Computers and Electronics in Agriculture 111 (2015) 179–185

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Prediction of subacute ruminal acidosis based on milk fatty acids: A comparison of linear discriminant and support vector machine approaches for model development

E. Colman [a], W. Waegeman [b], B. De Baets [b], V. Fievez [a,*]

[a] Laboratory for Animal Nutrition and Animal Product Quality, Ghent University, Proefhoevestraat 10, 9090 Melle, Belgium
[b] Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

ARTICLE INFO

Article history:
Received 26 August 2013
Received in revised form 4 December 2014
Accepted 2 January 2015

Keywords:
Acidosis
Classification
Support vector machines
Linear discriminant analysis

ABSTRACT

Subacute ruminal acidosis (SARA), characterized by low rumen pH, is one of the most important metabolic disorders in dairy cattle. As dairy cows experiencing SARA often do not exhibit overt clinical symptoms, diagnostic biomarkers in milk are of interest. Data of six acidosis induction experiments with rumen-fistulated dairy cows were combined to assess the potential of milk fatty acids (FA) to identify acidotic cases, based on three threshold values often reported in literature, i.e. time pH < 5.6 of 180 min/d and 283 min/d and time pH below 5.8 of 475 min/d (N = 442 cases, of which 111–165 acidotic cases, depending on the applied threshold value). Both linear discriminant analysis (LDA) as well as support vector machines (SVM) were used to develop classification models, with SVM based on two common types of kernel functions (linear kernels and Gaussian radial basis function kernels) and including either the whole milk FA profile (41–69 milk FA, depending on the experiment) or a selected number of milk FA (i.e. both odd and branched chain FA and biohydrogenation derivates of poly-unsaturated FA, 13–16 FA). Both evaluation of the performance of individual classification models as well as comparison of models was based on the area under the receiver operating characteristic (ROC) curve. Non-linear models developed through a radial kernel based SVM approach seemed of particular interest when including all milk FA as model features. However, linear models based on the selected group of milk FA most often performed as good as the non-linear models including all milk FA, with the former being least time consuming and more cost-effective, both from a computational as well as an analytical perspective. However, combination of all data sets only resulted in good classification models when including data of each dataset upon training the model, whereas model performance decreased dramatically in case of cross-dataset cross-validation. This indicates an important impact of the origin of the datasets on the performance of the model which should be taken into account in further exploration of prediction models of SARA.

1. Introduction

Subacute ruminal acidosis (SARA) is a well-recognized digestive disorder that is an increasing health problem in most dairy herds (Enemark, 2008). Dairy cows experiencing SARA often do not exhibit any clear, overt clinical symptoms and symptoms are often delayed from the time of incidence. The only reliable and accurate diagnostic test for SARA is measuring ruminal fluid pH (Keunen et al., 2002), although ruminal pH varies considerably at different locations in the rumen and during the day. As a result, rumen pH conditions that define SARA are still an issue and various parameters and associated threshold values have been proposed. Time or area (time × pH) below a certain pH as calculated from continuous pH monitoring, seems a potential indicator (Keunen et al., 2002; Dragomir et al., 2008) with thresholds of 283 and 475 min of pH below 5.6 or 5.8, respectively (AlZahal et al., 2007). Accordingly, there is an increasing interest in milk compounds, such as milk fatty acids (FA) as potential diagnostic tools of rumen function (Fievez et al., 2012). Milk odd and branched chain FA (OBCFA; i.e. iso FA either with an even and odd number of carbon atoms as well as anteiso and linear FA with an odd number of carbon atoms) are of particular interest in this respect as they have been linked to the

Abbreviations: FA, fatty acids; LDA, linear discriminant analysis; OBCFA, odd and branched chain fatty acids; ROC, receiver operating characteristic; SVM, support vector machines.

* Corresponding author.
E-mail address: Veerle.Fievez@Ugent.be (V. Fievez).

http://dx.doi.org/10.1016/j.compag.2015.01.002
0168-1699/

rumen fermentation pattern (Vlaeminck et al., 2006; Cabrita et al., 2009; Craninx et al., 2008) which strongly depends on rumen pH, e.g. decreasing rumen pH is related to decreasing molar proportions of acetate and butyrate and increasing molar proportions of propionate (Erfle et al., 1982).

Additionally, FA accumulating during the microbial conversion of poly-unsaturated FA (rumen biohydrogenation) depend on the rumen environment with e.g. predominant switches towards the C18:1 trans 10 pathway at low ruminal pH (e.g. Enjalbert et al., 2008).

In earlier work we suggested the potential of monitoring the milk FA profile for diagnosis of SARA (Colman et al., 2010, 2012, 2013). However, in these papers, models to effectively classify acidotic vs. non-acidotic cases either were not presented (Colman et al., 2012, 2013) or were based on data from a single experiment only (Colman et al., 2010), limiting the robustness of such models. Additionally, the whole milk FA profile was considered in most of these evaluations, whereas from an analytical perspective, SARA classification based on a selected number of milk FA would be beneficial. Moreover, the earlier classification model was developed using linear discriminant analysis (LDA), traditionally the most popular method to solve a classification problem. Nevertheless, recently, support vector machines (SVMs) and related kernel methods have become increasingly popular tools for learning tasks such as classification (Hastie et al., 2009). This relatively new technique is related to and contains elements of non-parametric applied statistics, neural networks and machine learning (Karatzoglou and Meyer, 2006). However, SVMs have several advantages over neural networks, including generalization ability, ease of training, a mechanism to model structured data and, most importantly, the generation of a unique solution (Hamm, 2007; Hastie et al., 2009).

As such, the goal of this study was to 1/develop and cross-validate classification models using data from multiple acidosis induction experiments, 2/compare two types of modeling approaches, i.e. the traditional LDA vs. the newer SVM approach and 3/assess whether models based on a selected number of milk FA, i.e. OBCFA and C18 biohydrogenation intermediates, are similarly performing to models based on the complete FA profile to diagnose SARA.

2. Material and methods

2.1. Acidosis induction experiments

Four datasets of acidosis induction experiments in dairy cows were available for this classification experiment based on the milk FA profile. In all datasets, rumen pH data originated from continuous measurements by an indwelling pH probe in rumen-fistulated dairy cows. Cases of SARA were classified for each dataset based on three different threshold values from literature: time rumen pH below 5.6 = 180 min/d (Gozho et al., 2005); time rumen pH < 5.6 = 283 min/d (AlZahal et al., 2007) and time rumen pH below 5.8 = 475 min/d (AlZahal et al., 2007). Milk samples of the evening and the morning after the pH registration day were sampled and pooled for further analysis. The milk fatty acid profile of each sample was determined based on gas chromatography. An overview and brief description of each dataset is given below and in Table 1.

Dataset 1. The experimental design, sampling and rations were described before by Colman et al. (2012). Acidosis induction was obtained through stepwise replacement of a standard concentrate A (conc A) by a concentrate rich in quickly fermentable carbohydrates (QFCH) (conc B) until 100% conc B, followed by an increase of the total amount of conc B. Two sub-experiments, performed in the Netherlands (Schothorst Feed Research, Lelystad, the Netherlands), were included in this dataset. In the first sub-experiment, the treatments were applied sequentially, whereas the second sub-experiment was designed according to a Latin square set-up. Briefly, in *sub-experiment 1.1*, three rumen-fistulated cows were subjected to five different treatments during 33 days: (1) control (forage/concentrate ratio (*F*/*C*) = 65/35), (2) stepwise replacement of the standard concentrate (conc A) by conc B until 100% (*F*/*C* = 65/35), (3) increase of the total amount of conc B (depending on the cow: *F*/*C* = 48/52, 42/58 and 24/76), (4) treatment with a buffer solution and (5) control ration. A 3 × 3 Latin square design was developed for *sub-experiment 1.2* in which three rumen-fistulated cows were used. Each period lasted 21 days. During the first 14 days of the period, the cows received a control diet (*F*/*C* = 68/32) while in the last 7 days conc A was stepwise replaced by conc B (from 100% conc A to 44% conc A and 56% conc B) and the amount

Table 1
Overview of four datasets from acidosis induction experiments in dairy cows. Datasets 1 and 3 consisted of two sub-experiments each (1.1, 1.2, 3.1 and 3.2, respectively).

Dataset.	Anim. (*N*)	DIM ($\mu \pm$ SD)[a]	Period (days)	Ration[a]	Measurements (sampling time) [a]	Samples (*N*)
1.1	3	127 ± 61.9	33	(1) Control diet (*F*/*C* = 65/35) (2) Standard conc. → [a] wheat-based conc. (3) Total amount of conc. increased (4) Buffer addition (5) Control diet	Rumen pH, milk FA(1 d ration 1; 4 d ration 2; 4 d ration 3; 2 d ration 4; 4–7 d ration 5)	49
1.2	3	170 ± 30.3	3 × 21	(1) d1–d13: control (2) d14–d21: standard conc. → wheat-based conc. + ↑total conc.	Rumen pH, milk FA (d13–d21 of each period)	80
2	12	236 ± 45	42	(1) Week 1: control ration (2) Week 2–5: standard conc. → wheat-based conc. (3) Week 6: ↑ total conc.	Rumen pH & VFA, milk FA (d2 and d7 of each week)	144
3.1	4	84 ± 29	2 × 42	(1) Week 1–5: control ration (2) Week 6: 21% DM → ground wheat and barley conc.	Rumen pH & LPS (5 d weeks 5 and 6), milk FA (2 d weeks 5 and 6), microbial composition	31
3.2	4	175 ± 75	42	(1) Week 1: control ration (2) Week 2–6: alfalfa hay → alfalfa pellets	Rumen pH & LPS (5 d each week), milk FA (2 d each week), microbial composition	46
4	3	146 ± 189	3 × 21	(1) d1–d13: starch-rich, sugar-rich or starch-and protein rich conc., *F*/*C* 75/25 (2) d14–d18: *F*/*C* 75/25 → 50/50 (3) d19–21: *F*/*C* = 50/50	Rumen pH, milk FA (d10–d21 of each period)	107

[a] DIM = days in milk; μ = average value; SD = standard deviation; conc. = concentrate; → = stepwise replacement; LPS = lipopolysaccharide; VFA = volatile fatty acids; FA = fatty acids.