Disease risk prediction based on an integrative data-methodological approach in dairy cattle

Lasser, J.[1], Matzhold, C.[2], **Egger-Danner, C.**[3], Steininger, F.[3], Fuerst-Waltl, B.[4], Wittek, T.[5] and Klimek, P.[2]

[1]Technical University, Graz, Austria

[2]Complexity Science Hub Vienna, 1080 Vienna, Austria / Medical University, Vienna, Austria

[3]ZuchtData EDV-Dienstleistungen GmbH, Dresdner Str. 89, 1200 Vienna, Austria, Rinderzucht AUSTRIA Innovation

[4]Department of Sustainable Agricultural Systems, Division of Livestock Sciences, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

[5] Veterinary University, Vienna, Austria

ICAR Annual Conference June 2, 2022

Background

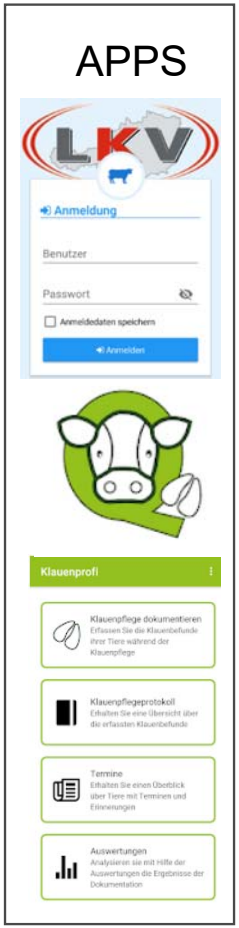
Huge amount of relevant (?) data is generated related to dairy operations

From a lot of different sources on farm and off farm (farmer, veterinarians, claw trimmers, labs, breeding and performance recording organisations, dairy processing operations, technology providers, ...)

Technological advances (genomics, robotic and AI, omics technologies, information technologies, ...)

Problem/challenges

- Separate viewing / analysis of data / little data integration (e.g. Rutten et al. 2013)
- no holistic perspective

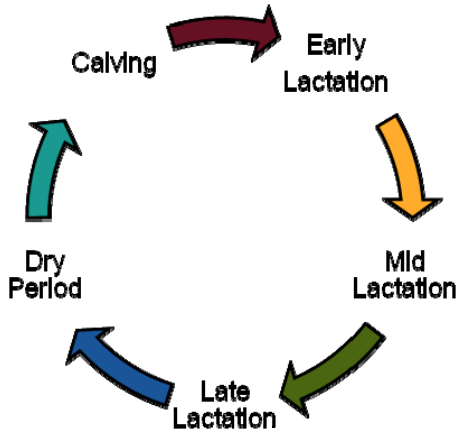


Pictures: SCR by Allflex, smaXtec, Wasserbauer, mpr Bayern, illumina

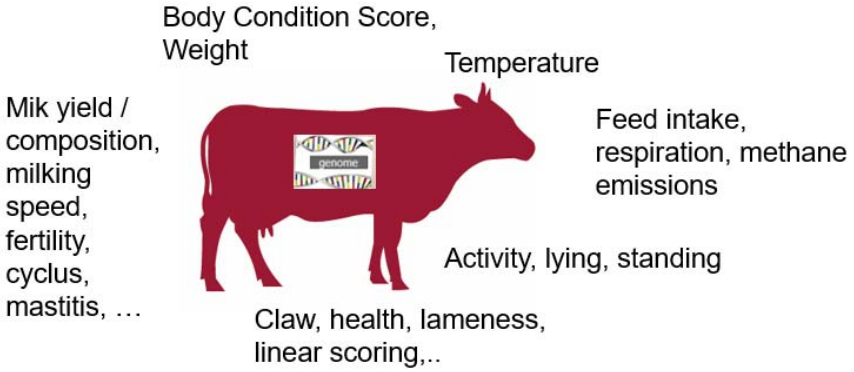
Holistic approach for prediction of disease

P. Klimek (Austrian Scientist of the year 2021: „ In Austria no other species has as much data as cattle“)

Farm- and management specific information (location, feeding, housing, climate,..)…



Animal specific information



Research questions

Digitalization holds potential to significantly improve the early detection and prevention of animal diseases

Research questions:

•Do complex integrated datasets enable and improve early detection of animal diseases?

•Do advanced methodological approaches improve prediction performance?

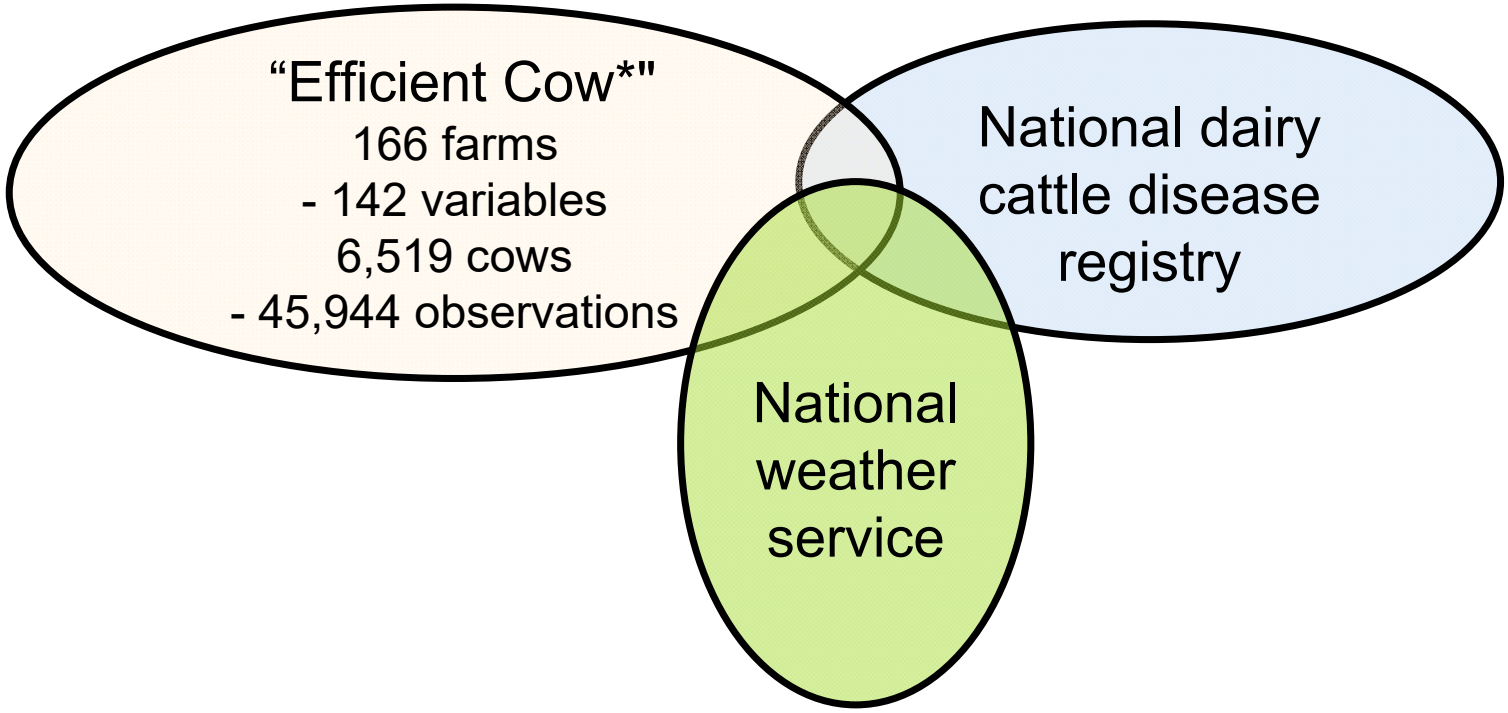
Prerequisite:

1. Data quality & quantity:

- Collecting, connecting data

2. Methodological approaches for the analysis of large data sets

Data “Efficient Cow project” – 1st step



*Nr. 100861 BMLFUW-LE.1.3.2/0083-II/1/2012

Individualized disease prediction by data integration

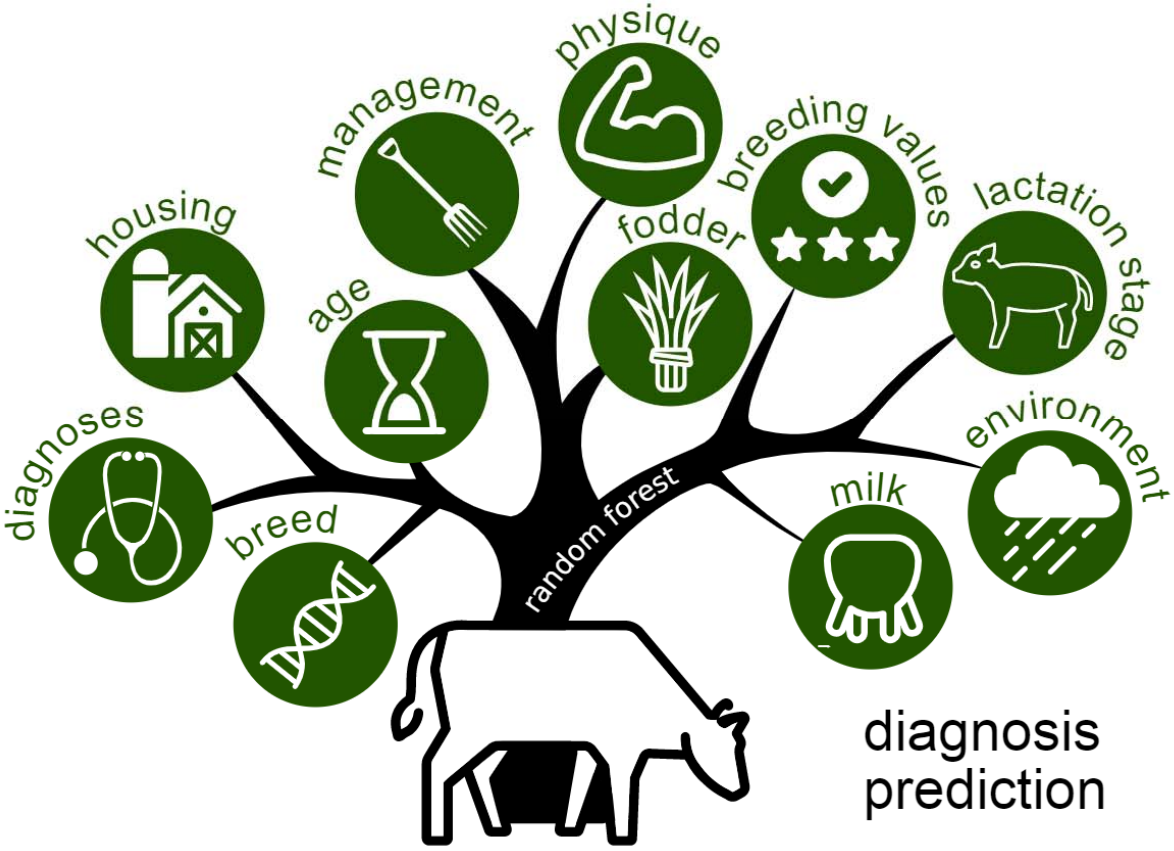


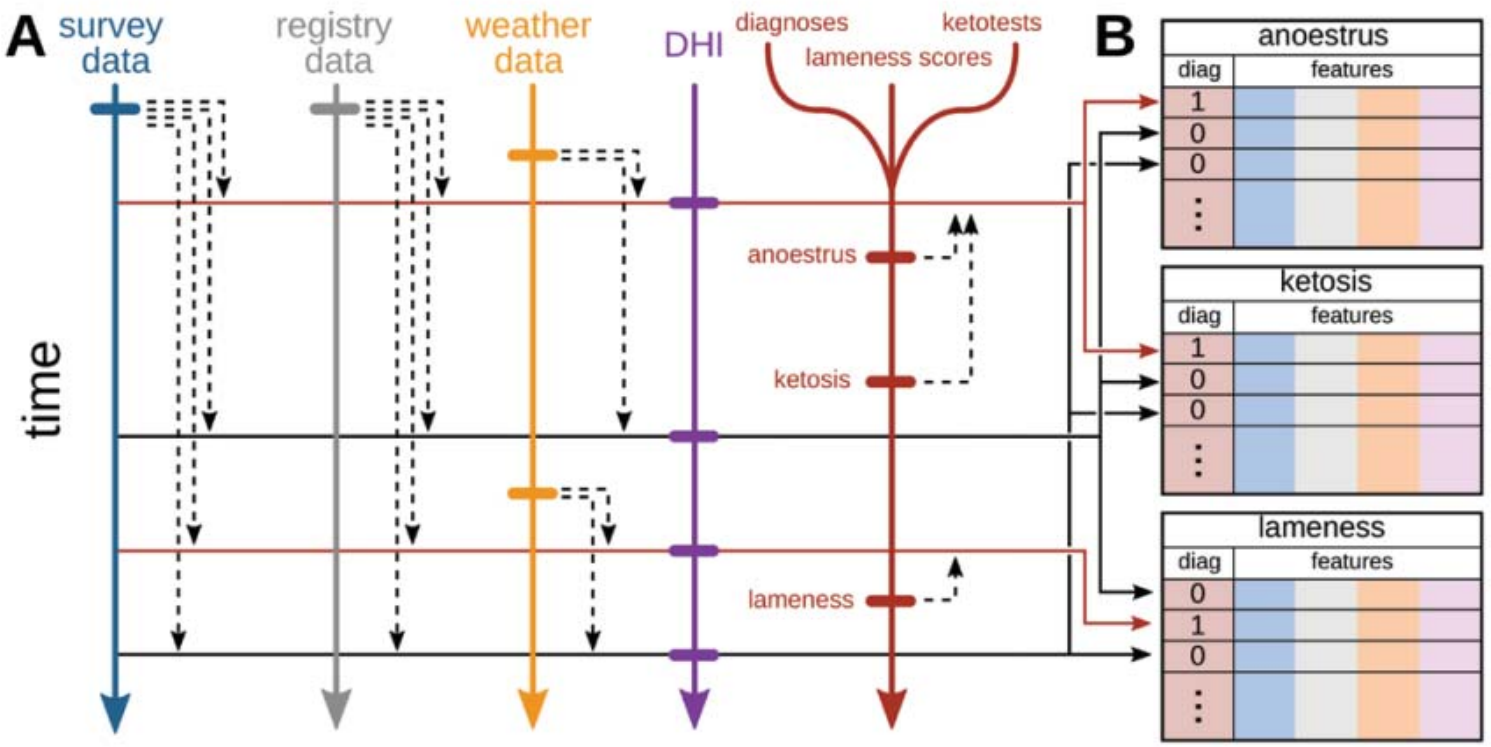
Table 1. Frequency of diagnoses for the eight most frequent diseases[1]

Diagnosis	N observations	%
Lameness	3,670	39.6
Acute mastitis	1,037	11.2
Anestrus	803	8.7
Ovarian cysts	705	7.6
Periparturient hypocalcemia	673	7.3
Ketosis	657	7.1
Chronic mastitis	335	3.6
Metritis	318	3.4
Other	1,062	11.5

[1]The percentages denote the share of the diagnoses among all diagnoses in the diagnoses dataset.

Lasser et al. 2021

Aggregation of data

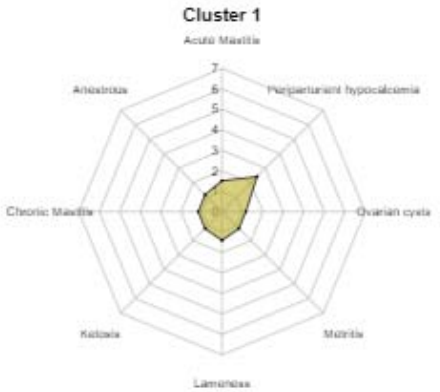


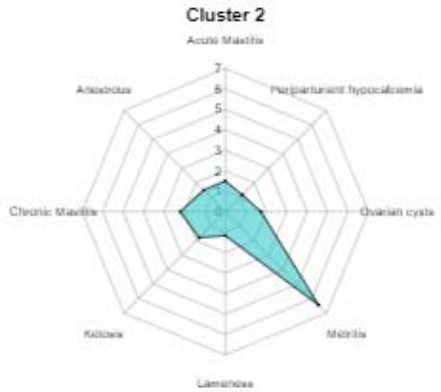
Lasser et al. 2021

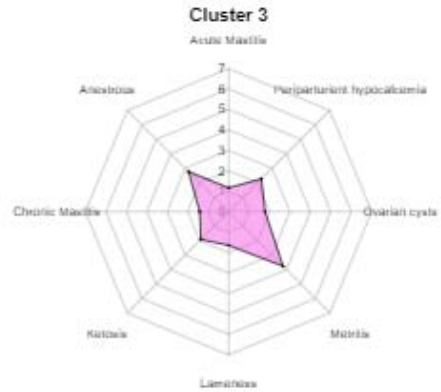
Farm-Risk-Profile

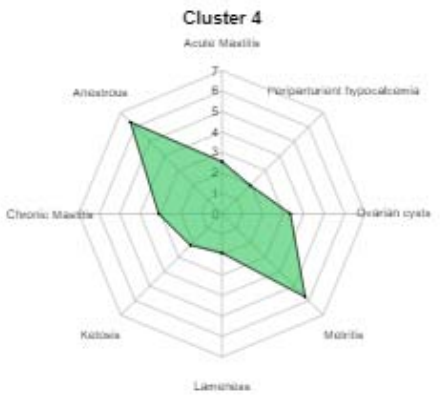
Cluster 1
lowest prevalence:
- o Anestrous
- o Ketosis
- o Chronic Mastitis
- o Metritis
- o Ovarian Cysts

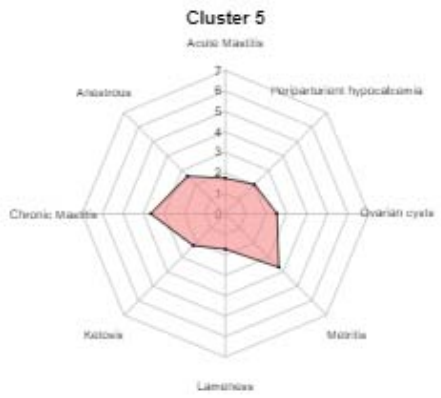
Cluster 4
highest prevalence:
- o Anestrous
- o Acute Mastitis
- o Lameness
- o Ovarian Cysts

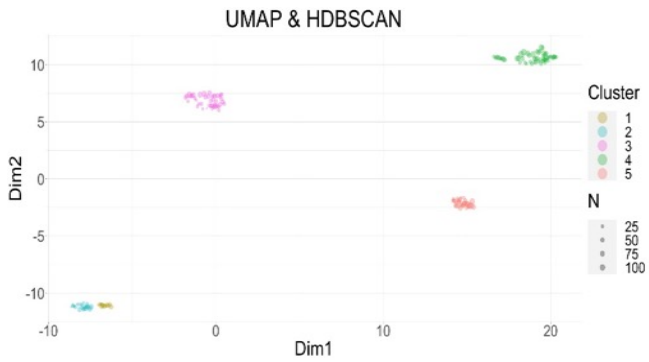












More information: Matzhold et al. (2021) A systematic approach to analyse the impact of farm-profiles on bovine health. Science Reports 11, 21152

Results: Logistic regression

Disease	F1 Score	Precision	Recall
Anestrus	0,739	0,763	0,729
Lameness	0,737	0,780	0,700
Ovarian cysts	0,616	0,675	0,543
Ketosis	0,521	0,651	0,437
Metritis	0,549	0,677	0,490
Hypercalcemia	0,482	0,576	0,420
Chronic mastitis	0,514	0,635	0,445
Acute mastitis	0,479	0,656	0,395

Precision: probability that a predicted disease will actually be diagnosed
Recall: probability that an actual disease was correctly predicted

Lasser et al. 2021

Based on Efficient Cow dataset promising first results for prediction of lameness and silent heat.

Feature category importance

Table 3. Feature category importance[1]

Feature category	Anestrus	Lameness	Ovarian cysts	Ketosis	Metritis	Periparturient hypocalcemia	Chronic mastitis	Acute mastitis
Age	4.9	36.7	2.8	4.7	2.2	28.4	7.1	11.3
Breed	1.8	2.6	1.2	1.5	2.4	0.4	0.3	2.5
Breeding values	3	2.4	2.7	1.7	2.7	1.9	4	2.5
Diagnosis source	16	15.6	29.6	11.9	33.9	7.1	18.9	16.3
Environment	14.3	4.9	16.5	10.5	15.6	13.8	19.8	10.1
Feed	27.4	11.8	17.5	22.4	12.7	7.5	18	19.9
Housing	11.5	12.9	11.3	19	15.6	5	7.4	13.6
Husbandry	9.1	5	4.5	8.8	7	2.7	6.4	6.4
Lactation stage	2.7	1.2	4.7	16.1	1.8	28.9	0.7	1.3
Milk indicators	5.9	2	3.6	0.7	2.2	1.4	10.4	12.5
Physical indicators	3.6	4.8	5.7	2.7	3.9	3	7.1	3.7

[1]Cumulative permutation feature importance contributions for the eleven feature categories. Values are given in % of the sum of all feature importances for a given disease.

Lasser et al. 2021

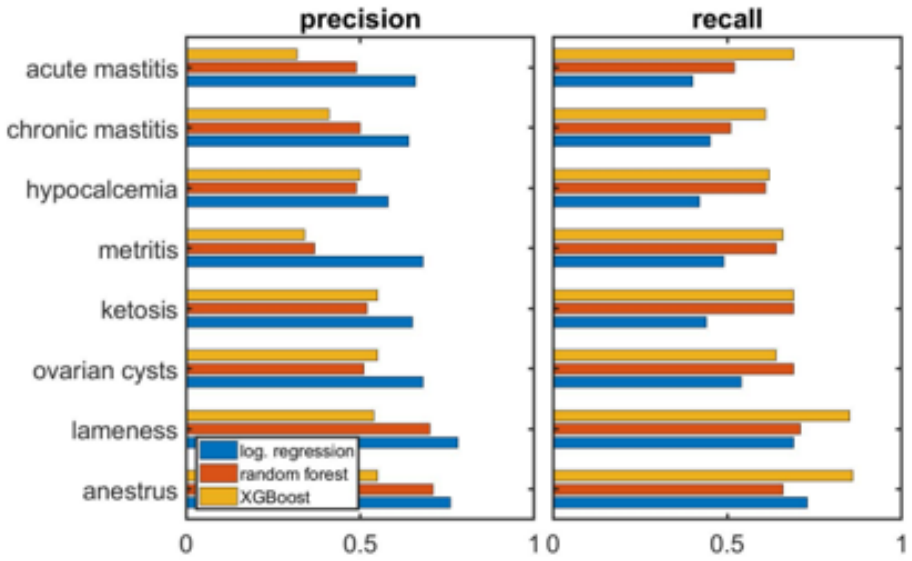
3 methods and 2 datasets

- … in increasing order of complexity
  - Logistic regression (with all ~120 variables)
  - Random forest
  - XGBoost (gradient boosting trees)
- …2 different datasets
  - „all“ data (Efficient Cow dataset, Egger-Danner et al. 2017)
  - Routinely recorded data

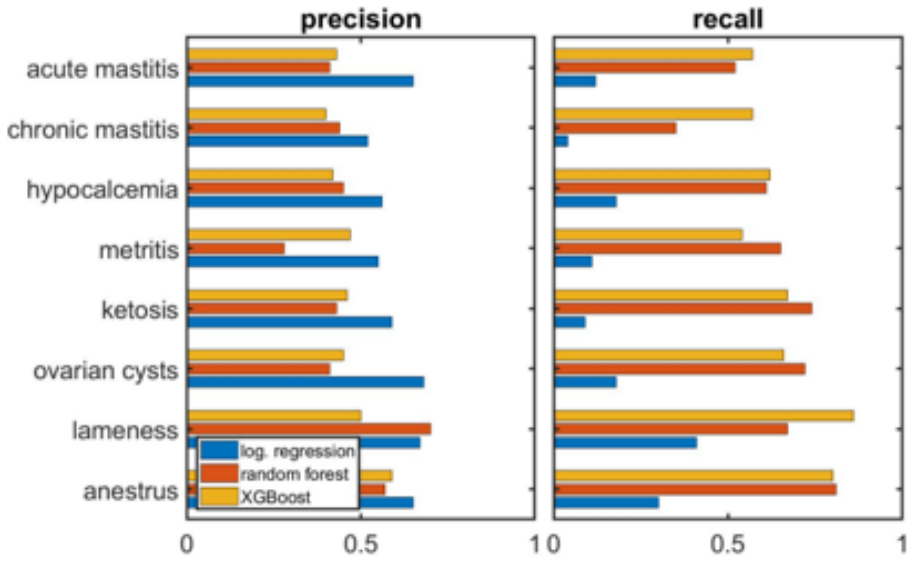
Lasser et al. (2021) Integrating diverse data sources to predict disease risk in dairy cattle – a machine learning approach. Journal of Animal Science, skab294

Results by different methods

„All“ data



Routinely recorded data



Lasser et al. 2021

Precision: probability that a predicted disease will actually be diagnosed
Recall: probability that an actual disease was correctly predicted

More complex methods perform better especially when using restricted datasets.

Can we increase prediction by more data?

D4Dairy – complex dataset created – 2nd step

Data recording 10.2019 - 06.2021

Data from pilot farms (DHI data, diagnoses, genetic and genomic information, pedigree, MIR-predictors, .. + specifically recorded data within D4Dairy

Daily milk yield from 18,261 cows

64,539 **Lameness scores**

64,865 **Body condition scores**

Sensor data from 5,439 cows

15,421 **Claw trimming** data sets

9,161 **Ketosis tests**

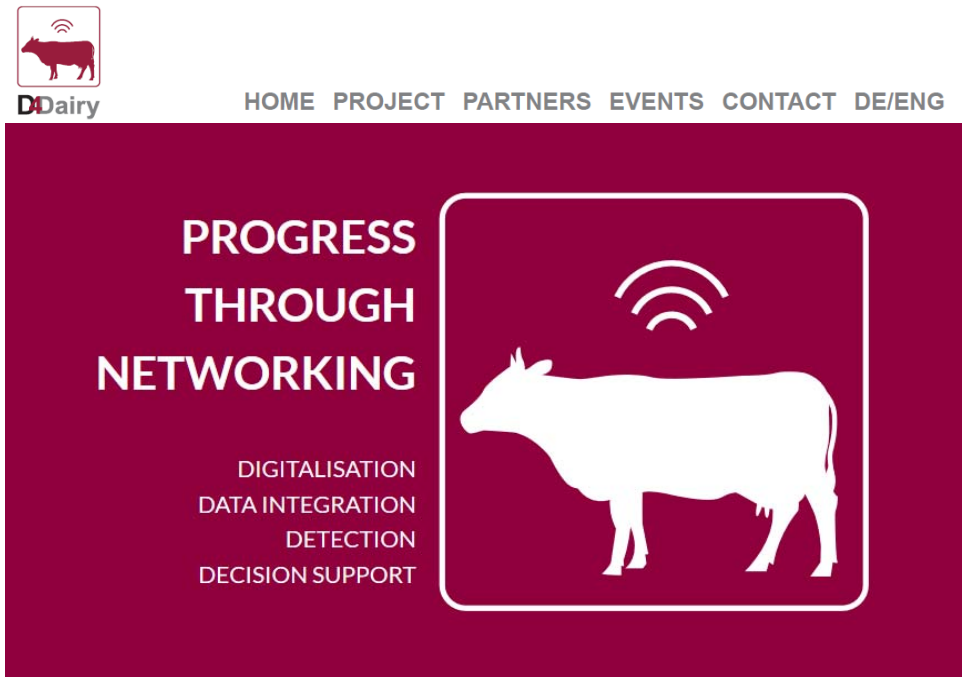
Research ongoing – aim decision support tool for early warning of diseases by data integration

Summary and next steps

- **Cattle – many different data sources available** – complexity with technological advances increasing
- **Disease risk prediction** based on an integrative data-methodological approach in dairy cattle **is promising**
- **Challenge data integration and preparation** with highly diverse datasets!
- **Machine learning approaches of varying complexity** (from logistic regression to gradient boosted trees) **show that to some extent the complexity of the algorithm can compensate for less diverse data**
- **D4Dairy – highly integrated dataset** with animal-, farm- and other environment and management specific information set up – **research ongoing**
- **AIM**: **elaboration of a data-driven decision support tool for early warning** to reduce the disease risk based on highly integrated datasets

D4Dairy – Progress through networking

Cooperation to increase benefits of digitalization along the value chain



Wissenschaftspartner:

Firmenpartner:

Cooperation partners for specific topics:

Duration: 10/2018 – 09/2022

www.d4dairy.com

Thank you!

Acknowledgement

Many thanks to farmers, veterinarians and colleagues from the partner organisations within D4Dairy for their cooperation and support.

This work was conducted within the **COMET-Project D4Dairy** (Digitalisation, Data integration, Detection and Decision support in Dairying). That is supported by BMVIT, BMDW and the provinces of Lower Austria and Vienna in the framework of COMET-Competence Centers for Excellent Technologies. The COMET program is handled by the FFG.