AARHUS UNIVERSITY

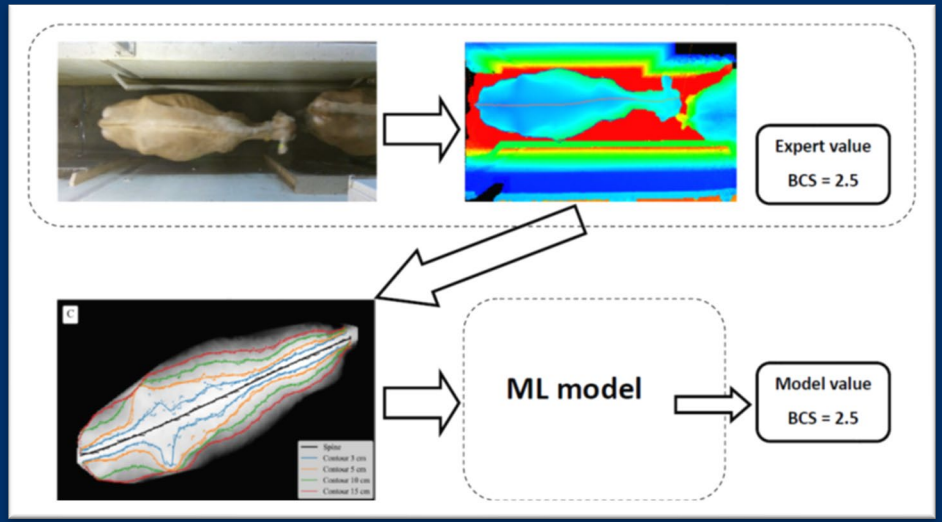
EAAP 2024 Congress
1st September 2024

RASMUS BAK STEPHANSEN
PostDoc

CENTER FOR QUANTITATIVE GENETICS AND GENOMICS QGG

Ensemble Learner Enhances Prediction Accuracies for Extreme Body Condition Classes in Jersey Dairy Cattle in a 3D Image High-Throughput System

R.B. Stephansen, G. Gebreyesus, G. Sahana, J. Lassen



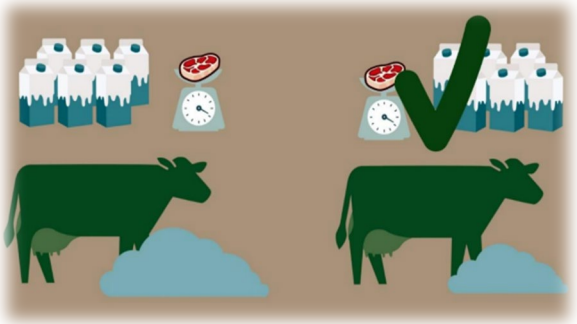
Why measure body condition?

Welfare indicator

- High management level of body condition gives
  - More functional cows (fertility, health, etc.)
  - Higher production

Important for **Metabolic Efficiency**

- Gives ability to model changes in energy
  - Distinguish between fat & protein



Previous lessons

Algorithm developed earlier

Regression analysis showed best predictive performance

Challenged in predicting **obese** classes (low frequency)

Journal of Animal Science, 2023, **101**, 1–14
https://doi.org/10.1093/jas/skad376
Advance access publication 7 November 2023
Technology in Animal Science

OXFORD

AMERICAN SOCIETY OF ANIMAL SCIENCE

Prediction of body condition in Jersey dairy cattle from 3D-images using machine learning techniques

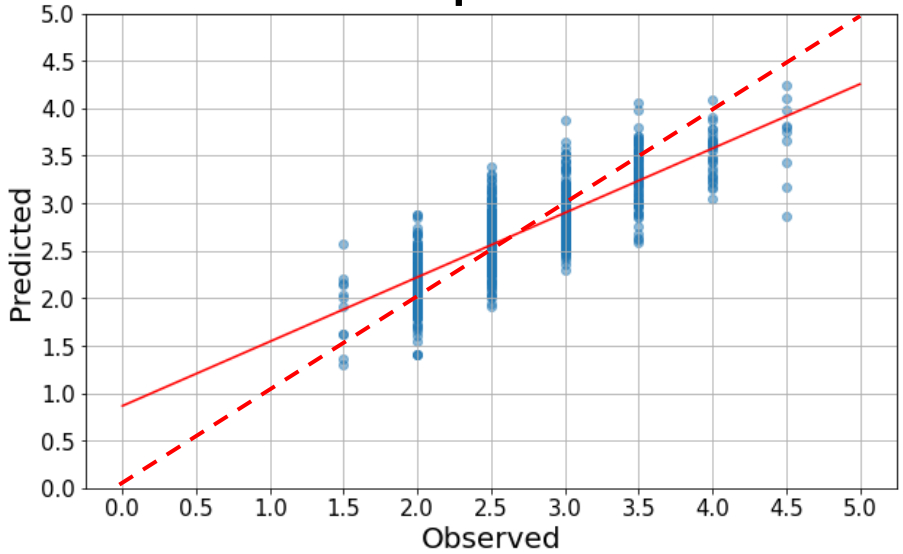
Rasmus B. Stephansen[†,1], Coralia I. V. Manzanilla-Pech[†], Grum Gebreyesus[†], Goutam Sahana[†], Jan Lassen[†,‡]

[†]Center for Quantitative Genetics and Genomics, Aarhus University, 8000-Aarhus C, Denmark
[‡]Viking Genetics, Assentoft, 8960-Randers, Denmark
[1]Corresponding author: rasmus.stephansen@qgg.au.dk

This project

Improve predictions in obese classes

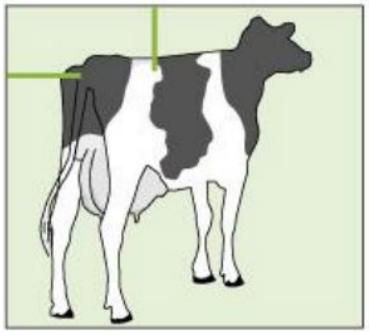
Reduce challenges with **bias** and **dispersion**



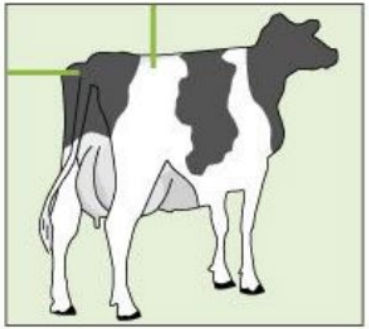
Aim - establish a more reliable BCS prediction

Data collection

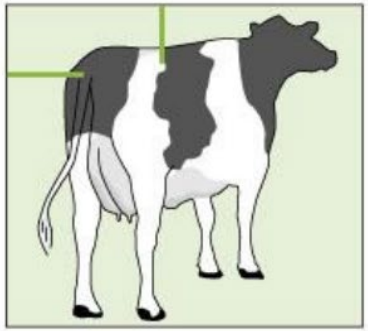
- 3 Jersey herds had all cows **classified** (cows: 794; n: 2167)



1
Lean



5
Intermediate



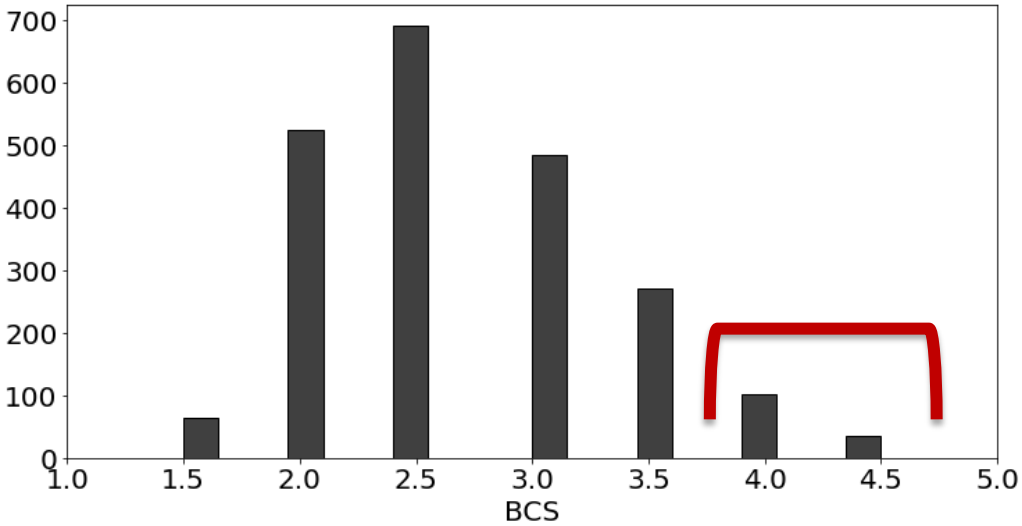
9
Fat

ICAR, 2023

Data collection

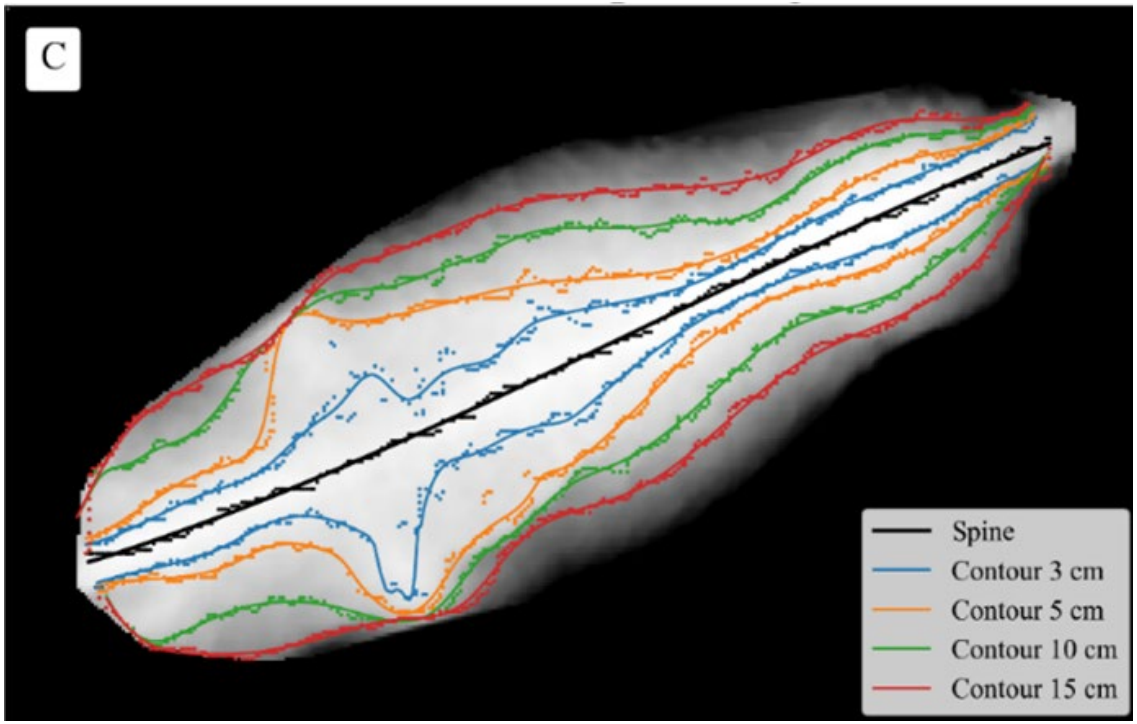
- 3 Jersey herds had all cows **classified** (cows: 794; n: 2167)

$$BCS = 0.5 \times score + 0.5,$$ Garnsworthy (2006)



Features for prediction

- 691 Features used from 3D **CFIT** system (VikingGenetics)



Machine Learning algorithms

Machine Learning **algorithms** determination

Identified with *LazyRegressor* in Python

- *BayesianRidge* from scikit.learn
- *LassoCV* from scikit.learn

Predictors

CFIT features (n = 691)

Fixed effect of evaluator + round x herd

Sub-model development

Non weighted models

BayesianRidge NoW

LassoCV NoW

Weighted models by class **frequency**: $\frac{N}{k \times f_c}$

N number of samples, k number of classes, f_c frequency of cth class

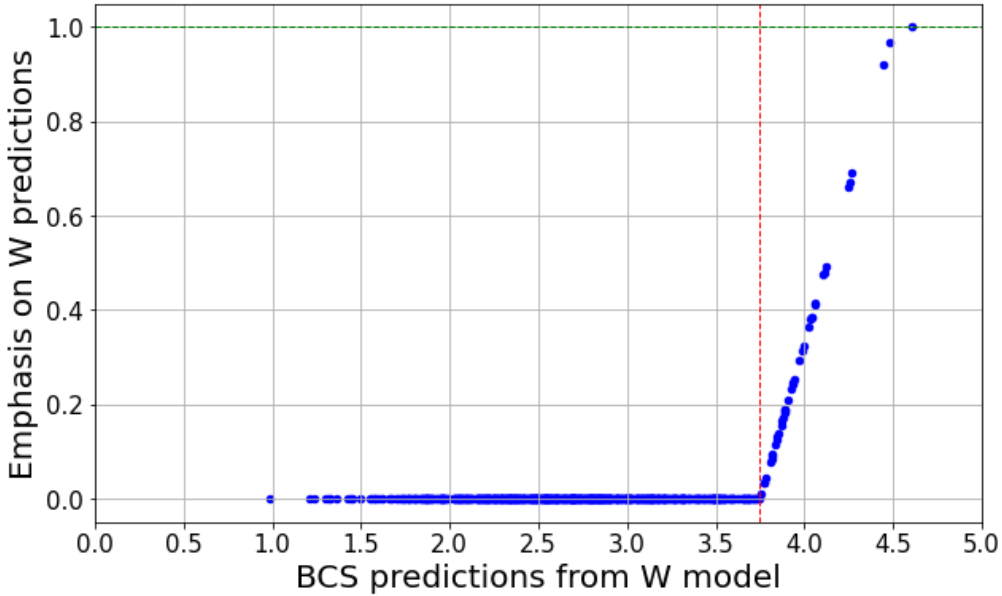
BayesianRidge W

LassoCVW

Ensemble model

Ensemble Model

1. Combine predictions equally within NoW and W models
2. Combine predictions of NoW and W by:



Training and validation

Defining Training and Validation data

Random **7:3** split, clustered by cows

Herd cross-validation

2 herds for training & 1 for validation

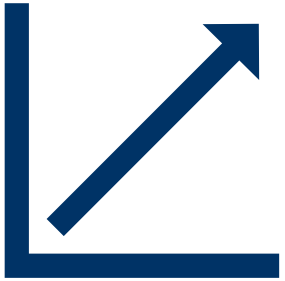
Evaluation parameters

Approximate **classification** using rounded predictions

Accuracy, %
F1-Score, %
Both assessed on **Exact** & **0.5-unit** deviation

Regression

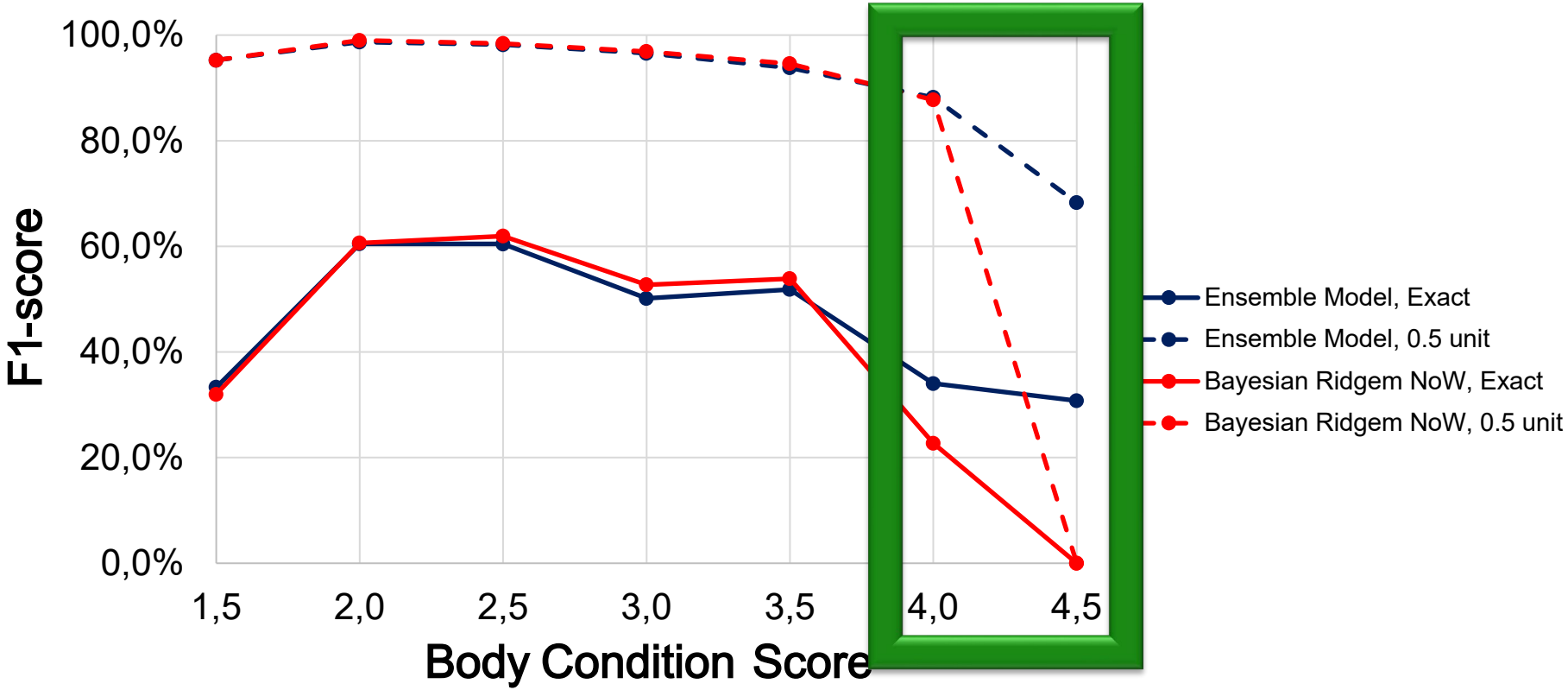
R^2, coefficient of determination
μ, bias (ideally 0)
ß, dispersion (ideally 1.0)



Validation results 7:3 split

	BayesianRidge NoW		Ensemble Model	
	Exact	0.5 unit	Exact	0.5 unit
Accuracy, %	56.0	96.8	55.1	96.5

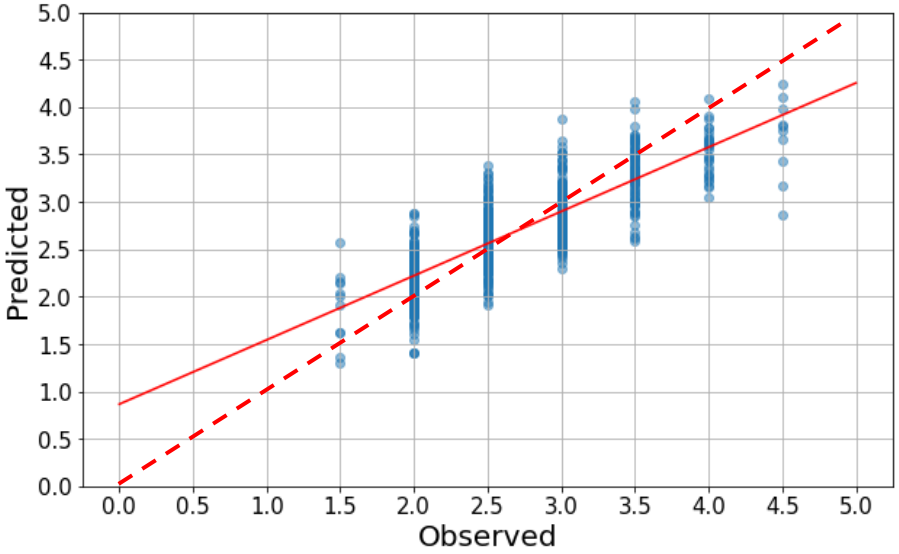
Validation results 7:3 split



Validation results 7:3 split

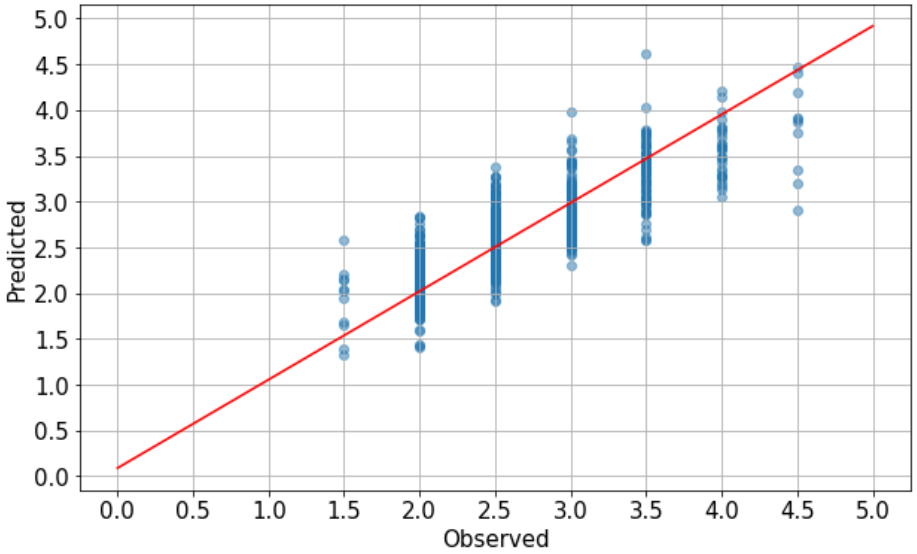
Bayesian Ridge, NoW

μ: 0.86 ß:0.68 R^2:0.67



Ensemble Model

μ: 0.09 ß:0.97 R^2:0.67



Validation results Herd CV

BayesianRidge NoW	Herd 1	Herd 2	Herd 3
Accuracy, %	Exact: 50.7 0.5 unit: 93.5	Exact: 44.9 0.5 unit: 90.5	Exact: 50.0 0.5 unit: 92.7

Ensemble Model	Herd 1	Herd 2	Herd 3
Accuracy, %	Exact: 52.2 0.5 unit: 93.0	Exact: 43.9 0.5 unit: 88.2	Exact: 50.2 0.5 unit: 92.9

Validation results Herd CV

BayesianRidge NoW	Herd 1	Herd 2	Herd 3
Accuracy, %	Exact: 50.7 0.5 unit: 93.5	Exact: 44.9 0.5 unit: 90.5	Exact: 50.0 0.5 unit: 92.7
μ, bias	0.74	1.29	1.00

Ensemble Model	Herd 1	Herd 2	Herd 3
Accuracy, %	Exact: 52.2 0.5 unit: 93.0	Exact: 43.9 0.5 unit: 88.2	Exact: 50.2 0.5 unit: 92.9
μ, bias	0.40	0.05	0.57

Validation results Herd CV

BayesianRidge NoW	Herd 1	Herd 2	Herd 3
Accuracy, %	Exact: 50.7 0.5 unit: 93.5	Exact: 44.9 0.5 unit: 90.5	Exact: 50.0 0.5 unit: 92.7
μ, bias	0.74	1.29	1.00
ß, dispersion	0.70	0.58	0.62

Ensemble Model	Herd 1	Herd 2	Herd 3
Accuracy, %	Exact: 52.2 0.5 unit: 93.0	Exact: 43.9 0.5 unit: 88.2	Exact: 50.2 0.5 unit: 92.9
μ, bias	0.40	0.05	0.57
ß, dispersion	0.87	0.94	0.76

Take-home messages

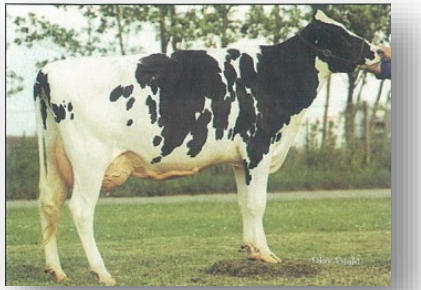
Ensemble model reduce **bias** and **dispersion** problems

Did **not** improve prediction accuracy

We could test Stacking Ensemble learning models

We can predict BCS in **Jersey** from 3D-images

Next step is to build a model for **Holstein**



Acknowledgements

CENTER FOR QUANTITATIVE GENETICS AND GENOMICS QGG

VIKINGGENETICS innovative breeding

SEGES INNOVATION






Goutam Grum Jan Mogens Jørgen Villy

Innovationsfonden

RASMUS BAK STEPHANSEN
PostDoc
CENTER FOR QUANTITATIVE GENETICS AND GENOMICS

Thank you for your attention