Moddicom: a Complete and Easily Accessible Library for Prognostic Evaluations Relying on Image Features

Nicola Dinapoli[1], Anna Rita Alitto[1], Mauro Vallati[2], Roberto Gatta[1],
Rosa Autorino[1], Luca Boldrini[1], Andrea Damiani[1] and Vincenzo Valentini[1]

***Abstract*— Decision Support Systems (DSSs) are increasingly exploited in the area of prognostic evaluations. For predicting the effect of therapies on patients, the trend is now to use image features, i.e. information that can be automatically computed by considering images resulting by analysis. The DSSs application as predictive tools is particularly suitable for cancer treatment, given the peculiarities of the disease –which is highly localised and lead to significant social costs– and the large number of images that are available for each patient.**

At the state of the art, there exists tools that allow to handle image features for prognostic evaluations, but they are not designed for medical experts. They require either a strong engineering or computer science background since they do not integrate all the required functions, such as image retrieval and storage. In this paper we fill this gap by proposing Moddicom, a user-friendly complete library specifically designed to be exploited by physicians. A preliminary experimental analysis, performed by a medical expert that used the tool, demonstrates the efficiency and the effectiveness of Moddicom.

I. INTRODUCTION

Decision Support Systems (DSSs) based on digital medical image-analysis have been increasingly accepted in the daily clinical practice, although they are considered as a second-best choice by some Specialists. The most popular application of such techniques is in diagnostic, where they have been largely applied (see, for example, their application for breast cancer diagnosis [1]). Recently, their exploitation is extending to the areas of prognostic evaluations or for predicting the effect of a given therapy on the patient, both in terms of disease control and/or iatrogenic effects of the treatments. Cancer treatment, in particular, is a topic where these techniques are strongly motivated: firstly, because of the nature of the disease (high social costs, direct effect on survival), and secondly due to the specific behaviour of the disease, normally "clearly located" and presenting different visual features according to used modalities: morphological –e.g., computed tomography (CT), magnetic resonance (MR) and Ultrasound– or metabolic –e.g., Functional magnetic resonance imaging, and Single-photon emission computed tomography–. Moreover, clinical decisions taken for treating cancer have a strong impact on quality of life of patients. This is due to toxicities induced by chemotherapy and radiotherapy; and the clinical path is often a painful trade-off, balancing probability of life expectancy and quality of life. To face this highly facet topic, a modern approach in medical image analysis is proposed in RADIOMICS [2], [3] where authors introduced the extraction from the images of a large number of features and an analysis inspired by genomics analysis and modern Machine Learning (ML) algorithms. Such analysis is now quite common and in many centre studies about prognosis and toxicities have been performed: Ravanelli et al. [4] investigate how to predict the effects of the 1st line of chemotherapy in lung cancer treatment by texture analysis; Bayanati et al. [5] try to identify, by analysing texture and shape, which features on CT images could predict the state of lymph nodes (benign vs malignant) in lung cancer.

At the state of the art, there exists a number of tools able to handle image features (e.g., [6], [7]) but they are usually designed for IT specialists, and are not easy to use by physicians, since they do not implement all the required basic infrastructure (e.g., databases, DICOM acquisition frameworks, etc..) allowing the user to focus on the the statistical analysis, only. The most famous solution which provides a complete framework, allowing non-IT experts to use it, is VODCA[1], which has been successfully used in numerous clinical studies (e.g., [8]). It implements all the infrastructure to acquire, store and extract Dose Volume Histograms (DVHs) and it provides the user with a friendly environment for handling data by the well-known and powerful statistical software R [9]. On the other hand, VODCA is an expensive commercial solution, it is not free or open source, it allows to handle only DVHs and does not implement the idea of data analysis outlined by the RADIOMICS approach, which considers also many other features.

In order to foster the development of widely exploitable free tools, in this paper we propose a library, called Moddicom, which has been designed for supporting user-friendly frameworks. Moddicom is able to handle DICOM/DICOM-RT objects, and can present to the user, in a R environment, data structures representing CT, MR and Radiation Dose Distribution voxels, Region of Interests, DVHs, and image features classified in morphological, textural and statistical. The library is free and open source and allows the user to work in R; this has been done in order to provide physicians a suitable environment to handle with an already available collection of image features. A further improvement of moddicom is already planned, and the aim is to propose a full architecture able also to connect Institutional Picture Achieving Communication Systems (PACS) and Treatment Planning System (TPSs) via DICOM protocol and preset to

[1]Radiotherapy Department, Università Cattolica del Sacro Cuore, Italy roberto.gatta.bs@gmail.com

[2]School of Computing and Engineering, University of Huddersfield, UK

[1]http://www.vodca.ch/

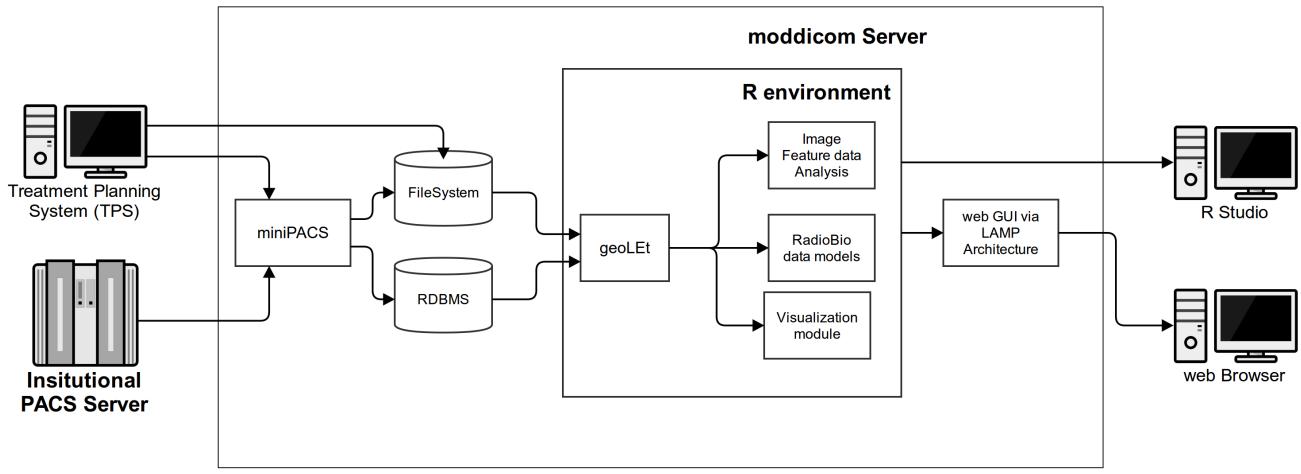

Fig. 1. The overall structure of the Moddicom architecture. The core, which is already available, is represented in the box "*R environment*".

the user a web based Graphical User Interface (GUI).

In the remainder of this paper we firstly describe the Moddicom architecture. Second, we show a real case where it was used to handle clinical data to find significant results in prognostic evaluation for patients treated for rectal cancer. Finally, conclusions are given.

II. STRUCTURE

The core of Moddicom is a R package composed by classes written using functions and S3 objects paradigm. It is tailored to interact with RStudio [10] in order to give to the user the most comfortable environment, using R. The main classes are shown in Figure 1. The main modules of Moddicom are:

- **geoLet**. It allows to load DICOM images from a given path. All the stored DICOM series/images are loaded and subsequently used for building a memory structure of imaging data (CT/MR), doses (RTDose) and structures (RTStruct). It requires the previous installation of DCMTK Offis libraries [11].
- **Image Feature data Analysis**. Given a path, this module loads recursively all the stored DICOM Studies and, for each of them, it extracts corresponding image features. The provided output includes a structure containing the calculated features, that can be easily exploited by R to allow statistical computing
- **RadioBio data models**. This module calculates DVHs (starting from RTDose built by geoLet) and fills six different radio-biological models designed to to build a predictor of Tumour Control Probability (TPC) and Normal Tissue Complication Probability (NTCP). It is also responsible for simulating DVHs to test fitting or simulate calculus.

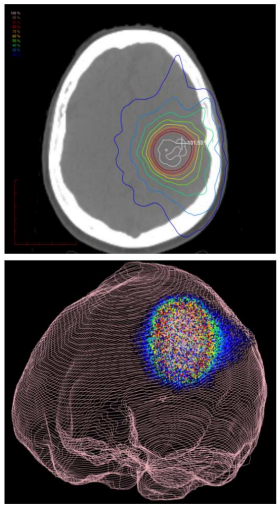

Fig. 2. Example of different views provided by Moddicom for representing dose distribution for a brain cancer treatment.

- **Visualization module**. The aim of this modules is visualising loaded images. It is able to plot CT/RM images in 3D, by mesh, to plot Region Of Interests on the corresponding image slices and show Radiation Doses either in 2D or 3D.

Figure 2 shows an example of different views provided by Moddicom of dose distribution for a brain cancer treatment.

III. EVALUATION

In our experimental analysis we exploited a preliminary version of the Moddicom library.

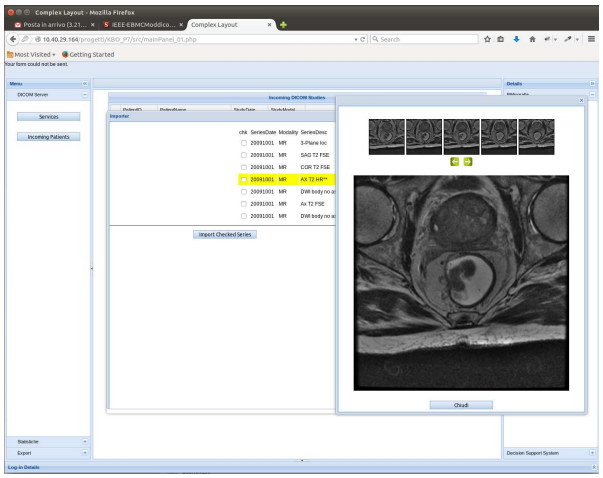

Fig. 3. A preliminary draft of the proposed graphical user interface.

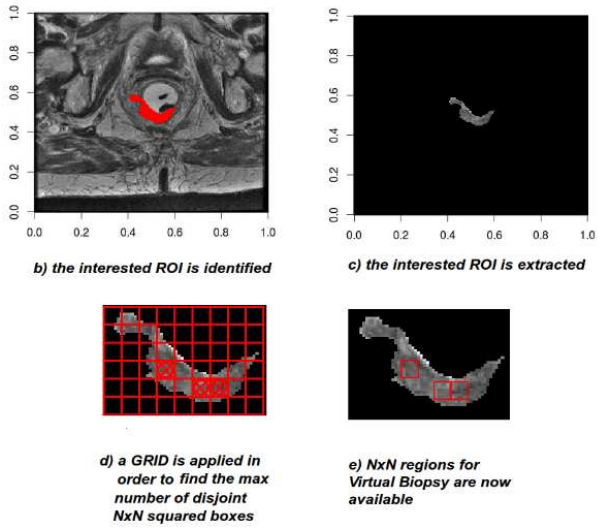

Fig. 4. Example of biopsied points.

Such version is publicly available on gitHub (https://github.com/kbolab/moddicom.git). In the available version it works on RStudio and require some statistical skill but we are currently actively working on a GUI (a mockup is shown in Figure 3).

In order to evaluate its usefulness and effectiveness in real-world scenario, we tested Moddicom on a clinical dataset of 26 MRs from patients treated for rectal cancer at the Department of Radiotherapy of the Gemelli hospital of Rome, Italy. The MRs were pre-treatment images and the aim was to extract image features about the Gross Tumour Volume (GTV) for predicting possible post-treatment recurrences without relying on any other clinical information. All the patient had the same RT treatment, in terms of techniques and GTV Doses. More formally, Moddicom has been tested with prognostic aims for a TCP estimation.

All the images has been normalised with respect to the same zero captation (black voxels in the dose-captation cube) and the higher captation in bladder. Particularly, the latter played a crucial role for our analysis because, even if all the MRs were scaled by an unknown and different factor (a not-declared rescale slope due to physics acquisition), we made them comparable by the estimation of such unknown rescale slope by the captation value in the bladder.

DICOM Studies were classified as negative (no recurrences, 19) or positive (recurrences, 6). Given the small number of recruited patients we considered few features, with the aim of avoiding over-fitting behaviours. Features considered are the follows:

- Kernel Density Function (KDF) of GTV voxels.
- A feature corresponding to the fact that for each slice of the GTV voxels, the 2D spectral analysis value is lower than the average. For each patient, this analysis has been conducted by using a set of "virtual biopsy" of the GTV, in order to avoid the inclusion of high gradients at the borders of the ROI, and to focus the attention on its internal texture.
- volume and volume/surface ratio from the Morphological set.

All the analysis, and the evaluation of the corresponding results, have been performed by a Radiation Oncologists skilled in the use of R, but without a strong engineering or computer science background.

IV. RESULTS

Most of the time required by Moddicom has been spent in loading the 26 DICOM studies (MR and the corresponding RTStruct). In particular, loading the studies required approximately 30 minutes. On the other hand, all the subsequent required computations (Kernel Density Function, Spectral Analysis, etc..) were extremely fast. They usually require a few tens of seconds. Figure 4 shows an example of biopsied points. The Spectral Analysis conducted on the virtual biopsies of the GTV did not show a significant difference, this is possibly due to the little number of recruited patients.

Similarly, a noticeable relation between volume size and expected clinical outcome has been observed but, since it is a well-known relation, it was not of interest for our purposes.

An interesting results was obtained by analysing the KDF. The application of this technique allows identifying a qualitative difference between the two clusters of patients. Such difference can be easily observed in Figure 5. Furthermore, an analysis performed on the average of KDFs of the two clusters, with the boundaries of 95% of Confidence Interval (CI), is shown in Figure 6. A Kolmogorov-Smirnov test [12] performed on such differences, revealed a significant statistical difference ($p = 0.05$) between the two distributions. The value of Normalised HR MRI voxel signals equals to 0.23244.

Since we are aware that 26 studies are not enough to support a reliable clinical evidence, we also tested the same features on the corresponding MRs post-treatment of the patients. Although this will not increase the number of studies, it can provide some useful insights about the outlined behaviour. Remarkably, we observed that the behaviour was

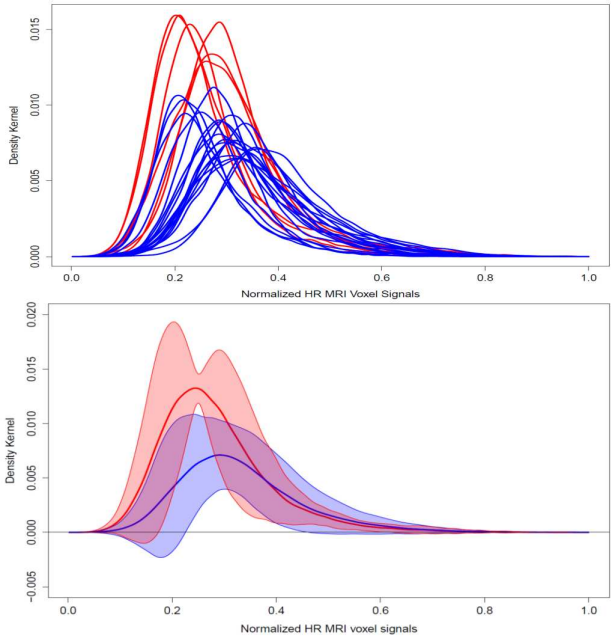

Fig. 5. The application of KDF allows to identify a qualitative difference between two clusters of considered patients.

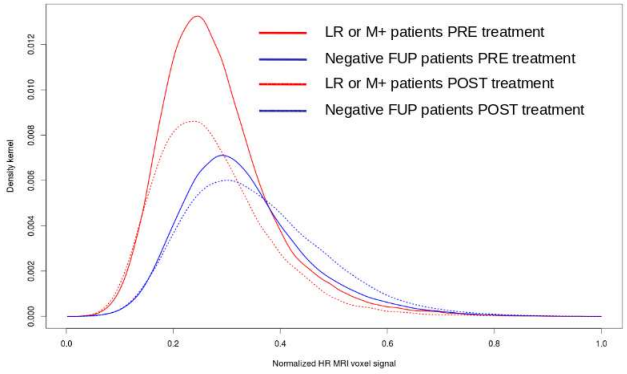

Fig. 6. The results of a KDF analysis of the two clusters, with the boundaries of 95% of Confidence Interval.

confirmed also after the treatment, and that it is still possible to distinguish the two classes of patients. Intuitively, such results support the behaviour but, more interestingly, clearly indicate the potential of the Moddicom library. Providing an easy-to-use tool for medical experts can foster the identification of trends, thus improving the use of medical treatments.

V. CONCLUSIONS

The area of image features analysis for clinical research is strongly affected by the poor availability of free user-friendly tools and software, designed for and exploitable by physicians. In order to fill this gap, in this paper we presented Moddicom, an user-friendly and freely available library. Moddicom has demonstrated to be capable to perform RADIOMICS investigation for research in radiation therapy. Remarkably, Moddicom is a great opportunity for modern Radiation Therapy Departments for a number of reasons: (i) it is free; (ii) it works on a common and well-know statistical software package, which is increasingly mastered by physicians, and (iii) Moddicom has shown to be effective and useful to find relations between image features and clinical outcome.

Finally, in order to foster the exploitability of Moddicom, we are currently working on a graphical user interface. Future works include the development of modules for collecting DICOM studies directly via DICOM protocol, and for easily handling a local mini-PACS repository that will allow grouping patients and studies in many clinical investigation tasks.

This job followed the principles outlined in the Helsinki Declaration of 1975, as revised in 2000.

REFERENCES

[1] K. Ganesan, U. Acharya, C. K. Chua, L. C. Min, K. Abraham, and K. Ng, "Computer-aided breast cancer detection using mammograms: A review," *Biomedical Engineering, IEEE Reviews in*, vol. 6, pp. 77–98, 2013.

[2] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. Aerts, A. Dekker, D. Fenstermacher *et al.*, "Radiomics: the process and the challenges," *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1234–1248, 2012.

[3] 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.

[4] M. Ravanelli, D. Farina, M. Morassi, E. Roca, G. Cavalleri, G. Tassi, and R. Maroldi, "Texture analysis of advanced non-small cell lung cancer (nsclc) on contrast-enhanced computed tomography: prediction of the response to the first-line chemotherapy," *European radiology*, vol. 23, no. 12, pp. 3450–3455, 2013.

[5] H. Bayanati, R. E. Thornhill, C. A. Souza, V. Sethi-Virmani, A. Gupta, D. Maziak, K. Amjadi, and C. Dennie, "Quantitative ct texture and shape analysis: can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer?" *European Radiology*, vol. 25, no. 2, pp. 480–487, 2015.

[6] A. H. Squillacote, "The paraview guide: A parallel visualization application. kitware inc., 2007," ISBN 1-930934-21-1, Tech. Rep.

[7] L. Zhang, M. Hub, S. Mang, C. Thieke, O. Nix, C. P. Karger, and R. O. Floca, "Software for quantitative analysis of radiotherapy: overview, requirement analysis and design solutions," *Computer methods and programs in biomedicine*, vol. 110, no. 3, pp. 528–537, 2013.

[8] A. Maggio, V. Carillo, C. Cozzarini, L. Perna, T. Rancati, R. Valdagni, P. Gabriele, and C. Fiorino, "Impact of the radiotherapy technique on the correlation between dose–volume histograms of the bladder wall defined on mri imaging and dose–volume/surface histograms in prostate cancer patients," *Physics in medicine and biology*, vol. 58, no. 7, p. N115, 2013.

[9] R. Statistical Package, "R: A language and environment for statistical computing," *Vienna, Austria: R Foundation for Statistical Computing*, 2009.

[10] J. Allaire, "RStudio: Integrated development environment for R," in *The R User Conference, useR!*, 2011, p. 14.

[11] M. Eichelberg, J. Riesmeier, T. Wilkens, A. J. Hewett, A. Barth, and P. Jensch, "Ten years of medical imaging standardization and prototypical implementation: the dicom standard and the offis dicom toolkit (dcmtk)," in *Medical Imaging 2004*, 2004, pp. 57–68.

[12] R. Wilcox, *Kolmogorov-Smirnov Test*. John Wiley & Sons, Ltd, 2005.