Development of a prediction model for automatic lameness detection in dairy cows

Isabella Lorenzini[1], Katharina Grimm[1], Bernhard Haidn[1], Eyal Misha[2]

[1]Bayerische Landesanstalt für Landwirtschaft, Institut für Landtechnik und Tierhaltung, Deutschland

[2]ENGS Dairy Solutions, Israel

Due to the average increase in the size of dairy farms, there has been a reduction in the time available to farmers for animal observation. Animal observation is an important part of the daily work routine; it is essential for the detection of health issues within the heard, for example cases of lameness. In 2015 9,1% of all losses on Bavarian dairy cattle were due to claw and limb disorders; in order to avoid unnecessary pain for the affected animals as well as economic loss for the farmers, it is essential to recognise lameness in its early stages. In a current project at the Institute for Agricultural Engineering and Animal Husbandry in Grub, Bavaria, data concerning animal behaviour and claw health will be collected for a year on four farms. It has been found in in a previous project at the same institute that the lying and feeding behaviour of the animals are very susceptible to changes due to lameness even in its early stages. The aim of the current project is to expand on these results and verify them on a larger number of animals. The behavioural data will be automatically collected by pedometers, whilst claw health will be extensively documented by carrying out regular locomotion scores followed by a clinical examination of the lame animals. The pedometers contain three-dimensional accelerometers that regularly register the position of the animal's front limb in space and thus identify whether the animal is lying, standing or moving. The feeding behaviour is recorded by means of an induction loop which is installed at the feeding table and that recognises every animal which steps into the induced magnetic field. All the collected data will then be summarized into one daily dataset containing the behavioural data as well as the claw health data for each animal. Locomotion scoring is the standard method for evaluating claw health; in this project a three-point locomotion scoring system was used, by which animals are either considered sound (1), suspected lame (2) or lame (3) based on various gait parameters. The locomotion scoring was performed on video recordings of the animals coming out of the milking parlour. The aim of this investigation is to use the behavioural changes of lame animals to develop an algorithm that could help farmers recognise lame animals earlier.

Die durchschnittliche Größenzunahme der Milchviehbetriebe hat zu einer Verminderung der für die Einzeltierbeobachtung zur Verfügung stehenden Zeit geführt. Die Einzeltierbeobachtung im Arbeitsalltag ist wichtig um Hinweise auf gesundheitliche Störungen, zum Beispiel Lahmheiten, zu erkennen. Nach LKV-Bayern waren 2015 mit 9,1 % aller Abgangsursachen die Klauen- und Gliedmaßenerkrankungen von besonderer Bedeutung; um unnötige Schmerzen der Tiere sowie die durch Lahmheit verursachten wirtschaftlichen Verluste zu vermeiden spielt die Früherkennung von Klauenerkrankungen eine wesentliche Rolle. In einem aktuellen Projekt am Institut für Landtechnik und Tierhaltung werden auf 4 Praxisbetrieben über ein Jahr Verhaltensdaten erfasst und gleichzeitig die Klauengesundheit aller Tiere dokumentiert. In einem Vorgängerprojekt wurde herausgefunden, dass vor allem das Liege- und Fressverhalten der Tiere sich in den frühen Stadien einer Lahmheit signifikant verändern. Diese Erkenntnisse sollen im aktuellen Projekt vertieft und an einer größeren Tierzahl überprüft werden. Die Verhaltensdaten werden über Pedometer gemessen, während die Klauengesundheit durch regelmäßige Gangbeurteilungen und anschließende Untersuchungen der lahmen Tiere ausführlich dokumentiert wird. Die Pedometer enthalten Beschleunigungssensoren, die in regelmäßigen Zeitabständen die Lage der Vordergliedmaße registrieren und somit das Verhalten Liegen, Stehen oder Aktivität erfassen. Das Fressverhalten wird über eine am Futtertisch angebrachte Induktionsschleife erfasst, die die Einzeltiere beim Betreten des induzierten magnetischen Feldes registriert. Die somit erfassten Daten werden zu einem aus den Verhaltensdaten sowie aus einem Locomotionscore bestehenden Tagesdatensatz für jedes Einzeltier zusammengefasst. Das Locomotionscore ist die Standardmethode zur Erfassung des Klauengesundheitsstatus. Dabei handelt es sich um ein Dreipunkte-System bei dem verschiedene Gangparameter ausgewertet werden und das Tier als gesund (1), verdächtig (2) oder lahm (3) eingestuft wird. Das Locomotionscore aller Tiere wird über Videoaufnahmen der Melkzeiten ausgeführt. Das Ziel dieser Untersuchungen ist es mithilfe der gemessenen Abweichungen im Verhalten lahmer

Tiere einen Algorithmus zu entwickeln, der Landwirte dabei helfen soll, im Arbeitsalltag lahme Tiere leichter und früher zu erkennen.

1. INTRODUCTION

Lameness is defined as a sign of a structural or functional disruption that affects one or more limbs and is evident when walking and/or standing [1].

The causes for lameness in dairy cows can be found proximally or distally in the limbs, although claw diseases represent the most common trigger for an abnormal gait pattern and behavioural changes [2]. Even though claw diseases are multifactorial, the underlying pathological mechanism is always a weakening of the structural integrity of the claw's horn and surrounding tissues [3].

Both intrinsic factors, such as metabolic illnesses and the lactation status, and external factors, such as management, weather conditions, housing etc., play an important part in the development of claw disorders [3]. The frequency and the timing of the claw trimming are also of great significance when it comes to the prevention of lameness, as is also the quality of the feed and the ration composition.

Lameness in dairy cows is a major welfare problem, causing pain to the animal, and at the same time it is also a significant economic factor; in 2015 in Bavaria 9,1% of all losses on dairy cattle were caused by lameness [4]. Despite the technological innovation and advancements in the field of genetics, there has been no significant reduction in the average lameness prevalence on European dairy farms in the last 20 years [5]. These statistics highlight the need for further research into early lameness recognition, which is the key to reducing economic losses and avoiding unnecessary pain to the affected animals.

One of the limiting factors in early lameness recognition is the time invested by farmers in animal observation; the structural transformation that has taken place in the last few years in German dairy farms has resulted in the increase of the average number of dairy cows per farm in Germany from only 16 animals per farm in 1986 to the 61 animals per farm today [6]. This increase negatively affects the time available to farmers to perform animal observation and lameness assessment.

Locomotion scoring systems are considered the gold standard for lameness detection. They use gait and posture as parameters to assess the presence of irregularities in the locomotion of the animal. There are many different types of locomotion scoring systems; some are more suitable for on-farm use, whilst others are relevant only for research purposes. Schlageter-Tello et al. [7] reviewed various different locomotion scoring methods and highlighted their common problem, namely their variable agreement and reliability. The percentage of agreement between two observers estimates range from 17 to 42 % for the locomotion scoring system according to Manson & Leaver [10] [11], 37 % for O'Callaghan et al.'s scoring system [12], and 83 % for the locomotion scoring system according to Sprecher et al. [9] [8].

Channon et al. determined the percentage of agreement between multiple observers using the Manson & Leaver score [10] and concluded that when the agreement was calculated using a lame/not lame rating threshold only, 88 % of observers agreed, whilst only 33.3 % agreed when using all 9 different scores [11], indicating that by reducing the number of available scores in a locomotion scoring system, a higher percentage of agreement and thus of reliability can be achieved between observers. For this reason a new score with fewer levels was developed in the context of a larger project at the Bavarian State Research Centre for Agriculture; the locomotion scoring system according to Grimm & Lorenzini [13]. The aim of this study is to verify the interrater and intrarater reliability of this three point locomotion scoring system and to assess its accuracy with the prospect of using it as a reference system for the development of an automatic lameness detection system. This system could help farmers recognise lame animals earlier, thus addressing both the welfare and economic issues connected to lameness in dairy cows.

2. ANIMALS, MATERIALS AND METHODS

2.1 Farms and animals

The aim of the current project at the Bavarian State Research Centre for Agriculture is to develop an algorithm that could automatically detect lameness in dairy cows by using data from activity sensors. In a previous project at the same institute Grimm et al. [14] determined that some behavioural traits such as the number of lying bouts (day/night ratio) and the feeding duration per meal are influenced by lameness even in its early stages. The current project expands on these results by combining behavioural data with claw health data of 375 animals over one year.

The data collection started in April 2017, spans over twelve months and involves 5 farms of which one research farm. All of the farms have loose housing; three of them have milking parlours and two have a milking robot, with an average of 83 (minimum = 68, maximum = 103) milking Simmental cows each.

2.2 Materials and methods

At the beginning of the project all lactating animals were fitted with "Track a Cow" Pedometers (ENGS Dairy Solutions, Israel); these three-dimensional accelerometers are attached to the right front limb of each animal and continuously record lying behaviour, activity and feeding behaviour. The feeding behaviour is registered by means of an induction loop which is installed in the feeding alley along the feeding table; every time an animal steps into the magnetic field generated by the induction loop it's recognised by the system and the information is then sent to a transmitter that displays it for the end user in a software environment and saves it in an MS Office Access Database.

The claw health is closely followed and documented by means of a locomotion score for each animal of the whole herd every two weeks. For this purpose cameras were installed at the milking parlour or milking robot exits, allowing for an indirect observation of the animals through video recordings. The choice of indirect video observation as opposed to direct observation at milking times was preferred as previous experience suggested an influence of the observer on the animals, who would often try to hide their lameness thus compromising the accuracy of the score.

2.3 Locomotion score according to Grimm & Lorenzini

The starting point and reference method for the claw health status of the animals during the project is the three point locomotion scoring system (1 = sound, 2 = unsound, 3 = lame) according to Grimm & Lorenzini [13]. This system only has 4 locomotion traits, chosen based on experience values and on an analysis of other commonly used scoring systems and locomotion traits, such as those described by Schlageter-Tello et al. in their study on the relation between locomotion traits and locomotion score [15]. The specific score for an animal is determined in two steps; firstly, gait, as the more general trait, is assessed as either regular or irregular. The animals with an irregular gait are directly classified as lame (score 3). In a second step, the locomotion of the animals with a regular, symmetric looking gait is examined for three more specific traits; back arch, compensatory posture and head bob. If one or more of these traits are present, the animal is considered a score 2 (unsound). If none of the traits is present, it is considered score 1 (sound). To increase the accuracy of the scoring system, the locomotion scoring is followed by a third step, consisting in a clinical examination. The claws of unsound (2) animals undergo a pain reaction test using hoof pincers. If the outcome is positive, regardless of whether there are any other clinical findings or not, these animals will be considered lame (3) and treated as such. The claws of all lame (3) animals are examined and treated if needed and the findings documented. The three point locomotion score can be represented graphically as a flow chart with a dichotomous decision boundary for each variable (**Fehler! Verweisquelle konnte nicht gefunden werden.**).

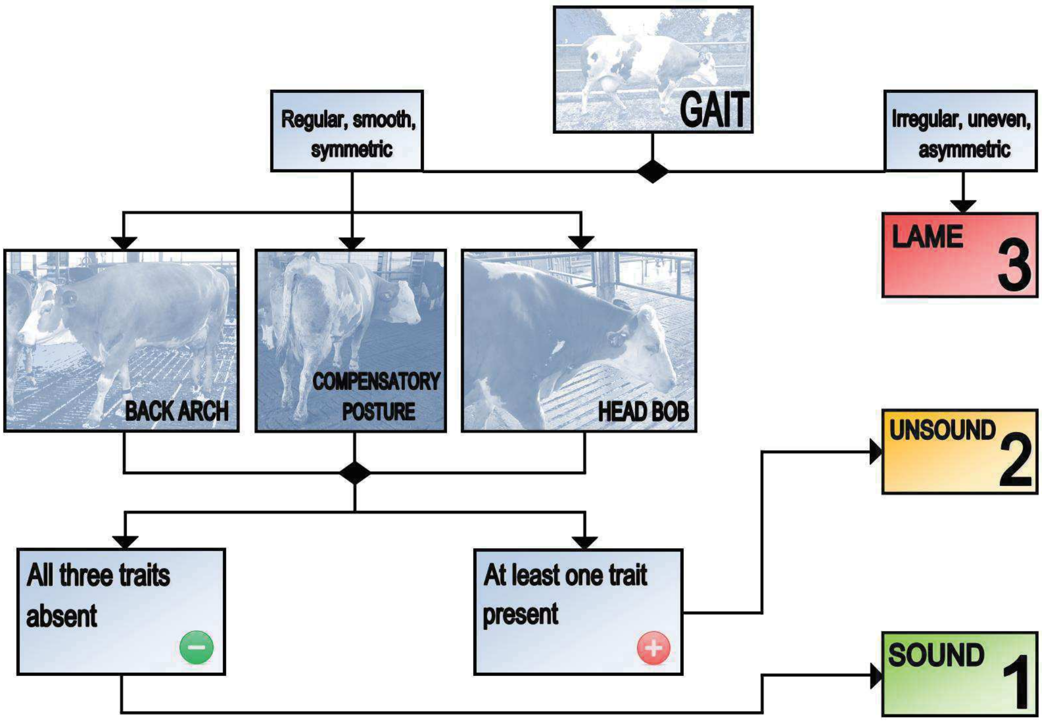

Figure 1: Locomotion score according to Grimm & Lorenzini

For this study, locomotion scores were performed on video as well as by direct observation; the scores performed by direct observation involved two observers scoring simultaneously without consulting each other. Two observers also simultaneously scored animals on video, and then repeated the viewing of the same video on a different day.

2.4 Inter- and intra-observer reliability

All data was analysed using R Studio (1.0.136 Version) [16]. To quantify the level of inter-observer agreement for the three point locomotion score both the percentage of agreement (PA), Cohen's Kappa (κ) and Kendall's Coefficient of Concordance (W) were calculated. Cohen's Kappa is a statistical measure that represents the level of agreement between two raters assessing

the same objects on a nominal or ordinal scale, and has a value between -1 and 1, where 0 corresponds to a level of agreement equivalent to chance and 1 represents perfect agreement. Kendall's Coefficient of Concordance on the other hand is a "measure of the agreement among several (*m*) quantitative or semi quantitative variables that are assessing a set of *n* objects of interest" [17] and also has a value between -1 and 1, where 1 indicates perfect agreement and -1 indicates a negative association. In the case of this study, the variables are the two observers and the objects observed are the animals being scored. The data is considered separately for video and direct observation and for different observer combinations, as there were in total three observers that scored the animals in pairs simultaneously.
The second reliability parameter which was tested for the three point locomotion score according to Grimm & Lorenzini [13] is the intra-observer reliability, which is a statistical measure that refers to an observer's consistency in evaluating the same object(s) twice. The same statistical tests were performed as for the inter-observer reliability.
To test the intra-observer reliability two observers independently performed locomotion scorings of cows on video recordings and rescored the same video recordings one week later. Also, a locomotion scoring performed by two observers by direct observation was then repeated by scoring the video recordings from after the same milking session.

2.5 Fortnightly locomotion score (FS) and Daily locomotion score (DS)

The locomotion score is performed fortnightly for all animals in this study as the reference for claw health. The aforementioned pain reaction test was only preformed when cows were scored 2 for three FS in a row (i. e. for 6 weeks) for reasons of feasibility due to the large number of animals that were scored at a time. To find the exact lameness onset date of the animals which were found to be lame in the FS, the daily video recordings of after each milking were watched and the animal was assigned a score for every day until the day was found, when the animal moved soundly, i. e. scored 1. This daily locomotion score (DS) enables a precise reconstruction of the development of the lameness from its onset to its discovery. The time after the FS 3 was not further analysed, as the prediction of lameness was the aim of the study and not the examination of the healing process.
Seeing as the ultimate aim of the study is to have a score per day per animal, at the beginning of the data handling process all the remaining days between two FS all animals will be assigned a daily locomotion score by interpolation on the basis of rules which have been previously defined. For instance, if an animal has the same score for two FS running, the fourteen days between the two FS will be filled with the same score. If an animal has score 1 and a score 2 at the following FS, the first seven days of the interval will be filled with a score 1, and the next seven with a score 2. If an animal has a score 3 on one FS and a score 1 on the next FS the days between the two FS will be divided into thirds and the scores assigned accordingly.

3. RESULTS

3.1 Locomotion scoring and clinical examination of unsound animals

Between April 2017 and January 2018, n = 283 clinical examinations of lame or unsound animals were performed. The claws of the animals which were scored 3 in the FS were examined and hoof trimming was performed if necessary, whereas the animals which were scored 2 for six weeks in a row, i.e. three FS in a row, were tested for a pain reaction with hoof pincers. In the case of a positive pain reaction, or in the presence of any evident clinical findings, the animal's score was then corrected to a score 3. The animals whose pain reaction test resulted negative and did not present any clinical findings remained a score 2. represents the results of these clinical examinations

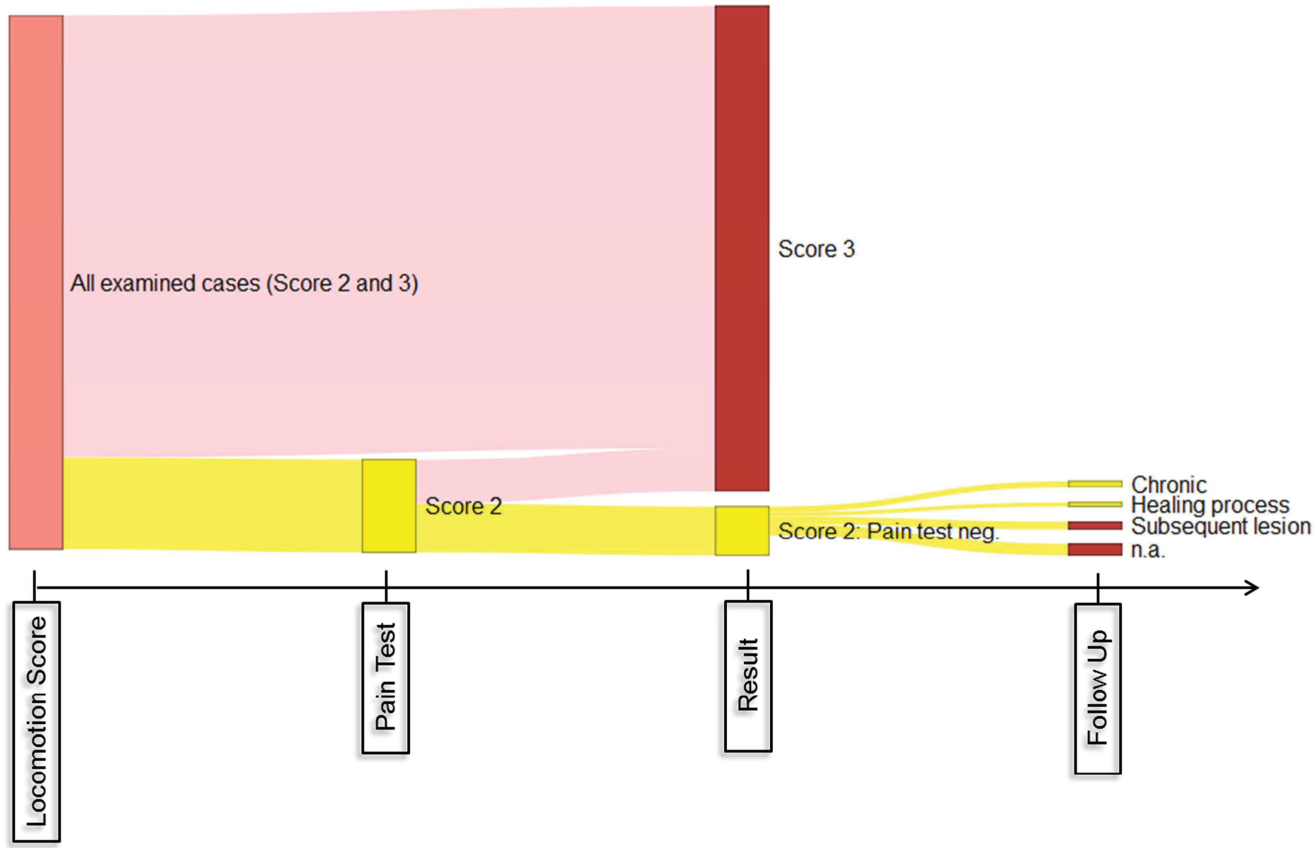

Figure 2: Sankey diagram illustrating the number of animals with score 2 and 3 as a result of the fortnightly locomotion scoring (FS), subsequent pain test and follow up.

Of the n = 283 cases, n = 234 were a score 3 to start and thus presented with clinical findings. Of the n = 49 remaining cases initially scored 2, 47 % (n = 23 cases) had a positive reaction to the pain test or had clinical findings and were therefore corrected to a score 3. 53 % on the other hand, had neither evident claw lesions nor reacted to the pain test with hoof pincers. These n = 26 cases correspond to n = 15 animals, as some of them were examined more than once in the above-mentioned period of time. The chronological sequence of the FS for these animals was analysed both before and after the negative pain test as illustrated in Figure 3. n = 4 animals developed a clinically apparent lameness after the first negative pain reaction (e. g. cow 457), and n = 2 animals were in the process of healing from a previous claw lesion (see cow 357). n = 3 animals were considered chronic, repeatedly presenting with scores 2 and 3, like cow number 499 in Figure 3. n = 5 animals' gait incongruences could not be explained through claw lesions as none were found after multiple examinations and pain tests. The locomotion score timeline of these animals typically resembles the timeline of cow number 485 in Figure 3, constantly switching between a score 1 and a score 2. One animal underwent the pain reaction test in January, so not enough data is present to make a hypothesis with regards to the cause of recurring gait or posture irregularities.

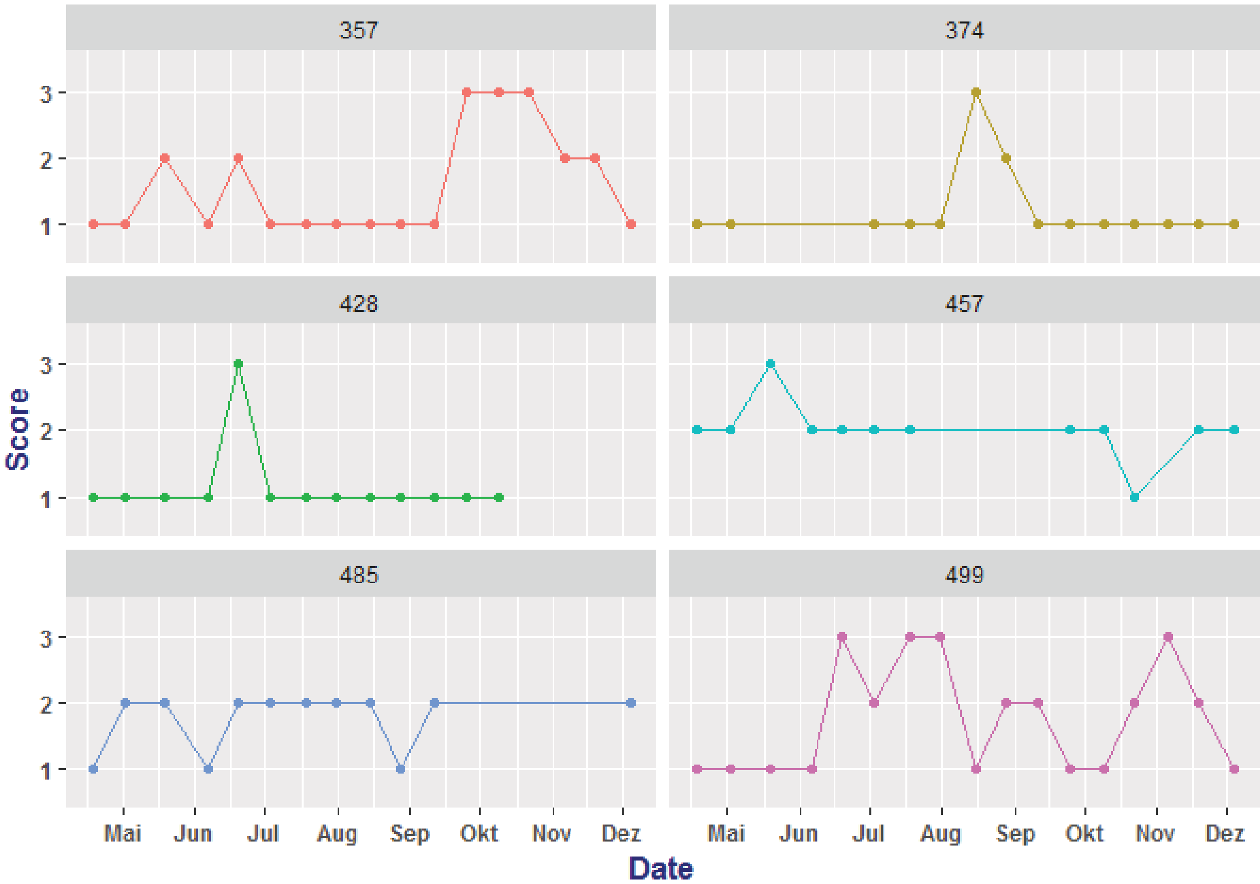

Figure 3: Locomotion score timelines for n = 6 chosen animals

3.2 Inter-observer reliability

The use of well-established locomotion scoring systems in previous projects proved problematic for various reasons; the level of agreement between two observers (inter-observer) as well as the level of agreement for one observer and multiple viewings of the same score (intra-observer) was not satisfactory and generally not very high, an observation which corresponds to Schlageter-Tello et al.'s findings in their review of manual locomotion scoring systems [7]. In particular the use of Sprecher et al.'s [9] five point locomotion scoring system proved difficult with Simmental cows in our recent research, as they do not consistently show lameness traits used in the system such as back arching when standing or moving. For these reasons a new score was developed to use in the current project (see also paragraph 2.3) and the level of agreement between raters was analysed.

In total n = 475 animals were scored by three different observers in pairs; the results show a slightly higher percentage of overall agreement for the video observation (80.1 %) compared to direct observation (77.9 %), although the percentage of agreement (PA) varies depending on the observer pairs (see Table 1). Kendall's Concordance Coefficient is highest ($W = 0.889$) for observer 1 and 2 in the direct observation and is considered "almost perfect" [18], whilst the lowest is between observer 1 and observer 3 ($W = 0.578$) in the direct observation, where it is only "moderate". The values calculated for Cohen's Kappa follow the same tendency as Kendall's Concordance Coefficient, but with slightly lower values (see Table 1). The results are graphically represented in Figure 4; in the bar plot there is no distinction between observer pairs.

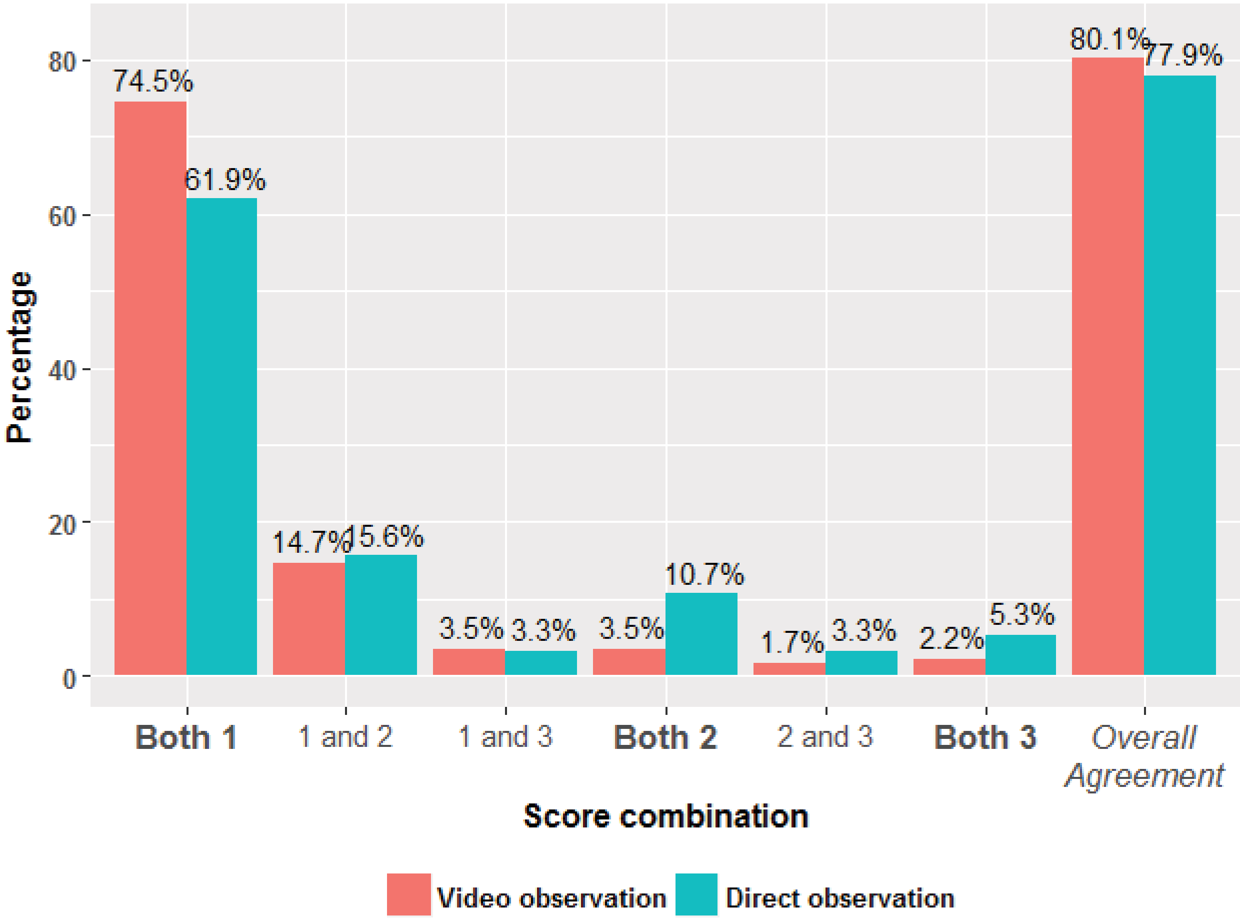

Figure 4: Bar plot showing the percentage of inter-observer agreement for different score combinations and the overall inter-observer agreement for both video and direct observation. n (observers) = 3, n (scored animals) = 475

Table 1: Percentage of agreement (PA) and inter-observer reliability (W and κ) for three observers.

	Observer 1 & Observer 2	Observer 1 & Observer 3	All observers
PA video observation	78 % (n = 159)	84.7 % (n = 72)	80.1 % (n = 231)
PA direct observation	85 % (n = 147)	67 % (n = 97)	77.9 % (n = 244)
W video observation	0.640 (n = 159)	0.784 (n = 72)	0.680 (n = 231)
W direct observation	0.889 (n = 147)	0.578 (n = 97)	0.786 (n = 244)
κ video observation	0.354 (n = 159)	0.635 (n = 72)	0.423 (n = 231)
κ direct observation	0.772 (n = 147)	0.221 (n = 97)	0.606 (n = 244)

3.3 Intra-observer reliability

The results of the statistical tests carried out to measure the level of intra-observer agreement, listed in Table 2, show a high percentage of agreement for the repeated viewing of the video recordings, and an overall intra-observer reliability which at $W = 0.768$ can be considered “substantial” [18] and $\kappa = 0.6$ which can be regarded as “moderate” [18]. The observers were not considered individually for these tests.

Table 2: Percentage of agreement (PA) for two observers and multiple viewings and intra-observer reliability (W and κ)

	Video observation	Direct obs./ video obs.	Overall
Percentage of agreement	90.3 % (n = 238)	72.4 % (n = 192)	82.3 % (n = 430)
Kendall’s W	0.856 (n = 238)	0.681 (n = 192)	0.768 (n = 430)
Cohen’s Kappa	0.77 (n = 238)	0.419 (n = 192)	0.600 (n = 430)

In Figure 5 the distribution of the scores in the different viewings is represented graphically in a jitter plot. Each dot represents an animal (n = 430) and its position in the graph represents whether the scores assigned to it in the two viewings were concordant or discordant. The green dots are the concordant scores, i.e. the score on the first and the second viewing given by the same observer were the same. The highest number of dots is to be found where the animal was a score 1 in both viewings, although this also depends on the fact that score 1 had an overall higher frequency in the locomotion scoring sessions. The highest number of discordant scores can be found at the score 1 with score 2 combinations. Interestingly, more animals had a score difference (meaning the absolute value of the score at the first viewing minus the score at the second viewing) of 2, than animals which were scored respectively a 2 or a 3 on the first viewing and a 3 or a 2 at the second (thus had a score difference of 1).

Figure 5: Jitter plot illustrating the combinations of scores for one observer and two different viewings of the same group of animals. n (scored animals) = 430

3.4 Daily locomotion score analysis

The animals which were scored 3 after directly at an FS or after a positive pain test were subsequently scored in the daily video recordings from the days preceding the FS to find the day of onset of the lameness.

Figure 6 illustrates the occurrence of different time lapses, from the last day in which the cow was scored 1 (DS1) in the daily locomotion score, to the first of the following days in which the same cow was scored 3 (DS3), so the day of lameness onset. For reasons of comparison, the time lapse between the last day in which the cow was scored 1 (DS1) to lameness discovery at the fortnightly locomotion score (FS3) is displayed in the same plot. A total of n = 50 lameness cases was analysed to generate the graph in **Fehler! Verweisquelle konnte nicht gefunden werden.**; the red density curve illustrates how in this study most lameness cases developed during the time of 1 to 7 days, with a frequency peak at 3 and at 1 day(s). The blue density curve shows how in contrast 10 to 15 days elapsed in most lameness cases between the animal being sound and lameness discovery at the fortnightly locomotion score.

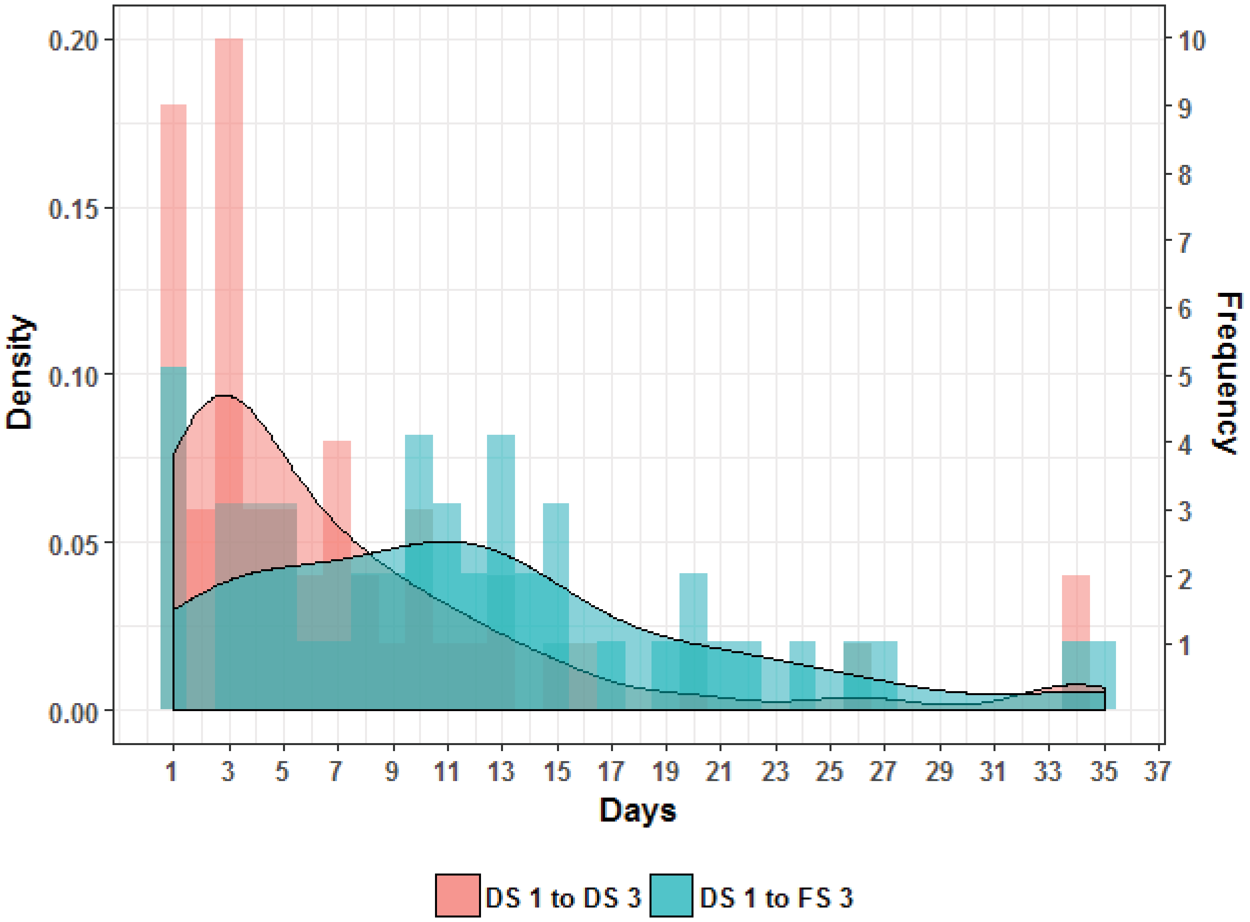

Figure 6: Density curve and histogram illustrating the occurrence of different time lapses between 1 (red): the last day the animal presented as sound (DS 1) and lameness onset (DS 3), and 2 (blue): the time lapse between the last day the animal presented as sound (DS 1) and the day the lameness was discovered at the fortnightly locomotion score (FS 3)

In a next step the time difference between DS3 and FS3 was calculated, in order to verify whether a fortnightly locomotion score such as the one carried out in this study is frequent enough to detect lame animals early. Figure 7 displays frequency and density of the time lapses in which the animals were lame for before being discovered. The highest frequency was encountered at 0 and 2 days, as well as at 10 and 11 days. As the data was not normally distributed the median was calculated and resulted in Mdn = 3 days, meaning that approximately half of the lameness cases were discovered within 3 days of the onset of lameness.

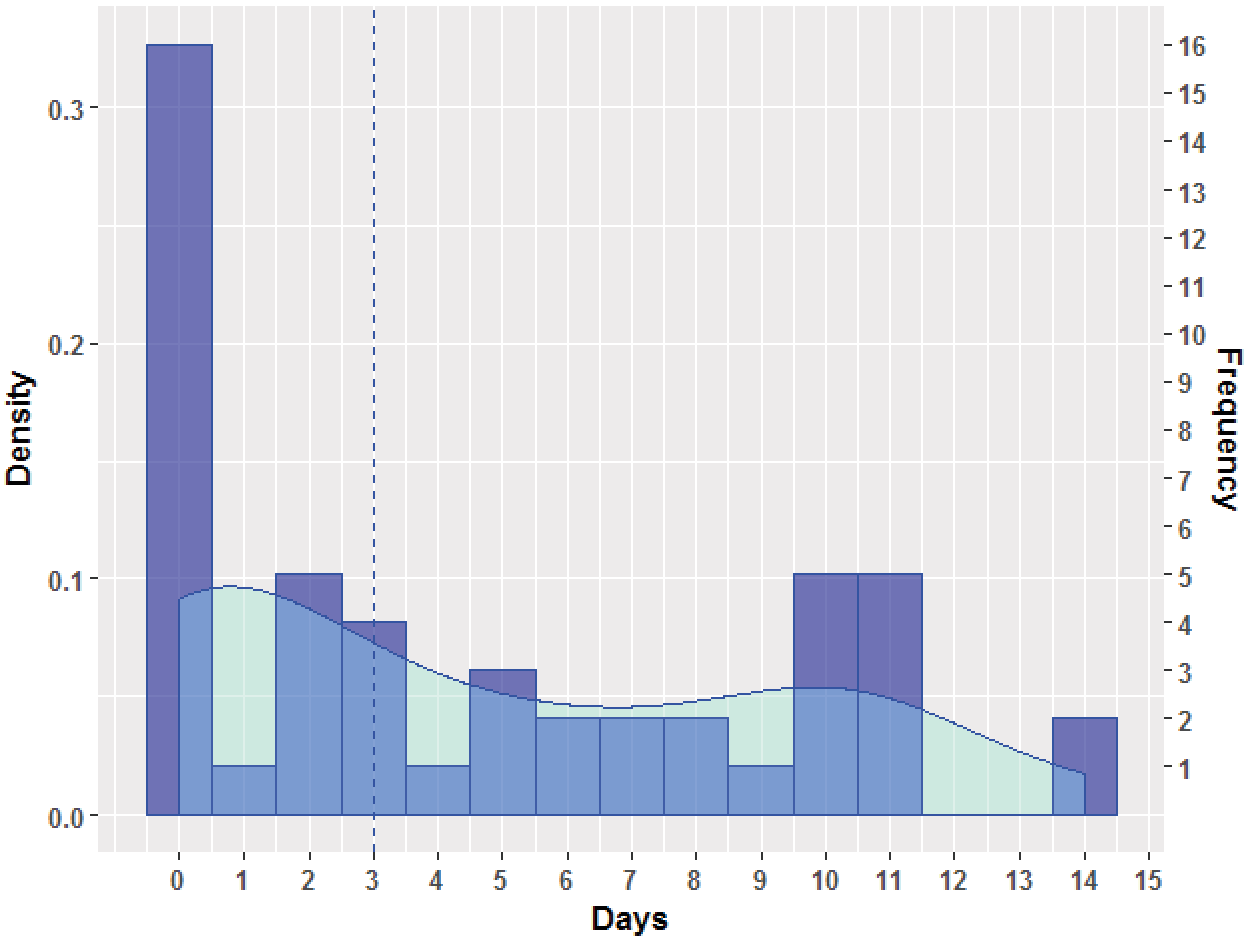

Figure 7: Histogram and density curve of the occurrence of the difference in days between the actual onset of lameness and lameness discovery at fortnightly locomotion score.

4. DISCUSSION

4.1 Locomotion score reliability and study limitations

The aim of including a clinical examination of lame animals after the locomotion score at the fortnightly farm visits is to improve the accuracy of the score. Due to time limitations it is not possible to examine the claws of all the animals in the herd after every locomotion score, meaning that the clinical examinations are restricted to animals that were considered lame in the first place or were a score 2 six weeks in a row. Therefore it is not possible to confirm the lack of clinical findings in animals considered sound and scored 1. This only limits the strength of the locomotion score to a certain extent, as animals that are actually lame and are erroneously scored 1 in a fortnightly score are expected to stand out at a later locomotion score when the lameness progresses; in this case the lameness onset is found by means of a daily locomotion score and the previous supposed "sound" animal's score is changed if necessary.

On the other hand, lameness cases which are examined without any clinical findings aren't retrospectively changed to a score 1, as some claw diseases, such as laminitis, can be extremely painful for the animal but produce no immediate lesions to the horn which would be visible to the naked eye.

One of the greatest difficulties in the process of locomotion scoring on video recordings is the amount of subjectivity involved in judging a cow's gait and posture. There are many influencing factors for video locomotion scoring including lighting conditions, which are especially difficult in the winter, and overcrowding of the area where cows are scored, which makes it difficult to see each animal individually. The way the animals exit the milking parlour differs on each farm; on one farm the cows exit in single file and walk on an even surface past the camera, on others the cows crowd the exit of the milking parlour, making it difficult to judge gait and posture.

The overall percentage of agreement (PA) illustrated in Figure 4 and Table 1, show an almost equal level of agreement between observers for the video observation (PA = 80.1 %) and for the direct observation (PA = 77.9 %).

A different situation presents itself when Kendall's Coefficient of Concordance (W) is calculated; in this case the scoring by direct observation has a higher level of agreement between observers ($W = 0.786$) than the video observation ($W = 0.680$). Cohen's Kappa (κ) was also calculated to measure the level of inter-observer and intra-observer reliability; this statistical test in particular was chosen for reasons of comparability. All other manual locomotion scoring systems listed in Schlageter Tello's review [7] are compared for inter- and intra-observer agreement using Cohen's Kappa and the percentage of agreement.

Whilst the PA only considers whether the scores given by the two observers have the same value, Kendall's *W* and Cohen's Kappa also take into consideration how far apart the discordant scores are; if observer 1 considers an animal to be lame (score 3) and observer 2 considers an animal to be completely sound (score 1), this influences the level of inter-observer reliability more than if the scores are only one rank apart.

Both *W* and κ measure the level of agreement between multiple observers rating the same objects or one observer reassessing the same objects in a separate viewing, and they both have values between –1 and +1. However, these two measures aren't directly comparable. Cohen's Kappa corrects the value for chance agreement, so that the value of κ can never be 1. In the case of this study, ordinal values with a ranking of 1 to 3 were used, therefore the possibility of chance agreement is very high, which increases the correction for chance agreement, thus decreasing the possibility of reaching higher values of κ in comparison to scores with more ranks. Cohen's Kappa was calculated using squared weighting in order to take into account the difference in ranking between scores and not just the score concordance or discordance. The lower κ value for the inter-rater reliability for the video observation can be explained by the fact that the "score 1 and score 3" combination (see Figure 4) is at 3.5 % the third most frequent score combination, and due to the squared weighting, this combination of scores particularly influences the outcome of the test. Although the percentage for this score combination is almost the same for the direct observation (at 3.3 %), in this case the "score 2 and score 2" combination was more frequent at 10.7 %, balancing the effect of the more heavily weighted "score 1 and score 3" combination and increasing κ's value.

The video scoring method is used nonetheless in this project, as previous experience and on-farm trials before the start of the project indicated that the presence of an observer standing at the milking parlour exit evidently affected the animals' behaviour.

4.2 Importance of regular locomotion scoring

The fluctuation of the locomotion scores in the course of this study (as illustrated in the timelines in Figure 3) demonstrates the need for regular locomotion scoring; one score every few months before or after claw trimming is not enough to capture most lameness bouts early enough, which can develop as quickly as illustrated in the density curves in **Fehler! Verweisquelle konnte nicht gefunden werden.**. To avoid chronic cases of lameness and not only economic loss but also concomitant severe pain for the cow it is crucial to detect lame cows in the early stages of the disease. According to the results of the data analysis in this study the chances of discovering lame animals soon after the onset of lameness are high if locomotion scoring takes place every 10 to 14 days; seeing as most analysed cases of lameness took between 1 and 9 days to develop (see Figure 6). It is especially noticeable in Figure 7 how in 16 cases there were 0 days difference between lameness onset and lameness discovery. This high frequency is a consequence of the combination of the locomotion score and the pain test. Most of the 0 day differences result from cases where animals presented with a score 2 in the DS for every day preceding the FS and originally also at the day of the FS itself, but were then corrected to a score 3 as a consequence of the positive pain test. These cases are especially relevant as they would not have been discovered if the locomotion score were not combined with a pain test. The closer examination of animals which repeatedly present with a score 2 offers an insight into the highly variable pattern of lameness development in the individual.

5. CONCLUSION

5.1 Automatic lameness detection

The conditions in which on-farm locomotion scoring takes place are often not ideal. The animals haven't always got a straight, even surface where they walk past the farmer allowing for accurate scoring. Also, time for locomotion scoring of the whole herd and animal observation isn't always available and most importantly not as often as suggested in the previous paragraph.

These issues highlight the need for automatic lameness detection on farms, which would shorten the crucial time lapse between lameness onset and recognition.

The three point locomotion score used in this study was developed with the intention of using it as a reference system for a predictive algorithm, meaning the algorithm would be able to distinguish between a severely lame animal which immediately needs treatment (score 3) and an animal which currently presents the conditions which speak for the development of a case of lameness and will become acute if not treated soon (score 2).

The claw health data for the months of September through to January have been collected but still not analysed, so many more lameness cases will be analysed in the future and eventually compared to automatically collected behaviour data in order to examine the correlations between the relevant parameters and the development of claw diseases.

6. LITERATURE

[1] T. S. Stashak und H. Wissdorf, *Adams' Lahmheit bei Pferden*. Schlütersche, 2008.

[2] R. D. Murray *u. a.*, „Epidemiology of lameness in dairy cattle: description and analysis of foot lesions“, *Vet. Rec.*, Bd. 138, Nr. 24, S. 586–591, Juni 1996.

[3] C. Mülling, J. Hagen, A. Fiedler, und S. Bystron, „Klauenerkrankungen beim Rind“, *Der Praktische Tierarzt*, Nr. 93 (Suppl. 1), 2012.

[4] Landeskuratorium der Erzeugerringe für tierische Veredelung in Bayern e.V., „LKV-Leistungsprüfung und LKV-Beratungsgesellaschaft mbH in der Milchviehhaltung in Bayern 2015, Ergebnisse und Auswertungen“, 2015. [Online]. Verfügbar unter: http://www.lkv.bayern.de/lkv/medien/Jahresberichte/mlp_jahresbericht2015.pdf. [Zugegriffen: 29-Mai-2017].

[5] European Food Safety Authority (EFSA), „Scientific report on the effects of farming systems on dairy cow welfare and disease“, *EFSA J.*, Bd. 7, Nr. 7, S. n/a-n/a, Juli 2009.

[6] Bundesministerium für Ernährung und Landwirtschaft, „Milchbericht 2017 des Bundesministeriums für Ernährung und Landwirtschaft“, Mai 2017.

[7] A. Schlageter-Tello *u. a.*, „Manual and automatic locomotion scoring systems in dairy cows: A review“, *Prev. Vet. Med.*, Bd. 116, Nr. 1–2, S. 12–25, Sep. 2014.

[8] A. C. Hoffman, D. A. Moore, J. R. Wenz, und J. Vanegas, „Comparison of modeled sampling strategies for estimation of dairy herd lameness prevalence and cow-level variables associated with lameness“, *J. Dairy Sci.*, Bd. 96, Nr. 9, S. 5746–5755, 2013.

[9] D. J. Sprecher, D. E. Hostetler, und J. B. Kaneene, „A lameness scoring system that uses posture and gait to predict dairy cattle reproductive performance“, *Theriogenology*, Nr. 47, S. 1179–1187, 1997.

[10] F. J. Manson und J. D. Leaver, „The influence of concentrate amount on locomotion and clinical lameness in dairy cattle“, *Anim. Sci.*, Bd. 47, Nr. 02, S. 185–190, 1988.

[11] A. J. Channon, A. M. Walker, T. Pfau, I. M. Sheldon, A. M. Wilson, und others, „Variability of Manson and Leaver locomotion scores assigned to dairy cows by different observers“, *Vet. Rec.*, Bd. 164, Nr. 13, S. 388, 2009.

[12] K. A. O Callaghan, P. J. Cripps, D. Y. Downham, und R. D. Murray, „Subjective and objective assessment of pain and discomfort due to lameness in dairy cattle“, *Anim. Welf.*, Bd. 12, Nr. 4, S. 605–610, 2003.

[13] I. Lorenzini, K. Schindhelm, B. Haidn, und E. Misha, „Using a three point lameness scoring system combined with a clinical examination to increase the reliability of visual locomotion scoring.“, in *Proceedings of the 19th International Symposium and 11th International Conference on Lameness in Ruminants*, Munich, Germany, 2017.

[14] K. Grimm, B. Haidn, M. Tremblay, und D. Döpfer, „Automatically Measured Performance and Behavior Parameters as Indicators of Lameness in Dairy Cattle“, in *Proceedings of the 19th International Symposium and 11th International Conference on Lameness in Ruminants*, Munich, Germany, 2017.

[15] A. Schlageter-Tello *u. a.*, „Relation between observed locomotion traits and locomotion score in dairy cows“, *J. Dairy Sci.*, Bd. 98, Nr. 12, S. 8623–8633, Dez. 2015.

[16] RStudio Team, *RStudio: Integrated Development for R (1.0.136 Version)*. Boston, MA, 2015.

[17] P. Legendre, „Coefficient of concordance“, *Encyclopedia of Research Design, N.J. Salkind*, Bd. 1. SAGE Publications, Los Angeles, S. 164–169, 2010.

[18] J. R. Landis und G. G. Koch, „The Measurement of Observer Agreement for Categorical Data“, *Biometrics*, Bd. 33, Nr. 1, S. 159, 1977.