Defining the biological and clinical basis of radiomics

Valorization addendum

VALORIZATION ADDENDUM

The central objective of this dissertation was to enable biological reasoning of radiomic prognostic models to facilitate the development of imaging-based biomarkers for clinical decision support. Such biomarkers will be crucial in the success of targeted therapies. As this thesis has shown, radiomics has unprecedented potential to provide actionable data for noninvasive and cost-effective biomarker for specific cancer treatments. Furthermore, this thesis has also demonstrated that advanced machine learning algorithms are necessary to provide the level accurateness and robustness required for clinical applications.

To translate research of this thesis into clinical tools, further validation of the proposed models must be performed. To this end, transparacy in the processed data and methods used is critical. Therefore, we made data and code publicly available to the wider scientific community wherever possible. In particular, we released radiomic, genomic, and clinical data, as well as open sourced our analysis code from Chapter 2 online (https://elifesciences.org/articles/23421), which is the core of our results that associate radiomic with biological data. In addition, results from Chapter 2 were extended in Chapters 3 and 4, and data underlying Chapter 3 and 4 were collected through The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov/) and The Cancer Imaging Archive (TCIA, http://www.cancerimagingarchive.net/), which are public repositories. Hence, reproducibility of those results can be assured, as well. Obviously, releasing data is not possible for all studies due to intellectual property; for instance, this was the case for data underlying Chapters 5 and 6, which analyzed data from clinical trials. Similar holds for the remaining Chapters 7-9. Hence, more efforts should be performed to achieve global data sharing.

One solution to facilitate data sharing could be distributed learning, an emerging infrastructural concept in which machine learning models are distributed to different sites (i.e., clients such as hospitals) to learn from data of each site. In this way, data never leaves a site and hence only learned model coefficients are shared across sites. Obiously, however, full transparency in underlying data can only be achieved with access to raw data. Regarding transparency in methods, three parts have to be considered: 1) Segmentation, 2) Feature extraction, and 3) Analysis. In this thesis, methods for segmenting tumors were already publicly available either commercially (e.g, Eclipse^{TM}) or open source (e.g., 3D Slicer). Methods for feature extraction were recently shared through an open-sourced Python package at pyradiomics.readthedocs.io. All algorithms used for the analyses in this thesis were implemented in 'R', an open source package for statistical analysis.

General employment of open source methods can significantly support reproducibility analysis even in clinical studies, because reasons for deviating results can more easily be deduced. This can also have an impact on standardization of radiomic methods, which also must be achieved to gain regulatory approval for clinical use. Currently, major international consortia, such as QIN, QIBA, EORTC, and ESR, are developing guidelines for standardization in image acquisition, registration, and analysis.

Results of this dissertation can also be leveraged to improve existing biomarker approaches, as our thesis also suggests complementary prognostic value of radiomics to volumetric, genomic, and clinical assessments. To achieve effective combinations of radiomics and traditional methods, one radiomic signature should be fixed first. This radiomic signature can then be combined with genomic or clinical signatures that have been extensively described in recent literature to understand how this combination can be reproducibly and robustly optimized for a predefined clinical endpoint. Ideally, training and validation of such novel combinations should be performed by independent groups, ensuring absence of data leakage and thus generating overoptimistic results. Should radiomic approaches continue to show complementary value, combined biomarkers will lead to more accurate and cost-effective diagnostic strategies, and hence safer treatments. Inherently, more accurate biomarkers can also save considerable amount of funds required to conduct a clinical trial, as the target population can be identified easier and thus less patients need to be recruited to gain statistical significance.

Major parts of this dissertation have gained broad attention for its innovative approaches. For example, parts of the study prestend in Chapter 2 were honored with the *Best in Physics* award at the Annual Meeting of the American Association of Physicists in Medicine (http://www.aapm.org/meetings/2015AM/PRAbs.asp?mid=99&aid=29354), and was also announced with an online press release about Hot Topics (http://www.aapm.org/meetings/2015AM/VirtualPressroom/NewsReleases.asp#HotTopics). In addition, the significance of this work has also been suggested by a recent press release: https://www.sciencedaily.com/releases/2017/08/170803135210.htm. Altogether this emphasizes the anticipated impact that radiomics will make in clinical decision support. Finally, it should be noted that radiomic approaches have translational potential beyond oncology; as radiomic processes standard of care imaging techniques, any disease that can routinely be assessed by medical imaging can also be a potential application of radiomics, including cardiovascular, neurodegenerative, and bone diseases.