialists and involved patients or relatives should have access to medical records per case. In those facilities where transport software such as PACS is in use, mistakenly moving data from one channel to another would lead to record incoherence and general failure. Another operational challenge is the need for algorithm repositories within hospitals and clinics. Even in the most sophisticated healthcare facilities, specialists generate verdicts with relatively few assisting quantifications performed by certified automatic tools [12, 13].

Once these difficulties are solved, developers might want to share numbers among several institutions, a task that requires administration and management input. Interinstitutional integration is desirable because it spreads the usability of AI to more extensive geographical regions [14], and generalizations in knowledge and verdicts can consider a bigger formulation space, yet with a feature space fixed by the certified implementations.

This manuscript presents the design and implementation of a data architecture integrating access to standard transport managers and algorithms to share the resulting quantifications with a data centralizer. The architecture automates the extraction of numbers that describe pathologies' verdicts and treatment-tracking responses inside the institutions and shares labeled numbers with the centralizer, where a higher hierarchy implementation runs population analysis to provide a myriad of new AI-based services to society.

This document is organized as follows: the materials and methods sections present the technicals behind the solution implemented to assert querying flexibility and, thus, enable the capability to retrieve data from any repository. Then, we explain how the final users can program jobs using smartphones or web interfaces by perfecting their querying interfaces and defining the algorithms for the programmed jobs. Finally, we present the schematics for the architecture capable of reading and executing the jobs while complying with confidentiality regulations. The materials and methods section ends with two detailed proofs of concepts where we produce analytics using the proposed architecture. In the results section, we display interface prototypes of the syncing utility since those views are the highest-level component of a solution for the problems that motivated this work. The results section also provides evidence for the two proofs of concept listed in the materials and methods section. Finally, we discuss the inclusion of analytics into medical pipelines and further work. In the conclusion, we concisely resume the more striking contributions of this work.

2. Materials and methods

The synchronizing interface will serve as a referee by signaling the functionalities required to provide analytics and automatic reporting inside clinics and hospitals. The proposed architecture should comply with protocols of specific appliances and be flexible enough to keep operative regardless of infrastructure size or running applications. Moreover, the design should consider adding new data sources, new sharing numerical data schemes, and new reporting templates without affecting the operation of the already established services. The syncing interface should provide redundant security mechanisms since this device is physically visible and accessible within hospital premises. Access to the configuration methods of this device would represent a security breach that might render void its usability under HIPAA terms [15].

2.1. Querying orders in a multi-user, multi-institution, multi-database environment

The presented architecture allows healthcare providers to schedule data processing tasks easily. The pipeline starts when a user's smartphone synchronizes with the Evalu@ service (E@) [16]. The syncing happens whenever the fingerprint, face, or login/pass credentials are correctly presented in E@'s native smartphone interfaces. Once authenticated, the user can navigate through hospitals and query instruments. Through the forms displayed in the query instruments, users create filtering definitions. Whenever a user fulfills a formulary and feds E@, a new data processing order is saved in the system. The institution and data repository identifiers are saved in variables cid and eitid. At the same time, the software builds the data filtering string with the responses captured by the querying instrument pointed with variable eid. Within the form, the authorized user must define the algorithm or set of algorithms to process the filtered data, and the system will propagate that selection in the string array aid. The designed protocol and signaling allow multiple institutions to be handled without the risk of merging orders or forwarding results to the incorrect institution or person.

Since the E@-Med – the local interface – initiates the communication with the E@ service, there is no space for security breaches, and keeping the link alive with distinct institutions remains under local control.

2.2. Accessing repositories in a local domain

Regardless of network security and complexity, hospitals and clinics rely upon TCP/IP for data transport. Applications such as File Transfer Protocol (FTP), its secure version SFTP, Pop3 or IMAP for emailing, or the specialized PACS are extensions of the platforms supporting the

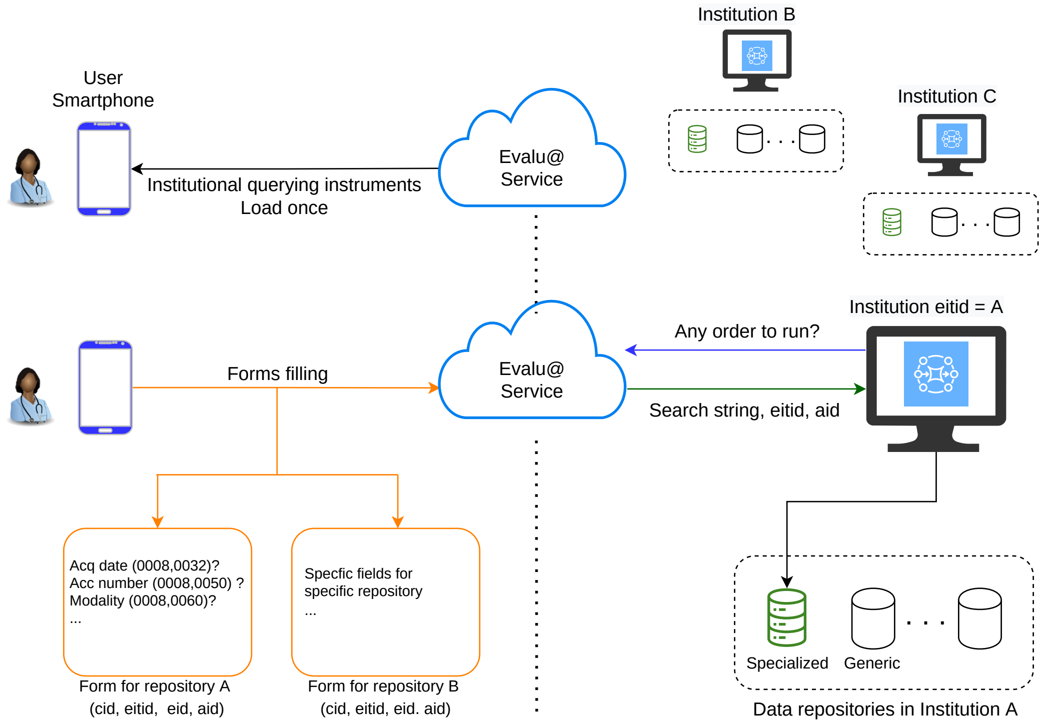

Figure 1: Creating and ordering online that will be executed by E@-Med interfaces working in local environments. The system provides identifiers for the institution (cid), the repository to be used (eitid), values to construct the querying string (eid), and the used algorithms (aid).

internet. Therefore, accessing a local or remote repository can be generalized as shown in **Figure** 2. In general, a user willing to retrieve data – human or robot – will need to provide the IP address of the machine where the information resides, namely the "host," a username, and a password. The host is an IP address or a human-readable name that will be translated to its correspondent IP address in networks with active domain name servers (DNS); otherwise, the IP address will be enough to activate the Address Resolution Protocol (ARP) that completes the TCP/IP requirements to establish a communication between computers in share media. At this moment, the syncing interface (E@+Med) will deliver the user and passwords through the recently established channel, and, in case access is granted, the syncing interface will need to provide the querying string that varies in extent and complexity depending on the repository or a path inside the host in less complex transport applications [17].

2.3. Data processing

The aid variable holds a unique identifier for tested and approved algorithms, in which docked executables are uploaded to E@ as in an application repository such as Android or Apple stores. Then, E@ shares those docked codes with every registered E@-Med, making the software solutions available to all users with accounts in at least one of the collaborating institutions. The architecture warrants data processing inside the institutions within a protected domain.

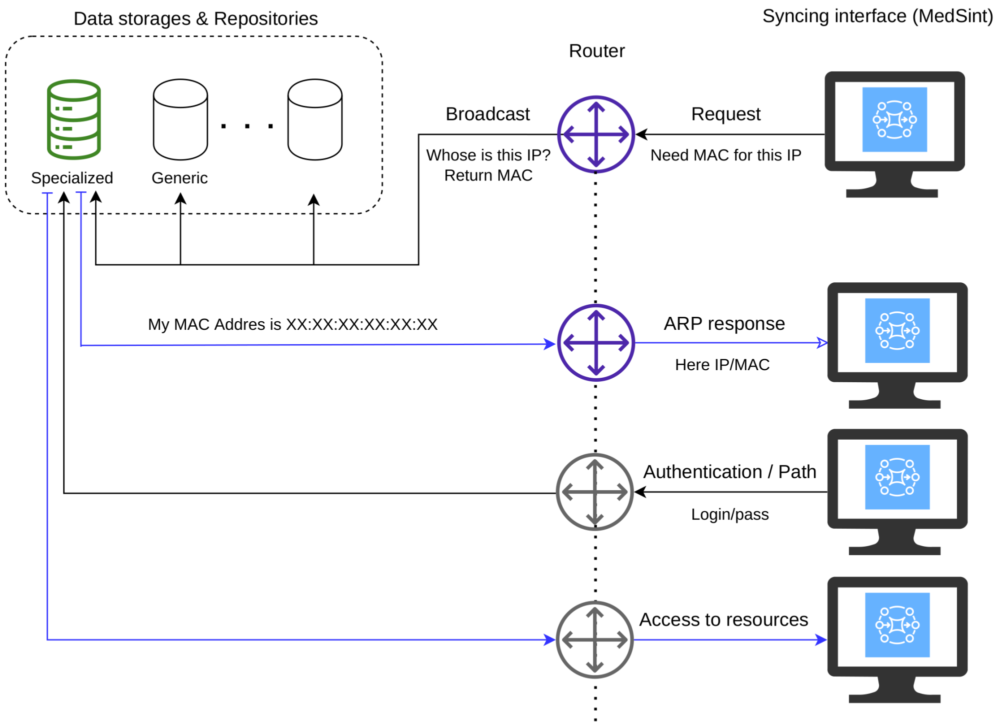

Figure 2: Generalization of virtual channel creation in TCP/IP networks. Once the channel is established, the E@-Med should use the protocol and provide the requirements to access different repositories.

When users define the querying string to filter data, they also select the algorithm(s) for processing. The algorithm developer defines the deliverables of each solution, and users accept the implementation as it is. With the querying string, the E@-Med brings the data, decrypts it, anonymizes it, and makes it available in a temporal directory, as we presented in [18]. Then, it calls the selected algorithms and waits for the outputs listed in a JSON file residing in the same folder where the docked code is. The E@-Med uses the algorithm description saved as a JSON file to define whether the resulting reports should be sent to the referring user by institutional email, back to the original repository, or both. Another mandatory section of the reporting JSON defines the API parameters and numerical variables that will be shared with the data centralizer E@.

2.4. Proofs of concept

2.4.1. Supporting Institutional Review Board (IRB) tasks

Only moving the data and making it available for research endeavors is one of the most solicited applications. The data should be anonymized and converted to a format that is easy to transport. The most critical case happens when DICOM data are required after screening for a specific pathology, patient's age, gender, or treatment. The user can also set aspects such as the number of studies and the range of creation dates. As explained in **Section** 2.2, each repository defines

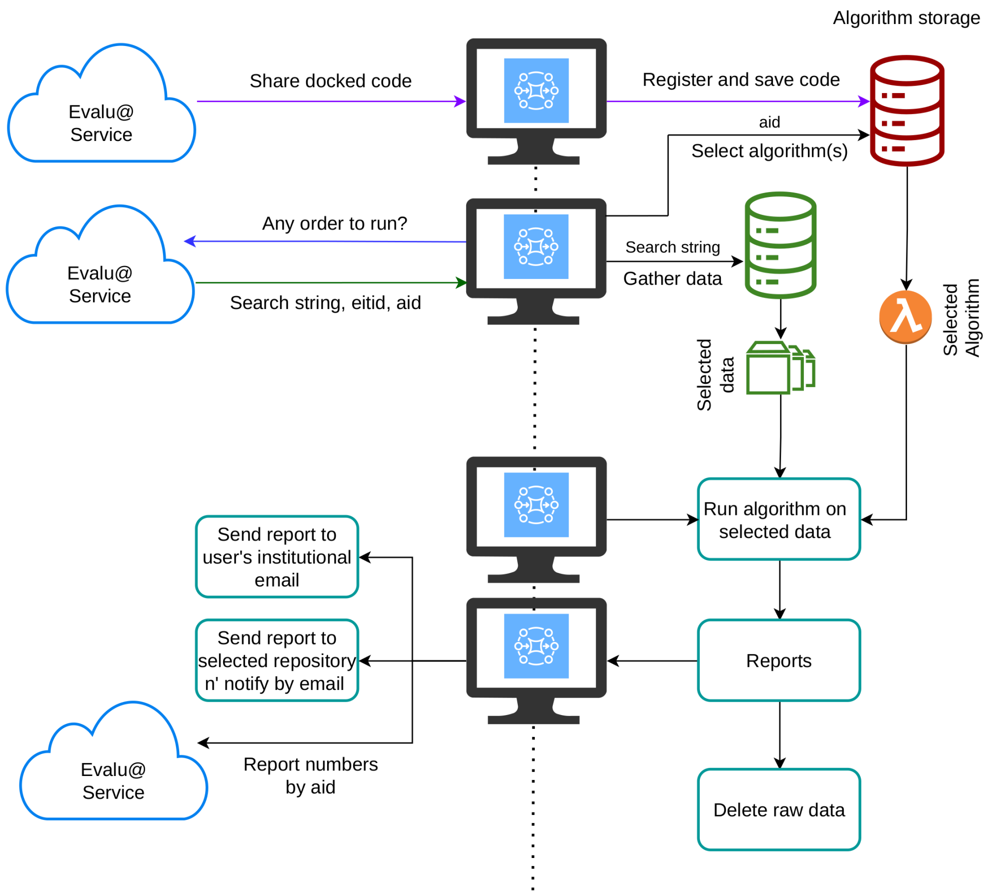

Figure 3: Test and approved algorithms are shared and registered system-wide with unique identifiers. The E@-Med application makes registered algorithms available, and healthcare personnel can post orders involving the new functionalities. After deleting the raw data gathered, the syncing interface will execute the orders and send the generated materials to institutional emails or back to the querying repositories. Additionally, some numbers can be sent back to Evalu@ to enable artificial intelligence applications with regional pertinence, personalize services to individuals, and serve prediction capabilities to adjust budgets.

the fields needed to create an order. For this purpose, a user employs the string query creation functionality to filter the data. What follows for this particular application does not need data processing; instead, it will request an action from an IRB member, which consists in authorizing the IRB study to be delivered to the requesting user. The system will provide emails to inform process flagging and availability.

2.4.2. Ventricular volume before and after Shunting

One common problem in radiology and neurosurgery units is monitoring the correct operation of installed CSF draining valves. Young hydrocephalus patients often receive an MRI before and after surgically placing the device. The medical images are taken at different moments, and the MRI technician defines imaging parameters to proceed with the critical constraint of favoring the patient's comfort, thus producing the images in the shortest possible time. This proof of concept shows the results of an algorithm that reports the differences in ventricular volumes between studies. The Querying string is specific about grabbing FLAIR T2 images produced in a range of rates over one patient id; therefore, the algorithm will try to segment the lateral ventricles in all found volumes and report them with the study acquisition date. Additionally, the final report displays a 3D reconstruction of the segmented structures with the read volumes in mm3.

3. Results

3.1. Syncing interface administration

The syncing interface displays eight environments to all authenticated users, except for the super-administrator, who will see the ambient to manage users. The signing-in process includes a tokenized key sent to the institutional email and a fingerprint reading, such as security robustness to respond to HIPAA regulations. The interfaces shown in **Figure 4** correspond to different displays where the administrators control the available repositories.

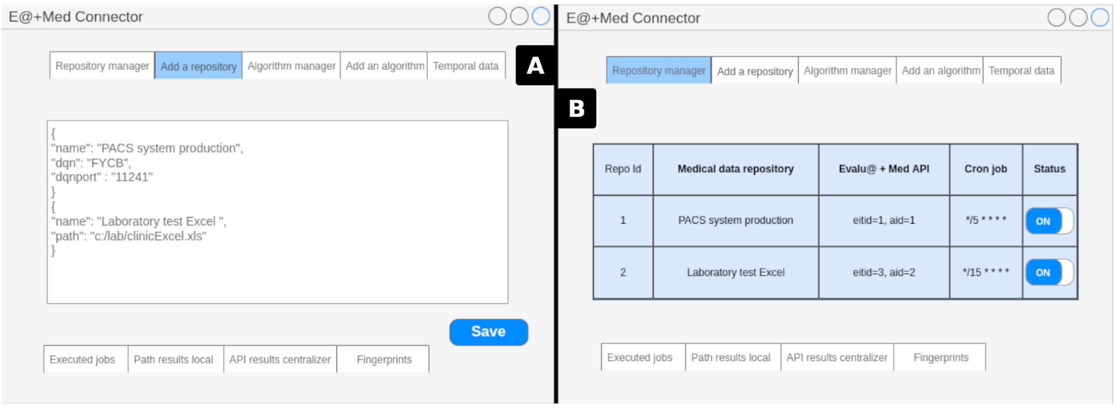

Figure 4: Synchronizing interface views. In panel A, a part of JSON structures connects to different repositories. In panel B, the administrator can turn on and off the configured repositories. The Cron job tells the E@-Med interface when it should attempt to connect Evalu@ for programmed orders.

The add repository interface allows the administrator to define unlimited databases. There is no fixed template to explain the connection parameters. Instead, the administrator is encouraged to provide sufficient information in the form key:value to have successful connections, and this flexibility also supports further changes to repositories' security. The E@-Med interface will use both keys and values to connect the storage as shown in **Figure 2**.

In **Figure 5**, panel A shows the superficial functionality for repositories. From here, an administrator can turn the access to algorithms on and off, and this action will be reflected

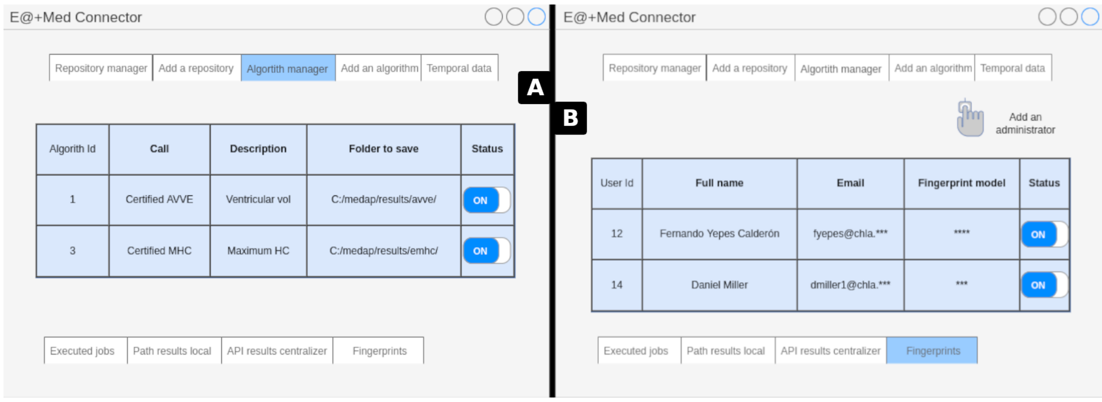

Figure 5: The syncing interface. In panel A, the algorithms manager from which the administrators make them available system-wide. In panel B, an administrative interface that is available only for super-administrators.

system-wide. To add an algorithm, the administrator should use the add algorithm function (screenshot not shown).

Panel B of **Figure** 5, presents the only screen for the exclusive use of super-administrators. From here, the authorized user can give access to the critical functions of E@-Med to second-level administrators after registering their fingerprints; therefore, someone will always be available in the hospital's facilities to control the system's behavior in local environments.

3.2. POC 1: Clinical data available for research purposes

After the IRB has authorized the use of clinical data, the files are delivered without encryption, anonymized, and in Nifti format. The new container concatenates the individual DICOM files and deletes the DICOM headers; therefore, the new files occupy less space in the disk and are easier to transport. **Figure** 6 displays a query run on the command line as the E@-Med would do with the configured Cron job. The presented connection involves a PACS system, and a video of this interaction is available at www.fernandoyepesc.com..

3.3. POC 2: Ventricular volume before and after Shunting

An algorithm based on the Automatic Ventricular Volume Estimator AVVE runs on MRI volumes of the same subject. The querying string used a patient number and study creation date constraints. Therefore, for the specific case, two volumes are compared. The Data gathering process occurs in the same manner presented in **Section** 3.2. After that, the system calls the modified version of the AVVE and provides the gathered volumes – two volumes due to the date constraint – for processing. The system creates the report shown in **Figure** 7, which can be dicomized and sent back to PACS. The processing also includes returning Evalu@ the patient's age in days and the ventricular volumes before and after the shunt procedure.

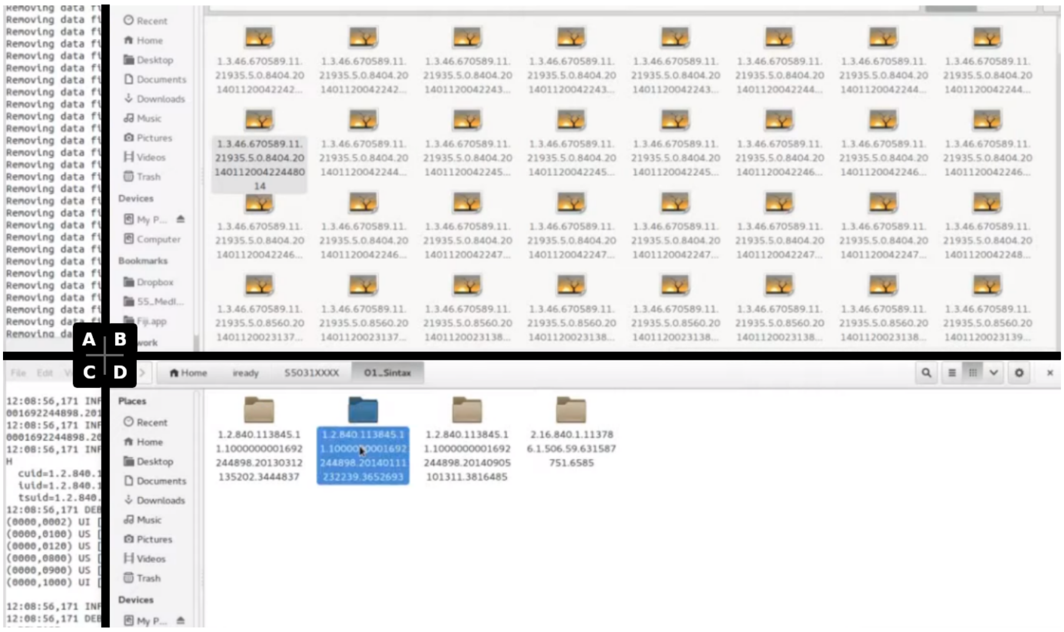

Figure 6: Displaying a background process that connects the PACS in a production environment. Panels A and C contain confidential information and are consoles where the code reports the channel status and actions (Panel A) and details about what is being moved (Panel C). Panes B and D are local directories showing the raw data as it arrives from PACS (Panel B) and the containers where the decrypted, anonymized, and Nifti formatted data reside temporarily before being sent to the requesting user (Panel D). The information in Panel B will be deleted right after Panel D is populated.

4. Discussion

Authors producing thousands of methods developed year by year in research units will not have the opportunity to implement their creations in a production environment. Such a situation precludes the use of technology to benefit many aspects of current healthcare methods. The transport utilities inside hospitals – as specialized as PACS or simple FTP-based ones – could be adjusted to move the data to temporal storages where algorithms perform diverse quantifications. With such a level of Generalization, proposals such as integrating the Radiology Information Systems (RIS) and PACS to complement the storing and data-sharing capabilities and provide this new platform with image processing capabilities [19, 20], become a specific use of the proposed architecture. Additionally, working on these transport applications with no view to interinstitutional collaboration forces the proliferation of developments with few or no probabilities of unification. Such a scheme precludes the golden opportunity of collecting tons of data from particular abnormalities with context pertinence that artificial intelligence experts highly appreciate.

Instead, the proposed architecture covers the need for analytics in hospitals and clinics where algorithms are used globally. The authors can test their creations in real arenas and gain prestige. When an algorithm is valuable and infallible, healthcare institutions will use it to obtain all the benefits that several authors have agreed on regarding the need for analytics within hospitals

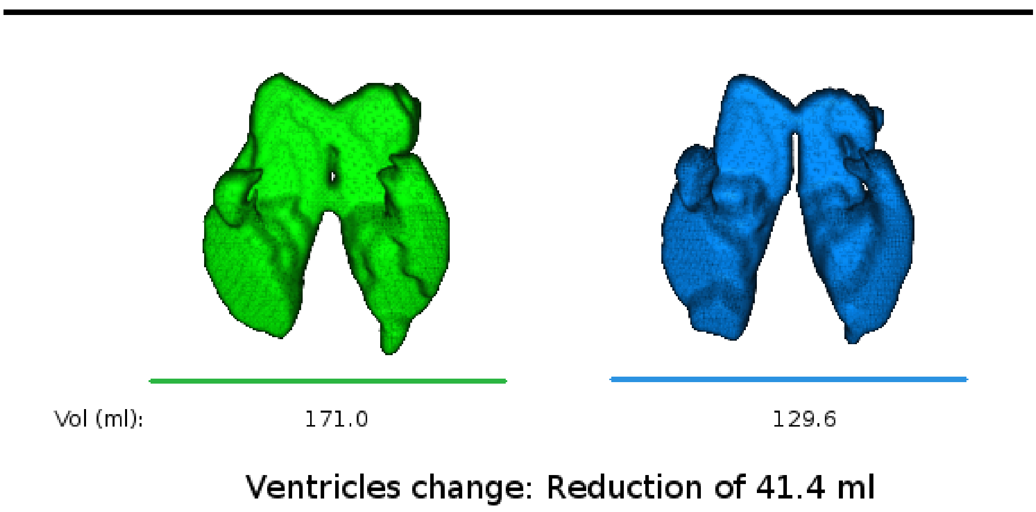

Figure 7: Automatically generated report on a patient with MRI studies before and after a shunt procedure. The automation considers discrepancies in resolution, scaling, rotations, and displacements to calculate the volume in the two time-separated volumes.

[21, 22]. Moreover, the architecture envisages an App store exclusive for medical applications that will collect data on all sorts of abnormalities to run AI and provide inter-institutional feedback to improve the decision capabilities of worldwide health systems.

The presented architecture integrates several communication systems and is flexible to incorporate new repositories with cutting-edge security strategies of today and further ones. The presented protocol assures that external communication does not propagate confidential information and algorithm execution happens in local facilities using local resources. Additionally, the system facilitates job programming by integrating smartphone technology.

Further developments involve the creation of a testing environment where authors can check feasibility while the platform assures the rightness of the presented codes. Also, we envisage the creation of new algorithms and the possibility of connecting them in compounded pipelines.

5. Conclusion

The presented interface is a flexible system that enables analytics inside hospitals, respecting HIPAA and implementing a protocol that allows interinstitutional collaboration to support artificial intelligence implementations. Confidentiality compliance is accomplished when the system only allows human interaction in the tasks' programming and when receiving the reports. Additionally, the proposed architecture gives local administrators the faculty to turn off the components of the system at will.

References

[1] W. H. Roach, Medical records and the law, Jones & Bartlett Learning, 2006.
[2] E. Kim, S. M. Rubinstein, K. T. Nead, A. P. Wojcieszynski, P. E. Gabriel, J. L. Warner, The evolving use of electronic health records (ehr) for research, in: Seminars in radiation oncology, volume 29, Elsevier, 2019, pp. 354–361.
[3] A. Bohr, K. Memarzadeh, The rise of artificial intelligence in healthcare applications, in: Artificial Intelligence in healthcare, Elsevier, 2020, pp. 25–60.
[4] R. B. Parikh, Z. Obermeyer, A. S. Navathe, Regulation of predictive analytics in medicine, Science 363 (2019) 810–812.
[5] S. Secinaro, D. Calandra, A. Secinaro, V. Muthurangu, P. Biancone, The role of artificial intelligence in healthcare: a structured literature review, BMC medical informatics and decision making 21 (2021) 1–23.
[6] W. L. Bi, A. Hosny, M. B. Schabath, M. L. Giger, N. J. Birkbak, A. Mehrtash, T. Allison, O. Arnaout, C. Abbosh, I. F. Dunn, et al., Artificial intelligence in cancer imaging: clinical challenges and applications, CA: a cancer journal for clinicians 69 (2019) 127–157.
[7] U. Raghavendra, U. R. Acharya, H. Adeli, Artificial intelligence techniques for automated diagnosis of neurological disorders, European neurology 82 (2020) 41–64.
[8] M. DeCamp, J. C. Tilburt, Why we cannot trust artificial intelligence in medicine, The Lancet Digital Health 1 (2019) e390.
[9] R. E. Cooke Jr, M. G. Gaeta, D. M. Kaufman, J. G. Henrici, Picture archiving and communication system, 2003. US Patent 6,574,629.
[10] O. S. Pianykh, O. S. Pianykh, DICOM Security, Springer, 2012.
[11] F. Yepes-Calderon, S. Bluml, S. Erberich, M. D. Nelson, J. G. McComb, Improving the picture archiving and communication system: towards one-click clinical quantifying applications, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 7 (2019) 154–161.
[12] H. Kaur, S. K. Wasan, Empirical study on applications of data mining techniques in healthcare, Journal of Computer science 2 (2006) 194–200.
[13] Z. Lv, L. Qiao, Analysis of healthcare big data, Future Generation Computer Systems 109 (2020) 103–110.
[14] Z. Shao, S. Yuan, Y. Wang, Institutional collaboration and competition in artificial intelligence, IEEE Access 8 (2020) 69734–69741.
[15] W. Moore, S. Frye, Review of hipaa, part 2: limitations, rights, violations, and role for the imaging technologist, Journal of nuclear medicine technology 48 (2020) 17–23.
[16] F. Yepes-Calderon, J. F. Yepes Zuluaga, G. E. Yepes Calderon, Evalu@: An agnostic web-based tool for consistent and constant evaluation used as a data gatherer for artificial intelligence implementations, in: H. Florez, M. Leon, J. M. Diaz-Nafria, S. Belli (Eds.), Applied Informatics, Springer International Publishing, Cham, 2019, pp. 73–84.
[17] D. Hercog, D. Hercog, Some application layer protocols in ip networks, Communication Protocols: Principles, Methods and Specifications (2020) 349–363.
[18] J. G. M. Fernando Yepes Calderon, Enabling the centralization of medical derived data for artificial intelligence implementations, 2020. URL: https://patents.google.com/patent/US20200273551A1, uS Patent US20200273551A1.

[19] S. S. Boochever, His/ris/pacs integration: getting to the gold standard., Radiology management 26 (2004) 16–24.
[20] L. Faggioni, E. Neri, F. Cerri, F. Turini, C. Bartolozzi, Integrating image processing in pacs, European journal of radiology 78 (2011) 210–224.
[21] L. Wang, C. A. Alexander, Big data analytics in medical engineering and healthcare: methods, advances and challenges, Journal of medical engineering & technology 44 (2020) 267–283.
[22] A. Belle, R. Thiagarajan, S. Soroushmehr, F. Navidi, D. A. Beard, K. Najarian, et al., Big data analytics in healthcare, BioMed research international 2015 (2015).