



Multimodality Imaging and Artificial Intelligence for Tumor Characterization: Current Status and Future Perspective

Jérémy Dana, Vincent Agnus, Farid Ouhmich, Benoit Gallix

► To cite this version:

Jérémy Dana, Vincent Agnus, Farid Ouhmich, Benoit Gallix. Multimodality Imaging and Artificial Intelligence for Tumor Characterization: Current Status and Future Perspective. *Seminars in Nuclear Medicine*, 2020, 50, pp.541 - 548. <10.1053/j.semnuclmed.2020.07.003>. <hal-03493132>

HAL Id: hal-03493132

<https://hal.science/hal-03493132v1>

Submitted on 17 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

**Multimodality imaging and artificial intelligence for tumor characterization:
current status and future perspective**

Dana J ^{1, 2, 3}, Agnus V ^{1, 4}, Ouhmich F ^{1, 4}, Gallix B ^{1, 4, 5, 6}

¹ IHU of Strasbourg, France

² Inserm & University of Strasbourg UMR-S1110, France

³ Faculty of Medicine, University of Paris, France

⁴ Icube Laboratory, University of Strasbourg, France

⁵ Faculty of Medicine, University of Strasbourg, France

⁶ Faculty of Medicine, McGill University, Canada

Contact information:

Benoit GALLIX

IHU of Strasbourg,

1 place de l'Hôpital, 67000, Strasbourg, France

Benoit.gallix@ihu-strasbourg.eu

+33 6 19 66 06 13

Abstract

Research in medical imaging has yet to do to achieve precision oncology. Over the past thirty years, only the simplest imaging biomarkers (RECIST, SUV,...) have become widespread clinical tools. This may be due to our inability to accurately characterize tumors and monitor intra-tumoral changes in imaging. Artificial Intelligence, through machine learning and deep learning, opens a new path in medical research because it can bring together a large amount of heterogeneous data into the same analysis to reach a single outcome. Supervised or unsupervised learning may lead to new paradigms by identifying unrevealed structural patterns across data. Deep Learning will provide human-free, undefined upstream, reproducible and automated quantitative imaging biomarkers. Since tumor phenotype is driven by its genotype and thus indirectly defines tumoral progression, tumor characterization using machine learning and deep learning algorithms will allow us to monitor molecular expression non-invasively, anticipate therapeutic failure and lead therapeutic management. To follow this path, quality standards have to be set: standardization of imaging acquisition as it has been done in the field of biology, transparency of the model development as it should be reproducible by different institutions, validation and testing through a high-quality process using large and complex open databases and better interpretability of these algorithms.

Introduction

Artificial intelligence (AI) is a widely spread term referring to different fields and leading to different objectives. Medicine and patient care are at the dawn of a revolution. Future is personalized medicine, from diagnosis to treatment, and machine learning (ML) will be part of it because it can learn without explicit programming.¹ Because of the obvious relatively large amount of images and its impact in oncology, research in medical imaging has been one of the first to explore this new tool. Indeed, it is

known that tumor phenotype is driven by its genotype and can be assessed by the multiple imaging modalities, morphological and functional. Machine learning will help medical imaging analysis for tumor detection, segmentation, characterization, treatment and follow-up. Certainly, human assessment is precious and can evaluate different tumor, qualitative and semi-quantitative, features (size, shape, calcifications, necrosis, etc.). These features are part of the medical lexicon and are called “semantic”. However, semantic features are time-consuming and tend to be subjective and poorly reproducible. Therefore, their use remains limited. As opposed to semantic features, ML-based imaging biomarkers are quantitative, reproducible and automatically measurable. With the emergence of molecular targeting therapeutics, we urgently need accurate tools to propose the most suited treatments. Furthermore, molecular expression in tumor can change under treatment. Multimodal imaging may help in monitoring molecular expression non-invasively, anticipating therapeutic failure and leading therapeutic management.

To reach this objective, AI-algorithms will help in identifying new image biomarkers, inaccessible to human eyes. These new image biomarkers will reflect imaging tumor phenotype and indirectly its genotype. We aim to expose in this paper the state-of-the-art and future perspectives of AI driven by multimodality imaging.

Complex and reproducible imaging biomarkers beyond visible

In the era of personalized treatments (molecular targeted therapies, Immunotherapy, loco-regional therapy, etc.), the international consensus in oncology still advocates monitoring solid neoplasia, as varied as they are, according to the RECIST (Response Evaluation Criteria in Solid Tumors) criteria. These criteria are based on the sum of measurements of the longest axis of tumor lesions, chosen arbitrarily, without considering volume, shape, morphology, vascularity, and even less the internal structure reflecting cellular metabolism. If nuclear medicine may be an added value in the follow-up of solid

neoplasia (Standardize Uptake Value), its use remains limited. This may be due to our inability to accurately characterize tumors and monitor intra-tumoral changes in imaging. We need new parameters, human-engineered or free, to go beyond the visible image and more sophisticated statistical analysis to choose the most impacting features. Machine learning meets all these requirements. This approach consists of extracting patterns from a set of data in order to make predictions based on statistics (Figure 1). In the case of medical image analysis, these patterns are called imaging biomarkers.² This extraction can be performed based on hand-crafted descriptive mathematical models or directly learned from the images without any human intervention. These imaging biomarkers can be used in two types of machine learning algorithms, the supervised and the non supervised.

Supervised learning consists in building a predictive model thanks to outputs already known and labelled by the physician. One of the main applications lies in classification issues, which in oncology and tumor characterization can include the prediction of tumor grade (or tumor differentiation), molecular expression, risk of recurrence or even survival. Specific molecular expressions are impacting prognosis and treatment by inducing resistance profile or allowing targeted treatments (HER, KRAS, IDH,...). Different classifiers exist and can be more or less efficient according to the hypothesis-driven research. The most popular classifiers are random forests, support vector machines (SVM) and convolutional neural networks (CNN). On the contrary, unsupervised learning will apply to unlabelled data. The purpose of unsupervised learning will be to reveal new medical paradigms by identifying new structures in the data. A common application lies in clustering data and/or estimate its probability density. Unsupervised learning may help to group patients with different expressions of a same disease which can lead to a better understanding of it. There is a wide variety of algorithms, from the most classical K-means, through self-organizing maps to neural networks (auto-encoder).

Radiomics

Radiomics are non-invasive, reproducible and automatically calculated quantitative features that are supposed to reflect the heterogeneity of the tumor phenotype and thus indirectly its genotype. Radiomics corresponded to human-engineered and mathematically defined image descriptors either simple, such as size or shape, to more complex: first-order features based on intensity voxel histogram statistics, second-order or texture features (Gray-level co-occurrence matrix, gray-level run-length matrix,...) reflecting the spatial relationship between voxel values or higher-order statistics (fractals, wavelets,...) representing more complex patterns.

Recently, a new approach allowed us to extract new human-free image biomarkers, mathematically undefined upstream and sometimes inaccessible to our understanding.³ Some authors have even used the term “deep radiomics” by analogy to Deep learning and the use of complex neural networks.⁴ On the contrary of “traditional” hand-crafted radiomics, they are free of human intervention and can identify new representations and the most informative properties of the image to solve research hypotheses. Deep Learning is one of the aspects of machine learning using non-linear transformations based on convolutional neural networks imagined from the human neurons. Hidden layers are used to complexify the CNN model in order to extract and pool neural features with different levels of data abstraction. As the human brain, neural network can adjust its parameters to optimize its predictions by reducing the loss function (or error). This process is called back-propagation. The explainability of Deep learning-based features is very limited for the moment.

How to meet the need of a large amount of data?

Paradoxically, ML may be limited in medical imaging by the lack of data. Indeed, simple models such as linear models or SVMs can be built models with few parameters. They are therefore easier to learn and require less data. On the other hand, these models are often too simple to describe class distributions. This complexity can be learned with neural networks which contain millions of parameters. Their

learning requires a large amount of data. These millions of parameters can exactly record the complete dataset, which is called overfitting (Figure 2). It is therefore necessary during the training to make sure that there is enough data to avoid this phenomenon.

Several techniques exist to overcome the data limitation. First, a widespread computer science technique called “Data Augmentation” can be applied. This corresponds to the artificial creation of new data from the original dataset. The difficulty lies in respecting the original data. Especially in oncology and medical imaging, it can be particularly hazardous to create new data. Therefore, only simple geometric transformations of the image should be recommended. These can be rotated, mirrored or translated. A second technique lies in the existence of common images features. A second type of data augmentation is the generation of synthetic data using neural networks known as Generative Adversarial Network.

Because deep learning requires large databases, it can be smart to pre-train a learning model. It is not necessary to use medical imaging databases and large databases of nature or animal images can be used. As an example, ImageNet, a publicly accessible database, is commonly used to pre-train deep neural networks. The re-use of this pre-trained model is called transfer learning. The first learning layers will come from the pre-training, mimicking the visual primitive system, and this pre-trained model will be re-trained, or fine-tuned, on the study database.

Validation and testing

Two essential steps in the development of diagnostic and characterization algorithms are validation and testing (or external validation). The objective of these steps is to optimize the training of a model, not to obtain the best diagnostic performance on the training dataset, but to allow a generalization of the model to patient populations different from the training population.

The validation step allows the parameters of the training model to be adjusted to avoid overfitting the training data. It is obvious that reaching diagnostic performances on the training set alone is quite simple. At best, these performances will be overestimated. At worst, it will simply be wrong. In order not to sacrifice part of the training data at the validation step, it is common to use a technique called cross-validation. This technique consists in dividing the training set into k groups (usually 5 or 10), selecting one of these groups, training the model on the other $k-1$ groups and then validating on the selected group. This operation can be repeated k times. Especially in medicine, when creating these subgroups, it is necessary to ensure that percentages of classes are the same as that of the overall population.

The test step allows to evaluate the performances of the selected parameters. To do this, the dataset allocated to the test must be independent of the first dataset, ideally from an external database. It should never be used to train the model and it is therefore imperative to keep a strict separation between the training and test datasets. In a practical way, the model has to be locked before being tested on the external base without the possibility of a new learning iteration.

Tumor characterization by multimodality imaging

Medical imaging is rich in its diversity (X-ray, CT scan, ultrasound, MRI, metabolic imaging) and each technique produces multiple and complementary image datasets. If the diversity and complementarity of its modalities have proven their importance in patient care, they are different information for human but also for machine learning. The analysis, during the same diagnostic or therapeutic process, of multiple imaging modalities repeated over time, is a difficult objective to achieve. To do so, harmonization between the imaging techniques must be reached.

Tumor progression over time and response to treatment by automated methods should provide valuable information. This unstructured temporal integration of data is a new dimension for machine learning.

Figure 3 illustrates the potential impacts of AI on patient management.

Malignancy Risk Assessment

First objective in oncology is usually to determine the benignity or malignancy of a lesion. It may be a very challenging task with sometimes limited answers and unnecessary follow-ups. As already mentioned, semantic features are subjective and poorly reproducible leading to imperfect “Reporting and Data Systems” for assessing the risk of malignancy. Using “data augmentation”, transfer learning and neural features extraction techniques from ultrasound images, Chi et al.⁵ proposed an algorithm based on a random forest classifier with higher performance than TI-RADS, a US-based widespread malignancy risk stratification system, in distinguishing benign from malignant thyroid nodules. Thoracic oncology has also been a major field of interest. Indeed, diagnostic issues are encountered for infra-centimetric pulmonary nodules while non-enhanced CT lung contrast offers interesting possibilities for machine learning. These diagnostic issues are encountered since metabolic imaging techniques such as PET-scan are non-contributory and growth over time is the only diagnostic tool available. However, follow-up modalities and screening programs are still debated in terms of reduction of mortality.⁶ In this context, Lui et al.⁷ proposed a semantic (density and margins) and radiomics features-based nomogram to assess the risk of malignancy of small lung nodules. Semantic features slightly improved the diagnostic performance of the radiomics signature. Finally, it can be noted that interest has also been shown in pancreatic intraductal papillary mucinous neoplasms⁸ or in liver masses⁹ to achieve the same objective of assessing the risk of malignancy.

It is interesting to highlight that, so far, many machine learning studies have focused on this objective without achieving sufficient clinical impact to be used in practice.

Tumor characterization for precision oncology

Accurate non-invasive tumor characterization is the key for stratifying prognosis, predicting treatment response and optimizing patient management. It is supported by the proven assumption that imaging tumor phenotype is driven by its genotype and indirectly determines its evolution and our management. It starts by defining the histological subtype. This task was evaluated by Yasaka et al.⁹ between hepatocellular carcinoma and other liver malignancies using CT-scan. As it is the case from the physician's point of view, MRI should be superior to CT-scan to accomplish this task using machine learning. Hamm et al.¹⁰ developed a Deep Learning model based on multiphase contrast-enhanced T1-weighted imaging to discriminate hepatocellular carcinoma, intra-hepatic cholangiocarcinoma, colorectal metastasis, focal nodular hyperplasia, hemangioma and cyst achieving an accuracy of 0.92. The diagnostic performance of this model outperformed two radiologists' review (respective accuracy of 0.80 and 0.85). Hepatocellular carcinoma was diagnosed by the model with a sensitivity of 0.94 and a specificity of 0.98.

Patient management can also start by the discovery of metastatic lesions without the knowledge of the primary neoplasia. If it is sometimes possible to guide the diagnostic investigations of the primary lesion, they are usually exhaustive due to the absence of diagnostic orientation leading to a delayed therapeutic management. In a retrospective study including patients with brain metastases secondary to breast, lung, gastrointestinal cancers and melanoma, Kniep et al.¹¹ designed a multiparametric MRI-based (T1-weighted contrast material-enhanced, T1-weighted nonenhanced, and fluid-attenuated inversion recovery) model to discriminate the tumor type of the brain metastases and guide the diagnostic investigations of the primary lesion. Using radiomics features and a random forest classifier, they achieved higher diagnostic performance than radiologist but still relatively modest with an AUC value of 0.64 for non-small cell lung cancer. The highest diagnostic performance was observed for melanoma with an AUC value of 0.82 with a statistically significant difference compared to radiologist. The explication of such results for melanoma may lie in the presence of increased T1 signal areas (melanin)

insufficiently identified by the radiologist. Non-enhanced T1 weighted imaging first order maximum was in the top 10 most important radiomics features for melanoma. This could reflect the superiority of automated and reproducible quantitative radiomics features over semantic features.

Secondly, it is known that prognosis is well correlated with tumor grade. Thus, rapidity of treatment initiation should take tumor grade into account. When biopsy is not yet available, anticipating tumor grade with non-invasive imaging could impact therapeutic management and, for example, indicate neoadjuvant therapy. It is also important to note that tissue sample biopsy may not appreciate certain higher-grade focal areas within the tumor. Machine Learning may supplement this limitation. For example, if tumor grade of neuroendocrine pancreatic tumor is still defined according to ki-67 index, it can be well predicted by a machine learning nomogram.¹²

Whether treatment in oncology relies more and more on targeted therapies, pathology of tissue sample is still required to evaluate the expression of these targets within the tumor. Yet, tumor may be heterogenous with different areas with distinct molecular characteristics. Knowing the genetic status of the tumor through a virtual biopsy could aid pre-treatment decision-making. Genetic characterization of tumor by radiomics is called radiogenomics. Aerts et al. were among the first to identify CT-based radiomics features in lung and head-and-neck carcinoma associated with the underlying gene-expression patterns by reflecting tumoral heterogeneity¹³. These radiogenomics features outperformed TNM classification for predicting survival and may impact therapeutic management. Similar results have been reported in non-small-cell lung cancer using a CT-based deep learning model in comparison with standard methods such as TNM classification.¹⁴ In thoracic oncology, several molecular targeted therapies exist such as tyrosine kinase inhibitor (TKI)-sensitive mutations of the epidermal growth factor receptor (EGFR), ALK, ROS1 or MET genes. Presence of TKI-sensitive mutations of EGFR has already been the focus of studies which have shown that it can be predicted with high accuracy by ML algorithms.¹⁵ In Jia et al.'s study¹⁵, predicting TKI-sensitive mutations of EGFR using CT-based machine learning

algorithms had benefited from clinical and biological data. When clinical features (sex and smoking history) were added to the model, diagnostic performance were slightly improved. Thoracic oncology has not been the only field of interest. In Neuro-oncology, glioma can present with different molecular profiles impacting prognosis and driving patient management and treatments. Predicting tumor grade and mutational status of 1p19q, IDH1, MGMT and ATRX may be achieved with ^{18}F -FET PET-MRI-based ML algorithms.¹⁶

Finally, from tumor characterization results a prognosis and a probability of response to treatment. In addition to prognosis, risk of local recurrence can also be appreciated by machine learning models such as in patients with hepatocellular carcinoma on cirrhosis after local treatment.¹⁷

Obviously, therapeutic choices also depend on the extension of the tumor, from local invasion to lymph node status. Bladder cancer prognosis clearly correlates with bladder muscle invasion requiring a radical cystectomy instead of a transurethral resection. Preoperative accurate assessment of muscular invasion would prevent under or overtreatment. Combining clinical with radiomics features from T2 weighted MRI, Zheng et al.¹⁸ developed a highly performant nomogram for the preoperative assessment of muscular invasiveness with an AUC value of 0.88. On another note, some authors demonstrated that non-invasive imaging can also accurately predict deep myometrial invasive and lympho-vascular space invasion of endometrial carcinoma using MRI-based machine learning algorithms.¹⁹ The same is true for microvascular invasion in hepatocellular carcinoma, a difficult preoperative assessment, using contrast-enhanced US²⁰ or MRI-based^{21,22} algorithms.

The preoperative prediction of lymph nodes status has also been a major field of interest as much in breast cancer using ultrasound²³ as in colorectal carcinoma with CT-scan²⁴. While lymph node metastasis and extra-nodal extension may change operative planification or indicate adjuvant treatments in locally advanced cancer, preoperative assessment of extra-nodal extension remains poor. In response to this issue, Kann et al.²⁵ developed and validated across different institutions, a deep radiomics CT-based

algorithm achieving high performance in predicting extra-nodal extension in head-and-neck squamous cell carcinoma with an AUC of 0.84.

Response to treatment

Prediction of response is another challenging objective. As in breast cancer, pathological complete response to neoadjuvant chemotherapy is a major prognostic factor in oncology. Pre-therapeutic prediction would be a significant added value for patient management. This task was performed by Li et al.²⁶ focusing on tumor volume in a retrospective study of breast cancer patients prior to neoadjuvant chemotherapy. Peritumoral environment should also provide relevant prognostic information as it is illustrated by different immune score quantification developed in non-squamous non-small cell lung²⁷, colon²⁸ or gastric cancer²⁹. Jiang et al. designed a radiomics CT-based model predictor of the immune-score of gastric cancer that was significantly associated with disease-free and overall survival.³⁰ Peritumoral T-cell immune environment targeting specific antigens at the surface of the tumor cells is also the key of the efficacy of immunotherapy. Sun et al. designed a radiomics model to predict CD8 T-cell infiltration as an image biomarker for good response to immunotherapy.³¹ A high baseline radiomics score was associated with improved overall survival. These results are consistent with other radiomics model developed in patients with metastatic melanoma and non-small-cell lung cancer treated by PD-L1 immunotherapy for predicting response.^{32,33}

As explained above, RECIST criteria are still advocated despite its limitation. Machine learning should help to detect treatment failure at an early stage by monitoring intra-tumoral changes reflecting genotypic modifications. In patients with unresectable hepatic metastases of colorectal cancer treated with FOLFIRI and bevacizumab, Dohan et al.³⁴ developed a radiomics score for early prediction of good responders. It supplemented standard evaluation as it was able to predict a poor outcome at 2 months with the same performance as RECIST 1.1 at 6 months. Other authors reported interesting results for the

prediction of complete pathological response in triple-negative breast cancer at pre-treatment MRI based on Kurtosis, a traditional radiomics feature.³⁵

Furthermore, response evaluation of solid tumor under immunotherapy is particularly challenging in imaging with the concept of pseudo-progression that we are unable to differentiate from real progression, resulting in a delay in therapeutic management and the continuation of ineffective treatment. Tumor characterization for identifying pseudo-progression still needs to be studied.

Like the concept of pseudo-progression under immunotherapy, glioblastoma under Temozolomide, an alkylating agent, in association with radiotherapy may demonstrate pseudo-progression up to several weeks after the end of treatment. This pseudo-progression mimic true progression and diagnosis is usually made on spontaneous improvement or stabilization of imaging findings over several months. Therapeutic consequences can be important. According to Akbari et al., pseudo-progression has distinctive MRI-based radiomics features that could help for patient management.³⁶

Radiation therapy may also cause radiation injury regardless of the underlying tumor type resulting in new contrast enhancement. Thus, differential diagnosis with tumor recurrence may be challenging and impacts patient management. Subject to the small number of patients (52) and the lack of external testing, Lohmann et al.³⁷ support the contribution of radiomics features in distinguishing radiation injury from recurrent brain metastasis. The radiomics model using combined contrast-enhanced MRI and O-(2-[18F]fluoroethyl)-L-tyrosine PET-based features outperformed single-modality models (PET or MRI) reinforcing the interest of associating morphological and functional imaging modalities.

Perspectives

To be an integral part of medical imaging and patient management, several challenges remain. As explained, medical imaging data remains rare for Deep Learning requirements.³⁸ As long as large and complex databases from different institutions are not available, “Data augmentation/generation” and

“Transfer Learning” will serve as powerful tools but hardly compensate for the lack of data. Thus, it seems obvious that the quality of a machine learning model depends on the training database. It is therefore essential to use high quality images³⁹ and standardize imaging acquisition protocols which is even more necessary for multimodal algorithms. Also, training and validation databases must represent the full spectrum of the disease in order to make the algorithms generalizable and robust. The evaluation of the diagnostic performance of the model must therefore be carried out in an external and independent population to ensure this generalizability. However, an external and independent test of the model is frequently missing. Secondly, machine learning algorithms should be reproducible. Authors should use open source code packages (e.g. Pyradiomics) to standardize the extraction of data from medical images. Other research institutions should be able to reproduce any of the published machine learning models. The trained network or a network of identical architecture with the same training database and the same initialization parameters should be shared. Unfortunately, deep learning algorithms are not easily reproducible. Unlike traditional radiomics features, mathematically defined, deep radiomics features are usually represented by the concept of “black box”. They result from non-linear transformations and the deeper the convolutional neural network, the more difficult it becomes to interpret the deep radiomics from a physio-pathological point of view. A final condition for bringing machine learning into tomorrow's medicine is to reliably automate the segmentation of lesions. Segmentation using CNN (as U-net or cascaded architecture) are already proposed.⁴⁰ Most published studies rely on manual segmentation, implicating inter-observer variability or, at best, semi-automatic segmentation.

To meet reference quality standards, Lambin et al.⁴¹ proposed a radiomics quality score. In a recent review of radiomics in hepatocellular carcinoma using this score, all studies but one were scored below 18/36 (50%). Main reasons were the retrospective design, the lack of validation and open-access

scientific data resources. To guide authors, a Checklist for Artificial Intelligence in Medical Imaging (CLAIM) has also been proposed.⁴²

On another note, machine learning would benefit from an exhaustive exploitation of multimodal imaging techniques. Indeed, ultrasound remains the least studied imaging technique. A few reasons can be advanced. The main limitation lies in the complete absence of standardized acquisition. However, ultrasound brings real-time kinetics, elastography or doppler data. Regarding the published MRI-based algorithms, they usually do not associate different sequences losing multiple information on tissue characterization. Combining metabolic and molecular imaging modalities (such as ¹⁸FET PET-scanner) with conventional CT-scan or MRI will also contribute to tumor characterization and better understanding of underlying molecular mechanisms. As an example, comparison of metabolic imaging with Diffusion Weighted MRI is already useful for assessing cell density and proliferation. Furthermore, machine learning will learn from the different metabolic radiotracers reflecting intratumoral metabolism and heterogeneity.

At the difference of physicians who benefit from the interpretation of previous examinations, no studies have included temporality in the machine learning algorithms. However, changes in size, shape, limitation, enhancement and heterogeneity are crucial information for tumor characterisation.

In our opinion, machine learning will mostly prove helpful in the assessment of tumor response. Early detection of non-response is crucial to rapidly adapt therapeutic management and propose new treatments. The tumor phenotype, driven by the tumor genome, and molecular expression define the indications for targeted treatments and immunotherapy. As these tumor characteristics can change under treatment, machine learning can be used to monitor the expression of tumor targets, detect phenotypic changes and thus adapt treatments early. The contribution of medical imaging to personalized medicine will rely largely on the automation of image analysis through machine learning methods.

Conclusion

For nearly thirty years, multimodal cross-sectional imaging has been attempting to design reproducible and high-performance biomarkers. Only the simplest imaging biomarkers (RECIST criteria, SUV,...) have become widespread clinical tools. The advent of artificial intelligence brings new paradigms by identifying structural patterns across large and heterogenous data. However, automated image analysis using these new biomarkers will only become part of the clinical practice under several conditions: standardization of imaging acquisition as it has been done in the field of biology, transparency of the model development as it should be reproducible by different institutions, validation and testing through a high-quality process using large and complex open databases and better interpretability of these algorithms.

Figure 1 – Artificial Intelligence (machine learning and deep learning) processing according to quality standards. Clinical, biological and imaging data should be divided into two strictly distinct sets: training/validation and test. Training/validation dataset should be used to design the model. Only then will the performance metrics of this AI model be evaluated on the test dataset.

Figure 2 – Illustration of different two-dimensional decision boundaries: from (A) too simple or underfitting through (B) well-balanced complexity to (C) too complex or overfitting (C).

Figure 3 – Tumor characterization by AI impacts oncology patient management. Compared to human-based management, AI can predict the tumor progression profile by stratifying its prognosis, optimize treatment modalities for better performance and allows early detection of bad and good responders to treatment.

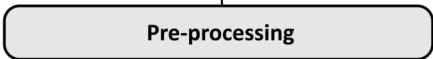
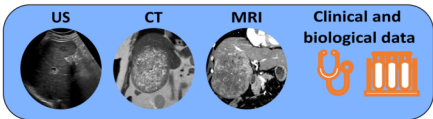
References

1. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nature Reviews Cancer*. 2018;18(8):500-510. doi:10.1038/s41568-018-0016-5
2. Savadjiev P, Chong J, Dohan A, et al. Image-based biomarkers for solid tumor quantification. *Eur Radiol*. Published online April 8, 2019. doi:10.1007/s00330-019-06169-w
3. Chartrand G, Cheng PM, Vorontsov E, et al. Deep Learning: A Primer for Radiologists. *Radiographics*. 2017;37(7):2113-2131. doi:10.1148/rg.2017170077
4. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. doi:10.1038/nature14539
5. Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. *J Digit Imaging*. 2017;30(4):477-486. doi:10.1007/s10278-017-9997-y
6. Oudkerk M, Devaraj A, Vliegenthart R, et al. European position statement on lung cancer screening. *The Lancet Oncology*. 2017;18(12):e754-e766. doi:10.1016/S1470-2045(17)30861-6
7. Liu Q, Huang Y, Chen H, Liu Y, Liang R, Zeng Q. The development and validation of a radiomic nomogram for the preoperative prediction of lung adenocarcinoma. *BMC Cancer*. 2020;20(1):533. doi:10.1186/s12885-020-07017-7
8. Hanania AN, Bantis LE, Feng Z, et al. Quantitative imaging to evaluate malignant potential of IPMNs. *Oncotarget*. 2016;7(52):85776-85784. doi:10.18632/oncotarget.11769
9. Yasaka K, Akai H, Abe O, Kiryu S. Deep Learning with Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-enhanced CT: A Preliminary Study. *Radiology*. 2017;286(3):887-896. doi:10.1148/radiol.2017170706
10. Hamm CA, Wang CJ, Savic LJ, et al. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. *Eur Radiol*. 2019;29(7):3338-3347. doi:10.1007/s00330-019-06205-9
11. Kniep HC, Madesta F, Schneider T, et al. Radiomics of Brain MRI: Utility in Prediction of Metastatic Tumor Type. *Radiology*. 2019;290(2):479-487. doi:10.1148/radiol.2018180946
12. Liang W, Yang P, Huang R, et al. A Combined Nomogram Model to Preoperatively Predict Histologic Grade in Pancreatic Neuroendocrine Tumors. *Clin Cancer Res*. 2019;25(2):584-594. doi:10.1158/1078-0432.CCR-18-1305
13. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006. doi:10.1038/ncomms5006
14. Hosny A, Parmar C, Coroller TP, et al. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PLOS Medicine*. 2018;15(11):e1002711. doi:10.1371/journal.pmed.1002711

15. Jia T-Y, Xiong J-F, Li X-Y, et al. Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling. *Eur Radiol.* 2019;29(9):4742-4750. doi:10.1007/s00330-019-06024-y
16. Haubold J, Demircioglu A, Gratz M, et al. Non-invasive tumor decoding and phenotyping of cerebral gliomas utilizing multiparametric 18F-FET PET-MRI and MR Fingerprinting. *Eur J Nucl Med Mol Imaging.* 2020;47(6):1435-1445. doi:10.1007/s00259-019-04602-2
17. Ji G-W, Zhu F-P, Xu Q, et al. Radiomic Features at Contrast-enhanced CT Predict Recurrence in Early Stage Hepatocellular Carcinoma: A Multi-Institutional Study. *Radiology.* 2020;294(3):568-579. doi:10.1148/radiol.2020191470
18. Zheng J, Kong J, Wu S, et al. Development of a noninvasive tool to preoperatively evaluate the muscular invasiveness of bladder cancer using a radiomics approach. *Cancer.* 2019;125(24):4388-4398. doi:10.1002/cncr.32490
19. Ueno Y, Forghani B, Forghani R, et al. Endometrial Carcinoma: MR Imaging-based Texture Model for Preoperative Risk Stratification-A Preliminary Analysis. *Radiology.* 2017;284(3):748-757. doi:10.1148/radiol.2017161950
20. Hu H-T, Wang Z, Huang X-W, et al. Ultrasound-based radiomics score: a potential biomarker for the prediction of microvascular invasion in hepatocellular carcinoma. *Eur Radiol.* 2019;29(6):2890-2901. doi:10.1007/s00330-018-5797-0
21. Banerjee S, Wang DS, Kim HJ, et al. A computed tomography radiogenomic biomarker predicts microvascular invasion and clinical outcomes in hepatocellular carcinoma. *Hepatology.* 2015;62(3):792-800. doi:10.1002/hep.27877
22. Xu X, Zhang H-L, Liu Q-P, et al. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. *J Hepatol.* 2019;70(6):1133-1144. doi:10.1016/j.jhep.2019.02.023
23. Zheng X, Yao Z, Huang Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun.* 2020;11(1):1236. doi:10.1038/s41467-020-15027-z
24. Huang Y-Q, Liang C-H, He L, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol.* 2016;34(18):2157-2164. doi:10.1200/JCO.2015.65.9128
25. Kann BH, Hicks DF, Payabvash S, et al. Multi-Institutional Validation of Deep Learning for Pretreatment Identification of Extranodal Extension in Head and Neck Squamous Cell Carcinoma. *J Clin Oncol.* 2020;38(12):1304-1311. doi:10.1200/JCO.19.02031
26. Li P, Wang X, Xu C, et al. 18F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. *Eur J Nucl Med Mol Imaging.* 2020;47(5):1116-1126. doi:10.1007/s00259-020-04684-3

27. Li B, Cui Y, Diehn M, Li R. Development and Validation of an Individualized Immune Prognostic Signature in Early-Stage Nonsquamous Non–Small Cell Lung Cancer. *JAMA Oncol.* 2017;3(11):1529-1537. doi:10.1001/jamaoncol.2017.1609
28. Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *The Lancet.* 2018;391(10135):2128-2139. doi:10.1016/S0140-6736(18)30789-X
29. Jiang Y, Zhang Q, Hu Y, et al. ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer. *Ann Surg.* 2018;267(3):504-513. doi:10.1097/SLA.0000000000002116
30. Jiang Y, Wang H, Wu J, et al. Noninvasive imaging evaluation of tumor immune microenvironment to predict outcomes in gastric cancer. *Annals of Oncology.* 2020;31(6):760-768. doi:10.1016/j.annonc.2020.03.295
31. Sun R, Limkin EJ, Vakalopoulou M, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *The Lancet Oncology.* 2018;19(9):1180-1191. doi:10.1016/S1470-2045(18)30413-3
32. Trebeschi S, Drago SG, Birkbak NJ, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Annals of Oncology.* 2019;30(6):998-1004. doi:10.1093/annonc/mdz108
33. Mu W, Tunali I, Gray JE, Qi J, Schabath MB, Gillies RJ. Radiomics of 18F-FDG PET/CT images predicts clinical benefit of advanced NSCLC patients to checkpoint blockade immunotherapy. *Eur J Nucl Med Mol Imaging.* 2020;47(5):1168-1182. doi:10.1007/s00259-019-04625-9
34. Dohan A, Gallix B, Guiu B, et al. Early evaluation using a radiomic signature of unresectable hepatic metastases to predict outcome in patients with colorectal cancer treated with FOLFIRI and bevacizumab. *Gut.* 2020;69(3):531-539. doi:10.1136/gutjnl-2018-316407
35. Chamming's F, Ueno Y, Ferré R, et al. Features from Computerized Texture Analysis of Breast Cancers at Pretreatment MR Imaging Are Associated with Response to Neoadjuvant Chemotherapy. *Radiology.* 2018;286(2):412-420. doi:10.1148/radiol.2017170143
36. Akbari H, Rathore S, Bakas S, et al. Histopathology-validated machine learning radiographic biomarker for noninvasive discrimination between true progression and pseudo-progression in glioblastoma. *Cancer.* 2020;126(11):2625-2636. doi:10.1002/cncr.32790
37. Lohmann P, Kocher M, Ceccon G, et al. Combined FET PET/MRI radiomics differentiates radiation injury from recurrent brain metastasis. *Neuroimage Clin.* 2018;20:537-542. doi:10.1016/j.nicl.2018.08.024
38. Savadjiev P, Chong J, Dohan A, et al. Demystification of AI-driven medical image interpretation: past, present and future. *Eur Radiol.* 2019;29(3):1616-1624. doi:10.1007/s00330-018-5674-x
39. Sabottke CF, Spieler BM. The Effect of Image Resolution on Deep Learning in Radiography. *Radiology: Artificial Intelligence.* 2020;2(1):e190015. doi:10.1148/ryai.2019190015

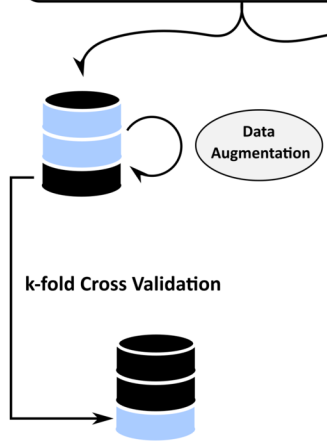
40. Ouhmich F, Agnus V, Noblet V, Heitz F, Pessaux P. Liver tissue segmentation in multiphase CT scans using cascaded convolutional neural networks. *Int J Comput Assist Radiol Surg*. 2019;14(8):1275-1284. doi:10.1007/s11548-019-01989-z
41. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762. doi:10.1038/nrclinonc.2017.141
42. Mongan J, Moy L, Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. *Radiology: Artificial Intelligence*. 2020;2(2):e200029. doi:10.1148/ryai.2020200029



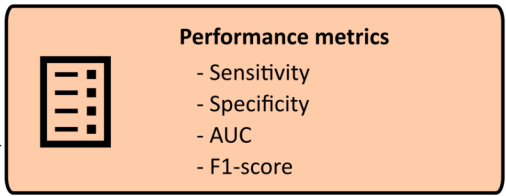
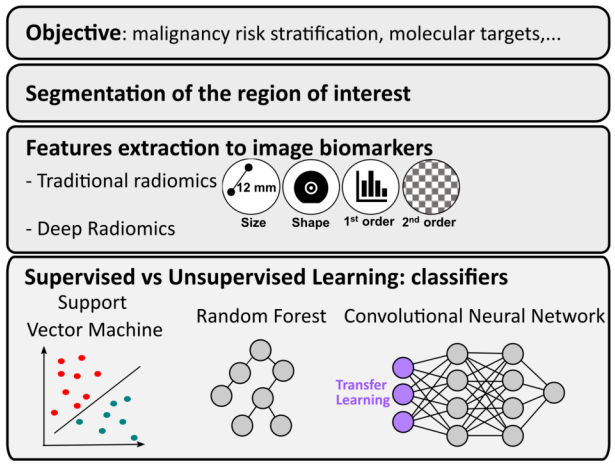
Training

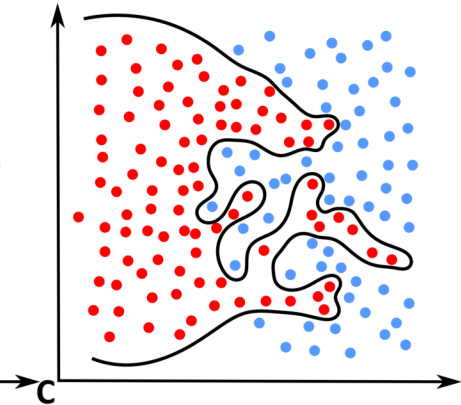
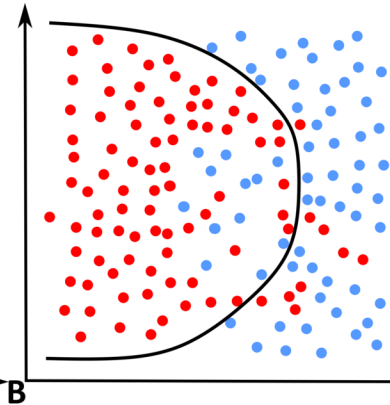
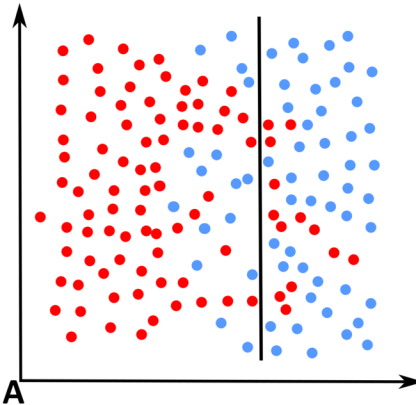
Validation

Test



Model Design





Malignancy Assessment

Early and accurate diagnosis of malignancy in order to avoid:

- Unnecessary follow-ups or treatments
- Reduce false positive

Tumor Characterization For A Prognostic Profile

Automated imaging biomarkers as a surrogate of tumor genotype and aggressivity (**Indolent** to **Aggressive**)

- Histology Subtype
- Tumor Differentiation
- Molecular expression
- Drug target

Reduce the need for biopsies

Treatment and Response Assessment

● Therapeutic Decision and Treatment

□ Imaging Response Assessment

— AI + Human-based Management

— Human-based Management (e.g. RECIST)

