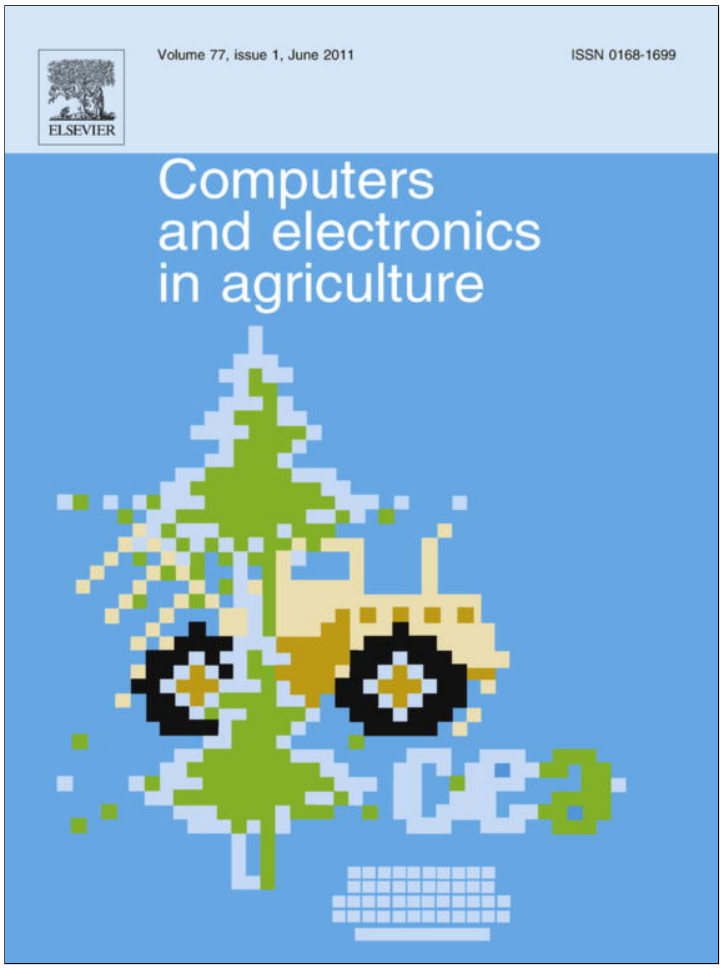
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Computers and Electronics in Agriculture 77 (2011) 86–94

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Sensor measurements revealed: Predicting the Gram-status of clinical mastitis causal pathogens

C. Kamphuis [a,*], H. Mollenhorst [a], H. Hogeveen [a,b]

[a] Department of Farm Animal Health, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 7, 3584CL Utrecht, The Netherlands
[b] Chairgroup of Business Economics, Wageningen University, Hollandseweg 1, 6706KN Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 27 October 2010
Received in revised form 23 January 2011
Accepted 29 March 2011

Keywords:
Gram-status prediction
Sensor measurements
Decision-tree induction

ABSTRACT

Automatic milking systems produce mastitis alert lists that report cows likely to have clinical mastitis (CM). A farmer has to check these listed cows to confirm a CM case and to start an antimicrobial treatment if necessary. In order to make a more informed decision, it would be beneficial to have information about the CM causal pathogen at the same time a cow is listed on the mastitis alert list. Therefore, this study explored whether decision-tree induction was able to predict the Gram-status of CM causal pathogens using in-line sensor measurements from automatic milking systems. Data were collected at nine Dutch dairy farms milking with automatic milking systems and included 140 bacteriological cultured CM cases with sensor measurements of electrical conductivity, colors red, green, and blue and milk yield for analyses. In total, 110 CM cases were classified as Gram-positive CM cases and 30 as Gram-negative. Stratified randomization was used to divide the data in a training set ($n = 96$) for model development, and a test set ($n = 44$) for validation. The decision tree used three variables to predict the Gram-status of the CM causal pathogen; two variables were based on electrical conductivity measurements, and one on measurements of the color blue. This decision tree had an accuracy of 90.6% and a kappa value of 0.76 based on data in the training set. When only those CM cases were considered with extreme high probability estimates for their Gram-status (either positive or negative), 74% of all records in the training set could be classified with a stratified accuracy of 97.1%. When validated, the decision tree performed poorly; accuracy dropped to 54.5% and the kappa value to −0.20. The stratified accuracy calculated for 75% of all records in the test set was 66.7%. Predicting the CM causal pathogen showed a similar poor result; the decision tree had an accuracy of 27.9% and a kappa of 0.12, based on data in the test set. Based on these results, it is concluded that decision-tree induction in conjunction with sensor information from the electrical conductivity, color, and milk yield provide insufficient discriminative power to predict the Gram-status or the CM causal pathogen itself.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Mastitis is one of the most frequent and costly diseases in the dairy industry (Halasa et al., 2007; Viguier et al., 2009), with clinical mastitis (CM) being responsible for about 75% of the total costs of mastitis for an average Dutch farm milking 65 cows and with a bulk somatic cell count (SCC) of 200,000 cells/ml (Hogeveen et al., 2010). The disease has negative effects on farm economics due to production losses, treatment costs, and costs of culling (McDougall et al., 2007; Huijps et al., 2008; Viguier et al., 2009). It is important that cows that suffer from CM are detected during the early onset of the disease to eliminate the infection and to prevent recurrence, and that the appropriate type and amount of antimicrobial is applied by the correct route and for the correct time to increase cure rate (Hillerton and Kliem, 2002).

When milking with an automatic milking system, also called a milking robot, CM is detected using two diagnostic tests: the first is the CM detection model on the automatic milking system itself. This model uses sensor measurements as input and gives a CM alert as output. These CM alerts appear on a mastitis alert list to warn the dairy farmer for cows that need attention (“management by exception”, Hogeveen and Ouweltjes, 2003). The second test involves the visual confirmation of CM and is conducted by the dairy farmer, who checks the alerts from the mastitis alert list he or she thinks necessary. If a CM case is visually confirmed, it is the responsibility of the dairy farmer to decide on antimicrobial treatment. Initially, the choice of antimicrobial treatment is based in absence of any knowledge about the CM causal pathogen, as the whole process from taking a milk sample, culturing it, and determining the pathogens involved usually takes 3 days. Choosing an

* Corresponding author. Tel.: +64 (0) 7 858 3750; fax: +64 (0) 7 858 3751.
E-mail addresses: claudia.kamphuis@dairynz.co.nz (C. Kamphuis), H.Mollenhorst@uu.nl (H. Mollenhorst), H.Hogeveen@uu.nl (H. Hogeveen).

0168-1699/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2011.03.012

Author's personal copy

inappropriate treatment protocol or an ineffective antimicrobial may be a reason cure rates of 75%, as suggested at the introduction of commercial mastitis tubes or mastitis syringes, are not met in practice (Erskine et al., 2003). In order to make a more informed decision on which antimicrobial to use, it would be beneficial for dairy farmers to have information about the CM causal pathogen at the same time a cow is listed on the mastitis alert list.

In the past, studies were conducted to use several information sources to provide or to predict the CM causal pathogen or the Gram-status of the pathogen involved. For example, Godden et al. (2007) and MacDonald et al. (2010) presented the potential of on-farm culture systems in order to replace bacteriological culturing that is done in the laboratory. Despite attractive benefits, major disadvantages of this approach are the need of experienced personnel and good on-farm laboratory conditions in order to achieve proper accuracy. Instead of moving the laboratory to the farm itself, other studies focused on using other information sources to predict CM causal pathogens or Gram-status; expert knowledge was used by Jones and Ward (1990) and Kim and Heald (1999), cow information by Milne et al. (2003) and Steeneveld et al. (2009), cow data and farm management practices by Heald et al. (2000), and somatic cell count patterns by De Haas et al. (2004). These studies showed varying predictive performances of the classification models developed, or concluded that collection of the necessary information would take too long for an appropriate prediction at cow-level. Kamphuis et al. (2008a) reported a difference in sensor measurement patterns of the electrical conductivity and the colors red, green, and blue between healthy quarters and quarters infected with CM, but they did not check whether different pathogens show different sensor measurement patterns. Espada and Vijverberg (2002) expected color sensor measurements to be useful for abnormal milk detection and potentially also for pathogen prediction. However, that study was based on a small data set, collected during 1 month at two farms, with only 6 cows showing abnormal milk in one quarter.

Kamphuis et al. (2010b) developed a CM detection model using decision-tree induction. Sensor measurements of the electrical conductivity and colors red, green, and blue were used as input for this detection model. The final model showed a good performance (a sensitivity of 40% at a specificity of 99%), considering the narrow time window in which the model should alert for CM and the inclusion of quarter milkings with a less clear mastitis status. Decision-tree induction could very well be used to extract knowledge from sensor data, amongst that of color sensors, for Gram-status prediction. The main objective of this study is to explore whether sensor measurements from automatic milking systems can be used by decision-tree induction to predict the Gram-status of the CM causal pathogen. In addition, the potential of predicting the actual CM causal pathogen itself is evaluated as well.

2. Materials and methods

2.1. Data collection

The collection of data used for this study has been described in detail by Kamphuis et al. (2010b). In short, raw sensor data and observations of CM were collected at nine commercial Dutch dairy farms milking automatically (version A2 (n = 10) or A3 (n = 2); Lely Astronauts N.V., Maassluis, The Netherlands) from November 2006 until March 2009. Farm characteristics are summarized in Table 1. Both versions of automatic milking systems use the same sensors: the electrical conductivity was measured with a sensor consisting of a collection bin with known content between two electrical pins. In addition, the colors red, green, and blue were measured with a light emitting diode (LED) sensor combination consisting of three light sources emitting red, green, and blue light, and a receiver to measure the reflected light intensity (Espada and Vijverberg, 2002). Raw sensor data of the electrical conductivity, color, and an estimation of quarter milk yield were collected by connecting a remote computer to each of the 12 automatic milking systems. From these raw sensor measurements, 1065 potentially descriptive variables were developed using a data flow diagram (Kamphuis et al., 2008a, 2010a). These variables described characteristics (level, variability, and shape) of sensor measurements patterns from each quarter milking. Cows that raised suspicion of being affected by CM, according to the own criteria of the participating dairy farmer, were visually checked to confirm a CM case. The dairy farmers' suspicion could be based on the mastitis alert lists, but also on other information, like SCC data, the presence of clots on the milk filter, or through direct observations of redness or swollenness of the udder. This approach resulted in different procedures per farm to check quarters for CM. By introducing a scoring protocol, however, the assessment of the actual CM status and the procedure to collect milk samples for bacteriological culturing was standardized: a scoring protocol instructed the dairy farmers to visually score the 5th and 6th squirts of milk of all quarters they checked using a clean black paddle as commonly used for the California Mastitis Test (without using the reagent normally used when applying this test). When visually normal, the milk was scored as (1). When abnormal, the milk was scored as (2) watery milk, (3) flakes, (4) clots, (5) serum-like milk, or (6) milk with blood. If a dairy farmer decided the CM infection was severe enough to start an antimicrobial treatment, they were asked to first take two milk samples for bacteriological culturing and to store these milk samples in their refrigerator. Furthermore, they were instructed to record the cow's identification number, quarter, date and time, and the CM score assigned to the visually checked quarter. Every 4–6 weeks, dairy farmers were visited to collect sensor data stored by the remote computers, scoring forms, and milk samples if any. Milk samples were bacteriological cultured by the Veterinary Microbiological Diagnostic Centre (VMDC, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands) according to the standards of the National Mastitis Council (Harmon et al., 1990).

2.2. Data preparation

In order to combine visual observations of CM with sensor data, each visual quarter milk assessment was linked with sensor data from the most recent quarter milking, within a 24 h time window prior to the assessment time, recorded for that same quarter by the remote computer (for a more detailed description, see Kamphuis et al., 2010b).

By the end of the data collection period, 2003 quarters were checked visually for having CM (see Table 1), of which 1593 could be combined with sensor measurements within a 24 h time window. Quarters that received a CM score from 2 through 5 were considered as quarters with CM (n = 348, see Table 1). These included 74 quarters with watery milk, 217 quarters with flakes, 50 quarters with clots and 7 serum-like quarters. There were 6 quarters containing blood (CM score 6). These were excluded from further analyses, as milk with blood does not necessarily imply a CM case but can also be caused by damage in the udder or of the teat. From the 348 CM cases, 243 quarters were sampled for bacteriological culturing. For those CM cases with only one milk sample (11.7% of 243 quarters), only this milk sample was evaluated for inclusion criteria for further analyses. For the remainder 89.3% of the quarters with two milk samples, both of these milk samples were evaluated for inclusion criteria. Only CM caused by major CM causal pathogens (*Escherichia coli*, *Klebsiella* spp. *Staphylococcus aureus*, *Streptococcus dysgalactiae*, and *Streptococcus uberis*) were included

Author's personal copy

Table 1
Farm characteristics of participating farmers included in the study. Per farm, characteristics included the number of automatic milking systems (AMS), start of data collection, the number of individual cows milked during the data collection period, the number of quarters checked for clinical mastitis (CM), the number of quarters with CM, and the number of CM cases with sensor measurements and bacteriological culturing results that are included in the analyses.

Farm	AMS (*n*)	Start of data collection	Individual cows milked (*n*)	Quarters visually checked (*n*)	Quarters with CM (*n*)	CM included for analyses (*n*)
1	1	October 25 2007	64	67	14	11
2	1	August 14 2007	71	94	58	29
3	1	December 8 2006	70	437	81	17
4	1	January 30 2007	97	325	77	32
5	1[a]	September 13 2007	55	18	5	1
6	1[a]	October 4 2007	79	143	7	1
7	2	January 30 2007	101	627	26	11
8	2	March 9 2007	129	72	28	15
9	2	December 7 2006	106	220	52	23
Total	12		772	2003	348	140

[a] Farms with an A3 Lely Astronaut automatic milking system (Lely Industries N.V., Maassluis, The Netherlands). All other farms used an A2 Lely Astronaut automatic milking system.

for further analyses. If a culture within one milk sample or over milk samples of the same quarter showed a combination of major CM causal pathogens (e.g., an *S. aureus* in combination with an *E. coli*), the quarter was excluded for further analyses. Culture results that were combinations of one major CM causal pathogen with a 'mixed culture', 'no growth', or '*Bacillus* spp.' within one milk sample or over milk samples of the same quarter were included. In total, 140 quarters with CM were included for further analyses (see Table 1); all of them were labeled according to the major CM causal pathogen cultured within a milk sample or over milk samples of the same quarter. This resulted in 26 quarters infected with *E. coli*, 4 with *Klebsiella* spp., 22 with *S. aureus*, 72 with *S. uberis*, and 16 with *S. dysgalactiae*. Quarters infected with *E. coli* or *Klebsiella* spp. were classified as Gram-negative CM cases ($n = 30$), and all others were classified as Gram-positive CM cases ($n = 110$). Table 2 summarizes the distribution of the major CM causal pathogens isolated from bacteriological culturing of the 140 CM cases, their Gram-status, and their CM score recorded by the participating dairy farmers.

2.3. Model development and validation for Gram-status prediction

In order to develop a decision tree model that predicts the Gram-status of a CM causal pathogen, a training set and a test set were constructed. Two thirds of all data were selected for training and the remaining third for testing, where quarters were randomly stratified according to the pathogen that caused the CM. This stratification resulted in a training set with 21 Gram-negative and 75 Gram-positive CM cases. The test set included nine Gram-negative and 35 Gram-positive CM cases.

All CM cases in the training set had 1065 independent variables based on sensor measurements. To prevent overfitting of the decision tree, first all independent variables that had an information gain ratio higher than 0.01 were selected. The information gain of an independent variable X is based on the change in information value (or entropy) of a dataset S with respect to the dependent variable Y, after partitioning S using the values of independent variable X. A decision tree that selects independent variables based on their information gain tends to prefer variables with a large number of possible classes. Therefore, the decision tree in this current study selects independent variables based on their information gain ratio, which takes into account the number of classes and records per class of X (Witten and Frank, 2005). Table 3 lists the 16 independent variables in the training set that met this information gain ratio requirement.

To develop a decision tree based on data in the training set, the J48 algorithm as implemented in WEKA (Witten and Frank, 2005) with default settings was used. A decision tree is a graphic representation of a divide-and-conquer approach of a classification problem and consists of nodes at which a variable is tested. Based on its information gain ratio, an independent variable is selected to split a data set at the first node. For each possible outcome of the test involved at that node, a branch is made ending in a daughter node. Next, the process can be repeated for each branch, using only those records that actually reach that branch (Kamphuis et al., 2010b; Witten and Frank, 2005). In an ideal situation, the tree stops developing at the time all records at a node have the same classification, e.g., all records involve a Gram-positive CM case. In reality, this is often not the case, and the records that flow into an end node are used by the decision tree to calculate a probability estimate for the Gram-status for each record in that node. It becomes even more difficult when the decision tree encounters records with missing values for variables used at test nodes. To solve this problem, such a record with missing values for variables used at test nodes is proportionally split into pieces – or weights – and the parts are sent down each branch and from there right on down to the leaves of the sub-trees involved. The split into pieces is accomplished by using a numeric weight between zero and one, and the weight for a branch is chosen to be proportional to the number of records going down that branch that do have a value

Table 2
Bacteriological culturing results of 140 clinical mastitis cases with at least one milk sample being bacteriological cultured, their Gram status, and their clinical mastitis score as recorded by the participating farmers.

Main pathogen isolated	Gram status	Score assigned to quarter by farmer				Total (*n*)
		Watery (*n*)	Flakes (*n*)	Clots (*n*)	Serum-like (*n*)	
Escherichia coli	Negative	6	13	6	1	26
Klebsiella spp.	Negative	2	1	1	–	4
Staphylococcus aureus	Positive	2	19	1	–	22
Streptococcus dysgalactiae	Positive	–	14	2	–	16
Streptococcus uberis	Positive	5	40	22	5	72
Total		15	87	32	6	140

Table 3
Abbreviations, full descriptions, and the information gain ratio value of the 16 independent variables that were selected from the training set for predicting Gram-status, and of the one independent variable (abbreviation and full description) that was selected from the training set for predicting the actual CM causal pathogen. The selected variables all had an information gain ratio higher than 0.01

Variable name (abbreviated)	Variable description	Information gain ratio
Predicting Gram-status		
wcdtmax	Maximum value of the electrical conductivity of a whole quarter milking	0.2923
fcdtmax	Maximum value of the electrical conductivity of the first 500 mL of quarter milk produced	0.1855
lcdtmax	Maximum value of the electrical conductivity of the last 500 mL of quarter milk produced	0.1693
lcdtmean	Mean value of the electrical conductivity of the last 500 mL of quarter milk produced	0.1693
wcdtrangeP2	The range (maximum–minimum value) of the electrical conductivity of the whole quarter milking compared with the mean value for the range of the two previous milkings for that same quarter	0.1511
fcdtmiddevPR36	The deviation of the electrical conductivity between the actual midpoint observation and the expected midpoint observation from the first 500 mL of quarter milk produced compared with the mean value for this variable of all milkings within the previous 36h for that same quarter	0.1419
wblueincrFLQ2	The increase from the first to the last sensor measurement of the whole quarter milking of the color sensor blue compared with the mean value of the increase of two other quarters within the same cow milking that showed the highest values for this increase	0.1075
startmilkflowdelayP2	Time difference between teat cup attachments and start of milk flow, compared with the mean time difference of the two previous milkings of that same quarter	0.1008
lredmiddevP2	The deviation of the color sensor red between the actual midpoint observation and the expected midpoint observation from the last 500 mL of quarter milk produced compared with the mean value for this variable of the previous two milkings for that same quarter	0.0732
wcdtrange	The range (maximum–minimum value) of the electrical conductivity of the whole quarter milking	0.0675
lcdtincrFM	The increase from the first to the middle sensor measurement of the electrical conductivity of the last 500 mL of quarter milking produced	0.0603
lcdtincrFL	The increase from the first to the last sensor measurement of the electrical conductivity of the last 500 mL of quarter milking produced	0.0464
lcdtslope	The increase from the first to the last sensor measurement of the electrical conductivity of the last 500 mL of quarter milk produced/the number of sensor measurement recordings in this last 500 mL of quarter milk produced	0.0464
lcdtmin	The minimum value of the electrical conductivity of the last 500 mL of quarter milk produced	0.0401
wcdtslope	The increase from the first to the last sensor measurement of the electrical conductivity of the whole quarter milking/the number of sensor measurement recordings in quarter milking	0.0313
wcdtincrFL	The increase from the first to the last sensor measurement of the electrical conductivity of the whole quarter milking	0.0260
Predicting CM causal pathogen		
lcombincrFMP2	The increase from the first to the middle sensor measurement of the mean color value (average of red, green, and blue sensor measurements) of the last 500 mL of quarter milk produced compared with the mean value for the increase from the first to the middle sensor measurements of the previous two milkings for that same quarter	0.1390

Author's personal copy

Author's personal copy

for the variable used at the test node. A split record may be further split at a lower node. Eventually, the various parts – or weights – of the record will reach end nodes and the final CM probability estimate for these split records is computed by multiplying the probabilities for CM at the end nodes by the weight of that record that reaches that specific end node (Witten and Frank, 2005).

The number of Gram-negative CM cases in the training set was much lower than the number of Gram-positive CM cases (21 vs. 75, respectively). Therefore, a cost matrix was applied to balance the data (Kamphuis et al., 2010a) in order to prevent the decision tree to simply classify all CM cases as caused by the majority group (Gram-positive CM causal pathogens). A cost matrix is a square matrix with its size dependent on the number of levels of the dependent variable. This means that a 2×2-sized cost matrix was used for model development with zeros at the diagonal. Values off the diagonal represent the 'costs' that are made by the decision tree for misclassifying a CM case. As there are 3.57 more Gram-positive CM cases in the training set than there are Gram-negative, the costs for misclassifying a Gram-negative case was set at 3.57 and the costs for misclassifying a Gram-positive case was set at 1. By doing this, the dataset was artificially balanced.

To prevent the selection of correlated variables in the development of a Gram-status predicting decision tree based on 96 CM cases in the training set, first 16 univariate decision trees (including the cost matrix) were developed; all 16 variables listed in Table 3 were used one by one to split the training data according to their Gram-status. To select the best univariate decision tree, Cohen's kappa value was used. Cohen's kappa value is a test of agreement between two tests, in absence of a gold standard, after chance agreement is removed from consideration. Cohen's kappa value is calculated as the actual agreement between two tests beyond chance divided by the potential agreement beyond chance; the actual agreement beyond chance is calculated as the observed agreement minus the expected agreement (chance) and the expected agreement beyond chance is calculated as one minus the expected agreement (Cohen, 1960; Dohoo et al., 2009). The decision tree resulting in the highest kappa value was selected, and a forward selection procedure started including the variable chosen in the first selection round and all 15 remaining independent variables added one by one. This procedure of forward selection was continued until the kappa value no longer improved. The decision tree that was developed at that point was used to predict the Gram-status of the CM cases in the test set, where the output was a probability estimate for a quarter to have a CM infection caused by a Gram-positive or a Gram-negative pathogen. Kappa value and accuracy of this test set were evaluated as well.

2.4. Model development and validation for pathogen prediction

Decision-tree induction was also applied to sensor data to develop a model that predicts the CM causal pathogen itself. However, as the number of CM cases caused by *Klebsiella* spp. was limited ($n = 4$), these CM cases were excluded from the training set ($n = 3$) and test set ($n = 1$). The remaining 93 records in the training set were used to select independent variables with an information gain ratio higher than 0.01. This selection resulted in one variable that fulfilled this requirement (see Table 3). Therefore, a further forward selection procedure was not performed. A final decision tree was trained using the single independent variable that had a gain ratio higher than 0.01 and a cost matrix that balanced the classes of different CM causal pathogens to the majority class (*S. uberis*) in a similar way as was done for the Gram-status predicting model. This means that a 4×4-sized cost matrix was used, with zeros on the diagonal, and with values of 2.72 as cost factor of misclassifying an *E. coli* infection, 3.27 for misclassifying an *S. aureus* infection, and 4.45 as cost factor for misclassifying an *S. dysgalactiae* as being *S. uberis*. Costs for misclassifying *S. uberis* for any of the other pathogens were set at 1.0. Also this decision tree was evaluated with data from the test set, containing 43 CM cases, with the kappa value and the accuracy as evaluation measures.

The data mining software WEKA (Witten and Frank, 2005) was used to select variables based on their information gain ratio, to perform the forward selection of independent variables, to develop the final decision tree, and to compute kappa values and accuracies. The PROC SURVEYSELECT procedure in SAS (version 9.1, SAS Institute Inc., Cary, NC) was used for stratified randomization of the data to create a training and a test set.

3. Results

The most frequently isolated pathogen was *S. uberis* (Table 2). Most quarters were scored as having flakes or clots in their milk (85%). Compared to Gram-positive CM cases, Gram-negative CM cases were more often scored as having watery milk (6.4% vs. 26.7% for Gram-positive and Gram-negative pathogens, respectively).

There were 16 independent variables selected for Gram-status prediction based on their information gain ratio being >0.01 (Table 3). The information gain ratios ranged from 0.0260 to 0.2923. Most of these selected variables are based on the sensor measuring the electrical conductivity, although the sensors measuring the colors red and blue also seem informative (based on their information gain ratio value). Selected variables are based on absolute values, but also on comparisons with previous milkings or on other quarters within the same cow milking.

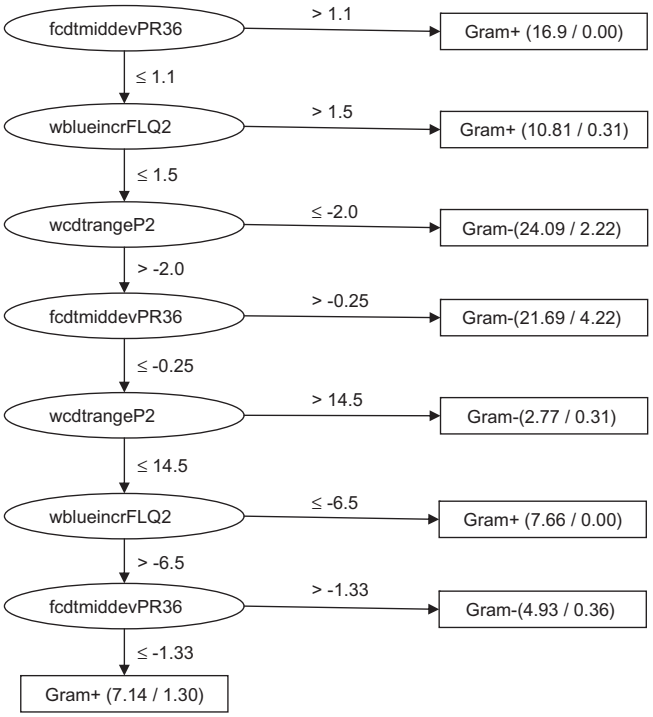


Fig. 1. Final decision tree using three different independent variables to classify 96 quarters in the training set for to their Gram-status. The ovals represent test nodes at which a variable is tested (variable names are explained in Table 3). The rectangles represent end-nodes at which a prediction is made (being infected with a Gram-positive (Gram+) or Gram-negative (Gram−) clinical mastitis causal pathogen) for those records reaching this rectangle. The first figure between brackets presented in these rectangles summarizes the total number of records reaching this rectangle; the second figure represents the records that are misclassified by the decision tree. The variables selected had a gain ratio >0.01.

Author's personal copy

The final decision tree (with the highest kappa value) for Gram-status prediction is graphically presented by Fig. 1. The accuracy of this decision tree based on data from the training set was 90.6%, and the kappa value was 0.76. The tree used seven tests (ovals in Fig. 1) to classify the 96 records in the training set based on three independent variables; two were based on electrical conductivity measurements, and one was based on the color sensor blue. The rectangles in Fig. 1 represent end nodes where records are classified as being caused by a Gram-positive or Gram-negative CM causal pathogen.

The figures between brackets in the end nodes can be used to calculate the probability estimates for a CM case to be caused by a Gram-positive or a Gram-negative CM causal pathogen. For example, if a CM case ends up in the second end node from the top, it will be classified by the decision tree as being caused by a Gram-positive CM causal pathogen. This end node contains 10.81 CM cases, of which 0.31 are misclassified (and thus were labeled by the gold standard or the bacteriological culturing results as Gram-negative CM cases). The decision tree will assign a probability estimate of 97.1% (calculated as $100 * ((10.81–0.31)/10.81)$ that this CM case is infected with a Gram-positive CM causal pathogen, and a probability estimate of 2.9% (calculated as $100 * (0.31/10.81)$) to be infected with a Gram-negative CM causal pathogen. The figures in the rectangles show decimals due to the way the J48 algorithm deals with records with missing values for the variables on which a test at a node applies.

The higher the probability estimate, the more accurate the prediction for a Gram-positive CM case gets (Fig. 2). But also, the lower the probability estimate, the more accurate the prediction for a Gram-negative CM case gets. There were 65 quarters out of the 96 (67.7%) that received a probability estimate of >0.60 for being a Gram-positive CM case. All of these CM cases were indeed Gram-positive cases. An extreme low probability estimate for a Gram-positive CM case ($\leqslant$0.10), which is thus similar to an extreme high probability estimates for a Gram-negative CM case (>0.90), also shows an accurate prediction: 12 out of the 96 CM cases (12.5%) received such a high probability estimate for a Gram-negative CM case, and 83.3% of these 12 CM cases were indeed Gram-negative CM cases.

The accuracy of the final decision tree for Gram-status prediction dropped to 54.5% when applied to the test set. The kappa value decreased to −0.20. Fig. 3 explains visually the decrease in both accuracy and kappa value. First of all, even at extremely high probabilities (e.g., >0.90) for a Gram-positive CM case, still 7 out of the 27 CM cases (25.9%) were Gram-negative. And at extremely low probabilities (e.g., $\leqslant$0.10) for Gram-positive CM cases, still 4 out of the 5 CM cases (80%) were Gram-positive.

There was only one independent variable that fulfilled the requirement of having an information gain ratio higher than 0.01 for predicting the actual CM causal pathogen (Table 3). The final decision tree used that independent variable at five test nodes, each time using different threshold values in order to classify the records according to the CM causal pathogen. Based on the 93 records in the training set, the accuracy of this model was 34.4% and the kappa value was 0.19. When applied on the test set, accuracy dropped to 27.9% and the kappa value dropped to 0.12.

4. Discussion

Only 16 out of 1065 potentially predictive variables had an information gain ratio higher than 0.01 for predicting the Gram-status of a CM case (Table 3). This indicates that all other variables had an information gain ratio value so low, that these had no discriminative power at all to predict the Gram-status of a CM causal pathogen. The prediction of the CM causal pathogen itself seemed even more difficult with the use of sensor information, as only one variable out of the 1065 potentially predictive variables had an information gain higher than 0.01 (Table 3). The 17 potentially predictive variables that were selected, however, were mainly based on electrical conductivity measurements. Already in the early beginnings of developing CM detection models using sensor data, the electrical conductivity has been used as detection tool (e.g., Maatje et al., 1992) although the electrical conductivity has not been suggested to be useful for pathogen detection before. The final decision tree for predicting the Gram-status of CM causal pathogens used two out of three variables that were based on electrical conductivity measurements (Fig. 1). The third one was based on the color sensor blue. A first study on the potential of color sensors for pathogen prediction was conducted by Espada and Vijverberg (2002), which served as basis for the study described in this paper. Although that study included just a very small number of cases (6 cows with abnormal milk in one quarter), they concluded that abnormalities in milk (e.g., clots) caused a significant change of milk color in comparison with milk from other quarters in the same milking or with previous milkings. Results from the current study do confirm the potential of using color as detection or prediction tool. The three independent variables used by the final decision tree (Fig. 1) characterized the variability or shape of sensor measurements, and they were based on comparisons with previous quarter milkings or with other quarters in the same cow milking. This finding is in line with results from Kamphuis et al. (2008a). They concluded that sensor data from the electrical conductivity and the color sensors blue and green contained the most information for abnormal milk or CM classification, and that variables based on the variability or shape (e.g., the range or increase) of sensor measurement patterns could be as predictive as variables based on the level of sensor measurement patterns (e.g., the mean value).

Fig. 2 and 3 show the probability estimates to be infected with a Gram-positive CM causal pathogen that the final decision tree assigned to the CM cases in the training set and the test set, respectively. This indicates that the decision tree is more confident for some CM cases to be Gram-positive CM cases than for others.

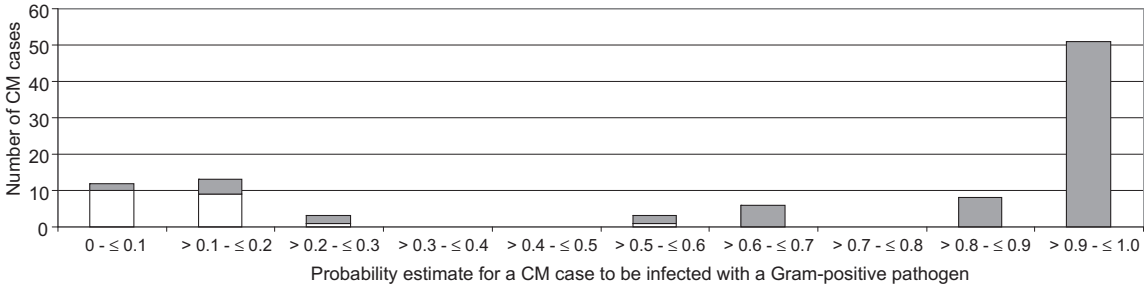


Fig. 2. The number of clinical mastitis (CM) cases in the training set infected with a Gram-positive or a Gram-negative CM causal pathogen (*y*-axis) per probability estimate for a CM case to be infected with a Gram-positive pathogen (*x*-axis). In total, 96 CM cases were included in the training set.

Author's personal copy

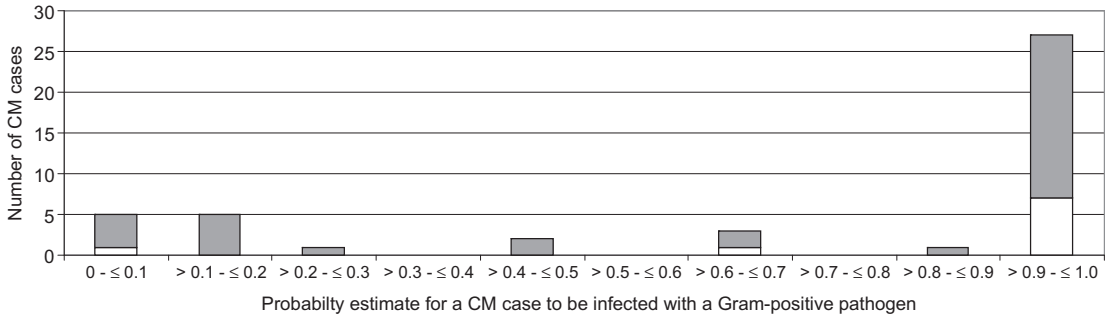


Fig. 3. The number of clinical mastitis (CM) cases in the test set infected with a Gram-positive or a Gram-negative CM causal pathogen (*y*-axis) per probability estimate for a CM case to be infected with a Gram-positive pathogen (*x*-axis). In total, 44 CM cases were included in the test set.

Results from both figures also indicate that Steeneveld et al. (2009) and Van der Gaag et al. (2009) made sensible suggestions that presenting a probability distribution for the causal CM pathogen would be more informative for dairy farmers as it reveals the uncertainty involved with a binary or multiclass classification. Both studies discuss the presentation of a stratified accuracy, a measure which is based upon different strata of the dataset under study. When this idea is applied to the current study, it would be possible to present to dairy farmers only those quarters with extreme high probability estimates for being a Gram-positive CM case (e.g., >0.80) or a Gram-negative CM case (e.g., >0.90, which is similar as a probability of $\leqslant$0.10 for being a Gram-positive CM case). By doing this, still 74% from the 96 quarters in the training set will have a probability estimate presented to the dairy farmer, with a stratified accuracy of 97.2%. For the test set, 75% will receive a probability estimate with a stratified accuracy of 66.7% (results not shown).

The three independent variables the final decision tree selected (Fig. 1) were not the three variables with the highest information gain ratios (Table 3). This somewhat surprising result may be explained by the fact that a cost matrix was applied to balance the distribution between Gram-positive and Gram-negative CM cases in the training set. In a situation where a decision tree is developed without a cost matrix, the figures – or weights – in the end nodes of a decision tree will add up to the number of records that are labeled Gram-positive or Gram-negative in the training set. However, when adding up the figures mentioned in the end nodes from Fig. 1, they add up to 48.0 for Gram-positive and for Gram-negative records. This shows that applying a cost matrix had effect in balancing the data, and this new balanced data may have an effect on how these records can be divided according to their Gram-status with the independent variables listed in Table 3.

After completing the forward selection procedure, a decision tree was developed based on three independent variables (Fig. 1) with the highest kappa value (0.76). This kappa value can be interpreted as a substantial agreement (Dohoo et al., 2009) between the classifications of the decision tree and the labeling of quarters according to their bacteriological culturing results as Gram-positive or negative CM cases. The accuracy of this final decision tree was 90.6%. For comparison, a simple model that would classify CM cases based on prevalence would have an accuracy of 66.7%, and a model that would classify all CM cases as Gram-positive cases would have an accuracy of 78.1%. These results suggest that the three sensor based variables used by the decision tree were able to discriminate between Gram-positive and Gram-negative CM cases. However, when the decision tree was applied to data not used for training to validate its performance, both the accuracy and the kappa value dropped dramatically to 54.4% and −0.20, respectively. These results indicate that the three selected variables based on data in the training set were unable to predict the Gram-status of the CM cases in the test set. So, although a forward selection procedure was used to prevent selection of correlated independent variables and a cost matrix was applied to balance the data in the training set, based on the results from the test set it has to be concluded that the developed decision tree was overfit to the training data. This conclusion was confirmed when two additional iterations to create training and test sets showed a similar trend of good performances with data from the training set and poor performances when validated with data from the test set (results not shown). The accuracy of 54.5% found when applying the decision tree to the test set is just slightly higher than the accuracy of 48% that White et al. (1986) observed for inexperienced clinicians when predicting the Gram-status of a CM case, and much lower than the accuracy of 79% reported by Milne et al. (2003) and 73% reported by Steeneveld et al. (2009). Accuracies and kappa values for predicting the CM causal pathogen itself were very poor in the current study: the decision tree showed an accuracy of 34.4% and a kappa of 0.19 based on data in the training set. The accuracy is lower than one would get with simply classifying all CM cases as *S. uberis* cases, which would result in an accuracy of 52.6%. Again performance dropped (an accuracy of 27.9% and a kappa of 0.12) when the developed decision tree was applied to data not used for training. This indicates that the decision tree was unable to use sensor information to predict the CM causal pathogen itself. The study of Steeneveld et al. (2009) also used cow information to predict the CM causal pathogen itself, and showed a model with an overall accuracy of 52%, and a stratified accuracy of 89% for 4% of all CM cases. Again, their results outperform the detection performance from the model developed in the current study.

Still, it may be a too harsh conclusion to exclude sensor measurements for future research in the field of these types of predicting models, and there are three reasons not to do so. First of all, the current study used only 140 CM cases. This is a low number in comparison with the 573 CM cases used by Milne et al. (2003) and the 3833 CM cases used by Steeneveld et al. (2009), but it is a high number of CM cases in the field of CM detection using sensor data. Studies in the field of automated CM detection often use a lot less CM cases for analyses (e.g., 19 CM cases used by Mollenhorst et al., 2010, and 36 CM cases used by Song et al., 2010). It is likely that the inclusion of more CM cases with sensor information and bacteriological results will improve the robustness of decision trees predicting the Gram-status or the CM causal pathogens itself.

Another reason causing the poor detection result is the classification problem itself and the distribution of pathogen species in the current study. Differentiating between quarter milkings with CM and without CM (healthy quarters) is probably easier for an algorithm than differentiating between already diseased quarter milkings, as the sensor measurement patterns are more likely to be different between truly healthy quarter milkings and truly

Author's personal copy

diseased ones. The distribution of pathogen species in the current study is strongly subject to which cows dairy farmers decided to check visually, and on how good they actually were in classifying the milk. For example, it has been suggested that watery milk without flakes is an indicator for an upcoming severe *E. coli* infection (Hogeveen et al., 1995). Results from Kamphuis et al. (2008b) showed that this category showed deviating mean sensor measurement values, especially for the sensors measuring the electrical conductivity and the color blue. However, it is also known that watery milk is a difficult to classify category (Rasmussen, 2005), and this may be the reason that just 6 out of 26 CM cases caused by *E. coli* are detected in the supposedly early stage of infection, and that more than half of the CM cases classified by the farmers as being watery to be caused by another pathogen species rather than *E. coli*. This misclassification of the CM status may cause the pathogen species linked to the wrong CM sensor measurement patterns, in this way introducing noise in the dataset and making it more difficult for the decision-tree algorithm to select independent variables more specific to the deviating sensor patterns specific for *E. coli* infection. Also, it is known that *S. aureus* infections are often subclinical, and develop slowly to clinical flare-ups (Harmon, 1994). When a dairy farmer knows a cow suffering from an *S. aureus* infection, he or she might decide to do nothing about these cases, causing the number of *S. aureus* CM cases in this study to be lower than they are in reality. In addition, due to the fact that the progress of infection is slow and mostly subclinical, with SCC levels tending to fluctuate at high levels long before but also long after a clinical flare-up (De Haas et al., 2002), the sensor measurement patterns may not be as distinct as for example the sensor measurement pattern of an early stage *E. coli* infection, where the SCC peaks within a short time window around the CM occurrence (De Haas et al., 2002). The expected lack of distinct sensor measurement patterns for an *S. aureus* infection therefore makes it more difficult for a decision-tree algorithm to select independent variables specific for this type of CM infection.

The third reason for not neglecting sensor measurements for future research is the ongoing development of new and improved sensors. Steeneveld et al. (2009) showed that whether a cow was sick or not, the color of milk, and the SCC from the 3- to 4-weekly Dutch national milk recording system were cow information sources that contributed significantly in the prediction for Gram-status or in the prediction of the CM causal pathogen itself. All these three aspects can be measured by sensors not used in the current study or by improved sensors compared to the ones used in the current study. For example, infrared thermography and sensors estimating SCC on-line are sensors that have been used for mastitis detection in previous research (e.g., Berry et al., 2003; Kamphuis et al., 2008b; Polat et al., 2010) and recent work from Song et al. (2010) indicated that the use of a new generation of color sensors (measuring LED transmittance) is expected to improve the detection of CM. So although the decision tree developed in the current study performed poorly on data from the test set, it could very well be that other sensors than used in the current study are able to contribute to predict the Gram-status or CM causal pathogens. And if such a Gram-status or CM causal pathogen predicting model becomes applicable in practice, this could affect the dairy farmer's choice on antimicrobial treatment. For example, a dairy farmer in practice being confronted with a quarter with an extreme high probability estimate for being a Gram-negative CM case (e.g., >0.90), might better start a supportive treatment to relieve the cow of systemic signs with fluids and other supportive care than to start an antimicrobial treatment (Pyörälä and Pyörälä, 1998; Erskine et al., 2003). If a reduction of antimicrobials used in the dairy industry is induced, the real or perceived concerns about antimicrobial use as a human health hazard by the regulatory sector and the milk consuming public may be reduced. Future work should include other sensors and more CM cases to develop a more robust decision tree for Gram-status prediction.

5. Conclusion

A decision tree for predicting the Gram-status of CM causal pathogens showed a kappa of 0.76 and an accuracy of 90.6% for the training set. The kappa value and accuracy decreased to −0.20 and 54.5%, respectively, when the decision tree was validated with data from the test set. These figures indicate that the developed decision tree was not a robust one. A similar result was found for a model predicting the CM causal pathogen itself: the kappa value and accuracy were 0.12 and 27.9% based on data from the test set. Based on these results, it is concluded that decision-tree induction in conjunction with sensor information from the electrical conductivity, color, and milk yield provides insufficient discriminative power to predict the Gram-status or the CM causal pathogen itself.

Acknowledgements

We acknowledge the participating dairy farmers for their contribution to the acquisition of the data used in this research. Also acknowledged is Lely Industries N.V. (Maassluis, The Netherlands) for providing addresses of dairy farmers working with automatic milking systems and their support with the installation of the data acquisition systems at the participating farms. This research is supported by the Dutch Technology Foundation STW, applied science division of NWO and the Technology Program of the Ministry of Economic Affairs.

References

Berry, R.J., Kennedy, A.D., Scott, S.L., Kyle, B.L., Schaefer, A.L., 2003. Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: potential for mastitis detection. Can. J. Anim. Sci. 83, 687–693.

Cohen, J., 1960. A coefficient of agreement for nomical scales. Educ. Physiol. Meas. 1, 37–46.

De Haas, Y., Barkema, H.W., Veerkamp, R.F., 2002. The effect of pathogen-specific clinical mastitis on the lactation curve for somatic cell count. J. Dairy Sci. 85, 1314–1323.

De Haas, Y., Veerkamp, R.F., Barkema, H.W., Gröhn, Y.T., Schukken, Y.-H., 2004. Associations between pathogen-specific cases of clinical mastitis and somatic cell count patterns. J. Dairy Sci. 87, 95–105.

Dohoo, I., Martin, W., Stryhn, H., 2009. Veterinary Epidemiology Research. In: McPike, S.M. (Ed.), vol. 2. VER Inc., Charlottetown.

Erskine, R., Wagner, S., DeGraves, F., 2003. Mastitis therapy and pharmacology. Vet. Clin. North Am. Food Anim. Pract. 19, 109–138.

Espada, E., Vijverberg, H., 2002. Milk colour analysis as a tool for the detection of abnormal milk. In: First North American Conference on Robotic Milking. Wageningen Pers, Wageningen, pp. 28–38.

Godden, S., Lago, A., Bey, R., Leslie, K., Ruegg, P., Dingwell, R., 2007. Use of on-farm culture systems in mastitis control programs. In: National Mastitis Council Regional Meeting Proceedings. Visalia, California, pp. 1–9.

Halasa, T., Huijps, K., Østerås, O., Hogeveen, H., 2007. Economic effects of bovine mastitis and mastitis management: A review. Vet. Q. 29, 8–31.

Harmon, R.J., 1994. Physiology of mastitis and factors affecting somatic cell counts. J. Dairy Sci. 77, 2103–2112.

Harmon, R.J., Eberhart, R.J., Jasper, D.E., Langlois, B.E., Wilson, R.A., 1990. Microbiological Procedures for the Diagnosis of Bovine Udder Infection. National Mastitis Council Inc., Arlington, VA.

Heald, C.W., Kim, T., Sischo, W.M., Cooper, J.B., Wolfgang, D.R., 2000. A computerized mastitis decision aid using farm-based records: an artificial neural network approach. J. Dairy Sci. 83, 711–720.

Hillerton, J.E., Kliem, K.E., 2002. Effective treatment of *Streptococcus uberis* clinical mastitis to minimize the use of antibiotics. J. Dairy Sci. 85, 1009–1014.

Hogeveen, H., Huijps, K., Halasa, T., Lam, T.J.G.M., 2010. Mastitis costs money: what's new? In: Hillerton, J.E. (Ed.), Mastitis Research into Practice: Proceedings of the 5th IDF Mastitis Conference, VetLearn, New Zealand, pp. 62–71.

Hogeveen, H., Noordhuizen-Stassen, E.N., Thysen, I., Van Werven, T., Lam, T.J.G.M., 1995. Automated pathogen diagnosis at low somatic cell count farms. In: Saran, A., Soback, S. (Eds.), proceedings of the Third International Mastitis Seminar, vol. I. Tel Aviv, Israel, pp. s2.47–s2.51.

Hogeveen, H., Ouweltjes, W., 2003. Sensors and management support in high-technology milking. J. Anim. Sci. 81, 1–10.

Author's personal copy

Huijps, K., Lam, T.J.G.M., Hogeveen, H., 2008. Costs of mastitis: facts and perception. J. Dairy. Res. 75, 113–120.
Jones, G.F., Ward, G.E., 1990. Evaluation of a scheme for predicting the Gram-staining reaction of organisms causal bovine mastitis. J. Am. Vet. Med. Assoc. 196, 597–599.
Kamphuis, C., Mollenhorst, H., Feelders, A.J., Pietersma, D., Hogeveen, H., 2010a. Decision-tree induction to detect clinical mastitis with automatic milking. Comput. Electron. Agric. 70, 60–68.
Kamphuis, C., Mollenhorst, H., Heesterbeek, J.A.P., Hogeveen, H., 2010b. Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision-tree induction. J. Dairy Sci. 93, 3616–3627.
Kamphuis, C., Pietersma, D., Van der Tol, R., Wiedemann, M., Hogeveen, H., 2008a. Using sensor data patterns from an automatic milking system to develop predictive variables for classifying clinical mastitis and abnormal milk. Comput. Electron. Agric. 62, 169–181.
Kamphuis, C., Sherlock, R., Jago, J., Mein, G., Hogeveen, H., 2008b. Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. J. Dairy Sci. 91, 4560–4570.
Kim, T., Heald, C.W., 1999. Inducing inference rules for the classification of bovine mastitis. Comput. Electron. Agric. 23, 27–42.
Maatje, K., Huijsmans, P.J.M., Rossing, W., Hogewerf, P.H., 1992. The efficacy of in-line measurement of quarter milk electrical-conductivity, milk-yield and milk temperature for the detection of clinical and subclinical mastitis. Livest. Prod. Sci. 30, 239–249.
MacDonald, K., Poole, D., Muckle, A., 2010. Preliminary assessment of the accuracy of on-farm diagnosis of clinical mastitis using 3M Petrifilms compared to standard microbiology. In: Hillerton, J.E. (Ed.), Mastitis Research into Practice: Proceedings of the 5th IDF Mastitis Conference, VetLearn, Wellington, pp. 481–482.
McDougall, S., Agnew, K.E., Cursons, R., Hou, X.X., Compton, C.R.W., 2007. Parenteral treatment of clinical mastitis with tylosin base or penethamate hydriodide in dairy cattle. J. Dairy Sci. 90, 779–789.
Milne, M.H., Biggs, A.M., Fitzpatrick, J.L., Innocent, G.T., Barrett, D.C., 2003. Use of clinical information to predict the characteristics of bacteria isolated from clinical cases of bovine mastitis. Vet. Rec. 152, 615–617.
Mollenhorst, H., Van der Tol, P.P.J., Hogeveen, H., 2010. Somatic cell count assessment at quarter or cow milking level. J. Dairy Sci. 93, 3358–3364.
Polat, B., Colak, A., Cengiz., M., Yanmaz, L.E., Oral, H., Bastan, A., Kaya, S., Hayirli, A., 2010. Sensitivity and specificity of infrared thermography in detection of subclinical mastitis in dairy cows. J. Dairy Sci. 93, 3525–3532.
Pyörälä, S.H.K., Pyörälä, E.O., 1998. Efficacy of parenteral administration of three antimicrobial agents in treatment of clinical mastitis in lactating cows: 487 cases (1989–1995). J. Am. Vet. Med. Assoc. 212, 407–412.
Rasmussen, M.D., 2005. Visual scoring of clots in foremilk. J. Dairy Res. 72, 406–414.
Song, X., Zhuang, S., Van der Tol, P. P. J., 2010. New model to detect clinical mastitis in Astronaut A3 Next™ milking robot. In: Hillerton, J.E. (Ed.), Mastitis Research into Practice: Proceedings of the 5th IDF Mastitis Conference, VetLearn, New Zealand, pp. 474–480.
Steeneveld, W., Van der Gaag, L.C., Barkema, H.W., Hogeveen, H.W., 2009. Providing probability distributions for the causal pathogen of clinical mastitis using naive Bayesian networks. J. Dairy Sci. 92, 2598–2609.
Van der Gaag, L.C., Renooij, S., Steeneveld, W., Hogeveen, H., 2009. When in doubt. Be indecisive. In: Sossai, C., Chemello, G. (Eds.), Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty. Springer Verlag, Berlin/Heidelberg, pp. 518–529.
Viguier, C., Arora, S., Gilmartin, N., Welbeck, K., O'Kennedy, R., 2009. Mastitis detection: current trends and future perspectives. Trends Biotechnol. 27 (8), 486–493.
White, M.E., Glickman, L.T., Barnes-Pallesen, F.G., Stem III, E.S., Dinsmore, P., Powers, M.S., Powers, P., Smith, M.C., Jasko, D., 1986. Accuracy of clinicians in predicting the bacterial cause of clinical bovine mastitis. Can. Vet. J. 27, 218–220.
Witten, I.H., Frank, E., 2005. Data Mining; Practical Machine Learning Tools and Techniques, vol. 2. Morgan Kaufmann Publishers, San Fransi.