Data integration for prediction of weight loss in clinically controlled dietary trials

Rikke Linnemann Nielsen[1,2*], **Marianne Helenius[1*]**, Sara Garcia[1], Henrik Munch Roager[3,4], Derya Aytan[1,4], Lea Benedicte Skov Hansen[1], Mads Vendelbo Lind[3], Josef Vogt[1], Marlene Danner Dalgaard[1], Martin I Bahl[4], Cecilia Bang Jensen[1], Rasa Muktupavela[1], Christina Warinner[5], Vincent Appel[6], Rikke Gøbel[6], Mette B Kristensen[3], Hanne Frøkiær[7], Morten H Sparholt[8], Anders F Christensen[8], Henrik Vestergaard[6], Torben Hansen[6], Karsten Kristiansen[9], Susanne Brix Pedersen[10], Thomas Nordahl Petersen[4], Lotte Lauritzen[3**], Tine Rask Licht[4**], Oluf Pedersen[6**], Ramneek Gupta[1**].

* These authors contributed equally
** Corresponding authors

rlni@dtu.dk, mahele@dtu.dk

1. Department of Health Technology, Technical University of Denmark. 2. Sino-Danish Center for Education and Research, University of Chinese Academy of Sciences, Beijing, China. 3. Department of Nutrition, Exercise and Sports, University of Copenhagen. 4. National Food Institute, Technical University of Denmark. 5. Department of Anthropology, Harvard University, Cambridge, USA 02138. 6. The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200, Copenhagen, Denmark. 7. Institute for Veterinary and Animal Sciences, University of Copenhagen. 8. Department of Radiology, Bispebjerg Hospital, Copenhagen, Denmark. 9. Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, DK-2100, Copenhagen, Denmark. 10. Department of Bioengineering, Technical University of Denmark.

Aim of study

Metabolic health was investigated in a randomized, controlled cross-over trial with two 8-week dietary intervention periods with a wholegrain-rich or gluten-poor and a refined grain diet. The global metabolic response reported in the wholegrain and gluten studies were not necessarily universal in all individuals [1]. In clinically controlled trials, people tend to loose weight independent of the intervention diet [2].

In this project we aimed to

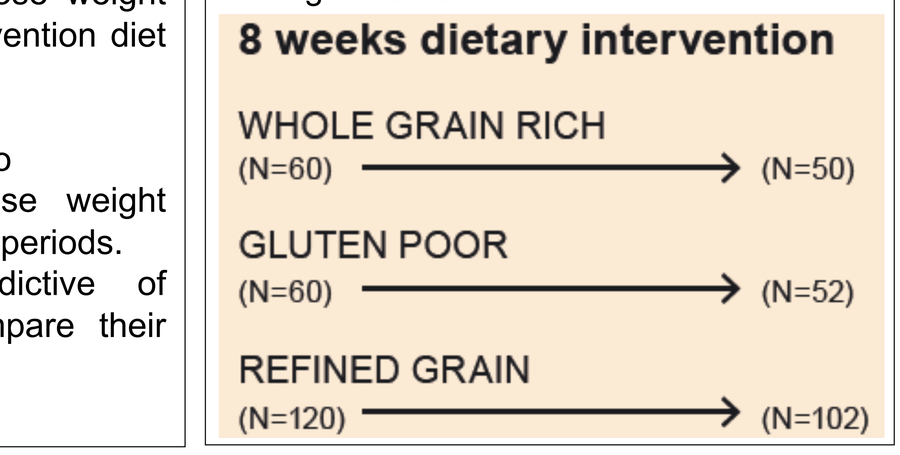
- Predict who will loose weight during the intervention periods.
- Identify factors predictive of weight loss and compare their predictiveness.

Study design

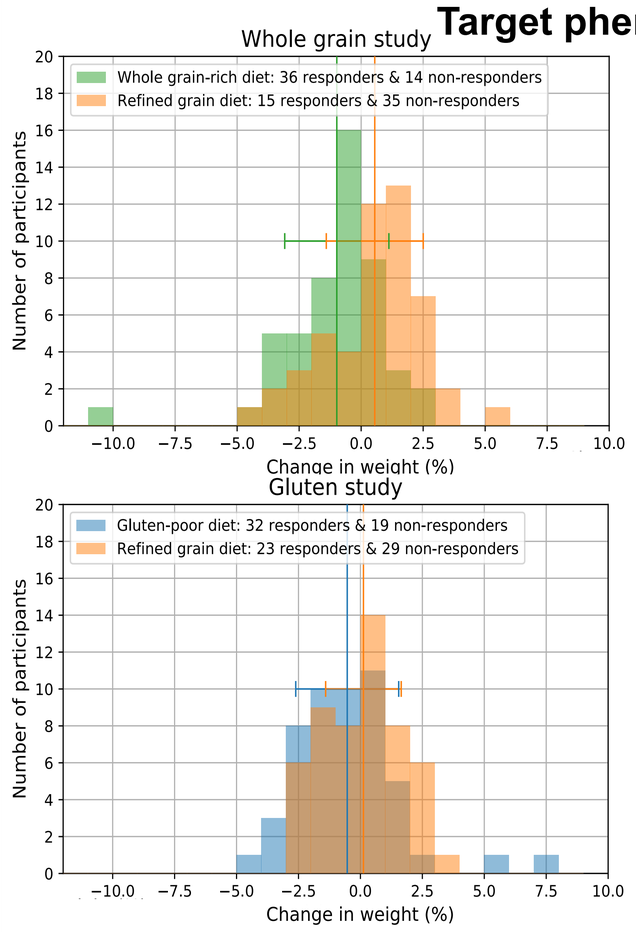
Study participants included healthy Danish men and women exhibiting a metabolic risk profile [3].

The criteria for participating in the studies included:

- Age 20-65 years
- Apparently healthy
- BMI 25-35 kg/m^2 or waist circumference ≥ 94 for men and ≥ 80 cm for women
- Weight stable



Target phenotype



The change in weight from before and after an 8-week intervention period was use to estimate and classify the study participants as responders or non-responders to any given weight loss.

$$\Delta_{weight} = \frac{weight_{after} - weight_{before}}{weight_{before}}$$

Any negative change, i.e. weight loss is considered a responder:

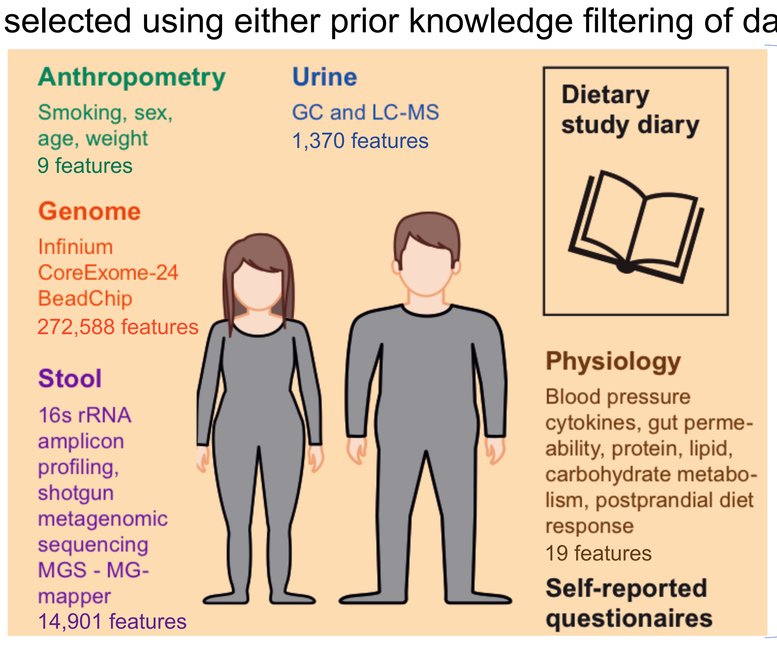
$$class = \begin{cases} \Delta < 0, & \text{Responder} \\ \Delta \geq 0, & \text{Non-responder} \end{cases}$$

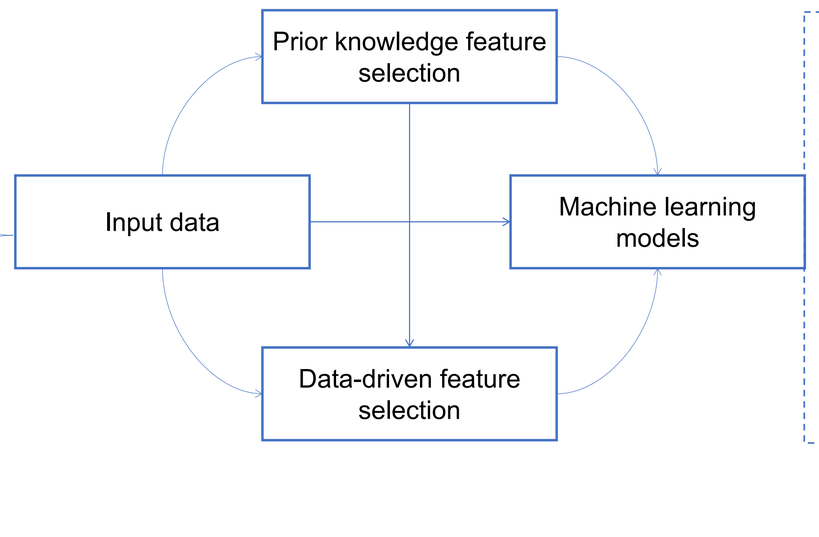
Therefore

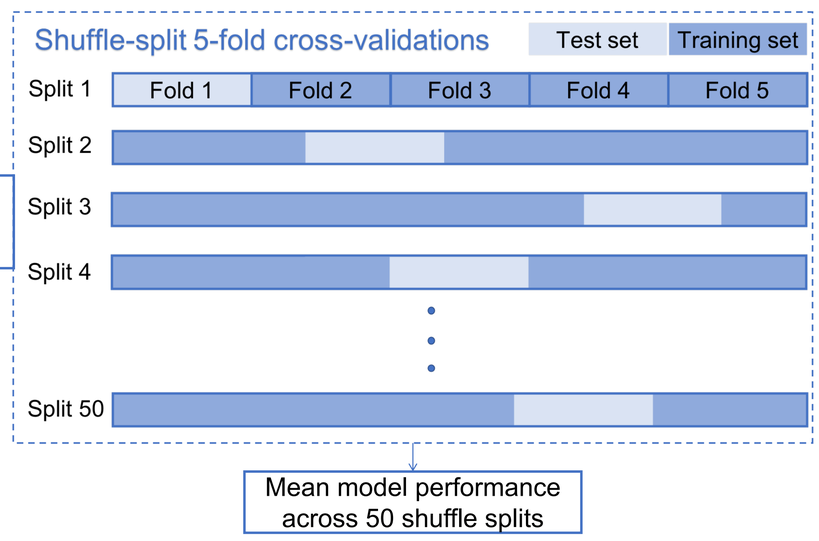
- $N_{Responders}$ = 106
- $N_{Non\text{-}responders}$ = 97

Machine learning framework and data integration strategy

Data sets were generated based on the data available for each participant before an intervention took place. Feature subsets for machine learning models were selected using either prior knowledge filtering of data or data-driven forward feature selection through cross-validation.







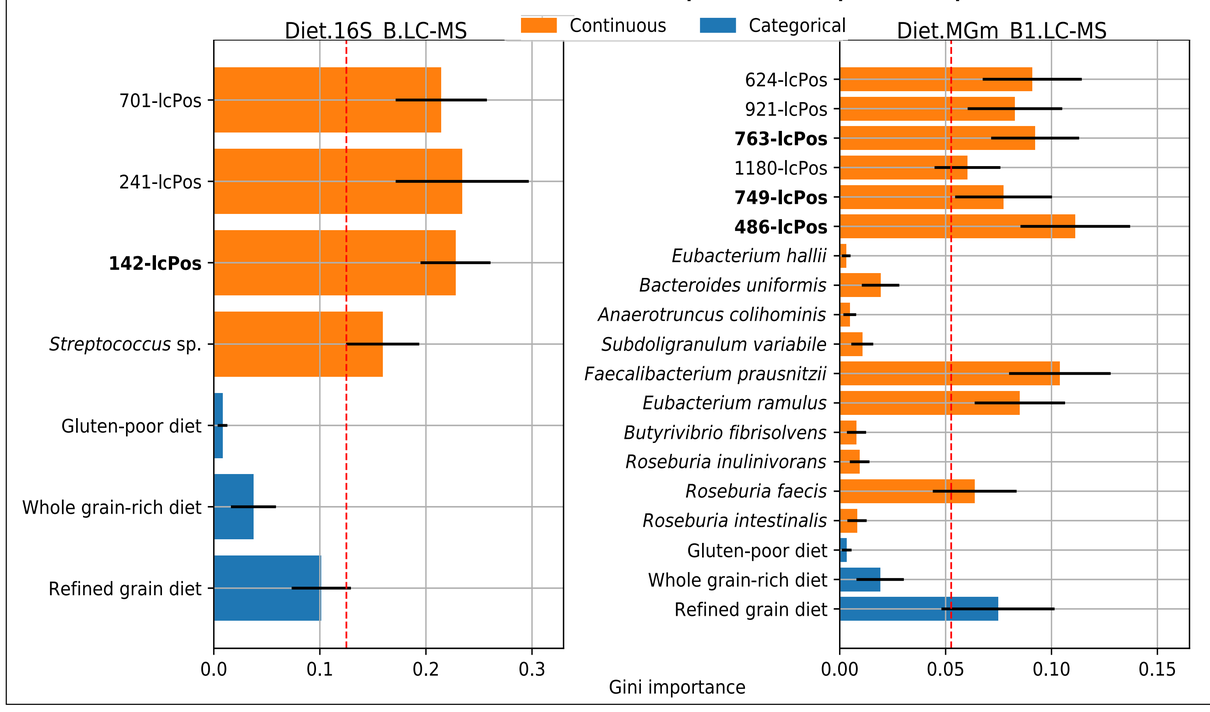
Performance of integrative random forest models

The most predictive models included urine metabolome identified by LC-MS and faecal microbiome features represented by 16S-based OTU clusters or prior knowledge selected MGmapper taxa.

	(samples|features)	ROC-AUC	Sensitivity	Specificity	MCC
Diet	(203|3)	0.62±0.01	0.64±0.00	0.66±0.00	0.30±0.00
Diet.AgeSex.EnergyIntake	(201|6)	0.65±0.02	0.67±0.03	0.58±0.03	0.25±0.05
Diet.AgeSex.VAS	(147|21)	0.56±0.04	0.54±0.05	0.57±0.05	0.11±0.06
Diet.ContinuousIntake	(201|5)	0.63±0.02	0.63±0.03	0.55±0.04	0.18±0.05
Diet.ClinicalA	(196|11)	0.57±0.02	0.60±0.03	0.54±0.03	0.14±0.04
Diet.ClinicalB	(196|7)	0.72±0.02	0.67±0.03	0.67±0.04	0.34±0.04
Diet.TransitTime	(195|4)	0.65±0.02	0.67±0.03	0.58±0.02	0.25±0.04
Diet.LitPathLD	(185|14)	0.77±0.02	0.77±0.04	0.65±0.05	0.42±0.05
Diet.GRS	(185|8)	0.60±0.02	0.66±0.03	0.48±0.04	0.14±0.05
Diet.16S_B	(179|14)	0.81±0.02	0.74±0.04	0.73±0.03	0.47±0.04
Diet.MGm_B	(183|12)	0.80±0.02	0.74±0.04	0.72±0.04	0.46±0.05
Diet.MGm_B1	(183|14)	0.67±0.02	0.64±0.04	0.61±0.03	0.25±0.05
Diet.topMGS	(185|31)	0.61±0.03	0.60±0.04	0.57±0.04	0.17±0.06
Diet.LC-MS[644-lcPos_127-lcPos]	(193|5)	0.77±0.02	0.73±0.02	0.69±0.04	0.42±0.04
Diet.PostPranFluc3_50	(203|8)	0.64±0.02	0.63±0.04	0.56±0.03	0.19±0.05
Diet.MGm_B1.LC-MS	(173|22)	0.88±0.02	0.83±0.04	0.78±0.04	0.62±0.05
Diet.16S_B.LC-MS	(169|8)	0.84±0.02	0.78±0.03	0.74±0.04	0.52±0.04

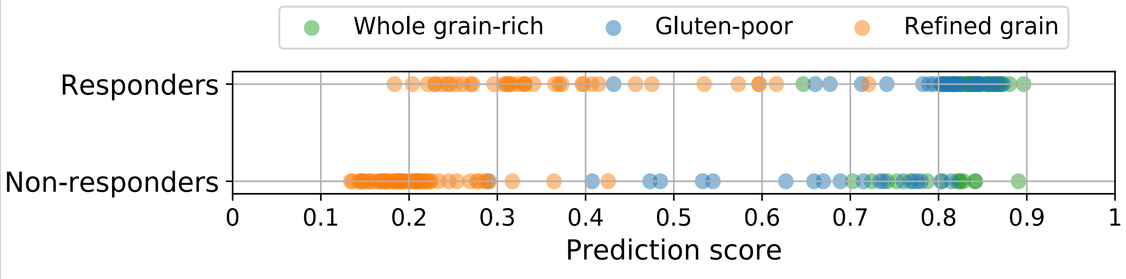
Feature importance

Urine metabolites and microbiome features proved important predictors.



Model ensemble

An ensemble of selected models (above in bold) was used in order to capture multiple aspects of biology and determine who loose weight or not with higher confidence. The ensembles performed at ROC-AUC: 0.84-0.86. Setting different score thresholds for dividing the classes enabled identification of highly confident groups, e.g. at score = 0.3, we can for 64% of the non-responders correctly classify 8/10 individuals.



Score	Sensitivity	PPV	Specificity	NPV
0.1	1.00	0.52	0.00	nan
0.2	0.99	0.63	0.37	0.97
0.3	0.88	0.73	0.64	0.83
0.4	0.74	0.70	0.66	0.70
0.5	0.69	0.72	0.70	0.67
0.6	0.65	0.72	0.72	0.65
0.7	0.61	0.74	0.76	0.64
0.8	0.56	0.86	0.90	0.65
0.9	0.00	nan	1.00	0.48

Conclusion

- The best predictors for weight loss response (ROC-AUC: 0.88) were based on selected gut microbiome features and urine metabolites identified by LC-MS.
- Without microbiome and metabolites, genotype, transit time and physiology (including post-prandial response) lead to a ROC-AUC: 0.72.
- AI frameworks can help understanding responders for diet and their place within comprehensive strategies for weight management.

Funding: This study was supported by the Innovation Fund Denmark (grant no. 11-116163/0603-00487B; Center for Gut, Grain and Greens (3G Center)). RLN was supported by a grant from DTU and the Sino-Danish Center for Education and Research.

References
[1] Roager et al, 2017. PMID: 29097438
[2] Johns et al, 2016. DOI: 10.1002/oby.21255
[3] Ibrügger et al, 2014. DOI: 10.4172/2167-0870.1000178