ARTIFICIAL INTELLIGENCE AND MEDICAL IMAGING APPLICATIONS

Leonardo RUNDO
lrundo@unisa.it
Leonardo RUNDO is an Assistant Professor at the Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, Italy. His research interests cover biomedical image analysis, radiogenomics, machine learning, computational Intelligence, high-performance computing.

Carmelo MILITELLO
carmelo.militello@cnr.it
Carmelo MILITELLO is a Research Scientist at the Institute for High-Performance Computing and Networking, Italian National Research Council (ICAR-CNR), Palermo, Italy. His research focuses on biomedical image analysis, radiomics, machine learning, digital architectures, biometrics, and hardware programmable devices.

Goal, Applications and Potential Impact to Clinical Practice

Artificial Intelligence (AI) provides considerable support to healthcare processes, allowing for the development of advanced technologies for improving the quality and personalization of medical care provided to patients. In the field of medical imaging, this has made it possible to develop and implement numerous tools that support physicians in different tasks of the treatment process: computer-assisted segmentation, support for diagnosis, assessment of response to treatment, and predictive models based on radiomics [1]. For these reasons, computer-assisted image analysis is considered an essential tool in the clinical workflow.

Despite the increasing diffusion of Information and Communication Technologies (ICT) in medicine, some medical activities are still performed manually. Manual procedures are highly dependent on the experience of the physician. Therefore, the operator-dependence is a strong, critical issue in terms of reproducibility and repeatability of results: in fact, considerable intra- and inter-operator variability may seriously affect the results.

In this scenario, computer-assisted approaches – i.e., automatic or semi-automatic region of interest (ROI) segmentation based on AI techniques allow us to mitigate some of the limitations typical of manual procedures. In fact, AI can offer clinical tools that allow us to reduce inter- and intra-observer dependence and improve the repeatability of the results. In particular, AI applied to clinical decision support systems (CDSS) could help doctors during all the stages of the healthcare processes, from diagnosis to treatment planning, as well as support for the prognosis.

In what follows, three aspects characterized by a potentially high impact in the definition and implementation of effective clinical tools will be analyzed in the context of medical imaging: (i) radiomics-powered predictive models, with the goal

of offering solutions in the clinical practice; (ii) computer-assisted ROI detection and segmentation, for dealing with result repeatability and accelerating the annotation of large-scale datasets, and (iii) feature robustness and reliability, which are fundamental aspects to define solid biomarkers to be adopted by AI-based clinical tools [2].

The general scheme for robust biomarker discovery is outlined in Figure 1.

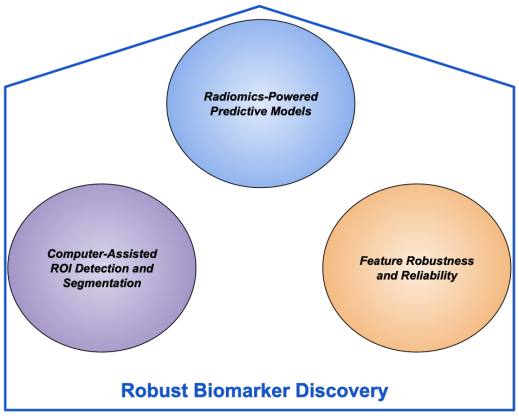

Figure 1. Key aspects for robust biomarker discovery: (i) radiomics-powered predictive models; (ii) computer-assisted ROI detection and segmentation; (iii) feature robustness and reliability.

Radiomics-Powered Predictive Models

Radiomics involves the extraction of mineable features from medical images to non-invasively characterize the in vivo phenotype of lesions or even simply of tissue portions (e.g., the apparently normal tissue surrounding a tumor), capturing the ROI characteristics. These can be morphometric measurements (size, shape, and diameter) or texture heterogeneity measurements of tissues [3], [4]. Starting from an input ROI, the radiomic features can be calculated in two manners: (i) voxel-based extraction (for each feature, a value is computed for each voxel, thus yielding feature maps as output), and (ii) segment-based extraction (a single, aggregated value per feature is computed for each ROI).

In [5] a study is proposed to develop and validate a radiomic model, with radiomic features extracted from breast Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI), for the prediction of mass-enhancement lesion malignancy. A total of 107 radiomic features were extracted from a manually annotated dataset of 111 patients, which was split into discovery and test sets. A feature calibration and pre-processing step was performed to find only robust non-redundant features. An in-depth discovery analysis was performed to define a predictive model: for this purpose, a Support Vector Machine (SVM) was trained in a nested 5-fold cross-validation scheme, by exploiting several unsupervised feature selection methods. The predictive model performance was evaluated in terms of Area Under the Receiver Operating Characteristic (AUROC), specificity and sensitivity, by considering the held-out set. The model combining Unsupervised Discriminative Feature Selection (UDFS) and SVMs on average achieved the best performance on the blinded test set: AUROC = 0.725±0.091, sensitivity = 0.709±0.176 and specificity = 0.741±0.114. The experimental findings demonstrate that the radiomic predictive model based on breast DCE-MRI, using only the strongest enhanced phase, got promising results in terms of accuracy and specificity in the differentiation of malignant from benign breast lesions.

The predictive capabilities of radiomic models can be useful both in diagnosis, as observed in previous work, but also in prognosis, i.e., in terms of treatment response prediction. Pathological response to neoadjuvant treatment for patients with high-grade serous ovarian carcinoma (HGSOC) was assessed using the chemotherapy response score (CRS) for omental tumor deposits. The main limitation of CRS is that it requires surgical sampling after initial neoadjuvant chemotherapy (NACT) treatment. Earlier and non-invasive response predictors could improve patient stratification. To this end, Computed Tomography (CT) radiomic features were adopted to predict neoadjuvant response before NACT using CRS as a gold standard. In [6] omental CT-based radiomics models, yielding a simplified fully interpretable radiomic signature, were developed using Elastic Net logistic regression and compared to predictions based on omental tumor volume alone. Models were developed on a single institution cohort of neoadjuvant-treated

HGSOC (n = 61; 41% complete response to NCT) and tested on an external test cohort (n = 48; 21% complete response). The performance of the comprehensive radiomics models and the fully interpretable radiomics model was significantly higher than volume-based predictions of response in both the discovery and external test sets, indicating high generalizability and reliability in identifying non-responders when using radiomics. Interestingly, the performance of a fully interpretable model was similar to that of comprehensive radiomics models.

Computer-Assisted ROI Detection and Segmentation

Radiomic image analysis relies upon quantitative features from medical images. The resulting quantitative models exploit these data for classification, prediction, prognostication and treatment response may be built. To this end, the tumor has to be segmented, which is mostly performed manually by the radiologist. This variability among readers is often recognized as a source of potential problems, as variability among readers leads to undermining the repeatability of results. Some relevant case studies will be outlined in what follows.

Prostate cancer is the most common malignant tumor in men but reliable prostate Magnetic Resonance Imaging (MRI) analysis remains challenging. Besides whole prostate gland segmentation, the capability to differentiate between the blurry boundary of the Central Gland (CG) and Peripheral Zone (PZ) can lead to differential diagnosis, since the frequency and severity of tumors differ in these regions. To tackle the prostate zonal segmentation task, in [7] a novel Convolutional Neural Network (CNN), called USE-Net, which incorporates Squeeze-and-Excitation (SE) blocks into a standard U-Net, was proposed. The SE blocks were added after every Encoder (Enc USE-Net) or Encoder-Decoder block (Enc-Dec USE-Net). This study evaluated the generalization ability of CNN-based architectures on three T2-weighted MRI datasets, each one consisting of a different number of patients and heterogeneous image characteristics, collected by different institutions. The following mixed scheme was used for training/testing: (i) training on either each individual dataset or multiple prostate MRI datasets and (ii) testing on all three datasets with all possible training/testing combinations. USE-Net was compared against three state-of-the-art CNN-based architectures, along with a semi-automatic continuous max-flow model. The results showed that training on the union of the datasets generally outperforms training on each dataset separately, allowing for both intra-/cross-dataset generalization. Enc USE-Net showed good overall generalization under any training condition, while Enc-Dec USE-Net remarkably outperformed the other methods when trained on all datasets. These findings reveal that the SE blocks' adaptive feature recalibration provided excellent cross-dataset generalization when testing is performed on samples of the datasets used during training. Therefore, we should consider multi-dataset training and SE blocks together as mutually indispensable methods to draw out each other's full potential. In conclusion, adaptive mechanisms (e.g., feature recalibration) may be a valuable solution in medical imaging applications involving multi-institutional settings.

As further development, the very recent Focus U-Net [8] combined efficient spatial and channel attention into a Focus Gate. Focus U-Net was applied to polyp segmentation during colonoscopy, by outperforming state-of-the-art results across five public polyp datasets. Moreover, loss functions play a crucial role for class-imbalanced medical imaging datasets. Therefore, the novel Unified Focal loss, which generalizes Dice and cross-entropy based loss functions, was recently introduced [9].

Regarding classic Machine Learning techniques, there are still highly relevant and successful case studies.

Multiparametric Magnetic Resonance Imaging (MRI) is the most sensitive imaging modality for breast cancer detection and is increasingly playing a key role in lesion characterization. In this context, accurate and reliable quantification of the shape and extent of breast cancer is crucial in clinical research environments.

Since conventional lesion delineation procedures are still mostly manual, automated

segmentation approaches can improve this time-consuming and operator-dependent task by annotating the regions of interest in a reproducible manner.

In [10] , a semi-automated and interactive approach based on the spatial Fuzzy C-Means (sFCM) algorithm is proposed, used to segment masses on dynamic contrast-enhanced (DCE) MRI of the breast. This method was compared against existing approaches based on classic image processing, namely (i) Otsu's method for thresholding-based segmentation, and (ii) the traditional FCM algorithm. A further comparison was performed against state-of-the-art CNNs for medical image segmentation, namely SegNet and U-Net, in a 5-fold cross-validation scheme. The results showed the validity of the approach, by significantly outperforming the competing methods in terms of the Dice similarity coefficient (84.47±4.75). A Pearson's coefficient of ρ=0.993 showed a high correlation between segmented volume and the gold standard provided by clinicians. Therefore, such a computer-assisted approach could be deployed into clinical research environments by providing a reliable tool for volumetric and radiomics analyses.

Feature Robustness and Reliability

Although radiomic features are well-established, there are still serious concerns about their stability and robustness. Indeed, radiomic features are often not robust against medical image acquisition parameters, such as spatial resolution (i.e., in-plane resolution and slice thickness, the latter also known as through-plane resolution) and image extraction settings (e.g., intensity quantization, voxel resampling).

Robust Machine Learning models based on radiomic features allow us to obtain biomarkers that are capable of modeling the disease and are able to support medical decision-making tasks, from diagnosis to prognosis. Recent studies have shown that it is fundamental that the computed features are robust and reproducible. Although several initiatives to standardize the definition and extraction process of biomarkers are ongoing, there is a lack of comprehensive guidelines. Therefore, no standardized procedures are yet available for ROI delineation, feature extraction, and processing, with the risk of undermining the effective use of radiomic models in clinical routine.

This kind of achievement could be realized by combining classic Machine Learning techniques with the latest advances in Deep Learning. For instance, in [11], the authors propose a Generative Adversarial Network (GAN)-based lesion-focused framework for CT image Super-Resolution (SR); for the lesion (i.e., cancer) patch-focused training, a Spatial Pyramid Pooling (SPP) was incorporated into GAN-Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE). At 2× SR, the proposed model achieved better perceptual quality with less blurring than the other considered state-of-the-art SR methods, while producing comparable results at 4× SR. Interestingly, the robustness of the radiomic features was evaluated in terms of quantization on a different lung cancer CT dataset using Principal Component Analysis (PCA). Relying upon this analysis, the most important radiomic features in the conducted PCA-based analysis were the most robust features extracted on the GAN-super-resolved images. These achievements pave the way for the application of GAN-based image Super-Resolution techniques for studies of radiomics for robust biomarker discovery.

In [12], the aim was to assess the impact that the different segmentation methods and the quantization level (defined by means of the number of bins used in the feature-extraction phase) may have on the robustness of the radiomic features. In particular, the robustness of texture features extracted by the PyRadiomics tool, and belonging to five categories – namely, GLCM, GLRLM, GLSZM, GLDM, and NGTDM – was evaluated using the intra-class correlation coefficient (ICC) and mean differences between segmentation raters. In addition to the robustness of each single feature, an overall index for each feature category was quantified. The analysis showed that the level of quantization (i.e., the 'bincount' parameter) plays a key role in defining robust features: in fact, in our study focused on a dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) dataset of 111 breast masses,

sets with cardinality varying between 34 and 43 robust features were obtained with `binCount' values equal to 256 and 32, respectively. Moreover, both manual segmentation methods demonstrated good reliability and agreement, while automated segmentation achieved lower ICC values. Considering the dependence on the quantization level, taking into account only the intersection subset among all the values of 'binCount' could be the best selection strategy. Among radiomic feature categories, GLCM, GLRLM, and GLDM showed the best overall robustness by varying segmentation methods.

Concluding Remarks

In conclusion, radiomics enables considerable support for clinical routines by providing tools that can aid clinicians' decision-making pipeline at various levels, from diagnosis to prognosis, as well as treatment support. Automatic ROI detection, reduction of intra/inter-reader variability, ensuring repeatability of results, and optimal setting in order to extract robust and descriptive features are just some of the issues that need to be addressed

These issues have been carefully addressed by the Image Biomarker Standardization Initiative [13], which aimed at providing standardized definitions, and recommended how to implement the different steps of a radiomic workflow, including data conversion in standardized units, post-acquisition image processing, image segmentation, data interpolation, resegmentation (i.e., procedure that involves only the pixels within a specified gray-value range for radiomic feature calculation within the ROI), and intensity quantization. However, the scientific community has to devote further attention to effectively translating this effort into the clinical routine.

References

[1] Rundo, L., Militello, C., Vitabile, S., Russo, G., Sala, E., & Gilardi, M. C. (2020). A survey on nature-inspired medical image analysis: a step further in biomedical data integration. Fundamenta Informaticae, 171(1-4), 345-365.
[2] Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., ... & Sardanelli, F. (2021). AI applications to medical images: From machine learning to deep learning. Physica Medica, 83, 9-24.
[3] Rundo, L., Militello, C., Tangherloni, A., Russo, G., Vitabile, S., Gilardi, M. C., & Mauri, G. (2018). NeXt for neuro-radiosurgery: a fully automatic approach for necrosis extraction in brain tumor MRI using an unsupervised machine learning technique. International Journal of Imaging Systems and Technology, 28(1), 21-37.
[4] Rundo, L., Beer, L., Ursprung, S., Martin-Gonzalez, P., Markowetz, F., Brenton, J. D., ... & Woitek, R. (2020). Tissue-specific and interpretable sub-segmentation of whole tumour burden on CT images by unsupervised fuzzy clustering. Computers in biology and medicine, 120, 103751.
[5] Militello, C., Rundo, L., Dimarco, M., Orlando, A., Woitek, R., D'Angelo, I., ... & Bartolotta, T. V. (2022). 3D DCE-MRI radiomic analysis for malignant lesion prediction in breast cancer patients. Academic Radiology, 29(6), 830-840.
[6] Rundo, L., Beer, L., Sanchez, L. E., Crispin-Ortuzar, M., Reinius, M., McCague, C., ... & Woitek, R. (2022). Clinically interpretable radiomics-based prediction of histopathologic response to neoadjuvant chemotherapy in high-grade serous ovarian carcinoma. Frontiers in oncology, 2423.
[7] Rundo, L., Han, C., Nagano, Y., Zhang, J., Hataya, R., Militello, C., ... & Cazzaniga, P. (2019). USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. Neurocomputing, 365, 31-43.
[8] Yeung, M., Sala, E., Schönlieb, C. B., & Rundo, L. (2021). Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy. Computers in biology and medicine, 137, 104815.
[9] Yeung, M., Sala, E., Schönlieb, C. B., & Rundo, L. (2022). Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Computerized Medical Imaging and Graphics, 95, 102026.
[10] Militello, C., Rundo, L., Dimarco, M., Orlando, A., Conti, V., Woitek, R., ... & Russo, G. (2022). Semi-automated and interactive segmentation of contrast-enhancing masses on breast DCE-MRI using spatial fuzzy clustering. Biomedical Signal Processing and Control, 71, 103113.
[11] de Farias, E. C., Di Noia, C., Han, C., Sala, E., Castelli, M., & Rundo, L. (2021). Impact of GAN-based lesion-focused medical image super-resolution on the robustness of radiomic features. Scientific reports, 11(1), 1-12.
[12] Militello, C., Rundo, L., Dimarco, M., Orlando, A., D'Angelo, I., Conti, V., & Bartolotta, T. V. (2022). Robustness Analysis of DCE-MRI-Derived Radiomic Features in Breast Masses: Assessing Quantization Levels and Segmentation Agreement. Applied Sciences, 12(11), 5512.
[13] Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J., Andrearczyk, V., Apte, A., ... & Löck, S. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology, 295(2), 328-338.