





UiO **Faculty of Medicine** University of Oslo

Forskningsgruppemøte 2022, Ullevål sykehus | 22. juni 2022

ComBat: En metode for harmonisering av multisenter radiomicsstudier med PET Etablering av stabile prediksjonsmodeller Henning Langen Stokmo

PhD-kandidat

Klinikk for Radiologi og Nukleærmedisin Avdeling for Nukleærmedisin Oslo Universitetssykehus / Universitetet i Oslo Overlege (Nukleærmedisin) Spesialist i Radiologi

Klinikk for Medisinsk Diagnostikk Avdeling for Bildediagnostikk Vestre Viken HF, Drammen Sykehus

Image credit: http://medium.com





### Bakgrunn

- Første radiomicsstudie med PET for å predikere behandlingsutfall i 2009
- Flere studier har demonstrert at radiomicsparametere er sensitive for PET
  - akkvireringsparametere
  - rekonstruksjonsparametere
- En radiomicsmodell laget for data for en PET-scanner kan ikke direkte brukes på data fra en annen PET-scanner
  - Problemer med multisenterstudier *sentereffekten*
- Mange radiomicsstudier med PET har få pasienter i forhold til antall radiomicsparametere (*curse of dimensionality*) og har dermed for dårlig kontroll av type 1-feil
  - Trenger pooling av multisenterstudier med flere pasienter for validering av modeller
- Fører til begrensninger i disseminering av modellene og overføring til klinisk praksis

El Naqa I, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit. 2009 Jun 1;42(6):1162–71. Orlhac F, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. Journal of Nuclear Medicine. 2018 Aug;59(8):1321–8.

#### Introduksjon – ComBat

- Metoden var først utviklet for micromatriseanalyser i genetikken; publisert i 2007
  - Problemer med *batcheffekt* (Ikke-biologisk, eksperimentell variasjon)
- Basert på et empirisk bayesiansk rammeverk
  - Robust for små utvalg med ekstremverdier
- Tilsvarende kan vi si at *batch* referer til akkvirerings- og rekonstruksjonsparametere i PET
- Fortin *et al.* 2017/2018 brukte ComBat for DTI/MRI
- Orlhac *et al*. 2018 foreslo å bruke tilsvarende metode for å fjerne *sentereffekten* fra radiomicsstudier
- Siden 2017 har minst 51 artikler brukt metoden
  - MRI 36%, CT 34%, PET 28%

sitetssykehus

Fortin JP, et al. Harmonization of multi-site diffusion tensor imaging data. NeuroImage. 2017 Nov 1;161:149–70. Fortin JP, et al. Harmonization of cortical thickness measurements across scanners and sites. NeuroImage. 2018 Feb 15;167:104–20. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007 Jan;8(1):118–27. Orlhac F, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. Journal of Nuclear Medicine. 2018 Aug;59(8):1321–8. Orlhac F, et al. A Guide to ComBat Harmonization of Imaging Biomarkers in Multicenter Studies. J Nucl Med. 2022 Feb;63(2):172–9.



https://www.quantstart.com/articles/Bayesian-Statistics-A-Beginners-Guide/

Oslo

universitetssykehus

### Pasientdata (1)

- To grupper pasienter (n = 137) med ikke-metastatisk brystcancer
- To forskjellige PET-scannere
  - Philips Gemini TF
    - Blob OS, 2i33s, TOF
    - Ingen post-rekonstrusjon
    - Voxelstørrelse 4 x 4 x 4 mm
  - Ge Discovery 690
    - OSEM, 2i24s,
    - Post-rekonstruksjon 6 mm Gaussfilter
    - Voxelstørrelse 2.7 x 2.7 x 3.3 mm

Parameter	Kohort A	Kohort B	
n	63	74	
Age (mean ± SD) [years]	55 ± 15	51 ± 14	
Molecular subtype (n, %)			
Luminal A	9 (14%)	11 (15%)	
Luminal B	35 (56%)	44 (59%)	
Triple-negative (TN)	16 (25%)	15 (20%)	
HER2-positive	3 (5%)	2 (3%)	
Unknown		2 (3%)	
Time inj. to scan (mean ± SD) [min.]	78 ± 9	74 ± 8	
18F-FDG Dose [MBq/kg]	3	3 – 3.5	
Rate of infusion per bed pos. [min.]	1.45	2.5	

Orlhac F, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. Journal of Nuclear Medicine. 2018 Aug;59(8):1321-8.



# Pasientdata (2)

- For hver pasient 2 VOIs
  - Primærtumor: Delineasjon 40% SUV<sub>max</sub>
  - Normal, frisk lever: 23 cm<sup>3</sup>
- For hver VOI 9 parametere
  - SUV<sub>max</sub>, SUV<sub>mean</sub>, SUV<sub>peak</sub>
  - Homogeneity, Entropy (Gray-Level Co-occurrence Matrix)
  - Short-Run Emphasis, Long-Run Emphasis (Gray-Level Run Length Matrix)
  - Low-Gray-Level Zone Emphasis, High-Gray-Level Zone Emphasis (Gray-Level Zone Length Matrix)

Orlhac F, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. Journal of Nuclear Medicine. 2018 Aug;59(8):1321–8. Im H-J, Bradshaw T, Solaiyappan M, Cho SY. Current Methods to Define Metabolic Tumor Volume in Positron Emission Tomography: Which One is Better? Nucl Med Mol Imaging. 2018 Feb 1;52(1):5–15.





### Harmoniseringsmetode (1)

• Alle parameterverdier y kan skrives på følgende måte



Orlhac F, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. Journal of Nuclear Medicine. 2018 Aug;59(8):1321-8.





### Harmoniseringsmetode (2)

- $\alpha, \gamma_i \text{ og } \delta_i$  estimeres ved hjelp av empirisk bayesiansk statistikk
- I kontrast til standard bayesianske metoder er a priori distribusjonen estimert fra dataene og ikke antatt fra før.
- Algoritmen estimerer en transformasjon for hver enkelt feature. Denne er spesifikk for vevstype og feature.

Fortin J-P, et al. Harmonization of multi-site diffusion tensor imaging data. NeuroImage. 2017 Nov 1;161:149–70. https://en.wikipedia.org/wiki/Empirical\_Bayes\_method



#### Harmonisering – friskt levervev (1)



- Figurene A og C viser sannsynlighets– tetthetsfunksjonene i friskt levervev før harmonisering
- Figurene B og D viser sannsynlighets– tetthetsfunksjonene i friskt levervev etter harmonisering

Orlhac F, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. Journal of Nuclear Medicine. 2018 Aug;59(8):1321-8.





#### Harmonisering – friskt levervev (2)

#### TABLE 2

*P* Values of Wilcoxon Tests Between Feature Values in Departments A and B and in Departments A and A-S Before and After Harmonization

	A vs	ь. В	A vs. A-S			
Feature	Before harmonization	After harmonization	Before harmonization	After harmonization		
Homogeneity	<0.0001*	0.7592	<0.0001*	0.9300		
Entropy	<0.0001*	0.7828	<0.0001*	0.9611		
Short-run emphasis	<0.0001*	0.8930	<0.0001*	0.7922		
Long-run emphasis	<0.0001*	0.4708	<0.0001*	0.8491		
Low-gray-level zone emphasis	0.5961	0.1319	0.9397	0.9650		
High-gray-level zone emphasis	0.2328	0.8100	0.0233*	0.8759		
SUV <sub>max</sub>	0.0522	0.7424	<0.0001*	1.0000		
SUV <sub>mean</sub>	0.4042	0.8409	0.9980	1.0000		
SUV <sub>peak</sub>	0.3407	0.9666	0.0614	0.9766		

\*P < 0.05.

Orlhac F, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. Journal of Nuclear Medicine. 2018 Aug;59(8):1321–8.





#### Harmonisering – klassifikasjon (1)

- Vurdering av verdien av harmonisering i multisenterstudier etablering av terskelverdier
- Etablering av en terskelverdi (cut-off) for senter A til bruk på pasienter i senter B
  - Youden indeks (sensitivitet + spesifisitet 1)
- Brukte molekylær subtype TN vs ikke-TN
- Bevaring av biologisk variasjon i radiomicsparametere til tross for veldig forskjellig pasientkarakteristika mellom sentere – valg av kovariater

Orlhac F, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. Journal of Nuclear Medicine. 2018 Aug;59(8):1321–8.



# Harmonisering – klassifikasjon (2)

#### TABLE 4

Youden Indices for Distinction Between TN and Non-TN Lesions from Department B with Department A–Based Threshold and Department B–Based Threshold

		Thresh. A	
Feature	Before harmonization	n After harmonization	Thresh. B
Homogeneity	0.23	0.28	0.36
Entropy	0.21	0.20	0.39
Short-run emphasis	0.12	0.35	0.38
Long-run emphasis	0.08	0.28	0.41
Low-gray-level zone emphasis	0.07	0.33	0.39
High-gray-level zone emphasis	0.16	0.21	0.39
SUV <sub>mean</sub>	0.15	0.30	0.37
SUV <sub>max</sub>	0.05	0.25	0.32
SUV <sub>peak</sub>	0.05	0.36	0.37



### Harmonisering – klassifikasjon (3)

#### TABLE 3

*P* Values of Wilcoxon Tests Between Feature Values in TN and Non-TN Lesions in Departments A and B Before and After Harmonization

	Before harmonization				After harmonization					
Feature	TN(A) vs. non-TN(A)	TN(B) vs. non-TN(B)	TN(A) vs. TN(B)	Non-TN(A) vs. non-TN(B)	TN(A+B) vs. non-TN(A+B)	TN(B) vs. non-TN(A)	TN(A) vs. TN(B)	Non-TN(A) vs. non-TN(B)	TN(A+B) vs. non-TN(A+B)	TN(B) vs. non-TN(A)
Homogeneity	0.0810	0.0078*	0.4232	0.0074*	0.0014*	0.4635	0.5986	0.8737	0.0015*	0.0093*
Entropy	0.0205*	0.0410*	0.5196	0.3906	0.0031*	0.0875	0.7405	0.9139	0.0027*	0.0254*
Short-run emphasis	0.2175	0.0091*	0.2995	0.0004*	0.0063*	0.9481	0.1294	0.8338	0.0062*	0.0061*
Long-run emphasis	0.2618	0.0072*	0.2814	0.0004*	0.0072*	0.9352	0.0055*	0.3871	0.0162*	0.0004*
Low-gray-level zone emphasis	0.0005*	0.0119*	0.0405*	0.0244*	5.69e-05*	0.3786	0.1102	0.3059	0.0002*	0.0003*
High–gray-level zone emphasis	0.0002*	0.0119*	0.0494*	0.0282*	3.20e-05*	0.2886	0.2814	0.3337	2.27e-05*	0.0058*
SUV <sub>max</sub>	0.0006*	0.0111*	0.0544	0.0278*	7.54e-05*	0.4058	0.5717	0.7943	4.47e-05*	0.0072*
SUV <sub>mean</sub>	0.0003*	0.0139*	0.0448*	0.0359*	3.20e-05*	0.2394	0.4463	0.7747	3.05e-05*	0.0052*
SUV <sub>peak</sub>	0.0004*	0.0167*	0.0267*	0.0306*	9.75e-05*	0.4736	0.3581	0.7894	4.99e-05*	0.0061*

\*P < 0.05.

Oslo

universitetssykehus

Orlhac F, et al. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. Journal of Nuclear Medicine. 2018 Aug;59(8):1321–8.



#### 16/21

#### Konklusjon

- Enkelt implementerbar
  - Pakker i R, Python™ og MATLAB<sup>®</sup>
- Senterspesifikk harmonisering basert bare på pasientdata
  - Ingen fantommålinger nødvendig
- Appliserbar direkte på parameterdata og ikke PET-bilder
  - Ingen reduksjon av kvalitet eks. EARL v2 –> EARL v1
- Bevarer biologisk informasjon i harmoniserte parametere
- Mulighet for å definere parametere som ikke skal harmoniseres – kovariater
  - Ved veldig forskjellige pasientkarakteristika mellom sentere

#### 17/21

#### ComBat for CT- og MR-radiomics

Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics. Radiology. 2019 Jan 29;182023.

MR

Orlhac F, Lecler A, Savatovski J, Goya-Outi J, Nioche C, Charbonneau F, et al. How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. Eur Radiol. 2021 Apr;31(4):





#### Applikasjoner ComBat

#### https://github.com/Jfortin1/ComBatHarmonization - MatLab, Python, R

#### https://forlhac.shinyapps.io/Shiny\_ComBat/ - Eget brukergrensesnitt hvor en kan laste opp data

Fortin J-P, et al. Harmonization of multi-site diffusion tensor imaging data. NeuroImage. 2017 Nov 1;161:149–70. Fortin J-P, et al. Harmonization of cortical thickness measurements across scanners and sites. NeuroImage. 2018 Feb 15;167:104–20.





### Videre utvikling og lesing

Da-ano R, Masson I, Lucia F, Doré M, Robin P, Alfieri J, et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. Sci Rep. 2020 Jun 24;10(1):10248.

> Da-ano R, Lucia F, Masson I, Abgral R, Alfieri J, Rousseau C, et al. A transfer learning approach to facilitate ComBat-based harmonization of multicentre radiomic features in new datasets. PLOS ONE. 2021 Jul 1;16(7):e0253653.

> > Orlhac F, Eertink JJ, Cottereau AS, Zijlstra JM, Thieblemont C, Meignan M, et al. A Guide to ComBat Harmonization of Imaging Biomarkers in Multicenter Studies. J Nucl Med. 2022 Feb;63(2):172–9.





# Takk for oppmerksomheten





21/21



Xkcd

BET YOU \$50

IT HASN'T.

For en nærmere forklaring:

https://www.explainxkcd.com/wi ki/index.php/1132:\_Frequentists \_vs.\_Bayesians

https://xkcd.com/1132/



