

Combining data science and medical imaging

Citation for published version (APA):

Coroller, T. P. (2017). *Combining data science and medical imaging: Advancing cancer precision medicine with radiomics*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20171214tc>

Document status and date:

Published: 01/01/2017

DOI:

[10.26481/dis.20171214tc](https://doi.org/10.26481/dis.20171214tc)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Societal Impact and Valorizations

In this thesis, we described the use of state-of-the-art advanced imaging features in combination with data science. The goal of this method is to investigate the association between image characteristics and tumor crucial information (relating to its nature, response, or evolution). This new method could be a potential tool for precision medicine. Any progress in precision medicine will benefit patients - such as avoiding surgery and tailoring treatment more effectively. In this context, several institutions have already started sharing tools to help. Published models can freely be found online. For example, MAASTRO clinic (<http://www.predictcancer.org/>) or the Memorial Sloan Kettering Cancer Center (<https://www.mskcc.org/nomograms>) have different models, from head and neck to rectum cancers, which are investigating overall survival or recurrence. However, these models have fairly simple metrics (such as patient age or tumor size). More advanced methods have emerged in the precision medicine toolbox to investigate relatively new field of research like genomics, proteomics, and radiomics. The major difference between these new methods and the more “conventional” one is the large number of parameters (from 1000 in radiomics to 20,000 in genomics). This distinction adds complexity to new tools and is prone to increased errors.

Several solutions can then be implemented to control those new techniques glitches. Open source initiative (<https://opensource.org/>) allows users, developers, and scientists to use and share technology at no cost while improving it. This ensure that researchers can widely share ideas and readers can replicate their results, at no cost nor legal issues. Our laboratory (<http://www.cibl-harvard.org/>) has been developing an open source code called PyRadiomics (<http://www.radiomics.io/>). The code allows extraction of radiomics features from medical images, which will help new teams investigate imaging analyses without the burden of starting from scratch, as well as allowing full flexibility. This project is done with the partnership of the Slicer development team (<https://www.slicer.org/>) which have extensive experience in developing open source software. Their product is a state-of-the-art software to visualize, segment, or quantify on any medical images (CT, PET, MRI) along with highly efficient open source libraries. While Slicer was mainly used for imaging management for this thesis, we also used open software such as R (<https://www.r-project.org/>) or python (<https://www.python.org/>) that allow easy sharing across institutions and scientists.

Open source is also valuable for information itself. Along with the performance and promise of new methods and software comes the curse of dimensionality. More complex methods will require larger sample size to train model with high accuracy and generality. Thus, several foundations have already helped increase data access such as the Cancer Imaging Archive (<http://www.cancerimagingarchive.net/>) with medical imaging, the Cancer Genome Atlas (<https://cancergenome.nih.gov/>) with genomics information. These platforms equipped scientists with direct access to highly curated medical information that can be immediately used to enhance science.