AI poWered cancer predIction, diagnosis and treatmeNt

Giampaolo Fiorentino [1], Mario Veniero[1], Ciro Romano[1]

[1] *Frontiere, Via Oslavia, 6, Rome, Italy*

Abstract

While the world is moving towards an increased adoption of AI algorithms which will increase exponentially over the next few years, the key challenge would be testing of AI-ML based systems. This is due to lack of large volumes of data to train them, including high-definition health images, as well the lack of structured or standard literature on the methodology or approach to be adopted while testing them. Two main challenges can be clearly observed in acquiring the required training data for based AI and medical image analysis: 1) gaining access to medical archives, located in closed proprietary databases in hospitals with privacy regulations impeding distribution and access to the data, and 2) obtaining validated and annotated image data in a systematic fashion (data values themselves, heterogeneity of the data sources, and provision of data labeling). Moreover, in the last years a large international scientific debate on ethics of AI in imaging is spreading in radiology and other medical fields. This paper illustrates an solution able to create and demonstrate an innovative, open, scalable and trustworthy platform enabling the development, training, validation, evaluation and deployment of innovative AI solutions based on the analysis of medical images, while encompassing an EU-wide health images federated and interoperable repository of a large volume of annotated medical images on selected types of cancer (breast, head&neck, brain, lung, prostate, thorax, …), acquired during diagnostic or therapy medical procedures, compliant with relevant ethics, security requirements and data protection legislation.

Keywords

Artificial Intelligence, Machine Learning, Deep learning

1. Introduction

Cancer was a major cause of death, averaging 1002 deaths per 100.000 inhabitants across the EU in 2016 [1]. As stated in the Cancer statistics - specific cancers [2] report issued by the Eurostat in July 2018, in 2015, 1.3 million people died from cancer in European member states which equated to more than one quarter (25.4%) of the total number of deaths. Cancer accounted for a somewhat higher share (28.7%) of deaths among men than among women (22.1%).

As for *The Global Cancer Observatory*, [3] the top 5 most frequent cancers excluding non-melanoma skin cancer ranked by cases in 2018 in Europe were: prostate, lung, colorectum, bladder and kidney for males; breast, colorectum, lung, corpus uteri and melanoma for females, with a total number of 1.085.582 deaths of men and 857.886 deaths of women. Considering both sexes, breast cancer was the first cause of death, followed by colorectum, lung, prostate, and bladder, with a total of 1.943.478 cancer death.

In this respect, it is of paramount importance for clinicians to use techniques and solutions that helps them to make early diagnosis of cancer, prediction of its evolution because early cancer detection and personalized monitoring is critical to prevent the creation of metastases and to increase survival and cost effectiveness of treatment, decreasing mortality rate as a result. Techniques and solutions are also needed to assist

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29-31, 2023, Pisa, Italy
EMAIL: giampaolo.fiorentino@frontiere.io (A.1); mario.veniero@frontiere.io (A.2); ciro.romano@frontiere.io (A.3)
ORCID: 0000-0002-6772-4885 (A. 1); 0009-0001-6926-7710 (A. 2); 0009-0007-6026-1797 (A. 3)

CEUR Workshop Proceedings (CEUR-WS.org)

clinicians in considering all treatment options and making informed and responsible clinical decisions.

Medical imaging techniques can accelerate diagnosis of cancers [4] and so that determine the adequate treatment per patient. The early detection of cancer, its prognosis and detailed information about the extent of the disease and its evolution, wouldn't be available to patients without medical imaging. All further treatment decisions are based on these findings. [5]

On the other hand, Artificial Intelligence (AI) can assist doctors (therefore, realizing intelligence in a "human in the loop" fashion) by enhancing the prediction, diagnosis, and treatment accuracy. By employing large amounts of medical images, it is possible to develop and train AI algorithms more and more capable of predicting, for instance, who is a likely candidate for developing cancer or establish a precise diagnosis quickly. Moreover, by analysing cancer-related datasets of images, the AI could predict which treatments work best for individual patients, and even forecast clinical endpoints such as complete clinical response and survival. In this sense, AI democratizes health care by boosting access for underserved communities and lowering costs, free overworked doctors and reduce the risk of medical errors.

However, while the world is moving towards an increased adoption of AI algorithms which will increase exponentially over the next few years, the key challenge would be testing of Artificial Intelligence (AI)-Machine Learning (ML) based systems. This is due to **lack of large volumes of data** to train them, at least regarding high-definition health images, as well the lack of structured or standard literature on the methodology or approach to be adopted while testing them. At the same time, in the last several years Information Technology (IT) solutions have been devised and are being developed for handling the big data issues, where huge archives containing various types of datasets have to be processed with high speed. AI methods applied to big datasets of medical images are recently under intense investigation by many groups of health professionals and scientists worldwide, and noteworthy in European Union member states, including teams of medical physicists, computer scientists, biomedical and informatics engineers, radiologists, nuclear medicine physicians, oncologists, radiotherapists. Given the specific nature of medical images of patients' data, experts of ethical and legal issues are important inclusions in those teams.

The European Federation of Organizations for Medical Physics (EFOMP) [6], the Federation of National scientific and professional Associations of Medical Physics in Europe, in their publication [7], addressed general aspects related to *big data and deep learning in medical imaging in relation to medical physics profession*. Here, deep learning (DL) methods refer to a class of AI-ML methods where the representation of the data information content is structured in a layered hierarchy of increasing complexity so that neural network architectures and corresponding high-level data abstraction description processes proceed to the data interpretation task via a cascaded processing system. In this EFOMP White Paper, a state-of-the-art analysis of the field produced a clear conclusion: two main challenges can be clearly observed in acquiring the required training data for DL based AI and medical image analysis:

1. gaining access to medical archives, located in closed proprietary databases in hospitals with privacy regulations impeding distribution and access to the data, and
2. obtaining validated and annotated image data in a systematic fashion (data values themselves, heterogeneity of the data sources, and provision of data labeling) [8]"

2. Concept

The paper proposes to **create and demonstrate an innovative, open, scalable and trustworthy platform enabling the training, validation, evaluation and deployment of innovative AI-based solutions** based on the analysis of medical images, while encompassing an **EU-wide health images federated and interoperable repository of a large volume of annotated medical images** on selected types of cancer (mainly **breast, head&neck, brain, lung, prostate, thorax, abdomen, rectal, pelvis, lower-limbs**), acquired during diagnostic or therapy medical procedures, **compliant with relevant ethics, security requirements and data protection legislation**). Along with a **pool of innovative validated and reliable AI solutions**, iWIN will contribute to better prediction, diagnosis and treatment of the most common forms of tumours, improving the treatment plans as well as **supporting medical experts through a Decision Support System on making transparent and trustworthy decisions and clearly communicate them to patients**.

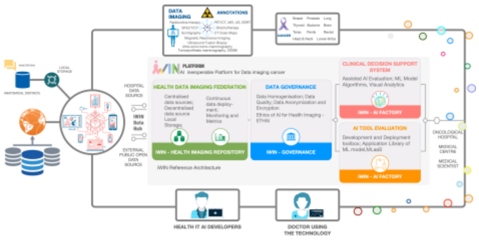

Figure 1: concept

This paper aims at directly focusing all the following challenges addressing to the iWin solution:

- Setting-up and populating an interoperable repository of medical doctors annotated health images on some common forms of cancers, including imaging from radiation treatment planning (e.g. dose maps) and pathological images, adopted for prediction, diagnosis, treatment and therapy of cancer, to securely share them among several Hospitals, while ensuring compliance with relevant legislations as well as ethics principles (including privacy).
- Supporting AI experts to train AI tools making use of the medical/doctors annotated stored health imaging. Data imaging harmonization will facilitate the use of those images by a developed pool of AI solutions.

2.1. Technical Solution

The Figure 2 shows the high-level architecture of iWIN.

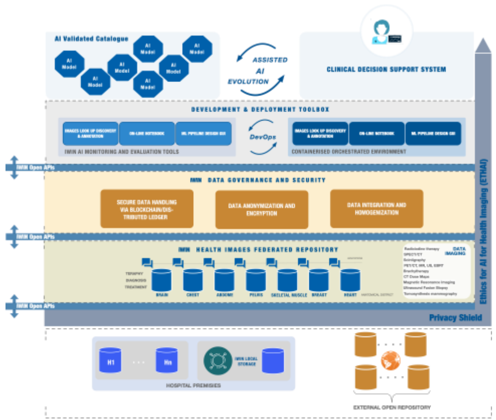

Figure 2: high-level architecture

The aim of the high-level architecture is to provide an overall understanding of the iWIN concept. First, by having a bottom-up overall picture (from health images data to AI-based solution and leveraging application), it is important to mention that the iWIN approach is considering different options to access and deal with health images data set:

1. Hospitals can donate health images data set directly to the iWIN repository while keeping control over their data even after data shipping. iWIN will support the shipment of data repositories by allowing data providers to create data repositories images and push them into a local storage from where they will be available for retrieval or shipment, allowing AI developers to securely benefit from data sharing.
2. iWIN will be committed to make health images data of External and Public Images Repositories accessible through the iWIN platform.

All the above options consider the most appropriate level of Privacy and Security as well as data anonymization by ensuring compliance with relevant legislations.

By means of the different option to access health images dataset, the iWIN Health Images Federated Repository contribute to populate a large interoperable repository of health images enriched with relevant metadata representing the most common forms of cancer. The aim of the iWIN federated repository is twofold: (i) to provide solutions to securely share health images across Europe in a way that is compliant with relevant legislation as well as considering the most appropriate level of privacy and security, (ii) to easily rely on such enabling the development, testing and validation of AI–based health imaging solutions.

On top of the Data Governance and Security, iWIN, through a Development & Deployment Toolbox allows multiple iterations in using of On-line notebook (based e.g. on iPython, Jupyter Notebook as a programming language independent tools) and a Machine Learning Pipeline Design GUI (based on e.g. KubeFlow) realizes a comprehensive solution for deploying and managing end-to-end ML workflows for rapid and reliable experimentation to schedule and compare runs, and examine detailed reports on each run of an AI experiment. The AI developer can monitor the progress of the experiment and (partially) interact with it. Multiple real-time statistics is displayed in various forms (digital displays, graphs, etc.) allowing the AI developer to gain live insights during the experiment. The DevOps continuous cycle allows the AI developer to run AI experiment making deployments of machine learning (ML) workflows on a Containerised Orchestrated Environment which makes the experiment replicable, trustable, portable and scalable. The iWIN DevOps cycle goal is to provide a straightforward way to deploy AI solutions to possible diverse infrastructures by defining in a declarative way the computational and storage resources of the infrastructure layer thus to realize the infrastructure-as-a-code principle. In addition, the Images annotation functionality will let medical staff the ability to browse through extensive lists of the available iWIN data imaging.

Once the iteration cycle followed by the AI solution developers is finished, the AI solution is made available through the iWIN AI Solutions Catalogue.

The AI Solutions Catalogue provides a transparent and reliable set of AI solution to facilitate the aim of the Clinical Decision Support System (DSS) whose purpose is to provide guidance to clinicians evaluating medical images. The Clinical DSS combines the input data with the selected AI solution output to suggest a set of intervention options to the clinician. Therefore, the Clinical DSS is based on the innovative AI solutions developed for iWIN use cases and a new inference engine. The AI solutions identifies, classifies, and identifies patterns in medical images, whereas the inference engine combines this information with patient-specific data input to the Clinical DSS to generate the expected results. Moreover, explanation methods are applied to AI algorithms to provide a clear picture of the relevant features affecting the performance of the selected AI solutions, their relations with the outcomes and with each-others and both their local/global effects on the problem under investigation. These steps help the specialists to perform supervised considerations, to adopt the best strategies in a decision-making clinical process and to communicate the rationale behind them to their patients.

All the technological development embrace different ethics approaches, and the main objective of this integration is the definition of the ETHAI (Ethics of AI for Health Imaging), a model of ethics standard of AI for health imaging.

3. AI-based solution for Use Case

Employing the development of AI solutions, iWIN is devoted to cancer prediction, diagnosis and treatment. These studies feed into and interact with all Medical centres to guarantee the improvements and results each other.

3.1. AI for segmentation of 3D breast images

State-of-the-Art (description): Nowadays, the research in novel imaging techniques as well as in testing and optimizing existing ones is inevitably related to the exploitation of medical patient images. These are used to evaluate the improved technologies in the x-ray techniques, to train and educate the medical specialists with the new technology, to extract features, which are basis for the development of advanced computer-aided detection systems or develop and train machine learning algorithms. For researchers, the use of images from databases may be the most flexible and time efficient approach, since data are summarized at one place and well documented. In case of breast imaging the demands are similar. All imaging modalities are designed, optimised and validated because of the need for earlier breast screening, better visualization and diagnosis of breast lesions. In a preliminary stage, computational tools are used for the evaluation of these new imaging techniques. In this case, besides the computational model of the imaging system, the other important elements are both the anthropomorphic model of the breast and realistic in form and shape 3D lesion models.

Beyond the State-of-the-Art (innovation): To have realistic breast models, different approaches and sources of information may be exploited. One such approach is based on the use of algorithms for segmentation of breast lesions (for instance based on convolutional neural networks) from 3D medical images: Computed Tomography (CT), Digital Breast Tomosynthesis (DBT), breast CT, Magnetic Resonance Imaging. With the introduction of the breast CT into clinical use, this modality may become a very popular approach for extracting 3D breast features. Such segmentation approaches, when validated well may represent a good approach in creation of computational models for the needs of breast imaging optimization. Very recently, it has been reported on the use of machine learning technique for generating 3D super-resolution (~60 μm voxel size dimensions) breast models based on breast CT data which are characterized with lower resolution compared to data from breast tomosynthesis. This approach may be successfully adopted as well as to segment breast lesions from similar data and generate super-resolution computational breast lesion models, which is an approach to be further explored and studied for its feasibility by researchers. Large databases of images would be needed to observe, model, validate and evaluate breast lesions. Therefore, the creation and the technical and scientific support of a database, dedicated to such purposes is of high interest.

Objective: We propose a novel approach for the realization of realistic in shape, size as well as in x-ray absorption properties 3D physical breast models for the development.

data augmentation with the aim of generating the first synthetic dataset of breast images.

3.2. AI in personalized dosimetry: Automatic extraction of body organs from the 3D dose volume

State-of-the-Art (description): Personalized dosimetry is an advanced method used to provide information about the 3D dose distribution for patients undergoing diagnostic and interventional X-ray examinations such as CT and fluoroscopically guided procedures. Input data required by the Monte Carlo software to start the dose computation procedure is:

- Input volume: A set of CT reconstructed images from an examination, in DICOM format.
- Scan parameters: Data for beam spectrum, filtration, and geometrical specifications.
- Simulation parameters: The number of x-rays depositing energy in the input volume is selected. Good statistical performance (<1% uncertainty) is obtained using a value in the order of 10^8 to 10^9 interacting x-rays.

The Monte Carlo software output after each computation is in the form of a 3D dose distribution, based on the physical properties (i.e. attenuation, composition and size) obtained from the input CT scan. Each slice in the dose volume corresponds to the same slice in the CT scan. Each pixel in a specific slice of the CT volume has a corresponding dose value in the 3D dose distribution output. Organ tissue dose information is extracted from 3D dose distributions through appropriate delineation. This is a very time-consuming procedure that prohibits personalized dosimetry to be used in everyday clinical practice.

Beyond the State-of-the-Art (innovation): Novel, sophisticated artificial intelligence, and machine learning algorithms, for example algorithms based on convolutional neural networks, develops to assist in automatic extraction of main radiosensitive organs of the chest from the 3D dose volume.

Objective: This approach allows automatic extraction of main radiosensitive organs of the chest needed for the implementation of patient-specific dosimetry

3.3. Diagnosis using CT images of head, lung and breast cancers using AI techniques

State-of-the-Art (description): Computed tomography (CT) is established as an excellent imaging technique for anatomy and pathology in various clinical areas. High resolution, three-dimensional image sets can be acquired with CT using x-ray beams in the energy range of 80 &140 kV. In addition, image sets can be acquired at multiple energies to enable dual energy computed tomography (DECT). DECT can be used to differentiate between different materials. AI approach for extracting effective atomic number is also particularly sensitive to the presence of artefact since the resultant small changes in the EMI number may be magnified into large errors. It is anticipated, however, that as techniques improve, the ability to carry out pathological processes in vivo will be an important application of computed tomography.

AI can minimize diagnostic errors during CT image interpretation [9].

Beyond the State-of-the-Art (innovation): The idea is to develop and train a smart software using the annotated DECT images from all partners of the project.

Objective: Use AI algorithms to maximize the accuracy of lung, breast and prostate cancer differentiation through in the ROIs across annotated CT images.

3.4. AI for optimization the radiation treatment planning process

State-of-the-Art (description): Radiotherapy process includes six major consecutive steps that encompass the entirety of treatment: patient assessment, simulation, planning, quality assurance, treatment delivery, and follow-up. For some of these steps (i.e. planning) the process may take hours or even days to complete. Areas where a data-centric approach using Machine Learning could improve the quality and efficiency of patient care were recently outlined. [10] For instance, the delineation of the target volume and the organs at risk by the radiation oncologist is a very complex task and is currently based on commercial auto-segmentation algorithms which are developed on atlas-based strategy rather than using Machine Learning. The main reason for this is the limited size of datasets available in radiation oncology. In addition, in some cases, such as rectal cancers, oncologists manually segment clinical target volumes. The literature review, however, shows a disagreement for target volume delineation in rectal cancer up to 1 cm, and this represents one of the most significant geometric uncertainties and causes of systematic error through the treatment. [11]

Beyond the State-of-the-Art (innovation): Artificial Intelligence algorithms will be trained to delineate target volumes and organs at risks for two specific cancer cases: rectal and lung cancers.

Objective: We propose to use artificial intelligence algorithms to automate and improve radiotherapy treatment planning process and develop suitable algorithms to support planners.

4. Conclusion

This solution leads toward two main challenges can be clearly observed in acquiring the required training data for DL based AI and medical image analysis:

1) gaining access to medical archives, located in closed proprietary databases in hospitals with privacy regulations impeding distribution and access to the data, and

2) obtaining validated and annotated image data in a systematic fashion (data values themselves, heterogeneity of the data sources, and provision of data labeling).

Moreover, in the last years a large international **scientific debate on ethics of artificial intelligence in imaging** is spreading in radiology as well as other medical fields. AI for health imaging has great potential to increase efficiency and accuracy.

5. References

[1] “Causes and occurrence of deaths in the EU,” 16 07 2019. [Online]. Available: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20190716-1.

[2] "Translate Cancer statistics," EUROSTAT, 05 2023. [Online]. Available: ec.europa.eu/eurostat/statistics-explained/index.php?title=Cancer_statistics.

[3] "International Agency for Research on Cancer," World Health Organmization, 2021.Available:http//gco.iarc.fr/today/data/factsheets/populations/908-europe-fact-sheets.pdf.

[4] Z. Mohammadzadeh, R. Safdari, M. Ghazisaeidi, S. Davoodi and Z. Azadmanjir, Advances in Optimal Detection of Cancer by Image Processing; Experience with Lung and Breast Cancers, 2015.

[5] S. Lee and M. Crean, "Making cancer visible the role of imaging in oncology," [Online].Available:https://www.internationaldayofradiology.com/app/uploads/2017/09/IDOR_2012_OncologyImaging_lowres.pdf.

[6] "European Federation of Organisations for Medical Physics," [Online]. Available: https://www.efomp.org.

[7] M. Kortesniemi, V. Tsapaki, A. Trianni, P. Russo, A. Mass, H.-E. Källman, M. Brambilla and J. Damilakis, "The European Federation of Organisations for Medical Physics (EFOMP)," in *European Journal of Medical Physics*, November, 2018.

[8] Geis, JR; et al, Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement., Radiology, 293(2), pp. 436-440 , 2019.

[9] H. I. K. C. Panagopoulos K., Intelligent System for Preoperative Evaluation of Parathyroid Gland for Surgical Intervention Prediction, Journal of Clinical and Experimental Radiology,, JCER-1000106, 2019.

[10] M. G. V. N. D. a. T. D. S. Feng, Machine Learning in Radiation Oncology: Opportunities, Requirements, and Needs, Front Oncol , 2018.

[11] Nijkamp, J., D. F. de Haas-Kock, J. C. Beukema, K., Target volume delineation variation in radiotherapy for early stage rectal cancer in the Netherlands., Radiother Oncol 102.