Lameness detection based on multivariate continuous sensing of milk yield, rumination, and neck activity

T. Van Hertem,*† E. Maltz,* A. Antler,* C. E. B. Romanini,† S. Viazzi,† C. Bahr,† A. Schlageter-Tello,‡ C. Lokhorst,‡ D. Berckmans,† and I. Halachmi*1

*Institute of Agricultural Engineering, Agricultural Research Organization, Volcani Center, PO Box 6, Bet-Dagan IL-50250, Israel
†Division Measure, Model & Manage Bioresponses, KU Leuven, Kasteelpark Arenberg 30, bus 2456, BE-3001 Heverlee, Belgium
‡WageningenUR Livestock Research, PO Box 65, NL-8200AB Lelystad, the Netherlands

ABSTRACT

The objective of this study was to develop and validate a mathematical model to detect clinical lameness based on existing sensor data that relate to the behavior and performance of cows in a commercial dairy farm. Identification of lame (44) and not lame (74) cows in the database was done based on the farm’s daily herd health reports. All cows were equipped with a behavior sensor that measured neck activity and ruminating time. The cow’s performance was measured with a milk yield meter in the milking parlor. In total, 38 model input variables were constructed from the sensor data comprising absolute values, relative values, daily standard deviations, slope coefficients, daytime and nighttime periods, variables related to individual temperament, and milk session-related variables. A lame group, cows recognized and treated for lameness, to not lame group comparison of daily data was done. Correlations between the dichotomous output variable (lame or not lame) and the model input variables were made. The highest correlation coefficient was obtained for the milk yield variable (rMY = 0.45). In addition, a logistic regression model was developed based on the 7 highest correlated model input variables (the daily milk yield 4 d before diagnosis; the slope coefficient of the daily milk yield 4 d before diagnosis; the nighttime to daytime neck activity ratio 6 d before diagnosis; the milk yield week difference ratio 4 d before diagnosis; the milk yield week difference 4 d before diagnosis; the neck activity level during the daytime 7 d before diagnosis; the ruminating time during nighttime 6 d before diagnosis). After a 10-fold cross-validation, the model obtained a sensitivity of 0.89 and a specificity of 0.85, with a correct classification rate of 0.86 when based on the averaged 10-fold model coefficients. This study demonstrates that existing farm data initially used for other purposes, such as heat detection, can be exploited for the automated detection of clinically lame animals on a daily basis as well.

Key words: dairy cow, lameness, behavior, sensor data

INTRODUCTION

Lameness is defined as a deviation in gait resulting from pain or discomfort from hoof or leg injuries and disease (Flower and Weary, 2009). Lameness is a major health and welfare issue in modern intensive dairy farming (Cha et al., 2010; Lievaart and Noordhuizen, 2011; Bruijnis et al., 2012). Prevalence rates depend on many different factors, such as housing, management, feed, and breed (Cramer et al., 2009; Dippel et al., 2009; Barker et al., 2010). Cramer et al. (2009) showed that 99.3% of the herds had at least one cow with a lesion that caused lameness. The mean herd-prevalence was 28.1% (range = 0–83.8%) and cow-level prevalence was 25.7%. Barker et al. (2010) found a within-herd prevalence of 36.8% (range = 0–79.2%). Farmers frequently underestimate the lameness prevalence in their farm (Leach et al., 2010; Potterton et al., 2011; Sarova et al., 2011).

The most common method to obtain a herd lameness prevalence rate is visual locomotion scoring (Flower and Weary, 2009). This procedure is subjective (Channon et al., 2009; Flower and Weary, 2009), time-consuming (Thomsen, 2009), and costly. Therefore, visual locomotion scoring is hardly done for large herd sizes; rather, it is often done on a sample of the entire herd (Main et al., 2010). On Israeli dairy farms, lame cows are treated after clinical diagnosis by the veterinarian or claw trimmer to whom suspected cows are presented. This procedure is preceded by visual identification of lame cows by a trained herdsman, who usually observes a deviation in gait when walking the cows to the milking parlor. The cows identified as suspiciously lame are brought to a trained veterinarian or claw trimmer, who will make the actual clinical diagnosis of the lameness. If the suspected cow is confirmed as lame, adequate treatment is applied.

Received September 23, 2012.
Accepted April 4, 2013.
1Corresponding author: halachmi@volcani.agri.gov.il
Lameness has a negative effect on fertility and reproduction rates (Walker et al., 2010; Alawneh et al., 2011). The effect of lameness on animal behavior and performance is also well described. Live weight (Alawneh et al., 2012), milk fat (van Straten et al., 2011), milk yield (Green et al., 2002; Archer et al., 2010; Reader et al., 2011), and behavior such as activity (Blackie et al., 2011; Pavlenko et al., 2011; Reader et al., 2011) were all found to change by a lameness incidence or foot lesions. Continuous monitoring of these parameters could result in an automated lameness detection tool.

The automation of lameness detection or gait scoring is an important topic in research and several different approaches have been developed. Studies focusing on ground reaction forces (Pastell et al., 2008; Liu et al., 2011), pressure sensitive walkways (Maertens et al., 2011), and accelerometers (Pastell et al., 2009; Chapinal et al., 2011) measured the asymmetry of the gait when walking. These approaches used the diversity in weight bearing on each leg for lameness detection. Other studies used computer vision to analyze the gait automatically. These studies focused on gait parameters such as back arch curvature (Poursaberi et al., 2010), step overlap (Pluk et al., 2010), and hoof release angle (Pluk et al., 2012). Each of those studies introduced an additional tool or sensor in the farm to detect lameness automatically.

The use of sensors is widespread in modern dairy farming, which makes continuous monitoring of the individual in the herd possible. However, the large pool of existing data on the farm is not extensively used. Estrus detection based on animal behavior is a well-described research item (de Mol et al., 2001; Holman et al., 2011; Jonsson et al., 2011) and found its way to the commercial market. Studies on mastitis prediction (de Mol and Ouweltjes, 2001; Kramer et al., 2009), calving time prediction based on a diurnal pattern analysis (Maltz et al., 2011), and lameness incidence (Kramer et al., 2009; Ito et al., 2010) were also reported. Ito et al. (2010) found that measures of lying behavior alone were not sensitive diagnostics for lameness detection, and, hence, they suggested using these automated measures of lying behavior in a multi-pronged approach for lameness detection. Kramer et al. (2009) reported high within- and between-cow variability, which made their fuzzy logic lameness detection model not suitable for practical use.

The diurnal rhythm affects animal behavior and farm routine. A correct management of the photoperiod affects dairy cow performance and health (Dahl and Petitclerc, 2003). Feeding and milking time influence the time budget of a dairy cow (Devries and von Keyserlingk, 2005; Belle et al., 2012). The photoperiodic effect on lying time was used to detect calving time of dairy cows (Maltz et al., 2011). Besides a diurnal analysis of the behavior data, discrimination was made between the daytime and nighttime period. Behavior during nighttime was found to be a better predictor for calving time in the study of Maltz et al. (2011).

To our knowledge, no attempt has been done to automatically detect lameness based on a combination of 3 sensor based variables (milk yield, neck activity, and ruminating time) and a diurnal pattern analysis of these variables. The aim of this study was to develop a mathematical model based on correlated variables measured with existing low-cost sensors to detect clinical lameness based on behavioral and milk performance data.

**MATERIALS AND METHODS**

**Animals and Housing**

Data were gathered from Refet HaEmek, a commercial Israeli dairy farm located in kibbutz Yifat. The milking herd consisted of 1,100 Israeli Holstein milking cows. The average parity in the herd was 2.6 ± 1.6 lactations, with a replacement rate of 33%. The annual milk production was 11,500 kg/cow. The herd contained 11 production groups according to parity, lactation stage, reproduction status, and health status (group size = 96 ± 12 cows). Cows with severe diseases, such as acute mastitis and clinical lameness, were housed in a small, separated group referred to as the hospital group (15 ± 4 cows). All cows were housed in a separate no-stall, fully roofed, open cowshed with dry manure bedding material. Each cowshed had a post-and-rail feed fence on 1 side of the barn and the area near this feeding lane was paved in concrete. Stocking rate was, on average, 19.3 m²/cow in each production group (Figure 1).

All cows were milked 3 times per day (at 0300, 1100, and 1900 h) in a 2 x 32 parallel milking parlor. Each milking session took about 6 h. The production groups were brought one-by-one to the waiting parlor to avoid group mixing. During each milking session, not more than 2 groups were in the waiting area. Cows in early lactation were milked first, then the primiparous groups, followed by multiparous cows, and lastly the hospital group.

Each milking session was done by 4 workers; 3 workers milked the cows and 1 worker (pusher) brought the cows to and from the milking parlor. Cows with a deviation in gait and suspected as lame were identified by the cow pusher, as he had the best view on cow’s locomotion. The cow pusher was a trained herdsman, but did not receive any specific training in locomotion scoring or lameness detection. The cows suspected as lame by the pusher were presented the next day to
the veterinarian. All cows presented to the vet were confirmed as lame and treated. Daily veterinary services were provided by the national veterinary service (Hachaklait Veterinary Services Ltd., Caesarea, Israel).

All cows were fed a TMR according to NRC recommendations (NRC, 2001). Trucks from the local feeding center delivered the TMR twice daily (around 0600 and 1100 h). The feed was pushed to the cows at least 6 times per day.

### Off-The-Shelf Sensors

The milking parlor was equipped with an electronic milk yield meter (FreeFlow, SCR Engineers Ltd., Netanya, Israel) at every individual stand. This gave a milk yield measurement, expressed in kilograms, during each milking. All cows in the farm were equipped with a neck collar tag (HR-Tag, SCR Engineers Ltd.) used for automated heat detection (Bar, 2010; Kamphuis et al., 2012). The tag had 3 functions: (1) cow identification based on optical signal transmission, (2) real-time individual neck activity level measurement, and (3) real-time individual ruminating time measurement. Activity measurement was based on the signal analysis of the head movements, and was expressed by a filtered activity index ranging from 0 to 255 for a 2-h interval. The activity index was proportional to the number, intensity, and direction of the neck movements (Figure 2). The rumination logger continuously recorded the time spent ruminating within a 24-h period in 2-h intervals, as was validated by Schirmann et al. (2009) and Burfeind et al. (2011). Ruminating time was expressed as minutes of ruminating per 2 h. All data were transferred automatically during each milking to the herd management software (DataFlow, SCR Engineers Ltd.), and reports were extracted from this software in a spreadsheet format (Microsoft Excel, Microsoft Corp., Redmond, WA).

### Database Building

Herd health reports were used to build a database. Data were acquired from August 2011 until May 2012. In this study, cows were considered as lame when they were first identified by the cow pusher and then diagnosed and treated by a veterinarian for one of the lameness diagnoses described in Table 1. This may imply that only severe lameness cases were considered in this research. If any metabolic disease (such as ketosis), digestive problems, or mastitis was reported before the lameness diagnosis, the cow was omitted from the database. Cows were considered not lame and included into analyses when (1) a lameness diagnosis (Table 1) was absent during their entire current lactation, (2) no metabolic disease was diagnosed during their current lactation, and (3) no unexplained drop in their lactation curve bigger than 10% on 2 consecutive days was observed.

To avoid the effect of postcalving stress on behavior and performance, only cows that were more than 40 d in lactation were considered in the further analysis. In total, 118 cows were used in this analysis: 44 lame cows and 74 not lame cows. Data were analyzed from 3 wk before until 3 wk after the diagnosis. For not lame cows, the day of diagnosis was a random day in the lactation period. The lame group (166 ± 99 d) and the not lame group (151 ± 72 d) were similar in the timing of the diagnosis day relative to calving.

### Model Variable Creation

The 2 behavioral variables, neck activity and ruminating time, and the performance variable, milk yield, were used to calculate different model input variables. These model variables included absolute values, stan-
standard deviations, differences and ratios, variables related to individual temperament, relative values, day to night ratios, slopes, and milking session-related input variables as described in Table 2.

**Statistical Analysis: Effect of Lameness on Cow Performance and Behavior**

All data were divided in 2 groups: lame and not lame. For each day during a 6-wk period (21 d before until 21 d after diagnosis and treatment), input data of both groups of all the above listed variables were compared against each other using the Kruskal-Wallis one-way ANOVA by ranks test (Kruskal and Wallis, 1952). This method performs a test of the null hypothesis that the data in both groups (lame group vs. not lame group) were independent and originated from the same distribution. A normal distribution of the data is not necessarily assumed. Significance level was set to $\alpha = 0.05$.

For both groups (lame and not lame), the daily group median value of each variable was calculated. The median value is a more robust measurement for group level than the average value due to the high intercow variability. Outliers have less influence on the median than on the average.

The 7-d average of the daily median value was calculated for all variables to compare both groups against each other. The 7-d average is a commonly used number in management software for health and performance monitoring purposes.

**Model Development**

A statistical model was developed to detect lameness based on behavioral and performance data. The model input variables with the highest correlation coefficients were compared with the dichotomous output variable `healthstatus`, which indicated to which group (lame or not lame) the cow belonged. The best fitting model was empirically chosen based on model accuracy. The model was developed with MatLab software (MatLab, The MathWorks Inc., Natick, MA).

Due to the dichotomous output (`healthstatus`: lame or not lame), a logistic regression model was used. The logistic regression model used a combination of the $n$ model variables to fit the response variable $z$ (Hosmer and Lemeshow, 2000). Before entering the logistic regression model, each model input variable ($var$) from Table 2 was transformed to a uniform scale (0–1), using

![Figure 2. Example of neck activity level of 1 lame cow, assessed by means of data loggers, and expressed by means of 2-h data points in function of time. Milking time of this specific cow, a member of the hospital group that was milked last, is shown by the vertical dashed lines. Once every 4 data points (8 h) an increase in neck activity level seems to occur. This recurring 8-h pattern of increased activity level coincides with the time of milking. Color version available in the online PDF.](image-url)
In the model training phase, the rescaled variables ($x_{\text{rescaled}}$) were used to estimate the model parameters $\beta_i$ by

$$x_{\text{rescaled}} = \frac{\text{var} - \text{minimum}(\text{var})}{\text{maximum}(\text{var}) - \text{minimum}(\text{var})},$$

$$z = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \ldots + \beta_n \times x_n.$$  \[[1]\]

The $z$-value resulting from equation 2 was used to calculate the probability of the model output being a positive model output. In this case, a positive model output
corresponds to being lame, and a negative model output to being not lame. The probability was calculated as
\[
\text{probability} = \frac{\exp(z)}{1 + \exp(z)}.
\]  \[3\]

An empirically chosen cutoff threshold classified the probability in 2 groups: not lame cows versus lame cows. The logistic regression model and its parameters were built from the complete database (n = 118). The MatLab Statistics Toolbox was used to estimate the model parameters from Equation 2 on a daily basis.

**Model Validation**

Model validation was done by a 10-fold cross-validation procedure on the complete database (n = 118). To evaluate the model, a confusion matrix was used (Fielding and Bell, 1997). This matrix compared the reference output (veterinary diagnosis and treatment) against the model output (logistic regression model). A positive output referred to the lame category, a negative output referred to the not lame category.

In the case of 10-fold cross-validation, the data set is randomly divided in 10 subsamples with an equal ratio of lame and not lame animals remaining in each subsample (Stone, 1977). One of the subsamples is retained as validation set, whereas the other 9 subsamples are used for model training. In each cross-validation step, the parameters, \( \beta_i \), of equation 2 are estimated with the model training data. This procedure is repeated until all subsamples were used 1 time for model validation. The resulting confusion matrices are combined to result the 10-fold cross-validation confusion matrix.

In the confusion matrices, 5 classification measures were presented in the bottom row (sensitivity and specificity; left to right) and the last column (correct classification rate or accuracy, positive predictive power and negative predictive power; top to bottom). All 5 classification measures were based on the number of true positives, true negatives, false positives, and false negatives.

The receiver operating characteristic-curve is a graphical plot that illustrates the performance of a classification system as its discrimination threshold is varied (Hosmer and Lemeshow, 2000). It is created by plotting the true positive rate (or sensitivity) versus the false positive rate (1 − specificity). The area under the curve is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

**RESULTS**

**Statistical Analysis: Effect of Lameness on Cow Performance and Behavior**

The effect of lameness on animal behavior and production performance is represented in Figures 3, 4 and 5 for the 3 most relevant variables. The difference between the not lame group of cows and the lame group in median daily milk yield was significant for all days from 3 wk before until 3 wk after diagnosis and treatment (Figure 3; Table 3). The 7-d average of daily milk yield on the day of diagnosis and treatment was significantly lower for the lame animals (37.7 ± 0.5 vs. 45.1 ± 0.5 kg; \( P < 0.001 \)).

Not lame cows had a lower night to day ratio of neck activity due to their higher neck activity during daytime compared with lame cows (Figure 4; Table 3). The 7-d average of the night-period to day-period neck activity ratio was significantly lower for lame animals (0.36 ± 0.01 vs. 0.39 ± 0.01; \( P < 0.001 \)).

Around the day of diagnosis and treatment (d0), the difference in ruminating time during the night period was more clearly visible than at the beginning and end of the analyzed period (Figure 5). A trend was observed for a decreased ruminating time during the night period. On d−6, d−4, d−3, d0, d1, and d7 this difference was significant (Table 3). The 7-d average of the nighttime ruminating time was significantly lower for the lame animals (208 ± 6 vs. 221 ± 3 min; \( P < 0.01 \)).

**Variable Correlations**

Table 2 shows the maximal correlation coefficients between the independent input variables and the dependent dichotomous output variable for the 7 d before until the day of diagnosis and treatment. The highest correlation of all variables was the milk-related variable milk yield 4 d before the diagnosis and treatment (\( r_{MY} = 0.45 \)). For the neck activity, the highest correlation with lameness was the night to day ratio on d 6 before diagnosis (\( r_{ANDR} = 0.38 \)). The highest correlation with a ruminating variable was found on d 6 before diagnosis for the nighttime ruminating time (\( r_{RNP} = 0.30 \)). The correlation coefficients for milk yield-related variables (0.33 ± 0.11) were on average higher than neck activity-related (0.23 ± 0.06) and ruminating time-related (0.21 ± 0.07) variables.

**Logistic Regression Model**

The best fitting model consisted of 7 model input variables with a correlation coefficient bigger than 0.30:
Figure 3. Graphical overview of median daily milk yield for 44 lame cows (●) and 74 not lame cows (▼) relative to day of diagnosis and treatment (vertical dashed line) for lame animals. The results of the statistical difference between the 2 groups can be found in Table 3.

Figure 4. Graphical overview of the night and day ratio of the neck activity level in function of time relative to day of diagnosis and treatment (dashed vertical line) for 44 lame cows (●) and 74 not lame cows (▼).
the daily milk yield 4 d before diagnosis; the slope coefficient of the daily milk yield 4 d before diagnosis; the nighttime to daytime neck activity ratio 6 d before diagnosis; the milk yield week difference ratio 4 d before diagnosis; the milk yield week difference 4 d before diagnosis; the neck activity level during the daytime 7 d before diagnosis; and the ruminating time during nighttime 6 d before diagnosis. Table 4 gives an overview of the estimated parameters of equation 2 in each of the 10 cross-validation steps. The confusion matrix resulting from the 10-fold cross-validation procedure is given in the left panel of Table 5. The logistic regression model obtained with the 10-fold cross-validation a sensitivity of 0.77, a specificity of 0.84, and a model accuracy of 0.81.

The average values of the model coefficients after the 10-fold cross-validation process are given in the bottom of Table 4. A validation on the complete data set based on these model parameters resulted in the confusion matrix presented in Table 5 in the right panel. The model obtained a sensitivity of 0.89, a specificity of 0.85, and an accuracy of 0.86. The area under the receiver operating curve was 0.89.

**DISCUSSION**

Identification of healthy and lame cows was entirely based on the herd health reports of the farm. The accuracy of these herd health reports were dependent on the identification of lameness by the herdsman, the reporting of the lameness cases and the expertise of the veterinarian. The time between the onset of lameness and veterinary diagnosis and treatment was not well known. However, the veterinarian visited the farm every day (except Saturday), so it could be assumed that the time delay was small in this farm. Moreover, veterinary services were provided by the Hachaklait Veterinary Services Ltd., and all associated veterinarians were trained to perform the same protocols in diagnosing and treating the animals.

The choice to use only data from strictly healthy cows with no reported metabolic or other diseases was done because it was never clear how long the recovery from such diseases took for each individual cow. Lame cows in this study were defined as cows that were diagnosed and treated for lameness issues and no other diseases. If a change in behavior was observed, this could be
expected to come from the lameness event and no other cause. In practice, the farmer deals with the entire herd where this strict distinction is not that clear. In this study, 118 cows agreed with the definitions of the lame or the not lame group on a complete milking herd of 1,100 cows.

Lame cows in this study had lower milk yield compared with not lame cows. The 7-d average of daily milk yield on the day of diagnosis and treatment was significantly lower for lame animals than for not lame animals. The effect of lameness on milk yield has been documented in literature (Green et al., 2002; Archer et al., 2010). Reader et al. (2011) found comparable results on lower milk yields in lame cows before the diagnosis and treatment for lameness. Lame cows tend to have a lower milk yield in the complete period of the analysis, not only around the day of diagnosis. This may indicate that lameness is a long-term developing

### Table 3.

<table>
<thead>
<tr>
<th>Day</th>
<th>MY (kg/d) Not lame</th>
<th>MY (kg/d) Lame</th>
<th>ANDR Not lame</th>
<th>ANDR Lame</th>
<th>RNP (min) Not lame</th>
<th>RNP (min) Lame</th>
<th>RM3 (min) Not lame</th>
<th>RM3 (min) Lame</th>
</tr>
</thead>
<tbody>
<tr>
<td>−21</td>
<td>46.3**</td>
<td>40.1</td>
<td>0.375*</td>
<td>0.406</td>
<td>219</td>
<td>222</td>
<td>64</td>
<td>58</td>
</tr>
<tr>
<td>−14</td>
<td>46.3†</td>
<td>41.1</td>
<td>0.365</td>
<td>0.381</td>
<td>230</td>
<td>219</td>
<td>58</td>
<td>60</td>
</tr>
<tr>
<td>−7</td>
<td>43.1†</td>
<td>38.7</td>
<td>0.359†</td>
<td>0.383</td>
<td>218</td>
<td>224</td>
<td>46</td>
<td>55</td>
</tr>
<tr>
<td>−6</td>
<td>42.7**</td>
<td>38.8</td>
<td>0.350†</td>
<td>0.394</td>
<td>233**</td>
<td>195</td>
<td>57</td>
<td>56</td>
</tr>
<tr>
<td>−5</td>
<td>45.4†</td>
<td>39.6</td>
<td>0.360†</td>
<td>0.390</td>
<td>229</td>
<td>204</td>
<td>54</td>
<td>56</td>
</tr>
<tr>
<td>−4</td>
<td>45.2†</td>
<td>38.1</td>
<td>0.354**</td>
<td>0.386</td>
<td>236*</td>
<td>212</td>
<td>48</td>
<td>51</td>
</tr>
<tr>
<td>−3</td>
<td>45.6†</td>
<td>37.2</td>
<td>0.372*</td>
<td>0.389</td>
<td>229</td>
<td>211</td>
<td>59</td>
<td>50</td>
</tr>
<tr>
<td>−2</td>
<td>44.4†</td>
<td>37.8</td>
<td>0.367</td>
<td>0.379</td>
<td>228</td>
<td>226</td>
<td>60*</td>
<td>50</td>
</tr>
<tr>
<td>−1</td>
<td>43.0†</td>
<td>37.0</td>
<td>0.356†</td>
<td>0.391</td>
<td>228</td>
<td>226</td>
<td>60*</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>46.2†</td>
<td>36.6</td>
<td>0.358</td>
<td>0.369</td>
<td>232**</td>
<td>203</td>
<td>51</td>
<td>47</td>
</tr>
<tr>
<td>2</td>
<td>44.2†</td>
<td>36.9</td>
<td>0.323†</td>
<td>0.390</td>
<td>225</td>
<td>221</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>44.7†</td>
<td>39.0</td>
<td>0.354*</td>
<td>0.376</td>
<td>234</td>
<td>230</td>
<td>53</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>45.1†</td>
<td>38.3</td>
<td>0.373</td>
<td>0.381</td>
<td>226</td>
<td>231</td>
<td>61</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>44.3†</td>
<td>38.4</td>
<td>0.356</td>
<td>0.374</td>
<td>226</td>
<td>221</td>
<td>48</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>42.9†</td>
<td>37.7</td>
<td>0.343**</td>
<td>0.370</td>
<td>225</td>
<td>221</td>
<td>53</td>
<td>55</td>
</tr>
<tr>
<td>7</td>
<td>44.1†</td>
<td>38.0</td>
<td>0.344</td>
<td>0.354</td>
<td>238*</td>
<td>205</td>
<td>53</td>
<td>51</td>
</tr>
<tr>
<td>14</td>
<td>42.9†</td>
<td>38.8</td>
<td>0.306†</td>
<td>0.392</td>
<td>219</td>
<td>220</td>
<td>48</td>
<td>53</td>
</tr>
<tr>
<td>21</td>
<td>43.8†</td>
<td>38.4</td>
<td>0.327**</td>
<td>0.372</td>
<td>236</td>
<td>220</td>
<td>48*</td>
<td>59</td>
</tr>
</tbody>
</table>

1 The presented model input variables are the daily milk yield (MY), nighttime to daytime ratio of the neck activity (ANDR), nighttime ruminating time (NRNP), and the ruminating time during the 3 milking session periods (RM3). Group difference between the not lame and lame group was statistically tested with Kruskal-Wallis significance test.

*Not lame group significantly different from lame group (α < 0.05).

**Not lame group significantly different from lame group (α < 0.01).

†Not lame group significantly different from lame group (α < 0.001).

Lame cows in this study had lower milk yield compared with not lame cows. The 7-d average of daily milk yield on the day of diagnosis and treatment was significantly lower for lame animals than for not lame animals. The effect of lameness on milk yield has been documented in literature (Green et al., 2002; Archer et al., 2010). Reader et al. (2011) found comparable results on lower milk yields in lame cows before the diagnosis and treatment for lameness. Lame cows tend to have a lower milk yield in the complete period of the analysis, not only around the day of diagnosis. This may indicate that lameness is a long-term developing

### Table 4.

<table>
<thead>
<tr>
<th>Fold</th>
<th>a0</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>a5</th>
<th>a6</th>
<th>a7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.99</td>
<td>−6.43</td>
<td>−4.64</td>
<td>4.44</td>
<td>−12.26</td>
<td>14.61</td>
<td>−3.45</td>
<td>−1.18</td>
</tr>
<tr>
<td>2</td>
<td>−1.56</td>
<td>−5.55</td>
<td>−9.48</td>
<td>6.85</td>
<td>−23.83</td>
<td>30.78</td>
<td>−3.36</td>
<td>0.86</td>
</tr>
<tr>
<td>3</td>
<td>1.71</td>
<td>−5.94</td>
<td>−4.66</td>
<td>4.83</td>
<td>−12.99</td>
<td>15.84</td>
<td>−3.13</td>
<td>−0.59</td>
</tr>
<tr>
<td>4</td>
<td>0.85</td>
<td>−7.57</td>
<td>−5.54</td>
<td>8.50</td>
<td>−13.11</td>
<td>16.81</td>
<td>−2.28</td>
<td>−0.91</td>
</tr>
<tr>
<td>5</td>
<td>5.13</td>
<td>−8.01</td>
<td>−5.25</td>
<td>3.88</td>
<td>−8.29</td>
<td>10.07</td>
<td>−4.06</td>
<td>−0.59</td>
</tr>
<tr>
<td>6</td>
<td>2.53</td>
<td>−6.20</td>
<td>−4.36</td>
<td>5.45</td>
<td>−8.81</td>
<td>10.16</td>
<td>−2.97</td>
<td>−0.34</td>
</tr>
<tr>
<td>7</td>
<td>3.87</td>
<td>−7.89</td>
<td>−4.28</td>
<td>4.75</td>
<td>−16.64</td>
<td>17.69</td>
<td>−3.37</td>
<td>−1.35</td>
</tr>
<tr>
<td>8</td>
<td>1.36</td>
<td>−6.27</td>
<td>−5.33</td>
<td>4.94</td>
<td>−13.27</td>
<td>16.75</td>
<td>−2.89</td>
<td>−0.17</td>
</tr>
<tr>
<td>9</td>
<td>1.55</td>
<td>−6.10</td>
<td>−4.91</td>
<td>4.98</td>
<td>−13.41</td>
<td>17.10</td>
<td>−2.94</td>
<td>−1.15</td>
</tr>
<tr>
<td>10</td>
<td>2.72</td>
<td>−7.05</td>
<td>−7.18</td>
<td>5.12</td>
<td>−13.64</td>
<td>16.52</td>
<td>−3.14</td>
<td>0.41</td>
</tr>
<tr>
<td>Average</td>
<td>2.12</td>
<td>−6.70</td>
<td>−5.56</td>
<td>5.37</td>
<td>−13.63</td>
<td>16.63</td>
<td>−3.17</td>
<td>−0.50</td>
</tr>
</tbody>
</table>

1 The logistic regression model consists of 7 model input variables and has the following equation:

$$z = a_0 + a_1 \times MY_{−4} + a_2 \times MYS_{−4} + a_3 \times ANDR_{−6} + a_4 \times MWR_{−4} + a_5 \times MWD_{−4} + a_6 \times ADP_{−7} + a_7 \times RNP_{−6},$$

with the variables daily milk yield ($MY_{−4}$), slope coefficient of 7 consecutive daily milk yield measurements ($MYS_{−4}$), the nighttime to daytime ratio of the neck activity ($ANDR_{−6}$), the ratio between 2 consecutive weeks of daily milk yield measurements ($MWR_{−4}$), the absolute difference between 2 consecutive weeks of daily milk yield measurements ($MWD_{−4}$), the neck activity during daytime ($ADP_{−7}$), and the ruminating time during the night period ($RNP_{−6}$); all subscripts represent day relative to diagnosis.
The daytime (data not shown). Ito et al. (2010) stated to the decreased neck activity of the lame cows during milks (0.39 ± 0.01 vs. 0.36 ± 0.01). This was mainly due to the neck activity ratio being significantly higher for lame animals (Figure 4). The 7-d average of the nighttime to daytime ratio of the neck activity showed a trend to be higher for lame cows than for not lame cows. The 7-d average of the nighttime ruminating time during this period may be due to variations in the night compared with the not lame cows around the day of diagnosis. The fluctuations in average ruminating time during the night period (RNP,) and the milking time during the night period (RNP,) all subscripts represent day relative to diagnosis. The confusion output matrix of the 10-fold cross-validation is presented in the left panel; the confusion matrix based on the averaged cross-validation model parameters is given in the right panel. In the bottom row of each panel, model sensitivity (ability to detect lame animals; left) and model specificity (ability to detect not lame animals; middle) are given. Model accuracy (ability to correctly classify observations) is given in the top right corner of the panel.

Differences in cow behavior were found by comparing the daytime and nighttime periods. Normal, healthy dairy cow behavior is affected by a diurnal pattern that includes the photoperiod (duration of the light span) and scotoperiod (duration of the dark span; Dahl and Petitclerc, 2003; Provolo and Riva, 2008; Piccione et al., 2011), as well as by the management routine in the dairy farm, such as feeding (Devries and von Keyserlingk, 2005; Belle et al., 2012) and milking (O’Driscoll et al., 2010; Piccione et al., 2011). During the night, only 1 milking was done and the feed was pushed twice to the cows. The results in this study showed that, on average, lame cows had a reduced ruminating time during the night compared with the not lame cows around the day of diagnosis. The fluctuations in average ruminating time during this period may be due to variations in ruminating as response to lameness between individual cows. The 7-d average of the night-period ruminating time was significantly lower for the lame animals. The ratio of nighttime and daytime neck activity showed a trend to be higher for lame cows than for not lame cows (Figure 4). The 7-d average of the nighttime to daytime neck activity ratio was significantly higher for lame animals (0.39 ± 0.01 vs. 0.36 ± 0.01). This was mainly due to the decreased neck activity of the lame cows during the daytime (data not shown). Ito et al. (2010) stated that lame cows had an increased lying time and length of lying bouts compared with healthy cows. This may suggest a reduced activity level of the lame cows. In the study of Ito et al. (2010), no differentiation between nighttime and daytime was observed.

The neck activity measurement by head movements on this farm could be dependent on the production group of the cow. The distance between the cowshed and the milking parlor was not the same for each group (Figure 1). Therefore, the activity was different in each group when they walked to the milking parlor 3 times per day. The severity and expression of lameness can also affect the activity level, as severely lame cows tend to take smaller steps, in some cases accompanied with an expressive head bob, and hence they make more steps to cover the same distance.

Correlation coefficients for milk yield-related variables were on average higher than neck activity-related and rumination time-related variables, indicating that performance data (milk yield) were more correlated with lameness than overall behavior (rumination and neck activity). Nevertheless, all presented coefficients in Table 2 indicated a weak correlation between the individual variables and the health status parameter. Hence, not one single variable was clearly related to lameness. This result agrees with the conclusions from other studies that suggest using a combination of the available variables to obtain an adequate lameness detection model (Kramer et al., 2009; Ito et al., 2010).

A logistic regression model was found to fit the data in the data set. For model validation, a 10-fold cross-validation procedure was chosen because (1) the data set was small (n = 118) and (2) this validation procedure is commonly used. If holdout validation was done, the initial data set was split randomly into 2 parts: 1

<table>
<thead>
<tr>
<th>Model</th>
<th>Reference</th>
<th>10-fold cross-validation output</th>
<th>10-fold cross-validation averaged parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lame</td>
<td>34</td>
<td>12</td>
<td>39</td>
</tr>
<tr>
<td>Not lame</td>
<td>10</td>
<td>62</td>
<td>5</td>
</tr>
<tr>
<td>Performance</td>
<td>0.77</td>
<td>0.84</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Table 5. Confusion matrices of the logistic regression model validation by 10-fold cross-validation

The logistic regression model was based on 7 input variables: daily milk yield (MY), slope coefficient of 7 consecutive daily milk yield measurements (MYS), the nighttime to daytime ratio of the neck activity (ANDR), the ratio between 2 consecutive weeks of daily milk yield measurements (MWR), the absolute difference between 2 consecutive weeks of daily milk yield measurements (MWD), the neck activity during daytime (ADP), and the ruminating time during the night period (RNP). All subscripts represent day relative to diagnosis. The confusion output matrix of the 10-fold cross-validation is presented in the left panel; the confusion matrix based on the averaged cross-validation model parameters is given in the right panel. In the bottom row of each panel, model sensitivity (ability to detect lame animals; left) and model specificity (ability to detect not lame animals; middle) are given. Model accuracy (ability to correctly classify observations) is given in the top right corner of the panel.
usually bigger part used for model calibration and the remaining part was used for model validation. Hold-out validation is more sensitive than cross-validation because each observation is randomly assigned to the calibration data set or to the validation data set, and not used for both, as it is for cross-validation.

Compared with the fuzzy logic model of Kramer et al. (2009), all obtained specificity values (0.84 and 0.85 vs. 0.76) and sensitivity values (0.77 and 0.89 vs. 0.76) in the current study were better. This may suggest that the presented logistic regression model is better, or more robust, than the fuzzy logic model by Kramer et al. (2009) because it deals better with the huge variability between and within cows that, according to Kramer et al. (2009), might have caused the high error rates in their study. The strict selection of lame and not lame animals in the present study might have caused the high sensitivity and specificity values. In practice, the chance of working with imperfect data are higher, so it will be less clear if behavioral changes occur by lameness or any other disease.

Some points should be considered when using continuous behavioral sensor data to detect lameness. Dairy cattle are descendants from ranging wild cattle that were prone to predator attack. To improve their chance of survival, they masked any signs of pain and its implied weakness (Phillips, 2002). This can imply that cows will only show pain when their level of pain is high and their suffering above a certain threshold. This might be a drawback in the early detection of lameness based on existing sensor information, as only a severely lame cow will show a significant change in behavior, such as activity due to pain. It is also possible that cows hide their weakness from other herd mates in order not to lose their rank in the social hierarchy of the herd. A lower rank implies less eating and therefore less ruminating, less lying, and more standing, which in turn leads to an increased chance of lameness (Galindo and Broom, 2002).

The present model incorporated 1 specific day before the diagnosis of each variable, according to the highest correlation with the healthstatus output variable. The differences between these day-related correlations were small. A more extended data set could possibly reveal other more influencing days.

Sensitivity and specificity are always a trade-off between each other. The selection of the cutoff threshold in the logistic regression model influenced the false positive and false negative rates, and hence the sensitivity and specificity of the model. A high false positive value meant that more healthy cows will be presented to the veterinarian. Conversely, a high false negative rate implied that some lame cows were missed by the model. Compared with mastitis detection or heat detection, the time tolerance for detecting lameness is much longer. Mastitis has to be detected during the first milk flow; otherwise the whole tank will be spoiled. For heat detection, the timing is also crucial, as insemination on the correct moment is a key factor for insemination success. Lameness detection might be considered less urgent than other factors. This implies that the sensitivity (model’s ability to detect positive outcome) of a lameness detection model does not have to be as high as a mastitis or heat detection model. In this study however, only severely lame cows that were recognized by the pusher were included. Less lame cows also suffer from pain and do not produce at full potential, resulting in an economic loss for the farmer. The well-being of these cows is impaired as well, and hence an early detection is favorable. Other studies have shown that early lameness detection is a key factor regarding the improvement of the well-being of the animals in the herd (Bruijnis et al., 2012), and the associated economic losses (Bruijnis et al., 2010).

Combination with other automated lameness detection systems based on gait could improve the model efficiency of both detection methods. Gait-related lameness detection methods were found in literature, such as the use of load sensors for the measurement of ground reaction forces (Rajkondawar et al., 2006; Liu et al., 2009, 2011), pressure sensitive walkways (Maertens et al., 2011), and the use of image processing (Pluk et al., 2010; Poursaberi et al., 2010; Pluk et al., 2012). The predictive lameness model of Liu et al. (2011) had a sensitivity of 0.52 and a specificity of 0.89. Compared with the present study, they had a bigger data set, but the sensitivity was lower (0.52 vs. 0.77 and 0.89). The area under the receiver operating curve of the logistic regression model was also bigger than the best values found by Rajkondawar et al. (2006; area under the curve = 0.89 vs. 0.84). The work of Poursaberi et al. (2010) showed a correct classification rate of 0.96 of an automatic lameness detection method based on the analysis of the back curvature of the cow in a video. Their correct classification rate was considerably higher than found in this study. The work of Maertens et al. (2011) showed a correct classification rate of 0.84 for a 3-point lameness classification. The sensitivity ranged from 0.76 to 0.90 for all 3 lameness categories, whereas the specificity had a range of 0.86 to 1.00. Their measurement success rate was comparable to this study, but their setup was highly dependent on cow traffic and cow behavior. It may be incorrect to compare the results to all these studies, as the references were not the same: a scaled reference based on the locomotion score versus a dichotomous reference based on the combination of visual lameness recognition by the pusher and veterinary diagnosis and treatment.
Other models, such as the classification tree model (Breiman et al., 1984), were also tested in the analysis. However, the classification performance was lower than for the logistic regression model (unpublished data).

To our knowledge, the present study is the first attempt to use existing data on behavior and performance from no-stall, open cowshed farms in the automatic detection of dairy cattle lameness. The sensors in this farm were mainly used for heat detection and monitoring of the nutritional status of the herd. This study showed that the combination of existing farm data containing milk yield, neck activity, and ruminating could also be used for the automatic detection of lameness. In future research, combining the measures used in the current study with image-processing techniques should be tested for improving lameness detection accuracy.

**CONCLUSIONS**

The present study shows that lameness affected dairy cow behavior and milk production performance. Milk yield and ruminating time during the night period were significantly lower for lame animals. The night to day ratio of neck activity was higher for lame animals. The correlation between the healthstatus output variable (lame or not lame) and the milk yield input variable was the highest ($r_{MY} = 0.45$). The logistic regression model for automated clinical lameness detection based on performance and behavioral data obtained a correct classification rate of 0.86, sensitivity of 0.89, and specificity of 0.85 in a 10-fold cross-validation procedure. This study showed that the combination of existing farm data containing milk yield, neck activity, and ruminating time could also be used for the automated detection of lameness.

**ACKNOWLEDGMENTS**

The authors thank all farm staff for their help on the farm. The help from Doron Bar and Rony Meir from SCR Engineers Ltd. (Netanya, Israel) and Shuli Cohen and Michael van Straten from Hachalait Veterinary Services Ltd., (Caesarea, Israel) is also very much appreciated. This study is part of the FP7-PEOPLE-ITN-2008 Marie Curie Initial Training Network BioBusiness (Leuven, Belgium). This study is contribution number 459-4398-951, funded by the Agricultural Research Organization (Bet Dagan, Israel).

**REFERENCES**


