

94 **Machine learning algorithms for early prediction of clinical mastitis.** L. Fadul-Pacheco\*, H. Delgado, and V. E. Cabrera, *University of Wisconsin, Madison, WI.*

Somatic cell count is the most used metric to detect CM. However, SCC data are only available once a month. Integrated data on a continuous basis for every milking are available through the “Dairy Brain” project at the University of Wisconsin-Madison. These data could be used to more accurately predict the onset of CM on a permanent basis. We analyzed records from 2 different data streams in a Wisconsin farm (2016–2018): 1) milking system (milk production [kg] and milk conductivity [mS/cm], and 2) management system (CM, metritis, and retained placenta cases, lactation, and DIM). Days in milk were limited to 1 to 150 and lactations were grouped between 1, 2 and 3+ (n = 681,759 records from n = 3,319 cows). One of the limitations of the data was the low number of CM cases (981), which accounted for only 1% of the records. Therefore, balancing the data was necessary. The SMOTE technique gave the best results for balancing the data. With the resulting data set, various classification machine learning algorithms were tested using 75% of the data as training data. Included variables were the difference of milk production and milk conductivity between

milking, lactation groups, DIM, and previous cases of CM, retained placenta, and metritis. Significant variables were the difference of milk production and milk conductivity between milkings, lactation groups, and DIM. Two algorithms, random forest and gradient boosting, showed the best performance. Also, best results were achieved using data from the 5 previous milkings before the reported case of CM. For random forest the specificity was 0.74 and the sensitivity was 0.70, whereas for the gradient boosting results were the opposite, a lower specificity, but a higher sensitivity (0.58 and 0.82, respectively). Cows with CM compared with healthy cows had higher absolute mean difference for milk conductivity (0.92 vs. 0.86 mS/cm) and milk production (2.22 vs. 1.95 kg), respectively. Results show that the algorithms can predict well cases of CM; however, additional validation is required. Likewise, the integration of other data streams such as genetics, sensors and diet changes, could help improve the prediction accuracy.

**Key Words:** data integration, clinical mastitis, machine learning