

Supervised machine learning-based prediction of bovine anaplasmosis in Florida beef cattle: A 10-year retrospective study

A. Megahed,¹ DVM, PhD; R. Bommineni,² DVM, PhD, DACVM, DACPV; M. Short,² DVM; J.H.J. Bittar,¹ DVM, MS, PhD

¹Department of Large Animal Clinical Sciences, College of Veterinary Medicine, University of Florida, Gainesville, FL 32608

²Division of Animal Industry, Florida Department of Agriculture and Consumer Services, Kissimmee, FL 34741

Introduction

The lack of updated information on the prevalence and risk factors of bovine anaplasmosis (BA), across the United States poses challenges in accurately assessing production losses and developing effective prevention and control strategies. Supervised machine learning (SML) algorithms are potentially powerful tools that may be used for identifying risk factors associated with infectious diseases such as BA. Here, we compared 6 different SML models to predict *Anaplasma marginale* seropositivity in Florida beef cattle.

Materials and methods

We used a dataset of 1,289 beef blood sample records from the Bronson Animal Disease Diagnostic Laboratory that were submitted for *Anaplasma marginale* antibody testing from 2012 to 2022. The SML algorithms that we used were logistic regression (LR), decision tree (DT), gradient boosting (GB), random forest (RF), neural network (NN) and support vector machine (SVM).

Results

A total of 199 (15.4, 95% CI 13.5 to 17.4) serum samples were positive for *Anaplasma marginale*. The RF model was the best model for predicting *Anaplasma marginale* seropositivity in beef cattle indicated by the highest Kolmogorov-Smirnov (KS) statistic of 0.38, area under the receiver operating characteristic (AUROC) of 0.68, gain of 2.1; and lowest misclassification rate of 0.14. The LR model was the worst. The RF model showed that *Anaplasma marginale* seropositivity can be best predicted by testing dairy cows of age ≥ 5 years raised in southern Florida during the wet season.

Significance

We concluded that the RF and other SML algorithms hold promise for predicting *Anaplasma marginale* seropositivity in beef cattle. It seems that the wet season, are highly predictive risk factor for BLV seropositivity in beef cattle of 5 years or older raised in southern Florida. These findings could help to develop predictive tools for effective screening for BA, and to develop early warning systems, and implement effective preventive strategies.

