FREIE UNIVERSITÄT BERLIN

Fachbereich Veterinärmedizin der Freien Universität Berlin
INSTITUT FÜR VETERINÄR-EPIDEMIOLOGIE UND BIOMETRIE

Time series data analysis to predict the status of mastitis in dairy cows by applying machine learning models to automated milking systems data

Muhammad N. Dharejo, Olivier Kashongwe, Thomas Amon, Tina Kabelitz, Lukas Minoque and Marcus G. Doherr

Introduction

- Mastitis is a significant disease affecting dairy cows, leading to economic losses and welfare concerns
- Early detection and management of mastitis are crucial for:

  > minimizing the risks to animal health & welfare > maximizing economic gains
- Traditional detection methods rely on manual inspections, which are labor-intensive
- Automated Milking Systems (AMS) generate large datasets that can be leveraged using Machine Learning (ML) for early mastitis prediction
- The aims of this study were to apply ML techniques to AMS data to predict mastitis occurrence:
  - one day prior to its observation and
  - on the day of mastitis observation.

Materials & methods

Dataset:

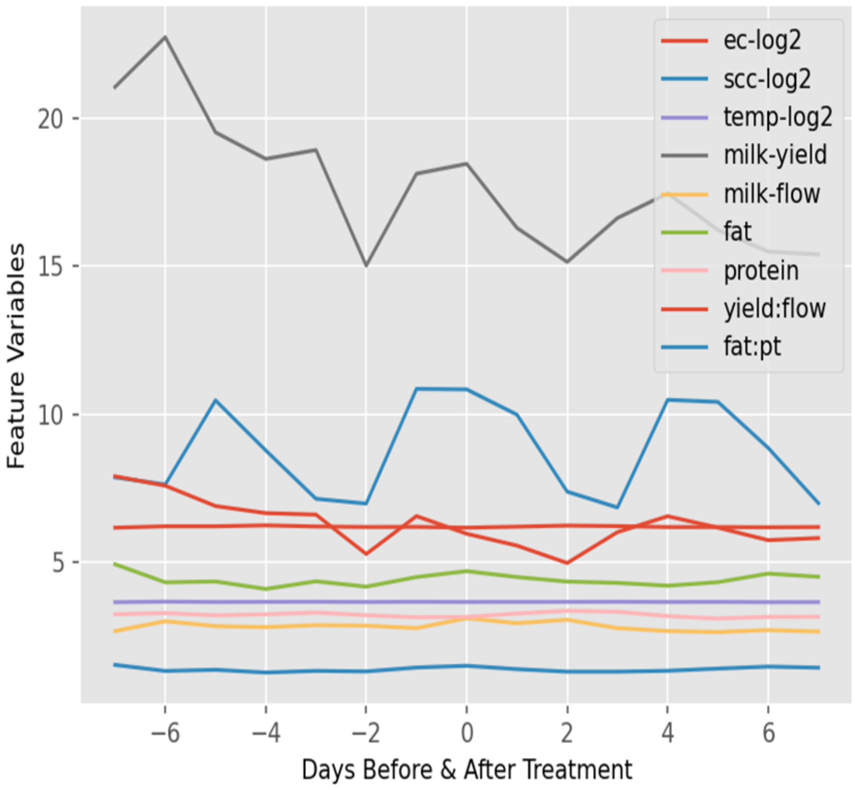
- Original data: AMS milk records 3-4 times per day per cow: 2.73 million observations
- Number of Individual Cows: 1790
- Number of Farms: 2
- Time Duration: 2019-2022 (4 years)
- X variables 7 (details on next slide…)
- Y variable 1
- Time resolution adjusted to one time (average value per day per cow)
- Total Observations 953,270
- Total negative obs. 942290 (days without treatment)
- Total Positive obs. 2250 (days with treatment)

Data format tranformation (auto-regressive order)

Seven predictor (X) variables

1. E-conductivity
2. Somatic cell count
3. Milk temperature
4. Milk yield
5. Milk flow
6. Fat content
7. Protein content

Individual pattern over time



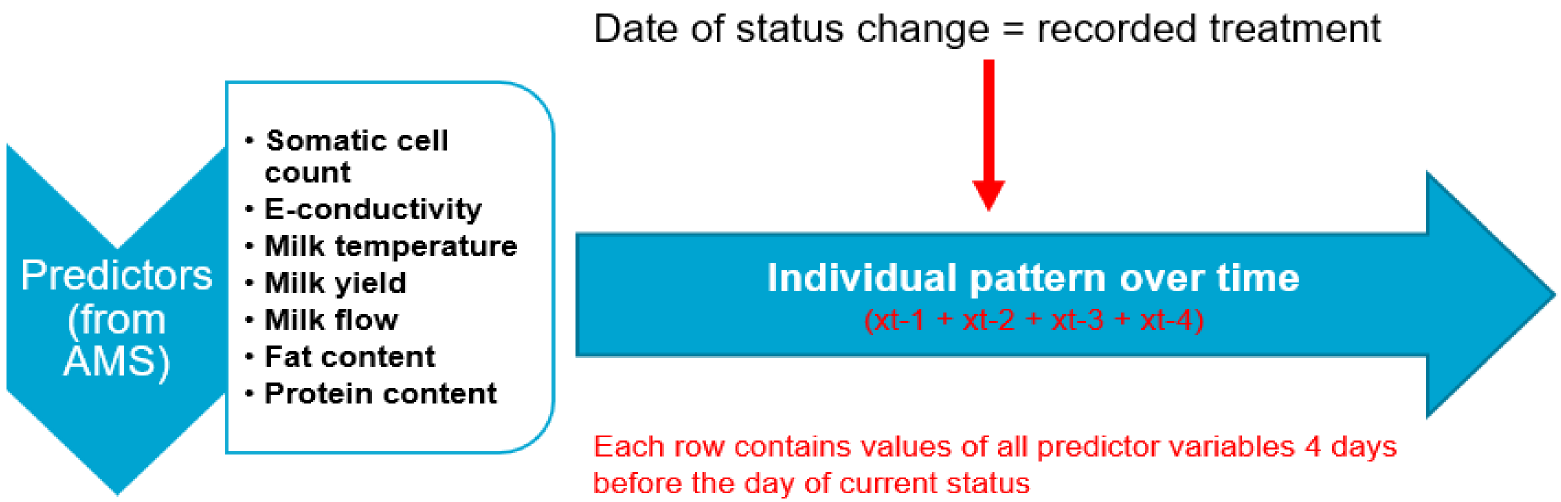
Auto-regressive data format

e-conduct	econd-1	econd-2	econd-3	econd-4
70.33	72.75	72.00	73.00	74.25
69.67	70.33	72.75	72.00	73.00
71.00	69.67	70.33	72.75	72.00
73.67	71.00	69.67	70.33	72.75
70.33	73.67	71.00	69.67	70.33

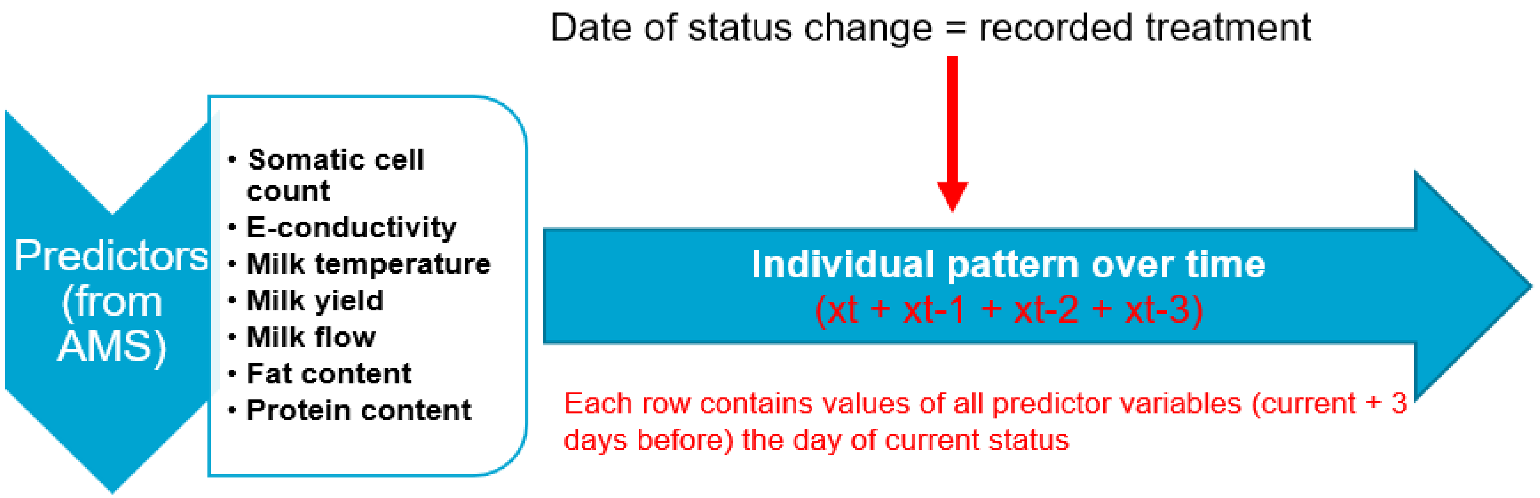
xt + xt-1 + xt-2 + xt-3 + xt-4

Model Formats

Format-1: Mastitis prediction one day prior



Format-2: Mastitis prediction on same day of observation



Training and testing of ML models

- Dataset was split into: >training (75%) and >testing (25%)
- Training dataset (highly class imbalance) >Negative = 722662 >Positive = 1800
- Synthetic minority oversampling technique (SMOTE)
  - After oversampling >Negative = 722662 >Positive = 722662
- Testing dataset (No oversampling) >Negative = 208299 >Positive = 450
- Six ML models applied:

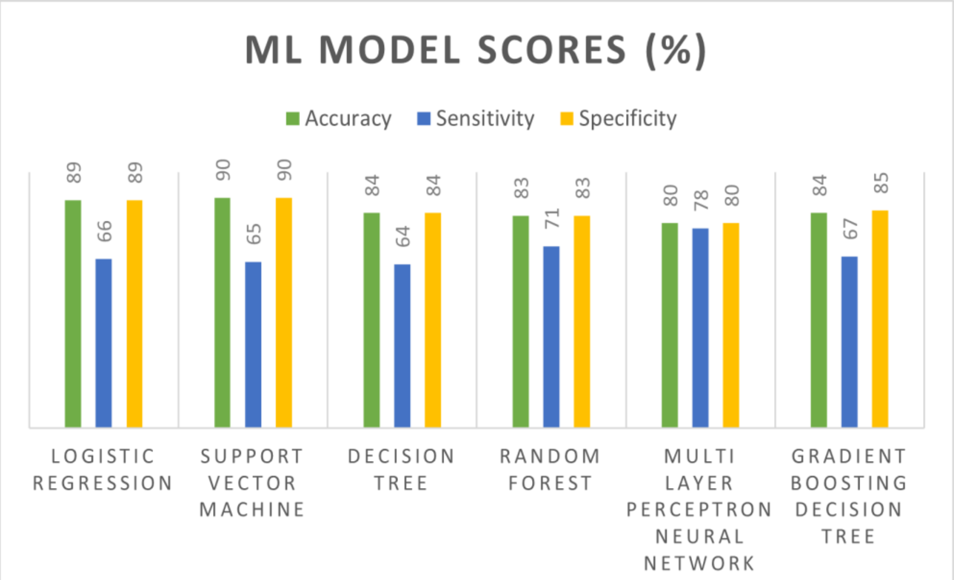
  > Logistic regression (LR) > Support vector machines (SVM) > Decision tree (DT)

  > Random forest (RF) > Gradient boosting decision tree (GBDT)

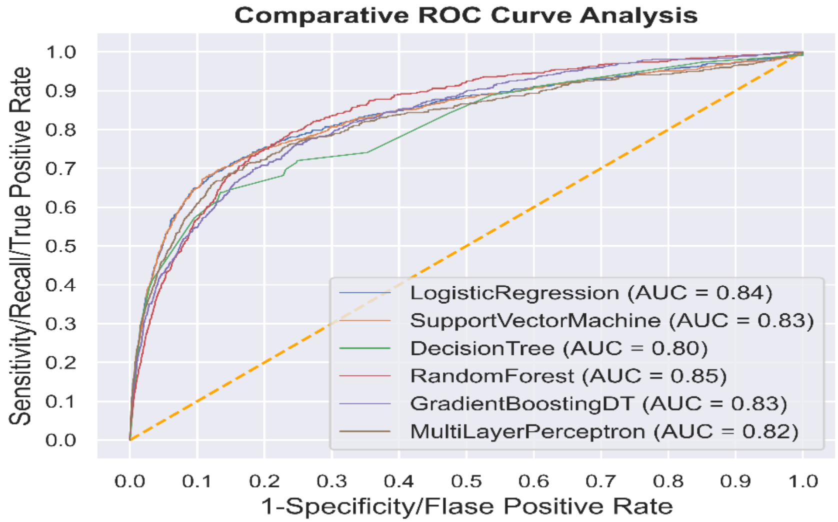
  > Multi-layer perceptron neural network (MLP-NN)
- GridSearchCV: > Hyper-parameter tuning > 5-fold cross validation

Results part-1 (mastitis prediction one day before)

- Over all
  - Accuracy 80-90%
  - Sensitivity 64-78%
  - Specificity 80-90%
- Highest accuracy and specificity > support vector machine
- Highest sensitivity > multi layer perceptron

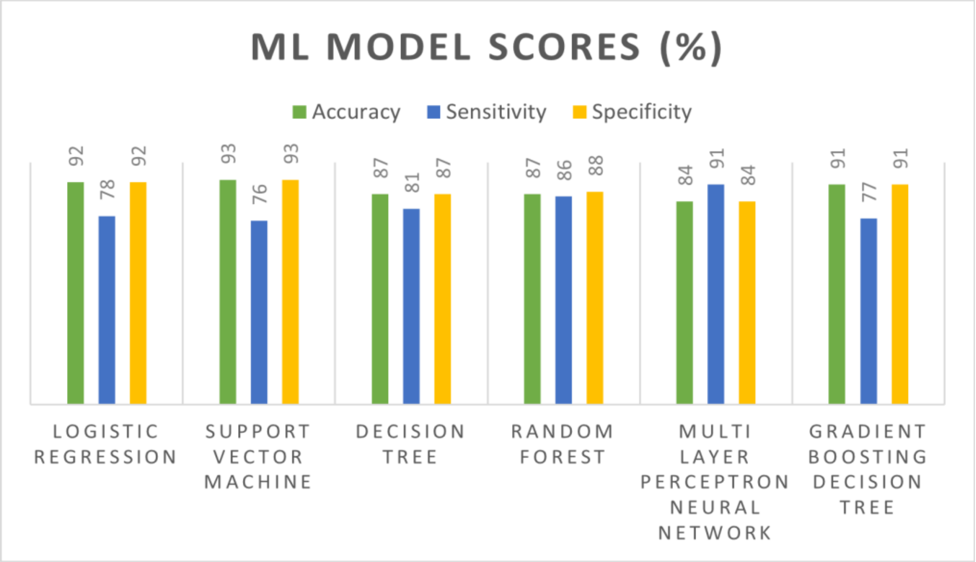


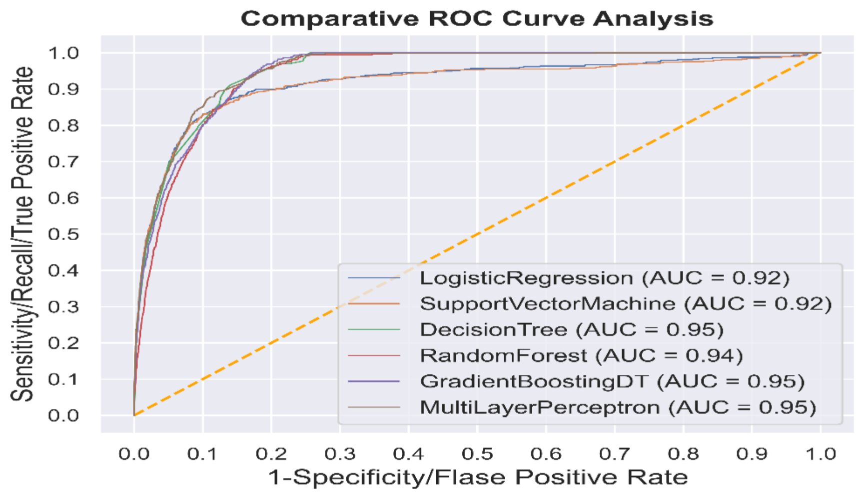
- Overall moderate predictive accuracy
- Area under curve (AUC) score 0.8-0.85
- Random forest classifier had greatest AUC
- Decision tree scored lowest



Results part-2 (mastitis prediction on the day of observation)

- Over all
  - Accuracy 84-93%
  - Sensitivity 76-91%
  - Specificity 84-93%
- Highest accuracy and specificity > support vector machine
- Highest sensitivity > multi layer perceptron
- Overall all models scored an area under curve (AUC) greater than 0.9 (high accuracy)
- Decision tree, gradient boosting and multi layer perceptron had greatest AUC
- Logistic regression and support vector machine scored lowest AUC





Discussion

- This study developed a framework of ML models for mastitis prediction on a time series dataset generated by AMS
- Since this is an ongoing area of research, a targeted level of sensitivity and specificity for prediction models on AMS data is not yet decisively defined.
- Some studies have suggested that Mastitis detection models should have Sensitivity of ≥80% (Hillerton, 2000; Hogeveen et al., 2010, 2021)
- International Standard Organization (ISO, 2007) recommends that Sensitivity be >70% with an Specificity level of 99%.
- However, the requirements mentioned above have rarely been met to date (Khatun et al., 2018; Anglart et al., 2021; Hogeveen et al., 2021).

Conclusion

- Our findings indicated moderate to high accuracy of ML models to predict mastitis
- Study also demonstrated the robustness of time series AMS data to predict mastitis events in future
- However, as each model had its own strengths and weaknesses, therefore these findings have certain limitations.
- We propose inclusion of additional variables from AMS records and integration of other sensorial data for further improvement of ML models in future studies.

Acknowledgments

ATB
Leibniz-Institut für
Agrartechnik und Bioökonomie

Gefördert durch

Bundesministerium
für Ernährung
und Landwirtschaft

aufgrund eines Beschlusses
des Deutschen Bundestages

Projektträger

Bundesanstalt für
Landwirtschaft und Ernährung