ABSTRACT

The objective of the current study was to develop a predictive model for calf disease detection in the preweaning period using data from automated milk feeders (AMF). A deep convolutional neural network (CNN) architecture for the detection of respiratory disease and diarrhea in dairy calves was developed. German Holstein calves were fed milk replacer either ad libitum (up to 25 L/d; n = 32) or restrictively (6 L/d; n = 32) via AMF from 10 ± 3 d of life on. Concentrate, hay, and water were freely available. Calf health parameters were scored daily. The AMF measured milk replacer (MR) intake, number of rewarded visits, number of unrewarded visits, and drinking speed. A calf was considered sick if its fecal score was 3 or 4 and its respiratory score was 2 or 3. Only data from AMF up to 47 d of age were included in the analysis. This cut in the data was made to avoid data from the weaning period. Data were split in 80:20 ratios for training and testing data sets according to the Pareto principle. A minimum sensitivity of 80% was considered an appropriate requirement for the prediction models. Considering all calves in group housing, cross-validation of the test data set showed a sensitivity of 83% and a specificity of 79%, with a positive predictive value and a negative predictive value of 37 and 97%, respectively (for ad libitum-fed calves), and 82 and 87%, respectively (for restricted-fed calves), with good area under the curve-receiver operating characteristic (0.77 to 0.87), indicating that the CNN models can predict calf disease in both groups with different MR allowances. The permutation feature importance was measured by the decrease in model accuracy, and features (behaviors) were summarized in descending order of their relative importance to the CNN model. Drinking speed and MR intake were the main factors to predict calf disease in calves fed ad libitum. The number of unrewarded visits to the milk feeder and MR intake were the main factors to predict calf disease in restricted-fed calves. Despite the relatively small sample size, the results provide strong evidence that daily feeding behavior data from AMF can be used to identify calves at risk for disease. In conclusion, despite a very good testing performance of the CNN model, the relatively low daily prevalence of calf disease in the present study resulted in a high proportion of false-positive alarms.

Key words: calf, deep learning, feeding behavior, disease

INTRODUCTION

Young dairy calves are particularly susceptible to disease in early life (Renaud et al., 2018). Respiratory disease and diarrhea remain the most common infectious causes of morbidity and mortality in preweaning calves (Uetake, 2013; Windeyer et al., 2014). It has been shown that respiratory disease causes decreased calf growth in the short term (Virtala et al., 1996), decreased survival to the first year of life, and reduced milk production in the long term (Adams and Buczinski, 2016; Teixeira et al., 2017; Dunn et al., 2018). Calves with diarrhea suffer from dehydration and decreased appetite (Trefz et al., 2017). Earlier detection of calves with diarrhea and...
respiratory disease may allow alternative interventions to reduce antibiotic use.

Calf health status can be assessed using clinical scoring systems (Love et al., 2014). In health scoring systems, numeric scores are indicators of specific calf health traits such as nasal discharge, ear position, cough, rectal temperature, and fecal consistency (McGuirk and Peek, 2014; School of Veterinary Medicine, 2017). Nevertheless, visual assessment of calf diseases, particularly respiratory disease, can be subjective, leading to delayed or overlooked detection of sick animals, as well as unnecessary antimicrobial treatment (Abutarbush et al., 2012; Maier et al., 2019). In many applications, it is desirable to have an automated algorithm that can identify animals at higher risk for disease. Behavioral data from an automated milk feeder (AMF) system may help producers identify sick calves (Conboy et al., 2021; Costa et al., 2021; Morrison et al., 2022). Calves are typically housed in groups when an AMF system is used, and it can be difficult for calf producers to monitor the health of each calf or identify sick calves in large groups (Medrano-Galarza et al., 2018; Conboy et al., 2021), so automation of disease detection is needed. Given the numerous benefits of automated feeding and the ability to track individual behavioral data provided by AMF, further research is needed on the suitability of AMF data for disease detection in dairy calves.

Time series and multivariate data generated by sensors and AMF systems (Svensson and Jensen, 2007; Belaid et al., 2020) can be used through machine learning (ML) algorithms for analyzing, predicting, and notifying farmers of abnormal conditions (Neethirajan, 2020). Conventional ML and deep learning, as a subset of artificial intelligence, are used in dairy cattle research due to their advantages in various tasks such as early detection of diseases (e.g., diarrhea and respiratory diseases; Belaid et al., 2020) and, furthermore, to classify dairy cattle based on their metabolic status (Ghaffari et al., 2020) or feeding behavior (Borderas et al., 2009; Knauer et al., 2017; Ghaffari et al., 2021). Deep learning techniques based on convolutional neural networks (CNN) have the advantage of automating the feature extraction process, unlike ML techniques that require manual development of the feature extractor (Villaruz, 2021). In cattle research, CNN are commonly used for classification and identification in areas such as animal behavior detection and cattle disease prediction (Kamilars and Prenafeta-Boldú, 2018; Denholm et al., 2020). The use of these technologies requires data processing and validation, which in turn requires additional resources. Therefore, there is a need to demonstrate the utility of this new technology in relation to calf health programs and discuss how to interpret the expected information.

It was hypothesized that detection of respiratory disease and diarrhea in preweaning calves would be possible with a predictive model based on AMF data. Therefore, the main objective of the present study was to develop a predictive model using deep CNN for the detection of respiratory disease and diarrhea in preweaning dairy calves using AMF data [intake of milk replacer (MR), number of rewarded visits, number of unrewarded visits, and drinking speed]. In addition, the contribution of each predictor variable (behaviors) to the predictive ability of CNN models was examined separately in calves fed either ad libitum or restricted milk using permutation feature importance (PFI) analysis.

**MATERIALS AND METHODS**

Experimental data from a companion study (Frieten et al., 2017) conducted at the Educational and Research Centre for Animal Husbandry (Hofgut Neu-ühle, Germany) were used for the current study. The experimental methods were approved by the relevant Department for Animal Welfare Affairs (Landesuntersuchungsamt Rheinland-Pfalz, Koblenz, Germany, registration number 23 177-07/G 13-20-069) according to the German Animal Welfare Act (Federal Republic of Germany; Ruhdel et al., 2014).

**Calves, Management, and Diets**

This is a companion study to previous studies (Frieten et al., 2017; Gerbert et al., 2018; Koch et al., 2019) reporting on growth performance, blood metabolism, immune system, gastrointestinal tract development, and behavior of 64 German Holstein calves (32 male, 32 female) fed either ad libitum (up to 25 L/d; n = 32) or restrictively (6 L/d; n = 32) MR (12.5% solids, 21.7% CP, 18.6% crude fat, 0.2% crude fiber, and 18.3 MJ/kg of DM metabolizable energy; Trouw Nutrition Deutschland GmbH). The previous study was used to estimate power and sample size. Based on a power analysis with α = 0.05 and power = 0.80 to estimate sample size (Morris, 1999; Hintze, 2008), the expected sample size was approximately 32 calves per treatment, which provides sufficient power to detect treatment effects. Calves were randomly assigned to 2 treatments that differed in the amount of milk allowance (Frieten et al., 2017). Each calf received 10 mL of an iron suspension (115 mg Fe3+/mL, Sinta fer-o-bac, Sinta GmbH). To prevent bacterial infections, the navels were disinfected with an iodine lotion (Albrecht GmbH). All calves received a portion (2.5 ± 0.09 kg, mean ± SD) of colostrum within 2 h of birth from their dams using a calf speedy feeder (Shoof International Ltd.; Figure 1).
Depending on the feeding regimen (6 L/d vs. maximum 25 L/d), calves received acidified transition milk (2 mL acidifier/l milk, H. W. Schaumann GmbH, Pinneberg, Germany) from their dams during the next 5 meals (2.5 d). Approximately 2 to 3 h after birth, calves were moved to calf hutches (Flixbox, Mayer Maschinenbaugesellschaft mbH) and housed individually in pens bedded with straw for the first 10 ± 3 d of life (mean ± SD). An electronic scale (Sartorius AG) was used to weigh the colostrum, milk, and residues.

In individual pens, all calves were fed twice daily via a bucket with an artificial teat. Calves in the ad libitum group were offered a full teat bucket containing 7 to 8 L of liquid feed at morning feeding and the same amount at afternoon feeding. The buckets were refilled as needed around noon to ensure continuous access to liquid feed for the ad libitum-fed calves. Ad libitum-fed calves consumed 9.2 L/d ± 0.2 in wk 1 and 8.8 L/d ± 0.2 in wk 2 of life. Buckets were cleaned twice daily. After the initial period in the individual pens (10 ± 3 d), all calves were moved to an open, bipartite barn with a 30.5 m² straw bedding area with automatic feeding systems for MR and concentrate (Förster-Technik GmbH). On each side of the barn was an AMF with 2 drinking stations for calves younger than 3 wk of age and calves 4 wk of age and older to ensure that young calves had adequate access to the MR feeding station.

If calves did not drink on their own on their first day of access to the AMF, staff helped them 2 h after they registered at the AMF. The number of calves in the 2 experimental pens varied due to the continuous intake of calves throughout the experiment. However, a maximum of 15 calves were allowed in each pen at 1 time. The lower limit for intake of MR was set at 1.5 L/meal, and individual portions of MR per day were limited to 5.5 L/meal with a 30-min break, and 2 L/meal with a 2-h break after the end of the meal for ad libitum- and restricted-fed calves, respectively. The maximum amount of milk for ad libitum-fed calves was 25 L/d MR and 6 L/d for the restricted-fed calves until d 56 of life. Calves received the full amount of their feeding regimen until 56 d of age, and MR was linearly stepped down from 57 to 70 d of age in both feeding groups and they received 2 L/d MR until the end of the experiment at 77 d of age (Gerbert et al., 2017). Further details on the ingredients and nutrient composition of MR and concentrate as well as feeding management have been reported previously (Frieten et al., 2017).

When calves were fed with the AMF (d 1 of the study), data on total daily MR intake and feeding behavior (number of rewarded visits, number of unrewarded visits, and drinking speed) were collected with the PC-Program Calf Manager WIN (Förster-Technik GmbH), which was linked to the AMF. Calves had free access to hay and water, and pelleted starter feed (21.0% CP, 4.2% crude fat, 5.9% crude fiber, and 13.2 MJ/kg of DM metabolizable energy; Raiffeisen Warenzentrale Rhein-Main eG) was offered ad libitum.

**Definition of Sick Calves**

Health status of all calves was checked 2 times per day and health scores were recorded by 2 veterinarians using a modified version of the University of Wisconsin calf health scoring system (McGuirk, 2008; McGuirk and Peek, 2014; School of Veterinary Medicine, 2017). All diseases, including respiratory disease, were evaluated and treated by well-trained veterinarians. No formal inter-rater reliability tests were performed in this study. Feces (score 1 = well-formed; score 2 = pasty, but formed; score 3 = smooth, but persist on bedding; score 4 = watery, runs through bedding) were scored daily (McGuirk, 2008). For the respiratory disease examination, the complete respiratory health assessment included clinical examination of ocular discharge, nasal discharge, cough, ear tilt, upper and lower respiratory tract auscultation, and rectal temperature, which were scored daily from 0 to 3, where 0 = normal, 1 = mildly abnormal, 2 = moderately abnormal, and 3 = severely abnormal (McGuirk and Peek, 2014). A calf was considered sick if its fecal score was 3 or 4 and its respiratory score was 2 or 3.

For illnesses, severity, duration, and treatment were documented. Calves with illnesses were monitored and, if necessary, treated immediately by veterinarians. In the cases of illness, calves were treated immediately by a veterinarian. For diarrhea (fecal score = 3 or 4), calves received an oral electrolyte replacement (Bewilyt Elektrolyttränke, Bewital GmbH & Co. KG) and essence of spruce (Stullmisan vet. Pulver, MSD Tiergesundheit, Intervet Deutschland GmbH) until the fecal score improved to 2 or 1. Due to an increase in respiratory disease in the fall of 2014 at the Educational and Research Center for Animal Husbandry (Hofgut Neumühle), calves were vaccinated against bovine respiratory syncytial virus and parainfluenza type-3 virus (Rispoval RS + PI3 IntraNasal, Zoetis Deutschland GmbH) from December 2014 to March 2015. For respiratory disease, antibiotics were provided in severe cases of respiratory diseases, and calves were treated with a combination product containing an antibiotic (florfenicol) and a nonsteroidal anti-inflammatory drug (flunixin meglumine; Resflor, Intervet Deutschland GmbH). In some cases, a mucolytic agent (Bisolvon, Boehringer Ingelheim Vetmedica GmbH) was added. Treatment of respiratory disease was based on the severity and clinical signs of the individual animal.
Figure 1. Schematic representation of the experimental plan. Sixty-four German Holstein calves (32 male, 32 female) were individually housed in straw-bedded pens for the first 10 ± 3 d (mean ± SD) of life. Calves were fed milk replacer (MR) either ad libitum (up to 25 L/d) or restrictively (6 L/d) via automated milk feeders. The automatic milk feeding device measured MR intake, MR allowance, the number of rewarded visits, unrewarded visits, and drinking speed in group-housed calves. Calf illness was defined as respiratory disease (respiratory score ≥ 2) and diarrhea (fecal score ≥ 3). A deep convolutional neural network architecture for the detection of respiratory disease and diarrhea in dairy calves was developed. Data were split into 80:20 ratios for training and testing data sets according to the Pareto principle (Zhu and Boiarskaia, 2012). The basic trial as previously described by Frieten et al. (2017).
Data Processing, Model Development, and Model Evaluation

Only data from AMF up to 47 d of age were included in the analysis. This cut in the data was made to avoid data from the weaning period. The data set was defined as an array \((N, Q, M)\), where \(N\) is the number of samples in the data set \((N = 1,844)\), \(Q\) is the maximum number of time steps of all variables \((Q = 5)\), and \(M\) is the number of variables processed per time step \((M = 5)\). In both groups (restricted or ad libitum), the CNN models included all AMF data, including total daily MR intake, the number of rewarded visits, the number of unrewarded visits, and drinking speed. Milk allowance was not included in the CNN models for restricted and ad libitum-fed calves. The CNN models were based on the convolutional block architecture. Each block consists of a convolutional layer, followed by a batch normalization layer, followed by an activation layer equipped with a rectified linear unit activation function. In total, the network consisted of 3 stacked blocks with filter sizes of 128, 256, and 128, respectively. After the last block, an average pooling layer was applied before passing the tensor to the final softmax layer. The softmax layer has 2 neurons and outputs the probability for each label (healthy or sick). It was detected as a healthy calf if the output value of the first neuron was greater than the output value of the second neuron; otherwise, it was detected as a sick calf. A minimum sensitivity of 80% was considered an appropriate requirement for the prediction models. Data were split according to the Pareto principle (Zhu and Boiarskaia, 2012) in a ratio of 80:20 for training and testing data sets. Rather than using samples as the validation set, the standard 5-fold cross-validation method was used to ensure that the test set was large enough to refine the models. The cross-validation procedure was repeated 5 times and the results were averaged to produce a single estimate (Jung and Hu, 2015). A total of 1,844 records [207 sick days (ad libitum-fed calves = 121 d, sick event = 36; restricted-fed calves = 86 d, sick event = 31) and 1,637 healthy days] were used for the prediction model. The training set consisted of 1,483 records and the test set consisted of 361 records (46 sick days and 315 healthy days). The approximate value of the area under the receiver operating characteristic (ROC) curve (AUC) was calculated using a Riemann sum.

A class imbalance was resolved by rejection sampling, or as it is commonly called, the acceptance-rejection method, to generate examples of each class with equal probability (Pei et al., 2020). The CNN model was trained for 300 epochs (training cycles) using the Keras and TensorFlow deep learning libraries via the Adam optimizer with a learning rate of 0.004. In a binary classification model, a cross-entropy loss is often used to calculate the loss between the predicted value and the true value of the model. Therefore, this method was applied to quantify the difference between predicted and true values. Minimizing the cross-entropy loss \(L\)-value directly leads to more accurate predictions and better detection performance. For all feature data \((y \in \{0, 1\})\), the maximum likelihood function was calculated, and then the logarithm was averaged to obtain the cross-entropy loss. The following equation shows the expression of a cross-entropy loss function:

\[
\text{Loss} : L = \frac{1}{N} \left\{ \sum_{i}^{N} \left[ y^{i} \log y^{i} + (1 - y^{i}) \log (1 - y^{i}) \right] \right\},
\]

where \(N\) represents the number of samples, \(y^{i}\) represents the true label of the ith sample, and \(Y^{i}\) represents the probability value that the sample is predicted to be 1.

Diagnostic Test Characteristics and PFI

A PFI algorithm with 1,000 iterations was implemented using the feature importance function from the scikit-learn library (version 0.24.2; Pedregosa et al., 2011) and Python software (version 3.6), and was visualized as bar plots using the Seaborn Python packages (https://seaborn.pydata.org/). Descriptive statistics were performed for MR intake, drinking speed, rewarded visits, and unrewarded visits using JASP version 0.14 (JASP Team, 2019).

RESULTS

Figure 1 shows the schematic diagram of the experimental design.

Descriptive Statistics

Descriptive statistics, including mean, standard deviation, minimum, range, and maximum for MR intake, drinking speed, rewarded visits, and unrewarded visits of dairy calves on restricted and ad libitum MR feedings (from d 10 ± 3 to d 56 of life) are presented in Table 1. We found large differences in feeding behavior between feeding regimens. In the restricted-fed calves, MR intake was 5.88 ± 0.4 L/d (minimum = 2.1, maximum = 6 L/d) and in the ad libitum-fed calves, it was 11.95 ± 3.3 L/d (minimum = 2.6, maximum = 24 L/d) (Figure 2A). The number of rewarded visits in the restricted-fed calves was 5.75 ± 2.6 (minimum = 2, maximum = 24), whereas the number of rewarded visits in the ad libitum-fed calves was 6.0 ± 2.7 (minimum = 1, maximum = 21) (Figure 2B). In restricted-fed calves, the number of unrewarded visits was 16.9 ± 0.4...
and in ad libitum-fed calves, the number of unrewarded visits was 0.05 ± 0.33 (minimum = 0, maximum = 4) (Figure 2C). Drinking speed in calves fed restricted diets was 640 ± 286 mL/min (minimum = 12, maximum = 998 mL/min), whereas drinking speed in calves ad libitum was 549 ± 157 (minimum = 14, maximum = 860 mL/min; Figure 2D). Throughout the experiment, calves were fed their assigned amount of MR until weaning.

The percentage of sick days in all calves and calves fed restricted or ad libitum MR is shown in Figure 3. The percentage of sick days was higher during the first 2 wk of AMF. The criteria for defining a disease event was that calves had to be disease-free for 2 d before the data were considered a new disease event. We had a total of 139 d with diarrhea (49 events, 31 calves) and 75 d with respiratory disease (21 events, 13 calves). Three calves had both diseases.

### CNN Model Performance and PFI

The characteristics of the diagnostic test, including true positive, true negative, false negative (FN), false positive (FP), the AUC of the ROC curve, the cut-off point, prevalence, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are defined in Table 2. Figure 4 shows the training and validation processes of the deep CNN based on AMF data for the disease prediction task. The cross-entropy loss (Figure 4A) and the AUC of the ROC curves per epoch (Figure 4B) for training and validation were plotted against 300 training iterations. From the cross-entropy loss curve, the validation error reaches a minimum of 70 iterations around the training epochs. Figure 5 shows the confusion matrix representing the performance of the CNN classifiers trained to identify calf disease (healthy vs. sick) using data from AMF in the test data sets on all calves in group housing (Figure 5A), and calves fed ad libitum (Figure 5B) or restricted (Figure 5C) MR. Considering all calves in group housing, cross-validation of the test data set yielded a sensitivity of 83% and a specificity of 79%, with a PPV and NPV of 37 and 97%, respectively (Table 3). A calf day was considered true positive if there was a positive classification on at least 1 of the 5 d preceding the disease, and FN if no disease was observed on any of these 5 d.

For ad libitum-fed calves, cross-validation of the test data set yielded a sensitivity of 83% and a specificity of 71%, with a PPV and NPV of 39 and 95%, respectively (Table 3). For restricted-fed calves, cross-validation of the test data set achieved a sensitivity of 82% and a specificity of 87%, with a PPV and NPV of 31 and 99%, respectively (Table 3). Estimates of the ROC curve for the deep CNN model yielded an AUC of 0.81, 0.77, and 0.84 for all group-housed, ad libitum-fed, and restricted-fed calves, respectively (Table 3).

Estimates of the ROC curve for the deep CNN model yielded an AUC of 0.81, 0.77, and 0.84 for all group-housed, ad libitum-fed, and restricted-fed calves, respectively (Table 3).

To evaluate the degree of contribution of features (feeding behavior) in predicting calf disease, the permutation function was applied to the deep CNN model. The features (feeding behaviors) in descending order of importance are summarized in Figure 6 for restricted (Figure 6A) or ad libitum-fed calves (Figure 6B). For ad libitum-fed calves, drinking speed and MR intake at the feeder were the most important features for predicting calf disease. For restricted-fed calves, the number of unrewarded visits to the milk feeder and MR intake were the most important features for predicting calf sickness.

### DISCUSSION

Daily feeding behavior data recorded by AMF have shown promise in identifying calves at high risk for disease (Knauer et al., 2017; Cramer et al., 2020). In a retrospective case-control study, sick calves (treated for either diarrhea or respiratory disease) were found to change many aspects of their feeding behavior collected from AMF before illness (Morrison et al., 2022). The objective of this study was to develop a predictive model for detecting respiratory disease and diarrhea in the preweaning period. To this end, a deep CNN architecture was developed for detecting sick calves using AMF data. Convolutional neural networks are

---

**Table 1.** Descriptive statistics for milk replacer (MR) intake, drinking speed, rewarded visits, and unrewarded visits of dairy calves

| Item                  | MR intake (L/d) | Drinking speed (mL/min) | Rewarded visits (n) | Unrewarded visits (n) |
|-----------------------|----------------|------------------------|---------------------|-----------------------|
|                       | Restricted     | Ad libitum             | Restricted          | Ad libitum            | Restricted          | Ad libitum           |
| Mean                  | 5.88           | 11.95                  | 640                 | 549                   | 5.75               | 5.96                 | 16.9                 | 0.05                 |
| SD                    | 0.37           | 3.28                   | 286                 | 157                   | 2.57               | 2.71                 | 10.0                 | 0.33                 |
| Minimum               | 2.10           | 2.60                   | 12.0                | 14.0                  | 2.0                | 1.0                  | 0.00                 | 0.00                 |
| Range                 | 3.9            | 21.4                   | 986                 | 846                   | 22.0               | 20.0                 | 60.0                 | 4.0                  |
| Maximum               | 6.0            | 24.0                   | 998                 | 860                   | 24.0               | 21.0                 | 60.0                 | 4.0                  |

1Dairy calves were fed restricted or ad libitum MR from d 10 ± 3 until d 56 of life (d 0 to 47 at automated milk feeders).
widely used among deep learning models for various pattern recognition tasks such as image classification, biomimetic pattern recognition, and action recognition (Zhou et al., 2017; Chen et al., 2018; Obeso et al., 2022). Convolutional neural networks are superior to other methods ML because they recognize important features automatically without human intervention (Villaruz, 2021). Moreover, CNN uses backpropagation to update older assumptions with newly acquired knowledge during training (Cireșan et al., 2010). These networks are essential for extracting high-level features from abstract data to improve the predictability of deep classification layers (Cireșan et al., 2010).

In the current study, the predictive model was trained on feeding behavior data, including milk allowance, drinking speed, the number of rewarded and unrewarded visits, and was optimized through cross-validation. The PFI was applied to the deep CNN model to evaluate the contribution of features (feeding behavior) in predicting calf disease. In restricted-fed calves, the number of unrewarded visits to the milk feeder was one of the most important features in predicting calf disease in the current study. In agreement with our findings, previous studies found that the number of unrewarded visits to the milk feeders was a better indicator of illness than the number of rewarded visits in dairy calves.
suggesting that sick calves are still motivated enough to visit the milk feeder, but no more than necessary to consume their milk allowance (Knauer et al., 2017; Sutherland et al., 2018). The number of unrewarded visits to the milk feeder is considered a hunger-related behavior (De Paula Vieira et al., 2008) and could also be an indicator of a decrease in appetite (Lowe et al., 2019) in dairy calves. In contrast, Swartz et al. (2017) found no difference in total visits and rewarded visits of milk consumed in sick calves compared with their healthy controls.

According to PFI results, the MR intake was the most important factor for predicting calf disease in both ad libitum- and restricted-fed calves. Consistent with our findings, Sutherland et al. (2018) reported that group-housed calves fed 6 L of milk per day had lower milk intake in the 4 d before diarrhea occurred. In addition, Borderas et al. (2009) found that group-housed calves

---

**Table 2.** Definition and calculation of diagnostic test characteristics

| Item | Definition and calculation |
|------|-----------------------------|
| AUC  | Area under the curve        |
| ROC curve | Receiver operating characteristic curve |
| TP   | Sick calves correctly identified as sick |
| TN   | Healthy calves correctly identified as healthy |
| FN   | Healthy calves incorrectly identified as sick |
| FP   | Sick calves incorrectly identified as healthy |
| Cut-off point | Sensitivity by 0.8 |
| Prevalence | Total number with disease/population at risk for the disease |
| Sensitivity | The proportion of the correctly identified sick calves, out of all sick calves |
| Specificity | The proportion of the correctly identified healthy calves, out of all healthy calves |
| PPV  | The proportion of sick calves on the alarm list (all calves predicted sick) |
| NPV  | The proportion of healthy calves without alarms (all calves predicted healthy) |

$\text{TP} = \text{true positive}; \text{TN} = \text{true negative}; \text{FN} = \text{false negative}; \text{FP} = \text{false positive}; \text{PPV} = \text{positive predictive value}; \text{NPV} = \text{negative predictive value}.$

---

**Figure 3.** Percentage of sick days [defined as respiratory disease (respiratory score ≥2) and diarrhea (fecal score ≥3)] in all calves and calves fed restricted or ad libitum milk replacer as determined by the Calf Health Scoring System (McGuirk and Peek, 2014; School of Veterinary Medicine, 2017).
fed high milk (12 L/d) or ad libitum had lower milk intake; fewer visits, and longer duration of visits to the milk feeder on the day before illness, whereas calves fed low milk (4 L/d) had a shorter duration of visits to the milk feeder on the day of illness diagnosed as gastrointestinal disease, respiratory disease, or a combination of both. Therefore, behavioral changes such as a decrease in milk intake contribute significantly to the early detection of calf disease.

A decrease in drinking speed may also serve as an indicator of disease in dairy calves (Maatje et al., 1993). According to PFI results, the drinking speed was the most important factor for predicting calf disease in ad libitum-fed calves. In agreement with our findings, Knauer et al. (2017) found that calves with signs of disease (defined as an intestinal disease and respiratory disease) had a slower drinking speed per meal compared with healthy calves. This result is consistent with a previous study showing that drinking speed was slower in sick calves compared with healthy calves (Johnston et al., 2016). More recently, Cramer et al. (2020) found that calves with clinical respiratory disease had slower drinking speeds than calves with subclinical respiratory disease and healthy calves. However, Swartz et al. (2017) found no difference in drinking speed when comparing calves with respiratory disease and healthy control calves. The discrepant results can be attributed to the lack of proper validation of the measurement method (Morrison et al., 2021), or variation in nipple diameter and tube length between individual AMF (Jensen and Holm, 2003) and sample size (many individual variations), disease definition, disease severity, and calf age among studies.

In the present study, instead of using the Youden index to optimize cut-off points, the model was adjusted to achieve a sensitivity of at least 80% to reduce the risk of FN results in disease identification. It was found that determining the optimal cut-off point using the Youden index to maximize the sensitivity and specificity of blood metabolism indicators measured on arrival at the farm can lead to poor results in predicting future disease risk in calves (Goetz et al., 2021). In the present study, the AUC-ROC results showed good performance for the CNN model in all group-housed calves, indicating that the CNN models can predict calf disease in both groups with different MR allowances. In line with our findings, a cross-sectional study by Knauer et al. (2018) found that the combination of feeding behavior data (including drinking speed, milk consumption, and the number of unrewarded visits to the AMF) were useful indicators for predicting sick calves in the preweaning period, with a model performance of 75% sensitivity (95% CI: 65.5–82.6), specificity of 27% (95% CI: 21.7–33.2), PPV of 65% (95% CI: 52.5–75.1), and NPV of 37% (95% CI: 27.1–48.9).

Although sensitivity and specificity are prevalence-independent test properties and do not depend on disease prevalence in the population of interest (Leeflang...
et al., 2013), the predictive values (PPV and NPV) of a model or diagnostic test could be affected by disease prevalence in the population tested. When disease prevalence (e.g., respiratory disease and diarrhea) increases, PPV also increases, but NPV decreases (and vice versa). In the current study, a NPV was defined as the proportion of calves without signals that were correctly identified as healthy calves. The NPV was 97% for all group-housed calves, which means that the FN rate was less than 3% under the conditions of this study. Nevertheless, the PPV in this study was estimated to be 0.37 (corresponding to a margin of error of 0.63) in all calves housed in groups, which means that only 37% of calves with a positive alert were truly sick, whereas in the present study, 63% of calves without signs of disease (healthy) were identified as positive sick calves. In the current study, sick days ranged from 5% to between 1 to 2% (first 2 wk vs. 3–8 wk with AMF) and fluctuated during the trial and decreased with calf age. The prevalence of diarrhea and respiratory disease in young calves varies among studies. In group-housed dairy calves, diarrhea and respiratory disease were prevalent in 22 and 26% of the calves within 7 d, respectively (Cramer et al., 2016). Knauer et al. (2017) found that 31% of calves housed in the preweaning groups suffered from diarrhea and 12% from respiratory disease throughout the study. In the Conboy et al. (2021) study, diarrhea and respiratory disease were prevalent in group-housed dairy calves at 22 and 25%, respectively, throughout the study.

A critical point to note is that disease prevalence in several studies were reported as the frequency of disease per calf throughout the study period rather than the daily prevalence of disease, which is important for PPV calculation. The lower the prevalence of a disease per day, the lower the PPV or the higher the rate of FP. According to Berman et al. (2019), the PPV for diagnosing respiratory disease increased from 3 to 8% when the prevalence of active pneumonia in weaned calves increased from 2.5 to 20.0%. However, the direct relationship between the PPV and the frequency of occurrence of the predicted event (disease prevalence) or classification is rarely illuminated in dairy cattle studies (Post et al., 2021). Therefore, the implications of prediction models with high sensitivity and specificity, but low PPV, would be a high proportion of FP alarms (also called error rates), leading to an increase in the daily workload of farmers and probably even unnecessary treatments, which is not compatible with the general aspects of precision livestock farming.

Some limitations should be considered when interpreting the results of the current study. First, the sample size was relatively small with a total of 64 calves, so the results may not be generalizable to the

![Figure 5. Confusion matrix depicting the performance of deep convolutional neural networks classifiers trained to identify calf illness (healthy vs. sick) by use of data from automated milk feeders in the test data sets in (A) all group-housed calves, and in calves fed (B) ad libitum or (C) restricted milk replacer.](image)
The results of the current study cannot be extrapolated to the general population, and future studies with larger sample sizes are needed to validate in different dairy farms and discuss these preliminary insights. Second, calves suffering from respiratory disease and diarrhea may show different behavioral changes during their condition. Unfortunately, the current study did not provide sufficient data to evaluate these differences, and further studies are needed to determine the relevance of these differences. However, using AMF data to predict a particular disease, such as respira-

| Item                                | Sensitivity (%) | Specificity (%) | Accuracy (%) | PPV1 (%) | NPV2 (%) | AUC-ROC curve3 |
|-------------------------------------|-----------------|-----------------|--------------|----------|----------|----------------|
| All group-housed calves             | 82.6            | 79.0            | 80.0         | 36.5     | 97.0     | 0.81           |
| Ad libitum-fed calves               | 83.0            | 71.4            | 73.0         | 38.7     | 95.0     | 0.77           |
| Restricted-fed calves               | 81.8            | 87.0            | 87.0         | 31.0     | 99.0     | 0.84           |

1PPV = positive predictive value.
2NPV = negative predictive value.
3AUC-ROC curve = area under the receiver operating characteristic (ROC) curve.

A
Restricted-fed calves

Unrewarded visits
Milk replacer intake
Drinking speed
Rewarded visits

B
Ad libitum-fed calves

Drinking speed
Milk replacer intake
Unrewarded visits
Rewarded visits

Figure 6. Permutation feature importance scores with deep convolutional neural network in the (A) restricted- or (B) ad libitum-fed calves. The permutation feature importance was measured by the decrease in model accuracy, and the features (behaviors) were summarized in descending order of their relative importance to the convolutional neural networks model.
tory disease or diarrheal disease, results in a lower PPV because the prevalence of a particular disease on a daily basis is lower than what would be obtained by combining the 2 diseases.

Farmers are faced with many decision-making situations on a “daily basis,” and eventually predictive models should lead farmers to decide which calf to examine (and treat accordingly based on the details of the examination) on a daily basis. Using AMF data alone as the sole source of information (behavioral data) for calf disease detection in a low daily disease prevalence setting remains a challenge. wearable accelerometers for preweanling calves could measure differences in walking behavior to identify calves at risk for disease (diagnosed as respiratory, digestive, or rectal temperature >39.5°C; Belaid et al., 2020). Further data on other behavioral factors, such as visits to the feed (starter and grass hay) bunk, steps, and lying time (Belaid et al., 2020) may also be relevant to explain some of the FP results. According to Lowe et al. (2019), milk consumption, body temperature at the side and shoulder, duration of time spent at the water trough, and number and duration of lying bouts are related to the days relative to the clinical identification of diarrhea in newborn calves. It is important to consider how specific or non-specific the association of the new sensor data is with calf disease. The use of non-specific indicators for a particular disease or symptom is likely to have a negative effect on the accuracy of the predictive value and could lead to a high number of FP results (Stachowicz and Umstätter, 2021). Nevertheless, these arguments tend to raise the question of how a larger data set should be compiled if it reflects the real situation (low daily prevalence of disease) on the dairy farm, as shown in this study. In addition, the use of multiple automated monitoring technologies should be affordable, and data management systems should be able to integrate, analyze, and interpret data from sensors (Stachowicz and Umstätter, 2021).

To ensure the future success of predictive models in practical settings, reducing FP alarms before model implementation is critical, and the reasons for these FP alarms should also be clarified for future analyses. Furthermore, the predictive values of a model estimated in a particular setting (e.g., a dairy farm) cannot be unconditionally extrapolated to other settings with a different disease prevalence (Leeflang et al., 2013; Post et al., 2021).

CONCLUSIONS

Despite the relatively small sample size, the results provide strong evidence that daily feeding behavior data from AMF can be used to identify calves at risk for disease. Despite a very good testing property of the model, the relatively low prevalence of calf disease on a daily basis in the present study resulted in a high proportion of FP alarms. Based on the experimental conditions used in this study, deep CNN models may prove useful when monitoring young calves in group housing systems, whether they are fed restricted or ad libitum. Disease detection using feeding data from an AMF system remains a challenge because of the high proportion of FP when matched to the real data distribution in low prevalence environments.

ACKNOWLEDGMENTS

We thank the staff of the Educational and Research Centre for Animal Husbandry (Hofgut Neumühle, Alsenz, Germany) and the students and team of Animal Nutrition at the University of Applied Sciences Bingen (Bingen am Rhein, Germany). The authors have not stated any conflicts of interest.

REFERENCES

Abutarbush, S. M., C. M. Pollock, B. K. Wildman, T. Perrett, O. C. Schunicht, R. K. Fenton, S. J. Hannon, A. R. Vogstad, G. K. Jin, and C. W. Booker. 2012. Evaluation of the diagnostic and prognostic utility of ultrasonography at first diagnosis of presumptive bovine respiratory disease. Can. J. Vet. Res. 76:23–32.

Adams, E. A., and S. Buczinski. 2016. Short communication: Ultrasound exam for diagnosis of bovine respiratory disease. Can. J. Vet. Res. 76:1874–1883. https://doi.org/10.1162/NECO_a_00052.

Berman, J., D. Francoz, S. Dufour, and S. Buczinski. 2019. Bayesian estimation of sensitivity and specificity of systematic thoracic ultrasound exam for diagnosis of bovine respiratory disease in pre-weaned calves. Prev. Vet. Med. 162:38–45. https://doi.org/10.1016/j.prevetmed.2018.10.025.

Borderas, T. F., J. Ruschen, M. A. von Keyserlingk, and A. M. de Passille, 2009. Automated measurement of changes in feeding behavior of milk-fed calves associated with illness. J. Dairy Sci. 92:4549–4554. https://doi.org/10.3168/jds.2009-2109.

Chen, X., J. Weng, W. Lu, J. Xu, and J. Weng. 2018. Deep manifold learning combined with convolutional neural networks for action recognition. IEEE Trans. Neural Netw. Learn. Syst. 29:3938–3952. https://doi.org/10.1109/TNNLS.2017.2740518.

Cireşan, D. C., U. Meier, L. M. Gambardella, and J. Schmidhuber. 2010. Deep, big, simple neural nets for handwritten digit recognition. Neural Comput. 22:3207–3220. https://doi.org/10.1162/NECO_a_00052.

Consolvo, M. H., C. B. Winder, C. Medrano-Galarza, S. J. LaBlanc, D. B. Haley, J. H. C. Costa, M. A. Steele, and D. L. Renaud, 2021. Associations between feeding behaviors collected from an automated milk feeder and disease in group-housed dairy calves in Ontario: A cross-sectional study. J. Dairy Sci. 104:10183–10193. https://doi.org/10.3168/jds.2021-20137.

Costa, J. H. C., M. C. Cantor, and H. W. Neave. 2021. Symposium review: Precision technologies for dairy calves and management applications. J. Dairy Sci. 104:1203–1219. https://doi.org/10.3168/jds.2019-17885.
Post, C., C. Rietz, W. Buscher, and U. Muller. 2021. The importance of low daily risk for the prediction of treatment events of individual dairy cows with sensor systems. Sensors (Basel) 21:1389–1410. https://doi.org/10.3390/s21041389.

Renauld, D. L., T. F. Duffield, S. J. LeBlanc, S. Ferguson, D. B. Haley, and D. F. Kelton. 2018. Risk factors associated with mortality at a milk-fed veal calf facility: A prospective cohort study. J. Dairy Sci. 101:2659–2668. https://doi.org/10.3390/jds.2017-13581.

Renaud, D. L., T. F. Duffield, S. J. LeBlanc, S. Ferguson, D. B. Haley, and D. F. Kelton. 2018. Risk factors associated with mortality at a milk-fed veal calf facility: A prospective cohort study. J. Dairy Sci. 101:2659–2668. https://doi.org/10.3390/jds.2017-13581.

Ruhdel, I., C. Maisack, and K. Wagner. 2014. German animal welfare act in breach with Directive 2010/63/EU. ALTEX 31:219–222. https://doi.org/10.14573/altex.1404011.

School of Veterinary Medicine. 2017. University of Wisconsin-Madison, Calf health scoring chart. Accessed Oct. 12, 2022. https://www.vetmed.wisc.edu/fapm/svm-dairy-apps/calf-health-scorer-chs/.

Stachowicz, J., and C. Umstätter. 2021. Do we automatically detect health- or general welfare-related issues? A framework. Proc. Biol. Sci. 288:20210190. https://doi.org/10.1098/rspb.2021.0190.

Swartz, T. H., A. N. Findlay, and C. S. Petersson-Wolfe. 2017. Short communication: Automated detection of behavioral changes from respiratory disease in pre-weaned calves. J. Dairy Sci. 100:9273–9278. https://doi.org/10.3390/jds.2016-12280.

Teixeira, A. G. V., J. A. A. McArt, and B. C. Bicalho. 2017. Thoracic ultrasound assessment of lung consolidation at weaning in Holstein dairy heifers: Reproductive performance and survival. J. Dairy Sci. 100:2985–2991. https://doi.org/10.3390/jds.2016-12016.

Trefz, F. M., I. Lorenz, A. Lorch, and P. D. Constable. 2017. Clinical signs, profound acidemia, hypoglycemia, and hypernatremia are predictive of mortality in 1,400 critically ill neonatal calves with diarrhea. PLoS One 12:e0182938. https://doi.org/10.1371/journal.pone.0182938.

Uetake, K. 2013. Newborn calf welfare: A review focusing on mortality rates. Anim. Sci. J. 84:101–105. https://doi.org/10.1111/asj.12019.

Villaruz, J. A. 2021. Deep convolutional neural network feature extraction for berry trees classification. J. Adv. Inf. Technol. 12:226–233. https://doi.org/10.12720/jait.12.3.226-233.

Virtala, A. M. K., G. D. Mechor, Y. T. Grohn, and H. N. Erb. 1996. The effect of calfhood diseases on growth of female dairy calves during the first 3 months of life in New York State. J. Dairy Sci. 79:1040–1049. https://doi.org/10.3168/jds.S0022-0302(96)76457-3.

Windeyer, M. C., K. E. Leslie, S. M. Godden, D. C. Hodgins, and S. J. LeBlanc. 2014. Factors associated with morbidity, mortality, and growth of dairy heifer calves up to 3 months of age. Prev. Vet. Med. 113:231–240. https://doi.org/10.1016/j.prevetmed.2013.10.019.

Zhou, L., Q. Li, G. Huo, and Y. Zhou. 2017. Image classification using biomimetic pattern recognition with convolutional neural networks features. Comput. Intell. Neurosci. 2017:3792805. https://doi.org/10.1155/2017/3792805.

Zhu, W. M., and E. Boiarskaia. 2012. A Pareto principle (20–80 Rule) analysis of daily physical activity and diabetes impact. Med. Sci. Sports Exerc. 44:497.