Machine Learning to predict tuberculosis in cattle from the state of Sao Paulo, Brazil

Leyde Pereira

LEC Pereira[1], AS Ferraudo[2], AR Panosso[2], AAB Carvalho[1], LA Mathias[1], AC Saches[3], KS Hellwig[4], RA Ancêncio[5]

[1]Preventive Veterinary Medicine Department, São Paulo State University, School of Agricultural and Veterinarian Sciences, Jaboticabal, Jaboticabal, Brazil

[2]Science and Exact, São Paulo State University, School of Agricultural and Veterinarian Sciences, Jaboticabal, Jaboticabal, Brazil

[3]Engineering, Federal Institute of São Paulo, Votuporanga, Brazil

[4]Agricultural Defense, Secretariat of Agriculture and Supply, Campinas, Brazil

[5]Department of Nursing and Public Health, São Paulo University, Ribeirão Preto, Brazil

Contact: leyde.vet@gmail.com

Tuberculosis is a well-known and worldwide spread zoonosis. In Brazil 1.594.787 cases were confirmed cases since 2001, where, in Sao Paulo state, 8.226 deaths were reported. This study aims to present steps related to the use of machine learning algorithms for predictive analysis for bovine tuberculosis. For this, an application was made based on data from farms in state of São Paulo, Brazil, of an epidemiological survey, using a specific questionnaire, carried out on farms (n = 1,743). Response variable was presented by apparent prevalence of positive properties for disease, and predictors by (k = 77) predictors related to type of farm, type of lactation, number of animals on property. Application was organized according to following steps: division of data in training (75%) and testing (25%), pre-processing of predictors, learning and model evaluation. In the learning step, algorithm for adjusting gradient boosted trees models was used. The hyperparameters of algorithms were optimized by 10-fold cross-validation, to select those corresponding to best models. Models showed an accuracy of 88.07%, with an error in learning process equal to 3%. In the test / model validation procedure (n = 436), an error in 12% estimate was observed. Five important predictors were daily milk production, number of cows, type of farm, bovine breed and slaughter of adult animals. Proportion of false positives among all individuals whose response of interest was observed was 2.06%, and proportion of false negatives among those with a response of absent interest was 9.86%. It is hoped that, with increase in trained surveillance to detect the disease and availability of data, it will be possible to develop predictive models of machine learning with potential to efficiently assist professionals in disease control and assist in education program in animal health

Key messages:

- Predictive analyzes in health: application for tuberculosis in cattle from the state of Sao Paulo, Brazil.
- An infectious disease and zoonosis important to the world that needs support to develop means to control and consequently eradicate it.