Design, conduct, and reporting of radiomic analyses: Let’s not reinvent the wheel

Chaya S. Moskowitz, Ph.D.
Department of Epidemiology & Biostatistics

Financial disclosures: None
Outline

• Review of radiomics literature
• Sources of variability and bias at each step of radiomic pipeline
  – Variability and biases that are frequently pointed out
  – Additional sources and pitfalls
• How to ensure study conduct and reporting standards
Reviews of radiomics literature

- Sanduleanu and colleagues: radiomics in oncology
  - 70% of reviewed papers scored < 30% of ideal quality score

- Park and colleagues: radiomics in oncology papers published in journals with impact factors > 7.0
  - Mean quality score was 26% of ideal score

- Roberts and colleagues: radiomics in COVID-19 papers
  - 88% of papers had a high risk of bias
  - None of the models had potential for clinical use

Sanduleanu et al., Radiother Oncol 2018
Park et al., Eur Radiol 2020
Roberts et al., Nat Mach Intell 2021
Radiomics analysis workflow

Aerts et al., Nat Comm 2014
Lambin et al., EJC 2012
Limkin et al., Ann Oncol 2017
Nie et al., Int J Radiat Oncol Biol Phys 2017
Shaikh et al., JCCI 2017
Trebeschi et al., Ann Oncol 2019
Liu et al., Theranostics 2019
Radiomics analysis workflow

- Scanners
- Acquisition parameters
- Processing
Radiomics analysis workflow

- Manual
- Semi-automatic
- Automatic

Imaging → Segmentation → Feature extraction → Analysis

Lambin et al., EJC 2012
Radiomics analysis workflow

- Computation methods
- Nomenclature

Lambin et al., EJC 2012
Standardization Initiatives

Clarke et al., Transl Oncol 2014
Sullivan et al., Radiology 2015
Zwanenburg et al., Radiology 2020
Radiomics analysis workflow

- Feature selection methods
- Classification and model building methods
- Validation methods (or lack thereof)

Lambin et al., EJC 2012
Frequent Sources of Bias in Radiomic Analyses

- Bias due to overfitting
- Optimistic performance bias
- Multicollinearity
- Multiple testing
Radiomics analysis workflow
Radiomics analysis workflow

- Imaging
- Segmentation
- Feature extraction
- Analysis
- Reporting & interpretation

Study design

Prospective study

Retrospective study
Study design

Elements of study design include:

- Research hypotheses
- Study objectives
- Endpoints
- Reference information
- Patient population
- Inclusion/exclusion criteria for images
- Reader study design if applicable
Potential Sources of Variability and Bias

Prospective study

Retrospective study

Study design

- Outcome / reference standard
- Patient / Image selection
- Confounding
Outcome

• In diagnostic accuracy studies, outcome = reference or gold standard
• In radiomic analysis, there might not be a reference standard for a feature or signature; the outcome could be presence of abnormal condition either at the time the image is acquired or in the future (e.g. overall or progression-free survival)
• Essential to carefully define outcome
• Definition of the outcome should not rely on information from the imaging modality producing the radiomic features
  – If it does → incorporation bias
Incorporation Bias in Diagnostic Testing Framework

- The outcome uses information from the images being analyzed.
- Occurs most often when clinical judgment is needed to determine the outcome.
- Leads to overestimating measures of accuracy.

Diagnosing Alzheimer’s Disease

- Mental status testing
- Laboratory results
- Brain scans: MRI, CT, PET

Ransohoff and Feinstein, NEJM 1978
Whiting et al., Ann Intern Med 2004
Incorporation Bias in Diagnostic Testing Framework

- The outcome uses information from the images being analyzed
- Occurs most often when clinical judgment is needed to determine the outcome
- Leads to overestimating measures of accuracy

Is MRI accurate for detecting Alzheimer’s Disease?

Equivalent to

Is MRI accurate for detecting mental status testing?

- Laboratory results
- Brain scans: MRI, CT, PET
- Alzheimer’s Disease
Incorporation Bias in Radiomics

• Bias is greatest when one or more of the features are vital for identifying the condition

• Bias may still be present if the features under study play less of a role in identifying the condition

• Difficult to understand degree of overestimation

Is MRI radiomics signature accurate for detecting Alzheimer’s Disease?

EQUIVALENT TO

Is MRI radiomics signature accurate for detecting Mental status testing

Laboratory results = Alzheimer's Disease

Brain scans: MRI, CT, PET
Who to include: Spectrum Bias

- Study data not fully representative of the population of interest
  - Patients with outcome have severe disease or health conditions that are more obvious
  - “Healthy” patients are more healthy than typical patients
- Leads to overestimating accuracy
- Example
  - Using scans from patients with osteoarthritis referred to specialty care to develop a model for use in a primary care population
- Model is affected is by differences in patient characteristics or settings (spectrum effect)

Ransohoff and Feinstein, NEJM 1978
Williams, Int Musculoskelet Med 2010
Verification Bias in Diagnostic Testing Framework

- Results from non-random selection of images
- Study is limited to images with the outcome ascertained
- Extreme verification bias: *only* images that are suspicious for disease have the outcome ascertained and are included in the study
- Partial verification bias: a non-random portion of images have the outcome ascertained
- Direction of bias is difficult to know
  - Often results in increased sensitivity
  - Can result in increased or decreased specificity
  - AUC?

Example of extreme verification bias

Example of partial verification bias

Begg and McNeil, Radiology 1988
Kohn et al., Acad Emerg Med 2013
Verification Bias in Radiomics

- Path may be more indirect than in classical sense
- Bias may be diluted, but it is still present
- Bias is likely to be greatest when one or more of the imaging features are vital for informing decision to verify outcome
- Very difficult to understand how it affects our evaluation of model performance

Mammogram
Features $f_1, f_2, \ldots, f_P$

Radiomic signature
$g(f_1, f_2, \ldots, f_P)$

+ → Biopsy

-
Bias Due To Confounding

• “Mixing of effects”

• Confounding is a distortion of the estimated association between radiomic features and the outcome that occurs because the radiomic features are entangled with another factor associated with the outcome.

• Can lead to missing an association that is present or falsely finding an association when none is present

• Important to collect information on confounders and adjust for it in the analysis
Source of Confounding in Radiomics Studies

- Different imaging protocols based on clinical factors
- Imaging differences (protocol variations or artifacts) affect feature measurement
- If they are related to clinical factors associated with the outcome, this can lead to confounding

BMI

Different contrast dose used based on BMI

Radiomic features

Outcome

Different contrast dose used based on BMI
Moving Forward

What do we need to improve the conduct and reporting of radiomic studies?

I haven’t said anything new here!
Conclusions

• Potential of radiomic studies weighed down by bias and variability

• Many sources of bias and variability have been identified in radiomics and other literature

• Involving experts from necessary disciplines can help

• Do we need guidelines for conduct and reporting specifically for radiomic studies?
Thank You!

Acknowledgments
Amber Simpson Queen’s University
Mattea Welch Princess Margaret Cancer Centre
Brenda Kurland ERT

Follow-up questions or thoughts?
e-mail: moskowc1@mskcc.org
Twitter: @ChayaSMoskowitz