Data Integration and Use of Machine Learning Algorithms to Monitor Individual Cows' Health

Liliana Fadul-Pacheco, Hector Delgado and Victor E. Cabrera.

Department of Dairy Science, University of Wisconsin-Madison, Madison, WI.

BACKGROUND

- The "Dairy Brain" project consists on the daily collection and transformation of different data streams generated on dairy farms and the development of an integrated data repository, used for visualization and development of real-time Decision Support Tools (**DST**) (Fig. 1).

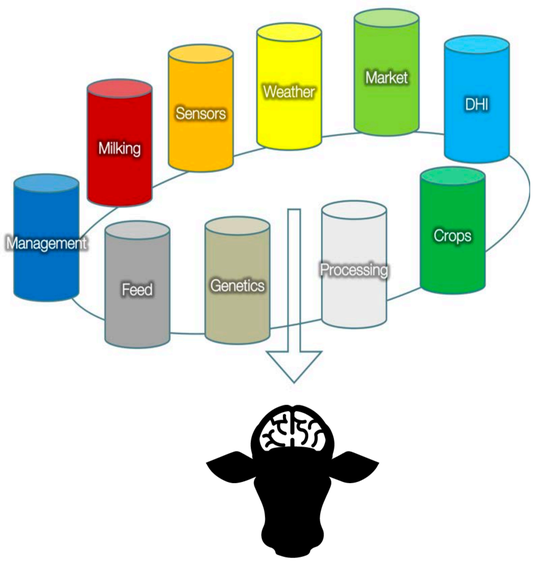


Figure 1. *Different data streams integrated through the Dairy Brain.*

- One DST aims to alert a possible Clinical Mastitis (**CM**) case before it happens.
- Somatic cell count is the most used metric to detect CM but is only available once a month.
- Integrating data from different data streams can be used to identify variables to predict the onset of CM on a daily basis.

AKNOWLEDGEMENTS

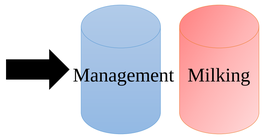
- Dairy Brain Team: Michael Ferris, Adam Christensen, Steve Wangen, Miron Livny, Derek Cooper, Andrew Maier, Ben Huebner, Heather White and Kent Weigel.
- Jessica Cederquist, UW-Madison.
- Mike Larson, Larson Acres.
- Mitch Breuning, Mystic Valley.

FOR MORE INFORMATION

https://dairybrain.wisc.edu/

MATERIALS & METHODS

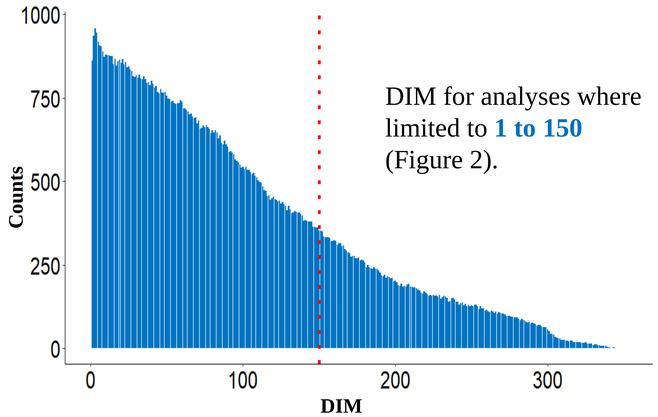
Records from 2 different data streams from 2016 -2018.



Integrated, edited and normalized (n=484,781 records from n=2,563 cows)

- **Management software**
  - Previous cases of CM
  - Previous cases of metritis, retained placenta, abortion, ketosis and/or metritis
  - Lactation
  - Days in milk (**DIM**)
  - Age at 1st calving
  - Pen
- **Milking software**
  - Milk production (L)
  - Milk conductivity (mS/cm)
  - Milk temperature

Figure 2. *Frequency CM cases from 1 to 350 days in milk (**DIM**)*



DIM for analyses where limited to **1 to 150** (Figure 2).

- One of the limitations of the data was the low number of CM cases: 715 (0.14% of the records), therefore, balancing was necessary.
  - Different techniques to balance the data were tested (i.e., down sampling, over sampling, Synthetic Minority Over-sampling Technique (SMOTE) and Random Over Sampling Examples (ROSE).
- Different classification machine learning algorithms were tested.
  - 75% of the data was used as training data.
- Variables included in the algorithms were:
  - Difference of milk production between milkings
  - Difference of milk conductivity between milkings
  - Lactation
  - DIM
  - Age at 1st calving
  - Pen
  - Previous cases of retained placenta, abortion, ketosis and/or metritis.

RESULTS

- The SMOTE technique for balancing the data gave the best results.
- The **random forest** algorithm had the best performance.
- Best results were achieved using data from the **7 previous milkings** before the reported case of CM.

Figure 3. *Δ* in milk conductivity between sick (1) and healthy (0) animals 7 milkings before the onset of CM.

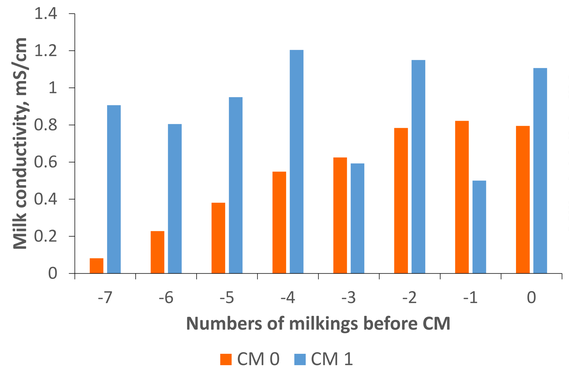
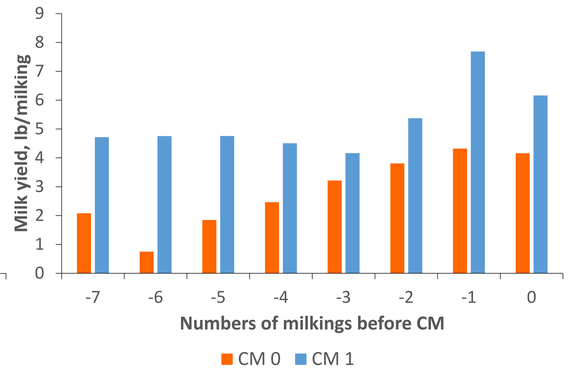
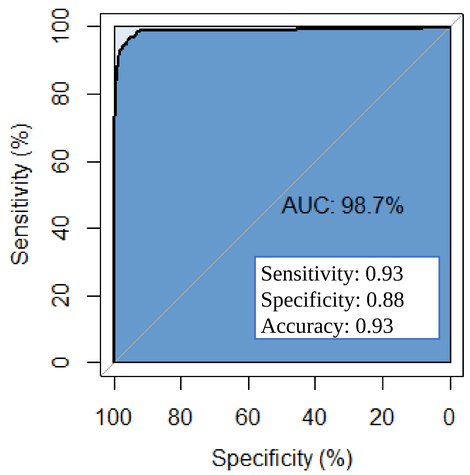


Figure 4. *Δ* in milk production between sick (1) and healthy (0) animals 7 milkings before the onset of CM.



- Cows with CM compared to healthy cows had higher absolute mean differences of milk conductivity (0.91 vs. 0.86 mS/cm; *P* <0.001) and milk production (4.70 vs. 4.29 lbs.; *P* <0.001), Figures 3 and 4, respectively.

Figure 5. *Receiver Operating Characteristics curve (ROC)* of the *prediction of CM with **Random Forest (RF)** algorithm*



- The RF algorithm will be certain in detecting **93%** of the CM cases with **93%** of true positive cases and **88%** of true negative cases.
- In other words, the algorithm will correctly classify:
  - 93% of the CM cases
  - 88% of the healthy cases

APPLICATION

- Farmer will receive daily list of cows at risk of contracting CM.
- Once the alert is emitted, a close-up follow up of the cow can be done.
- According to the health protocols and the evolution of the cow, farmer can better decide the course of actions.

TAKE HOME MESSAGES

- Real time data integration is an essential element to develop DST.
- The algorithms could be used as a monitoring tool to flag cows that are at risk of contracting CM and follow them closely on a daily basis.
- The integration of other data streams (e.g., sensors), could help improve the predictions