Hacking CASA to predict boar semen fertility

C. Kamphuis, B. Visser, P. Duenk, G. Singh, A. Nigsch, R.M. de Mol, **R.F. Veerkamp**, and M.L.W.J. Broekhuijse

WAGENINGEN UNIVERSITY & RESEARCH

100years 1918 — 2018

Topigs Norsvin PROGRESS IN PIGS

FARMHACK.NL

BREED4FOOD

Introduction

- World of Big Data:
  - machine learning, data driven,
  - people, experts, culture, none animal science ....

- Objective
  - Different way of working: hackaton
  - Concrete case for us.

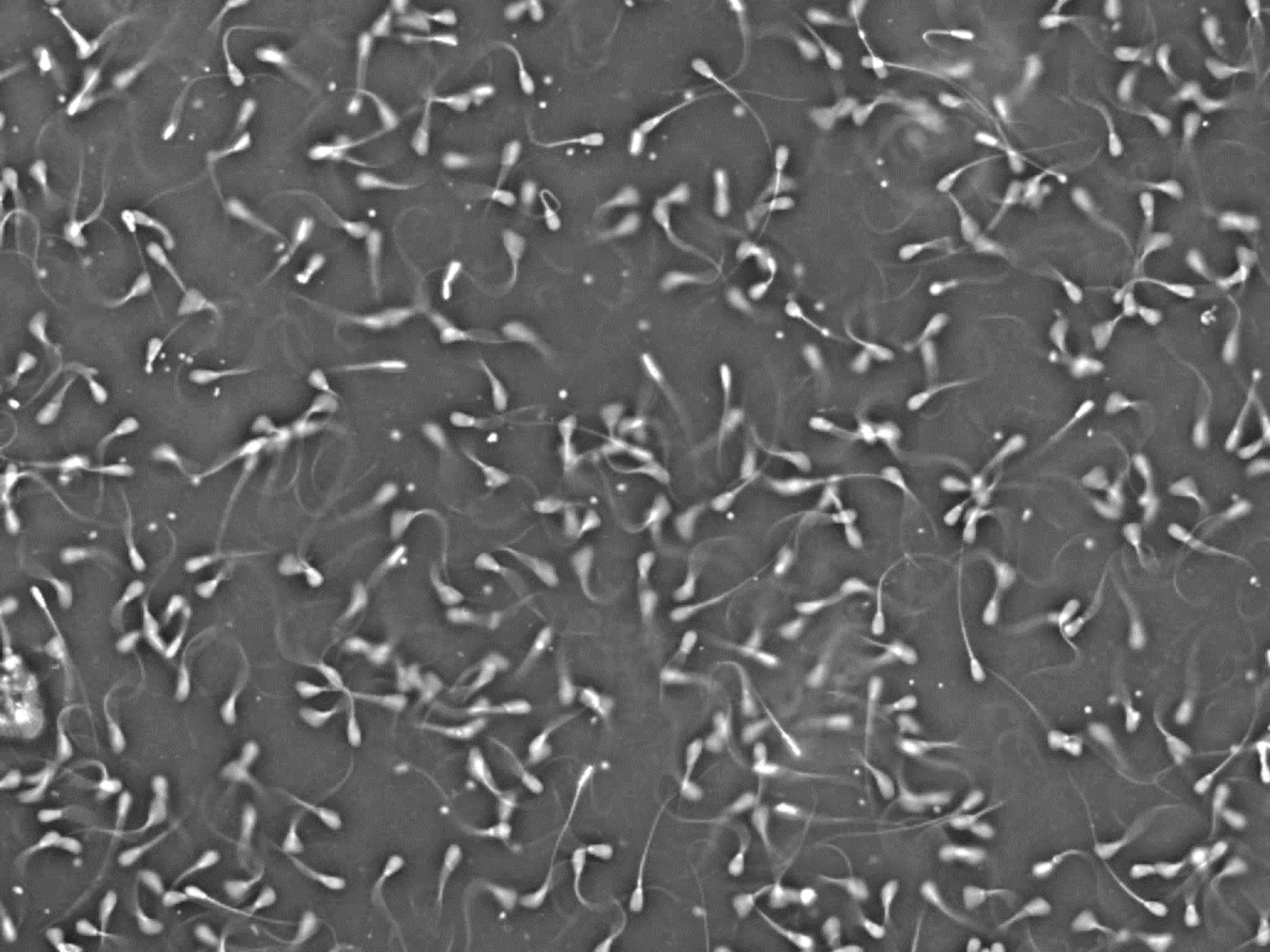
Case: predicting semen fertility

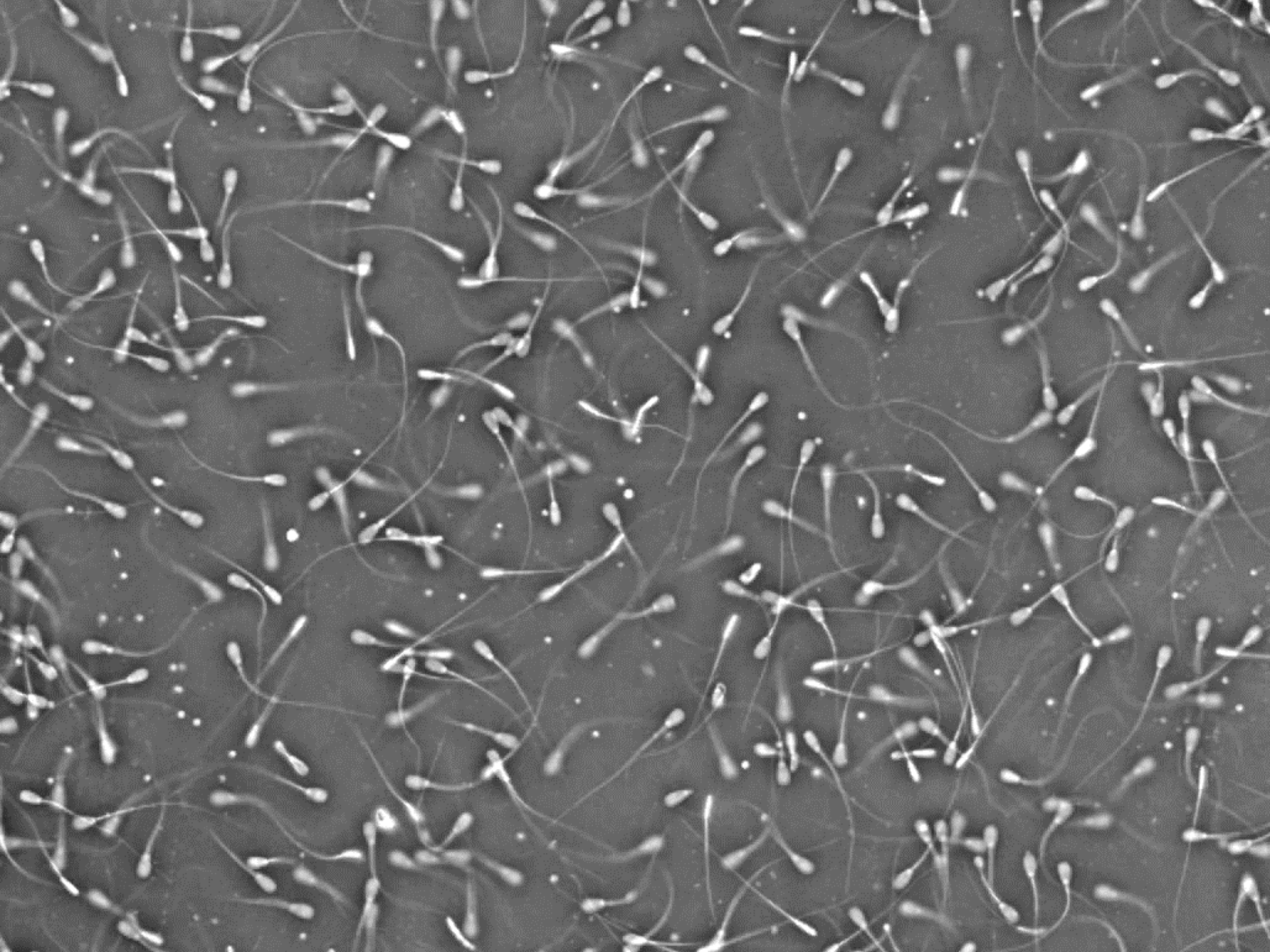
Subjective

Technician effect

Few parameters

Limited possibilities





Predicting semen fertility

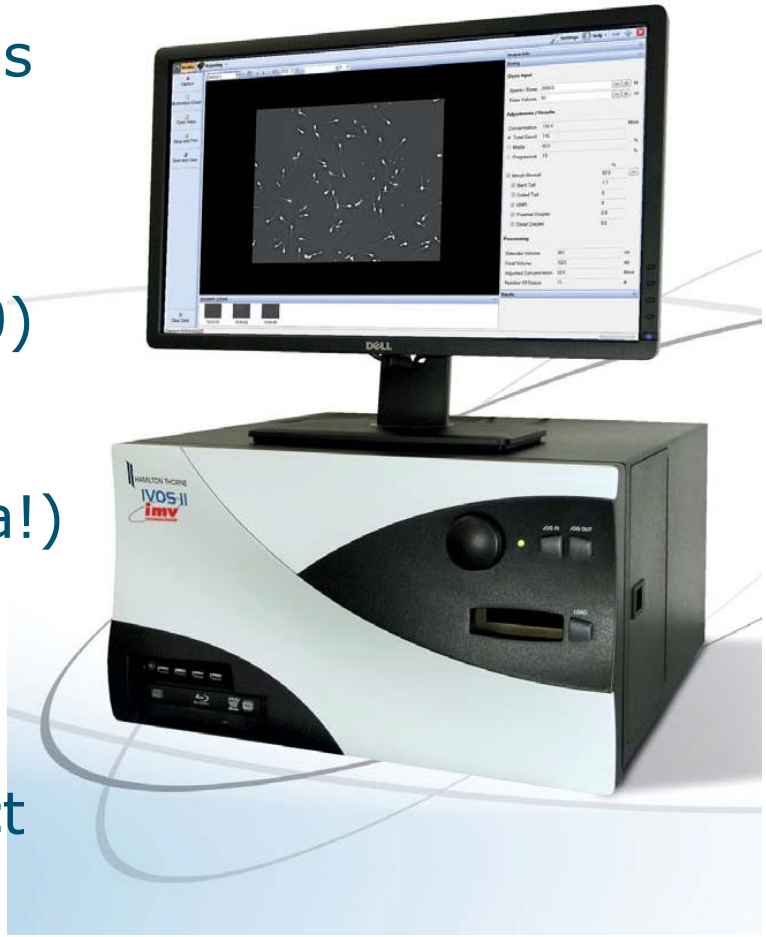
Computer Assisted Semen Analysis (CASA)

Many semen parameters (n > 400)

Used to exclude boars (not in data!)

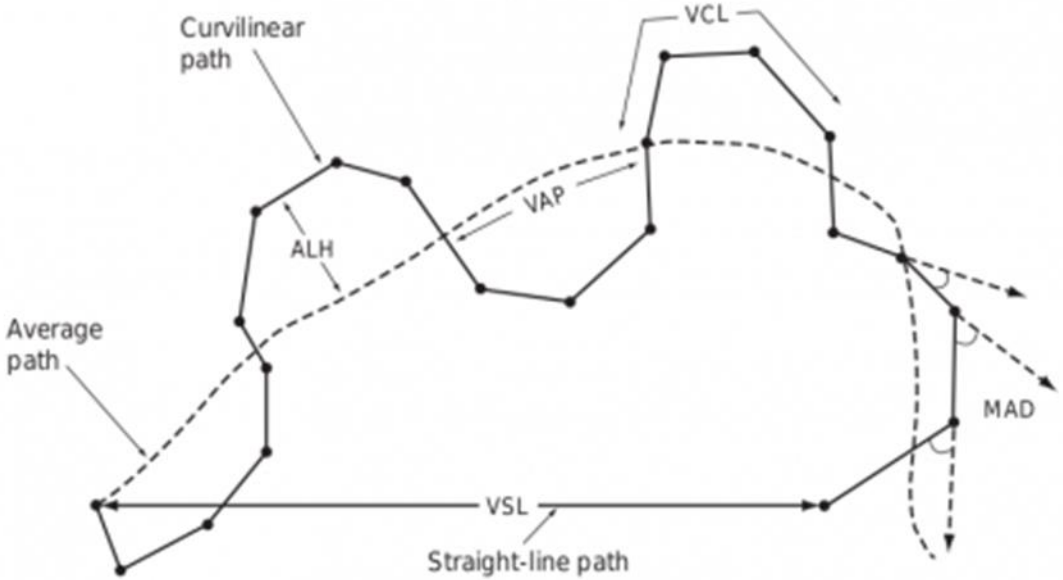
Objective:

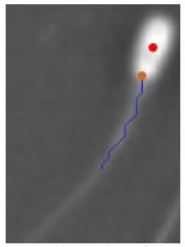
More information in data to predict fertility selected semen?

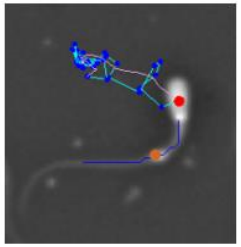


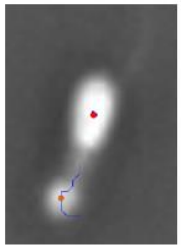
Predicting semen fertility with CASA

Metric	Value	Units
Kinematic Measures		
DAP	12.35	µm
DSL	7.80	µm
DCL	43.98	µm
VAP	16.80	µm/sec
VSL	10.61	µm/sec
VCL	59.81	µm/sec
ALH	3.62	µm
BCF	46.38	Hz
LIN	17.74	%
STR	63.17	%
WOB	28.08	%
FDM	1.66	
Morph Averages		
Head Length	12.69	µm
Head Width	2.41	µm
Head Elongation	0.20	
Head Perimeter	28.24	µm
Head Area	12.99	µm²
Tail Length	8.70	µm
Tail STR	66.83	
Droplet Distance	7.50	µm
Droplet Frame Count	39	
Bent Tail Count	9	
Coiled Count	0	
DMR Count	20	

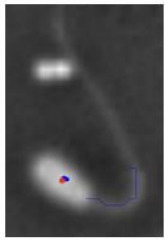



Proximal droplet


Distal droplet


Distal Midpiece reflex


Bent tail

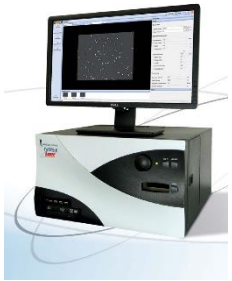

Coiled tail

Data

CASA

System 1: 2006 – 2015

System 2: 2016 – now



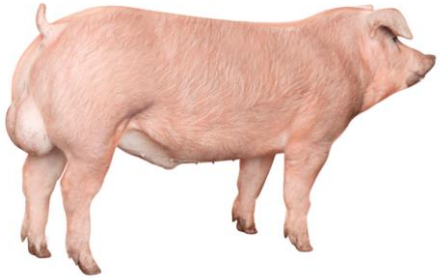
Fertility per boar/ejaculate

Total piglets born
Gestation length
Number of stillborn

Ejaculate and boar data

Fresh ejaculate

Storage (fresh)



Open source weather data

Daily data for 15yrs

KNMI

> 400 CASA parameters

>1.5 GB data

>1.2 billion ejaculates

12 years of data

9 boar stations

Method: Why a Farmhack

"Knowledge of the crowd"

- Other way of working: one (PhD) project, one model

Hackers / data enthusiasts / data analytics

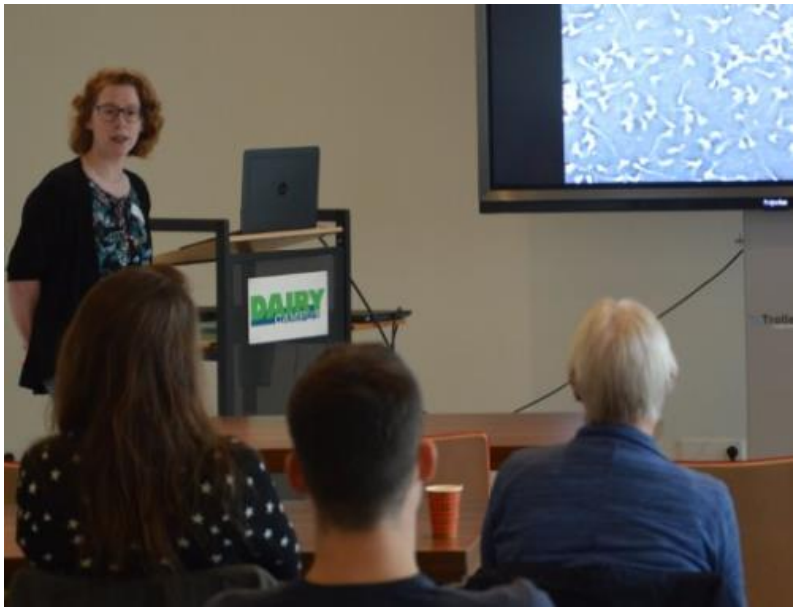
- Have often 'fake' data

Offers opportunity to

- Work with real data
- Meet people with expertise in machine learning
- Explore new ways to work / learn / collaborate

The start of a Farmhack

Introducing the domain of animal breeding



Introducing the problem, the challenge, and the expectations.

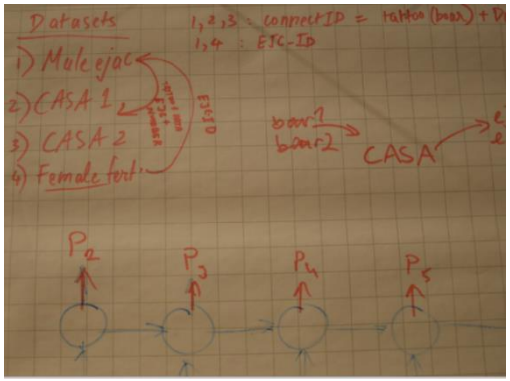
Creation of teams

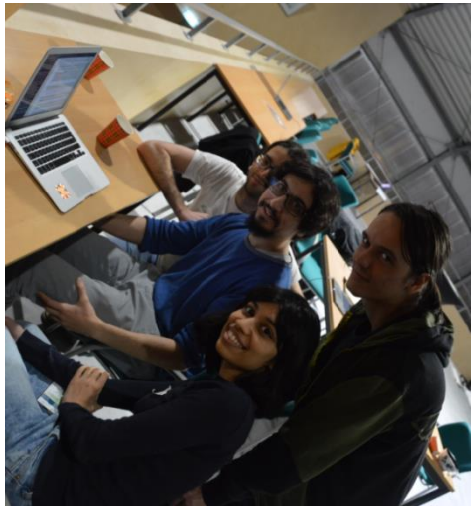
Pitch your idea

Decide whether you like it

Join the pitcher.....TEAM

Countdown: 24h hacktime





WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

Team 1st popcorn

Biology-informed machine learning

R package H2O

Gradient Boosting Machine (GBM)

sequentially builds regression trees on all variables

in a fully distributed way - each tree is built in parallel

The model

Predictors (n = 59)

CASA (parameters of interests, n=53)

Boar info

Age of Boar (in half years)

Boar temp at ejaculation (proxy fever)

Interval two ejaculates (in days)

Total morphological abnormalities

Weather info

Mean outside temperature

Daily outside temperature difference

(max – min temperature)

Responses

Total number piglets born

Gestation Length

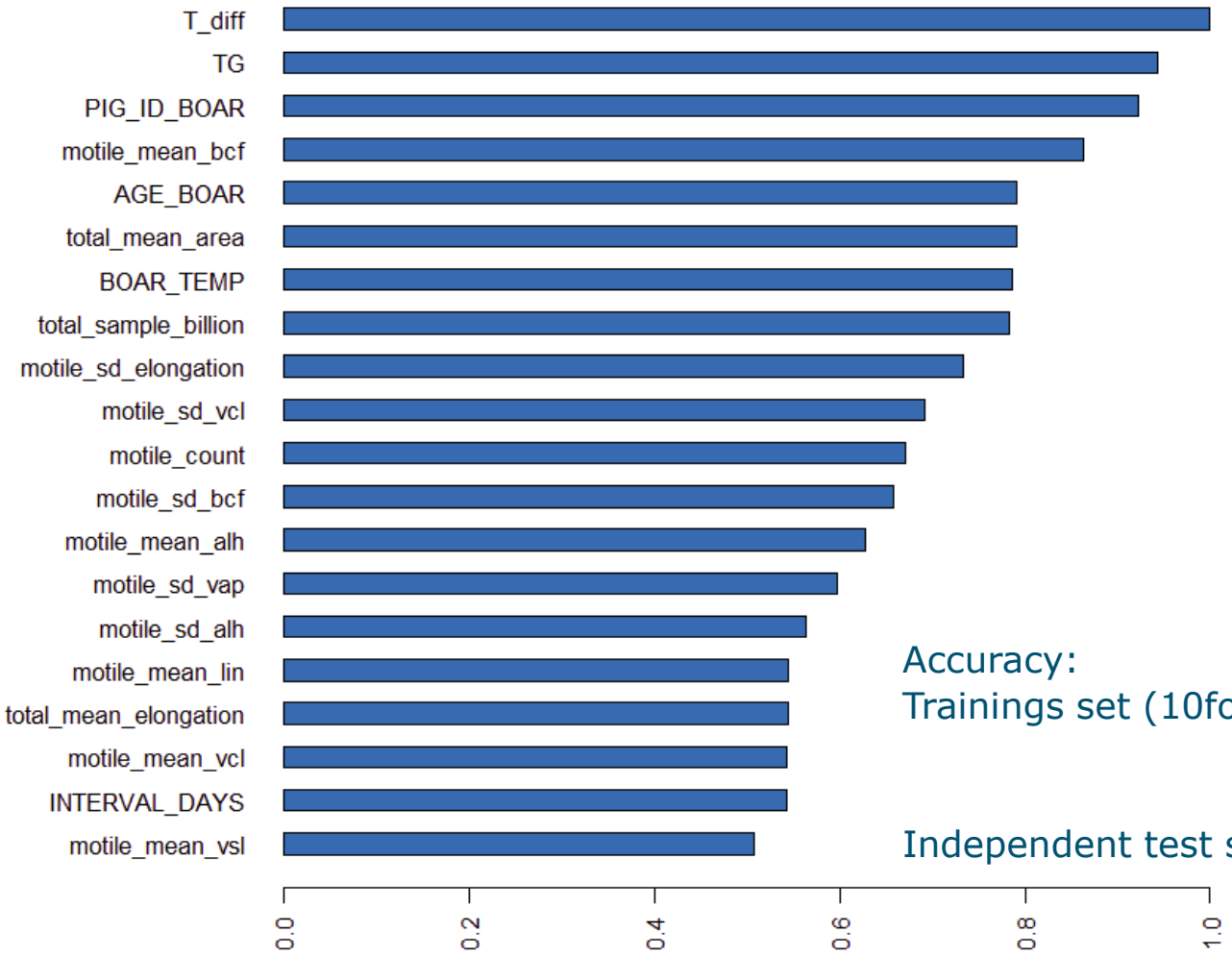
Number of stillborn

Total number piglets born **alive**

Number of stillborn (classes)

10% worst performing

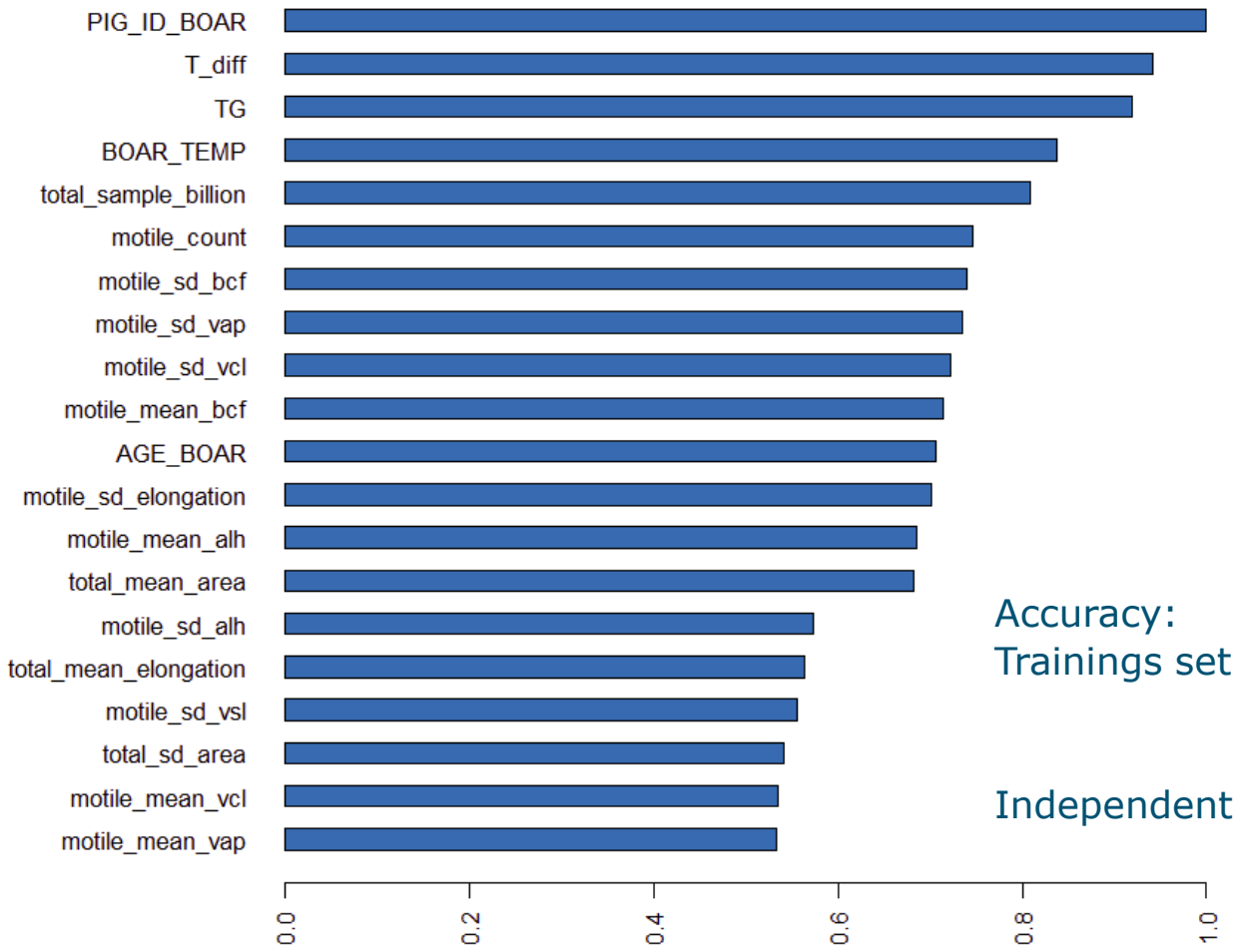
Predictor value for Gestation length



Accuracy:
Trainings set (10fold): 0.20

Independent test set: 0.03

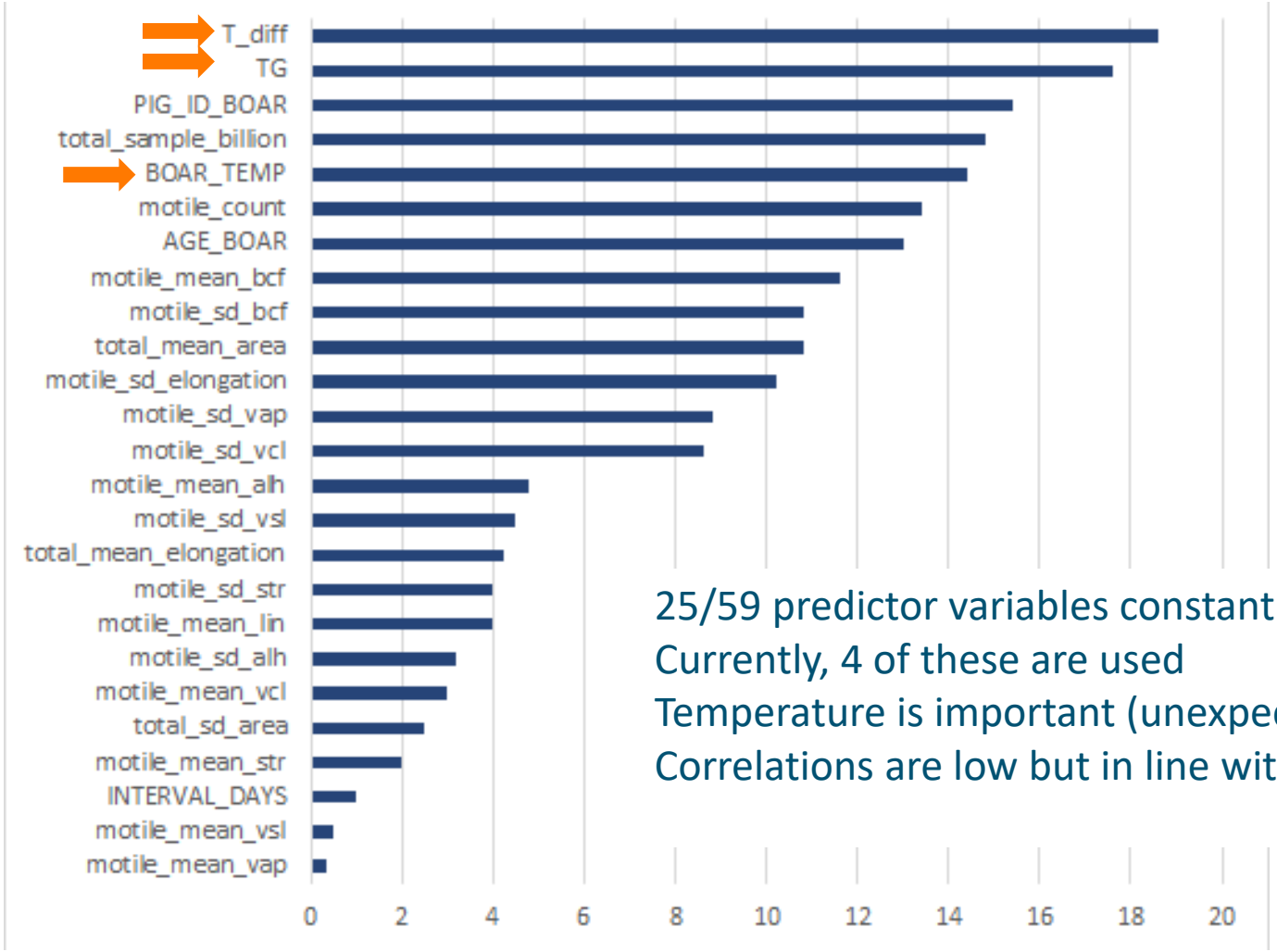
Results: Total number of piglets born



Accuracy:
Trainings set (10fold): 0.11

Independent test set: 0.11

Overall Score: summative ranking of 5 models



25/59 predictor variables constantly important

Currently, 4 of these are used

Temperature is important (unexpected high ranking)

Correlations are low but in line with expectations

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

Conclusion

Pressure cooker effect and interaction between people is an approach we could/should use more often

Combination of expert knowledge & data science essential

Machine Learning does not result in great prediction accuracy

Great networking and quick way to learn new packages/tools/approaches