J. Dairy Sci. 93:2559–2568
doi:10.3168/jds.2009-3020

Discriminating between true-positive and false-positive clinical mastitis alerts from automatic milking systems

W. Steeneveld,*[1] L. C. van der Gaag,† W. Ouweltjes,‡ H. Mollenhorst,* and H. Hogeveen*§
*Department of Farm Animal Health, Faculty of Veterinary Medicine, Utrecht University, PO Box 80151, 3508 TD Utrecht, the Netherlands
†Department of Information and Computing Sciences, Utrecht University, PO Box 80089, 3508 TD Utrecht, the Netherlands
‡Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad, the Netherlands
§Chair group Business Economics, Wageningen University, PO Box 8130, 6706 KN Wageningen, the Netherlands

ABSTRACT

Automatic milking systems (AMS) generate alert lists reporting cows likely to have clinical mastitis (CM). Dutch farmers indicated that they use non-AMS cow information or the detailed alert information from the AMS to decide whether to check an alerted cow for CM. However, it is not yet known to what extent such information can be used to discriminate between true-positive and false-positive alerts. The overall objective was to investigate whether selection of the alerted cows that need further investigation for CM can be made. For this purpose, non-AMS cow information and detailed alert information were used. During a 2-yr study period, 11,156 alerts for CM, including 159 true-positive alerts, were collected at one farm in the Netherlands. Non-AMS cow information on parity, days in milk, season of the year, somatic cell count history, and CM history was added to each alert. In addition, 6 alert information variables were defined. These were the height of electrical conductivity, the alert origin (electrical conductivity, color, or both), whether or not a color alert for mastitic milk was given, whether or not a color alert for abnormal milk was given, deviation from the expected milk yield, and the number of alerts of the cow in the preceding 12 to 96 h. Subsequently, naive Bayesian networks (NBN) were constructed to compute the posterior probability of an alert being truly positive based only on non-AMS cow information, based on only alert information, or based on both types of information. The NBN including both types of information had the highest area under the receiver operating characteristic curve (AUC; 0.78), followed by the NBN including only alert information (AUC = 0.75) and the NBN including only non-AMS cow information (AUC = 0.62). By combining the 2 types of information and by setting a threshold on the computed probabilities, the number of false-positive alerts on a mastitis alert list was reduced by 35%, and 10% of the true-positive alerts would not be identified. To detect CM cases at a farm with an AMS, checking all alerts is still the best option but would result in a high workload. Checking alerts based on a single alert information variable would result in missing too many true-positive cases. Using a combination of alert information variables, however, is the best way to select cows that need further investigation. The effect of adding non-AMS cow information on making a distinction between true-positive and false-positive alerts would be minor.
Key words: clinical mastitis, detection, automatic milking, dairy cow

INTRODUCTION

Mastitis is one of the most frequent and costly diseases in dairy cows (e.g., Halasa et al., 2007). Detection of clinical mastitis (**CM**) is important to maintain a high standard of milk quality. With conventional milking systems, detection of CM is based on visual inspection of the milk during milking, whereas farmers with an automatic milking system (**AMS**) have to rely on mastitis alert lists from the AMS for information on the udder health status of their cows (Hogeveen and Ouweltjes, 2003). These alert lists are generated from sensor measurements during milking [e.g., electrical conductivity (**EC**) and color measurements] and report the cows suspected of having CM. A general complaint of dairy farmers working with an AMS is the relatively large number of false-positive alerts on the mastitis alert lists. Several detection models were developed with the aim of reducing the number of false-positive alerts (e.g., de Mol and Ouweltjes, 2001; Cavero et al., 2008; Kamphuis et al., 2010). The sensitivity (**Se**) and specificity (**Sp**) of these models, however, remain too low to substantially reduce the number of false-positive alerts and at the same time retain a sufficient detection of true cases.

Deciding on which alerts have the highest priority to be visually checked is difficult. In essence, all mastitis

Received December 21, 2009.
Accepted March 9, 2010.
[1]Corresponding author: w.steeneveld@uu.nl

alerts given by an AMS have to be visually checked. Because of lack of time and the large number of fruitless visual checks, however, farmers do not check all alerts in practice (Claycomb et al., 2009; Neijenhuis et al., 2009). Results of a recent large survey on Dutch dairy farms using AMS showed that most farmers (65%) do not use any explicit rules for deciding upon which cows to check visually for CM (Neijenhuis et al., 2009). These farmers thus make their inspection decisions based on intuition. Furthermore, 12% of the farmers used cow information, such as a cow's SCC and CM history, and 23% of the farmers used alert information, such as the absolute EC value, to decide whether to check a particular cow (Neijenhuis et al., 2009). To what extent non-AMS cow information, alert information, or both can be used to discriminate between the true-positive and false-positive alerts of a mastitis alert list is unknown.

In 2 recent Danish studies, non-AMS cow information in addition to sensor measurements was taken into account for the detection of CM (Chagunda et al., 2006; Friggens et al., 2007). Steeneveld et al. (2010) presented a method in which a prior probability of CM (based on parity, DIM, season, SCC history and CM history) was combined with the test characteristics (Se and Sp) of the AMS detection system to discriminate between alerts in their likelihood of being a true-positive alert. In these studies, however, the additional value of non-AMS cow information to discriminate between true-positive alerts and false-positive alerts was not investigated. It has not yet been investigated whether the detailed alert information itself can be used to discriminate between true-positive and false-positive alerts.

The overall objective of the current study was to investigate whether alerted cows that need further investigation for CM could be selected. For this purpose, Bayesian networks were used; Bayesian networks are readily constructed and allow easy computation of posterior probabilities (Jensen, 2001). For our study, several Bayesian networks were constructed to compute the posterior probability of an alert being truly positive based on non-AMS cow information only, based on alert information only, and based on both types of information.

MATERIALS AND METHODS

Data Collection and Herd Description

From October 1, 2007, to October 1, 2009, data were collected at the Dutch research farm Waiboerhoeve (Lelystad, the Netherlands). During this period, the average herd size was approximately 500 cows. From the herd, 250 cows were milked with 4 AMS (Lely Industries N.V., Maassluis, the Netherlands). The other cows were milked in another barn with a conventional milking system. Because of different experiments on the farm, cows frequently changed between the 2 barns. All cows were housed in a free-stall barn and were of the Holstein Friesian breed. The average 305-d milk yield was 9,500 kg/cow, and the average bulk milk SCC during the study period was 185,000 cells/mL.

In the 2-yr study period, data were collected from all milkings of all cows milked with an AMS. These data included quarter-based EC and milk color reflection measurements (red, green, and blue; Lely Industries N.V.). Information on milk yield was available at the cow level. The EC measurements represented the mean value of the 20 highest measurements of a quarter milking and were available as index values. The color measurements represented the mean value of the milking. In addition to these sensor measurements, for each milking, information was collected about whether the quarter milking was alerted for mastitis based on EC, color measurement, or both. An EC alert was given for a quarter if the EC of the milk was 20% higher than that from the lowest quarter (interquarter ratio of 1.20) in 2 consecutive milkings. In addition, the interquarter ratio of the running average over the last 3 milkings had to exceed 1.20. Based on milk color measurements by the color sensor (Espada and Vijverberg, 2002), a quarter could receive the following color alerts: abnormal milk, mastitic milk, colostrum, or milk with blood. No color alerts for colostrum and milk with blood were given during the study period; therefore, only the alerts for abnormal milk and mastitic milk were used in the current study.

For all cows at the farm, information was collected about the occurrence of CM within the study period and in the preceding year. For the cows milked with the AMS, herd employees were instructed to check alerts twice a day. Cows appearing on the alert list were eligible for inspection. There was, however, no explicit protocol to guide employees in their decisions of which alerted cows to check visually. The employees used triggers other than the mastitis alert list to check cows for CM. These included clots on the filter sock and long milking intervals. These other triggers occasionally resulted in the detection of CM in cows for which no alert was given by the AMS detection system. Eight employees were involved in visually checking alerts. If CM was confirmed by obvious signs of wateriness or clots in any quarter or by other signs of CM such as swelling or redness, the infected quarter was treated with antibiotics. Immediately after detection of CM, the infected cow was removed from the AMS barn to the sick cow pen and was milked twice daily with the conventional milking system for an initially unknown period.

For all cows at the farm, the Dutch national milk production recording system (CRV, Arnhem, the Netherlands) provided information from the four-weekly milk recordings, which included date of milk recording, test-day milk yields (kg of milk, fat, and protein), and SCC measurements (cells/mL). Data from the milk production recording system were available from the study period and from the preceding year.

Data Preparation

During the study period, information from 511,744 milkings of 602 cows was collected. Milkings without any recorded milk yield (n = 2,842) were excluded from the data. In total, 227 CM cases were recorded for 148 different cows milked with the AMS.

For each milking, it was determined if any quarter of the cow had an alert based on EC, color, or both. For 11,314 milkings, at least one quarter milking received an alert based on EC or color measurements (abnormal milk or mastitic milk). The average number of alerts per day was 15, with a minimum of 1 and a maximum of 48. For each of these alerts, it was determined whether it was a true-positive, a false-positive, or an inconclusive alert for CM. If a cow had a single alert on the day on which CM was recorded, then that alert was assigned to be a true-positive alert. If a cow had multiple alerts on the day on which CM was recorded, then only the last alert of that day was assigned to be a true-positive alert. Because not all alerts are visually checked, it was not possible to know whether the other alerts during the 24 h preceding the true-positive alert were true-positive or false-positive alerts. Therefore, these alerts were considered inconclusive and excluded from the final data set (n = 110). Some cows with CM were detected in the early morning without having been milked yet on that day. If such a cow had an alert on the previous day, the last alert of that previous day was assigned to be a true-positive alert and the remaining alerts of that previous day were considered inconclusive and were again excluded from the final data set (n = 48). All other alerts from the data set were defined to be false-positive alerts.

In the resulting data set, information from 508,744 milkings was available. A total of 11,156 milkings were alerted for CM, of which 159 were true-positive alerts and 10,997 were false-positive alerts. For 68 CM cases, no alert was given by the AMS. The CM detection performance of the AMS can thus be summarized as having Se of 70%, Sp of 97.8%, and a predictive positive value of 1.4% (Table 1). Because not all milkings were checked, it must be noted that the Se and Sp values were not the exact values.

Table 1. Clinical mastitis detection performance of the automatic milking system (AMS)

	Clinical mastitis		
AMS alert	Yes	No	Total
Yes	159	10,997	11,156
No	68	497,520	497,588
Total	227	508,517	508,744

Six variables were defined to describe the detailed alert information. Only variables that were readily available and directly usable from the alert lists currently presented to dairy farmers using AMS were defined. First, for each alert, the highest EC of the 4 quarters was determined and classified into the intervals ≤80, 81–90, 90–100, and >100. Second, for each alert it was determined whether it originated from an increased EC, from a deviated color measurement, or from both. If, for instance, one quarter of a cow was alerted based on EC and another quarter was alerted based on color, then that alert was defined as being based on both EC and color. Third, for each alert, it was determined whether a color alert for mastitic milk was given. The fourth variable describes whether a color alert for abnormal milk was given. Fifth, for each alert, the deviation from the expected milk yield was determined. To determine the expected milk yield, the milk yield per hour was calculated by dividing the milk yield during a milking by the time since the last milking of a cow. The deviation from the expected milk yield was determined by comparing the milk yield per hour of the current milking with the mean milk yield per hour of the last 5 d. The deviation from the expected milk yield was classified into the intervals >40%, 30–40, 20–30, 10–20, or <10% less milk than expected. Finally, for each alert, the number of alerts (0, 1, 2, 3, ≥4) for that cow in the preceding 12 to 96 h was determined. A time-window starting with 12 h was chosen because herd employees were instructed to check alerts twice a day. Therefore, it was assumed that alerts in the 12 h before the current alert were checked at the same time as the current alert.

Non-AMS cow information was added to each alert in the data set, as suggested by Steeneveld et al. (2010). The parity of the cow, DIM, season of the year, SCC in the previous 30 d, SCC in the 30 d before the previous 30 d, and, for multiparous cows, the geometric mean SCC from all available test-day records from the previous lactation were added to each alert. In addition, the accumulated number of CM cases of the cow in the previous 30 d and the accumulated number of CM cases of the cow in the days before the previous 30 d were added to each alert.

Logistic regression was used to determine whether true-positive and false-positive alerts differed in their non-AMS cow information and in their alert information. Each variable was investigated individually using a random cow effect to account for the repeated measurements on cows. Data preparation and logistic regression were performed using SAS version 9.1 (SAS Institute Inc., Cary, NC).

Construction of a Naive Bayesian Network

Bayesian networks allow different graphical structures of varying complexity. Naive Bayesian networks (**NBN**) are the simplest type of Bayesian network. In veterinary research, they were described, for example, by Steeneveld et al. (2009). An NBN consists of a single output variable that represents the possible classes for the dependent variable of a study and a set of feature variables modeling the levels of the study's independent variables. An NBN further includes arrows from the output variable to each feature variable to describe the dependence of the latter on the output variable (Friedman et al., 1997). In the current study, the variable capturing whether a CM alert was a true-positive or a false-positive alert was the output variable. The variables describing detailed alert information and the variables capturing non-AMS cow information were feature variables.

Naive Bayesian networks are typically constructed from data that consists of determining prior probabilities for the output variable and of estimating conditional probabilities for the feature variables given the possible classes of the dependent variable. In this study, the prior probabilities for the output variable reflect the probability of an alert being a true-positive alert. Like these prior probabilities, the conditional probabilities for the feature variables are based on frequency counts in the data. For instance, the conditional probability of EC >100 given that the alert is truly positive; that is, the probability Pr (EC >100 | true-positive alert), was computed as the proportion of alerts with an EC >100 among the true-positive alerts.

In essence, all available feature variables can be included in an NBN. Methods exist, however, for selecting only those feature variables that best discriminate between the different classes of a dependent variable, thereby forestalling overfitting of the data (Langley and Sage, 1994). In the present study, wrapper-based backward elimination was used for selecting appropriate feature variables from all defined variables (Kohavi and John, 1997). With this method, feature variables were selected to optimize the area under the receiver operating characteristic (**ROC**) curve (**AUC**) (these will be described in more detail in the next section) of the NBN under construction. The method started with an NBN including the output variable and all feature variables. In each subsequent step, a single feature variable was removed. The variable chosen to this end was that which served to improve the AUC of the NBN the most, if any. The removal of feature variables was continued until the AUC of the NBN no longer improved.

To determine which information (non-AMS cow information, alert information, or both) was the most valuable for discriminating between true-positive and false-positive alerts, different NBN were constructed. These were an NBN with only feature variables with non-AMS cow information, an NBN with only feature variables with alert information, and an NBN with feature variables capturing both types of information.

For the purpose of constructing and subsequently validating the NBN, the available data set was split into a construction set and a validation set. From the data set, the alerts of two-thirds of the cows were selected randomly for construction; the alerts of the remaining cows were included in the validation data set. Constructing the different NBN, which included performing backward elimination and estimating prior and conditional probabilities, was done by using the Bayesian-network editing package Dazzle (Schrage et al., 2005).

Validation

For each alert from the validation data set, the constructed NBN was used to calculate the posterior probability of the alert being truly positive, given the available non-AMS cow information or the alert information. For computing posterior probabilities, an NBN builds upon Bayes' rule together with the assumption that all feature variables are mutually independent given the output variable (Friedman et al., 1997). More specifically, for computing the posterior probability Pr $(c_1 \mid f_1, \ldots, f_n)$ of the output c_1 given levels $f_1,\ldots,f_n$ for its n feature variables, the model uses

$$\Pr(c_1 \mid f_1,\ldots,f_n) = \frac{\prod_{i=1}^{n} \Pr(f_i \mid c_1) \times \Pr(c_1)}{\sum_{j=1}^{2} \prod_{i=1}^{n} \Pr(f_i \mid c_j) \times \Pr(c_j)}, \quad [1]$$

where $\Pr(c_1)$ is the overall prior probability of an alert being truly positive and $\Pr(c_2)$ is the prior probability of an alert being a false-positive alert. The probabilities $\Pr(f_i \mid c_1)$ are the conditional probabilities of finding the level f_i for the ith selected feature variable given that the alert is a true-positive alert. Note that the prior probabilities $\Pr(c_j)$ and the conditional probabilities $\Pr(f_i \mid c_j)$ for all levels have already been estimated from the data upon constructing the NBN and are therefore readily available in the NBN for the computation of the posterior probabilities using formula [1]. Computing the posterior probabilities was done using Dazzle (Schrage et al., 2005).

To evaluate the performance of the different constructed NBN, the posterior probabilities obtained for the validation data set were used to calculate the Se and Sp of the NBN over the whole range of possible threshold probabilities for classification. Receiver operating characteristic curves were constructed to visualize the performance. A ROC curve is a graphic representation of the Se versus 1 − Sp over the whole range of classification thresholds. To summarize the ROC curves into a single quantity, the AUC was computed (e.g., Detilleux et al., 1999; Dohoo et al., 2003). In the current study, the AUC can be interpreted as the probability that a randomly selected true-positive alert has a higher posterior probability to be truly positive than a randomly selected false-positive alert. The ROC curves were visualized using TIBCO S$^+$ version 8.1 (TIBCO Software Inc., Palo Alto, CA) and the AUC using the trapezoidal rule were calculated in SAS (%roc macro: http://support.sas.com/kb/25/017.html).

Subsequently, using the validation data set with the computed posterior probabilities, different threshold values on these probabilities were set such that 95, 90, and 80% of all true-positive alerts ended up being alerted. The effect of setting these thresholds on the number of false-positive alerts was investigated.

RESULTS

The descriptive statistics of the true-positive alerts and false-positive alerts are given in Tables 2 and 3. Most alerts were given for older cows and for cows later in lactation. From the non-AMS cow information, only the distribution of DIM was found to be significantly different between true-positive and false-positive alerts ($P = 0.002$). The SCC information captured in the 3 SCC variables did not significantly differ between true-positive and false-positive alerts. For instance, although 30% of the true-positive alerts were from cows with an SCC >500,000 cells/mL in the last 30 d, the same information was found in 24% of the false-positive alerts ($P = 0.135$; Table 2).

The distributions of 5 of the 6 variables capturing the detailed alert information were significantly ($P <$ 0.05) different between true-positive and false-positive alerts. The single exception was the variable whether a color alert for abnormal milk was given (Table 3). For instance, although 45% of the true-positive alerts had an EC value >100, this information was found in just 20% of the false-positive alerts. Although 40% of the true-positive alerts had a decreased milk production of more than 30%, just 9% of the false-positive alerts had such a large decrease in milk production compared with expected milk yield.

Upon constructing the 3 NBN, backward elimination resulted in the removal of feature variables. From the NBN with only non-AMS cow information, only the variable capturing the season of the year was removed. From the NBN containing alert information only, the variable capturing the number of alerts for the cow in the preceding 12 to 96 h and the variable describing whether a color alert for abnormal milk was given were removed. From the NBN containing both types of information, 2 variables were removed by backward elimination. These were the variable capturing the mean SCC in the previous lactation and the variable whether a color alert for abnormal milk was given.

The ROC curves of the 3 constructed NBN are presented in Figure 1. The AUC of the 3 NBN are reported in Table 4. The combination of non-AMS cow information and detailed alert information resulted in the highest AUC. The difference in AUC between the NBN with alert information only (AUC = 0.7499) and the NBN containing both information sources (AUC = 0.7792) was not significant ($P = 0.210$), but both differed significantly from the AUC of the NBN containing cow information only (AUC = 0.6175, $P = 0.014$ and <0.0001 respectively).

The validation data set contained 52 true-positive alerts and 3,636 false-positive alerts. Table 5 summarizes the classification results from the 3 NBN for these alerts, by using threshold values for which at least 95, 90, and 80% of all true-positive alerts were identified, respectively. A threshold value on the posterior probability computed from the NBN containing both cow and alert information, for example, for which at least 95% of all true-positive alerts were identified, resulted in 2,527 false-positive alerts. Compared with the original 3,636 false-positive alerts, this was a reduction of 31%. A threshold value set such that at least 95% of all true-positive alerts were identified but this time using the NBN with non-AMS cow information only resulted in 3,493 false-positive alerts—a reduction of 4%. A threshold value on the posterior probability computed from the NBN containing both types of information for which at least 80% of all true-positive alerts were identified resulted in 1,568 false-positive alerts—a reduction of 57% compared with the original 3,636 false-positive alerts.

DISCUSSION

The test characteristics of the AMS in the current study were an Se of 70% and an Sp of 97.8%. Because an

Table 2. Number of true-positive (tp) and false-positive (fp) alerts for different levels of cow information

Cow information	tp alerts, n (%) (n = 159)	fp alerts, n (%) (n = 10,997)	P-value[1]
Parity			0.121
1	8 (5)	941 (9)	
2	58 (37)	3,541 (32)	
3	45 (28)	3,325 (30)	
≥4	48 (30)	3,190 (29)	
DIM			0.002
1–30	9 (6)	273 (3)	
31–60	18 (11)	477 (4)	
61–90	18 (11)	726 (7)	
91–120	13 (8)	664 (6)	
121–150	12 (8)	733 (7)	
151–180	12 (8)	883 (8)	
181–210	24 (15)	822 (7)	
≥211	53 (33)	6,419 (58)	
Season			0.518
January–March	48 (30)	2,818 (26)	
April–June	31 (20)	2,374 (21)	
July–September	37 (23)	2,650 (24)	
October–December	43 (27)	3,155 (29)	
SCC in last 30 d			0.135
<500,000 cells/mL	79 (70)	6,911 (76)	
≥500,000 cells/mL	34 (30)	2,183 (24)	
SCC before last 30 d			0.864
<500,000 cells/mL	94 (73)	7,049 (77)	
≥500,000 cells/mL	35 (27)	2,091 (23)	
Mean SCC previous lactation			0.501
<500,000 cells/mL	132 (89)	9,045 (93)	
≥500,000 cells/mL	17 (11)	708 (7)	
CM[2] cases in last 30 d			0.054
0	148 (93)	10,271 (93)	
1	10 (6)	654 (6)	
2	1 (1)	72 (1)	
CM cases before last 30 d			0.158
0	84 (53)	5,026 (46)	
1	39 (24)	3,413 (31)	
2	36 (23)	2,558 (23)	

[1]Indicates whether the distribution over the levels for a cow information variable is different between true-positive and false-positive alerts.

[2]CM = clinical mastitis.

Se of at least 70% combined with an Sp of at least 99% is desired (Mein and Rasmussen, 2008), the CM detection performance of this AMS is suboptimal. Although new detection models were developed in several previous studies (e.g., de Mol and Ouweltjes, 2001; Cavero et al., 2008; Kamphuis et al., 2010), these models were unable to improve detection performance to the extent that the detection of CM cases remained satisfactory and the number of false-positive alerts was reduced to a reasonable level. Because of the suboptimal detection performance, and particularly the unsatisfactory Sp, interpretation of the alert lists is difficult. For instance, the current alert lists contain a large number of false-positive alerts and it is not possible to select those cows that have priority for visual checking. In our study, no new detection model was developed. The aim was to investigate whether cows needing further investigation for CM can be selected from the alert lists based on non-AMS cow information, on alert information, or on both.

For farmers milking with an AMS, handling the mastitis alert lists generated by the AMS is daily practice. Most Dutch farmers milking with an AMS make their inspection decisions based on intuition. Only a minority of farmers indicated that they use non-AMS cow information or detailed alert information to decide which cows to check visually (Neijenhuis et al., 2009). Results of our study show that checking alerts based on a single alert variable is not satisfactory. For instance, visual checking of just the alerts with an EC over 100 would result in missing 55% of the true-positive alerts; and checking of just the newly alerted cows without any alerts in the preceding 12 to 96 h would result in missing 50% of the true-positive alerts (Table 3). The results of our study show that a combination of variables capturing alert information is necessary to make a meaningful

Table 3. Number of true-positive (tp) and false-positive (fp) alerts for different levels of alert information

Alert information	tp alerts, n (%) (n = 159)	fp alerts, n (%) (n = 10,997)	*P*-value[1]
Electrical conductivity[2]			<0.0001
≤80	17 (11)	3,586 (33)	
81–90	34 (21)	2,270 (21)	
90–100	36 (23)	2,893 (26)	
>100	72 (45)	2,248 (20)	
Alert origin			<0.0001
Electrical conductivity	25 (16)	4,618 (42)	
Color	73 (46)	5,548 (50)	
Electrical conductivity + color	61 (38)	831 (8)	
Color alert: mastitic milk			<0.0001
Yes	75 (47)	2,346 (21)	
No	84 (53)	8,651 (79)	
Color alert: abnormal milk			0.685
Yes	63 (40)	4,256 (39)	
No	96 (60)	6,741 (61)	
Decreased milk production			<0.0001
>40%	46 (29)	609 (6)	
30–40%	17 (11)	303 (3)	
20–30%	22 (14)	702 (6)	
10–20%	24 (15)	1,770 (16)	
<10%	50 (31)	7,613 (69)	
Alerts of cow in preceding 12–96 h			0.020
0	80 (50)	3,856 (35)	
1	26 (17)	1,790 (16)	
2	19 (12)	1,291 (12)	
3	16 (10)	980 (9)	
≥4	18 (11)	3,080 (28)	

[1]Indicates whether the distribution over the levels for an alert information variable differs between true-positive and false-positive alerts.

[2]Reported in standardized units of the automatic milking system.

selection of alerts for visual checking. In fact, using a combination of 4 alert variables (height of EC, alert origin, color alert for mastitic milk, and expected milk production) proved to be very useful in discriminating between the true-positive and false-positive alerts from a mastitis alert list. This combination resulted in a reduction of the number of false-positive alerts by 33% at the cost of missing or detecting later 10% of the true-positive alerts (Table 5). Combining both non-AMS cow information and detailed alert information served to reduce further the number of false-positive alerts (Table 5).

Using only non-AMS cow information to discriminate between alerts was not very useful (Tables 4 and 5). The idea of using non-AMS cow information to discriminate between alerts originated from the observation that cows having CM differ in several aspects from cows not having CM (e.g., Barkema et al., 1998; Steeneveld et al., 2008, 2010). Based on that finding, we expected better performance from the NBN with only non-AMS cow information than was actually found in the present study. The true-positive alerts indeed differed in their cow information from the true-negative milkings. In particular, the SCC history was found to be significantly different between true-positive and true-negative milkings (data not shown). Within the group of alerted cows, however, the SCC histories of the true CM cases and of the false alerts proved to be no longer significantly different (Table 3). Most likely, the AMS had given many alerts for cows with subclinical mastitis. These alerts were not given without reason, but these cows were not (yet) clinically diseased. The lack of significant differences in non-AMS cow information between the true-positive and false-positive CM alerts explains the relatively low AUC of the constructed NBN based on non-AMS cow information only.

Upon constructing the 3 NBN, backward elimination resulted in the removal of some variables from the NBN. For instance, from the NBN containing only alert information, the variable capturing the number of alerts for the cow in the preceding 12 to 96 h and the variable describing whether a color alert for abnormal milk was given were removed. Because these variables were the least significant in the univariate analysis (Table 3), it was not surprising that they were removed. The presence of all other feature variables in the NBN indicates that each of these variables, in the presence of the previously selected variables, served to contribute

Table 4. Area under the receiver operating characteristic curve (AUC) with associated standard error and confidence interval of naive Bayesian networks including different types of information

Type of information	AUC	SE	95% CI
Cow information	0.6175	0.0387	0.5416–0.6934
Alert information	0.7499	0.0379	0.6756–0.8243
Cow + alert information	0.7792	0.0342	0.7121–0.8463

to the model's discriminative performance. It further indicates that the selected variables do not exhibit a large overlap in the information they contribute.

The method for discriminating between true-positive and false-positive alerts as presented in the current paper is based on the use of a relatively simple Bayesian network. Bayesian networks have been studied extensively and are being widely applied in human medicine (e.g., Chapman et al., 2005). Applications are not yet common in veterinary science, but they are gaining popularity (e.g., Otto and Kristensen, 2004; Jensen et al., 2009; Steeneveld et al., 2009). The NBN used in this study constitute the simplest type of Bayesian network. Despite their simplicity, these models are surprisingly effective, showing good classification performance even if the independence assumption for the feature variables does not hold in the data (Friedman et al., 1997). In the current study, Bayesian networks of increasing complexity were developed as well. More specifically, tree-augmented NBN, which include dependencies between the feature variables (Friedman et al., 1997), were constructed and validated. The more sophisticated dependency structures, however, did not result in significantly higher AUC than the ones obtained with the simple NBN (data not shown).

The posterior probabilities computed from the NBN in this study can be combined with the current detection algorithms in several different ways. For instance, it is possible to add a posterior probability to each of the alerts currently given by the AMS. The farmer is then provided with additional information to decide which alerts have the highest priority for visual checking. Another possibility is to present only the alerts from the current alert lists that have a posterior probability of being truly positive above a particular threshold value. In this way, fewer alerts are presented to the farmer—specifically fewer false-positive alerts. It is unavoidable that for some of the true CM cases no alert will be given ; consequently, these CM cases will be missed or later detected by the farmer. The risk attitude of the farmer, with respect to missing CM cases versus checking large numbers of false-positive alerts, can be incorporated by allowing the farmer to adjust the threshold value. For instance, the threshold values can lowered if the bulk milk SCC is trending upwards (Claycomb et al., 2009). For actual implementation, software is required to combine all available information into a posterior probability and subsequently comparing this probability against a preset farmer-specific threshold value.

For 68 CM cases, no alert was given by the AMS (Table 1). Most of these cases were not alerted because of the use of a narrow time-window. Only the final alert on the day on which CM was recorded was considered as a true-positive alert. The definition of an appropriate time-window is difficult and still under debate (Sherlock et al., 2008). In our study, using a wider time-window resulted in more true-positive alerts and fewer CM cases without an alert. For instance, considering the alerts from the day before the day on which CM was recorded as true-positive alerts resulted in 28 CM cases instead of 68 CM cases without an alert. We decided, however, to use a narrow time-window because of the systematic way of working on the research farm. Moreover, Sherlock et al. (2008) argued that using narrow time-windows for CM detection is more realistic and better for application in practice. The use of triggers other than the alert list for visually checking cows

Table 5. The effect of 3 probability threshold values on the number of true-positive (tp) and false-positive (fp) alerts[1]

	Cow information		Alert information		Cow + alert information	
Value[2]	tp	fp	tp	fp	tp	fp
Original number	52	3,636	52	3,636	52	3,636
Threshold value A	50	3,493	50	2,723	50	2,527
	(4)	(4)	(4)	(25)	(4)	(31)
Threshold value B	47	2,939	47	2,439	47	2,363
	(10)	(19)	(10)	(33)	(10)	(35)
Threshold value C	42	2,186	42	1,684	42	1,568
	(19)	(40)	(19)	(54)	(19)	(57)

[1]The percentage reduction in the number of tp and fp alerts compared with the original numbers is given in parentheses.

[2]Threshold values A, B, and C are the highest values such that at least 95%, 90%, and 80%, respectively, of all true-positive alerts were identified.

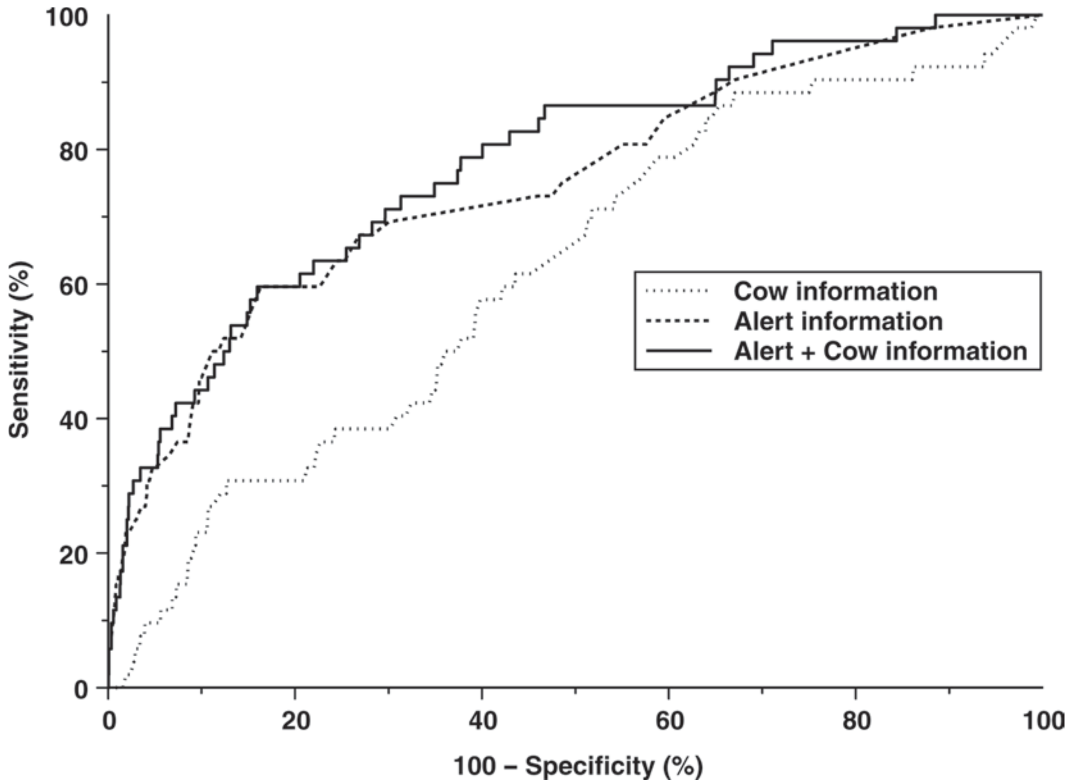

Figure 1. Receiver operating characteristic curves for the 3 developed naive Bayesian networks including different types of information: cow information, alert information, and cow + alert information.

for CM explains the detection of the nonalerted CM cases.

The herd employees indicated that approximately 40% of the alerts were visually checked. Because not all milkings and not all alerted cows were checked visually, it was not possible to calculate the exact values for Se and Sp. It is possible that nonalerted cows had CM and remained undetected. Therefore, in fact, some assigned true-negative milkings may have been false-negative. The herd employees did use triggers other than the alert list to detect CM, thereby minimizing the number of undetected CM cases. In addition, some true-positive alerts for CM may have been missed, and consequently, some alerts may have been incorrectly assigned as false-positive alerts. Although there were some missed CM cases, the effect of these missed cases on the conclusions would be minimal. Moreover, we believe that the systematic way of working on this research farm and the serious consideration of the alert lists have minimized the missing of true-positive alerts.

Because our research farm was used for several studies, the proportion of cows at risk for CM may have been somewhat different from that on commercial Dutch farms. For instance, heifers were more frequently housed in the other barn, cows did not enter the AMS barn in the colostrum-feeding period, and cows detected with CM were removed from the AMS barn. These observations explain the relatively small numbers of heifers and cows in the first weeks of lactation with alerts (Table 2). Most likely, the cows with chronic mastitis remained in the barn with the conventional milking system. In proportion, therefore, new CM cases occurred more frequently in cows milked with the AMS. Because of these specific characteristics of the research farm, it is not possible to generalize the developed model to other farms as it stands. For future implementation, it will be necessary to construct farm-specific models or a more generic model based on a variety of farms.

For the detection of CM cases on a farm milking with an AMS, checking all alerts is still the best option. Checking all alerts, however, is often not feasible because of the higher workload caused by the large number of false-positive alerts. Selecting alerts for visual checking based on a single alert information variable; for instance, only checking alerts with EC values >100, resulted in too many true-positive alerts being missed. To reduce both the workload and the annoyance of fruitless visual checks, selection of cows requiring further inspection is best based on a combination of alert information. The effect of using non-AMS cow information on making a distinction between true-positive and false-positive alerts proved to be minor in our study. Naive Bayesian networks can be used to combine the variables to select cows for visual checking.

ACKNOWLEDGMENTS

We gratefully acknowledge the employees of the Waiboerhoeve farm (Lelystad, the Netherlands) for their contribution to the acquisition of the data used in this research. Also greatly acknowledged is Lely Industries N.V. (Maassluis, the Netherlands) for their support with the installation of the data acquisition system at the farm. This research was supported by the Dutch Technology Foundation STW, applied science division of NWO, and the Technology Program of the Dutch Ministry of Economic Affairs (Utrecht, the Netherlands).

REFERENCES

Barkema, H. W., Y. H. Schukken, T. J. G. M. Lam, M. L. Beiboer, H. Wilmink, G. Benedictus, and A. Brand. 1998. Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. J. Dairy Sci. 81:411–419.

Cavero, D., K. H. Tolle, C. Henze, C. Buxade, and J. Krieter. 2008. Mastitis detection in dairy cows by application of neural networks. Livest. Sci. 114:280–286.

Chagunda, M. G. G., N. C. Friggens, M. D. Rasmussen, and T. Larsen. 2006. A model for detection of individual cow mastitis based on an indicator measured in milk. J. Dairy Sci. 89:2980–2998.

Chapman, W. W., J. N. Dowling, and M. M. Wagner. 2005. Classification of emergency department chief complaints into 7 syndromes: A retrospective analysis of 527,228 patients. Ann. Emerg. Med. 46:445–455.

Claycomb, R. W., P. T. Johnstone, G. A. Mein, and R. A. Sherlock. 2009. An automated in-line clinical mastitis detection system using measurement of conductivity from foremilk of individual udder quarters. N. Z. Vet. J. 57:208–214.

de Mol, R. M., and W. Ouweltjes. 2001. Detection model for mastitis in cows milked in an automatic milking system. Prev. Vet. Med. 49:71–82.

Detilleux, J., J. Arendt, F. Lomba, and P. Leroy. 1999. Methods for estimating areas under receiver-operating characteristic curves: Illustration with somatic-cell scores in subclinical intramammary infections. Prev. Vet. Med. 41:75–88.

Dohoo, I., W. Martin, and H. Stryhn. 2003. Veterinary Epidemiologic Research. AVC Inc., Charlottetown, Prince Edward Island, Canada.

Espada, E., and H. Vijverberg. 2002. Milk colour analysis as a tool for the detection of abnormal milk. Pages 28–38 in Proc. Conf. Robotic Milking, Toronto, Canada. Wageningen Pers, Wageningen, the Netherlands.

Friedman, N., D. Geiger, and M. Goldszmidt. 1997. Bayesian network classifiers. Mach. Learn. 29:131–163.

Friggens, N. C., M. G. G. Chagunda, M. Bjerring, C. Ridder, S. Hojsgaard, and T. Larsen. 2007. Estimating degree of mastitis from time-series measurements in milk: A test of a model based on lactate dehydrogenase measurements. J. Dairy Sci. 90:5415–5427.

Halasa, T., K. Huijps, O. Østeras, and H. Hogeveen. 2007. Economic effects of bovine mastitis and mastitis management: A review. Vet. Q. 29:18–31.

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

Jensen, F. V. 2001. Bayesian Networks and Decision Graphs. Springer-Verlag, New York, NY.

Jensen, T. B., A. R. Kristensen, N. Toft, N. P. Baadsgaard, S. Østergaard, and H. Houe. 2009. An object-oriented Bayesian network modeling the causes of leg disorders in finisher herds. Prev. Vet. Med. 89:237–248.

Kamphuis, C., H. Mollenhorst, A. Feelders, D. Pietersma, and H. Hogeveen. 2010. Decision-tree induction to detect clinical mastitis with automatic milking. Comput. Electron. Agric. 70:60–68.

Kohavi, R., and G. H. John. 1997. Wrappers for feature subset selection. Artificial Intelligence J. 97:273–324.

Langley, P., and S. Sage. 1994. Induction of selective Bayesian classifiers. Pages 199–406 in Proc. 10th Conf. Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers, San Francisco, CA.

Mein, G. A., and M. D. Rasmussen. 2008. Performance evaluation of systems for automated monitoring of udder health: Would the real gold standard please stand up? Pages 259–266 in Mastitis Control—From Science to Practice. T. J. G. M. Lam, ed. Wageningen Academic Publishers, Wageningen, the Netherlands.

Neijenhuis, F., J. Heinen, and H. Hogeveen. 2009. Automatic milking: Risk factors for the udder health status, report 257 (in Dutch). Wageningen UR Livestock Research, Lelystad, the Netherlands.

Otto, L., and C. S. Kristensen. 2004. A biological network describing infection with *Mycoplasma hyopneumoniae* in swine herds. Prev. Vet. Med. 66:141–161.

Schrage, M. M., A. Van IJzendoorn, and L. C. Van der Gaag. 2005. Haskell ready to Dazzle the real world. Pages 17–26 in Proc. 2005 ACM SIGPLAN workshop on Haskell. ACM Press, New York, NY.

Sherlock, R., H. Hogeveen, G. Mein, and M. D. Rasmussen. 2008. Performance evaluation of systems for automated monitoring of udder health: Analytical issues and guidelines. Pages 275–282 in Mastitis Control—From Science to Practice. T. J. G. M. Lam, ed. Wageningen Academic Publishers, Wageningen, the Netherlands.

Steeneveld, W., H. Hogeveen, H. W. Barkema, J. Van den Broek, and R. B. M. Huirne. 2008. The influence of cow factors on the incidence of clinical mastitis in dairy cows. J. Dairy Sci. 91:1391–1402.

Steeneveld, W., L. C. van der Gaag, H. W. Barkema, and H. Hogeveen. 2009. Providing probability distributions for the causal pathogen of clinical mastitis using naive Bayesian networks. J. Dairy Sci. 92:2598–2609.

Steeneveld, W., L. C. van der Gaag, H. W. Barkema, and H. Hogeveen. 2010. Simplify the interpretation of alert lists for clinical mastitis in automatic milking systems. Comput. Electron. Agric. 71:50–56.