Multimodality radiomics and deep learning for outcome modeling: application in head & neck cancer

Jan Seuntjens, Ph.D., FCCPM, FAAPM, FCOMP
Medical Physics Unit, McGill University
Montreal, Canada

McGill

AAPM July 2020

Centre universitaire de santé McGill / McGill University Health Centre

1

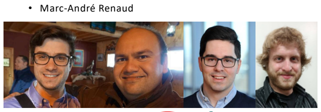
Acknowledgements

MEDomics

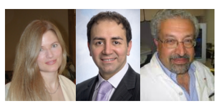
NSERC CRSNG

CIHR IRSC

- Post docs - RAs
  - Avishek Chatterjee
  - Martin Carrier-Vallières
  - Alexandre Berryman
- Former grad students
  - André Diamant-Boustead
  - Marc-André Renaud



McGill clinician collaborators



C. Reinhold R. Forghani G. Shenouda

McGill MP

International collaborators

I. Levesque

I. El Naqa, H. Lee Moffitt Cancer Center & Research Institute

O. Morin, UCSF

GRAY

Medical Physics Research Training Network

AAPM July 2020

2

Outline

- Radiomics for outcomes prediction
- Repeatability and reproducibility
- Radiomics and head & neck cancer – benchmark study
- Deep learning radiomics
- Extension to multiple modalities and clinical data
- Summary

McGill

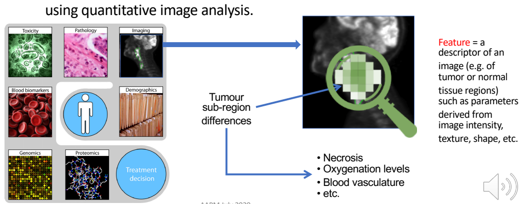
AAPM July 2020

3

Radiomics = use of "texture" information in images

→**Hypothesis 1**: The genomic heterogeneity of aggressive tumours translates into heterogeneous characteristics at the anatomical scale.

→**Hypothesis 2**: Intratumoral heterogeneity at the anatomical scale can be captured using quantitative image analysis.



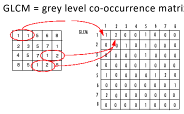
Adapted from (Lambin P *et al.*, *Eur J Cancer* **48**, 2012)

AAPM July 2020

4

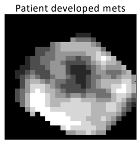
Example of *GLCM* textures

GLCM = grey level co-occurrence matrix



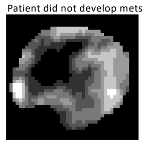
Vallières et al 2017

Patient developed mets



Energy: 0.03
Contrast: 1.61
Entropy: 3.57
Homogeneity: 0.64
Correlation: 0.80
Variance: 0.39

Patient did not develop mets

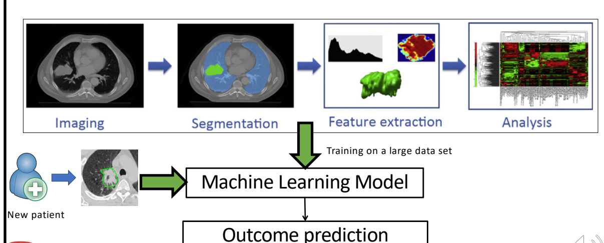


Energy: 0.06
Contrast: 0.73
Entropy: 3.09
Homogeneity: 0.74
Correlation: 0.87
Variance: 0.28

Texture analysis is concerned with the *spatial distribution (patterns) of gray level variations* within an image.

5

The radiomics world



McGill

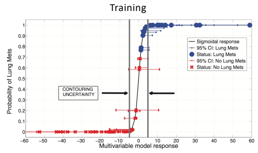
AAPM July 2020

Lambin P *et al.*, *Eur J Cancer* **48**, 2012

7

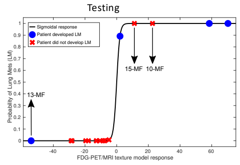
Soft tissue sarcoma – lung metastases prediction model

Training



Vallieres et al 2015 Phys. Med. Biol. 60: 5471

Testing



*Vallieres et al 2018 PHIRO **6**, 53–60*

McGill

AAPM July 2020

8

Repeatability and reproducibility

- Repeatability = measure of precision under identical or near-identical conditions and acquisition parameters
  - evaluated by "test-retest" analysis
  - 31 CT datasets Reference Image Database to Evaluate Therapy Response (RIDER)
- Reproducibility = better to assess overall robustness
  - Imaging system
  - Imaging parameters
  - Reconstruction
  - ROI delineation
  - Feature extraction and feature qualification

McGill

AAPM July 2020

10

Measuring CT scanner variability of radiomics features

Dennis Mackin, PhD[1], Xenia Fave, BS[1,2], Lifei Zhang, PhD[1], David Fried, BS[1,2], Jinzhong Yang, PhD[1], Brian Taylor, PhD[3,4], Edgardo Rodriguez-Rivera, MS[5], Cristina Dodge, PhD[6], A. Kyle Jones, PhD[7], and Laurence Court, PhD[1,7]
[1]Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030

→ CT scanner variability is large compared to the interpatient variability in the NSCLC tumors for some features.

OPEN

Reproducibility of radiomics for deciphering tumor phenotype with imaging

Received: 22 September 2015
Accepted: 04 March 2016

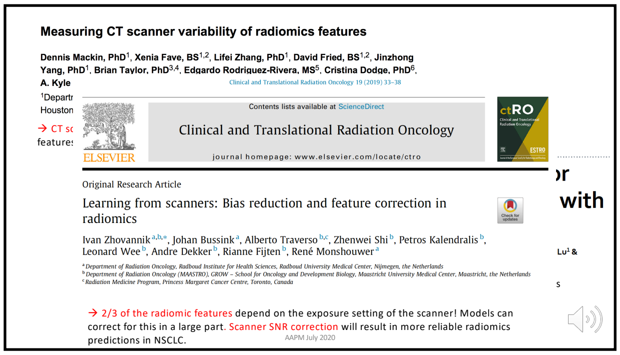
Binsheng Zhao[1], Yongqiang Tan[1], Wei-Yann Tsai[2], Jing Qi[1], Chuanmiao Xie[1], Lin Lu[1] & Lawrence H. Schwartz[1]

→ Imaging parameters that affect edge sharpness significantly affect radiomic features

McGill

AAPM July 2020

11


Measuring CT scanner variability of radiomics features
Dennis Mackin, PhD[1], Xenia Fave, BS[1,2], Lifei Zhang, PhD[1], David Fried, BS[1,2], Jinzhong Yang, PhD[1], Brian Taylor, PhD[3,4], Edgardo Rodriguez-Rivera, MS[5], Cristina Dodge, PhD[6], A. Kyle
Clinical and Translational Radiation Oncology 19 (2019) 33–38
Contents lists available at ScienceDirect
Clinical and Translational Radiation Oncology
journal homepage: www.elsevier.com/locate/ctro
ELSEVIER
Original Research Article
Learning from scanners: Bias reduction and feature correction in radiomics
Ivan Zhovannik[a,b,*], Johan Bussink[a], Alberto Traverso[b,c], Zhenwei Shi[b], Petros Kalendralis[b], Leonard Wee[b], Andre Dekker[b], Rianne Fijten[b], René Monshouwer[a]
→ 2/3 of the radiomic features depend on the exposure setting of the scanner! Models can correct for this in a large part. Scanner SNR correction will result in more reliable radiomics predictions in NSCLC.
AAPM July 2020

12

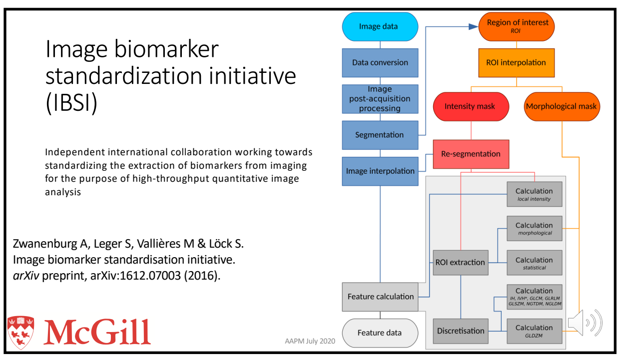
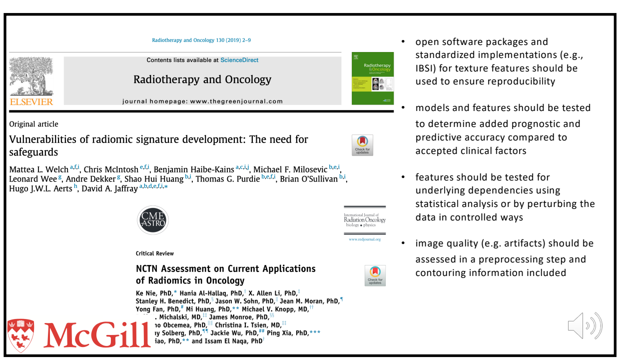


Image biomarker standardization initiative (IBSI)
Independent international collaboration working towards standardizing the extraction of biomarkers from imaging for the purpose of high-throughput quantitative image analysis
Zwanenburg A, Leger S, Vallières M & Löck S. Image biomarker standardisation initiative. arXiv preprint, arXiv:1612.07003 (2016).
Image data
Data conversion
Image post-acquisition processing
Segmentation
Image interpolation
Feature calculation
Feature data
Region of interest
ROI
ROI interpolation
Intensity mask
Morphological mask
Re-segmentation
ROI extraction
Discretisation
Calculation
Calculation
Calculation
Calculation
Calculation
McGill
AAPM July 2020

14


Radiotherapy and Oncology 130 (2019) 2–9
Contents lists available at ScienceDirect
Radiotherapy and Oncology
journal homepage: www.thegreenjournal.com
ELSEVIER
Original article
Vulnerabilities of radiomic signature development: The need for safeguards
Mattea L. Welch, Chris McIntosh, Benjamin Haibe-Kains, Michael F. Milosevic, Leonard Wee, Andre Dekker, Shao Hui Huang, Thomas G. Purdie, Brian O'Sullivan, Hugo J.W.L. Aerts, David A. Jaffray
Critical Review
NCTN Assessment on Current Applications of Radiomics in Oncology
• open software packages and standardized implementations (e.g., IBSI) for texture features should be used to ensure reproducibility
• models and features should be tested to determine added prognostic and predictive accuracy compared to accepted clinical factors
• features should be tested for underlying dependencies using statistical analysis or by perturbing the data in controlled ways
• image quality (e.g. artifacts) should be assessed in a preprocessing step and contouring information included
McGill

17

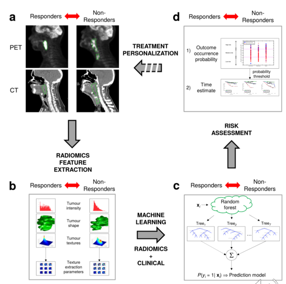
Head & neck radiomics

Benchmark study

Table 3: Patient/outcome distribution[25, 40]

	Training cohort	Validation cohort
Total	194	106
Outcome		
Distant metastasis (DM)	26 (13%)	14 (13%)
Loco-regional failure (LRF)	29 (15%)	16 (15%)
Death	32 (16%)	24 (23%)
Institution		
Hôpital général juif (HGJ)	92 (47%)	-
Centre hospitalier universitaire de Sherbooke (CHUS)	102 (53%)	-
Hôpital Maisonneuve-Rosemont (HMR)	-	41 (39%)
Centre hospitalier de l'Université de Montréal (CHUM)	-	65 (61%)
Tumor type		
Oropharynx	129 (66%)	77 (73%)
Hypopharynx	5 (3%)	7 (7%)
Nasopharynx	20 (10%)	8 (7%)
Larynx	36 (19%)	9 (8%)
Unknown	4 (2%)	5 (5%)

CANCER IMAGING ARCHIVE | Head-Neck-PET-CT | Head and Neck Cancer | PT, CT, RTSTRUCT, RTPLAN, RTDOSE | 298 | Head-Neck | Yes | Public | Complete | 2018/06/07

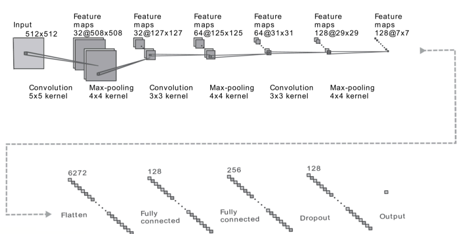


AAPM July 2020

Vallières et al 2017

18

Example of deep learning radiomics model



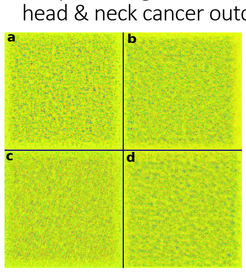
McGill

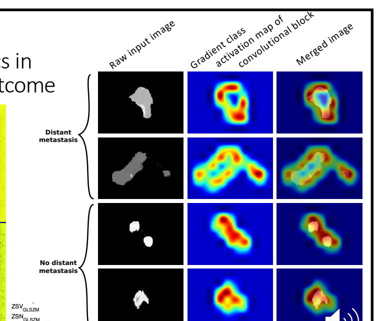
AAPM July 2020

Diamant et al 2019

20

Deep learning radiomics in head & neck cancer outcome



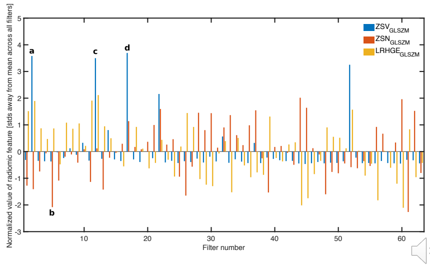


AAPM July 2020

Diamant et al, 2019

21

CNN filter activation and texture features



McGill

AAPM July 2020

Diamant, A. et al (2019)

22

Performance

Table 2: Validation set results compared to Vallières *et al.'s*[25] testing set results on the same patient cohort. Balanced accuracy is defined as the average of the specificity and sensitivity. It is noted that this study calculated specificity and sensitivity based on thresholds optimized in the training set, while the benchmark study[25] performed imbalance adjustments during training and then used a single probability threshold of 0.5 in the testing phase. DM: Distant metastasis; LRF: Loco-regional failure; OS: Overall survival.

	Specificity		**Sensitivity**		**Balanced Accuracy**	
	Present study	*Vallières et al.*[25]	*Present study*	*Vallières et al.*[25]	*Present study*	*Vallières et al.*[25]
DM	0.89	0.77	0.86	0.79	88%	77%
LRF	0.67	0.61	0.65	0.39	66%	58%
OS	0.67	0.67	0.68	0.55	68%	62%

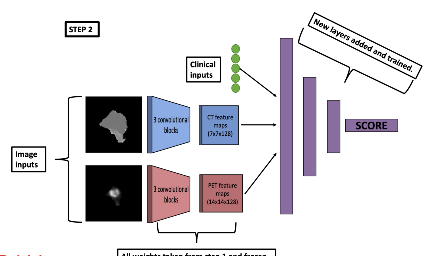
McGill

AAPM July 2020

Diamant, A. et al (2019)

23

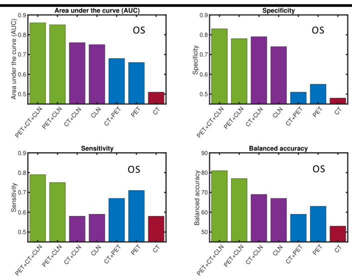
Extension to multiple imaging modalities



McGill

Diamant *et al* 2020

24



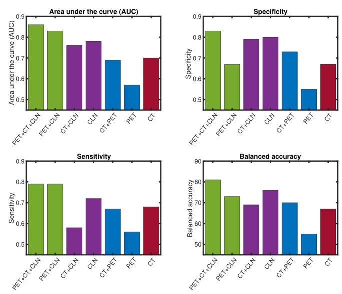
Diamant et al 2020 (preliminary)

Overall survival

Model trained with *all* data present; effect of missing data tested.

	Distant metastasis (DM)	Loco-regional failure (LRF)	Overall survival (OS)
AUC			
Method A	0.93 [0.27]	**0.78 [10^{-3}]**	**0.86 [10^{-3}]**
Method B	0.85	0.70	0.78
Diamant et al[27]	0.88	0.65	0.70

25



Overall survival

Model trained with only the data set present on the x-axis

Figure 2: Results when evaluating overall survival prediction performance on the validation set whereupon each model was trained with solely the information depicted on the *x*-axis. Balanced accuracy is described as the average of the specificity and sensitivity.

AAPM July 2020

Diamant et al 2020 (preliminary)

26

Summary

- We are only in the early stages of outcome modeling using these newer techniques, and far away from clinical implementation – data federation
- We emphasized standardization in the radiomics steps with the goal of better reproducibility
- We may build successful models but we have to recognize that there is a large variability of factors influencing clinical outcomes. We have to be careful with early generalizations.
- Outcome modeling hinges on the quality of the data. Each patient experience must be carefully documented and stored to contribute to accurate models for future patients' outcome.

McGill

AAPM July 2020

27

MEDomics

medomics.ai

https://youtu.be/2030Pdgm3_4

Synergy between medical image analysis, machine learning, deep learning, natural language processing and distributed learning

28