Identifying risk factors for ovine respiratory processes by using Bayesian networks

Javier Galapero [a,*], Sara Fernández [a], Carlos J. Pérez [b], Fernando Calle-Alonso [b], Joaquín Rey [c], Luis Gómez [a]

[a] Histology and Pathological Anatomy, Department of Animal Medicine, Faculty of Veterinary Medicine, University of Extremadura, Avda de la Universidad s/n, 10003 Cáceres, Spain
[b] Biostatistics Unit, Department of Mathematics, Faculty of Veterinary Medicine, University of Extremadura, Avda de la Universidad s/n, 10003 Cáceres, Spain
[c] Infectious Pathology and Epidemiology Unit, Department of Animal Health, Faculty of Veterinary Medicine, University of Extremadura, Avda de la Universidad s/n, 10003 Cáceres, Spain

ARTICLE INFO

ABSTRACT

A proposal is put forward to use Bayesian networks to identify risk factors for pulmonary consolidation. An experiment was conducted with 410 fattening lambs from five feedlots in Extremadura (southwestern Spain). Environmental conditions (temperature, relative humidity, and ammonia concentration) were recorded during the study period. In a microbiological study, *Mycoplasma* spp. and *Pasteurellaceae* were obtained by conventional culture and identified by nested polymerase chain reaction. After slaughter, lungs were collected and examined macro- and microscopically (histological type and pulmonary consolidation). To the best of the authors' knowledge, Bayesian networks have not before been used to relate the presence/absence of pulmonary consolidation to environmental conditions, *Mycoplasma* spp., *Pasteurella* spp., and histological changes. The results showed that the main factors causing ovine inflammatory respiratory processes and pulmonary consolidation were temperature, relative humidity, and *Mycoplasma* spp. Control of these factors may help reduce the incidence of pulmonary consolidation.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Analyzing animal health data is a complex task since the relationships between variables are usually not known. In fact, determining information about the way the variables are related is usually an objective of analysis. Lewis et al. (2011) discussed the potential of using Bayesian networks as analytical tools for processing complex animal health data. Specifically, they proposed the use of structure discovery to identify variables that may be associated with health status.

Bayesian networks belong to a family of probabilistic graphical models (see, e.g., Jensen, 1995). These graphical structures are mathematically rigorous and intuitively understandable, and can be used to represent knowledge about an uncertain domain. A Bayesian network is a directed acyclic graph (a graph with directed edges between vertices) with an associated set of probability distributions that enables an effective representation and computation of the joint probability distribution over a set of random variables. The visual representation of the random variables' dependency structure is especially useful. Bayesian networks combine principles from graph theory, probability theory, computer science, and statistics (Jensen and Nielsen, 2007).

These networks can be implemented to search for the structure between variables by finding an optimal directed acyclic graph for the dataset in hand, providing information about the possible relationships linking the variables involved. The network can then be used to infer conditional probabilities. Bayesian networks can also be built using reliable subjective information provided by an expert instead of searching for the structure. Moreover, a mixed approach can be taken by first searching for the structure based on data, and then adjusting the structure using subjective information.

Bayesian networks have been extensively used in many fields of study, especially in artificial intelligence. They are gradually being introduced for the analysis of data in the veterinary field. McKendrick et al. (2000) applied Bayesian networks to aid in the differential diagnosis of tropical bovine diseases. Otto and
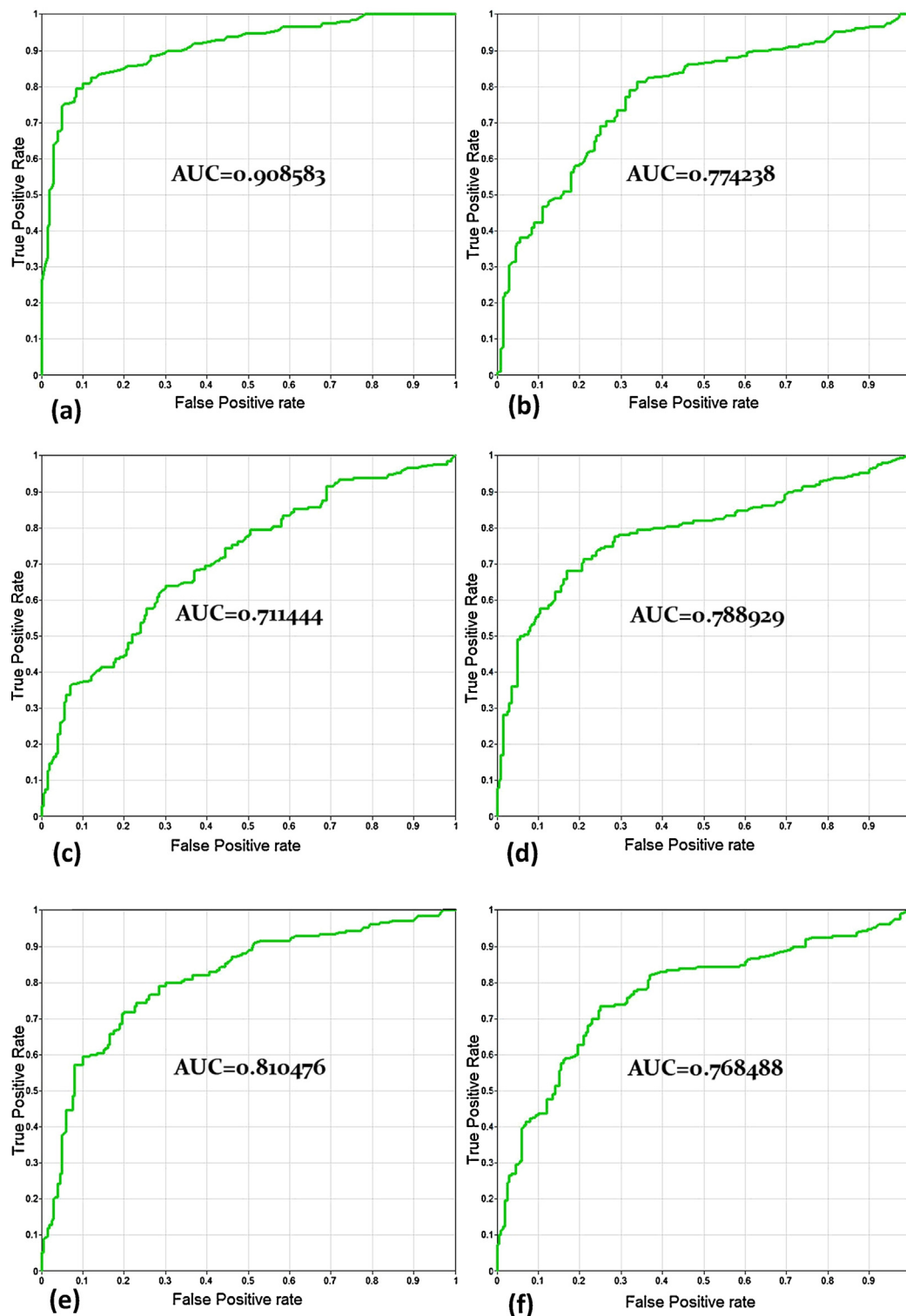
Fig. 1. ROC curves and AUC for Bayesian networks built using six algorithms: (a) Greedy Thick Thinning; (b) Bayesian Search; (c) Essential Graph Search; (d) Tree Augmented Naive Bayes; (e) Augmented Naive Bayes; (f) Naive Bayes.

Kristensen (2004) proposed a biological model based on a Bayesian network to determine risk factors for infection with *Mycoplasma hyopneumoniae* in swine for slaughter. Ettema et al. (2009) used a Bayesian network to estimate the probability of claw and digital skin diseases by combining cow- and herd-level information. Jensen et al. (2009) used a Bayesian network to model the causes of leg disorders in finisher herds. McCormick et al. (2013) applied these networks to the identification of environmental conditions

that influence disease in pigs, and Firestone et al. (2013) analyzed the associations between risk factors and the infection status of horses in an equine influenza outbreak in Australia.

In the present study, we focus on pathological respiratory processes that lead to direct and indirect losses such as mortality, weight loss, low conversion rates, and greater numbers of pulmonary lesions at slaughter (Goodwin et al., 2004; Lacasta et al., 2008). The agents and risk factors that may influence these pro-