Master thesis

Automatic Cattle Lameness Assessment with Markerless Cattle Pose Estimation

by

Yu Yang Quek

4729420

ME-BMD MSc Project (ME51032)

Master Mechanical Engineering

Delft University of Technology

To be defended on

28th November 2019

Thesis Committee: Prof. dr. ir. J. Harlaar
Dr. Y. Li
J. E. Geelen
Dr. D. M. J. Tax

Contents

Abstract

Lameness is characterized by abnormal gait and is an indicator of various hoof diseases in cattle. Not only does this raise animal welfare issues, it also causes significant economic loss from reduced milk yield and fertility. Despite that, prevalence of lameness in dairy farms is high because farmers are unable to dedicate time and labor to identify lame cattle. Many automatic lameness detection solutions have been proposed in literature. Machine vision solutions using cameras are especially attractive because cameras do not require much space and is relatively low cost. However, none of the machine vision solutions so far have been robust enough to be useful to dairy farmers. This thesis attempts to remedy that by applying deep learning methods to the pose estimation of cattle to analyze their gait and detect the presence of lameness.

313 videos of cattle walking were recorded at a dairy farm. Images were randomly extracted from those videos and 17 body parts were manually annotated on the images to fine-tune a deep neural network pretrained on ImageNet. The fine-tuned network is then used to automatically find the trajectories of the 17 body parts in all 313 videos. 84 gait features were extracted from each video based on these trajectories. Each video was also manually given a locomotion score between 1-5 by 2 experts. Due to the small number of locomotion score 5 cows, locomotion score 4 and 5 were merged into one group for analysis. Two experiments were conducted with these gait features and locomotion scores: a) Data analysis to test significant differences between locomotion score groups and b) Automatic locomotion score classification.

Data analysis was done using ANOVA followed by Bonferroni correction. Stance time related features were the best at differentiating locomotion score groups, but were unable to differentiate between locomotion score pair 2<>3. Step length related features were also relatively good at differentiating different locomotion score groups, but have trouble differentiating between locomotion score pairs 1<>2 and 2<>3.

For the automatic classification, the 84 gait features were first reduced to 3 features using LDA. Then, various classifiers were trained with these 3 features and locomotion score as labels. The linear discriminant classifier achieved the highest classification rate at 85.6%, but this number was heavily skewed by the high classification rate of locomotion score 1 group (95.3%), which also makes up the largest portion of the dataset. Classification rates of locomotion score 2 and 3 were much lower at 54.2% and 59.6%. Another way to look at this result is by viewing the locomotion score 1 cows as healthy cows and the rest as lame cows. The sensitivity of the classifier is then 61.3% and specificity is 95.3%. To correct this imbalance in the dataset, the prior probabilities were set equal for each locomotion score group and the classifiers were trained again. This resulted in much better classification rates with the linear discriminant classifier for locomotion score 2 and 3 (71% and 84% respectively) at the expense of lowering the classification rate of locomotion score 1 (83.4%). In terms of sensitivity and specificity, 78.3% and 83.4% was achieved respectively. Notably, locomotion score 4 (and 5) cows were classified with 100% accuracy.

The results showed that the proposed machine vision method can detect mild lameness relatively well and can be used as an early warning system to single out possibly mildly lame cows for inspection. If severely lame cows are on the farm, they can be detected with very high certainty. To be even more useful to farmers, improvements should be made in the sensitivity and specificity of the classifier.

1. Introduction

In the dairy industry, cattle health is extremely important for optimizing milk productivity and elongating their productive lifespan. The three biggest cattle health problems in the dairy industry are mastitis, infertility and lameness. Cattle lameness is the third largest factor of economic loss in the dairy industry [1] and it is an indicator of various hoof diseases in cattle, which manifests as abnormal gait. Lameness prevalence varies a lot from farm to farm, with various studies estimating a high mean lameness prevalence of 19-37% [2-6]. This has significant economic consequences, averaging at a cost of $4899 per year for a herd of 65 cows [7], and adversely impacts animal welfare [8]. One of the factors contributing to the prevalence of cattle lameness is the limited time and labor that the farmers can allocate to tackle the lameness problem [9]. Notably, the effective detection of lameness is particularly time and labor intensive because of the need to regularly inspect each individual cattle in a herd of dozens or hundreds of cattle. Various automatic solutions have been proposed in literature with the goal of saving time and labor for the farmers, but most of these solutions remain in the research phase and not yet commercialized, presumably because they are not yet commercially feasible.

1.1 Skeletal Anatomy of Cattle

An animal's gait is normally thought of as the absolute and relative movement of the bones of its limbs rather than the entire limbs including the muscles and skin. Thus, a basic understanding of the bones of cattle limbs is necessary for further discussion of gait analysis.

Figure 1 shows the names and locations of the bones of cattle limbs. In the order of increasingly distal limb bones, the thoracic limbs (fore limbs) are mainly composed of the scapula, humerus, ulna, radius, metacarpus and phalanges. The pelvic limbs (hind limbs) are mainly composed of the femur, tibia, fibula, metatarsus and phalanges.

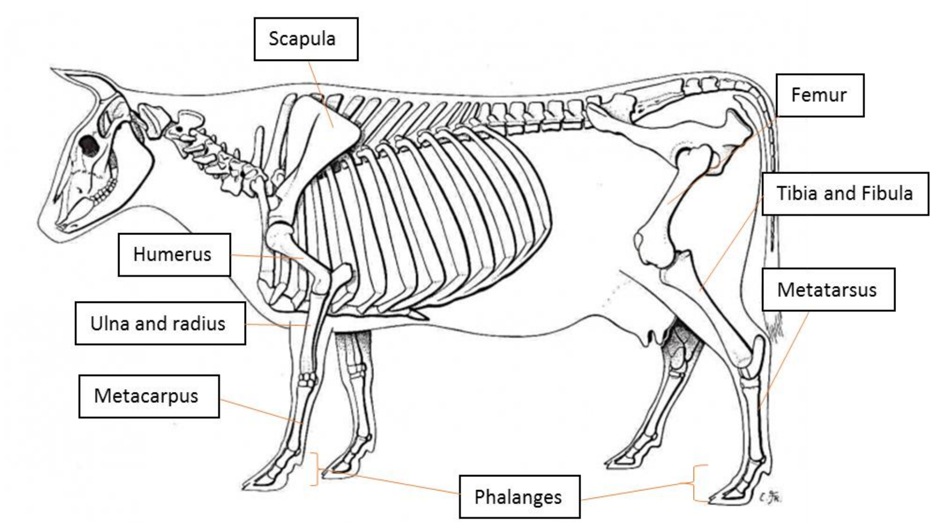

Figure 1 – Illustration of cattle skeleton with limb bones labeled. Image (without labels) is taken from https://www.purposegames.com/game/cattle-skeletal-system-quiz on 9/4/2019.

The phalanges are mainly composed of the proximal phalanx, middle phalanx and distal phalanx as shown in Figure 2. The bottom of the foot is protected by a hoof which covers the distal phalanx.

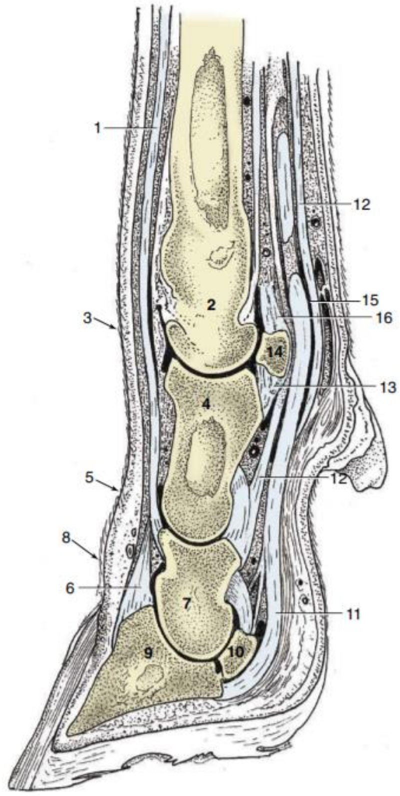

Figure 2 – Anatomy of distal end of cattle limb. 2, metacarpus/metatarsus. 4, proximal phalanx. 7, middle phalanx. 9, distal phalanx. Image taken from [10].

1.2 Cattle Gait

The normal walk cycle of each leg of the cattle goes through a stance phase and a swing phase [11, 12]. The stance phase is when the foot is in contact with the ground to support the movement of other legs. It starts when the foot first contacts the ground with a heel strike and ends when it leaves the ground. Midstance is defined as the moment when the metacarpus is vertical for the front limbs. For the hind limbs, midstance is defined as the moment when the hip joint, which is the joint at the proximal head of the femur, is vertical to the hoof. The swing phase starts when the foot leaves the ground and lasts until it contacts the ground again.

The movement of each leg during a complete gait cycle of cattle is illustrated in Figure 3 [13]. The horizontal bars represent the temporal distribution of the stance phase and swing phase of each leg, where the darker areas represent the stance phase and the white areas represent the swing phase. It is interesting to note that healthy cattle will try to put the hind feet where the front feet had stepped before during normal walking. Thus, cattle will start the swing phase of the front legs right before the hind legs steps in as shown in Figure 4. This can also be seen in Figure 3 where the left front (LF) and right front (RF) legs executes hoof off (HO) before the hoof strike (HS) of the left rear (LR) and right rear (RR) legs respectively.

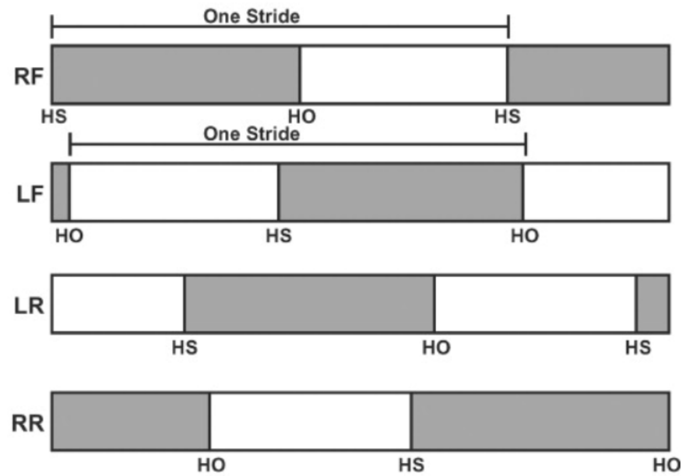

Figure 3 – Temporal distribution of stance phase and swing phase for each leg of a cattle. RF: right front. LF: left front. LR: left rear. RR: right rear. HS: Hoof strike. HO: Hoof off. Illustration taken from [13].

Figure 4 – Cattle left front leg lifting to make space for left hind leg (left) and cattle right front leg lifting to make space for right hind leg (right).

1.3 Cattle Diseases Associated with Lameness

Lameness is an indicator of many diseases which manifests as abnormal gait. Most diseases that induces lameness consists of infections or defects in the hooves such as sole ulcer, foot rot and white line disease [14]. Sole ulcer is caused by the destruction of part of the hoof tissue covering the dermis [15]. The exposed dermis inflicts pain when pressure is applied to the area. Another cause of lameness is foot rot, which is an infection affecting the skin near the hooves. It is known that the causes of lameness are associated with environmental factors and farming practices [16]. Lameness results in lower milk yield [17, 18], presumably because cattle are less motivated to feed when in pain [19]. Other consequences of lameness are lower reproductive performance [20, 21], lower welfare [8] and in severe cases, the cattle is deemed unproductive, resulting in early culling [22]. Current practice for treatment includes antibiotic sprays or footbaths. More sophisticated methods involve hoof trimming to expose affected area and attaching a block on the hoof such that the affected part is elevated and subject to less stress to speed recovery.

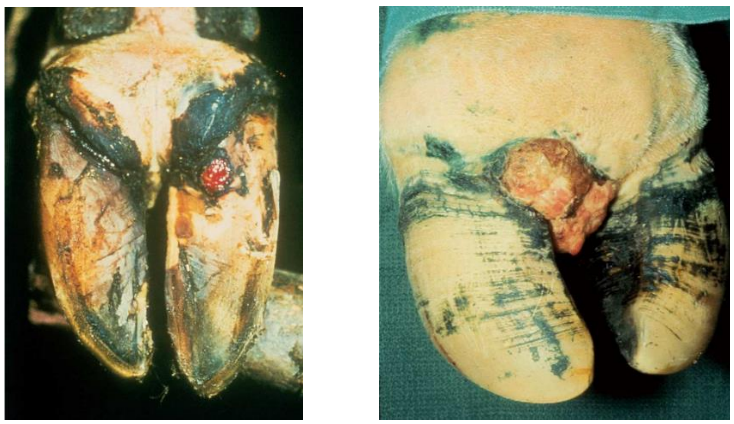

Figure 5 – Sole ulcer (left) and foot rot (right). Images taken from [15].

1.4 Locomotion Scoring

Many visual assessment systems, called locomotion scoring systems, has been devised in literature to define visual indicators of cattle lameness and its severity. For example, one popular locomotion scoring system by Sprecher et al. defines arched back, short strides, unnatural strides, favoring certain limbs and reluctance to bear weight on certain limbs as visual indicators of lameness (Table 1) [23]. In this system, a score of 1 corresponds to normal gait and a score of 5 corresponds to severe lameness.

Table 1 – 5 point locomotion scoring system from [23]

Lameness score	Clinical description	Assessment criteria
1	Normal	The cow stands and walks with a level-back posture. Her gait is normal.
2	Mildly lame	The cow stands with a level-back posture but develops an arched-back posture while walking. Her gait remains normal.
3	Moderately lame	An arched-back posture is evident both while standing and walking. Her gait is affected and is best described as short-striding with one or more limbs.
4	Lame	An arched-back posture is always evident and gait is best described as one deliberate step at a time. The cow favors one or more limbs/feet.
5	Severely lame	The cow additionally demonstrates an inability or extreme reluctance to bear weight on one or more of her limbs/feet.

Figure 6 shows examples of cattle of locomotion scores 1,3 and 5. Although some of the visual indicators of lameness are undetectable in still images, it can be seen that the back arches more with higher locomotion scores. The example image for locomotion score 5 also suggests that the cattle may be leaning to one side of its body to avoid bearing weight on the other.

Figure 6 – Cattle with locomotion score 1(left), 3(middle) and 5(right). Images taken from https://vet360.vetlink.co.za/locomotion-scoring-crucial-component-lameness-reduction-programs/ on 6-5-2019.

Another locomotion scoring system is one proposed by Manson and Leaver which focuses on abduction/adduction, gait asymmetry, tender hoof placement and difficulties in movement [24]. Breuer et al. used a 4 point system with an emphasis on head bobbing [25].

It is important to emphasize that lameness is the manifestation of pain and hence locomotion scoring can only detect how much pain is suffered by the cattle but not necessarily the severity of diseases. In fact, it has been found that higher locomotion scores are only associated with certain diseases but not with others [26]. Reliability of locomotion scoring systems varies widely but they are generally found to be moderately reliable [27-29].

1.5 Automatic Lameness Detection

Due to lack of time and labor, farmers are generally very reluctant to take action to control lameness in their herds [9]. To tackle this problem, some automatic lameness detection solutions have been proposed in literature. These solutions all involve some way of measuring something from the cattle and detecting patterns that distinguish lame cattle from healthy cattle. They can be categorized based on the method used, namely, accelerometers, kinetics, marker-based kinematics and machine vision.

Accelerometer based methods involve attaching accelerometers to anatomically important body parts of cattle and measuring their acceleration [30-32]. Kinetics based methods involve force sensing platforms that can measure the amount of force exerted by the feet of cattle and determine the location where the force is exerted [33, 34]. Kinematics based methods with markers involve the attachment of markers that can be easily detected by cameras on anatomically important body parts of cattle and track their movements in 2D or 3D space [35, 36]. Machine vision based methods involve video recordings of walking cattle and complex algorithms to extract information from the videos, like the back arch [37] and limb movements [38, 39].

Most of the current proposed methods have not been commercialized yet, presumably due to their disadvantages. Accelerometers and marker based kinematics require attachments on various body parts of the cattle, which is a very laborious task even for small farms with dozens of cattle. Moreover, the attachments would not only need to be secured tightly on the cattle so as to not fall off, but they would also need to be resistant to pressure, collisions, water, slurry and other environmental effects in cattle farms. Kinetics based methods require significant investment and space on the farm. This may be more feasible at very large farms that have separate buildings for milking and feeding. A force sensing walkway can be placed between the buildings such that the cattle will walk over it every time they go for milking. However, smaller farms typically have one barn where both feeding and milking take place, leaving barely any space for a force sensing walkway.

Machine vision based methods, are very attractive options from a feasibility point of view because they only require one or a few cameras which can be easily installed in farms and do not require attaching anything to the cattle. However, the proposed machine vision methods so far have not been successful in achieving robust and accurate lameness detection. Over the recent years, the advancement of deep learning methods have revolutionized the field of computer vision, significantly improving the performance of typical computer vision tasks such as object detection and segmentation to rival and even surpass human performance. Recently, an open source deep learning framework named DeepLabCut has received much attention from animal behaviorists and neuroscientists because it enables automatic, markerless animal pose estimation at human level accuracy [40]. DeepLabCut is based on the feature detector of a human pose estimation framework, DeeperCut [41], which itself is a variant of deep residual neural networks (ResNet) [42]. By using a ResNet pretrained on ImageNet [43], relatively few additional images (hundreds instead of tens of thousands) are needed to fine tune DeepLabCut for the pose estimation of arbitrary objects.

1.6 Research Questions

To strive towards an affordable system that can detect lameness without the hassle of attachments on cattle, the following research question is posed:

- How accurately can cattle lameness be detected based on features extracted from pose estimation using DeepLabCut?

With the following sub-questions:

- Can DeepLabCut accurately detect the location of cattle body parts and estimate their poses?
- Are the gait variables calculated from pose estimation different between healthy cows and lame cows?
- What are the sensitivity and specificity of lameness detection using information from the gait variables and manual locomotion score as ground truth?

Firstly, although DeepLabCut has been shown to achieve high accuracies in estimating the poses of various animals such as mice, horses and cheetahs in complex environments, it should be confirmed that it also works for cattle in a farm environment. Then, using the pose estimation results, gait variables can be calculated. These gait variables should be analyzed to find out if they differ between healthy cows and lame cows. Finally, an attempt will be made at automatically classifying the recorded cows based on the gait variables calculated.

2. Methods

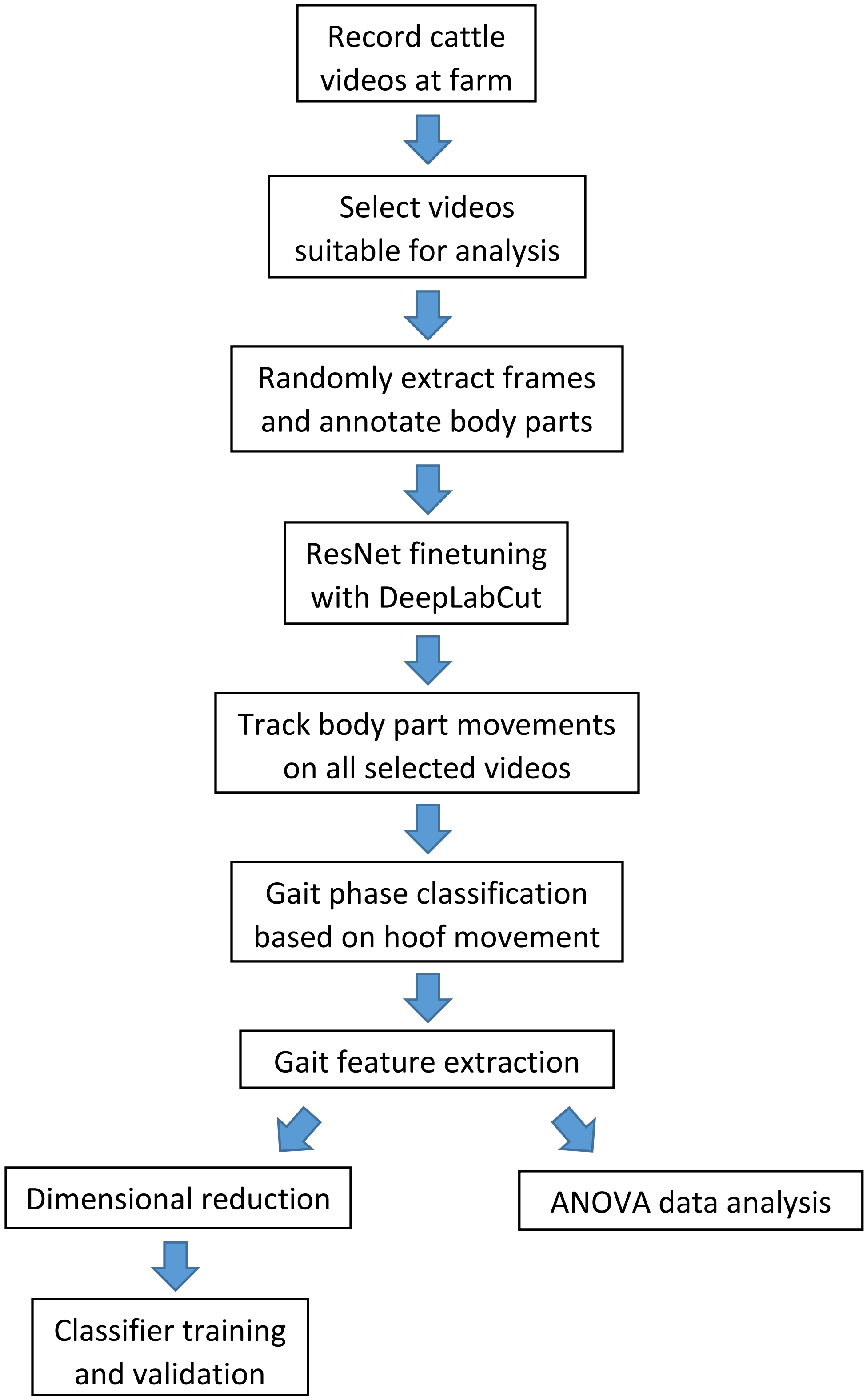

Figure 7 – Project process flowchart

2.1 Cattle and Farm Conditions

The farm where videos were recorded is in Tilburg, Netherlands. The herd from which videos were recorded consist of 70 lactating cows. The cows were milked with an automatic milking system (Lely Astronaut, Lely) where the cows may choose when they want to be milked.

At the time of year when the videos were recorded, the cows were allowed access to the pasture located a small distance away from the barn. Access to the pasture was allowed between 9am to 5pm. Within this time period, an automatic gate system manages the cows' exit and entry. This system reads the identification tag of the cows to determine when they were last milked. Cows were only allowed to exit the barn if they were recently milked within the last few hours. Upon exiting the barn, a walkway with electric fences leads to the pasture. Return to the barn was also through this same walkway. Before 5pm, the cows were allowed to return any time they wish to be milked in the barn. At 5pm, the cows that were still outside were led back into the barn by farm staff.

2.2 Recording Setup

Figure 8 – ZED camera (left). Recording setup with camera and a box to store laptop (right).

The recording setup used in this project consists of a depth camera (ZED, StereoLabs) mounted on a tripod (Figure 8). Although the camera used can capture depth images in addition to regular color images, only the color images were used in this project. The height of the pole holding the camera is adjustable and was set such that the camera is 2 meters from the ground. A metal box was attached to the tripod for storing a laptop and an external hard drive. The laptop is required to operate the camera and run online compression of recorded videos. This setup was placed 4 meters from the walkway leading from the barn to the pasture.

2.3 Motion Detection and Recording

With the camera recording at 1080p and 30fps, single images were retrieved from the camera and resized to 1/10 of its original size via bilinear interpolation. The images were downsampled to reduce processing burden such as not to interfere the video recording process. Then, the area of the image covering the walkway and the cows were segmented, converted to grayscale, and filtered with a 3-by-3 Gaussian kernel to reduce noise.

Motion is detected by background subtraction. First, the absolute difference between the retrieved image and a reference background image was calculated by element wise subtraction and taking the absolute values. Pixel locations where there is an absolute difference of 20 or more were segmented and further dilated with a 3-by-3 structuring element. This will result in several blobs in the image where the large blobs are cows and small blobs are noise or small movements of the environment such as swaying grass and passing birds. These segmented blobs were then detected individually and each of their contours were computed using the algorithm by Suzuki and Abe [44]. The areas within these contours were then calculated using Green's theorem.

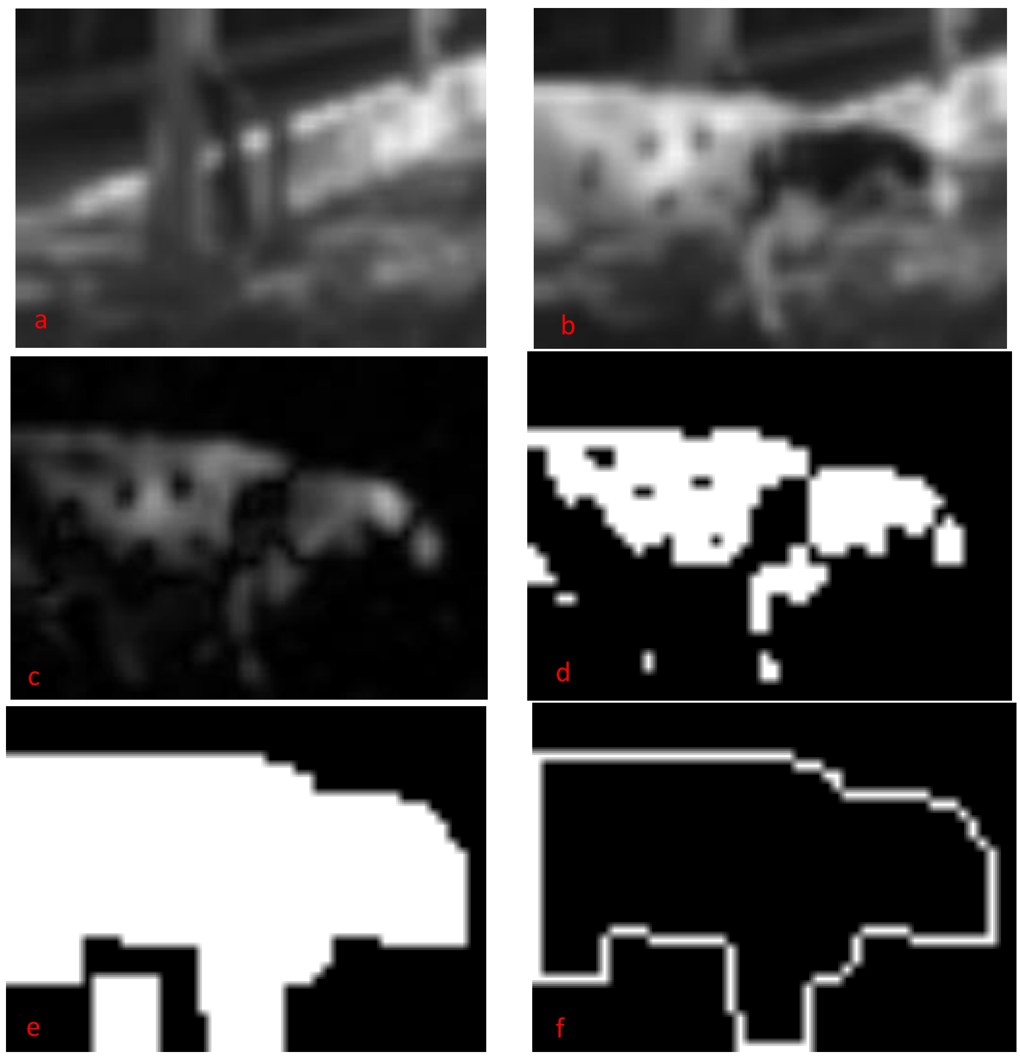

Figure 9 – (a) Background image. (b) Image with cow. (c) Background subtraction. (d) Pixels with large difference with background image. (e) Dilation (f) Contour of larger than threshold size

Every time an object with an area of more than 500 pixels but less than 1000 pixels was detected, a 12-second video was recorded.

Due to changing lighting conditions influenced by the sun, the reference background image used for background subtraction cannot simply be a single static image. Thus, a running average of recent images was used as a dynamic “background” image:

$$A_t = (1 - \alpha)A_{t-1} + \alpha I_t$$

where I_t is the tth frame captured by the camera, A_t is the running average of images up till the tth frame, and α controls how quickly the running average is updated with new images. The α value used in this thesis project was set to 0.1, which was found by trial and error. When the α value was too small, the “background” image was not adjusted quickly enough to rapidly changing lighting conditions. When the α value was too large, cows captured by the camera, especially slow moving cows, were quickly incorporated into the “background”, making them undetectable.

The above algorithm was implemented with python and set to automatically start every day at 9am and stop at 6pm. Interfacing with the ZED camera was done through an API for python provided by the developers (https://github.com/stereolabs/zed-python-api). The image processing algorithms were implemented with OpenCV.

2.4 Dataset Creation

From the recorded videos of the previous segment, a subset of videos was used as the dataset for this thesis project with the following inclusion criteria

- Only a single cow in the video
- The cow walks from the left of the camera view to the right
- The cow walks normally without stopping, slipping, slowing down or running
- The cow takes a total of more than 3 steps
- Lightning from the sun is not too bright nor too dim

Having only a single cow in the videos avoids the need to separate multiple cows in the video and greatly simplifies the later tasks. Having videos with cows walking in only one direction also greatly simplifies the later tasks. Cows in the videos should walk with their natural gait and thus videos containing cows stopping, slipping, slowing down or running were excluded. Videos of cows taking less than 3 steps were also excluded because it is difficult to manually score a cow without observing multiple steps. Videos with too much or too little lighting must also be excluded to ensure similar environmental conditions such that they do not influence the gait analysis results too much.

Figure 10 – Videos with high lighting intensities such as this were excluded

Videos satisfying the first criterion were found using YOLO, an object detection system [45]. Using the default weights provided in YOLO, which were pre-trained on the COCO dataset [46], YOLO finds the bounding boxes of a wide variety of objects and labels them.

A video is only included in the dataset if a single cow, dog or horse with a bounding box of more than 100,000 pixels was detected in one or more frames of the video. When two or more cows, dogs and horses with a bounding box of more than 100,000 pixels were detected in any frame in a video, the video was excluded from the dataset. Bounding boxes for dogs and horses were included because sometimes YOLO mistakenly detects cows as dogs and horses. Only bounding boxes with an area of more than 100,000 pixels were of interest because bounding boxes smaller than that were cows behind the walkway.

Then, each video was screened manually to select the ones that satisfy the rest of the criteria.

2.5 Locomotion Scoring

Each video of a walking cow was given a locomotion score of 1 to 5 after discussion by two people knowledgeable in animal science and have experience with assessing cow health. The scorers were first given a small set of 17 cow videos for a test run. These videos include 12 cows displaying common lameness indicators like back arch and gait asymmetry. A meeting was then set up with the scorers, where a set of general criteria for each locomotion score was discussed and agreed on (table 2) using the 17 test videos to facilitate the discussion. Locomotion scores were mostly whole numbers, except locomotion score 1.5 because the scorer's felt the need to have a score between 1 and 2 for cows that they suspected to be slightly lame but weren't sure about it. However, the cows with locomotion score 1.5 were later relabeled as score 2 for simplification and because it was more important to detect signs of lameness rather than differentiating between locomotion scores 1.5 and 2.

Table 2 – Locomotion scores and their general criteria

Locomotion Score	Criteria
1	Walks normally with a straight back and no visible signs of problems.
1.5	Has a slightly arched back but no problem in the limbs.
2	• Has an arched back OR • has a problem in one of the limbs OR • has a slightly arched back and a problem in one of the limbs
3	• More serious and obvious version of score 2 criteria OR • problems in multiple limbs OR • has a stretched neck
4	Walks with great difficulty due to severe problems in limbs.
5	Can barely walk.

These criteria were only intended as a reference and not adhered to strictly. The final judgement given by the scorers ultimately depended on their intuition. There were two reasons for this. The first reason is that locomotion scoring in common practice solely relies on visual assessment and not objective measurements. Thus, as with other locomotion scoring methods [23], it is difficult to describe how a locomotion score was given in non-ambiguous words. For example, it is up to the scorer's intuition to decide where to draw the line between a slight back arch and a more serious back arch.

The second reason is that locomotion scoring, like health examinations in general, requires more flexibility than a strict set of criteria. For example, a cow may exhibit indicators that fit the criteria of locomotion score 2, but the scorers might conclude that the cow is perfectly healthy. Such cases may involve cows that has an unusual build, causing it to appear lame or cows that seems to be happy and lively despite exhibiting lameness indicators.

2.6 Body Part Tracking

17 key points were automatically detected and tracked in the recorded videos using a deep learning framework, DeepLabCut [40]. These keypoints are illustrated in Figure 11.

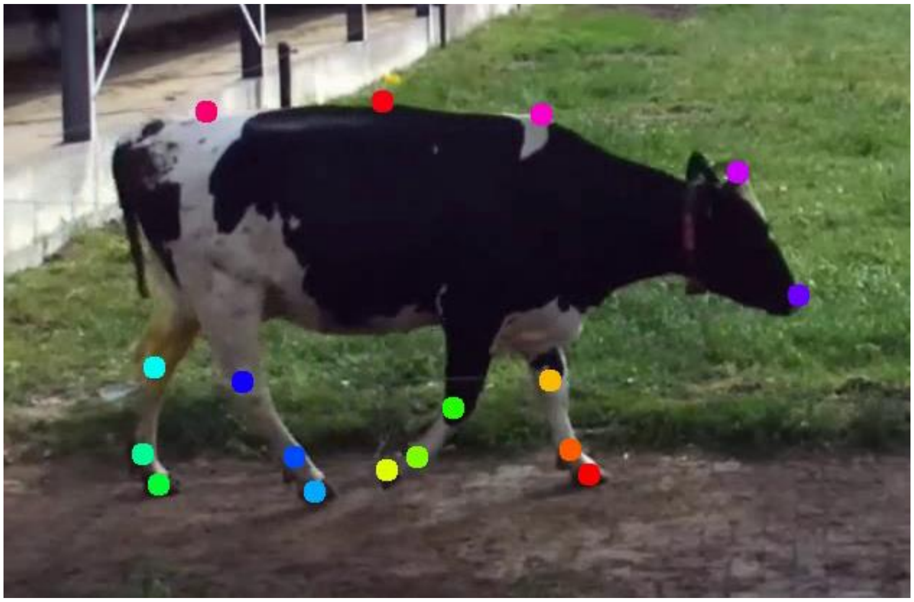

Figure 11 – Tracked body parts.

3 points were labeled for each limb, the hoof, the fetlock joint and the carpal/tarsal joint. 2 points were labeled for the head: the top of the head and the nose. Lastly, 3 points were labeled on the back of the cow: one point around the connection between neck and torso, one point above the hip bone, and one point in the middle of the two. These points on the back were defined ambiguously because there are no easily distinguishable features on the back.

The hooves, head, and back were chosen as key points because the movement patterns of these body parts are commonly used in locomotion scoring. The fetlock joint and the carpal/tarsal joint were also chosen because they are visually distinguishable body parts.

495 images were randomly extracted from 55 videos recorded on 3rd June. The images were then manually annotated and used to train a deep neural network for 850,000 iterations on a GPU (GTX 1080 Ti). The network was then used to find the pixel coordinates of the target body parts across all frames in 335 videos recorded during the time period ranging from 3rd June to 15th July. All steps from the random frame extraction to the points detection were facilitated by the DeepLabCut framework [47].

For each video, the positions of each body part key point across time were output as two vectors: the horizontal positions x in pixels, the vertical positions y in pixels. The length of both vectors are equal to the number of frames in the video. The specific position of a key point in a particular frame will be denoted as (x_t, y_t) for the tth frame in the video.

2.7 Valid intervals

In most of the videos, not all the frames contain a cow (e.g. the cow crossed the edge of the camera's view but the video was still recording). In such frames, DeepLabCut will still try and detect nonexistent cow body parts and outputting meaningless values for the positions of the body parts. Thus, for every video an interval where the full body of the cow is in the frame had to be determined. Specifically, the start of the interval should be the first frame which includes the full body of the cow, while the end of the interval should be the last frame before part of the cow crosses the right edge of the frame.

This was implemented in python where YOLO [45] was used again for detecting and finding the position of the cow. The first frame where the bounding box's left edge was more than 50 pixels away from the left edge of the image, and had an area of more than 100,000 pixels was set as the first frame of the valid interval. If at any frame a bounding box's left edge was more than 1300 pixels away from the left edge of the image, and had an area of more than 100,000 pixels, the corresponding frame was set as the last frame of the valid interval. If none of the frames of the video satisfied the criterion for the last frame of the valid interval, the last frame of the video was used instead.

Figure 12 – Example of frame used for interval start (left) and interval end (right)

2.8 Gait phase classification

The calculation of most gait variables require that the timing of the gait phases of each limb be known such that the swing phase and the stance phase can be separated. A simple classification algorithm based on the velocity of each hoof key point was made.

First, the horizontal velocities v of a hoof were estimated by calculating the differences in the horizontal positions adjacent in time.

$$v_t = x_t - x_{t-1}$$

If v_t is more than 8, the tth frame is classified as swing phase and if it is less than or equal to 8, the tth frame is classified as stance phase. If a swing or stance phase has a duration of less than 4 frames, it is assumed to be an error caused by the misdetection of the key points and is reclassified as the other phase.

Figure 13 shows an example of the classification of swing and stance phase based on hoof velocity.

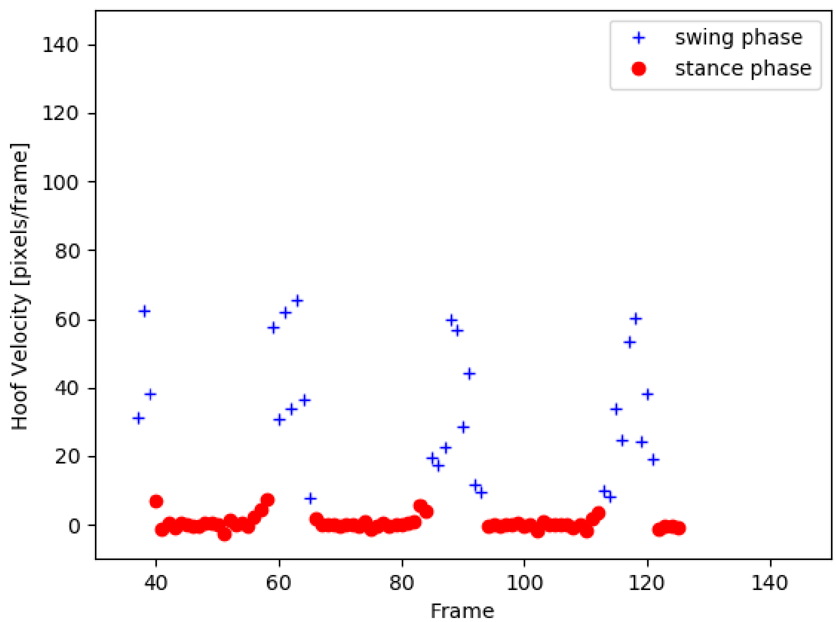

Figure 13 – Plot of hoof velocity and its classified phase

2.9 Gait Features

In total, 84 gait features were calculated from each video. The full list of these features can be found in appendix A1, but most of these features were similar to each other and can essentially be summarized into the features in table 3.

Table 3 – Gait features and their methods of calculation.

Gait features	Calculation
Step length	Difference in pixel in horizontal position between the start and end of a swing phase
Step time	Duration in number of frames of a swing phase
Stance time	Duration in number of frames of a stance phase
Asymmetry in step length	Difference in pixel between step length of left and right limbs
Asymmetry in step time	Difference in number of frames between step time of left and right limbs
Asymmetry in stance time	Difference in number of frames between stance time of left and right limbs
Head bobbing	Variance in pixel^2 of vertical position of head
Nodding	Variance of angle between top of the head and nose
Back arch	Radius in pixel of circle fitted on the three back key points

Some of these essential features were calculated multiple times but at different times in the video and for different limbs. For example, 19 different step length features were calculated: "LF step length 1" is the length of the first step of the left front limb, "LF step length 2" is the length of the second step of the left front limb, "RF step length 1" is the length of the first step of the right front limb, etc.

2.10 DeepLabCut performance validation

To find out how well DeepLabCut performs on our particular farm setting, the 495 labeled images were randomly split into a training set and a test set (80% and 20% respectively) for validation. The training set was used to refine a ResNet-50 network pretrained on ImageNet for 200,000 iterations on a GPU (GTX 1080 Ti). The refined network was then tested on the test set and the average euclidean error between the manual labels and the labels predicted by DeepLabCut was calculated. This process was repeated 5 times and the mean error was calculated.

2.11 Data Analysis

To gain an insight on how the gait features differ between locomotion scores, ANOVA (analysis of variance) followed by Bonferroni corrected tests were performed on most gait feature. Gait features calculated multiple times were excluded because their means were sufficient for the purpose of this multiple comparison test. For example, features from individual steps "LF step length 1", "LF step length 2" and LF step length 3" were excluded, but the mean of these individual steps, "LF step length" was included in the analysis. This step was implemented using MATLAB R2019a, mainly for its data analysis functions, 'anova1' and 'multcompare' (See appendix A2).

2.12 Lameness classification

For the lameness classification, linear discriminant analysis (LDA) was first used to reduce the 84 gait features to 3 features. LDA constructs 3 features that best differentiates the cows from different locomotion score groups based on the 84 gait features.

A linear discriminant classifier, quadratic discriminant classifier, k nearest neighbors classifier, decision tree classifier, support vector machine, parzen classifier and neural network classifier was trained on the reduced features dataset and the cows were classified into a locomotion score group. Prior probabilities were set to the frequency of each locomotion score. To assess the classifiers, the classification rate of each classifier was calculated with 10-fold cross validation. This was repeated 10 times to calculate the means and standard deviations of the classification rates.

The dataset was expected to be very unbalanced and dominated by cows with locomotion score 1, because the number of cows in each locomotion score group usually decreases with higher locomotion scores. The overall classification rate may be heavily influenced by the dominating group and can be misleading. Thus, it may be more informative to calculate the confusion matrices and get a clearer picture of how the classifiers performed on each locomotion score group.

A confusion matrix was calculated with each 10-fold cross validation. Since the 10-fold cross validation was performed 10 times, resulting in 10 confusion matrices, the mean and standard deviation of each element in the confusion matrix were calculated. This was repeated with the prior probabilities set equal for all locomotion scores to find out how it affects the classification rate and confusion matrix and whether it can correct the imbalance of the dataset.

Everything in this section from dimensional reduction to cross validation was implemented with a pattern recognition toolbox for MATLAB called PRTools (http://37steps.com/37-steps/). Specifically, PRTools5 was used in MATLAB R2019a (See appendix A3).

3. Results

3.1 DeepLabCut pose estimation performance

The mean error of 5 randomly split datasets as described in section 2.10 was 4.4 ± 1.4 pixels. Figure 14 shows examples of key points detected by DeepLabCut.

Figure 14 – Key points detected by DeepLabCut

3.2 Final dataset

The dataset created consisted of 335 videos recorded on dates ranging from 3rd June to 15th July. 22 videos were further excluded because they had poor body part tracking performance, had less than 3 full step phases or had less than 3 full stance phases. This resulted in a final dataset size of 313 videos. Due to the small number of locomotion score 5 cows (only two), they were relabeled and added to the locomotion score 4 group. The final distribution of locomotion scores was 224, 53, 25 and 11 videos for locomotion score 1, 2, 3 and 4 respectively.

3.3 Data Analysis

Table 4 presents the means of the analyzed gait features and their significant differences between different locomotion score groups according to the bonferroni corrected test. Different alphabetical letter superscripts represent groups that are significantly different. For example, in the entry for 'LF step length': 278^{a}, 268^{ab}, 258^{b}, 227^{c}, the numbers that do not have the same letter in their superscript are significantly different. In this particular example, locomotion score 1 is different from locomotion score 3 and 4, locomotion score 2 is only different from locomotion score 4, locomotion score 3 is different from locomotion score 1 and 4.

Table 4 - Mean of gait features and significant differences between different locomotion score groups

Features	Locomotion Score 1	Locomotion Score 2	Locomotion Score 3	Locomotion Score 4
LF step time [frames]	9.3^{a}	9.3^{a}	9.3^{a}	8.5^{b}
RF step time [frames]	9.3^{a}	9.4^{a}	9.1^{a}	9.5^{a}
LH step time [frames]	9.3^{a}	9.3^{a}	9.7^{a}	8.5^{b}
RH step time [frames]	9.2^{a}	9.2^{a}	8.9^{a}	10.0^{b}
Overall step time [frames]	9.3^{a}	9.3^{a}	9.2^{a}	9.1^{a}
Minimum step time [frames]	7.8^{a}	7.5^{ab}	7.6^{ab}	6.8^{b}
Maximum step time [frames]	10.7^{a}	11.1^{a}	11.0^{a}	11.8^{a}
LF step length [pixels]	278^{a}	268^{ab}	258^{b}	227^{c}
RF step length [pixels]	281^{a}	272^{ab}	259^{b}	232^{c}
LH step length [pixels]	276^{a}	267^{a}	263^{a}	226^{b}
RH step length [pixels]	286^{a}	276^{ab}	264^{bc}	239^{c}
Overall step length [pixels]	280^{a}	271^{b}	261^{b}	231^{c}
Minimum step length [pixels]	233^{a}	229^{ab}	223^{ab}	187^{b}
Minimum step length [pixels]	341^{a}	327^{a}	316^{a}	273^{a}
LF stance time [frames]	20.4^{a}	22.5^{b}	23.6^{b}	29.2^{c}
RF stance time [frames]	20.3^{a}	22.1^{b}	23.4^{b}	27.9^{c}
LH stance time [frames]	20.4^{a}	22.6^{b}	23.4^{b}	29.4^{c}
RH stance time [frames]	20.3^{a}	22.4^{b}	23.6^{b}	27.1^{c}
Overall stance time [frames]	20.3^{a}	22.4^{b}	23.4^{b}	28.3^{c}
Minimum stance time [frames]	16.2^{a}	18.2^{ab}	18.4^{ab}	21.8^{b}
Maximum stance time [frames]	23.4^{a}	25.5^{ab}	27.4^{b}	32.7^{c}
Front step length asymmetry [pixels]	17^{a}	17^{a}	13^{a}	12^{a}
Hind step length asymmetry [pixels]	17^{a}	20^{a}	24^{a}	24^{a}
Front step time asymmetry [frames]	0.8^{a}	0.9^{a}	0.9^{a}	1.5^{b}
Hind step time asymmetry [frames]	0.9^{a}	1.2^{ab}	1.4^{b}	2.4^{c}
Front stance time asymmetry [frames]	1.4^{a}	1.6^{a}	1.6^{a}	3.0^{b}
Hind stance time asymmetry [frames]	1.5^{a}	1.7^{ab}	2.7^{bc}	3.7^{c}
Nose bobbing [$pixels^2$]	1478^{a}	1887^{a}	872^{a}	805^{a}
Nodding	0.0107^{a}	0.0086^{a}	0.0072^{a}	0.0117^{a}
Back arch [pixels]	5389^{a}	1636^{b}	1211^{b}	794^{ab}

3.4 Dimension reduction and classification

Figure 15, 16 and 17 show scatterplots of the three features after dimensional reduction with LDA. Locomotion scores 1,2,3 and 4 were represented with a blue plus, red asterisk, green circle and black cross respectively.

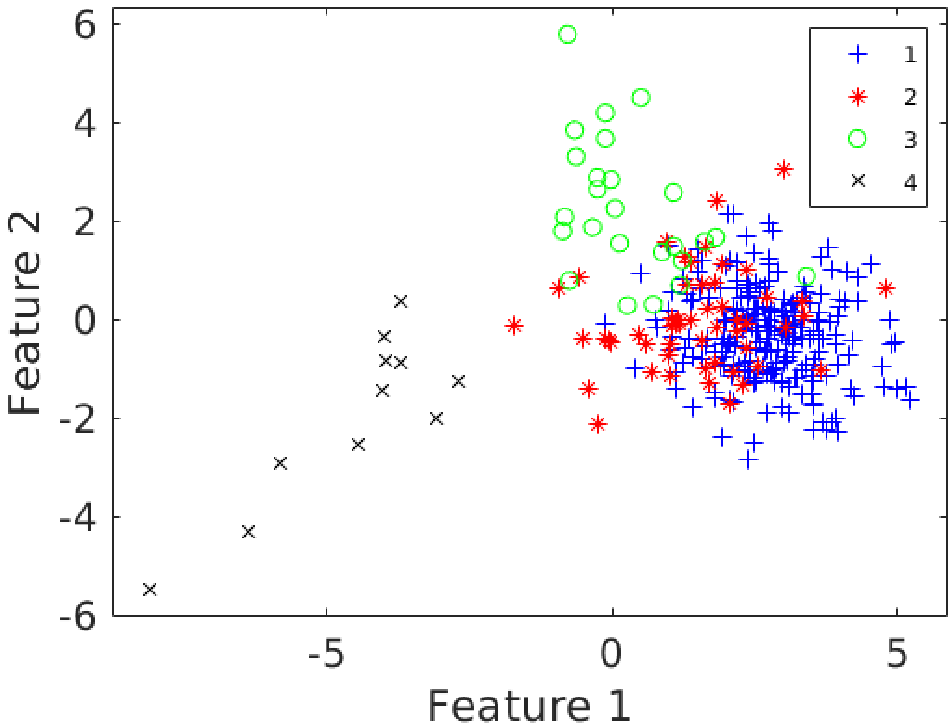

Figure 15 – Distribution of dataset after dimensional reduction (first and second feature)

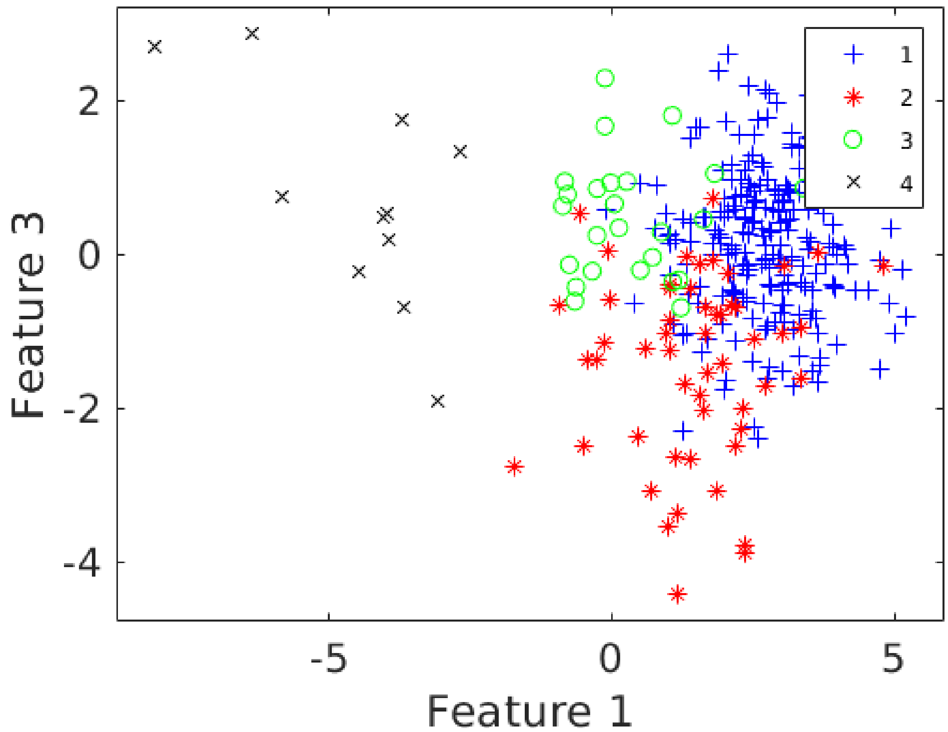

Figure 16 – Distribution of dataset after dimensional reduction (first and third feature)

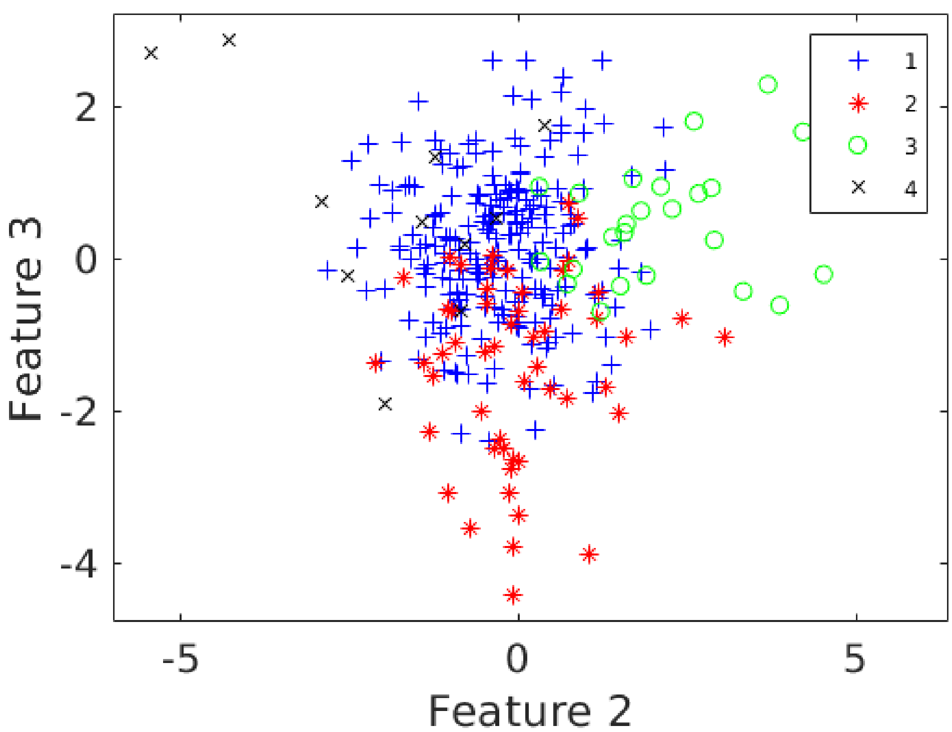

Figure 17 – Distribution of dataset after dimensional reduction (second and third feature)

The mean and standard deviation of the error rate after 10 repetitions of 10-fold cross validation is listed in table 5 for each classifier tested. The highest classification rate was from the linear classifier with $85.6\% \pm 0.5\%$.

Table 5 – Comparison of error rates of each classifier

Classifier	Classification Rate
Linear Discriminant	85.6% ± 0.5%
Quadratic Discriminant	84.5% ± 0.3%
K Nearest Neighbors	82.6% ± 0.6%
Decision Tree	76.7% ± 1.1%
Support Vector Machines	84.0% ± 0.5%
Parzen	84.5% ± 0.4%
Neural Network	75.2% ± 1.2%

Tables 6-12 below are the average confusion matrices of each classifier from the 10-fold cross validations. The rows represent the true score and the columns represent the predicted score by the classifier. For example, in table 6, the element in the first row and first column means that there were on average 213.4 locomotion score 1 cows classified as locomotion score 1. The element in the first row and second column means that there were on average 10 locomotion score 1 cows classified as locomotion score 2. The element in the second row and first column means that there were on average 22 locomotion score 2 cows classified as locomotion score 1.

Table 6 – Average confusion matrix from linear classifier after LDA

True Score	Predicted Score				Total
	1	2	3	4	
1	213.4 ± 0.8	10.0 ± 0.7	0.6 ± 0.5	0	224
2	22.0 ± 1.1	28.7 ± 1.1	2.3 ± 0.8	0	53
3	8.6 ± 0.8	1.5 ± 0.5	14.9 ± 0.6	0	25
4	0	0	0	11.0	11

Table 7 – Average confusion matrix from quadratic classifier after LDA

True Score	Predicted Score				Total
	1	2	3	4	
1	220.2 ± 0.6	3.6 ± 0.5	0.2 ± 0.4	0	224
2	31.6 ± 0.8	19.7 ± 0.8	1.7 ± 0.5	0	53
3	9.0	0.3 ± 0.5	15.7 ± 0.5	0	25
4	0	2.0	0	9.0	11

Table 8 – Average confusion matrix from K nearest neighbors classifier after LDA

True Score	Predicted Score				Total
	1	2	3	4	
1	207.7 ± 1.8	13.9 ± 1.7	2.4 ± 0.7	0	224
2	26.8 ± 1.5	24.4 ± 1.0	1.8 ± 0.8	0	53
3	7.8 ± 0.4	1.8 ± 0.4	15.4 ± 0.5	0	25
4	0	0	0	11.0	11

Table 9 – Average confusion matrix from decision tree classifier after LDA

True Score	Predicted Score				Total
	1	2	3	4	
1	195.9 ± 2.8	23.4 ± 2.4	4.7 ± 0.7	0	224
2	26.2 ± 2.6	23.7 ± 2.3	3.1 ± 0.6	0	53
3	7.5 ± 1.2	2.7 ± 0.8	14.8 ± 1.2	0	25
4	2.4 ± 0.8	2.2 ± 1.0	0.8 ± 0.4	6.0 ± 1.0	11

Table 10 – Average confusion matrix from support vector machine after LDA

True Score	Predicted Score				Total
	1	2	3	4	
1	216.2 ± 1.1	5.7 ± 0.9	2.1 ± 0.6	0	224
2	29.6 ± 1.5	21.2 ± 1.5	2.2 ± 0.8	0	53
3	9.5 ± 0.8	0	15.5 ± 0.8	0	25
4	0	1.0	0	10.0	11

Table 11 – Average confusion matrix from parzen classifier after LDA

True Score	Predicted Score				Total
	1	2	3	4	
1	219.5 ± 0.7	3.6 ± 0.7	0.9 ± 0.3	0	224
2	31.1 ± 1.0	19.7 ± 0.7	2.2 ± 0.4	0	53
3	9.7 ± 0.5	1.0	14.3 ± 0.5	0	25
4	0	0	0	11.0	11

Table 12 – Average confusion matrix from neural network classifier after LDA

True Score	Predicted Score				Total
	1	2	3	4	
1	192.1 ± 3.6	28.0 ± 3.4	3.9 ± 1.7	0	224
2	25.6 ± 1.4	24.3 ± 1.4	2.8 ± 1.0	0.3 ± 0.7	53
3	7.3 ± 1.3	5.1 ± 1.7	12.4 ± 0.8	0.2 ± 0.4	25
4	1.5 ± 1.4	3.0 ± 1.6	0	6.5 ± 1.6	11

The confusion matrices after setting the prior probabilities of all locomotion scores to be equal is shown in tables 13-19.

Table 13 – Average confusion matrix from linear classifier after LDA and equalizing prior probabilities

True Score	Predicted Score				Total
	1	2	3	4	
1	186.8 ± 1.0	28.5 ± 1.2	8.7 ± 1.1	0	224
2	10.6 ± 1.0	37.7 ± 1.2	4.2 ± 0.4	0.5 ± 0.5	53
3	1	3	21	0	25
4	0	0	0	11	11

Table 14 – Average confusion matrix from quadratic classifier after LDA and equalizing prior probabilities

True Score	Predicted Score				Total
	1	2	3	4	
1	182.4 ± 2.2	31.1 ± 1.4	10.5 ± 1.6	0	224
2	12.9 ± 1.2	33.9 ± 1.4	5.2 ± 0.8	1	53
3	1.1 ± 0.3	1.8 ± 0.6	22.1 ± 0.7	0	25
4	0	1.1 ± 0.3	0	9.9 ± 0.3	11

Table 15 – Average confusion matrix from K nearest neighbors classifier after LDA and equalizing prior probabilities

True Score	Predicted Score				Total
	1	2	3	4	
1	207.3 ± 0.9	14.3 ± 0.9	2.4 ± 0.7	0	224
2	26.8 ± 1.5	24.4 ± 1.0	1.8 ± 0.8	0	53
3	7.8 ± 0.4	1.8 ± 0.4	15.4 ± 0.5	0	25
4	0	0	0	11	11

Table 16 – Average confusion matrix from decision tree classifier after LDA and equalizing prior probabilities

True Score	Predicted Score				Total
	1	2	3	4	
1	196.1 ± 2.5	23.1 ± 2.1	4.8 ± 0.8	0	224
2	26.4 ± 2.8	23.6 ± 2.3	3.0 ± 0.7	0	53
3	7.4 ± 1.1	2.9 ± 1.1	14.7 ± 1.4	0	25
4	2.4 ± 0.8	2.3 ± 1.1	0.8 ± 0.4	6.0 ± 1.0	11

Table 17 – Average confusion matrix from support vector machine after LDA and equalizing prior probabilities

True Score	Predicted Score				Total
	1	2	3	4	
1	215.2 ± 1.2	6.7 ± 1.3	2.1 ± 0.6	0	224
2	28.3 ± 1.8	22.3 ± 1.8	2.4 ± 0.7	0	53
3	8.9 ± 0.7	0	16.1 ± 0.7	0	25
4	0	1	0	10	11

Table 18 – Average confusion matrix from parzen classifier after LDA and equalizing prior probabilities

True Score	Predicted Score				Total
	1	2	3	4	
1	184.2 ± 1.6	25.4 ± 1.4	14.4 ± 1.3	0	224
2	12.6 ± 0.8	33.8 ± 0.6	6.6 ± 0.7	0	53
3	1.4 ± 0.5	1.5 ± 0.7	22.1 ± 0.9	0	25
4	0	0	0	11	11

Table 19 – Average confusion matrix from neural network classifier after LDA and equalizing prior probabilities

True Score	Predicted Score				Total
	1	2	3	4	
1	192.6 ± 5.4	25.8 ± 4.3	5.5 ± 1.9	0.1 ± 0.3	224
2	25 ± 1.2	25.1 ± 1.3	2.9 ± 0.7	0	53
3	7.4 ± 1.8	4.7 ± 1.8	12.9 ± 1.4	0	25
4	2 ± 1.1	2.3 ± 1.3	0.1 ± 0.3	6.6 ± 1.6	11

4. Discussion

The aim of this project was to find out if markerless pose estimation can feasibly detect lameness in cattle. Using DeepLabCut, markerless pose estimation of cows could be done accurately and gait features could be extracted reliably from walking cow videos. Tests for significant differences revealed that some gait features were able to differentiate between cows of different locomotion scores, except between locomotion scores 2 and 3. Classifiers trained with the acquired data initially achieved modest sensitivity and very high specificity, but with some tweaks a higher sensitivity could be achieved at the expense of lower specificity.

4.1 DeepLabCut performance

Given that the hooves of the cow in the videos were around 15x25 pixels in size, it can be concluded that with a mean error of 4.4 pixels, the predictions by DeepLabCut were accurate and sufficiently captured the positional information of the body parts.

4.2 Data Analysis

In most of the gait features tested, the locomotion score 4 group was significantly different compared to the rest. This was not surprising considering that locomotion score 4 cows walk with such difficulty that they are easily noticed to be lame even to the untrained eye.

Significance tests among locomotion score groups 1,2 and 3 had more mixed results. In general, stance time gait features were significantly different between healthy cows (locomotion score 1) and lame cows (locomotion score 2 and 3), but no difference was found between locomotion score 2 and 3. The step length gait features performed similarly, except that locomotion score 1 and 2 were less distinguishable. This agrees with the results from the force sensor system by Maertens et al [34] because they also found that stance time and step length were features that were significantly different between different locomotion scores.

Step time showed no difference in general among locomotion score groups 1,2 and 3 (although locomotion score 4 was different from the rest). This may seem to contradict the results from Maertens et al [34], but they calculated step time differently by taking the time difference between the middle of each stance phase, not the time difference between hoof lift off and ground strike. The former may be influenced by other features like stance time (a longer stance time increases the time between the middle of each stance phase).

Despite other studies showing asymmetry in the left and right limbs [30-32, 34], this analysis did not find much differences between the locomotion score groups other than the locomotion score 4 group. One possible reason is the lower sampling rate of 30 Hz used in this project compared to others like Maertens et al. (60 Hz) [34] and Alsaaod et al. (400 Hz) [32]. Another possible reason is that most of those studies measured the asymmetry of acceleration [30-32] and not the asymmetry of common gait variables like step time and stance time used in this project.

Back arch was another feature that showed lackluster results despite being an important feature commonly used in manual locomotion scoring. Notably, the average value of the locomotion score 4 group was very different from the rest, but was not found to be significantly different from any of them.

The standard errors from the back arch of locomotion score groups 1, 2, 3 and 4 were 452 pixels, 928 pixels, 1352 pixels and 2038 pixels respectively. Thus, it is likely that the sample size of the locomotion score 4 group was too small to accurately estimate the mean.

Head bobbing and its related feature, nodding showed no difference between the locomotion scores despite head bobbing being a well-known indicator of lameness. Generally, head bobbing is associated with fore limb lameness [48]. Thus, one possible explanation is that the relationship between head bobbing and lameness was changed because this project does not differentiate between fore and hind limb lameness.

4.3 Dimensional reduction and lameness classification

It can be seen in figures 15, 16 and 17 that the 3 resulting features after LDA contained enough information to partially differentiate the different locomotion scores with various degrees of success. Feature 1 alone was capable of completely separating locomotion score 4 cows from the rest. Locomotion score 3 cows seemed partially separable from locomotion score 1 cows, mostly with feature 2, and feature 1 to a lesser degree. Locomotion score 2 cows also seemed partially separable from locomotion score 1 cows, mostly with feature 3, and feature 1 to a lesser degree.

The linear discriminant classifier achieved the highest classification rate of 85.6%. However, the locomotion score 1 group alone contributed 68.2% to 85.6% of this classification rate because the group constitutes about 72% of the whole dataset. Thus, only looking at the classification rate is misleading. The confusion matrices reveal how the classifiers actually performed with each locomotion score group. All 11 locomotion score 4 cows were detected with 100% accuracy. This was not surprising because score 4 was clearly separated from the rest in figure 15. The performance of classification among cows of locomotion score 3 were mixed, with an average of 14.9 out of 25 locomotion score 3 cows (59.6%) were correctly classified as such. A significant 8.6 out of 25 locomotion score 3 cows (34.4%) were wrongly classified as locomotion score 1. As we look at milder states of lameness, it becomes even harder for the classifier to differentiate between healthy and lame cows. Only 28.7 out of 53 locomotion score 2 cows (54.2%) were correctly classified as such. A significant 22 out of 53 locomotion score 2 cows (41.5%) were wrongly classified as locomotion score 1. These classifier performance is also reflected in the visualization in figure 15 where a portion of the locomotion score 3 cows (green circles) and a portion of the locomotion score 2 cows (red asterisks) were both mixed with the cluster of locomotion score 1 cows (blue plus).

Unfortunately, even after examining the confusion matrices, none of the other classifiers really outperformed the linear discriminant classifier. Distribution of classification rates of other classifiers were similar to the linear discriminant classifier in that locomotion score 1 and 4 cows had very high classification rates while locomotion score 2 and 3 had mixed results. Neural network did especially badly with a classification rate of 75.2%. This was most likely the result of overfitting because additional tests revealed that the neural network classifier achieved a 96% classification rate on the training set.

Setting the prior probabilities to be equal for all locomotion scores significantly increased the classification rates of locomotion score 2 and 3 at the expense of decreasing the classification rate of locomotion score 1 cows. This was expected because equal weights were given to all locomotion scores such that the classifiers were no longer biased against low population locomotion score groups.

If we view the locomotion score 1 cows as healthy cows and other cows as lame cows, the initial configuration of prior probabilities resulted in 61.3% sensitivity and 95.3% specificity with the linear discriminant classifier. Setting the prior probabilities of all locomotion scores to be equal resulted in 78.3% sensitivity and 83.4% specificity with the linear discriminant classifier. While the proposed lameness detection method is nowhere near being able to give assessments with 100% confidence, it can be used as an early warning system to alert farmers about cows that are potentially locomotion score 2 or 3 and may need attention. Whether to favor detecting more potentially lame cows early (high sensitivity), or avoid wasting time and labor examining false positive cows (high specificity), depends on their economic values and the preferences of the farmer.

4.4 Recommendations for future work

One way to improve the classification rate is by adding more features that can differentiate between healthy and lame cows. The movements of the fetlock joints and carpal/tarsal joints were not included in the extracted features despite being successfully detected by DeepLabCut. With these key points, the angles of the joints of each limb can be calculated, which may turn into useful gait features as shown by Pluk et al. They showed that joint angle features such as touch and release angles (angle of fetlock joint when touching and leaving the ground) were different between healthy and lame cows [38]. It may also be possible to add more features by adding new key points on other body parts. From the results in this and many other projects, DeepLabCut seems capable of detecting arbitrary objects with high accuracy. Thus, as long as the key points of interest are visually distinguishable enough such that they can be manually labeled by a human, it is very likely that DeepLabCut can handle these new key points.

Another obvious way to improve the classification rate is by increasing the size of the dataset. The overfitting problems encountered by the neural network classifier might have been caused by the small size of the current dataset. Of course, to verify that the proposed method works well in general, additional data should be gathered at different farms with different cows and different farm conditions.

Instead of trying to differentiate between healthy cows and lame cows, some works in literature have proposed a different approach: detecting lameness by detecting changes in gait variables in individual cows [38, 49]. It has been argued that every cow is different and not all cows have the same gait variables when they are lame. Thus, it may be better to monitor individual cows over a long period and detect any significant changes in gait variables, which may indicate that the cow has gotten lame. The methods described in this thesis is very suitable for such a task because as long as a camera can be set up at a good location in a farm, walking cows can be automatically recorded and monitored for a very long period of time. This should be coupled with a system that automatically records the ID number of cows when they leave and enter the barn, like the Lely Grazeway (Lely, the Netherlands), to enable the identification of individual cows.

5. Conclusion

Lameness is a big problem in cattle farms because farmers lack the time and labor to control it. Automatic lameness detection methods proposed in literature so far are costly or unfeasible for commercial farms. This thesis proposed a method to detect lameness by extracting gait features with a camera. Pose estimation of cows was done using DeepLabCut. Gait features were then extracted from the information acquired via pose estimation and analyzed. Notable gait features with significant differences between healthy cows and lame cows were step length and stance time. Classification attempts resulted in an overall classification rate of 85.6%. However, upon further analysis, it was found that the specificity was 95.3%, but sensitivity was only 61.3%. Adjusting the prior probabilities of all the locomotion score groups to be equal significantly boosted the sensitivity to 78.3% at the expense of reducing specificity to 83.4%. Future works should focus on improving the classification rate (both sensitivity and specificity) by adding more relevant features and looking into the change in a cow's gait as it becomes lame.

References

1. Enting, H., et al., *Economic losses due to clinical lameness in dairy cattle.* Livestock production science, 1997. **49**(3): p. 259-267.
2. Barker, Z.E., et al., *Assessment of lameness prevalence and associated risk factors in dairy herds in England and Wales.* Journal of Dairy Science, 2010. **93**(3): p. 932-941.
3. Cook, N.B., *Prevalence of lameness among dairy cattle in Wisconsin as a function of housing type and stall surface.* Journal of the American Veterinary Medical Association, 2003. **223**(9): p. 1324-1328.
4. Espejo, L.A., M.I. Endres, and J.A. Salfer, *Prevalence of lameness in high-producing Holstein cows housed in freestall barns in Minnesota.* Journal of Dairy Science, 2006. **89**(8): p. 3052-3058.
5. Šárová, R., et al., *Farm managers underestimate lameness prevalence in Czech dairy herds.* Animal Welfare, 2011. **20**(2): p. 201-204.
6. Sadiq, M.B., et al., *Prevalence of lameness, claw lesions, and associated risk factors in dairy farms in Selangor, Malaysia.* Tropical Animal Health and Production, 2017. **49**(8): p. 1741-1748.
7. Bruijnis, M.R.N., H. Hogeveen, and E.N. Stassen, *Assessing economic consequences of foot disorders in dairy cattle using a dynamic stochastic simulation model.* Journal of Dairy Science, 2010. **93**(6): p. 2419-2432.
8. Bruijnis, M.R.N., et al., *Assessing the welfare impact of foot disorders in dairy cattle by a modeling approach.* Animal, 2012. **6**(6): p. 962-970.
9. Leach, K.A., et al., *Working towards a reduction in cattle lameness: 1. Understanding barriers to lameness control on dairy farms.* Research in Veterinary Science, 2010. **89**(2): p. 311-317.
10. Dyce, K.M., W.O. Sack, and C.J.G. Wensing, *Textbook of veterinary anatomy-E-Book*. 2009: Elsevier Health Sciences.
11. Leach, D., *Recommended terminology for researchers in locomotion and biomechanics of quadrupedal animals.* Cells Tissues Organs, 1993. **146**(2-3): p. 130-136.
12. Meyer, S.W., M.A. Weishaupt, and K.A. Nuss, *Gait Pattern of Heifers Before and After Claw Trimming: A High-Speed Cinematographic Study on a Treadmill.* Journal of Dairy Science, 2007. **90**(2): p. 670-676.
13. Flower, F.C., D.J. Sanderson, and D.M. Weary, *Hoof pathologies influence kinematic measures of dairy cow gait.* Journal of Dairy Science, 2005. **88**(9): p. 3166-3173.
14. Sanders, A.H., J.K. Shearer, and A. De Vries, *Seasonal incidence of lameness and risk factors associated with thin soles, white line disease, ulcers, and sole punctures in dairy cattle.* Journal of Dairy Science, 2009. **92**(7): p. 3165-3174.
15. Greenough, P.R., *Bovine Laminitis and Lameness*. Bovine Laminitis and Lameness. 2007.
16. Ranjbar, S., et al., *Identifying risk factors associated with lameness in pasture-based dairy herds.* Journal of Dairy Science, 2016. **99**(9): p. 7495-7505.
17. Green, L., et al., *The impact of clinical lameness on the milk yield of dairy cows.* Journal of dairy science, 2002. **85**(9): p. 2250-2256.
18. Hernandez, J., et al., *Comparison of milk yield in dairy cows with different degrees of lameness.* Journal of the American Veterinary Medical Association, 2005. **227**: p. 1292-6.
19. Weigele, H.C., et al., *Moderate lameness leads to marked behavioral changes in dairy cows.* Journal of Dairy Science, 2018. **101**(3): p. 2370-2382.
20. Lucey, S., G. Rowlands, and A. Russell, *The association between lameness and fertility in dairy cows.* The veterinary record, 1986. **118**(23): p. 628-631.
21. Collick, D.W., W.R. Ward, and H. Dobson, *Associations between types of lameness and fertility.* Veterinary Record, 1989. **125**(5): p. 103.

22. Booth, C.J., et al., *Effect of Lameness on Culling in Dairy Cows.* Journal of Dairy Science, 2004. **87**(12): p. 4115-4122.
23. Sprecher, D.J., D.E. Hostetler, and J.B. Kaneene, *A lameness scoring system that uses posture and gait to predict dairy cattle reproductive performance.* Theriogenology, 1997. **47**(6): p. 1179-1187.
24. Manson, F.J. and J.D. Leaver, *The influence of concentrate amount on locomotion and clinical lameness in dairy cattle.* Animal Production, 1988. **47**(2): p. 185-190.
25. Breuer, K., et al., *Behavioural response to humans and the productivity of commercial dairy cows.* Applied Animal Behaviour Science, 2000. **66**(4): p. 273-288.
26. Tadich, N., E. Flor, and L. Green, *Associations between hoof lesions and locomotion score in 1098 unsound dairy cows.* Veterinary Journal, 2010. **184**(1): p. 60-65.
27. Schlageter-Tello, A., et al., *Manual and automatic locomotion scoring systems in dairy cows: a review.* Prev Vet Med, 2014. **116**(1-2): p. 12-25.
28. Schlageter-Tello, A., et al., *Effect of merging levels of locomotion scores for dairy cows on intra- and interrater reliability and agreement.* Journal of Dairy Science, 2014. **97**(9): p. 5533-5542.
29. Thomsen, P.T., L. Munksgaard, and F.A. Tøgersen, *Evaluation of a Lameness Scoring System for Dairy Cows.* Journal of Dairy Science, 2008. **91**(1): p. 119-126.
30. Pastell, M., et al., *A wireless accelerometer system with wavelet analysis for assessing lameness in cattle.* Biosystems Engineering, 2009. **104**(4): p. 545-551.
31. Chapinal, N., et al., *Measurement of acceleration while walking as an automated method for gait assessment in dairy cattle.* Journal of Dairy Science, 2011. **94**(6): p. 2895-2901.
32. Alsaaod, M., et al., *The cow pedogram—Analysis of gait cycle variables allows the detection of lameness and foot pathologies.* Journal of Dairy Science, 2017. **100**(2): p. 1417-1426.
33. Rajkondawar, P.G., et al., *A system for identifying lameness in dairy cattle.* Applied Engineering in Agriculture, 2002. **18**(1): p. 87-96.
34. Maertens, W., et al., *Development of a real time cow gait tracking and analysing tool to assess lameness using a pressure sensitive walkway: The GAITWISE system.* Biosystems Engineering, 2011. **110**(1): p. 29-39.
35. Flower, F.C. and D.M. Weary, *Effect of hoof pathologies on subjective assessments of dairy cow gait.* Journal of Dairy Science, 2006. **89**(1): p. 139-146.
36. Blackie, N., et al., *Impact of lameness on gait characteristics and lying behaviour of zero grazed dairy cattle in early lactation.* Applied Animal Behaviour Science, 2011. **129**(2-4): p. 67-73.
37. Poursaberi, A., et al., *Real-time automatic lameness detection based on back posture extraction in dairy cattle: Shape analysis of cow with image processing techniques.* Computers and Electronics in Agriculture, 2010. **74**(1): p. 110-119.
38. Pluk, A., et al., *Automatic measurement of touch and release angles of the fetlock joint for lameness detection in dairy cattle using vision techniques.* Journal of Dairy Science, 2012. **95**(4): p. 1738-1748.
39. Zhao, K., et al., *Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique.* Computers and Electronics in Agriculture, 2018. **148**: p. 226-236.
40. Mathis, A., et al., *DeepLabCut: markerless pose estimation of user-defined body parts with deep learning.* Nature Neuroscience, 2018. **21**(9): p. 1281-1289.
41. Insafutdinov, E., et al. *Deepercut: A deeper, stronger, and faster multi-person pose estimation model*. in *European Conference on Computer Vision*. 2016. Springer.
42. He, K., et al. *Deep Residual Learning for Image Recognition*. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
43. Deng, J., et al. *ImageNet: A large-scale hierarchical image database*. in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009.

44. Suzuki, S. and K. be, *Topological structural analysis of digitized binary images by border following.* Computer Vision, Graphics, and Image Processing, 1985. **30**(1): p. 32-46.
45. Redmon, J., et al. *You only look once: Unified, real-time object detection*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
46. Lin, T.-Y., et al. *Microsoft COCO: Common Objects in Context*. 2014. Cham: Springer International Publishing.
47. Nath, T., et al., *Using DeepLabCut for 3D markerless pose estimation across species and behaviors.* Nature Protocols, 2019: p. 1-25.
48. Whay, H., *Locomotion scoring and lameness detection in diary cattle.* In Practice, 2002. **24**(8): p. 444-449.
49. Viazzi, S., et al., *Analysis of individual classification of lameness using automatic measurement of back posture in dairy cattle.* Journal of Dairy Science, 2013. **96**(1): p. 257-266.

Appendix

A1. Full list of gait features

Table – Full list of gait features

Gait Feature	Description
LF step time 1	Step time of first step of left front limb
LF step time 2	Step time of second step of left front limb
LF step time 3	Step time of third step of left front limb
LF step time mean	Mean step time of left front limb
RF step time 1	Step time of first step of right front limb
RF step time 2	Step time of second step of right front limb
RF step time 3	Step time of third step of right front limb
RF step time mean	Mean step time of right front limb
LH step time 1	Step time of first step of left hind limb
LH step time 2	Step time of second step of left hind limb
LH step time 3	Step time of third step of left hind limb
LH step time mean	Mean step time of left hind limb
RH step time 1	Step time of first step of right hind limb
RH step time 2	Step time of second step of right hind limb
RH step time 3	Step time of third step of right hind limb
RH step time mean	Mean step time of right hind limb
Overall step time mean	Mean step time of all limbs
Minimum step time	Shortest step time
Maximum step time	Longest step time
LF step length 1	Step length of first step of left front limb
LF step length 2	Step length of second step of left front limb
LF step length 3	Step length of third step of left front limb
LF step length mean	Mean step length of left front limb
RF step length 1	Step length of first step of right front limb
RF step length 2	Step length of second step of right front limb
RF step length 3	Step length of third step of right front limb
RF step length mean	Mean step length of right front limb
LH step length 1	Step length of first step of left hind limb
LH step length 2	Step length of second step of left hind limb
LH step length 3	Step length of third step of left hind limb
LH step length mean	Mean step length of left hind limb
RH step length 1	Step length of first step of right hind limb
RH step length 2	Step length of second step of right hind limb
RH step length 3	Step length of third step of right hind limb
RH step length mean	Mean step length of right hind limb
Overall step length mean	Mean step length of all limbs
Minimum step length	Shortest step length
Maximum step length	Longest step length
LF stance time 1	Stance time of first step of left front limb
LF stance time 2	Stance time of second step of left front limb

LF stance time 3	Stance time of third step of left front limb
LF stance time mean	Mean stance time of left front limb
RF stance time 1	Stance time of first step of right front limb
RF stance time 2	Stance time of second step of right front limb
RF stance time 3	Stance time of third step of right front limb
RF stance time mean	Mean stance time of right front limb
LH stance time 1	Stance time of first step of left hind limb
LH stance time 2	Stance time of second step of left hind limb
LH stance time 3	Stance time of third step of left hind limb
LH stance time mean	Mean stance time of left hind limb
RH stance time 1	Stance time of first step of right hind limb
RH stance time 2	Stance time of second step of right hind limb
RH stance time 3	Stance time of third step of right hind limb
RH stance time mean	Mean stance time of right hind limb
Overall stance time mean	Mean stance time of all limbs
Minimum stance time	Shortest stance time
Maximum stance time	Longest stance time
Front step length asymmetry 1	Absolute difference between LF step length 1 and RF step length 1
Front step length asymmetry 2	Absolute difference between LF step length 2 and RF step length 2
Front step length asymmetry 3	Absolute difference between LF step length 3 and RF step length 3
Front step length asymmetry mean	Mean front step length asymmetry
Hind step length asymmetry 1	Absolute difference between LH step length 1 and RH step length 1
Hind step length asymmetry 2	Absolute difference between LH step length 2 and RH step length 2
Hind step length asymmetry 3	Absolute difference between LH step length 3 and RH step length 3
Hind step length asymmetry mean	Mean hind step length asymmetry
Front step time asymmetry 1	Absolute difference between LF step time 1 and RF step time 1
Front step time asymmetry 2	Absolute difference between LF step time 2 and RF step time 2
Front step time asymmetry 3	Absolute difference between LF step time 3 and RF step time 3
Front step time asymmetry mean	Mean front step time asymmetry
Hind step time asymmetry 1	Absolute difference between LH step time 1 and RH step time 1
Hind step time asymmetry 2	Absolute difference between LH step time 2 and RH step time 2
Hind step time asymmetry 3	Absolute difference between LH step time 3 and RH step time 3
Hind step time asymmetry mean	Mean hind step time asymmetry
Front stance time asymmetry 1	Absolute difference between LF stance time 1 and RF stance time 1
Front stance time asymmetry 2	Absolute difference between LF stance time 2 and RF stance time 2
Front stance time asymmetry 3	Absolute difference between LF stance time 3 and RF stance time 3
Front stance time asymmetry mean	Mean front stance time asymmetry

Hind stance time asymmetry 1	Absolute difference between LH stance time 1 and RH stance time 1
Hind stance time asymmetry 2	Absolute difference between LH stance time 2 and RH stance time 2
Hind stance time asymmetry 3	Absolute difference between LH stance time 3 and RH stance time 3
Hind stance time asymmetry mean	Mean hind stance time asymmetry
Nose bobbing	Variance in pixel^2 of vertical position of head
Nodding	Variance of angle between top of the head and nose
Back arch	Radius in pixel of circle fitted on the three back key points

A2. ANOVA and Bonferroni test

```
% features: 313*84 array, 84 gait features from 313 cow videos
% labels: 313*1 array, true locomotion score of 313 cows
for i = 1:size(features,2)
    % ANOVA
    [p,t,stats] = anova1(features(:,i),labels);

    % Comparison between groups using Bonferroni
    [c,m,h,nms] = multcompare(stats,'CType','bonferroni');
end
```

A3. Dimensional Reduction, classification and cross validation

```
% Create dataset with features and their corresponding labels
% features: 313*84 array, 84 gait features from 313 cow videos
% labels: 313*1 array, true locomotion score of 313 cows
locomotion_data = prdataset(features,labels);

% Set prior probabilities to the frequency of each locomotion score group
locomotion_data = setprior(locomotion_data,getprior(locomotion_data));

% Dimensional reduction using linear discriminant analysis (Fisher mapping)
dim_reduc_map = fisherm(locomotion_data);
locomotion_data_reduc = locomotion_data*dim_reduc_map;

% Scatter plots for figures 13-15
scatterd(locomotion_data_reduc,'legend');
scatterd(locomotion_data_reduc(:,[2,3]),'legend');
scatterd(locomotion_data_reduc(:,[1,3]),'legend');

% Initialize linear discriminant classifier, quadratic discriminant
% classifier, K=3 nearest neighbors classifier, decision tree classifier,
% support vector machine, parzen classifier and neural network classifier
classifiers_untrained = {ldc,qdc,knnc([],3),treec,svc,parzenc,neurc};

% Perform cross validation 10 times
iterations = 10;

% Initialize array to store error rate of each classifier over 10
% iterations
err_list = zeros(iterations,length(classifiers_untrained));

% Initialize array to store confusion matrix of each classifier over 10
% iterations
confusions = zeros(4,4,iterations,length(classifiers_untrained));


for i=1:iterations
    % 10-fold cross validation to calculate error rate and predicted labels
    [err,~,nlab_out] = prcrossval(locomotion_data_reduc,classifiers_untrained,10);
    err_list(i,:) = err;

    % Calculate confusion matrices
    for j = 1:length(classifiers_untrained)
        [confusion,~,~,~] = confmat(labels,nlab_out{j});
        confusions(:,:,i,j) = confusion;
    end
end

% Calculate mean and standard deviation of error rates
err_mean = mean(err_list,1);
err_std = std(err_list,0,1);

% Calculate mean and standard deviation of confusion matrices
confusion_mean = mean(confusions,3);
confusion_std = std(confusions,0,3);
```