Investigation into the use of sensor and production data to predict lameness in dairy cows

Heather Hemingway-Arnold BVMedSci (Hons)

Thesis submitted to the University of Nottingham for the degree of Master of Research

May 2022

Abstract

The object of this research was to investigate the effect of lameness on individual sensor and production parameters and combine this data to predict lameness in dairy cows. Lameness is an extremely important welfare issue currently facing the worldwide dairy industry. Traditional lameness detection by mobility scoring is subjective, variable and has low sensitivity: rendering it an insufficient tool in combating the high UK dairy lameness prevalence. Automatic lameness detection is undergoing constant investigation to support early lameness detection, but currently there is no system that offers an accurate and practical solution for implementation on farm. This study used a neck-mounted accelerometer, leg-mounted accelerometer and an automatic milking system to collect a plethora of data of 110 cows over a 3-month period. Mobility scoring occurred twice-weekly using an adapted version of the AHDB mobility score (0 = Perfect mobility, 1 = Imperfect mobility, 2A = Mildly Lame, 2B = Moderately Lame, 3A= Severe Lame, 3B = Non-Weight Bearing). There were considerable differences observed between lying time, activity, average weight and parity. Four random forest models were constructed using 1. ICEQube data (lying time and activity), 2. Lely Qwes-H data (activity and rumination), 3. Production data and 4. All data combined. The combined model achieved the best sensitivity, and specificity, at 0.74 and 0.75, and AUC at 0.82. This demonstrates that combining data from multiple sources improves predictive accuracy, although care must be taken when including confounding variables such as weight and parity. Model performance was improved when detecting severely lame cows over mildly lame cows. Combining multiple sensor technologies shows promise in improving detection of lameness in dairy cows, although challenges such as generalisability and variability must be overcome to improve sensor performance.

Acknowledgements

This MRes has been the single most amazing year of my veterinary degree so far. First and foremost, I would like to extend a massive thank you to the staff and students of the Ruminant Population Health Group at Nottingham. My experience with the team has been exceptional, and it has been a pleasure getting to know everyone. Thank you for being supportive, friendly, kind, hilarious and all around fantastic as I have undertaken this project with you.

Thanks especially must be extended for Rachel Clifton, Luke O'Grady and Martin Green for your guidance and supervision throughout this project. It has been utterly inspirational working with you all, and I could never express enough gratitude for how much you have taught me over the past year. Thank you for always being there when I needed help, and especially thank you for giving up so much of your time for me. I couldn't have asked for better mentors within this project and I am so excited to work with you all again in the future.

Thank you also to Bobby and Rachel for teaching me and so many other students the mysterious ways of R. Your help has been fundamental in producing this MRes. Thank you to Nikki, for your constant help and support with everything during my data collection! A special thank you must go to the team at the dairy centre, for your practical help during the data collection portion of the project and even now when the odd sensor turns up!

Finally, thank you to my loved ones, friends and family who have supported me throughout this project. Mum, Hattie, Holly, Mirelle, Emma, Archie and so many more: I am so appreciative of you. Specifically, to Sam, whose love I am very lucky to have. Jess Burridge, I literally couldn't have done this without you sitting opposite to me- you're going to do some amazing things.

Table of Contents

Table of Tables

Table of Figures

1 Introduction

In an age of increasing national and global animal welfare standards, lameness is one of the most prominent welfare issues facing the dairy industry in the present day (Cattle Health and Welfare Group Report., 2018; Farm Animal Welfare Council., 1997). Commonly defined as the manifestation of pain and injury to the foot (Archer et al., 2010; Bicalho et al., 2009), it has massive impacts on cattle health, welfare and production at an individual, farm and a national herd level. With regards to the industry, many UK milk buyers are now requiring lameness prevalence below 10% as part of a contract, although UK national herd prevalence is still estimated to be between 30-40% (Barker et al., 2010). Extensive research and education into lameness prevention is required to reduce prevalence on a farm level and in the national herd.

1.1 Impact of lameness

1.1.1 Welfare

Welfare is significantly reduced when cows suffer from lameness. The fundamental definition of lameness denotes a disease of extreme pain (Whay et al., 1998), so much so it results in altered gait (Bicalho et al., 2009). Lameness presents significant challenges to the UK dairy herd, where it is estimated that prevalence of lameness ranges from 15%-79% (Barker et al., 2010). Prevalence is further discussed in Section 1.3, but this demonstrates the enormity of the population suffering from the disease. Lameness also impacts directly on each of the 5 freedoms of the UK Farm Animal Welfare Council (Whay and Shearer, 2017). Lameness impacts the freedom from hunger and thirst, from discomfort, from pain, injury and disease, to express normal behaviour and the freedom from fear and distress (Whay and Shearer, 2017).

1.1.2 Behaviour

Additional to the substantial welfare challenges presented by lameness are negative impacts on behaviour, manifesting as increased lying time, reduced social interactions and reduced productivity (Whay and Shearer., 2017). Lame animals are less likely to exhibit normal behaviour than their non-lame counterparts (Galindo and Broom, 2002; O'Leary et al., 2020). Galindo and Broom (2002) used visual observation over a group of lame and non-lame cows for 32 hours to sample and compare behaviour. They describe a visible difference in behaviour, with lame cows exhibiting reduced aggressive behaviour, longer lying times, reduced feeding times and more time being licked by other cows. Other studies have also seen a difference in behaviour using visual observation of lame and non-lame cows, with lame cows exhibiting reduced feeding and watering times (Cramer et al., 2009) and an increased time moving the same distance as a non-lame cow (Espejo et al., 2006; Hut et al., 2021) and a decreased survival rate for those lower-ranking in the herd (Galindo and Broom, 2000). More recently, with increased digital recording technology available, behaviours have been quantified using sensor data (O'Leary et al., 2020). Sensor technology is increasingly used within the industry to record behaviours such as number of steps and lying times (Rutten et al., 2014). Several studies describe using sensor data to determine behavioral differences between lame and non-lame cows (Alsaaod et al., 2019a). In lame cows, they observed longer lying times, longer lying bout duration, reduced locomotor activity and reduced feeding times (Alsaaod et al., 2019a; O'Leary et al., 2020). These behaviours exhibit a large disparity between lame cows and non-lame cows, illustrating the debilitating effect of lameness on normal expression of behaviour in cattle.

1.1.3 Feeding and body condition score

Low body condition score (BCS) is also associated with an increased risk of lameness (Randall et al., 2015), but it is contentious as to which is the causal factor. Lame animals experience altered feeding and drinking behaviours compared to their non-lame counterparts (Shane et al., 2016; Barker et al., 2018). This is due to a combination of factors: pain manifesting in altered standing behaviour (Walker et al., 2008), less time spent at the feed face (Bach et al., 2007; Yunta et al., 2012; Thorup et al., 2016; Weigele et al., 2018) and competition from other cows resulting in less access to resources (Galindo and Broom, 2002). However, several studies have explored whether a low BCS or a drastic change in BCS can be a predisposing factor (Green et al., 2014; Lim et al., 2015; Newsome et al., 2017; Randall et al., 2016). Green et al. (2014) demonstrated a low BCS being associated with an increased risk of treatment for lameness over the following 4 months. Lim et al. (2015) showed a lower BCS (BCS < 2.25) at calving or a change of +- 1 BCS had a higher probability of becoming lame in the following 15-day period. They also found that these cows had a lower probability of returning to the non-lame. Randall et al. (2015) supports these findings, with the greatest risk of lameness at a BCS < 2. This suggests a bidirectional relationship between lameness and BCS in terms of causality. Several studies have associated a decreased BCS with a decreased digital cushion volume (Bicalho et al., 2009; Green et al., 2014; Newsome et al., 2017), resulting in decreased support of the structures in the foot. Newsome et al. (2017) described the digital cushion as an impact absorber to protect the sensitive corium and the pedal bone. Low BCS or changes in BCS could affect digital cushion volume, causing increased association of the pedal bone with the sensitive corium and thereby predisposing a cow to claw horn lesions.

1.1.4 Yield

Like the link between BCS and lameness, the relationship between milk yield and lameness has been explored in many previous studies. It is important to state that the effects of lameness on milk yield are complex and not easily explained. Higher yielding animals are more at risk for lameness than their lower-yielding herd mates (Amory et al., 2008; Archer et al., 2010; Green et al., 2002; Maxwell et al., 2015), meaning that although lameness can reduce milk yield, lame animals may still have a higher 305-day yield than those than are non-lame (Green et al., 2014). However, lame cows experience a decrease in production compared to their potential milk yield before a lameness event (Green et al., 2014). Green et al. (2002) reported a reduced potential milk yield up to 4 months before a lameness event, whilst Amory et al. (2008) reported a reduction in yield 2 months before a sole ulcer was diagnosed. (Randall et al., 2016) also reported a reduced milk yield of 2.68 kg/day in heifers during a lameness event in the first lactation. A reduction in milk yield occurs prior to the lameness event and varies with lesion type (Bicalho et al., 2008; Reader et al., 2011). Attempts to quantify this loss range from 350kg-360kg per case per lactation for a general case of lameness (Archer et al., 2010b; Green et al., 2002a), 573kg for a sole ulcer (Amory et al., 2008) and 369kg for a white line lesion (Amory et al., 2008), although quantified estimates may be poorly generalisable depending on the farm or system. Huxley. (2013) summarises estimates of between 270 to 574 kg per lactation. Following a lameness event, (Green et al., 2014) demonstrated a failure to return to full productive potential, with yield stabilising lower than the previous plane of production. Other studies have corroborated a lower yield than expected post-recovery (Amory et al., 2008; Reader et al., 2011). This evidence shows that whilst production losses associated with lameness are often hidden, they are extremely high.

1.1.5 Fertility

Another huge contributor to dairy production efficiency is fertility. The ability to successfully breed cows with a calving interval as close as possible to 365 days reduces daily costs associated with reduced yield at the end of lactation, increased number of inseminations to conception and smaller percentage of lactation spent at peak yield (Cattaneo et al., 2015). Lameness affects all aspects of fertility, including recognition of oestrus behaviour, time to conception and holding of an early pregnancy (Omontese et al., 2020. Whilst Walker et al. (2008) demonstrated similar proportions of lame and non-lame cows exhibiting oestrus behaviour, a significant reduction in time spent exhibiting oestrus behaviour by lame cows was observed (36%). In a subsequent study where oestrus was induced with prostaglandin, there was a reduction in the intensity and duration of oestrus behaviours in lame cows, although not the incidence of oestrus (Walker et al., 2010). Reduced duration and intensity of expression of oestrus behaviour could result in reduced recognition of oestrus in lame cows, and therefore increase the chance of missing a service. Fürst-Waltl et al. (2021) demonstrated that lameness during the dry period is associated with a prolonged period from calving to conception, and lameness early in lactation is associated with delayed first service during lactation. Many studies corroborate these findings, and all report increased time to first service and increased calving to conception interval (Lucey, Rowlands and Russell, 1986; Alawneh, Laven and Stevenson, 2011; Aungier et al., 2014 and Somers et al., 2015). Huxley (2013) summarised studies that investigate the effect of lameness on fertility and concluded that lame cows have a mean increase of 7 days for time to first service and 1.2 more services per conception, with a 20% lower conception rate overall. Although the cost of a lameness case is likely to be dependent on the type of lesion, Cha et al. (2010) estimates the cost of lameness on fertility to be upwards of £70 per case for a sole ulcer, and £40 per case for digital dermatitis. Although a hidden cost often masked by other factors, the effects of lameness on fertility are multi-faceted and significant.

1.1.6 Culling

Evidence from the literature regarding the impact of lameness on risk of culling is ambiguous (Cramer et al., 2009). Some papers have negated the risk of culling increasing due to lameness (Hultgren, Manske and Bergsten, 2004, Bandeau et al 1994), however other, more recent papers have described an increased culling risk due to lameness (Booth et al., 2004; Cramer et al., 2009; Randall et al., 2016). Cramer et al. (2009) reported a median time from foot trimming to culling of 157 d in cows with a foot lesion present, compared to 188d for cows with no foot lesions present. Booth et al. (2004) determined the hazard of culling for cows diagnosed as lame during the first half of lactation as nearly double that of non-lame cows. The costs of early culling are significant, including not only the loss of production from premature ejection from the herd, but also that of a replacement animal and the reduction in milk yield of a heifer compared to an adult cow. Lameness has a potential to increase culling risk, and costs associated with this should be considered when estimating cost of lost productivity.

1.1.7 Cost of prevention and treatment

Reduced productivity, fertility and increased culling risk result in substantial hidden losses for the UK dairy industry. Furthermore, substantial detrimental economic implications associated with lameness detection, prevention and treatment. At the individual animal level, lameness treatment is expensive (Puerto et al., 2021), not including expense for labour for carrying out treatment, coupled with the cost of detection (Dolecheck et al., 2019). Further costs include those incurred in lameness detection by a professional mobility scorer, visits of a hoof trimmer to the herd and visits from a vet for lameness cases with poor cure rates. Losses were summarised in a review by Dolecheck et al. (2019) and include potential susceptibility to other diseases and risk of lameness recurrence. It is important to note that the authors concluded that further research is required into each category to accurately estimate the true losses. Willshire and Bell. (2009) gives estimates of a cost per lameness case depending on disease: a total of £518.73 loss for a sole ulcer, a £300.05 loss for a white line lesion, and a £75.57 loss for a digital dermatitis lesion. Costs to the individual farmer are poorly generalisable, although this study estimates a cost of over £7000 per year. Understanding of the costs of lameness are lacking within the dairy industry, even though producers cite lameness as one of the three major health concerns along with mastitis and reduced fertility (Leach et al., 2010). Failure to appreciate the costs of lameness are detrimental to individual herd prevalence (Willshire and Bell., 2009), as understanding of the costs associated with increased lameness in the herd can be a driver for improved lameness treatment and detection (Dolecheck and Bewley, 2018).

1.2 Aetiology of lameness

Lameness has several main causes in dairy cows, which can be divided into 2 main categories: infectious causes and non-infectious causes (Barker et al., 2010; Shearer et al., 2012). Infectious causes of lameness include digital dermatitis, a bacterial infection affecting the skin on the heel bulbs of cattle (Palmer and O'Connell., 2015), and heel erosion, which is a bacterial infection primarily affecting the dermis of the heel bulbs (Monrad et al., 1983). These are estimated to account for up to 20% of lesions (Murray et al., 1996). Non-infectious causes include mainly claw horn lesions such as sole ulcers, sole haemorrhages and white line disease. These are recognised as being the most common causes of lameness (Miguel-Pacheco et al., 2016). Shearer et al. (2012) defines a sole ulcer as a full-thickness ulcerated section of the sole epidermis. An ulcer is thought to be a latter disease stage of the sole haemorrhage, which is a discoloured portion of sole horn where the corium haemorrhages during horn production (van Amstel and Shearer, 2008). A while line lesion (WLD) is described by van Amstel and Shearer. (2008) to be any disruption of the white line. Trauma or injury to the upper limbs can also be causes of lameness, although these are significantly less common than claw horn lesions (Newcomer and Chamorro, 2016). The causes of lameness are therefore diverse, yet all cause significant pain and so are equally important to detect and treat.

1.3 Prevalence of lameness

Many estimations of lameness prevalence in UK dairy herds have been attempted (Barker et al., 2010b Clarkson et al., 1996; Dippel et al., 2009; Griffiths et al., 2018; Randall et al., 2019; Rutherford et al., 2009). Most estimations support an alarmingly high national prevalence of 39% (Barker et al., 2010; Griffiths et al., 2018; Haskell et al., 2006), with some herds reaching a prevalence of up to 80%. Prevalence estimates have increased since earlier studies have been conducted, with a prevalence of 22.1% reported in 2003 (Whay et al., 2003), but more recent studies suggesting a national prevalence of around 40% (Randall et al., 2019). Caution must be taken when comparing time periods due to the disparity in scoring systems, however it is evident that UK lameness prevalence is only increasing. Drivers of this high prevalence are multifactorial, with risk factors including herd-level factors, such as management and housing strategies, cow-level factors such as parity, previous lameness events or negative energy balance, and environmental factors such as high stocking densities or housing on concrete flooring (Amory et al., 2006; Barker et al., 2010; Chapinal et al., 2014; Dippel et al., 2009; Pérez-Cabal and Alenda, 2014; Wells et al., 1999). Prevalence is lower for grazed herds than herds housed year-round, with (Haskell et al., 2006) finding a mean prevalence of 15% in extensive systems, compared to 38% in intensive systems. As discussed in Section 1.1.4, higher yielding cows are more at risk of lameness, and this could support the observed difference in prevalence between systems (Haskell et al., 2006), where intensively farmed cattle are generally more high yielding than their extensively farmed counterparts (Balmford et al., 2018). Evidence suggests several other factors can contribute to increased lameness prevalence observed in intensive systems. Griffiths et al (2018) suggest that increased standing times on hard surfaces, increased exposure to slurry causing subsequent reduced hygiene, and increased stocking density are all risk factors for increased lameness prevalence. Other studies support this even in extensive herds, observing increased lameness prevalence over the winter housing period

(Haskell et al., 2006, Rutherford et al., 2009). These high prevalence estimates demonstrate a growing need for extensive research and education in order to reduce lameness prevalence at a national level.

1.4 Lameness management

1.4.1 Prevention and treatment of lameness

In order to reduce lameness prevalence, decrease associated costs and improve welfare, herd level prevention and treatment strategies are instrumental in improving mobility. Treatment given depends on the type of lesion present. Infectious causes require topical treatment and foot bathing, whereas non-infectious lesions require therapeutic foot trimming and placement of an orthopaedic block. Early detection and treatment are paramount to a successful treatment outcome (Thomas et al., 2016). The majority of treatment is carried out by stockmen or farm workers (Murray et al., 1996), however treatment of lameness by farmers is often delayed (Barker et al., 2010). Alawneh et al. (2012) estimates the mean length of time between detection and treatment to be 28 days, but other estimates of time to treatment can be up to 135 days (Thomas et al., 2015). Delayed treatment results in a prolonged period of lameness, with a reduced chance of recovery (Thomas et al., 2016; Whay, 2002). Productivity losses and poor welfare implications are also enhanced with an increased period of lameness, which has detrimental effects to profit and yield as previously discussed. In addition to reducing productivity, untreated claw horn lesions lead to development of bony projections from the pedal bone (Newsome et al., 2016). These are irreversible and cause increased likelihood of lesion recurrence (Leach et al., 2012; Thomas et al., 2015). Early detection and thereby early intervention are essential for increased chances of recovery, with early treatment reducing presence of lameness up to 4 weeks after treatment (Leach et al., 2012). Reader et al. (2011) also demonstrated that early treatment invokes an increased chance of recovery, reduces recovery time and decreases risk of lameness recurrence. Therefore, intervening early is essential as part of improving treatment outcome, and is an important part of implementing a prevention and prevalence reduction strategy on farm.

1.4.2 Detection of lameness

For a lame cow to be treated, she must first be identified as lame. For the most part, lameness detection is performed by visual observation, using different aspects of a cow's gait. Many methods of assessing gait have been developed for easy identification of lame cows, and to thereby assess lameness prevalence across the wider herd. Flower and Weary. (2006) concluded that symmetry, tracking up, spinal curvature, head bobbing, speed, abduction and equal weight bearing are all important factors in assessing a cow's mobility. Several five-point scoring systems have been developed, such as the Sprecher et al., (1997) "locomotion score". This system assesses gait and back posture on a scale from 1-5, with lame cows being measured as score 3, 4 or 5. This system is commonly used in research and across the globe. Due to their stoic nature, however, cows could be less likely to express pain when under observation (Marino and Allen, 2017). A 5 score system may be too sensitive to distinguish between mild and moderately lame cows. Furthermore, this system focuses primarily on back posture, which has tentative associations with mobility and may not be the best indicator of lameness (Hoffman et al., 2014). Bell and Huxley. (2009) proposed a 4-point "mobility score" (MS), with scores ranging from 0-3. MS 0 represents perfectly non-lame cows, MS 1 represents cows with imperfect mobility, MS 2 represents lame cows where the lame leg can be identified, and MS 3 represents severely lame cows that cannot keep up with the healthy herd. This method has been widely adopted across the UK and by the Agricultural and Horticultural Development Board (AHDB) in order to standardise mobility scoring across the industry.

1.4.3 Challenges with visual lameness detection

Although the mobility score may be the industry standard, other countries and other research groups use other gait-scoring systems such as the Sprecher et al. (1997) mobility score mentioned above, which leads to poor generalisability when applied to UK research, who predominantly use the UK AHDB Mobility Score (https://dairy.ahdb.org.uk/technical-information/animal-health-welfare/lameness/husbandry-prevention/mobility-scoring/#.WXnhULuFOr8). Milk buyers also have different scoring systems, with Arla using a standard 2-point score to represent lame or not-lame (https://www.arlafoods.co.uk/). Overall, this ambiguity in different scoring systems needs to be addressed, and systems developed to achieve better standardisation across the global industry.

In addition to the use of different scoring systems, a variety of scoring intervals have been used, ranging from weekly (Alawneh et al., 2012), to fortnightly and even to monthly (Schlageter-Tello et al., 2014). Most milk contracts require mobility scoring at least quarterly ("Mobility scoring – A simple step to stamp out lameness - Promar International,"), although evidence suggests this is not sufficient for early detection and treatment of lame cows (Thomas et al., 2016). On many farms, regular lameness detection is rarely performed by a qualified individual (Adams et al., 2017), with the most common method of on-farm lameness detection being ad-hoc detection during routine management practice

(O’Leary et al., 2020a). This method is significantly flawed and presents a challenge to detecting and intervening with early stage or mild lameness (O’Leary et al., 2020a). Historically, farmers have underestimated the lameness prevalence in their herd (Leach et al., 2010; Fabian, Laven and Whay, 2014, Sadiq et al 2019), due to difficulty in detecting behaviour changes at an early stage (Espejo et al., 2006; Leach et al., 2010; Whay et al., 2003). Leach et al. (2010) found 90% of farmers in their study underestimated mobility scores. Fabian et al., (2014) estimated that farmers can detect as few as 27.3% of lameness cases. Furthermore, ad-hoc lameness detection leads to poor prevalence estimates and an inability to provide herd statistics (O’Leary et al., 2020a).

The poor sensitivity of traditional ad-hoc lameness detection has led to a need for a qualified register of individuals who have had some formal training in providing a mobility scoring service. The Register of Mobility Scorers (RoMS, Register of Mobility Scorers Limited, www.roms.org.uk) is a UK-based register for individuals who have had centralised, approved training and examinations based on performing and delivering the AHDB mobility score. Training improves an individual’s sensitivity when scoring, so detects a higher proportion of lame cows in the herd (Garcia et al., 2015). Schlageter-Tello et al. (2014) demonstrated that individuals that have undergone some professional training had a higher probability of assigning the same cow the same mobility score. Similarly, Weigele et al. (2018 also displayed high intra-observer agreement (93%) in locomotion scoring due to intensive training.

However, even in trained individuals, there remains a large variation and a poor reproducibility in mobility scoring (Channon et al., 2009; K A O’Callaghan et al., 2003; Kottner et al., 2011), with Garcia et al. (2015) demonstrating only a 72% probability of trained individuals assigning the same score to the same cow. Even with training, mobility scoring remains subjective (Martinez-Ortiz, 2013), and lameness prevalence is still being underestimated by farmers within their herds (Fabian et al., 2014). Because cows are a prey animal, they may not display a dramatic change in posture or gait in the presence of minor lesions (Schlageter-Tello et al., 2014). Therefore, mobility scoring by human observation will have poor sensitivity for identifying early-stage foot disease (Tadich et al., 2010), a clear failing in a preventative strategy focussing on intervention at an early stage (Thomas et al., 2016).

The practical limitations of mobility scoring must also be considered. Mobility scoring is time consuming, expensive and labour intensive (Atkinson et al, 2014). It is also disruptive to the cow's daily routine and can cause accidents on farm due to this disruption, with cows falling on a slippery surface or becoming aggressive when being faced with an unexpected human observer. Furthermore, lack of facilities, skills and time make detection challenging (Dutton-Regester et al., 2019; Horseman et al., 2014).

Mobility scoring, in summary, is a less than ideal way to carry out sensitive and timely lameness detection. It is time time consuming, labour intensive and depends on the skill of the observer to detect mild lesions (Winckler and Willen, 2001), which are those we hope to target with early detection. Therefore, there is significant need to develop alternative methods using technology for lameness detection (Schlageter-Tello et al., 2014). This will improve accuracy of lameness detection, allowing substantial progress with reducing national herd prevalence.

1.5 Technology and lameness detection

Precision livestock farming (PLF) combines livestock farming with engineering and technology (Lovarelli et al., 2020). PLF can be applied to increase efficiency regarding management of livestock health and disease by using automatic systems to provide longitudinal measurements of cattle health and behaviour (Riaboff et al., 2022). These systems apply permanent automatic monitoring technologies, resulting in real-time reporting on cattle behaviour that may be able to accurately detect fluctuations of normal behaviour in a timely fashion. This could allow early detection of any welfare challenges an individual may face (Chapa et al., 2020).

Considering the problems with traditional lameness detection, there is an increasing demand from farmers in the industry to be supported by automatic lameness detection systems (ALDS) (T. van de Gucht et al., 2017). ALDS are undergoing constant development and improvement, and few are available commercially (Alsaaod et al., 2019). There are many different technologies available, these either directly measure attributes of locomotion, or indirectly measure cow behaviour (Alsaaod et al., 2019). These systems provide support to the farmer by providing prompt, objective information where changes in behaviour can be measured over time (Beer et al., 2016a). Regular measurements allow longitudinal analysis of gait and performance, providing the farmer with up-to-date information on every cow (King et al., 2017).

As outlined in Section 1.4.3, traditional methods of mobility scoring have demonstrated poor sensitivity and impractical implementation. The use of ALD on farm has potential to overcome these pitfalls and provides an opportunity to improve detection of mild lameness (O'Leary et al., 2020). This can support high cattle welfare standards by providing prompt, effective treatment and intervention: preventing deterioration in lameness state and providing the cow with the highest chance of recovery (Fabian, Laven and Whay., 2014; Thomas et al., 2016). ALD also provides the opportunity for data collection and monitoring recovery post-treatment, potentially providing farmers with motivation to monitor lameness as a key performance indicator (KPI) and adopt practices to reduce prevalence in their herd (Horseman et al., 2014). Different types of ALD systems and their uses are discussed below.

1.5.1 Image processing

Direct observation of cattle mobility has been observed in several studies through analysis of image processing from video information (Abdul Jabbar et al., 2017; Flower et al., 2005; Pluk et al., 2012; Poursaberi et al., 2011; Schlageter-Tello et al., 2018; Viazzi et al., 2013). Image processing techniques involve analysis of a cow at walking speed, and anatomical features being identified by the image processor. Analysis of gait and body movement patterns (BMPs) by machine learning algorithms are performed to detect abnormalities in movement (Alsaaod et al., 2019). Relatively high sensitivity has been achieved in multiple studies: (Viazzi et al., 2013) achieved a sensitivity of 76% when analysing back curvature from a dorso-ventral direction and (Martinez-Ortiz et al., 2013) achieved a sensitivity of 99% by mapping gait patterns and walking speed from a lateral view. Care needs to be taken when comparing visual detection methods, as different angles and locomotion assessment methods are used between different studies. Chapinal et al. (2011) did report the limitations of using walking speed in predicting lameness, as slower walking speed can be induced by increased parity, interference from other cows in the race or fear of an object in the environment. Difficulty in cow identification is reported (Schlageter-Tello et al., 2018), with increased cow traffic meaning inaccuracy in detecting an individual. Whilst visual locomotion evaluation is fairly generalisable between farms, improvement is needed regarding implementing a system practically on farm, with farmers preferring animal-mounted technology over race-mounted or walk-over systems (T. van de Gucht et al., 2017).

1.5.2 Pressure sensors

Static and dynamic pressure sensor usage has been investigated for its ability to predict lameness (Dunthorn et al., 2015; Maertens et al., 2011; van Nuffel et al., 2015; Rajkondawar et al., 2002; van der Tol et al., 2003, 2002). Static pressure sensors primarily measure weight distribution during standing (Chapinal et al., 2010; Pastell et al., 2007). (Chapinal and Tucker, 2012) investigated the use of a weighing platform to measure the relationship between lameness and weight distribution between legs. They observed an increase in weight shifting of the hind limbs (1.6 steps for lame and 1.0 steps for non-lame), as well as a correlation between steps/minute and lameness. Pastell and Kujalaf, (2007) compared parameters of weight distribution between legs and used these to predict lameness using neural networks. They achieved a sensitivity of 100% but a specificity of 57%, meaning just under half of non-lame cows were being misclassified. Nechanitzky et al. (2016) calculated limb weight ratio, mean limb difference, and the standard deviation (SD) of the weight applied on the limb taking the least weight and analysed the difference between lame and non-lame cows. They found significantly lower weight bearing in the lame group than the non-lame group. Although both studies were able to observe differences between lame and non-lame, they all concluded that limb-weight ratio would be useful in combination with other sensor technologies to predict lameness. All studies further reported having difficulty keeping cows on the force plates for the required time of 90 seconds; this could be a practical limitation on farms not using a robot milking system.

Dynamic pressure sensors have also been evaluated in their ability to predict dairy cow lameness. Ground force reaction plates are sensor systems that measure force exerted onto the plate and the biomechanical process of walking. Several studies have demonstrated a difference in hind-leg symmetry between lame and non-lame cows (Bicalho et al., 2007a; Rajkondawar et al., 2006). Dunthorn et al. (2015) used an upgraded force plate system to measure longitudinal ground forces in three directions. They developed a logistic regression model built from 76 variables with a sensitivity of 90% and a specificity of 94%. High sensitivity in this study was reported due to no separation of test and train dataset, meaning a model could have been potentially overfitted to the data used. However, this did improve on the sensitivity and specificity achieved by (Liu et al., 2011), who investigated the same technology but measuring forces in a single dimension with 346 cattle, with a sensitivity and specificity which was 51.92% and 88.84% respectively. Limited sensitivity of this study was cited to be due to locomotion abnormalities in study cows causing false positive results. This does demonstrate, however, how maximizing the amount of data available in predictive modelling may improve model performance. This technology, similar to vision technology, is very expensive and difficult to practically implement on farms (Tim van de Gucht et al., 2017; van de Gucht et al., 2018). Further work is being conducted into improving the practicalities of these systems for industrial use, but in the meantime there is significant justification into investigating more practical solutions of predicting lameness, such as cow-mounted accelerometers.

1.5.3 Thermal imaging cameras

Although the exact pathology behind foot lesions is not known and is likely to be multifactorial, a substantial theory is that inflammation evokes changes in the corium and therefore the sole horn (Newsome et al., 2017). In the milking parlour, thermal imaging cameras have been used to detect a thermal signal from inflammation in the foot (Schafer and Cook, 2013). Alsaaod et al., 2015 used thermal imaging to collect data from 24 cows before and after claw trimming, finding a significant difference in the temperature of the coronary band in feet with a lesion present to those without. Similarly, further studies show an increased temperature in the hoof with presence of a lesion (Alsaaod et al., 2015b, 2014; Alsaaod and Büscher, 2012; Cockcroft et al., 2000; LokeshBabu et al., 2018; Main et al., 2012a; Nikkhah et al., 2005; Wood et al., 2015). The highest sensitivity and specificity was achieved by Main et al. (2012), with a sensitivity of 72% and a specificity of 73%. This results in 28% of cows are still being incorrectly identified as having a lesion present. Inability to distinguish between the type of claw horn lesion present was displayed in all these studies, indicating that further research needs to be completed in this area to improve accuracy (Alsaaod et al., 2019). Digital dermatitis detection also yields results with high sensitivity and specificity, with a sensitivity and specificity of 89.1% and 66.8% respectively (Alsaaod et al., 2014). However, this is still no use for detection of claw-horn lesions. Thermal imaging is a promising tool for detection of lameness in dairy cattle, although displays relatively poor sensitivity and specificity compared to other technologies mean that further development is required before practical implementation is possible.

1.5.4 Animal-mounted accelerometers

1.5.4.1 Background

Despite the obvious place for other ALD systems as detailed above, animal-mounted systems are both affordable and practical (Delagarde and Lamberton, 2015), and cause less disruption to cows' daily routine (Chapa et al., 2020). Farmers also prefer animal-mounted systems to walk-over or camera systems (T. van de Gucht et al., 2017). Most animal-mounted systems use accelerometers, with 3-Dimensional Accelerometers being the most extensively studied and widely used (Riaboff et al., 2022). Accelerometers collect raw movement data which is processed by machine learning methods to classify behaviour patterns such as standing, walking, and lying with high sensitivity, specificity, and accuracy (O'Driscoll et al., 2008). Chapa et al., (2020) reviewed accelerometer use across 219 research papers and concluded that they are a valuable tool in assessing various welfare parameters: namely activity level, oestrus behaviour, rumination and feeding behaviour, and lying times. Accelerometer use in ALD is categorised into measurements of behaviour and measurements of gait (O'Leary et al., 2020). Both are discussed in more detail below.

1.5.4.2 ALD using gait

Several studies have investigated the use of gait data from accelerometers for predicting lameness (Alsaaod et al., 2017a; Beer et al., 2016; Chapinal et al., 2011). Most notably, Alsaaod et al. (2017) achieved a 100% accuracy in classification, at a sensitivity and specificity of 100%. This method investigated only 12 cows without foot lesions, detailed “non-lame”, and 12 cows with foot lesions, detailed “lame”. High frequency accelerometers were used on both hind limbs to measure gait cycles, swing and stance phase and peak acceleration at toe off and toe on the floor. A positive aspect of this study is that the nature of the data means it is likely to be generalisable, meaning the measurements of a cow walking can be generated in a similar manner with every cow. However, this study was undertaken in a controlled environment with a small sample size and hence repeatability is likely to be poor (Alsaaod et al., (2017). Other studies also report high accuracies (Beer et al., 2016; Mangweth et al., 2012), but small sample sizes make evaluating these systems at herd level difficult. This technology is arguably more generalisable and more accurate than behavioural measures (O’Leary et al., 2020), since gait is fairly standard between cattle. However, gait analysis is very expensive and difficult to implement in a farm setting (Alsaaod et al., 2017), meaning measuring behaviour may be a more practical and simple solution to ALD.

1.5.4.3 Monitoring behaviour

Measurements of behaviour normally involve a single 3-D accelerometer which measures features such as standing and lying times, and step counts. Monitoring of animals in real time may provide valuable insights into cow-specific factors and day-to-day behaviour, and how these behaviours are affected by lameness (Shane et al., 2016). Different behaviours derived from accelerometers such as activity, lying time and feeding data have been measured and analysed, and comparisons drawn between lame and non-lame cows in previous studies (Borderas et al., 2011; Kamphuis et al., 2013; Weigele et al., 2018).

Furthermore, the ability to predict lameness based on the range of features mentioned above has been investigated with promising results (O'Leary et al., 2020). This technology is certainly preferable to farmers when animal-mounted (T. van de Gucht et al., 2017) and is a useful tool to overcoming the limitations of mobility scoring mentioned above. A full analysis of features used to predict between lame and non-lame cows is discussed below.

The use of accelerometer systems may be more desirable than those that require animals to be placed in a controlled position, as this may evoke a natural response to conceal disease in cows with mild lameness. Guiding animals through or across a sensor may trigger their instinct to hide signs of pain and weakness (Byabazaire et al., 2019). Therefore, there is significant justification in using behavioural features over other technology systems to support prompt and effective lameness detection.

1.5.4.3.1 Lying time

Across the literature, there is a majority agreement to the fact that lame cows have increased lying time to non-lame cows (Blackie and Maclaurin, 2019; Ito et al., 2010; King et al., 2017b; Nechanitzky et al., 2016; Schindhelm et al., 2017; Weigele et al., 2018). Several studies have also investigated the ability of lying time alone to predict lameness. Alsaaod et al. (2012) used purely variables related to lying time: Total lying bouts, bout duration and total lying time. They demonstrated a 65% accuracy using these variables, which is very low. Byabazaire et al. (2019) also investigated lying time in combination with other activity variables, such as step count and number of lying bouts. Using a clustering-based approach to identify individuals of high, normal and low activity levels, they achieved an accuracy of 87% with a random forest model based upon individuals of normal activity. They did not report the results of the other two clusters due to the small number of cows classified into these clusters. Similarly, they had a small sample size of lame cows, with only 32 lameness cases, of which only a proportion were available in the training data. Furthermore, lameness was identified by the farmer or on some occasions the agricultural scientists. Judging by the challenges with visual lameness detection presented in Section 1.4.3, this means that only severe lameness cases were likely to have been detected. This means that lying time alone is useful, but performance needs to be optimized using other features in tandem.

Consistently, there is large variation in measures of lying time, with mean lying time varying within a cow, between cows and between farms. O'Leary et al. (2020) speculates that measures of lying time are not dependable as predictors of lameness alone, because they are influenced by many factors (Thorup et al., 2015). Parity, DIM, farm management practices (grazing or zero grazing), cow comfort and lameness-causing pathology are all cow-specific or farm-specific factors affecting lying times (O'Leary et al., 2020; Thorup et al., 2015). Although lying times might not be generalisable as a lameness predictor across multiple farms, it could be used in combination with other factors as a predictor of lameness. It would also be beneficial to examine how much lying time is affected by mild lameness, as there are significant gaps in the literature addressing lying time and how it differs with mild lesions or lameness (Weigele et al., 2018).

1.5.4.3.2 Activity

Similar conclusions can be drawn between activity (total steps) and lying time in terms of evaluating their effectiveness as predictors. In general, literature demonstrates a reduction in activity due to lameness (Alsaaod et al., 2012;

Grimm et al., 2019; Kamphuis et al., 2013; de Mol et al., 2013; Thorup et al., 2015). However, activity alone has been shown to be a poor predictor of lameness. For example, Kamphuis et al. (2013) demonstrated an area under the curve (AUC) of only 0.66 when using just activity data to predict lameness. Borghart et al. (2021) also built a random forest model using only activity data, achieving an AUC of only 0.54. This demonstrates a poor ability of activity to predict lameness, although it can be used in combination with other factors. Reasons for this could vary. Like lying time, activity varies greatly between cows and within cows which makes thresholds between lame and non-lame cows difficult to define (Alsaaod and Büscher, 2012; Byabazaire et al., 2019). Furthermore, activity decreases as lameness severity increases (Reader et al., 2011; Thorup et al., 2015) meaning mildly lame cows may not show significant differences in activity levels. Parity and stage of lactation also have confounding effects on activity (Thorup et al., 2015). The average change of activity associated with increased mobility score was 1% by Reader et al. (2011), parity and DIM had a much bigger effect on activity. Although activity has poor performance in predicting lameness alone, there is lots of literature investigating its use in lameness prediction (O'Leary et al., 2020). The majority of papers investigating behavioural change with lameness uses activity as a feature (Alsaaod et al., 2019). Activity clearly has some merit for predicting lameness, albeit maybe just improving model prediction when combined with other variables. It would be beneficial to include it in combined models with other behaviours, as well as investigating the effect of mild lameness on activity.

1.5.4.3.3 Rumination

Rumination has been investigated in tandem with feeding behaviour data in predicting lameness, often in conjunction with differences in activity. General differences in feeding behaviour between lame and non-lame cows have been described, with lame cows spending less time eating, less time at the feed face and being less likely to eat during the day (A. Bach et al., 2007; Barker et al., 2007; Blackie and Maclaurin, 2019; Grimm et al., 2019; Norring et al., 2014; Weigele et al., 2018; Yunta et al., 2012). Rumination, however, are not likely to contribute to lameness prediction (O'Leary et al, 2020). Several studies have investigated the effect of lameness on rumination, showing no difference between rumination in lame and non-lame cattle (King et al., 2017; Thorup et al., 2016; Walker et al., 2008; Weigele et al., 2018).

Limited studies have analysed the use of rumination alone in predicting lameness, although van Hertem et al. (2013) did determine high correlation between rumination at night and lameness. Subsequently, when combined with other features such as milk yield and activity, the resulting logistic regression model achieved a sensitivity of 89% and a specificity of 85%. However, rumination was only highly correlated 6 days before diagnosis of lameness via farmer detection, and rumination of both lame and non-lame cattle was highly variable. Furthermore, they were only predicting moderate to severely lame cattle, not based upon those with mild lesions. Contrary to the above, they do, however, suggest that rumination could be useful in combination with other factors to achieve adequate lameness prediction (van Hertem et al., 2013). Rumination is also cited as an important tool in assessing presence of other diseases such as metritis (Cocco et al., 2021), and rumination drop is often used on farm as a signal for disease. However, this could be attributed to the acute nature of a disease such as metritis, compared to a more chronic disease such as lameness. Overall, rumination between lame and non-lame cattle could be useful in predicting lameness.

1.5.4.4 Production data

As discussed in above, accelerometer-based variables give poor predictive ability individually. However, much greater accuracies have been achieved by using accelerometer-based variables alongside production data (O'Leary et al., 2020). For example, accelerometer-based variables have been used in tandem with production features such as parity, DIM, milking order and milking yield to improve accuracy of machine-learning models (Beer et al., 2016b; Kamphuis et al., 2013; Schindhelm et al., 2017; van Hertem et al., 2013a). The increasing use of automatic milking systems (AMS) within the industry allows for a greater capacity of data collection in terms of production features (King et al., 2017). Most AMS collect a plethora of data, including visits to the AMS, milk yield, milk constituents and body weight. Some even collect data on milk temperature and somatic cell count estimates. More studies are investigating lameness prediction using features generated from the AMS (Byabazaire et al., 2019; de Mol et al., 2013; King et al., 2017b; Miguel-Pacheco et al., 2017; Westin et al., 2016). This technology has potential to alert to early indicators of disease when combined with other features of behaviour denoted from activity sensors and accelerometers. Use of specific production features for predicting lameness are discussed in more detail below.

1.5.4.4.1 Yield

As discussed in Section 1.1.4, decreased milk production in lame cows has been seen in multiple studies. The performance of milk yield in predicting lameness has also been been evaluated (Cramer et al., 2009b; Grimm et al., 2019; Kamphuis et al., 2013). Grimm et al., (2019) used an elastic net regression model to predict lameness using variables associated with liveweight, feeding behaviour, activity, lying time and milk yield. During univariate analysis there was no association of milk yield with lameness, although when combined with other features such as feeding behaviour and lying time milk yield had a substantial contribution to their predictive model. They concluded that it had a high reliance on other factors (Grimm et al., 2019) although could be useful in lameness prediction. Schindelm et al. (2017) used elastic net regression in with milk yield, feeding time, lying time and lying bouts. They achieved a sensitivity and specificity of 80% and 100% respectively. High model sensitivity and specificity can be attributed to the fact they were predicting only severe clinical lameness cases. However, they do conclude that milk yield in conjunction with lying time and feeding time was a useful predictor of lameness. In contrast, Kamphuis et al. (2013) and (van Hertem et al. (2013) excluded milk yield from their final classification models because it had no association with lameness in their data. Other factors such as parity and DIM both playing a massive part in determining yield differences between animals (Grimm et al., 2019; Schindhelm et al., 2017), which were not accounted for in either of these studies. Clearly, there is a reduction to milk yield due to lameness (Section 1.1.4), although there are other factors to account for differences in yield between herd mates, independent of lameness. The use of milk yield in predictive models has so far yielded ambiguous results, and further work is required to evaluate the effect of lameness on milk yield before definitive conclusions can be drawn. However, these studies do indicate that milk yield could be useful in predicting lameness when combined with other features, and if available it would be interesting to analyse the effect of mild lameness on milk yield.

1.5.4.4.2 Milk constituents

The use of milk constituents as a predictor for lameness has not been previously investigated. Furthermore, associations between lameness and milk constituents (fat percentage, protein percentage and lactose percentage) has been little documented. The physiology behind the process of lameness in the body is little known and there are likely numerous risk factors, although some papers discuss negative energy balance being a causative factor (Olechnowicz and Jacekowski., 2010; Schindelm et al., 2017). Negative energy balance and subsequent metabolic imbalance in cows with higher milk yields could cause potentially poor assimilation of nutrients, leading to lower protein, fat and lactose contents (Olechnowicz and Jacekowski., 2010). Antanaitis et al. (2021) did not use milk protein in a predictive model, but they demonstrated significant a higher milk fat to protein ratio, and lower milk protein in lame and non-lame cows. Olechnowicz and Jaskowski. (2012) supports this, with a lower milk protein in lame cows compared to non-lame cows. Borghart et al. (2021), however, displayed poor importance values for milk fat, protein and lactose content in their random forest models, suggesting they may not be good predictors of lameness. Since there are other contributing factors of other factors on constituent levels in milk (diet, parity, DIM), definitive conclusions cannot be drawn from the impact of lameness on these levels directly, but they may be useful in future analysis. Limited studies have concluded the effect of milk constituents in lameness detection. Therefore, it may be useful to collect this data and evaluate the differences of milk constituents between lame and non-lame cows, and their subsequent use in lameness prediction.

1.5.4.4.3 Visits to AMS

No studies have evaluated the use of number of AMS visits as a predictor for lameness. However, associations between number of visits to the AMS has and lameness have been reported (de Mol et al., 2013; Miguel-Pacheco et al., 2014). Miguel-Pacheco et al. (2014) described lame cows making 0.5 fewer visits to the AMS per day than non-lame cows, with the reduction of visits highly significant between 00:00 and 06:00. Borderas et al., (2011) displayed gait abnormalities and subsequent lameness as being a significant factor in reducing number of AMS visits per day. Therefore, visit data to the AMS could have fundamental differences between lame and non-lame cows and may improve accuracy in conjunction with other features when used in a predictive model.

1.5.5 Combining data from different sources

Combining multiple sources of data can lead to better accuracy of lameness prediction. Recent studies have shown that combining features obtained from multiple sensor technologies improves model performance compared to using features from a single sensor technology (Borghart et al., 2021; Grimm et al., 2019; Kamphuis et al., 2013; Schindhelm et al., 2017; van Hertem et al., 2013a). Kamphuis et al. (2013) achieved an AUC at 0.74 when combining activity data, liveweight and milking order, whereas models using a single data source produced a maximum AUC of 0.66. Borghart et al. (2021) demonstrated an AUC of 0.85, with a sensitivity and specificity both of 0.78 when combining multiple data sources, whilst individual models had poor sensitivity and specificity in comparison. These features could be weak predictors individually but when combined provide better information about a cow's behaviour (de Mol et al., 2013; Kamphuis et al., 2013).

Whilst associations between lameness and individual variables measured by sensor technologies is commonly reported, there is considerably less literature available evaluating predictive ability of various sensor technologies on lameness (Alsaaod et al., 2019). Lying time is the most heavily evaluated, and activity is also investigated frequently due to the dual purpose of accelerometers often measuring step count as well as lying time. Production data such as parity and DIM is often included in models as it is commonly available on farm (Borghart et al., 2021). Milk yield data is available both through automatic parlours and AMS systems, so has been reviewed in predictive models as in Section 1.5.4.4.1. Increasing use of the AMS mean that variables such as number of milking visits and milk constituents are increasingly available, so they have not been so heavily investigated. Each study uses different combinations of features from different sensor technologies in their models, with no consistency across the literature (O'Leary et al., 2020). Due to the variation of all behavioural and production traits between farms (Thorup et al., 2015), ability to predict lameness of each of the technologies investigated within literature is incomparable with each other, and there is limited literature that evaluates multiple systems in conjunction with one another. To the author's knowledge, there is no literature available evaluating the use of all common types of sensor technology, and so investigation into this on farm could be useful at complimenting current literature.

1.5.6 Shortcomings of automatic lameness detection

Despite the benefits of automatic lameness detection discussed in Section 1.5, there are also several shortcomings. Most notably, some methods of automatic lameness detection are significantly more expensive than manual mobility scoring. Specifically, pressure sensors and vision detection systems have historically been both expensive and difficult to install on farm (Tim van de Gucht et al., 2017; van de Gucht et al., 2018). Farm-specific factors can also hinder installation and setup, for example power availability, positioning of technology on parlor exit and race, and WiFi or connection available for data storage and transfer. Furthermore, most literature has demonstrated a low sensitivity and specificity to be considered practical to be useful on farm (Schlageter-Tello *et al.*, 2014). Behavioral monitors are considered the most affordable form of automatic lameness detection (Alsaaod, Fadul and Steiner, 2019), although there are specific consequences surrounding generalizability of the data between animals and between farms.

1.5.7 Technology summary

ALD systems are more likely to detect severe lameness than mild lameness (summarised by O'Leary et al., (2020). Severely lame cows are likely to be detected by observation, meaning the requirement is for ALD to detect mildly lame animals. A failure in early detection leaves mildly lame cows without treatment, and they are more susceptible to becoming chronically lame. Therefore, technology is required to improve detection of mild lameness, to intervene prior to these cows becoming chronically lame. However, no technology available in the literature have detected mild lameness repeatedly, and with good accuracy.

In summary, many studies have evaluated the effectiveness of using accelerometers to predict lameness. Most common behaviours measured with accelerometers are lying times, feeding behaviour, rumination behaviour and activity (Alsaaod et al., 2019). Although better accuracy is obtained in studies with gait measurements over behavioural measurements (summarised by O'Leary et al., 2020) this promotes several substantial cost implications to the farmer (Kaniyamattam et al., 2020). Combined with the fact that farmers prefer cow-mounted ALD systems (T. van de Gucht et al., 2017b), there is a justification for investigation into collecting behavioural data from multiple systems and evaluating their use in lameness detection. Several studies have shown improvement to model performance when predicting lameness by combining activity data, rumination data and data from an automatic milking system, as summarised in Section 1.5.5. However, each study uses a range of features collected from each sensor type, and all with differing results. Therefore, this gap in the literature needs to be addressed.

1.6 Methods used for predicting lameness

O'Leary et al. (2020) summarises the variables generally used for predicting lameness. They report that the most common behavioural data analysis computes daily summary totals of activity data to give a more succinct dataset that is easy to analyse (Beer et al., 2016b; Byabazaire et al., 2019; Kamphuis et al., 2013; King et al., 2017b; Thorup et al., 2015). Some studies have also computed day:night ratios of some variables, for example Grimm et al. (2019). However, environmental influences of day:night ratios may confound results, for example if milking takes place during the day then cows will naturally be more active during the day, not necessarily due to lameness. Furthermore, Miguel-

Pacheto et al., (2014) found a significant decrease in AMS visits during the hours of 00:00 and 06:00. Therefore, whilst day:night ratios could give interesting insights into differences between lame and non-lame cows, daily summaries are used most commonly for predictive models.

Supervised classification models are most often used to predict lameness, with a binary outcome of either lame or not lame dependent on mobility score. The most common models used are logistic regression (Beer et al., 2016; van Hertem et al., 2013), random forest (Byabazaire et al., 2019) and support vector machines (Alsaaod et al., 2012; Haladjian et al., 2018). The use of these models was summarised by O'Leary et al. (2020), who concluded that support vector machines can often be unpredictable. Grimm et al. (2019) used elastic net regression due to the inter-variable relationships between features in their dataset; elastic net allows preservation of features only directly associated with the outcome. Borghart et al. (2021) used gradient boosted decision trees in their models, reporting high sensitivity and specificity and good accuracies. Byabazaire et al. (2019) investigated various models including decision trees, random forest, support vector machines and k-nearest neighbours, concluding that random forest and k-nearest neighbours offered the best performance. Overall, decision tree algorithms such as random forest appear to have been widely used and yield some of the best results in terms of sensitivity and specificity (Borghart et al., 2021; Byabazaire et al., 2019).

Various measures have been implemented to evaluate model performance in classification of lame and non-lame cows. There is a variation in statistical methods used in model evaluation, but the most common are a confusion matrix, model metrics such as sensitivity, specificity, positive predictive value, negative predictive value and accuracy (van Hertem et al., 2013; de Mol et al., 2013). A receiver operator characteristic (ROC) curve, subsequent area under the curve (AUC), and balanced accuracy are also common (O'Leary et al., 2020), with many papers using AUC as a principal evaluator in their models (Bicalho et al., 2007b; Borghart et al., 2021; Grimm et al., 2019; Kamphuis et al., 2013). F-score and precision have also been used in previous research investigating lameness in other species (Kaler *et al.*, 2020). Borghart et al. (2021) also demonstrates calibration curves providing a good visual device to assess model accuracy. They also use variable importance graphs to determine which features contribute the most to the overall model. Judging by the literature, good evaluators of model performance appear to be a confusion matrix and subsequent sensitivity, specificity, accuracy and positive predictive value, a ROC curve and AUC, calibration curves and relative variable importance.

1.7 Introduction summary

With an increasing lameness prevalence in the UK dairy herd (Griffiths et al., 2018), investigation into more sensitive lameness detection for earlier lameness intervention is even more important for the modern dairy industry. The common limitations of mobility scoring; including subjectivity, poor sensitivity, and impracticality could be overcome using automatic lameness technology. Evaluation of the use sensor data for lameness detection would be of substantial benefit to prevention at a cow-level and a national herd level. The main aim of ALD is to construct a predictive model with high performance. As discussed above, and there are substantial grounds for investigation into using behavioural and production data to predict lameness, and combinations of multiple sensor data sources to improve model performance.

1.8 Aims of study

The principal aim of this study was to investigate the performance of models for predicting lameness in dairy cows using three different technologies: automated milking systems, neck-worn sensors and leg-worn accelerometers. My hypotheses were:

- Sensor-based and production data can be used to accurately predict lameness in lactating dairy cows
- Combining multiple sensor technologies will produce a more accurate predictive model for lameness than a single sensor or a single range of features

To address the above hypotheses, the following objectives were developed:

1. Obtain longitudinal data from a cohort of lactating dairy cows on a single farm over a three-month period including regular mobility scores, behavioural features from both neck-worn and leg-worn accelerometers and production data from an AMS
2. Examine differences in behaviour between non-lame and lame cattle using the data collected from the different sensor technologies
3. Develop and evaluate three supervised machine learning models for predicting mild, moderate and severe lameness using data from individual technologies described Objective 1
4. Develop a supervised machine learning model using data from all three technologies combined and compare performance of this model to the models developed in Objective 2
5. Evaluate which of the features and therefore technologies described in Objective 1 are the most useful in predicting lameness

2 Methods

2.1 Study farm

Data collection took place from the 14th June 2021 to 10th September 2021 at the University of Nottingham Centre for Dairy Science Innovation in Sutton Bonington, Leicestershire. The dairy herd comprises of 330 year-round calving Holstein-Friesian dairy cows, averaging a 305-day milk yield of 13,000 litres/cow and is housed year-round.

This study was carried out in accordance with the University of Nottingham's guidelines and approval on ethical animal research. It was an approved by an Ethics Committee based at the School of Veterinary Medicine and Science, Sutton Bonington Campus.

Study animals were split between two groups subject to the same management conditions. Each group were housed in a free-stall barn with 60 deep sand cubicles. Water access was provided using 4 water troughs placed either end of the pen space. Feeding consisted of a total mixed ration offered ad libitum and concentrate fed in the automatic milking system during milking according to milk yield (0.45 kg/kg milk yield above 32 kg/d, up to maximum of 12 kg/d or 3 kg/AMS visit). Milking was performed on an individual basis in an automatic robotic milking station (MS; Lely Astronaut A3; Lely UK LTD., St Neots, UK).

Lameness management on farm consisted of a routine mobility score (MS) and foot trim according to farm's herd health plan. Cattle were mobility scored by a RoMS accredited technician fortnightly according to a modified version of the AHDB 4-point mobility score (Agricultural and Horticultural Development Board, www.ahdb.org.uk). Cattle experiencing their first lameness case were seen and treated by the farm's routine foot trimmer at a fortnightly interval. A standard 5-step Dutch foot trim was performed on all 4 feet. If indicated, a block was applied to the contralateral claw and the cow treated with a licensed non-steroidal anti-inflammatory drug.

2.1 Recruitment

Prior to the start date of the study (14^{th} June) according to the inclusion and exclusion criteria below.

To be eligible for recruitment cattle had to be greater than 21 days in milk (DIM) and be between parity 1 and 6. Cows less than 21 DIM were excluded to minimise the risk that cows would be suffering from transition or post calving disease, which could have affected behavioural data. Cow were also considered ineligible for recruitment if they had; undergone treatment for any lameness-causing pathology in the precious four weeks, diagnosed with any upper limb lameness or were undergoing a course of treatment for any disease at the time of recruitment. Cattle were also excluded form recruitment if they were due to be sold or dried-off during the study period. Based on these criteria 115 cows were eligible for recruitment.

2.1.1 Recruitment method

To ensure study animals in both groups demographically similar, cows were grouped into strata before being randomly assigned to one the groups. Strata were based upon parity (1, 2 and 3+) and mobility score (MS) history. A MS of 0 or 1 was considered sound, and an MS of 2 or 3 was considered lame. MS history was evaluated for the previous 4 MS dates prior to the recruitment date. Strata comprised of three categories, "Non-lame"; where the cow had been sound, "Fluctuating"; where the cow had received both sound and lame mobility scores, and "Chronic"; where the cow had only received lame mobility scores.

Each stratum was then divided at random to fulfil cohort requirements. DIM distribution was plotted to ensure DIM did not differ.

2.1.2 Preparing recruited cows

Cows were moved into the study environment one week prior to the study start date to prevent excess activity affecting data collection. Cattle were fitted with wearable sensors and sensor function was checked to ensure correct data collection as described in Section 2.2.3.

2.1.3 Removal of cows from study

Seven cattle were removed from study for a variety of health reasons. Removed cows were replaced with an eligible cow of similar lameness history, same parity, and a DIM of +- 10 days if possible. If cattle of +-10 days DIM were not available, then a cow as close as possible would be chosen.

2.2 Data collection

Data collection occurred from the 14th June-10th September 2021. Data were collected from multiple sources and using a variety of modalities as outlined below.

2.2.1 General animal data

General animal data was captured from each cow using the on-farm management system Uniform Agri (Uniform Agri, https://www.uniform-agri.com/en/). Uniform collects data on all significant events in the animal's lifetime from birth to present, including information relating to the cow's current lactation; cow identification number, date of birth, calving date, number of calves at present calving, current parity, any incidence of disease and courses of treatment.

2.2.2 Mobility scoring

Mobility scoring was carried out by visual assessment according to a modified version of the AHDB mobility score (Table 3). This method has been used in other studies to increase the sensitivity of mobility scoring by allowing further classification of MS 2 cows into "mild" (2A) and "moderate" (2B), and MS 3 cows into "severely lame" (3A) and "non weight bearing" (3B) (Wilson et al., 2022).

MS data were captured twice weekly (Monday and Thursday) during the study period by two individual Register of Mobility Scorers (RoMS, Section 1.4.2) accredited mobility scorers. RoMS-accredited scorers were used to ensure the scorers had similar basic training and a qualification. Two scorers were used to reduce familiarity to the cows and their previous mobility history. No intra-observer reliability was carried out in this study although the author acknowledges this would have been useful to determine similarity between scorers. Scores were recorded on pre-prepared score sheets containing all the cow identification number (ID) present in the group on that day to ensure no cow was missed from a mobility score. These were then translated weekly into Microsoft Excel (Microsoft Corporation, 2018) by the principal researcher responsible for data management.

The following variables were recorded: Date, Cow ID, Robot Group, Operator, Mobility Score, Lame Limb (if identifiable), Comments. Comments included; "no score", if a cow was not present at the mobility score; "block", if a cow had an orthopaedic block on one claw or more claws.

Table 1 A modified version of the AHDB Mobility Score (Bell and Huxley, 2009) as described in Section 2.2.2. This version has been used in previous research papers to increase sensitivity of mobility scoring (Wilson et al., 2022).

Mobility score	Category of mobility score	Description
0	Perfect mobility	Even weightbearing, flat back, long, fluid strides
1	Imperfect mobility	Uneven steps or shorted strides, lame or affected limb or limbs not obvious
2A	Impaired mobility – mildly lame	Uneven weightbearing on one or multiplelimbs that are subtle but visible to the trained eye
2B	Impaired mobility- moderately lame	Uneven weightbearing on one or multiplelimbs that is immediately obvious
3A	Severely impaired mobility	Unable to keep up with the healthyherd or walk as fast as a fast human pace
3B	Non weight bearing	Unable to keep up with the healthyherd as not weight bearing on one or more limbs

2.2.3 Sensor data

2.2.3.1 The ICEQube sensors

The ICEQube by ICERobotics sensor (IceTag3D, ICERobotics, Edinburgh, UK) is a leg-worn accelerometer, that provides information on various behavioural features using accelerometers that measure acceleration across multiple axes. These behavioural features include: lying time, number of steps, motion index, transitions up (i.e. from lying to standing) and transitions down (standing to lying). The sensors operate by interpreting and summarising the accelerometer data into behavioural metrics over 15-minute intervals where it is stored on the CowAlert by ICERobotics (CowAlert, ICERobotics, Edinburgh, UK) online data management portal. These data were then downloaded from CowAlert weekly by the principal researcher, to ensure a backup was kept in case of any data loss.

Study animals were fitted with one sensor during the selection period prior to the study. Sensors were fitted on either the left or the right hind leg. Initially, sensors were fitted by a trained technician employed by ICERobotics. During the study, the ICEQube sensors were fitted and managed by the principal researcher and sensor function was evaluated weekly using the CowAlert portal. Dysfunctional sensors were replaced.

2.2.3.2 Lely Qwes-H system

All cattle on the farm were fitted with a Lely Qwes-H system (Lely Industries N.V., Maassluis, the Netherlands) collar at their first calving. This equipment is a neck-worn identification device that is essential for every cow to be able to use the Lely Astronaut A3 automatic robotic milking system (AMS). It also links with the Lely “Time for Cows” (T4C, Lely Industries N.V., Maassluis, the Netherlands) database to collect production data from the AMS on each individual. This system is managed by the farm staff as part of the daily routine, and any dysfunctional sensors are identified using the T4C system and replaced immediately.

The Lely Qwes-H system gives information on several behavioural features using a 3-dimensional accelerometer which detects movement duration, movement intensity and total steps. Daily rumination time is collected through a rumination microphone, where vocal signals are recorded and combined with general activity index to calculate daily rumination time. Sensors capture these behavioural metrics in 2-hour time intervals which is transmitted and stored in the T4C online portal.

2.2.4 Production data

The Lely Qwes-H system allows electronic identification of each individual cow when using the Lely Astronaut robot milkers. The Lely Astronaut robot milkers use a variety of sensors to collect production data at each milking, including; milking time, milk yield (L and kg), milk constituents (fat %, protein %, lactose %), milk speed per quarter, conductivity per quarter and milk temperature. The robot milkers also collect other production data, such whether the milking was successful or a failure, whether the robot refused the cow, kilograms of feed consumed (kg) per milking and liveweight (kg) per milking. All production data is stored using the T4C system.

2.2.5 Data collation

Sensor and production data was downloaded from the relevant source at weekly intervals by the principal researcher and stored using Microsoft OneDrive (Micro) as a backup in case of any data loss from the sensor's online storage system.

2.3 Data handling and preparation

All datasets were subject to data preparation before analysis to ensure they were suitable for data handling and compatible with the handling software. All data handling was completed in R (Core Team 2022) using R Studio version 4.3.1 (R Studio Core Team 2022, Version 4.3.1). Data were downloaded from various sources as outlined below into Microsoft Excel. The Excel (xls) file converted to a Comma Delimited (CSV) version and individually imported into R Studio.

2.3.1 Mobility Scores

2.3.1.1 Initial handling

The mobility scores were first arranged by cow identification number and score date in ascending order. To maintain consistency during analysis, all lame scores that contained letters were capitalised, so were written in the format of 2A, 2B, 3A and 3B. A range of comments were provided from each operator. Any comments were visually inspected, but none were significant.

2.3.1.2 Defining lame/non-lame

Definitions of lame and non-lame in reference to MS was confirmed prior to analysis. Non-lame (NL) mobility scores were defined as either a MS 0 or a 1. Lame (L) mobility scores were defined as either a MS 2A, 2B, 3A or 3B.

Cows of MS 2A were classed as “lame” as there were proportionally less cows MS 2B and 3A (n = 255) than MS 2A (n = 547). Including MS 2A as lame resulted in a more balanced dataset. Furthermore, when cows of MS 2A were presented to the foot trimmer, lameness-causing lesions were found.

2.3.2 ICEQube and Lely Qwes-H sensor data

2.3.2.1 Generating daily summary totals

ICEQube and sensor data was initially arranged in cow identification number and date in ascending order.

ICEQube data was transmitted from the sensors in 15-minute intervals. Lely Qwes-H data was transmitted from the sensors in 2-hour intervals. This led to large and harder to manage data frames. It was also difficult to join these two data sets due to the differing time frames. Therefore, all features were combined into daily summary totals to provide a more manageable dataset. A ‘day’ was defined as a 24-hour period commencing at midnight. All the separate variables were totalled to provide a daily total for each cow per day. If any daily values were missing, a mean value was generated from the time period before and the time period after.

2.3.2.2 Removal of cows from study

Time periods where cows were removed from the study were recorded by the principal researcher. Periods of time these cows were absent from the study environment were excluded from each dataset from midnight of the previous day before their removal date. Several cows also joined the study to replace those lost. Sensor data was included from midnight of the day they entered the study.

Cows with <14 consecutive days of sensor data were excluded from the study as insufficient MS data was captured. 6 cows total were excluded from the study for reasons mentioned above.

2.3.2.3 Feature generation

2.3.2.3.1 ICEQube sensor data

The ICEQube monitors record data in 15-minute time periods. Several features were computed from the activity data based on an individual cow in a 24-hour period, defined as a 'day'. Daily lying time, total steps per day, total daily motion index were all computed by adding the total number of each feature in each 15 minute block. Number of transitions up (act of standing) and number of transitions down (act of lying down) were combined into the number of lying bouts. The difference of each cow from the herd daily mean lying time was calculated. The difference of each cow from the herd daily mean steps was calculated. The difference of each cow from the herd daily mean motion index was also calculated.

2.3.2.3.2 Lely Qwes-H system

Similarly, to the ICEQube sensor data, several features were computed from the Lely Qwes-H system. Number of heat alerts, steps per day and rumination minutes per day were calculated. The difference of each cow from the herd daily mean total rumination time was calculated. The difference of each cow from the herd daily mean total steps was calculated. Heat alerts were converted from a TRUE/FALSE character format into a numerical categorical value (1 or 0).

2.3.3 Production data

Production data was arranged by cow identification number and milking date order. Data was available per milking, but milking times varied greatly between cows. A data frame in this format was large and hard to manage and combine with other variables. Therefore, a daily summary dataset was created prior to analysis. If any production data was missing, a mean value was created from the day before and the day after. Similar to the method described in Section 2.3.2.1, data were taken for the previous 24 hours prior to a mobility score, to ensure that the same 24-hour periods were analysed.

Features that could average daily were defined. This included each cow's milk yield, milk constituents (fat %, protein %, lactose %) and average weight. These were combined into a daily average per cow for each feature. Visits to the AMS were categorized into daily number of milkings, milking failures and daily refusal: if they were not in need of milking but still visited the AMS. These were combined into daily totals. Parity and DIM was also combined with this dataset. For every day of the study, DIM was computed by calculating the number of days between the calving date, and the current date.

2.4 Descriptive statistics

Initial analysis comprised of determining the number of each MS observation, the number of data points for each data frame and the number of data points per cow.

The spread of each feature was plotted as a simple histogram, with the feature on the x axis and count on the y axis. The distribution of each feature was examined and evaluated. If there was an uneven, non-standard distribution of a feature, a natural logarithm was applied.

Descriptive statistics were calculated using the R packages "Psych" (Revelle., 2022) and "ggplot2" (H Wickam, 2016). A summary analysis was conducted for each feature, determining the mean, minimum, maximum, and interquartile range for each variable by lameness group and by mobility score. These were then plotted by mobility score to visualise any obvious differences in distributions. Mean and maximum standard deviations (SD) per cow per day were calculated for each feature. SDs were calculated to understand the variability within an individual cow's activity and behaviour.

2.5 Machine learning

2.5.1 Model definitions

2.5.1.1 Initial analysis

Four datasets were prepared for statistical analysis. These were the ICEQube sensor data, the Lely Qwes-H system sensor data, the Lely AMS production data and all of these data combined.

MS data was linked to each data frame from the 24-hour period previous to the MS. This was done to reduce the effect of the any variability in the behavioural data caused by the act of mobility scoring the cows. Data was excluded for any date where the foot trimmer attended the farm (n = 4). This avoided possible inaccuracies in behavioural metrics where only certain cows were disrupted. Blocks can alter gait due to elevation of one claw (Haladjian et al., 2018). Mobility scores from blocked cows were unlikely to be reliable, so they were excluded from the analysis completely.

2.5.1.2 ICEQube model

48 features were used in the ICEQube model, as described in Appendix 1. These variables were not adjusted or reduced and were combined with variables from the other models. The following variables were calculated; total lying time in seconds per day, total steps per day, total motion index per day, total lying bouts per day, difference from the daily mean herd lying time per day, difference

from the daily mean herd steps per day, difference from the daily mean herd motion index per day and difference from the daily mean herd lying bouts per day. To investigate changes over time, values of total lying time, total steps per day, total motion index per day and total lying bouts per day for the previous 7 days was included. A 7-day average and a 7-day standard deviation was included for each feature.

2.5.1.3 Lely Qwes-H model

27 features were used in the Lely Qwes-H model as described in Appendix 2. These variables were not adjusted or reduced and were combined with variables from the other models. The following variables were calculated, total steps per day, total rumination minutes per day, total heat alerts per day, difference from the daily mean herd rumination time per day and difference from the daily mean total steps per day. Total steps and total rumination minutes for the previous 7 days were also included in the model. A 7-day average and a 7-day standard deviation was included for total steps and total rumination minutes.

2.5.1.4 Production model

86 features were analysed as described in Appendix 3. These variables were not adjusted or reduced and were combined with variables from the other models. The following variables were created; parity, DIM, daily yield, average daily yield, average weight, average daily fat percentage, average daily protein percentage, average daily lactose percentage, average daily ISK, number of successful milkings per day, number of failures per day, number of refusals per day. Values for the previous 7 days, a 7-day average and a 7-day standard deviation was included for each feature apart from parity and DIM.

2.5.1.5 Combined model

The combined model comprised of all of the features discussed in Sections 2.5.1.2, 2.5.1.3 and 2.5.1.4. Overall, 157 features were included in this model.

2.5.2 Model building

Supervised classification analysis was performed to evaluate the ability of each of the four data sources to predict lame or non-lame cows.

Using the R packages "Caret" (M Kuhn., 2022) and "RandomForest" (Liaw and Wiener., 2002), four random forest models were created on one of the datasets detailed in Section 2.5.1, with lame or non-lame as the outcome classifier. Random forest algorithm uses decision trees created from a subset of features sampled from the full set of features at each split in the tree. The "forest" is created by building multiple decision trees using bootstrap aggregation ("bagging"), where random bootstrapped samples of the training data are used to build each tree. The predictions of these multiple uncorrelated trees are then combined by averaging the predictions from each tree in the forest.

2.5.2.1 Test and train dataset

Data were divided into test and train datasets randomly using the createDataPartition() function in the Caret package (M.Kuhn., 2022). Data from an individual cow could only be in either the test or train dataset, so cows were randomly assigned to the test or the train dataset. The proportion of lame and non-lame points in each model was evaluated to ensure that an equal proportion of lame and non-lame cows were available in both the test and train set. The train dataset was used for the development and tuning of each of the models and consisted of 60% of the study observations. Once training and tuning was complete, the remaining training dataset of 40% of observations were used to evaluate the performance of the models on data it has not been exposed to. The number of observations in the training dataset was 1098, and in the test dataset was 730.

2.5.3 Model training

2.5.3.1 Cross validation

Ten-fold cross validation was used during the model training and tuning process (Erbe et al., 2013). Ten-fold cross validation randomly splits each train dataset into ten equally sized subsamples. One subsample is then held out and used to evaluate the predictions of a model trained on the remaining nine subsets. This process is then repeated until predictions are evaluated on all ten subsets. The ten-fold cross validation process was repeated 10 times. This aims to avoid the problem of over fitting a model on a non-representative subset of data. Balanced accuracy was used to evaluate model performance during the training and tuning process at an accuracy cut off of 0.5. Balanced accuracy is a measure of model accuracy and is calculated by adding the sensitivity (true positive rate) to the specificity (true negative rate), and then dividing by two. It gives information on how the model performs in terms of both sensitivity and specificity, in case of a high value of each individually (Velez et al., 2007).

2.5.3.2 Hyperparameter tuning

The hyperparameter "mtry" was used in each random forest model. "mtry" denotes the number of features considered at each decision node in each decision tree. To optimise model accuracy, hyperparameter tuning was performed, where a manual grid was set up to evaluate different values of "mtry": from 2 to the number of features used in each model. The best "mtry" was selected and used in the model based on which gave the highest balanced accuracy.

2.5.3.3 Variable importance

Variable importance was calculated for each feature used in each model. A variable importance graph was determined for each model. This method ranks the variables by how much they contribute to the outcome of the model. Variable importance is calculated by determining how much the squared error on all the trees of a model decreased when that variable is chosen to make a classification decision at a decision node ("Variable Importance in Random Forests - Code and Stats"). This is then scaled up to 100. Features with a high variable importance are considered to have a considerable effect on the outcome values. The relative variable importance for each feature was then plotted using a variable importance graph. Visual inspection of the variable importance graph was performed, and the most important model features determined.

2.5.4 Evaluating model performance

Each model's performance was tested on both the training and the 40% hold out test datasets. The predictive performance of each model compared to the actual lameness class was evaluated using a variety of measures as described in this section.

2.5.4.1 Confusion matrix and discrimination statistics

Each model was evaluated using a confusion matrix, which summarises performance of the classification algorithm. Prediction output of the classification at a probability cut-off value of 0.5 was generated from each model and compared to the actual lameness class. This determines number of true positives, false positives, true negatives and false negatives. From these metrics, several discrimination metrics can be calculated. A sensitivity, specificity, positive predictive and negative predictive value was calculated. Accuracy and Cohen's kappa statistic was also calculated. The formulas for these metrics are displayed in Table 2.

Metric	Formula
Sensitivity	True positives / (True positives + False negatives)
Specificity	True negatives / (False positives + True negatives)
Positive Predictive Value	True positives / (True positives + False positives)
Negative Predictive Value	True negatives / (False negatives + True negatives)
Cohen's Kappa statistic	Total accuracy – Random Accuracy / 1 - Random Accuracy
Balanced accuracy	Sensitivity + Specificity / 2

Table 2 Formulas for discriminant statistics. This table shows the formulas for the discriminant statistics calculated from the confusion matrix detailed in Section 2.5.4. Sensivity, specificity, PPV, NPV, Cohen's kappa statistic and balanced accuracy were all calculated to evaluate model performance.

2.5.4.2 ROC Curve and area under the curve

Using the R packages “pROC” (Robin et al., 2011) and “ggplot2”, a receiver-operator characteristic (ROC) curve was generated for each model (Robin et al., 2011). The performance of each model was illustrated by plotting the sensitivity on the y axis and specificity on the x axis as the probability threshold was varied. An ideal ROC curve takes an immediate steep increase before plateauing close to the y axis (Robin et al., 2011).

Area under the curve (AUC) was calculated from each model by determining the area under the ROC curve. This is the probability that a lame cow would be given a higher predicted probability than a non-lame cow. Perfect accuracy is associated with an AUC of 1 (Liu and Wu, 2003).

2.5.4.3 Calibration curve

Performance of each model was also evaluated using a calibration curve from the package “PresenceAbsence” (Freeman and Moisen, 2008). A calibration curve compares the distribution of actual lame cows for groups based on binned predicted probabilities. The test dataset was divided into uniform sized bins of predicted probabilities (e.g. 0-0.1, 0.1-0.2). The average predicted probability of each bin (also called the Observed Event Percentage) was plotted on the x axis, and the proportion of true positives in each bin was plotted on the y axis. The ideal calibration curve demonstrates a direct linear relationship (x = y) and an intercept of 0 (Borghart et al., 2021).

2.5.4.4 Optimising the discrimination threshold

Using the R package “OptimalCutpoints” (López-Ratón et al., 2014) and the ROC Curve, an optimum probability threshold was determined for each model using

Youden’s index (Youden., 1950). An optimum threshold that produces the highest combined sensitivity and specificity. New confusion matrices were then generated for each model using the optimum threshold to determine the classification performance of each model at the optimum sensitivity and specificity. The discrimination variables detailed in Table 2 were calculated.

Further confusion matrices were constructed comparing the performance of the model with only non-lame (MS 0 and MS 1), moderately lame (MS 2B) and severely lame (MS 3A) cows included. This was to evaluate the performance of the model on only moderate and severe lame cows. The discrimination variables detailed in Table 2 were calculated from each confusion matrix for each model.

2.5.4.5 Evaluation of Predicted probabilities by Mobility Score

To allow a visual assessment of each model’s ability to correlate predicted probabilities with each MS group, the predicted probabilities generated by each model were compared visually using boxplots, with individual MS on the x axis. Two horizonal intercepts were added to the plots, one at the default probability threshold (0.5) and one at the calculated optimum probability threshold value for the respective model.

3 Results

3.1 Mobility scores

Overall, 2430 individual MS observations were captured: 146 (6%) observations of MS 0, 1424 (58%) of MS 1, 547 of MS 2A (23%), 212 of MS 2B (9%), 43 of MS 3A (2%) and 0 of MS 3B (0%). There were 56 observations where cows were wearing blocks (2%) and 2 observations with no score recorded; these were all excluded prior to merging mobility scoring data with the sensor data. Once combined with the sensor data, 1828 individual MS events were available for analysis. 111 MS 0, 1091 MS 1, 431 MS 2A, 162 MS 2B and 33 MS 3. There were 626 (33%) lame (L) MS observations and 1202 (67%) non-lame (NL) MS observations.

There were 664 observations for cows in parity 1, 538 for parity 2 and 626 for parity 3+. Of cows in parity 1, there were 505 non-lame and 159 lame observations. In parity 2, there were 400 non-lame and 138 lame observations. In parity 3+, there were 297 non-lame and 329 lame observations. Severely lame animals (scores of 2B or 3A) were more likely to be multiparous, with 141 out of 195 severe lame mobility scores being from cows with parity 3+ (73%).

3.2 ICEQube data

881,970 individual ICEQube observations were captured. These were summarised into 12,863 full 24-hour periods.

3.2.1 Summary statistics split by lameness group

3.2.1.1 Lying time

Table 3 summarises the ICEQube data by lameness group: non-lame (NL) or lame (L). The mean lying time of the non-lame and lame groups were 13.26 hours/day (h/d) and 14.03 h/d respectively. The SD of lying time in the non-lame and lame groups were 1.93 h/d and 2.28 h/d respectively and the inter-quartile (IQR) range of lying time in both groups was large, at 2.90 h/d for non-lame and 2.32 h/d for lame, showing that lying times were highly variable. The distribution of lying time in each MS group is shown in Figure 1 and summary statistics are presented in Appendix 1. There was a trend for lying time to increase with increasing MS (Figure 1 and Appendix 1), and the greatest difference in mean lying times was observed between cows with normal gait (MS 0, mean lying time 12.8 h/d) and cows that were severely lame (MS 2B and MS 3A, mean lying time 14.5 h/d and 15.1 h/d respectively). The SD of the lying time in these groups were 1.6 for MS 0, 2.3 h/d for MS 2 and 2.1 h/d for MS 3. However, differences in mean lying times between sequential MS groups were small (Appendix 1) and there was a large range of lying time for all MS groups meaning there was significant overlap between groups (Figure 1). In particular, the distribution of lying times for cows with MS 1 and MS 2A were very similar, whilst the biggest difference in distribution was between the MS 0 and MS 3A groups.

3.2.1.2 Total steps

There was a difference in the mean of total steps between non-lame and lame cows (944.01 steps/day for non-lame cows compared to 835.57 steps/d for lame cows). SD was large in both non-lame and lame groups, at 322.27 steps/d and 311.62 steps/d respectively, meaning this difference was not significant. When looking at individual MS groups there was a more obvious difference in mean total steps between cows with normal gait (MS 0, 1056 steps/d) and cows classed as lame (MS 2A, MS 2B and MS 3, 814.0, 841.2 and 827.7 steps/d respectively, Appendix 1).

The distribution of total steps for each MS group is displayed in Figure 2. There was a trend for total steps to decrease with increasing MS. Notably, the distribution of total steps between MS 2A and 2B was very similar, with a similar median and range (Appendix 1).

Motion Index was highly correlated with total steps (Appendix 10) and similar trends were observed in motion index data as reported for total steps data (Appendix 1 and).

3.2.1.3 Lying bouts

The mean number of lying bouts per day between the non-lame and lame cows were very similar at 11.51 and 11.52 bouts/day respectively, with a SD of 3.03 for non-lame and 2.98 for lame. The range of lying bouts per day was also similar between non-lame and lame cows at 24 and 29 bouts/day respectively.

There was very little difference in mean daily lying bouts between MS groups (Appendix 1) or in the distribution of daily LBs between groups (Figure 3).

3.2.1.4 Difference from the mean

The mean difference in mean for lying time of both groups are -0.30 h/d for non-lame and 0.52 h/d for lame. The SD is similar and large for non-lame and lame groups, at 2.92 h/d and 2.91 h/d respectively. The mean difference in mean for steps for both groups is 31.57 for non-lame and -84.67 for lame. Similar to lying time difference in mean, the SD is large for both groups at 317.73 and 307.15 respectively. Both groups have a large range of lying time and steps difference in mean. Difference between the means of lying bouts difference in mean is negligible, with a difference of 0.06 lying bouts/d. Due to the correlation between total steps and motion index, motion index difference from the mean follows a similar trend as total steps difference in mean.

Table 3 Descriptive statistics for variables from the ICEQube sensor. This table gives the mean, standard deviation, median, minimum, maximum and range for each daily value by lameness group. This data was collected from ICEQube sensor and averaged into daily values. Non-lame cows are MS 0 and MS 1 and lame cows are MS 2A, MS 2B and MS 3A.

Variable	Group	Mean	SD	Median	Minimum	Maximum	Range
Total Steps (steps/day)	Non-Lame	944.01	322.27	919.50	0.00	5658.00	5658.00
	Lame	835.57	311.62	807.50	0.00	4390.00	4390.00
Total Steps Diff from mean (steps/day)	Non-Lame	31.57	317.73	6.55	-1010.69	4637.14	5647.83
	Lame	-84.67	307.15	-106.69	-1020.86	3379.31	4400.17
Lying Bouts (bouts/day)	Non-Lame	11.51	3.03	11.00	0.00	24.00	24.00
	Lame	11.52	2.98	11.00	0.00	29.00	29.00
Lying Bouts Diff from mean (bouts/day)	Non-Lame	-0.11	2.92	-0.26	-11.22	11.63	22.85
	Lame	-0.05	2.91	-0.37	-11.37	16.62	27.99
Lying time (hours/day)	Non-Lame	13.26	1.93	13.27	2.80	24.00	21.20
	Lame	14.03	2.28	14.03	5.16	24.00	18.84
Lying time Diff from mean (hours/day)	Non-Lame	-0.30	1.85	-0.31	-10.51	11.64	22.15
	Lame	0.52	2.15	0.54	-8.14	10.48	18.62
Motion Index	Non-Lame	3856.34	1312.93	3758.00	0.00	20902.00	20902.00
	Lame	3576.96	1366.70	3429.50	0.00	18850.00	18850.00
Motion Index Diff from mean	Non-Lame	65.60	1289.32	-72.56	-4201.80	16613.74	20815.54
	Lame	-252.22	1353.61	-383.05	-4288.26	14641.20	18929.46

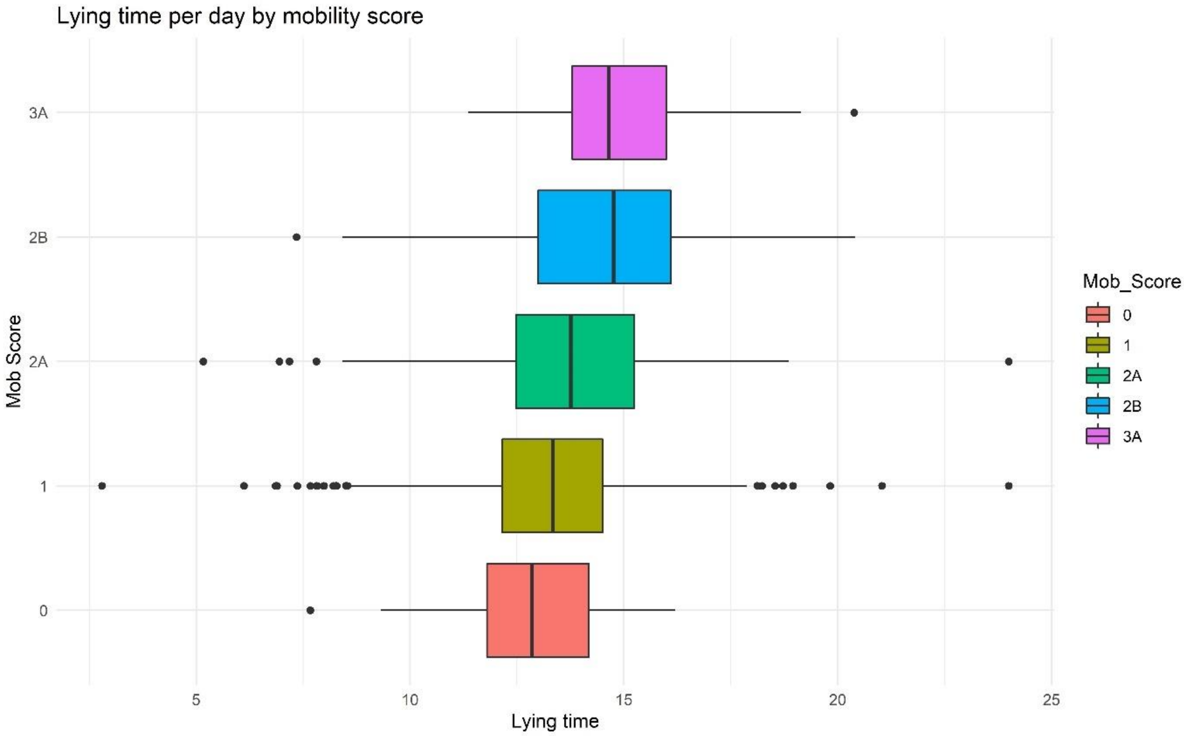

Figure 1 Total lying time from the ICEQube system by mobility score. This graph displays the lying time on the x axis, and mobility score on the y axis. Daily lying time was collected using a pedometer built in to the ICEQube system and summarised into daily values. Mobility scores were collected by two RoMS accredited mobility scorers on Monday and Thursday during the study period.

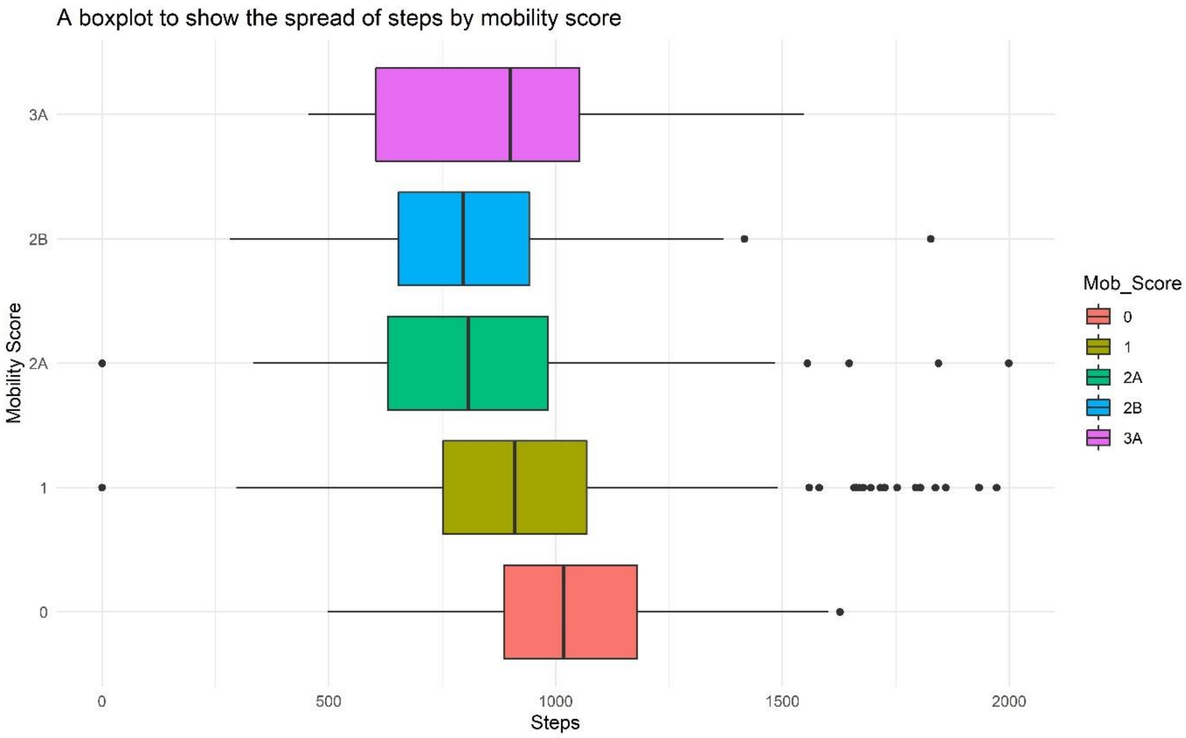

Figure 2 Total steps from the ICEQube system by mobility score. This graph displays the total steps on the x axis, and mobility score on the right axis. Number of steps were collected using a pedometer built in to the ICEQube system and summarised into daily values. Mobility scores were collected by two RoMS accredited mobility scorers on Monday and Thursday during the study period.

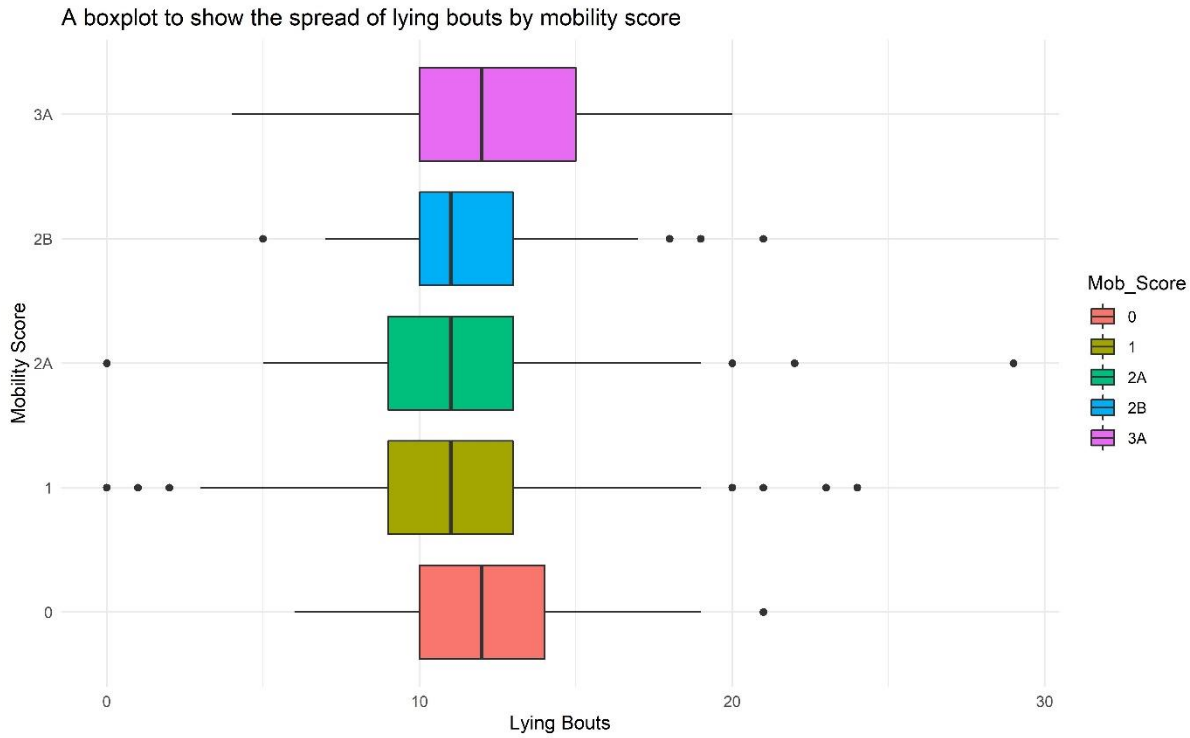

Figure 3 Lying bouts from the ICEQube system by mobility score. This graph displays the daily lying bouts on the x axis, and mobility score on the right axis. Daily lying bouts were collected using the transitions up (from lying to standing) and the transitions down (from standing to lying) data collected by the ICEQube system and summarised into daily values. Mobility scores were collected by two RoMS accredited mobility scorers on Monday and Thursday during the study period.

3.2.2 Model outcomes

3.2.2.1 Confusion matrix

Table 6 compares the results of model prediction by creating a confusion matrix using the test dataset comparing the actual classification of lame/non-lame and the predicted values of lame/non-lame.

The model building on the features created from the ICEQube data (ICE model) had sensitivity = 0.26, specificity = 0.94, kappa statistic = 0.23 and balanced accuracy = 0.60 when evaluated using the test dataset (730 observations). The final value of mtry for this model was 2.

3.2.2.2 Variable importance

0 shows the relative variable importance of each feature for the ICE model. For the ICE model, the most important feature was the lying time 7-day average; all lying time-related features were in the top ten contributors to the model. The least important features in the model related to lying bouts, with all lying bout-related features having a relative variable importance value of below 10% (Appendix 7).

3.2.2.3 Model calibration

The calibration for the ICE model was good: Figure 15 shows that Observed Event Percentages (OEPs) are close to the y=x line, and a line of best fit applied fits this line exactly. For P=0.1, 0.2 and 0.3, the OEP is such that the bin midpoints lie on the y=x line, showing good calibration. There is a small drop in the OEP in this curve at P=0.6. The confidence intervals in this curve increase when P>0.6 and are very wide at P>0.8, although the OEPs of the bin midpoints follow the y=x line. Wide confidence intervals are likely to be due to small numbers of observations in higher predicted probability categories.

3.2.2.4 ROC curves, AUC and optimum threshold

The AUC for the ICE model was 0.72 indicating moderate accuracy. The optimum threshold for the ICE model was 0.3 and applying this threshold increased sensitivity from 0.26 to 0.50, and decreased specificity from 0.94 to 0.82. PPV increased (+24) and NPV decreased (-0.27) whilst the kappa statistic and balanced accuracy increased from 0.23 to 0.27 and from 0.61 to 0.67 respectively (Table 7).

3.2.2.5 Predicted probabilities by MS group

Appendix 15 shows the distribution of predicted probabilities (P) for each model per mobility score group. Descriptive statistics including the mean, SD, median, maximum, minimum and range P is detailed in Appendix 13. With the ICE model there was an increasing trend of P per MS group. The mean P for MS 0 is 0.19, and the mean P for MS 3A is 0.57. The P for MS 1, 2A and 2B is 0.30, 0.41 and 0.50 respectively. This displays an increase in mean P as MS increases. The SD ranges from 0.13 for MS 3A to 0.20 for MS 2B. The minimum and maximum P for MS 0 is 0.024 and 0.756, meaning a range of 0.732. For MS 1 the minimum and maximum values are 0.03 and 0.80, with a range of 0.77. For MS 2A the range is 0.84, for 2B the range is 0.88 and for MS 3 the range is 0.46. The maximum values of P for MS 2A and 2B and 3A are 0.92, 0.94 and 0.72 respectively. The median values of P increased with MS. The median for MS 0 was 0.11, for MS 1 was 0.27, for MS 2A was 0.41, for MS 2B was 0.47 and for MS 3A was 0.59.

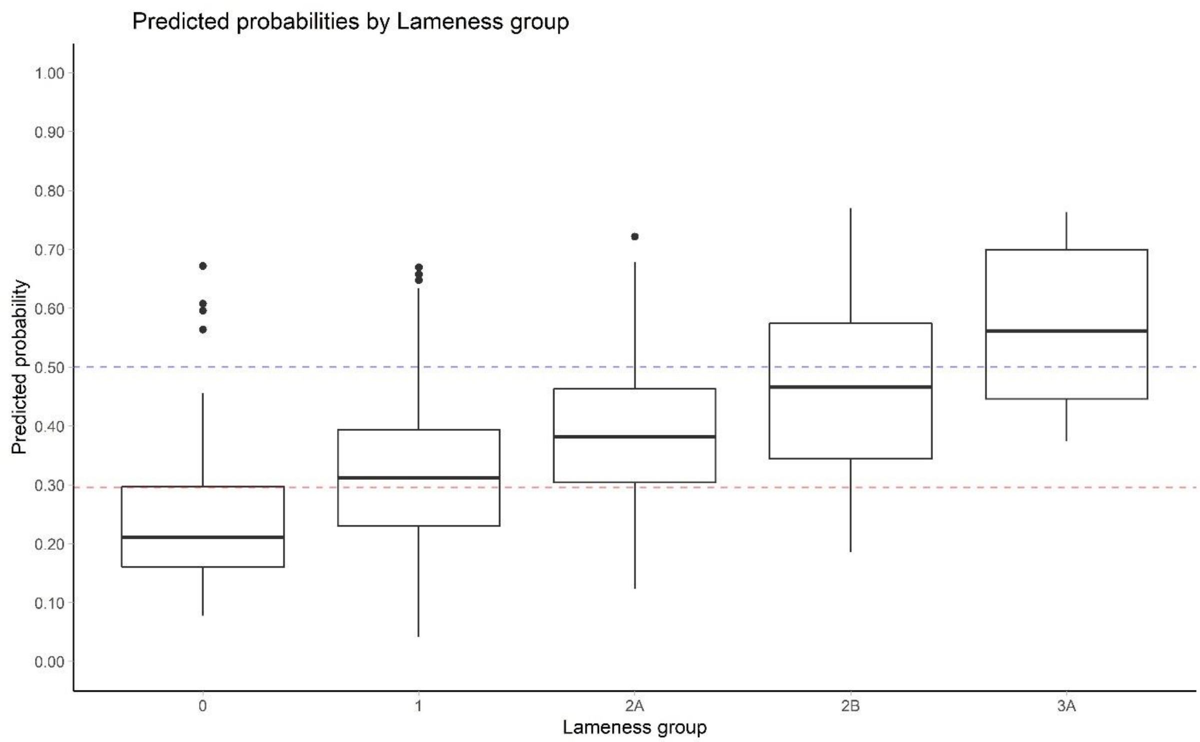

Figure 4 The distribution of the predicted probability mobility score for the ICEQube model. This graph shows the spread of predicted probabilities by lameness group from the ICEQube data model. The predicted probability is the probability of a cow receiving a lame mobility score based on the features included in the model. The predicted probabilities were generated on the test dataset containing 40% of the initial model dataset. The predicted probabilities have been plotted on the y axis and the mobility score on the x axis. The blue line demonstrates a probability cut-off of 0.5, whilst the red line shows the optimum probability threshold for highest combined sensitivity and specificity.

3.3 Lely Qwes-H data

90,966 individual Lely datapoints were captured. This was reduced to 7,158 when summarised in 24-hour periods.

3.3.1 Summary statistics split by lameness group

3.3.1.1 Total steps

Table 4 denotes the descriptive statistics from the Lely Qwes-H data by lameness group. The mean total steps for the non-lame group was 612.37 (steps/d) compared to 620.42 steps/d for the lame group, meaning the mean total steps was higher for the lame group than the non-lame. SD was large for both groups (147.81 for non-lame and 357.72 for lame) meaning the difference in total steps between groups was not likely to be significant.

Figure 5 shows the distribution of total steps by MS. In general, there was a decrease in total steps as MS increased. The range of total steps for MS 1, 2A and 2B was larger than for MS 0 and 3A. There was much less variability in total steps in the MS 0 group than the other MS groups. There was significant overlap in the distribution of the MS 1 group and the MS 2A group, whereas the MS 2B and MS 3A group had slightly lower IQR and mean. All of the MS 3A group had a total steps less than 600 steps per day.

3.3.1.2 Rumination minutes

The mean rumination minutes per day for the non-lame group was 474.05 minutes/day (m/d), and for the lame group was 466.65 m/d. The difference between these mean values is much smaller than the SD values of 50.87 m/d and 48.43 m/d for the non-lame and lame groups respectively. There was very little evidence of differences in rumination minutes between MS groups except for a tendency for slightly higher rumination minutes in the MS 0 group (Figure 5 and Appendix 6). All MS groups have a similar range, and are all highly variable with a large distribution in all MS groups (Figure 5).

Variable	Group	Mean	SD	Median	Minimum	Maximum	Range
Heat Alerts	Non-Lame	0.43	0.50	0.00	0.00	1.00	1.00
	Lame	0.29	0.45	0.00	0.00	1.00	1.00
Total steps	Non-Lame	612.37	147.81	600.00	200.50	2307.00	2106.50
	Lame	620.42	357.72	554.00	216.50	3036.00	2819.50
Total steps Diff from mean	Non-Lame	-0.44	140.80	-15.01	-359.39	1703.19	2062.58
	Lame	3.77	353.86	-61.25	-342.03	2431.81	2773.84
Rumination	Non-Lame	474.05	50.87	478.00	252.00	609.00	357.00
	Lame	466.65	48.42	470.00	294.00	593.00	299.00
Rumination Diff from mean	Non-Lame	4.69	46.53	7.29	-221.00	127.75	348.74
	Lame	-3.09	44.97	-1.31	-187.25	112.85	300.10

Table 4 Descriptive statistics for variables from the Lely Qwes-H sensor. This table gives the mean, standard deviation, median, minimum, maximum and range for each daily value by lameness group. This data was collected from Lely Qwes-H sensor and averaged into daily values. Non-lame cows are MS 0 and MS 1 and lame cows are MS 2A, MS 2B and MS 3A.

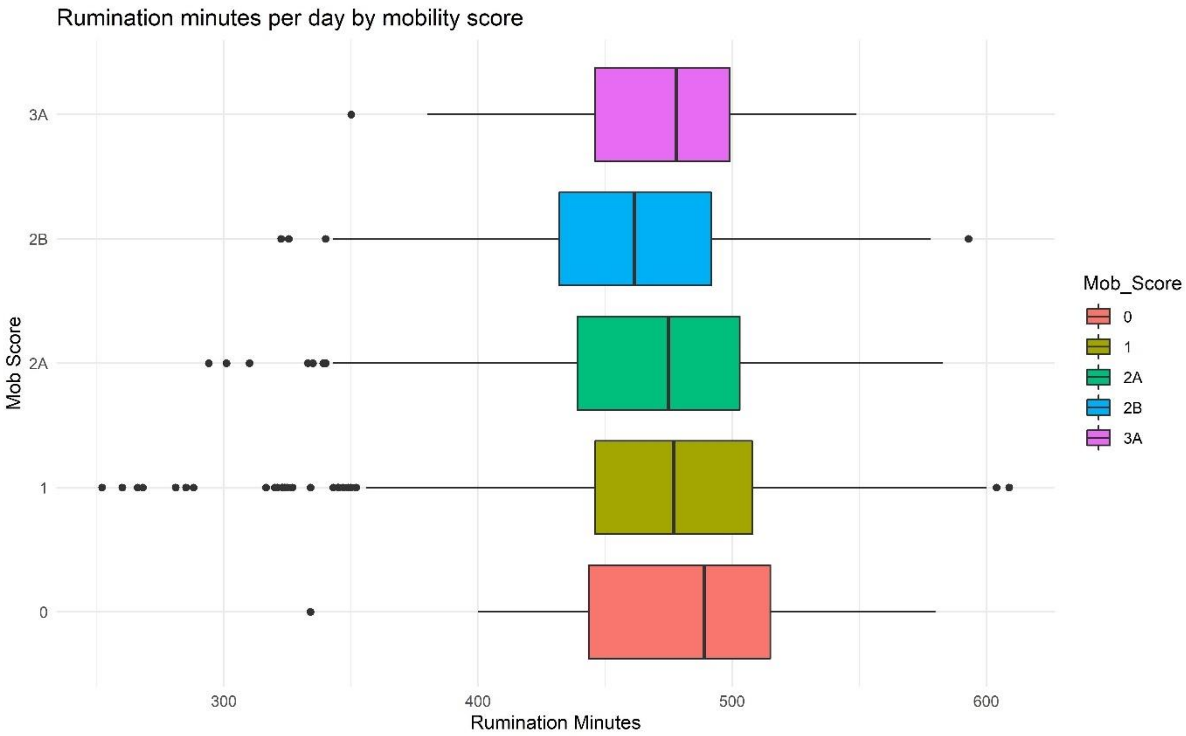

Figure 5 Daily rumination minutes from the Lely Qwes-H system by mobility score. This graph displays the daily rumination minutes on the x axis, and mobility score on the y axis. Rumination minutes were recorded using a rumination microphone built into the Lely Qwes-H system and summarised into daily values. Mobility scores were collected by two RoMS accredited mobility scorers on Monday and Thursday during the study period.

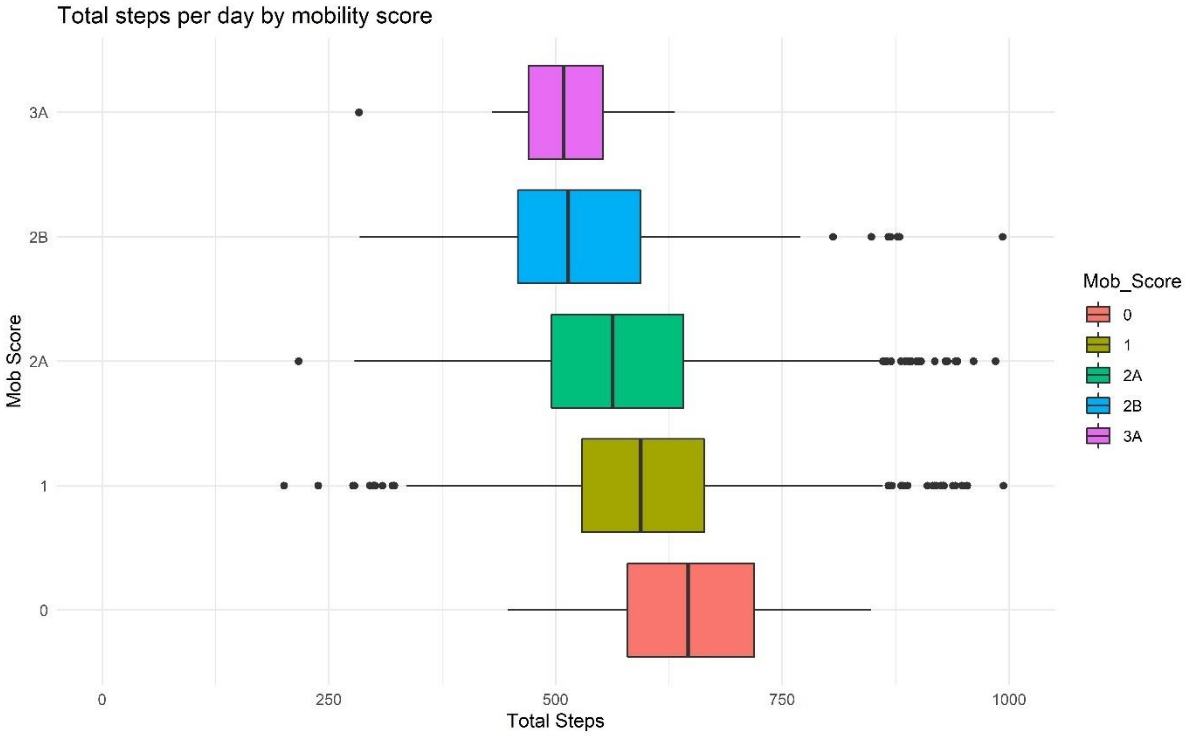

Figure 6 Total steps from the Lely Qwes-H system by mobility score. This graph displays the total steps on the x axis, and mobility score on the y axis. Number of steps were collected using a pedometer built in to the Lely Qwes-H system and summarised into daily values. Mobility scores were collected by two RoMS accredited mobility scorers on Monday and Thursday during the study period.

3.3.2 Model outcomes

3.3.2.1 Confusion matrix

The model built using features from the Lely Qwes-H sensor data (Lely model) did not perform as well as the ICE model on the test dataset, with sensitivity = 0.35, specificity = 0.89, kappa statistic = 0.27 and balanced accuracy = 0.61. The final mtry value for this model was 14. Figure 7 shows the distribution of predicted probabilities generated on the test dataset for this model. The poor sensitivity of this model is driven by the variable distribution of predicted probabilities of the lame group. The spread of predicted probabilities generated by this model are displayed in 0. The model does not manage to separate the MS groups well, with an overlap of P between the MS 1 and MS 2A groups.

3.3.2.2 Variable importance

Appendix 12 displays the relative importance of each feature for the Lely model. This graph details the individual features and their relative variable importance value to the model. Features relating to total steps were the most important in the Lely model, specifically, total steps for each of the previous 5 days which all had relative variable importance of above 90%. The relative importance of total steps on the day of mobility scoring was 73.66%. Rumination minutes and rumination difference from the mean also had high importance values, at 68.49 and 67.81 respectively. The least important variables in this model related to heat alerts, with importance values ranging from 0-17.41.

3.3.2.3 Calibration plot

The calibration of the Lely model is displayed in Figure 15. Calibration was relatively good, with P generally increasing as OEP increases. At P=0.1, there is a large confidence interval and a bin midpoint of 25%. At 0.2<P<0.5, the bin midpoints are close to the y=x line and the confidence intervals are much smaller, showing good calibration. At P=0.6-0.7, the OEP of the bin midpoints decreases slightly to below the y=x line and the confidence intervals increase in size. At P=0.8, OEP increases sharply to around 95%. At P>0.8, the confidence intervals increase to nearly 100% and the calibration is poor.

3.3.2.4 ROC curve, AUC and optimum threshold

As displayed in Figure 16, the Lely model had an AUC = 0.66, lower than the ICE model at 0.72. As shown in Table 7, the OT for the Lely model was 0.43 and applying this increased specificity from 0.27 to 0.55 and decreased specificity from 0.86 to 0.71. The NPV decreased (-0.29), the PPV increased (+11) and the balanced accuracy increased from 0.57 to 0.63.

3.3.2.5 Predicted probabilities by MS group

Figure 7 shows the distribution of predicted probabilities (P) for the Lely model per mobility score group. Descriptive statistics including the mean, SD, median, maximum, minimum and range P is detailed in Appendix 15. The generated P of Model 2 were not as polarising as for the ICE model. The mean P for MS 0 was 0.20, and for MS 3A was 0.43. The mean P for MS 1 was 0.27, for MS 2A was 0.33 and for MS 2B was 0.49. A range of P of 0.51 is shown in MS 0, and a range of 0.702 for MS 1. The range for MS 2A, 2B and 3A were large, at 0.76, 0.77 and 0.68 respectively. This was driven by high maximum values, at 0.54

for MS 0, 0.75 for MS 1, 0.78 for MS 2A, 0.86 for MS 2B and 0.79 for MS 3A. The median P for MS 2B is higher than that of MS 3, at 0.51 and 0.45 respectively. The median P for MS 0 is the lowest at 0.20, but close to MS 1 at 0.25 and MS 2A at 0.29.

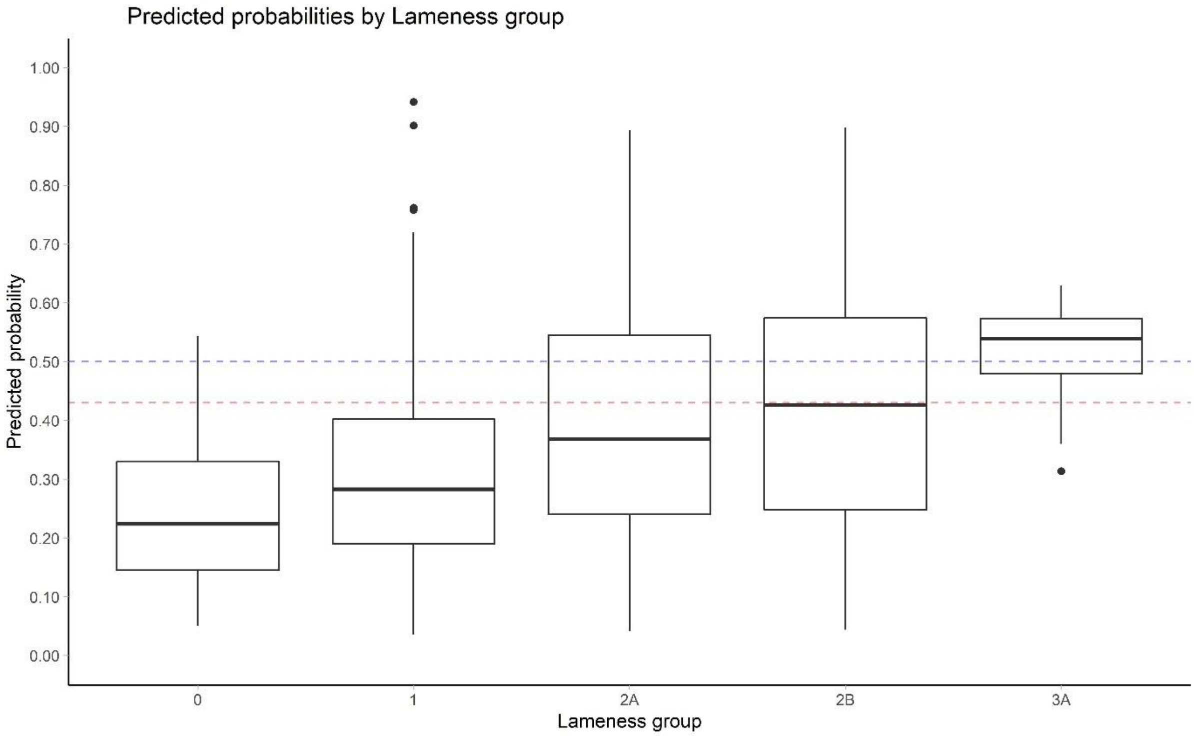

Figure 7 Predicted probabilities by mobility score from the Lely Qwes-H model. This graph shows the spread of predicted probabilities by lameness group from the Lely Qwes-H data model. The predicted probability is the probability of a cow receiving a lame mobility score based on the features included in the model. The predicted probabilities were generated on the test dataset containing 40% of the initial model dataset. The predicted probabilities have been plotted on the y axis and the mobility score on the x axis. The blue line demonstrates a probability cut-off of 0.5, whilst the red line shows the optimum probability threshold for highest combined sensitivity and specificity.

3.4 Production data

Data from 31,846 individual milkings were collected. This was summarised to include 10,291 full 24-hour periods.

3.4.1 Summary statistics split by lameness group

Table 5 shows summary statistics of the production data split by lameness group. The daily yield of the lame group was 44.01 litres/day (L/d), similar to that of the non-lame group at 42.09 L/d. SD of for both lame and non-lame groups was high (10.55 and 11.42 L/d respectively) indicating significant variability in daily yield for both groups. The largest variability was seen in the MS 3A group, with an IQR of 33.4 (Figure 8). Figure 8 describes the daily yield of study cows by mobility score over the study period. Large variability within each MS group is shown by the large range of milk yield in each group. The range of MS 1 is the smallest at 16.3 (Figure 8). There is a slight increase of mean with an increase in MS (Appendix 6 and Figure 8), however the standard deviations of each overlap. Large variability in the distribution of each group means there is only a slight increase of MDP with an increase in MS.

3.4.1.1 Milk constituents

Fat percentage, lactose percentage and protein percentage were similar between non-lame and lame groups (Table 5). When examining milk constituents by MS group (Figure 9, Figure 10, Figure 11), distributions of all three constituents were similar between the MS 1, MS 2A and MS 2B groups. Protein percentage does increase slightly with mobility score, with MS 2B and MS 3A having had a slightly higher protein percentage than MS 1, and MS 0 having the lowest mean in milk protein percentage. In the MS 0 group there was less variability in fat percentage with an IQR of 0.63. The MS 0 group also had a slightly higher lactose percentage than the other MS groups, with a mean of 3.92 (Appendix 6). The MS 3 group had a greater spread of fat percentage and protein percentage values and a smaller spread of lactose values than the other MS groups. The IQR for MS 3 of fat percentage is 0.63, for protein percentage is 0.27 and for lactose percentage is 0.09. These differences may reflect differences in the parity of cows in different MS groups, with a higher proportion of heifers in the MS 0 group and a higher proportion of older cows in the MS 3 group. Overall, there is a high degree of variability of fat percentage, lactose percentage and protein percentage values within MS groups and the similarity in distributions of these variables between MS groups (Figure 9, Figure 10, Figure 11) suggests that milk constituents are not likely to be strong predictors for lameness.

3.4.1.2 Milking behavioral data

Milking behavior differs slightly between the two groups. The mean number of milkings per day for the lame group was similar to the non-lame group, at 3.09 and 3.26 (milkings/d) respectively. When examining the distribution of milkings/day by MS groups, the MS 0, MS 1 and MS 2A groups were very similar (Figure 12), with 75% of cows in these groups having had at least 3 milkings/d. There was a tendency for cows in the MS 2B and MS 3 groups to have fewer milkings/d, with a greater proportion of cows in these groups having had 1 or 2 milkings/d compared to the MS 0, MS1 and MS 2A groups (Figure 12).

Number of refusals per day number of refusals was negatively skewed with most cows having 0 or 1 refusals/d (Appendix 8 - histogram). There was a slight difference between non-lame and lame groups, with median values of 1 and 0 refusals/d respectively (Table 6). Similarly, NR tended to increase with decreasing MS (Figure 13 and Appendix 6).

Table 5 Descriptive statistics per lameness group for each production feature. This table gives the mean, standard deviation, median, minimum, maximum and range for each daily value by lameness group. This data was collected from the robots at each milking and averaged into daily values. Non-lame cows are MS 0 and MS 1 and lame cows are MS 2A, MS 2B and MS 3A.

Variable	Group	Mean	SD	Median	Minimum	Maximum	Range
Average Weight	Non-Lame	717.36	87.56	715.00	528.00	994.00	466.00
	Lame	763.28	99.00	757.00	542.00	1060.00	518.00
Daily yield	Non-Lame	42.09	10.55	41.30	10.80	70.10	59.30
	Lame	44.01	11.42	43.45	14.60	72.20	57.60
Daily yield average	Non-Lame	42.31	10.01	41.30	14.80	69.90	55.10
	Lame	44.40	10.86	43.30	19.70	67.30	47.60
DIM	Non-Lame	160.98	66.18	150	29	324	295
	Lame	160.39	62.28	148	35	331	296
Failures	Non-Lame	0.06	0.28	0.00	0.00	3.00	3.00
	Lame	0.04	0.23	0.00	0.00	2.00	2.00
Fat Percentage	Non-Lame	3.92	0.85	3.84	2.04	7.04	5.00
	Lame	3.92	0.93	3.80	1.86	6.95	5.09
ISK	Non-Lame	62.97	11.48	62.90	16.50	98.80	82.30
	Lame	62.21	11.97	63.75	24.90	89.60	64.70
Parity	Non-Lame	2.01	1.15	2.00	1.00	5.00	4.00
	Lame	2.69	1.32	3.00	1.00	5.00	4.00
Lactose Percentage	Non-Lame	5.02	0.12	5.04	4.39	5.31	0.92
	Lame	4.99	0.12	5.00	4.68	5.26	0.58
Milkings	Non-Lame	3.26	0.92	3.00	1.00	9.00	8.00
	Lame	3.09	0.98	3.00	1.00	8.00	7.00
Protein percentage	Non-Lame	3.32	0.24	3.30	2.57	4.06	1.49
	Lame	3.34	0.26	3.34	2.72	4.17	1.45
Refusals	Non-Lame	2.75	4.44	1.00	0.00	57.00	57.00
	Lame	1.91	3.94	1.00	0.00	53.00	53.00

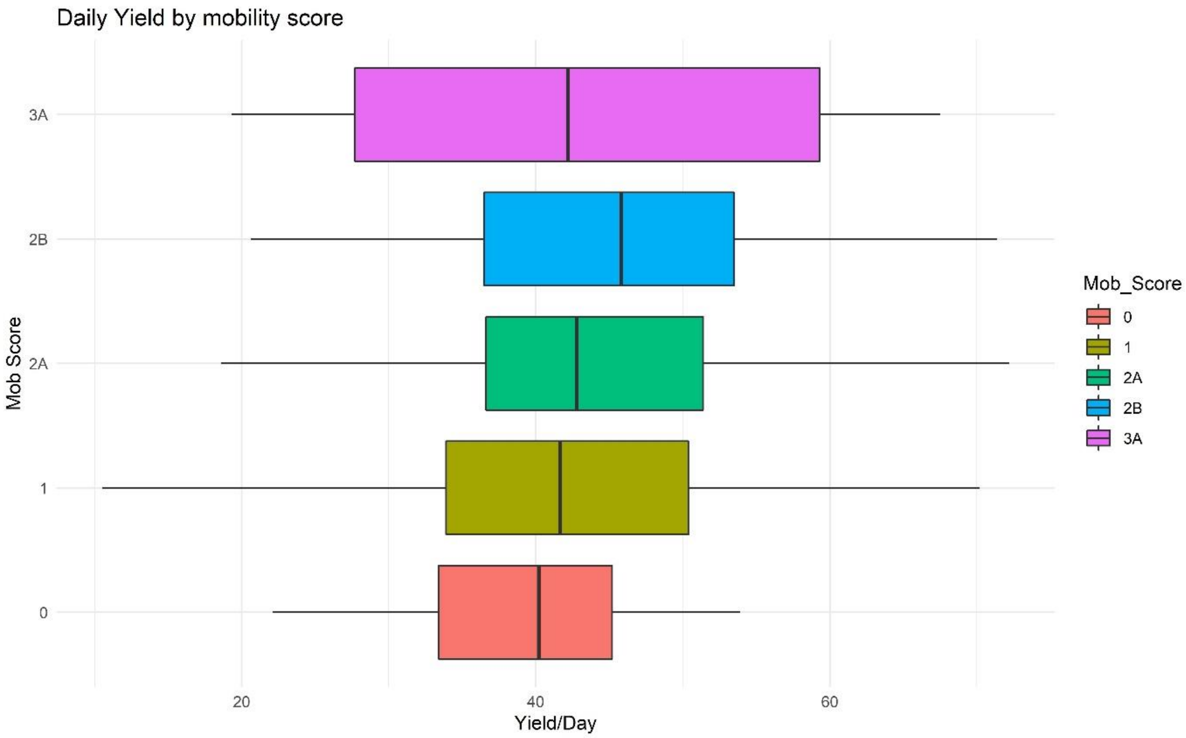

Figure 8 Yield per day (L/d) by mobility score. This graph shows the daily yield per cow per day by mobility score, with yield in L/d on the x axis and mobility score on the y axis. Yield was collected from the AMS at each milking and averaged over the milkings within the day. Mobility score was collected by one of two RoMS qualified operators on every Monday and Thursday during the trial period.

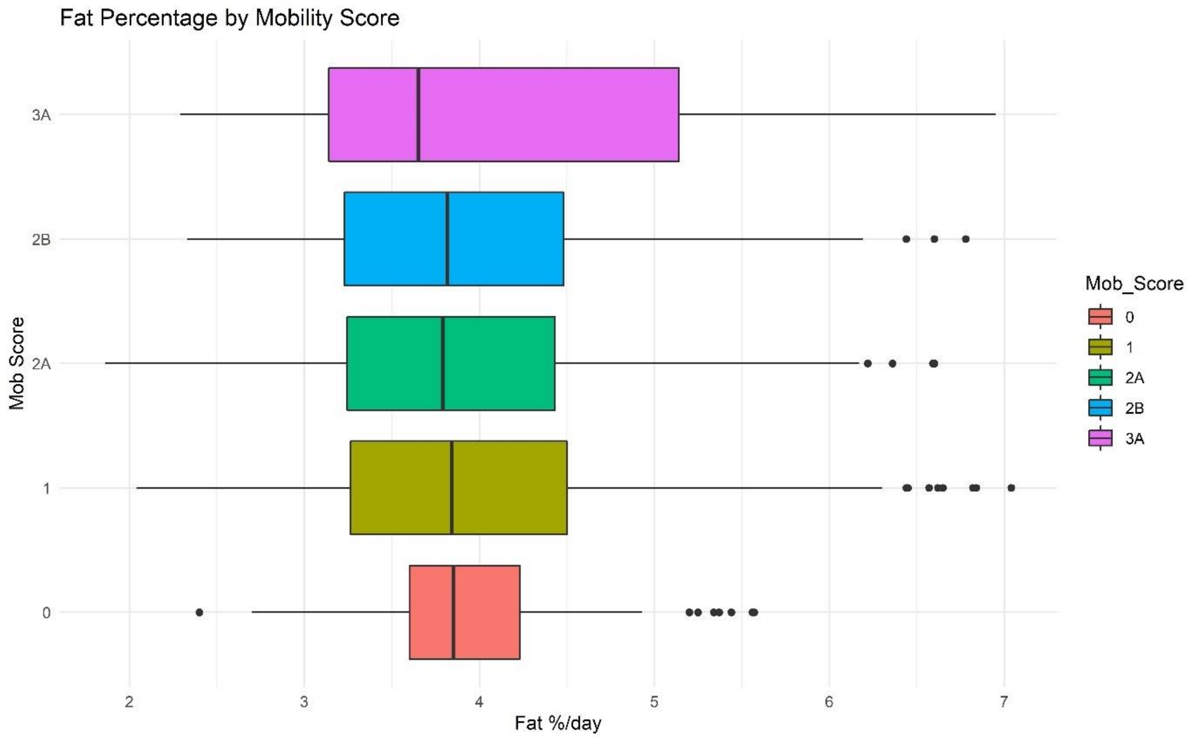

Figure 9 Fat percentage per day by mobility score. This graph shows the fat percentage per cow per day by mobility score. Fat percentage is displayed on the x axis and mobility score on the y axis. Fat percentage was collected from the AMS at each milking and averaged over the milkings within the day. Mobility score was collected by one of two RoMS qualified operators on every Monday and Thursday during the trial period.

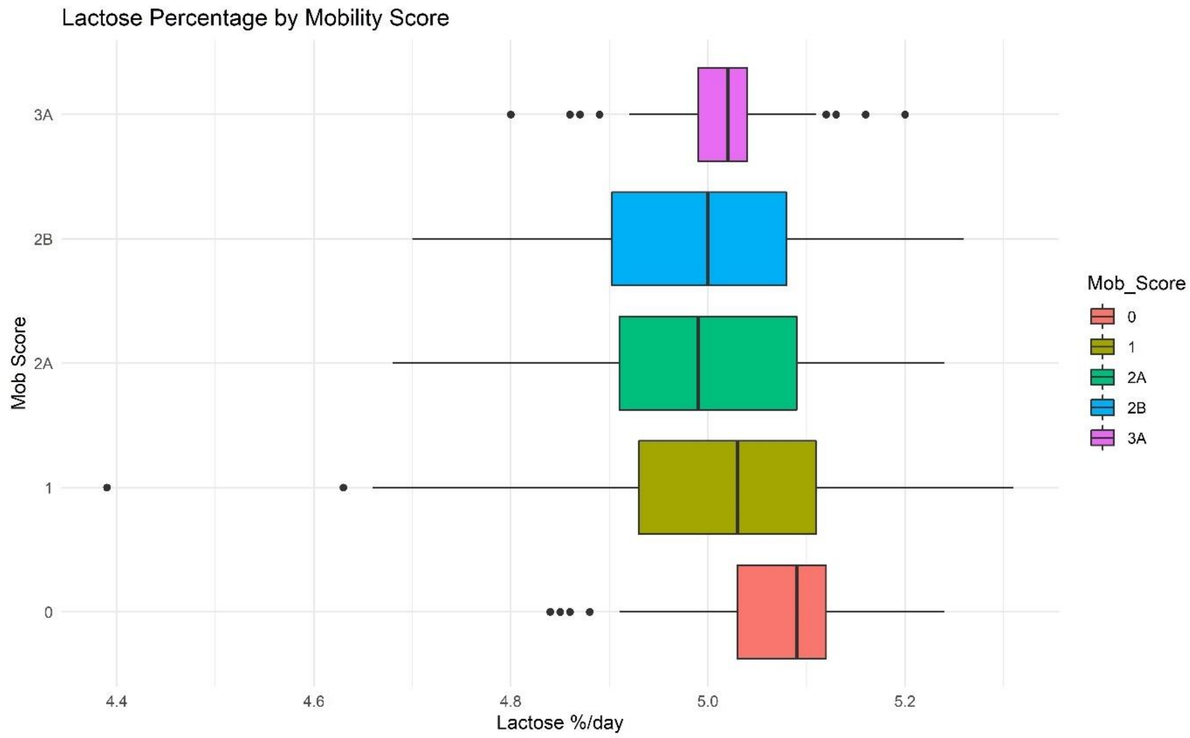

Figure 10 Lactose percentage per day by mobility score. This graph shows the lactose percentage per day by mobility score. Lactose percentage is displayed on the x axis and mobility score on the y axis. Lactose percentage was collected from the AMS at each milking and averaged over the milkings within the day. Mobility score was collected by one of two RoMS qualified operators on every Monday and Thursday during the trial period.

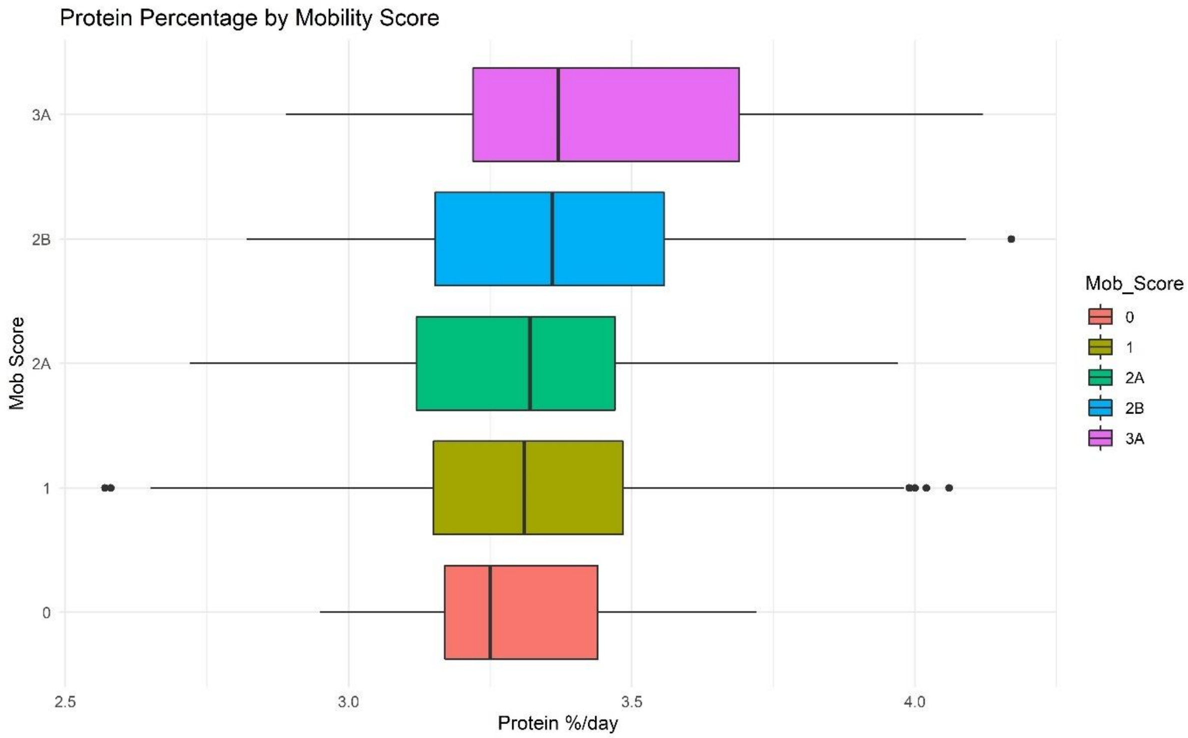

Figure 11 Protein percentage per day by mobility score. This graph shows the protein percentage per day by mobility score. Protein percentage is displayed on the x axis and mobility score on the y axis. Protein percentage was collected from the AMS at each milking and averaged over the milkings within the day. Mobility score was collected by one of two RoMS qualified operators on every Monday and Thursday during the trial period.

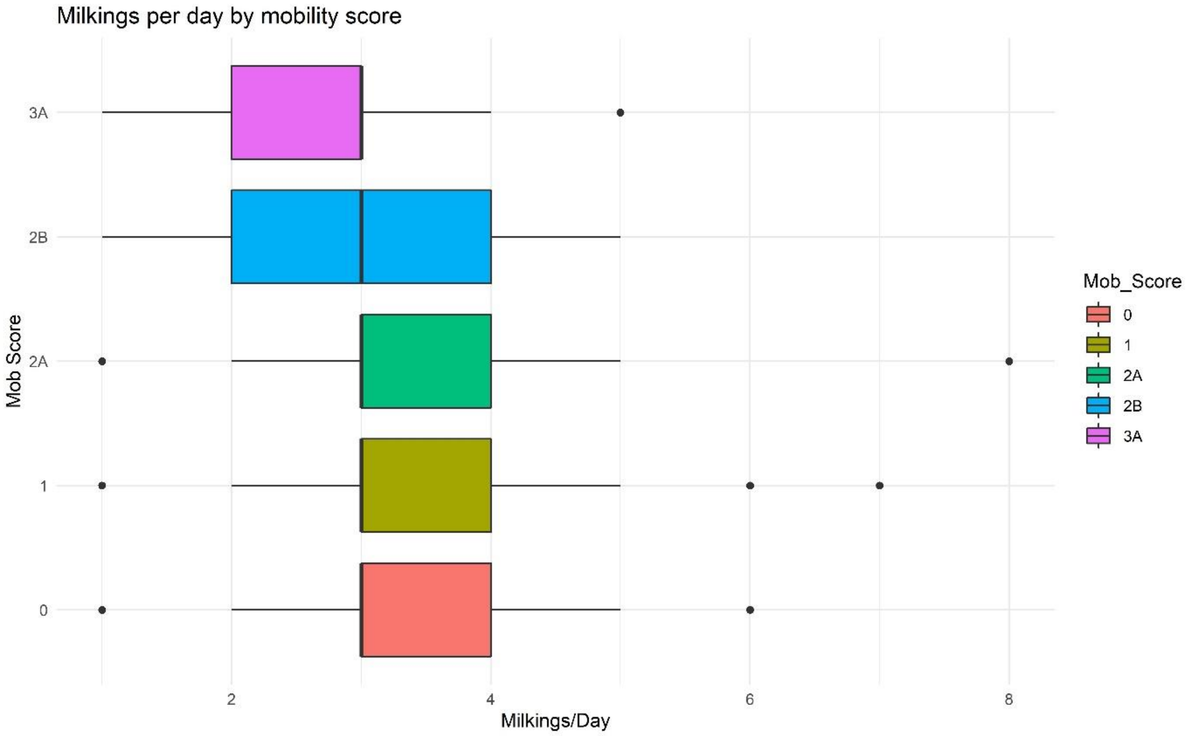

Figure 12 Number of milkings per day by mobility score. This graph shows the number of milkings per cow per day by mobility score. Mobility score is plotted on the y axis and number of milkings is plotted on the x axis.

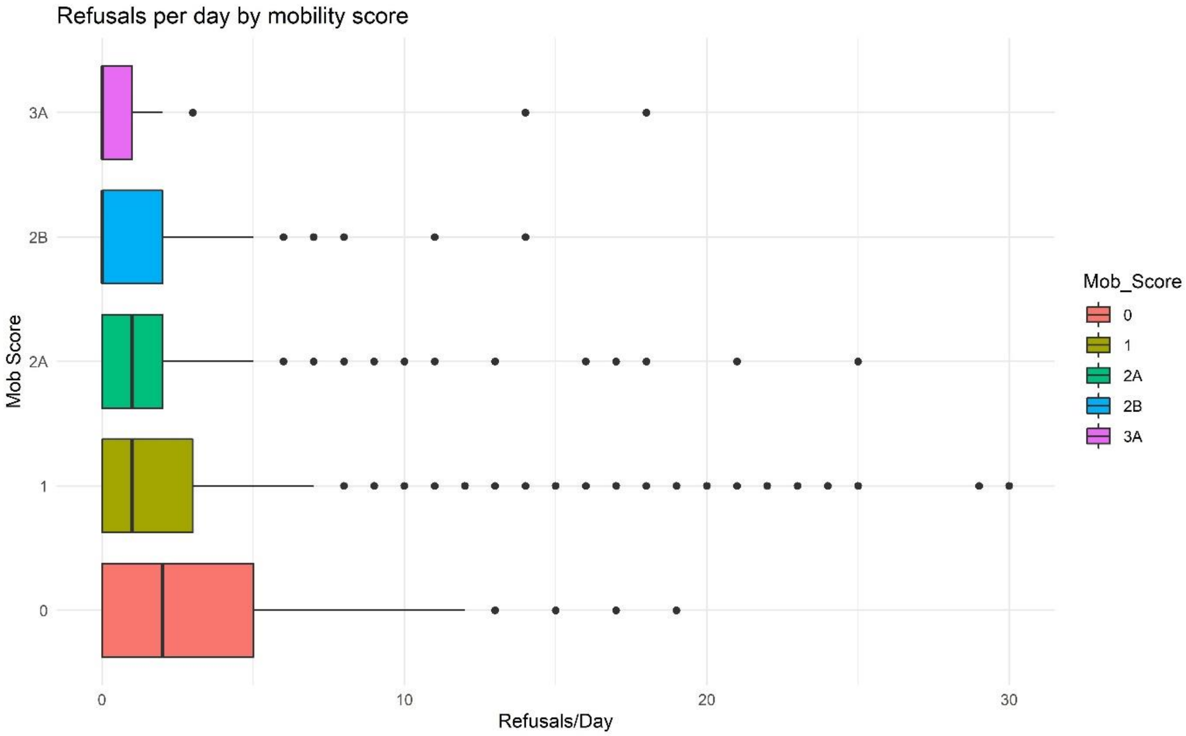

Figure 13 Number of refusals per day per cow by mobility score. This graph demonstrates the number of refusals per cow per day by mobility score. Mobility score is plotted on the y axis and number of refusals on the x axis.

3.4.2 Model outcomes

3.4.2.1 Confusion matrix

The model built from production data (Production model) performed slightly worse than the ICE model on the test dataset, but better than the Lely model (Table 6), with sensitivity = 0.40, specificity = 0.90, kappa statistic = 0.33 and balanced accuracy = 0.65. The final mtry value for this model was 3. The generated predicted probabilities on the test data are displayed in Appendix 15.

3.4.2.2 Variable importance

Appendix 13 details the individual features and their importance value relative to this model. Overall, variables relating to Average Weight were the highest contributors to the Production model. The 7-day weight average was the most important, and the average weight for the previous 7 days were also contributing highly to the model, with importance values > 88%. Several features not relating to average weight were also important, such as 7-day SD of MDP, as well as the 7-day average of number of refusals. These have an importance value of 89.97, 86.58 and 87.75 respectively. The features with the lowest importance values are relating to milkings per day, number of failures and number of refusals. The relative importance for the milkings per day is 29.3, the failures per day is 0 and the number of refusals is 41.01.

3.4.2.3 Calibration curve

Calibration for the Production model was worse than the ICE or Lely models. At P=0.1, the confidence interval of the OEP is large, and the OEP bin midpoint is at 0. From P=0.2 to P=0.5, calibration is good as the OEPs of the bin midpoints follow the y=x line and the confidence intervals are small. When P=0.6 and when P=0.7, there OEP is very high and does not fit the y=x line. This leads to a steep increase in the calibration curve and poor calibration. This remains high, until at P=0.8 and P=0.9, the OEP=100 and the confidence intervals are large. Calibration becomes poor at these probability values.

3.4.2.4 ROC curve, AUC and optimum threshold

As shown in Figure 16, the Production model had an AUC of 0.79, this was higher than either the ICE or Lely models (0.72 and 0.66 respectively). As shown in Table 7, the optimum threshold for the Production model was 0.31, and when this was applied sensitivity increased from 0.40 to 0.64 and specificity decreased from 0.90 to 0.80. The PPV increased from 0.68 to 0.86, and the NPV decreased from 0.74 to 0.54. The balanced accuracy increased from 0.65 to 0.72, and kappa increased from 0.33 to 0.40.

3.4.2.5 Predicted probabilities by MS group

Figure 14 shows the distribution of predicted probabilities (P) for each model per mobility score group. Descriptive statistics including the mean, SD, median, maximum, minimum and range P is detailed in Appendix 15. For the Production model, the mean P for MS 0 is 0.19. The mean P for MS 3A is 0.63. The mean P for MS 1, 2A and 2B increase as MS increases, at 0.30, 0.40 and 0.53. The minimum and maximum P for MS 0 is 0.05 and 0.43, with a range of 0.38. This means at an OT of 0.5, no MS 0 cows were classified as lame. The range of MS 3A was 0.33, with a minimum of 0.45 and a maximum of 0.78. The range for MS 1 was larger at 0.63, for MS 2A was 0.59 and for MS 2B was 0.58. The median for each MS increased as MS increased, so for MS 0 was 0.18, for MS 1 was 0.29, for MS 2A was 0.39, for MS 2B was 0.54 and for MS 3A was 0.65.

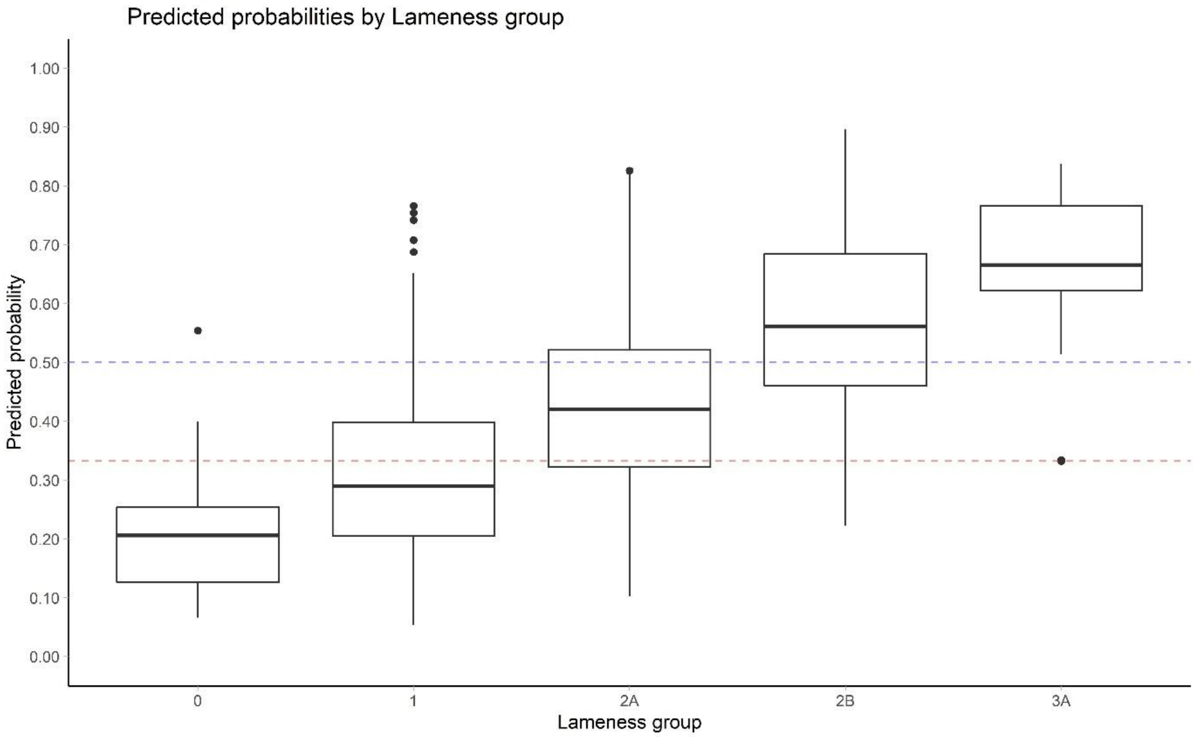

Figure 14 Predicted probabilities by mobility score for the Production data model. This graph shows the spread of predicted probabilities by lameness group from the ICEQube data model. The predicted probability is the probability of a cow receiving a lame mobility score based on the features included in the model. The predicted probabilities were generated on the test dataset containing 40% of the initial model dataset. The predicted probabilities have been plotted on the y axis and the mobility score on the x axis. The blue line demonstrates a probability cut-off of 0.5, whilst the red line shows the optimum probability threshold for highest combined sensitivity and specificity.

3.5 Combined model

3.5.1 Confusion matrices

The model built combining features from all three datasets (Combined model) was the best performing model on the test data in terms of sensitivity and specificity (Table 6). The sensitivity and specificity of this model was 0.36 and 0.95 respectively. This model had a kappa statistic of 0.36, and a balanced accuracy of 0.66. The distribution of predicted probabilities generated on the test dataset for this model is shown in Figure 17. Generally, the non-lame group are displayed to have predicted probabilities below the threshold value of 0.5. Half of the lame group, the majority being score 2A, also have a predicted probability of below this threshold, generating the low sensitivity, or false negative rate, of this model. The spread of predicted probabilities generated by this model are displayed in Appendix 15. The mean predicted probability generated by this model for the non-lame group is 0.30 for the lame group is 0.41, showing a difference in mean that means accuracy could be improved by applying an optimum threshold.

Table 6: Discrimination statistics for all 4 models. This table shows the discrimination statistics for all 4 models, including the value of mtry used in the model, the sensitivity, specificity, positive predictive value, negative predictive value, balanced accuracy and Cohen's kappa statistic. This was created using a default probability cut-off of 0.5.

	Model			
Value	ICEQube	Lely Qwes-H	Production	Combined
Mtry	2	14	3	10
Sensitivity	0.26	0.35	0.40	0.36
Specificity	0.94	0.89	0.90	0.95
PPV	0.68	0.63	0.68	0.79
NPV	0.71	0.72	0.74	0.74
Balanced Accuracy	0.60	0.61	0.65	0.66
Kappa	0.23	0.27	0.33	0.36

3.5.2 Variable importance

Appendix 14 displays the individual importance value for each variable in this model. The features with the highest importance value for this model are related to average weight. The average weight Lag 1:4 are the most important features, with an importance value ranging from 91.13-100%. The next most important feature is lying time, with an importance value of 72.43%. Several other features relating to lying time are also contributors to the model with high importance values, such as lying time 7-day average and lying time difference from the daily herd mean, with importance values of 68.76% and 70.24% respectively. The features with the lowest importance value were related to lying bouts, failures per day and milkings per day. Features relating to failures per day had the lowest importance values overall, with an importance ranging from 0-3.60. Milkings per day-related features had a range of importance values from 7.03%-12.71%. Lying bouts had a range of importance values from 15.90%-19.49%.

3.5.3 Calibration curve

Good calibration was observed for the Combined model (Figure 15), where the OEPs were often further away from the y=x line. Although the curve increases as the bin increases, it does not increase as steep as the others and does not follow the y=x line as closely. The confidence intervals are large but do not get as large as the other curves at P>0.8. The model follows good calibration until P=0.4, where the OEP does not change from P=0.3. At P=0.5, the OEP of the bin midpoint increases towards the y=x line. At P=0.6, there is an increase in the OEP of the bin midpoint, leading the midpoint to be further away from the y=x line. At P>0.6, the confidence intervals are relatively small compared to the other models, and the OEP of the bin midpoints follow the y=x line well.

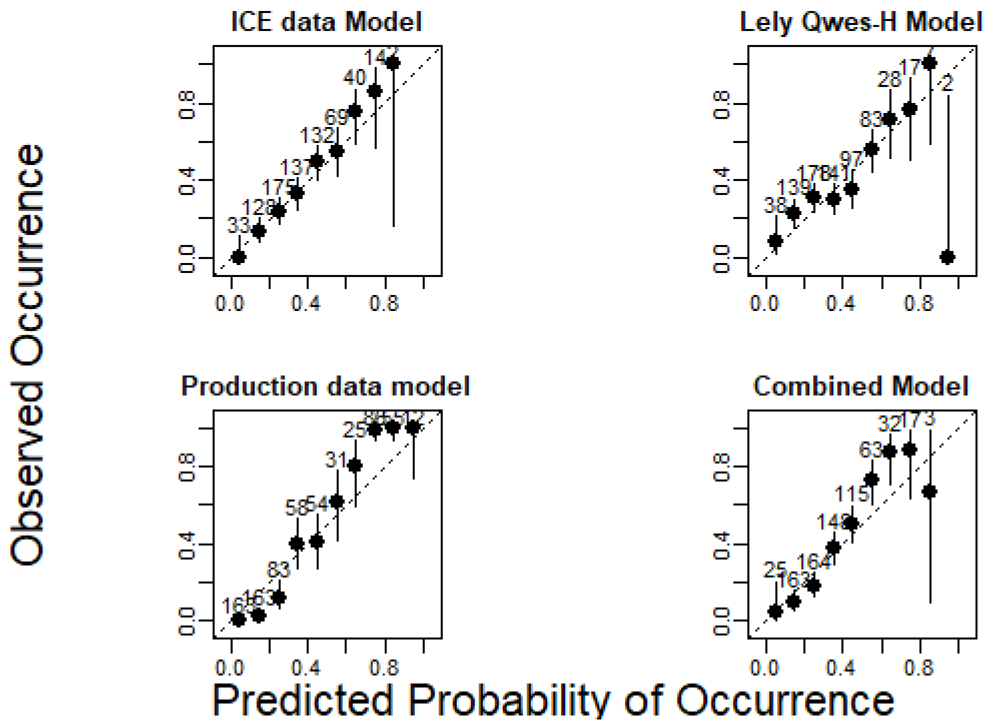

Figure 15 Calibration curves for each model. This figure shows a calibration curve for each model. A calibration curve is a measure of model accuracy. It divides the predicted probabilities of 0-1 in 10 bins and plots their mid points on the x axis. The y axis denotes the number of samples whose class is positive in that bin. Each dot is plotted with the bin midpoint on the y axis, and the average predicted probability of all the cows in that group on the x axis. The numbers denote the number of cows in each bin. A well calibrated model should have a curve that hugs the y = x line, as the predicted probability of occurrence and observed event percentage should be equal to achieve perfect calibration.

3.5.4 ROC Curve, AUC and optimum threshold

The AUC for the Combined model was 0.8. The OT for this model was 0.34; applying this threshold increased sensitivity from 0.36 to 0.75 and decreased specificity from 0.95 to 0.71. This model also had a PPV of 0.85 (+0.16) and a NPV of 0.57 (-0.127). Kappa statistic increased from 0.36 to 0.43, and balanced accuracy saw a small increase from 0.66 to 0.72.

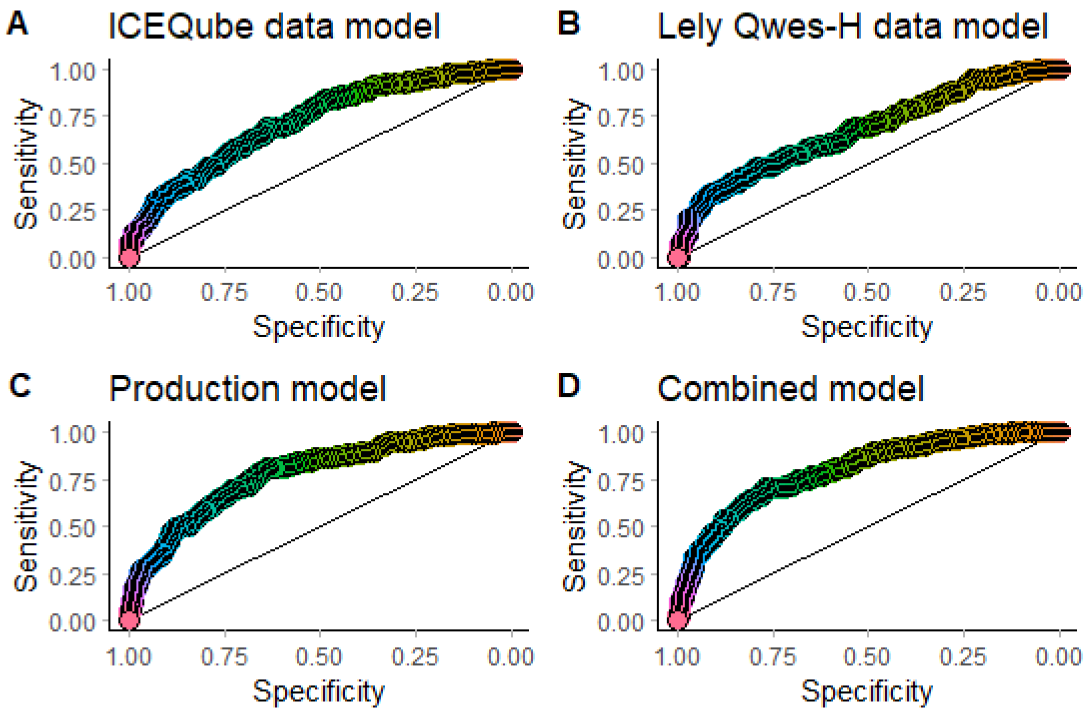

Figure 16 Receiever-Operator Characteristic (ROC) Curves for each model This figure shows a ROC Curve for each model, with each point denoting a probability cutoff value, with the sensitivity and specificity of that cutoff value on the x and the y axis respectively. A shows a ROC curve for the model based on the ICEQube data, B shows a ROC curve for the Lely Qwes-H model, C shows a ROC curve based on the Production data model and D shows a ROC curve based on the combined model.

Table 7 Discrimination statistics after optimum cut-off This table shows the sensitivity, specificity, positive predictive value, negative predictive value, balanced accuracy and Cohen's kappa statistic for each of the 4 models constructed after an optimum probability cut-off was used.

	Model			
Value	ICEQube	Lely Qwes-H	Production	Combined
Optimum cutoff	0.30	0.43	0.31	0.34
Sensitivity	0.50	0.80	0.64	0.71
Specificity	0.82	0.46	0.80	0.75
PPV	0.84	0.74	0.86	0.85
NPV	0.46	0.54	0.54	0.57
Balanced accuracy	0.66	0.63	0.72	0.72
Kappa	0.27	0.28	0.40	0.43

3.5.5 Predicted probabilities by lameness group

Figure 17 shows the distribution of predicted probabilities (P) for the model per mobility score group. Descriptive statistics including the mean, SD, median, maximum, minimum and range P is detailed in Appendix 15. This model achieved the best separation in terms of distribution of P across all MS groups. For the Combined model, the mean P with MS 0 is 0.17, and the mean P of MS 3A is 0.68, and a standard deviation of 0.11 and 0.19 respectively. MS 2B also has a high mean P at 0.59, and the mean P of MS 2A is slightly lower at 0.41. The mean P of MS 1 is 0.29. The SD of MS 1, 2A and 2B range between 0.16 and 0.19. There is a large range of P in all MS groups. The maximum P for MS 0 is 0.46, for MS 1 is 0.85, for MS 2A is 0.84, for MS 2B is 0.88 and for MS 3A is 0.90. The minimum P for MS 0 is 0.01, and the minimum P for MS 3A is 0.328. The median P for MS 1 is 0.16, for MS 2 is 0.28, for MS 2A is 0.42, for MS 2B is 0.61 and for MS 3A is 0.75.

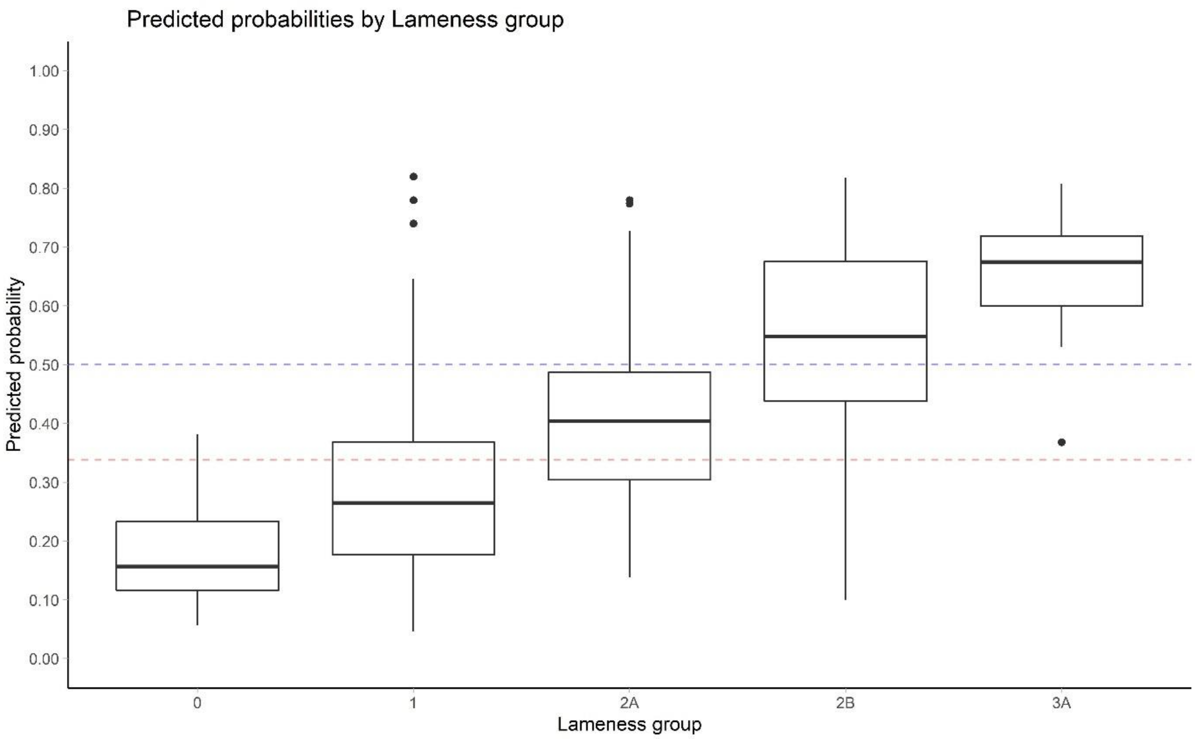

Figure 17 Predicted probabilities by mobility score for the combined model. This graph shows the spread of predicted probabilities by lameness group from the ICEQube data model. The predicted probability is the probability of a cow receiving a lame mobility score based on the features included in the model. The predicted probabilities were generated on the test dataset containing 40% of the initial model dataset. The predicted probabilities have been plotted on the y axis and the mobility score on the x axis. The blue line demonstrates a probability cut-off of 0.5, whilst the red line shows the optimum probability threshold for highest combined sensitivity and specificity.

3.5.6 Comparison of model performance

All models observed have relatively good calibration, as they all follow a general linear trend. Calibration in all plots is poor at higher predicted probabilities, leading to large confidence intervals at Probability (P)>0.8 for all models. There are limited observations with a P>0.8 in all models, with 4 in the ICE model, 6 in the Lely model, 2 in the Production model and 12 in the combined model. There was better calibration at higher predicted probabilities in the Combined model than the other models, as the Observed Event Percentages (OEPs) of the bin midpoints were closer to the y=x line and the confidence intervals were smaller than for the other models. In the Production model, there were no observations with a P>0.9, leading to a steep decrease in the calibration curve, as the OEP was 0. The ICE and Lely models followed the same pattern of large confidence intervals in the OEP at P>0.8, leading to poor calibration above these values. The combined model had the highest AUC at 0.8 Lely model received the lowest AUC at 0.66. None of the ROC Curves demonstrated the ideal L shaped curve that means good balance between specificity and sensitivity, although the combined model achieved the highest combined sensitivity and specificity when an optimum threshold was applied. This had an associated sensitivity and specificity of 0.75 and 0.71. The Lely model had the lowest sensitivity and specificity at the optimum of 0.43 at 0.80 and 0.46 respectively.

Across all the models, there was a general trend for predicted probabilities to increase with mobility score. The Combined model showed the greatest differentiation between mobility score groups.

3.6 Performance of all models on severely lame cows

Table 8 demonstrates confusion matrices based on only non-lame and severe lame cows with a new optimum threshold. All models have an improved performance when being used to predict only severe lame cows. The combined model remains the model with the highest performance, at a sensitivity of 0.91 and a specificity of 0.80. It also has the highest kappa statistic at 0.6. The Lely model has had a large improvement in performance, from a balanced accuracy of 0.63 to a balanced accuracy of 0.7. The ICEQube model has also had increased balanced accuracy, at 0.73 from 0.66. The negative predictive value of both the ICEQube model and the Lely model are both low, at 0.25 (-0.21) and 0.3 (-0.24 respectively. **Error! Not a valid link.**

4 Discussion

4.1 Key research findings

This study aimed to improve sensitivity, specificity, and accuracy of lameness prediction from accelerometer data by combining with production data. Overall, my study demonstrated differences in both behavioural and productive features between lame and non-lame cows, although these differences were rarely significant. Results indicate substantial differences between several features: lying time, total steps from the ICEQube sensor, number of refusals, number of milkings, parity and average weight. There were slight differences in ISK and daily yield. Behavioural features such as lying time varied greatly within an individual, with an average SD of 1h/day and a SD of 250 steps/day for total steps. Variables associated with lying time, total steps, average weight and parity were the most significant for predicting lameness overall, but several features were important to each model independently. For example, milk protein percentage and ISK were important features in the production model alone. Total steps was important to the Lely model alone. Increased sensitivity and specificity were obtained by combining production data with activity and behavioural data, at values of 0.75 and 0.71 respectively. Model performance also increased when predicting moderate or severe lameness over mild lameness, driven by the fact that mild lameness causes mild changes in the behavioural features evaluated in this study.

4.2 Combined sensor and production data

A key finding from our study was that the model built using combined sensor and production data performed better than models built using sensor or production data alone. The combined model with the optimum threshold applied had the highest sensitivity, specificity, and AUC of any model in this study (0.71, 0.75 and 0.82 respectively). Other studies have also achieved better performance by combining measures of behaviour and production with activity than by using one set of data alone (Alsaaod et al., 2019; O'Leary et al., 2020). The performance of the combined model in this study was similar to that achieved by Kamphuis et al. 2013) who combined milking order, liveweight and activity data to achieve an AUC of 0.74, higher than that achieved using each feature on their own. Borghart et al (2021) developed four different random forest models comprising various activity and production features. Their best model achieved a higher sensitivity and specificity than the model created in this study, at 0.78 and 0.78 respectively, with an AUC of 0.89. It must be highlighted, however, that Borghart et al. (2021) used Sprecher locomotion scoring, which focuses on back arch rather than gait (Sprecher et al., 1997). Furthermore, their classification of lameness was different to that used in the current study, with mildly lame cows (Sprecher locomotion score > 1) being included in the "non-lame" category due to insufficient number of moderate to severe lame cows (Borghart et al., 2021).

When comparing the three models built on individual datasets, the model based on production data achieved a higher sensitivity and specificity than the models based on sensor data alone. Whilst it appears that production data provides improved predictive accuracy compared to activity data, many of the production variables are confounded by factors such as DIM and parity, which show a strong association with lameness (Kamphuis et al., 2013; Reader et al., 2011; Schindhelm et al., 2017). Illustrating this point, Borghart et al. (2021) found that a random forest model based on parity and DIM had better sensitivity, specificity and AUC for predicting lameness than a random forest model based on behavioural metrics such as activity. This is discussed in more detail in Section 4.4.4.

Performance of the model for predicting lameness from ICEQube behavioural data was poor, particularly with respect to sensitivity which was 0.50. Our models have similar performance than those investigated in previous literature. Kamphuis et al. (2013) achieved similarly poor performance with a model based only on activity and lying time data across five study farms, with sensitivity of 0.4 and a specificity of 0.8; this is also corroborated by earlier work (Bicalho et al., 2007; Miekley et al., 2013). Alsaaod et al. (2012) reported low accuracy (0.66) when predicting lameness using only lying time data. These authors described high variation in lying time between cows and within individual cows independent from locomotion score, which made it difficult to find universal thresholds for lameness. High variation within a cow was also observed in this study, and this is likely to have contributed to the poor predictive performance of the ICEQube model. There was a mean daily standard deviation of 1 hour/day lying time, and a mean daily standard deviation of 250 steps. These differences present challenges for a model when attempting to predict lameness, as a reduced daily activity or a reduced daily lying time could be just a standard variation in activity, rather than being due to lameness. Other studies also report highly variable behaviour both between and within cattle (Alsaaod et al., 2012;

Byabazaire et al., 2019; Cramer et al., 2009b; Grimm et al., 2019; Thorup et al., 2015), of which very little is attributable to lameness (Reader et al., 2011).

Various strategies have been implemented to attempt to overcome variation in activity data. Alsaaod et al. (2012) used deviation from normal behaviour in their models with an improvement in accuracy of their final model. Another study conducted by Byabazaire et al. (2019) applied a clustering-based approach, where they categorised animals by their behaviour every 2 weeks. Cattle were either considered normal, dormant or active, depending on their behaviour patterns. Models were then trained on each cluster, meaning animals were in the same behaviour range. This reduced absolute mean error by 8% compared to using absolute values in the model. Garcia et al. (2014) also discuss a clustering-based approach, but concluded that although some clusters were identified, within-cow variation was too high to cluster behaviour between cows. However, the dynamic clustering approach discussed by Byabazaire et al. (2019) potentially negates this affect as clustering will be based in real time, so is accounting for variations of behaviour within a cow over time. These methods display promise in increasing accuracy by classifying or mapping behaviours and investigating difference from the norm within an individual. Certainly, variation in behaviour (lying time and activity) was too variable to reliably and accurately predict lameness in this study, even in cows with moderate or severe lameness.

The Lely Qwes-H sensor data model had the worst performance of any model. Overall, rumination was not an important predictor in our study, and activity had higher contribution to the final model than rumination. This supports the conclusions of Beer et al. (2016) and van Hertem et al. (2013), who demonstrated that fluctuations in accelerometer behaviours were more successful than rumination behaviours at predicting lameness. Other studies have investigated a combination of activity, feeding behaviour and rumination with promising results (Barker et al., 2018; Beer et al., 2016; Norring et al., 2014; Thorup et al., 2015; Weigele et al., 2018), however these include other aspects of feeding behaviour such as feeding times and frequency that were not available in our dataset. In our study, rumination times were only marginally different between lame and non-lame cows, with a large standard deviation. This supports previous work that rumination time is not affected by lameness (King et al., 2017b; Thorup et al., 2015; Weigele et al., 2018).

The ICE model showed better performance for predicting lameness than the Lely model. This may be in part due to the different types of data available from each sensor: whilst both sensors measured steps, the ICEQube measured lying time which was shown in the combined model to be a more important predictor for lameness than the rumination time measured by the Lely sensor. As discussed in Section 1.5.4.3, activity is likely to not be a reliable predictor alone, and lying time has a better performance than activity at predicting lameness. However, there was also a large difference in the range of steps data from both the Lely Qwes-H system and the ICEQube system, and they did not correlate with each other (Appendix 7). Mean steps/day of the ICE data was 906.88, whereas the mean steps/day of the Lely data was 615.13. Several studies have validated the accuracy of the ICEQube sensors (Charlton et al., 2022; Shepley et al., 2017), however, to the author’s knowledge, there are no papers available validating the accuracy of the Lely Qwes-H system, although it has been used on many farms for heat detection with good results (Müschner-Siemens et al., 2020). Variation in steps data between the two systems could be due to differences between the neck-mounted and leg-mounted accelerometers. Although the accuracy of neck-mounted accelerometers has been reported (Pavlovic et al., 2021), leg-mounted accelerometers are thought to be more accurate and of more use in lameness prediction (O'Leary et al., 2020). This could be due to differences in movement between the neck and the legs, although conclusions regarding the accuracy of either pedometer cannot be drawn from this study. However, the results of this study do suggest that variables produced by the ICE sensors were more useful in predicting lameness than those from the Lely Qwes-H system.

4.3 Performance for early detection

The performance of the combined model improved when evaluated on only cows that were severely lame (MS 2B and MS 3A; **Error! Reference source not found.**). The sensitivity improves to nearly 0.91, with specificity of 0.8. Positive predictive value increased to 0.97, showing a strong ability for the model to detect severely lame cows. Similarly, De Mol et al. (2013) reported high sensitivity and specificity (0.855 and 0.888% respectively) for detecting severely lame cows (locomotion score > 5). In contrast, Kamphuis et al. (2013) found only a slight increase in AUC from 0.74 to 0.75 when mildly lame cows were removed from their model. However, direct comparison of these studies with the current study is difficult due to the different mobility scoring systems employed.

The improvement in model performance when mildly lame cows were removed can be explained by associations between predictors and mobility score (Figure 1 to Figure 13). Decrease in activity occurred with increasing lameness severity, and severely lame cows had longer lying times than both non-lame and mildly lame cows (Figure 1). Other studies have also shown decreased activity with increasing lameness score (Kamphuis et al., 2013; de Mol et al., 2013; Thorup et al., 2015; Weigele et al., 2018) with differences in behaviour more evident with severe lameness. Mild lameness could indicate early-stage lesions or less painful lesions (Griffiths et al, 2018), meaning cattle do not need to lie down for as long to alleviate pain. This means that whilst lying time could be a good predictor for more severe lameness, it may be less useful for detecting mild lameness caused by minor lesions.

Differences in behavioural data are more pronounced in our study with increasing lameness score. Decrease in activity is seen with increasing lameness severity, and severely lame cows had longer lying times than both non-lame and mildly lame cows (Figure 1, Figure 2). This is the shortcoming of existing lameness detection systems, that they are successful with predicting lameness but only in severe lame cows (Rutten et al., 2013), which can also be detected visually. Given the aim of ALD to detect mild lameness for early intervention and treatment, current model performance (including model performance in this study) is unsatisfactory at flagging mildly lame cows and further work should focus on this area.

In Section 3.5.5, predicted probabilities from each model by lameness group are displayed. As mentioned above, poor model performance in this study is driven by an inability for the model to distinguish between non-lame and mildly lame cows. Although there is a slight increase in the mean and distribution of predicted probability in MS 2A to MS 1, there is still significant overlap between these MS groups. This could be due to the inaccuracy of mobility scoring even in trained individuals, which is further discussed in Section 4.9. However, it should not be discounted that a general trend of increasing predicted probabilities across all MS groups is shown, displaying that despite overlap all models can differentiate between each mobility score to some degree. Demonstrated by the best sensitivity and positive predicted value in Table 6, combining multiple predictors improves detection of mildly lame cows. However, as discussed in Section 4.2, this performance is driven by involvement of confounding variables such as parity and weight. However, judging by the increased importance values of other production features in the combined model (Appendix 14), combining sensor and production data achieves a better separation of predicted probabilities than using activity data or production data alone.

4.4 Importance of model features

Features that contributed strongly to model prediction included average weight, lying time, step count and parity; these are discussed in more detail below.

4.4.1 Average weight

Average weight was higher in lame cows than non-lame cows, at 828.26 and 803.4kg respectively, and was a strong contributor to both the combined model and the production model. Kamphuis et al. (2013) also found a significant difference in liveweight between lame and non-lame cows and demonstrated that liveweight had the same predictive ability for lameness as activity data, with an AUC of 0.66. Schindelm et al. (2017) also described liveweight as an important factor in lameness detection. There are differences in feeding behaviour between lame and non-lame cows (Grimm et al., 2019; Thorup et al., 2015), which could result in a lower BCS and therefore lower weight in lame cows compared to non-lame cows (Randall et al., 2019). Furthermore, increased pressure on the sole is considered to be an important risk factor for claw horn lesion development, so increased weight could result in more pressure on the sole of the foot (Newsome et al., 2019), resulting in a higher chance of developing a lesion. However, in the current study there was a strong positive correlation between parity and average weight (Figure 32). The association between lameness and liveweight may therefore be confounded by parity, with lame cows more likely to be multiparous and therefore heavier.

Body condition score may be a more appropriate approach to measuring changes in weight or condition caused by lameness, as it is less likely to be confounded by parity and is more generalisable between different demographics of cow. Alternatively, longitudinal measurements of change in weight over time could be measured in accordance with changes in lameness status, although weight is still likely to fluctuate with factors independent of lameness status such as yield, DIM, gestational status and withers height (Schindelm et al., 2017). Models could also be created for different parity groups to reduce variations in weight due to factors not associated with lameness. Miguel-Pacheco et al. (2014) matched pairs of lame and non-lame cows by parity and DIM to minimise confounding and did not see any difference in weight between lame and non-lame cows. Ultimately, these findings suggest that average weight is not a reliable predictor for lameness, as too much variation occurs between parity, DIM and diet composition.

4.4.2 Lying time

Lying time was also a feature of high importance in the combined model. There was a difference in the lying time recorded from the ICEQube sensors between lame and non-lame cows of 45 minutes. This agrees with the findings of several previous studies (Alsaaod et al., 2012; Beer et al., 2016b; Blackie and Maclaurin, 2019; Chapinal et al., 2010; Ito et al., 2010; Solano et al., 2016; Thompson et al., 2019; Weigele et al., 2018). For example Sepúlveda-Varas et al. (2014) determined that lying behaviour in lame cows increased by 1.7h/d compared to non-lame. Lying times are likely to increase when a cow is lame because it is painful to stand, meaning they remain recumbent in longer lying bouts. There are also studies that report no association between lying time and lameness (Blackie and Maclaurin, 2019; Charlton et al., 2016). However, there are factors which may account for this lack of association. Charlton et al. (2016) utilised stationary lameness assessment in conjunction with Leach et al. (2009) a method that has been shown to have lower accuracy than mobility scoring. Likewise, work by Blackie and Maclaurin. (2019) may be confounded by the high stocking density recorded in the straw yards studied. Cows with less space may be forced into increased activity and lower lying times (King et al., 2017). Overall, lying time could be a reliable predictor of lameness.

4.4.3 Activity

Activity was also an important feature in the combined model in our study. It did contribute as much to the combined model as lying time, but it was still ranked as having a higher feature importance than other production behaviours. Activity differed between lame and non-lame cows in our study. Using the ICEQube sensors, the mean number of steps in lame cows was 829.2 (+-307 SD) and for non-lame cows was 946.2 (+-319.2 SD). A decrease in total steps is shown due to the pain associated with standing and walking, meaning lame cows have a lower activity. This supports the results of other work where a decrease in activity occurs with lameness (Beer et al., 2016b; Kamphuis et al., 2013; Nechanitzky et al., 2016; Thorup et al., 2015).

4.4.4 Parity

As mentioned above, mean parity of lame cows was higher than that of non-lame cows because there was lower prevalence of lameness in primiparous cows in this study compared to multiparous cows. Lameness in primiparous cows also tended to be less severe than in multiparous cows. Other studies have also demonstrated increased lameness prevalence in multiparous cattle (King et al., 2017b; Reader et al., 2011c; Thompson et al., 2019). King et al. (2017) demonstrated an increased risk of lameness with increasing parity, with those in first and second lactation having a 2.1- and 2.7-times lower risk of lameness than cows of parity 3+. Reader et al. (2011) undertook a longitudinal study of mobility scoring a single herd for one year. They describe increased chance of multiparous cows becoming lame, and these cows were less likely to recover from lameness. Beer et al. (2016) contradicts these findings, finding no increased risk of lameness due to parity. However, they describe a small sample size of only 12 non-lame cows, which may not be sufficient data to represent lameness prevalence in multiparous cows at herd level. Multiparous cows have a greater period of being at risk to have contracted an initial lameness lesion, after which lameness recurrence is much more likely (Green et al., 2014). This could be due to pedal bone proliferation, causing uneven pressure and disruption to the horn-producing corium, predisposing a cow to further lameness cases in the future (Newsome et al., 2016). Furthermore, an increase in 305-day yield is seen with increasing parity, which in itself is a risk factor for lameness (Huxley et al., 2013) and could predispose these animals to a higher chance of lameness than their younger counterparts. Garcia et al., (2014) concluded that modelling these parity groups separately could overcome the factors discussed above with differing behaviour, production, and physiology. This could be implemented with our study, although it would be more ideal to investigate with a bigger cohort size as there were only 30-35 cows in each parity group available for analysis.

Furthermore, cows in our study were more likely to be predicted lame from the combined model if they were of a higher parity, even if they were non-lame (Figure 33). The distribution of predicted probabilities of parity 3+ cows was higher than that of parity 1 and 2, even if they were mobility scored sound. This could be due to the confounding variables having a big effect on model output, as discussed in Section 4.2.

4.4.5 Milk yield

Milk yield for the cows in this study was highly variable and there was negligible difference in mean yield between lame and non-lame cows. In fact, lame cows produced 2L more per day on average. Features related to yield, for example daily yield and ISK, had relatively low importance values in the combined model over other, stronger predictors such as average weight and lying time. However, features relating to ISK did have a relatively good importance in the production model. ISK is a standardised measure of yield computed by the T4C online portal, and stands for “Individuele Standaard Koe”, which translates to “Individual standard cow”. This feature standardises yield between parity and DIM in order to compare between different demographics of cow. The fact that this had relatively high importance in the production model could indicate that it is useful in predicting lameness, although it did not have high importance values when combined with other sources of data. This could mean that ISK or a standardised yield could be a weak predictor of lameness, that is not as important to model accuracy as stronger predictors such as lying time, parity or average weight.

The differences in milk yield between lame and non-lame cows are complex, with most longitudinal studies showing a reduction in yield for an individual before a lameness event, rather than a direct comparison of daily yield amounts between lame and non-lame (Green et al., 2014, 2002; van Hertem et al., 2014; Reader et al., 2011). Yield drop in lame cows also differs depending on type of lameness-causing lesion (Hernandez et al., 2002), with lesions such as digital dermatitis causing a negligible yield drop, although still causing significant lameness (Green et al., 2014). Other papers have made direct comparisons of milk yield between lame and non-lame cows (Grimm et al., 2019; Kamphuis et al., 2013; Schindhelm et al., 2017). Kamphuis et al. (2013) showed no difference in mean values of lameness between lame and non-lame cows. Schindler et al. (2017) also showed no contribution of daily yield to the overall predictive model. Grimm et al. (2019) found weak associations between milk yield and lameness during univariate analysis, but when combined with other factors yield became a strong contributor to their model. However, it was actually higher in lame cows than non-lame cows, similar to our study. This could represent the overarching theme that when combined, features have a stronger ability to detect lameness than when used individually.

The relationship between yield and lameness is multi-faceted. The negligible difference in yield could have been confounded by differences in parity or DIM. As discussed in Section 4.4.4, a higher proportion of lame cows were multiparous than primiparous. Furthermore, the average daily yield of parity one cows was 36.21, of parity two was 43.21 and of parity three+ was 49.38. Higher yielding animals are more susceptible to lameness than their lower yielding herd mates (Green et al., 2002), and experience a yield drop to a lower plane of production than those that are non-lame (Green et al., 2014). Even if lameness caused a yield drop in our study, it would not be below those that were more non-lame but would bring them to a plane of production equal to those that are non-lame. Equally, if a multiparous cow experienced lameness, it would not bring their yield to below that of a primiparous counterpart.

Furthermore, there is high variability of yield amongst all MS groups (Figure 8)**Figure 8 Yield per day (L/d) by mobility score.** This graph shows the daily yield per cow per day by mobility score, with yield in L/d on the x axis and mobility score on the y axis. Yield was collected from the AMS at each milking and averaged over the milkings within the day. Mobility score was collected by one of two RoMS qualified operators on every Monday and Thursday during the trial period. Milk yield of MS 3A cows are the most variable, with an IQR of 33.4L/day and a varied distribution (Figure 6). The significant variability could potentially stem from the skew in DIM in the dataset. Since yield drops according to a lactation curve, after peak yield at 6-8 weeks, mean daily yield reduces with increased DIM (Deming et al., 2013). Beer et al. (2016), Grimm et al. (2019) and King et al. (2017) all showed that lame cows on average had a higher DIM than non-lame. In our study, completely non-lame cows (MS 0) have a lower DIM than those that are MS 1, 2A, 2B or 3A. The mean DIM of lame and non-lame cows did not differ in our study (160 and 161 days respectively), but the mean DIM of MS 0 is 30 days less than any of the other MS groups. This could confound results, with some non-lame cows having a higher milk yield due to stage in lactation, not caused by lameness. Although it did not appear that DIM affected our data, parity had a confounding effect on our results in terms of yield/day, and therefore this is not a reliable predictor. Suggestions to improve further work on milk yield are discussed in Section 4.6.

4.4.6 Other features

Other features with moderately high importance include protein percentage, ISK, and motion index. Since motion index is heavily correlated with steps (Appendix 10), this is expected. Variables associated with ISK have a high importance in the production model, although in the combined model they have less importance than variables such as activity and lying time from the ICEQube sensors in the combined model. The 7-day average for number of refusals has a high importance in the production model, although other features related to refusals per day have very low importance, so this is likely to be an anomaly.

Aside from these features, features relating to protein percentage have a relatively higher variable importance in both the combined and the production model. There is a small difference in the protein percentage between lame and non-lame cows (Figure 11), although the standard deviations are small and it is not as variable as fat or protein. Our work contradicts that of Antanaitis et al. (2021) and Olechnowicz and Jaskowski. (2012) who both display reduced milk protein in lame cows than non-lame, although did not use milk protein as a predictor for lameness. In our study, a higher milk protein was observed in moderately and severe lame cows. As metabolic process behind lameness is not known, we can only speculate whether a lower milk is a causal factor for lameness. Most likely is reduced feeding behaviour in lame cows (Thorup et al., 2016) causes reduced protein intake, and therefore reduced protein levels in milk (Olechnowicz and Jaskowski., 2010). This could suggest milk protein may be a predictor of lameness, albeit a weak one compared to features such as lying time.

Features with a low model importance included features relating to lying bouts, rumination and AMS visits: number of failures, number of refusals and number of milkings. Milk constituents, aside from milk protein, also had relatively low performance in the combined model. This is concurrent with other studies reported (Milk constituents: (Borghart et al., (2021), Rumination: (Beer et al., 2016), Lying bouts: (Grimm et al., 2019)), suggesting these features are not likely to be good predictors in a wider model and may not be related to lameness. The number of AMS visits has been reported to have significant differences in other literature (Miguel-Pacheco et al., 2014; de Mol et al., 2013), however due to the minimal differences between lame and non-lame cows in our dataset it is evident that they are not strong contributors to either the production or the combined model. Considering the high yielding nature of study cows (Section 2.1), and the higher yield of lame cows compared to sound cows (Figure 8) it is not surprising that even lame cows have to visit the AMS regularly in order to relieve pressure caused by a full udder (Miguel-Pacheco et al., 2014).

4.5 Practical applications of ALD

Naturally, behaviour data differs between farms due to separate management practices affecting behaviour (O'Leary et al., 2020). Several cross-farm studies have shown differences in behavioural metrics between farms (King et al., 2017b; Thorup et al., 2015; Weigele et al., 2018). Thorup et al. (2015) demonstrated a difference in mean activity variables between each of the 4 farms and concluded that farm-specific factors need to be considered when comparing behavioural data between farms. Lying times specifically differ largely, with Solano et al. (2016) demonstrating a higher lying time with cows based in a sand environment. Behaviour will also vary within a farm between seasons if management practices differ (Kamphuis et al., 2013), with cows at grazing having a different range of behavioural metrics than housed cows (Navarro et al., 2013). Cows in our study were housed all year round (Section 2.1), but the majority of UK dairy farms experience housing in the winter and out at pasture in the summer, each with a different range of activity (Thompson et al., 2019). Implications of this denote differences and potential inaccuracies of models based on cows housed indoors, and cows housed at pasture. Furthermore, different environmental factors will affect behavioural variables on farm, so model generalisability will be poor from farm-to-farm when considering threshold values to indicate lameness. These differences require consideration when assessing model generalisability based on behavioural data. A potential solution for this would be to train each model based upon the cows on that specific farm, although this may be impractical as it requires regular mobility scoring whilst data is being gathered on that farm. Generalisability could also be improved by using different systems for ALD, such as vision technology or gait measurements (O'Leary et al., 2020).

Irrespective of the type of system, many factors need to be considered when implementing ALD. Cost is a significant factor, with many systems being expensive to implement and maintain (O'Leary et al., 2020). However, improved lameness detection could result in a reduction of costs associated with lameness (Section 1.4.1), so cost-benefit analysis is important. Farmer attitudes to ALD also need to be given consideration. Even with a sensitivity and specificity of >0.9 for lameness detection, 10% of cows will be misclassified. In a herd of 300 cows, this could be up to 30 cows per day. A sensitivity of 99% is required to be practically applicable on farm (O'Leary et al., 2020), although as evidenced in this study, no technology is reliably and repeatedly reaching this figure. Lameness treatment is time consuming, labour intensive and costly, and farmers could lose faith in a system that produces too many false positives. Additionally, because farmers may underestimate a cow's mobility score (Leach et al., 2010c), they may conclude a false positive result for that cow (Kamphuis et al., 2013). Therefore, farmer education and continued training is important in conjunction with implementation of ALD.

4.6 Study limitations

Several study limitations have already been discussed in this section, for example the limited availability of feeding behaviour, the pitfalls of analysing milk yield and the contribution of average weight to the final combined model. In summary, alternative methods are more reliable for analysing milk yield data to provide further insights into how lameness affects yield. Additionally, the inclusion of a more generalisable variable such as BCS may have been more appropriate than weight. This could reduce the confounding effects of parity, DIM, stature and pregnancy on average weight, and the increased likelihood for older, heavier cows to be lame.

Another limitation of this study was the basis of classifying lameness using mobility score. Mobility scoring is heavily subjective and has poor sensitivity to detect minor lesions (Engel et al., 2003), even in trained individuals. This could explain the similarity of data distributions between MS 1 and MS 2A cows (Appendix 1, 0, Appendix 6). MS 1 cows were considered "non-lame" in the study. However, the definition of these cattle is having "imperfect" mobility, showing some disparity in their gait compared to perfectly non-lame cows. There is possibility they were suffering from or masking minor lesions, due to their stoic nature when under observation. This could cause behaviour patterns close to those of MS 2A. A better method of studying lameness would have been regular inspection of feet and recording of any lesions present. Therefore a "lame" cow would be "lesion-positive", improving the poor sensitivity and subjectivity of mobility scoring. This could also allow analysis of behaviour for specific lesions.

Whole-lactation observation would be more useful in investigating the longitudinal effects of lameness on variables such as milk yield and activity. This study gave some valuable insights into how lameness affects several different variables; however, a longitudinal approach would mean analysis could be undertaken on individual cow attributes over a whole lactation. It could give more cow-specific observations as to how things like activity and lying time are affected by DIM status. It would allow more information to be gathered per cow and per lactation, as a longer study would mean a higher number of lameness events were achieved. This could invoke within-cow analysis, so a lame period and a non-lame period can be gathered from an individual and she acts as her own control. This could not be attempted in this study due to the short time frame of data collection, so a limited number of lame periods were available.

This study also was conducted on a single farm where there were numerous data sources available. This is not the case for much of the UK dairy herd, with the majority of systems opting to graze cows in the summer and milk on a parlour-based unit. This study is not generalisable to these herds due to the sheer amount of data collected, and so the study design would be hard to replicate.

Due to the confounding variables discussed in Section 4.4, the reliability of this model requires further analysis. The model also places high importance on parity and weight. These variables are not useful for comparison between lame and non-lame cows, as there are fundamental differences between these variables not just explained by lameness. With the high importance values based on features that are related more to predicting parity than lameness, the model's accuracy must be called into question. The relatively high sensitivity and specificity of this model are not reliable and removing average weight as a predictor would likely resulting in a reduction of sensitivity, specificity, and AUC.

5 Conclusion

Data from a combination of three sensor technologies (AMS, neck-worn and leg-worn accelerometers) were used to predict lameness with an improved model sensitivity, specificity and AUC compared to the use of each sensor technology independently. Therefore, a combination of weak predictors from features obtained from each sensor were able to predict lameness better than that achieved by any feature individually. Features obtained from the leg-worn accelerometers and data obtained from the AMS appear to be the best combination for predicting lameness. Variables such as weight and parity were found to make a substantial contribution to model accuracy, but care should be taken when including these variables in future analysis since the effect may not be generalisable.

Model sensitivity and specificity were higher when predicting only severely lame cows. However, traditional mobility scoring remains both cheaper and more sensitive than implementing these technologies for early lameness detection.

Practically speaking, behavioural data is too variable within an individual cow and between cows to reliably predict lameness. No model created in this study had sufficient sensitivity to be of practical value, especially considering the expense of the systems used to collect the data. Furthermore, very few farms have the technology used to collect the data used in this study available, making this poorly generalisable to other farms. Generalisability and detection of mild lameness should be a priority when researching future systems, which indicates using another system, such as a vision-based automatic lameness detection system, may have more practical merit than those using behavioural data.

6 Appendices

Appendix 1 A table to define all of the features included in the ICE model

Table 8 All of the features included in the IceQube model. This table details all of the features included in the IceQube model, and a definition for each, and how each was calculated.

Feature	Description
Lying time	Total lying time per day (hours/day)
Steps	Total number of steps (Steps/day)
Motion index	Total motion index per day (Motion index/day)
Lying bouts	Total number of lying bouts per day (bouts/day)
Lying time diff	Difference of an individual's total daily lying time from the mean lying time of all the cows in the group for that day
Steps diff	Difference of an individual's total daily steps from the mean steps of all the cows in the group for that day
Motion index diff	Difference of an individual's total daily motion index from the mean motion index of all the cows in the group for that day
Lying bouts diff	Difference of an individual's total daily lying bouts from the mean lying bouts of all the cows in the group for that day
Lying.Time_7da	Average lying time per day over the previous 7 days
Lying.Time_7sd	Standard deviation for the lying time per day over the previous 7 days
Lying.Timelag1	Total daily lying time for 1 day before
Lying.Timelag2	Total daily lying time for 2 days before
Lying.Timelag3	Total daily lying time for 3 days before
Lying.Timelag4	Total daily lying time for 4 days before
Lying.Timelag5	Total daily lying time for 5 days before

Lying.Timelag6	Total daily lying time for 6 days before
Steps_7da	Average lying time per day over the previous 7 days
Steps_7sd	Standard deviation for the lying time per day over the previous 7 days
Stepslag1	Total daily Steps for 1 day before
Stepslag2	Total daily Steps for 2 days before
Stepslag3	Total daily Steps for 3 days before
Stepslag4	Total daily Steps for 4 days before
Stepslag5	Total daily Steps for 5 days before
Stepslag6	Total daily Steps for 6 days before
Motion.Index_7da	Average motion index per day over the previous 7 days
Motion.Index_7sd	Standard deviation for the motion index per day over the previous 7 days
Motion.Indexlag1	Total daily motion index for 1 day before
Motion.Indexlag2	Total daily motion index for 2 days before
Motion.Indexlag3	Total daily motion index for 3 days before
Motion.Indexlag4	Total daily motion index for 4 days before
Motion.Indexlag5	Total daily motion index for 5 days before

Motion.Indexlag6	Total daily motion index for 6 days before
Lying.Bouts_7da	Average lying bouts per day over the previous 7 days
Lying.Bouts_7sd	Standard deviation for the lying bouts per day over the previous 7 days
Lying.Boutslag1	Total daily lying bouts for 1 day before
Lying.Boutslag2	Total daily lying bouts for 2 days before
Lying.Boutslag3	Total daily lying bouts for 3 days before
Lying.Boutslag4	Total daily lying bouts for 4 days before
Lying.Boutslag5	Total daily lying bouts for 5 days before
Lying.Boutslag6	Total daily lying bouts for 6 days before

Appendix 2 A table to define all of the features included in the Lely model

Table 9 All of the features included in the Lely model. This table details all of the features included in the Lely model, and a definition for each, and how each was calculated.

Feature	Description
Steps	Total number of steps (Steps/day)
Rumination minutes	Total rumination minutes per day (Rumination minutes/day)
Heat alerts	Total number of heat alerts per day (bouts/day)
Steps diff	Difference of an individual's total daily steps from the mean steps of all the cows in the group for that day
Rumination minutes diff	Difference of an individual's total daily rumination minutes from the mean rumination minutes of all the cows in the group for that day
Heat alerts diff	Difference of an individual's total daily heat alerts from the mean heat alerts of all the cows in the group for that day
Activity.Counter_7da	Average steps per day over the previous 7 days
Activity.Counter_7sd	Standard deviation for the steps per day over the previous 7 days
Activity.Counterlag1	Total daily steps for 1 day before
Activity.Counterlag2	Total daily steps for 2 days before
Activity.Counterlag3	Total daily steps for 3 days before
Activity.Counterlag4	Total daily steps for 4 days before
Activity.Counterlag5	Total daily steps for 5 days before
Activity.Counterlag6	Total daily steps for 6 days before
RuminationMinutes_7da	Average rumination minutes per day over the previous 7 days

RuminationMinutes_7sd	Standard deviation for the rumination minutes per day over the previous 7 days
RuminationMinuteslag1	Total daily rumination minutes for 1 day before
RuminationMinuteslag2	Total daily rumination minutes for 2 days before
RuminationMinuteslag3	Total daily rumination minutes for 3 days before
RuminationMinuteslag4	Total daily rumination minutes for 4 days before
RuminationMinuteslag5	Total daily rumination minutes for 5 days before
RuminationMinuteslag6	Total daily rumination minutes for 6 days before
HeatAlert_7da	Average heat alerts per day over the previous 7 days
HeatAlert_7sd	Standard deviation for the heat alerts per day over the previous 7 days
HeatAlertlag1	Total daily heat alerts for 1 day before
HeatAlertlag2	Total daily heat alerts for 2 days before
HeatAlertlag3	Total daily heat alerts for 3 days before
HeatAlertlag4	Total daily heat alerts for 4 days before
HeatAlertlag5	Total daily heat alerts for 5 days before
HeatAlertlag6	Total daily heat alerts for 6 days before

Appendix 3 A table to define all of the features included in the Production model

Table 10 All of the features included in the Production model. This table details all of the features included in the production model, and a definition for each, and how each was calculated.

Feature	Description
AverageWeight	Average weight per day (kg)
AverageWeight_7da	Average weight over the previous 7 days (kg)
AverageWeight_7sd	Standard deviation for weight over the previous 7 days (kg)
AverageWeightlag1	Average weight for the day before (kg)
AverageWeightlag2	Average weight for 2 days before (kg)
AverageWeightlag3	Average weight for 3 days before (kg)
AverageWeightlag4	Average weight for 4 days before (kg)
AverageWeightlag5	Average weight for 5 days before (kg)
AverageWeightlag6	Average weight for 5 days before (kg)
DIM	Days in milk
FatPercentage	Fat percentage per day
FatPercentage_7da	Fat percentage over the previous 7 days
FatPercentage_7sd	Standard deviation for weight over the previous 7 days
FatPercentagelag1	Fat percentage for the day before
FatPercentagelag2	Fat percentage for 2 days before
FatPercentagelag3	Fat percentage for 3 days before
FatPercentagelag4	Fat percentage for 4 days before
FatPercentagelag5	Fat percentage for 5 days before
FatPercentagelag6	Fat percentage for 5 days before
ISK	Standardised yield, adjusted for DIM and parity, total per day (L)
ISK_7da	Average ISK over the past 7 days (L)
ISK_7sd	Standard deviation for ISK over the past 7 days (L)
ISKlag1	ISK for the day before
ISKlag2	ISK for 2 days before
ISKlag3	ISK for 3 days before
ISKlag4	ISK for 4 days before
ISKlag5	ISK for 5 days before
ISKlag6	ISK for 6 days before
LactosePercentage	Lactose percentage per day

LactosePercentage_7da	Lactose percentage over the previous 7 days
LactosePercentage_7sd	Standard deviation for lactose percentage over the previous 7 days
LactosePercentagelag1	Lactose percentage for the day before
LactosePercentagelag2	Lactose percentage for 2 days before
LactosePercentagelag3	Lactose percentage for 3 days before
LactosePercentagelag4	Lactose percentage for 4 days before
LactosePercentagelag5	Lactose percentage for 5 days before
LactosePercentagelag6	Lactose percentage for 6 days before
MdpFailures	Number of failures in the robot per day
MdpFailures_7da	Average number of failures per day over the past 7 days
MdpFailures_7ds	Standard deviation for the number of failures per day over the past 7 days
MdpFailureslag1	Failures for the day before
MdpFailureslag2	Failures for 2 days before
MdpFailureslag3	Failures for 3 days before
MdpFailureslag4	Failures for 4 days before
MdpFailureslag5	Failures for 5 days before
MdpFailureslag6	Failures for 6 days before
MdpMilkings	Number of Milkings in the robot per day
MdpMilkings_7da	Average number of Milkings per day over the past 7 days
MdpMilkings_7sd	Standard deviation for the number of Milkings per day over the past 7 days
MdpMilkingslag1	Milkings for the day before
MdpMilkingslag2	Milkings for 2 days before
MdpMilkingslag3	Milkings for 3 days before
MdpMilkingslag4	Milkings for 4 days before
MdpMilkingslag5	Milkings for 5 days before
MdpMilkingslag6	Milkings for 6 days before
MdpRefusals	Number of Refusals in the robot per day
MdpRefusals_7da	Average number of Refusals per day over the past 7 days

MdpRefusals_7sd	Standard deviation for the number of Refusals per day over the past 7 days
MdpRefusalslag1	Refusals for the day before
MdpRefusalslag2	Refusals for 2 days before
MdpRefusalslag3	Refusals for 3 days before
MdpRefusalslag4	Refusals for 4 days before
MdpRefusalslag5	Refusals for 5 days before
MdpRefusalslag6	Refusals for 6 days before
MilkDayProduction	Total daily yield (L)
MilkDayProduction_7da	Average daily yield over the previous 7 days (L)
MilkDayProduction_7sd	Standard deviation for daily yield over the previous 7 days (L)
MilkDayProductionlag1	Total daily yield for the day before
MilkDayProductionlag2	Total daily yield for 2 days before
MilkDayProductionlag3	Total daily yield for 3 days before
MilkDayProductionlag4	Total daily yield for 4 days before
MilkDayProductionlag5	Total daily yield for 5 days before
MilkDayProductionlag6	Total daily yield for 6 days before
Parity	Current lactation number
ProteinPercentage	Protein percentage per day
ProteinPercentage_7da	Protein percentage over the previous 7 days
ProteinPercentage_7s	Standard deviation forProtein percentage over the previous 7 days
ProteinPercentagelag1	Protein percentage for the day before
ProteinPercentagelag2	Protein percentage for 2 days before
ProteinPercentagelag3	Protein percentage for 3 days before
ProteinPercentagelag4	Protein percentage for 4 days before
ProteinPercentagelag5	Protein percentage for 5 days before
ProteinPercentagelag6	Protein percentage for 6 days before

Appendix 4 Table to show the spread of ICEQube variables by mobility score

Table 11 Descriptive statistics for variables from the ICEQube sensor by mobility score. This table gives the mean, standard deviation, median, minimum, maximum and range for each daily value by mobility score. This data was collected from ICEQube sensor and averaged into daily values. Non-lame cows are MS 0 and MS 1 and lame cows are MS 2A, MS 2B and MS 3A.

Variable name	MS Group	Mean	SD	Median	Minimum	Maximum	Range
ICE total steps	MS 0	1056.43	266.82	1020.00	497.00	2596.00	2099.00
	MS 1	932.58	325.32	910.00	0.00	5658.00	5658.00
	MS 2A	841.68	332.37	815.00	0.00	4390.00	4390.00
	MS 2B	811.75	256.17	791.50	282.00	2123.00	1841.00
	MS 3A	872.70	278.02	900.00	455.00	1547.00	1092.00
Ice total steps difference from mean	MS 0	150.98	259.78	110.81	-389.94	1707.17	2097.11
	MS 1	19.42	320.66	-4.01	-1010.69	4637.14	5647.83
	MS 2A	-80.21	328.04	-104.65	-1020.86	3379.31	4400.17
	MS 2B	-106.07	252.00	-130.58	-593.69	1108.10	1701.79
	MS 3A	-37.89	268.72	-29.18	-555.69	559.81	1115.50
Lying bouts	MS 0	12.09	3.20	12.00	6.00	21.00	15.00
	MS 1	11.45	3.01	11.00	0.00	24.00	24.00
	MS 2A	11.43	2.95	11.00	0.00	29.00	29.00
	MS 2B	11.69	2.86	11.00	5.00	21.00	16.00
	MS 3A	11.79	3.89	12.00	4.00	20.00	16.00
Lying bouts difference from mean	MS 0	0.34	3.02	0.12	-5.37	8.97	14.34
	MS 1	-0.15	2.91	-0.26	-11.22	11.63	22.85
	MS 2A	-0.12	2.88	-0.26	-11.37	16.62	27.99

	MS 2B	0.09	2.74	-0.37	-6.83	9.17	16.00
	MS 3A	0.15	3.93	-0.26	-7.88	7.63	15.51
Lying time	MS 0	12.89	1.64	12.85	7.67	16.20	8.53
	MS 1	13.29	1.96	13.34	2.80	24.00	21.20
	MS 2A	13.75	2.20	13.76	5.16	24.00	18.84
	MS 2B	14.54	2.37	14.76	7.35	20.42	13.07
	MS 3A	15.17	2.20	14.65	11.36	20.38	9.02
Lying time difference from mean	MS 0	-0.62	1.50	-0.69	-4.86	2.43	7.29
	MS 1	-0.27	1.88	-0.26	-10.51	11.64	22.15
	MS 2A	0.25	2.06	0.29	-8.14	10.48	18.62
	MS 2B	1.03	2.25	1.17	-6.09	6.32	12.41
	MS 3A	1.63	2.10	1.50	-2.25	6.63	8.88
Motion index	MS 0	4190.66	1128.66	4042.00	2174.00	9784.00	7610.00
	MS 1	3822.33	1325.98	3731.00	0.00	20902.00	20902.00
	MS 2A	3537.94	1415.36	3380.00	0.00	18850.00	18850.00
	MS 2B	3588.93	1229.03	3434.50	1409.00	10573.00	9164.00
	MS 3A	4027.97	1316.20	4161.00	1937.00	7061.00	5124.00
Motion index difference from the mean	MS 0	439.34	1098.96	339.46	-1418.60	6099.57	7518.17
	MS 1	27.58	1301.57	-109.63	-4201.80	16613.74	20815.54
	MS 2A	-298.68	1399.88	-404.64	-4288.26	14641.20	18929.46

	MS 2B	-229.72	1215.23	-342.72	-2508.26	6331.97	8840.23
	MS 3A	244.14	1318.29	234.25	-2271.80	2991.20	5263.00

Appendix 5 Table to show the spread of Lely Qwes-H values by mobility score

Table 12 Descriptive statistics for variables from the Lely Qwes-H sensor by mobility score. This table gives the mean, standard deviation, median, minimum, maximum and range for each daily value by mobility score. This data was collected from Lely Qwes-H sensor and averaged into daily values. Non-lame cows are MS 0 and MS 1 and lame cows are MS 2A, MS 2B and MS 3A.

Variable name	MS Group	Mean	SD	Median	Minimum	Maximum	Range
Heat alerts	MS 0	0.62	0.49	1.00	0.00	1.00	1.00
	MS 1	0.41	0.49	0.00	0.00	1.00	1.00
	MS 2A	0.32	0.47	0.00	0.00	1.00	1.00
	MS 2B	0.24	0.43	0.00	0.00	1.00	1.00
	MS 3A	0.03	0.17	0.00	0.00	1.00	1.00
Lely total steps	MS 0	649.49	89.41	646.00	447.50	848.00	400.50
	MS 1	608.59	152.03	596.00	200.50	2307.00	2106.50
	MS 2A	622.18	300.92	568.50	216.50	3036.00	2819.50
	MS 2B	638.29	501.16	521.00	284.00	3036.00	2752.00
	MS 3A	509.68	66.62	509.00	283.00	631.50	348.50
Lely total steps difference from the mean	MS 0	32.01	82.12	20.84	-133.16	226.94	360.10
	MS 1	-3.74	145.07	-18.81	-359.39	1703.19	2062.58
	MS 2A	5.09	298.40	-49.28	-317.78	2425.94	2743.72
	MS 2B	20.88	495.11	-86.49	-342.03	2431.81	2773.84
	MS 3A	-97.49	60.45	-93.83	-207.39	-5.89	201.50
Rumination	MS 0	483.48	46.05	489.00	334.00	580.00	246.00
	MS 1	473.09	51.26	477.00	252.00	609.00	357.00

	MS 2A	469.42	47.64	475.00	294.00	583.00	289.00
	MS 2B	458.63	50.49	461.50	322.50	593.00	270.50
	MS 3A	469.91	45.02	478.00	350.00	549.00	199.00
Rumination difference from the mean	MS 0	16.90	43.45	16.93	-87.07	108.43	195.50
	MS 1	3.44	46.67	6.77	-221.00	127.75	348.74
	MS 2A	-0.71	44.71	-0.51	-187.25	109.31	296.56
	MS 2B	-10.16	46.65	-7.19	-166.82	112.85	279.67
	MS 3A	0.57	36.69	3.74	-83.25	56.68	139.93

Appendix 6 Table to show the spread of production data by mobility score

Table 13 Descriptive statistics for variables from the AMS (Lely Astronaut A3) sensor by mobility score. This table gives the mean, standard deviation, median, minimum, maximum and range for each daily value by mobility score. This data was collected from automatic milking system (AMS) sensor and averaged into daily values. Non-lame cows are MS 0 and MS 1 and lame cows are MS 2A, MS 2B and MS 3A.

Variable name	MS Group	Mean	SD	Median	Minimum	Maximum	Range
Average Weight	MS 0	652.32	61.91	633.00	543.00	786.00	243.00
	MS 1	723.98	87.10	721.00	528.00	994.00	466.00
	MS 2A	747.68	89.82	746.00	542.00	964.00	422.00
	MS 2B	789.00	108.86	781.50	576.00	1014.00	438.00
	MS 3A	840.82	102.84	883.00	616.00	1060.00	444.00
Daily yield	MS 0	39.92	6.98	40.60	19.10	53.90	34.80
	MS 1	42.32	10.83	41.40	10.80	70.10	59.30
	MS 2A	43.28	10.62	42.80	17.90	72.20	54.30
	MS 2B	45.61	11.85	45.80	14.60	71.40	56.80
	MS 3A	45.67	17.38	47.00	18.20	68.10	49.90
Daily yield average	MS 0	39.77	6.89	40.70	21.10	52.90	31.80
	MS 1	42.56	10.25	41.50	14.80	69.90	55.10
	MS 2A	43.62	10.17	42.40	19.70	67.30	47.60
	MS 2B	46.18	11.01	46.85	21.10	66.30	45.20
	MS 3A	45.81	16.68	49.60	19.90	65.80	45.90
Failures	MS 0	0.08	0.33	0.00	0.00	2.00	2.00
	MS 1	0.06	0.27	0.00	0.00	3.00	3.00
	MS 2A	0.04	0.23	0.00	0.00	2.00	2.00
	MS 2B	0.04	0.22	0.00	0.00	2.00	2.00

	MS 3A	0.06	0.24	0.00	0.00	1.00	1.00
Fat percentage	MS 0	3.92	0.62	3.85	2.40	5.57	3.17
	MS 1	3.92	0.87	3.84	2.04	7.04	5.00
	MS 2A	3.90	0.87	3.79	1.86	6.60	4.74
	MS 2B	3.95	0.99	3.82	2.33	6.78	4.45
	MS 3A	4.14	1.35	3.65	2.29	6.95	4.66
ISK	MS 0	63.55	11.46	61.90	33.70	92.10	58.40
	MS 1	62.91	11.48	63.10	16.50	98.80	82.30
	MS 2A	62.42	11.50	63.20	31.60	89.60	58.00
	MS 2B	62.21	11.92	64.45	24.90	87.80	62.90
	MS 3A	59.48	17.34	70.50	26.30	79.80	53.50
Lactation number	MS 0	1.28	0.47	1.00	1.00	3.00	2.00
	MS 1	2.09	1.17	2.00	1.00	5.00	4.00
	MS 2A	2.50	1.32	2.00	1.00	5.00	4.00
	MS 2B	3.06	1.21	3.00	1.00	5.00	4.00
	MS 3A	3.42	1.30	3.00	1.00	5.00	4.00
Lactose percentage	MS 0	5.07	0.08	5.09	4.84	5.24	0.40
	MS 1	5.01	0.12	5.03	4.39	5.31	0.92
	MS 2A	4.99	0.12	4.99	4.68	5.24	0.56
	MS 2B	4.98	0.12	5.00	4.70	5.26	0.56
	MS 3A	5.02	0.09	5.02	4.80	5.20	0.40
Milkings	MS 0	3.41	0.79	3.00	1.00	5.00	4.00
	MS 1	3.24	0.93	3.00	1.00	9.00	8.00
	MS 2A	3.16	0.95	3.00	1.00	8.00	7.00

	MS 2B	2.99	1.03	3.00	1.00	5.00	4.00
	MS 3A	2.70	1.13	3.00	1.00	5.00	4.00
Protein percentage	MS 0	3.31	0.19	3.25	2.95	3.72	0.77
	MS 1	3.33	0.25	3.31	2.57	4.06	1.49
	MS 2A	3.31	0.23	3.32	2.72	3.97	1.25
	MS 2B	3.38	0.29	3.36	2.82	4.17	1.35
	MS 3A	3.45	0.35	3.37	2.89	4.12	1.23
Refusals	MS 0	3.81	6.50	2.00	0.00	57.00	57.00
	MS 1	2.65	4.16	1.00	0.00	39.00	39.00
	MS 2A	2.16	4.44	1.00	0.00	53.00	53.00
	MS 2B	1.46	2.40	0.00	0.00	14.00	14.00
	MS 3A	0.97	2.49	0.00	0.00	14.00	14.00

Appendix 7 Graphs to show the distribution of the ICEQube variables

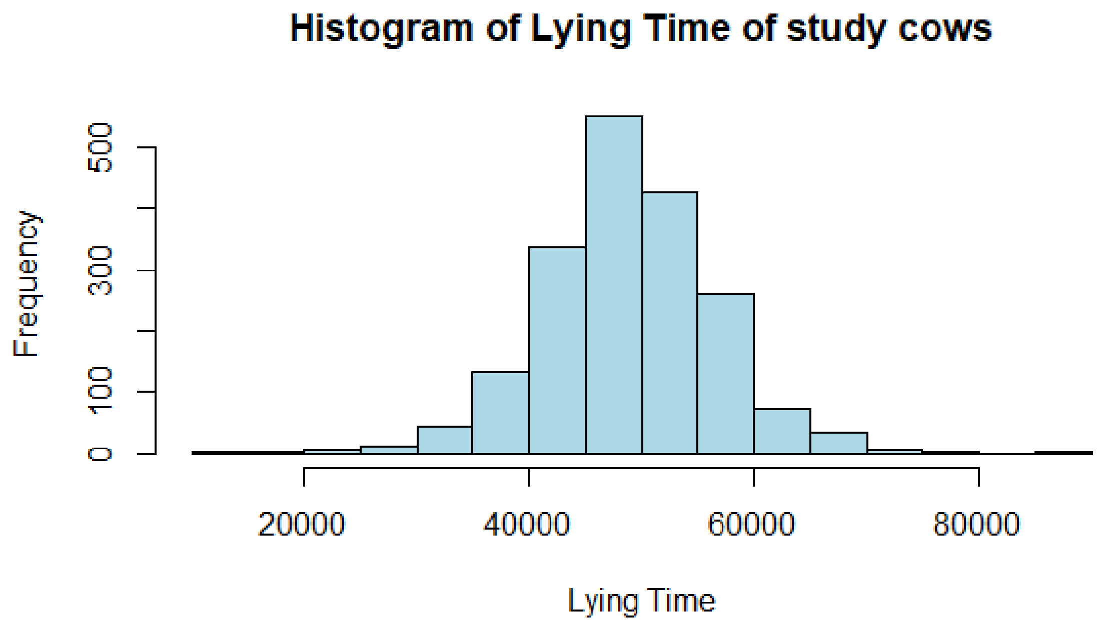

Figure 18 A histogram to show the distribution of total lying time in seconds collected from the ICEQube sensors. Lying time was collected from the ICEQube sensors and computed into total lying time per day. Lying time is kept in seconds for this graph as it gives a better spread of lying time than if plotted in hours. Lying time is plotted on the x axis and frequency on the y axis. This graph shows an even distribution of the variable and so no natural log was applied for analysis.

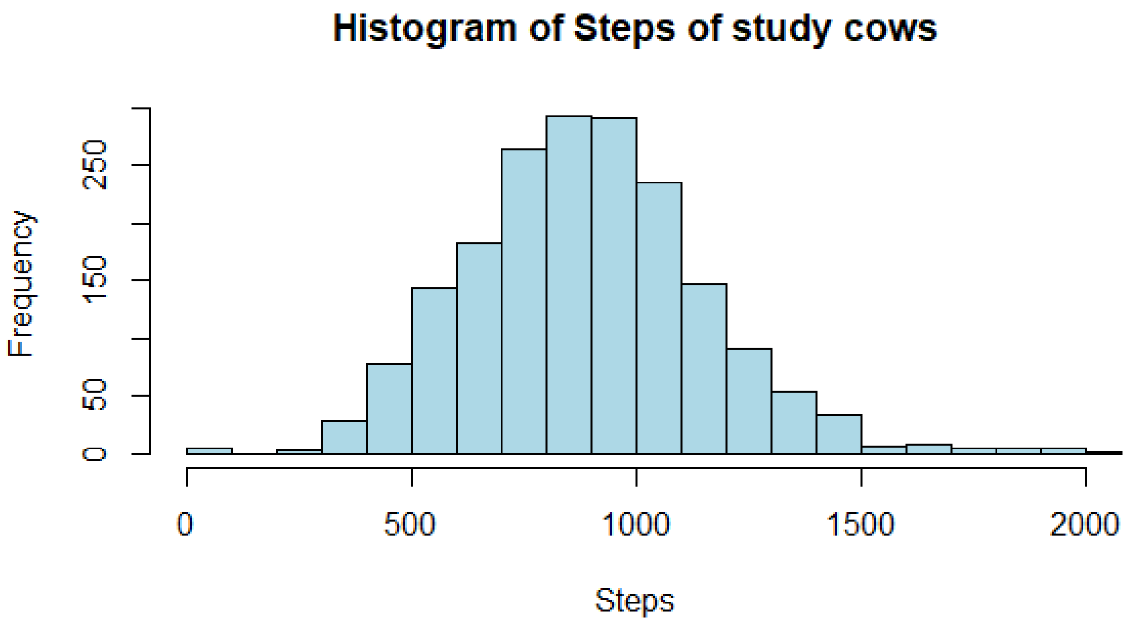

Figure 19: A histogram to show the distribution of total steps collected from the ICEQube sensors. Number of steps was collected from the ICEQube sensors and computed into total number of steps per day. Number of steps is plotted on the x axis and frequency on the y axis. This graph shows an even distribution of the variable and so no natural log was applied for analysis.

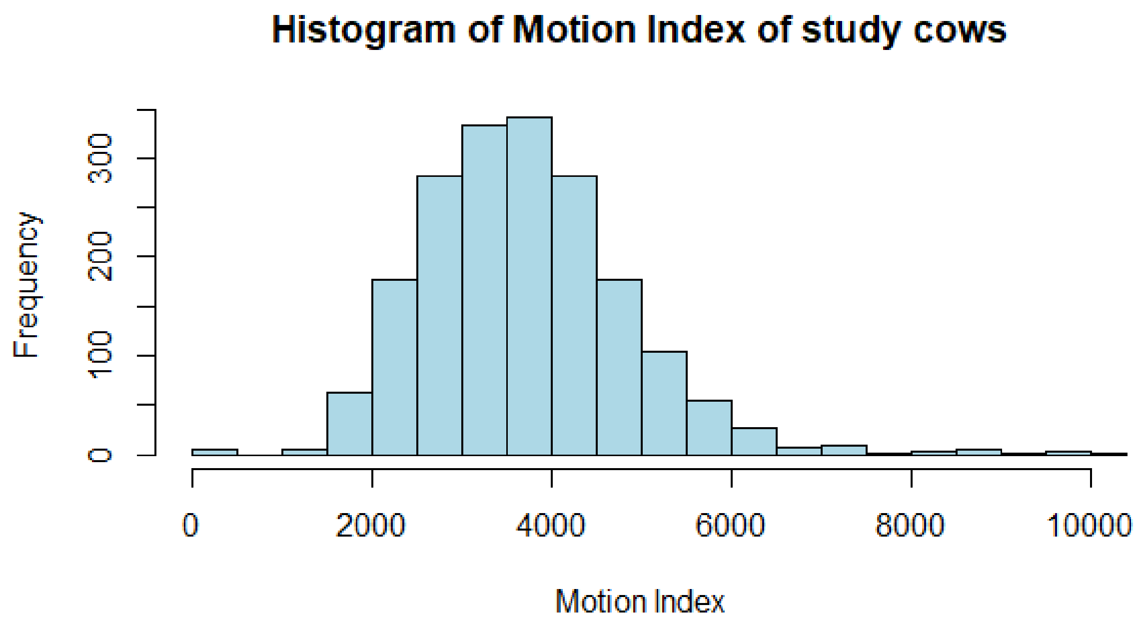

Figure 20 A histogram to show the distribution of motion index over the study cows. Motion index was collected from the ICEQube sensors and given as an absolute value that equates to energy expended per cow per day. Since it is heavily correlated with steps, we focussed our analysis on total step count over motion index, since it seemed a more generalisable feature. It was analysed as part of the ICEQube and combined models. Motion index is plotted on the x axis, and frequency on the right axis. This feature follows an even, bell-shaped curve and so no natural logarithm was applied to this feature.

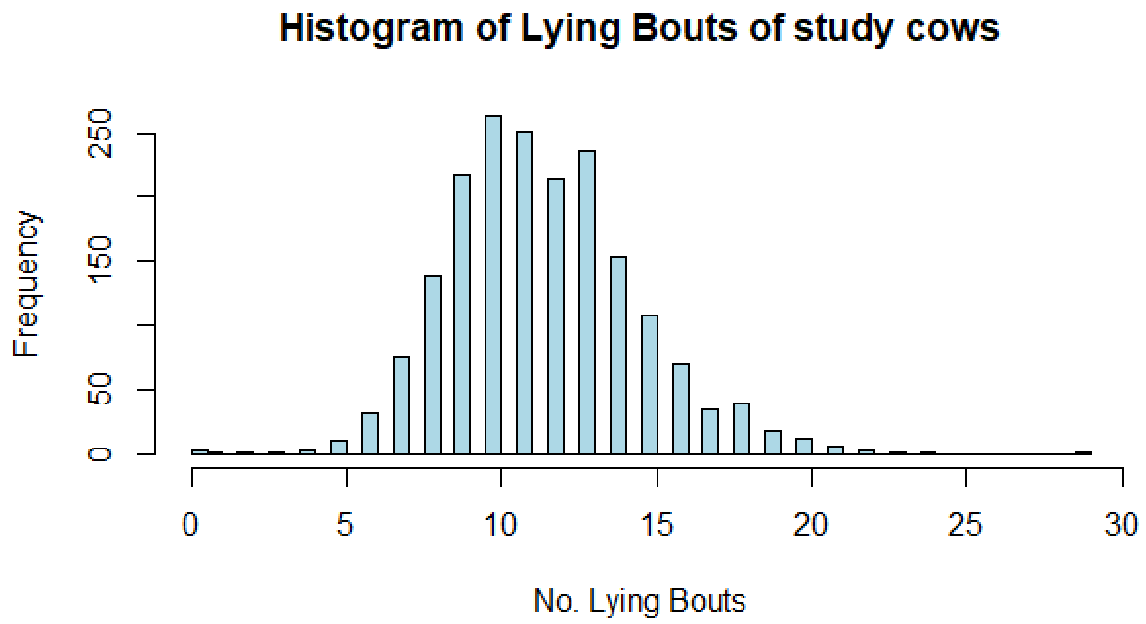

Figure 21 A histogram to show the distribution of the number of lying bouts per cow per day. Number of lying bouts was collected from the ICEQube sensors in the form of transitions up (lying to standing) and down (standing to lying) per day, and computed into a daily total of lying bouts. It was analysed as part of the ICEQube model. Lying bouts is plotted on the x axis, and frequency on the right axis. This feature follows an even, bell-shaped curve and so no natural logarithm was applied to this feature.

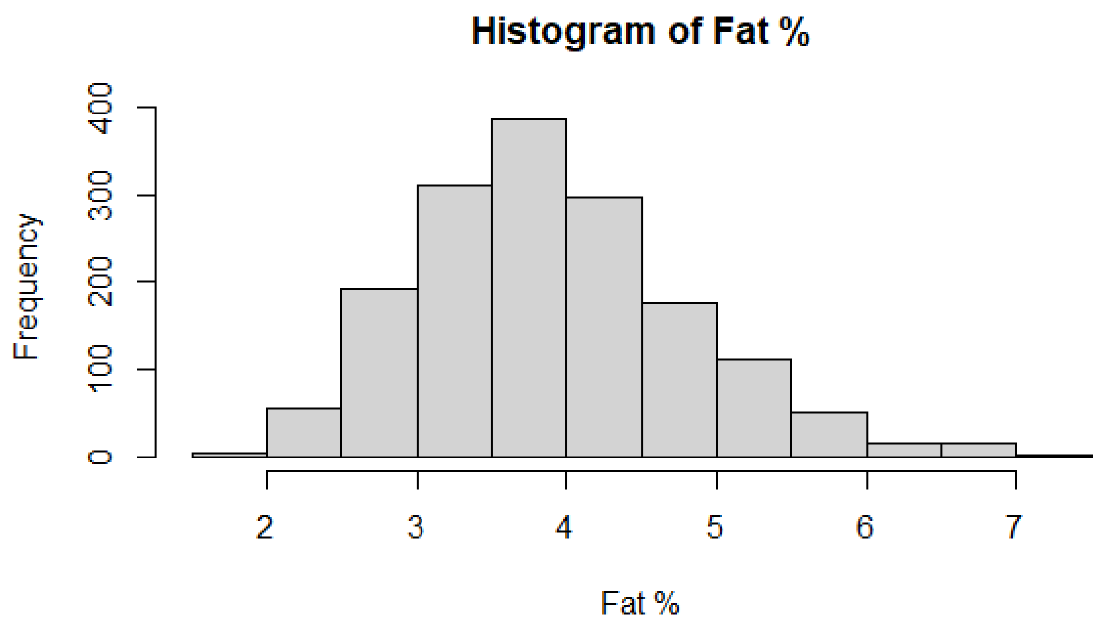

Figure 22 This graph shows a histogram of fat percentage in study cows. Fat percentage was collected from the Lely Astronaut A3 AMS system and averaged over all of the milkings in a day to give a daily value. Fat percentage is included across the x axis in bins of size 0.5, and frequency is plotted across the y axis. This demonstrates an evenly distributed, bell shaped curve and so no natural logarithm was applied to this feature.

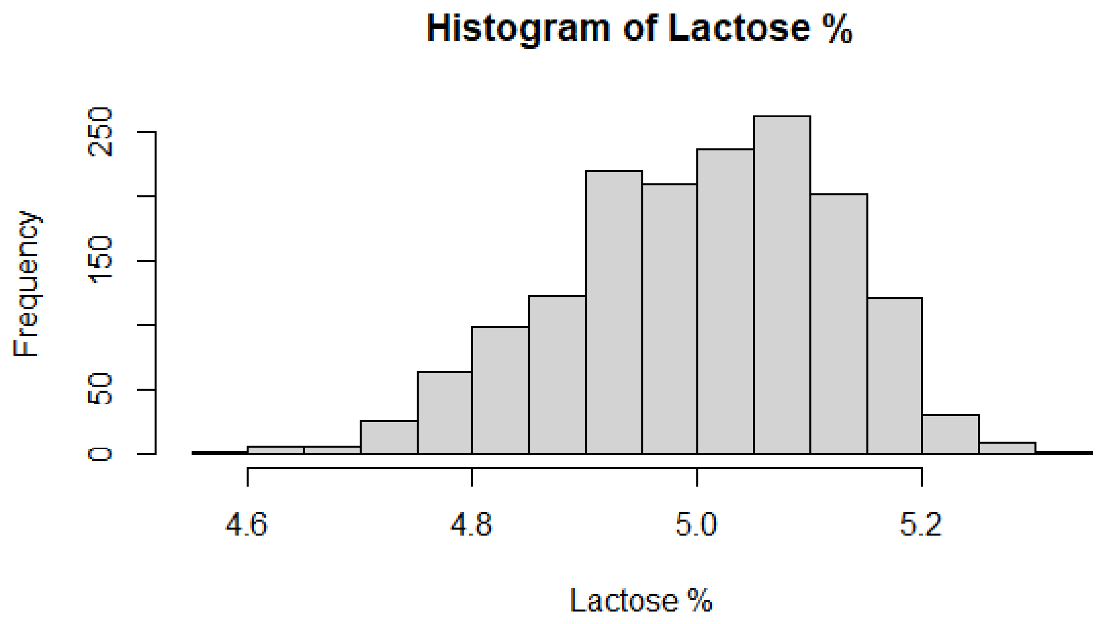

Figure 23 This graph shows a histogram of lactose percentage in study cows. Lactose percentage was collected from the Lely Astronaut A3 AMS system and averaged over all of the milkings in a day to give a daily value. Lactose percentage is included across the x axis in bins of size 0.05, and frequency is plotted across the y axis. This demonstrates an evenly distributed, bell shaped curve and so no natural logarithm was applied to this feature.

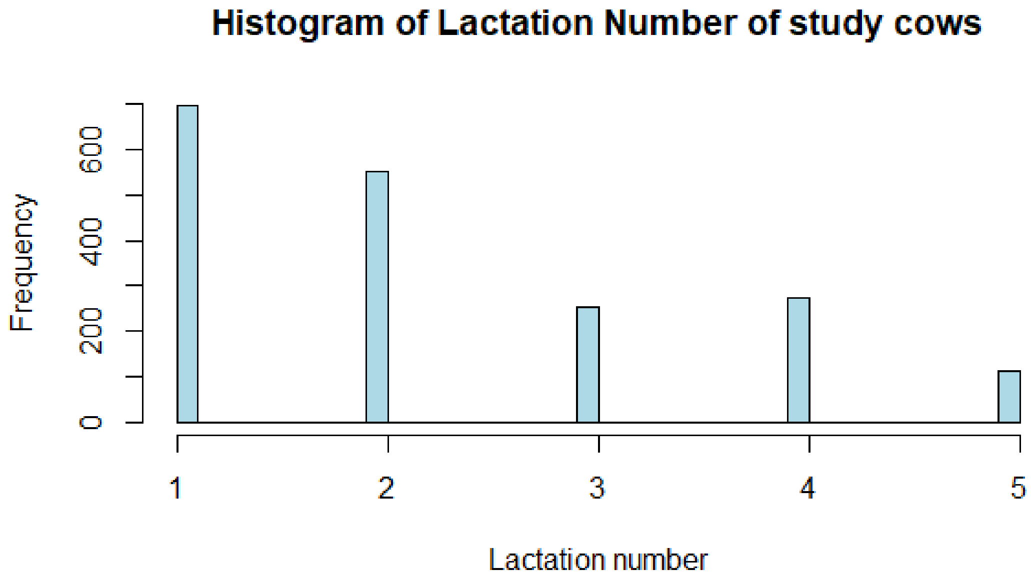

Figure 24 A histogram to show the distribution of parity in my study sample dataset. Parity was obtained from the Uniform Agri farm management software and plotted on the x axis with frequency on the y axis. There was a higher number of parity one cows than any other parity, so we combined parity 3, 4 and 5 together into one category during analysis. Multiparous cows have similar behaviour and so we concluded this to be an appropriate solution to the smaller number of each parity > 2. Number of cows per lactation is detailed in graph (x). There were 697 observations for cows in PARITY 1, 553 observations for cows in PARITY 2 and 640 observations for cows in PARITY 3+.

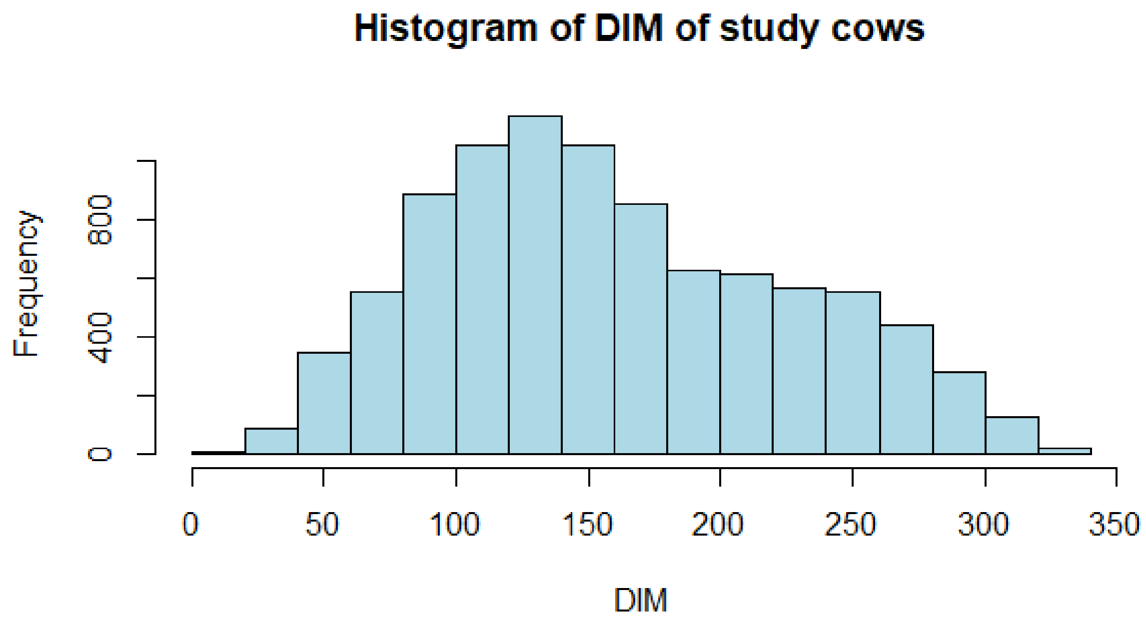

Figure 25 A histogram to show the distribution of DIM in the study dataset. Calving date was collected from Uniform Agri and the days between calving date and current date was calculated for all of the days included in the model. DIM is plotted on the x axis and frequency on the y axis. Mean DIM was high, at 160 for both lame and non-lame cows. Although this graph follows an even trend for DIM < mean, there is a skewed distribution for a higher DIM and so a natural logarithm was applied to this feature.

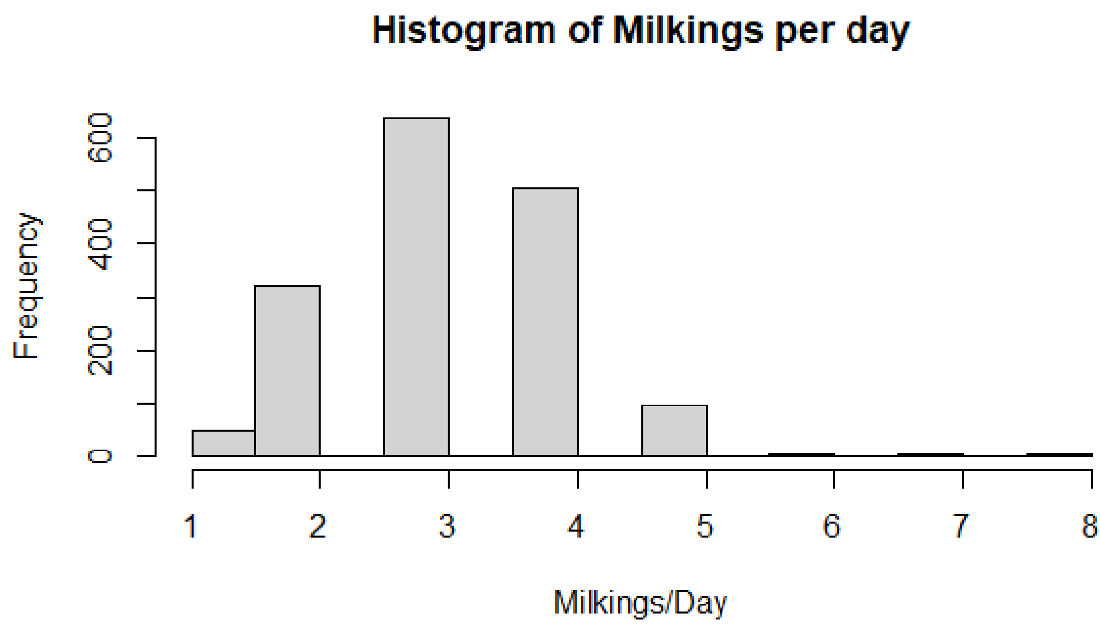

Figure 26 This graph shows a histogram of milkings per day. Milkings per day were collected from the milk visit data from the Lely Astronaut A3 milking system. Milkings per day is included across the x axis in bins of size 1, and frequency is plotted across the y axis. The mean of milkings/day is around 3, with several outliers at 6-8 milkings per day. For the most part, this feature is evenly distributed and so no natural logarithm was applied.

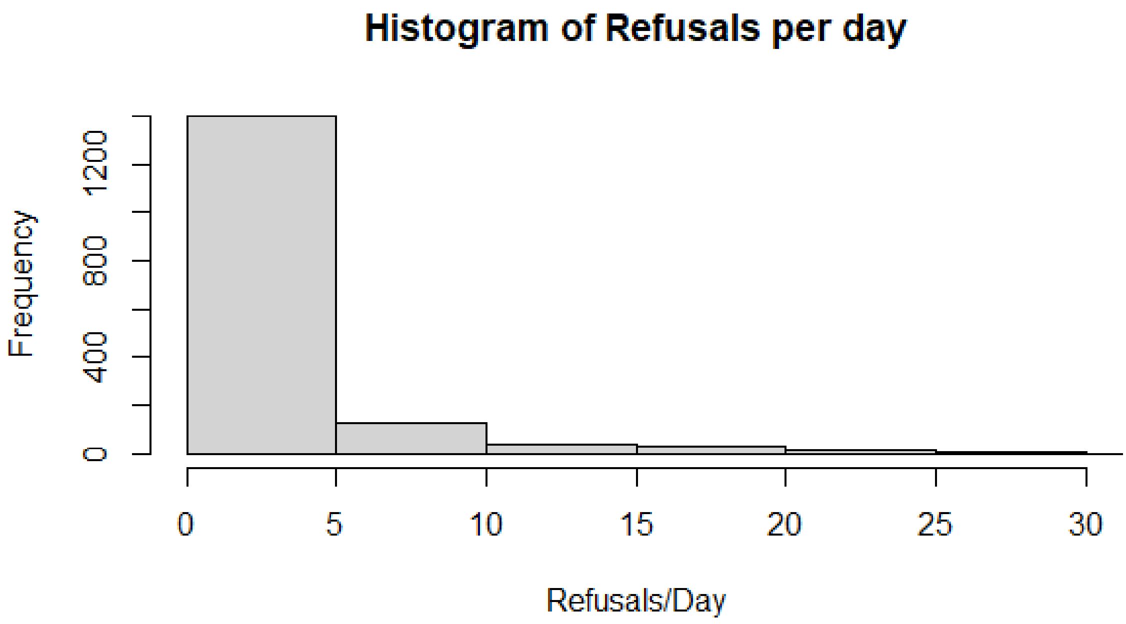

Figure 27 This graph shows a histogram of refusals per day. Refusals per day were collected from the milk visit data from the Lely Astronaut A3 milking system. Refusals per day is included across the x axis in bins of size 5, and frequency is plotted across the y axis. The majority of refusals/day is 0, with a mean of 0 and a median of 0. Because of this fact, no natural logarithm was applied to this feature as we thought it important to observe the difference of number of refusals between lame and non-lame cows.

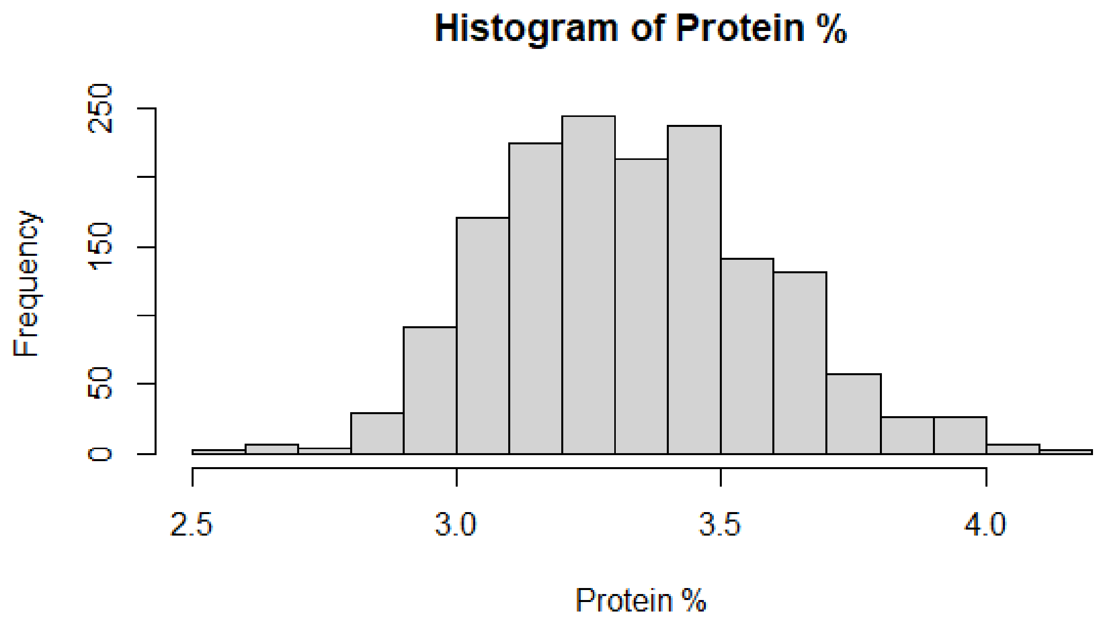

Figure 28 This graph shows a histogram of protein percentage. Protein percentage is included across the x axis in bins of size 0.1, and frequency is plotted across the y axis. This demonstrates an evenly distributed, bell shaped curve and so no natural logarithm was applied to this feature.

Appendix 9 Boxplots to show distribution of each variable by mobility score

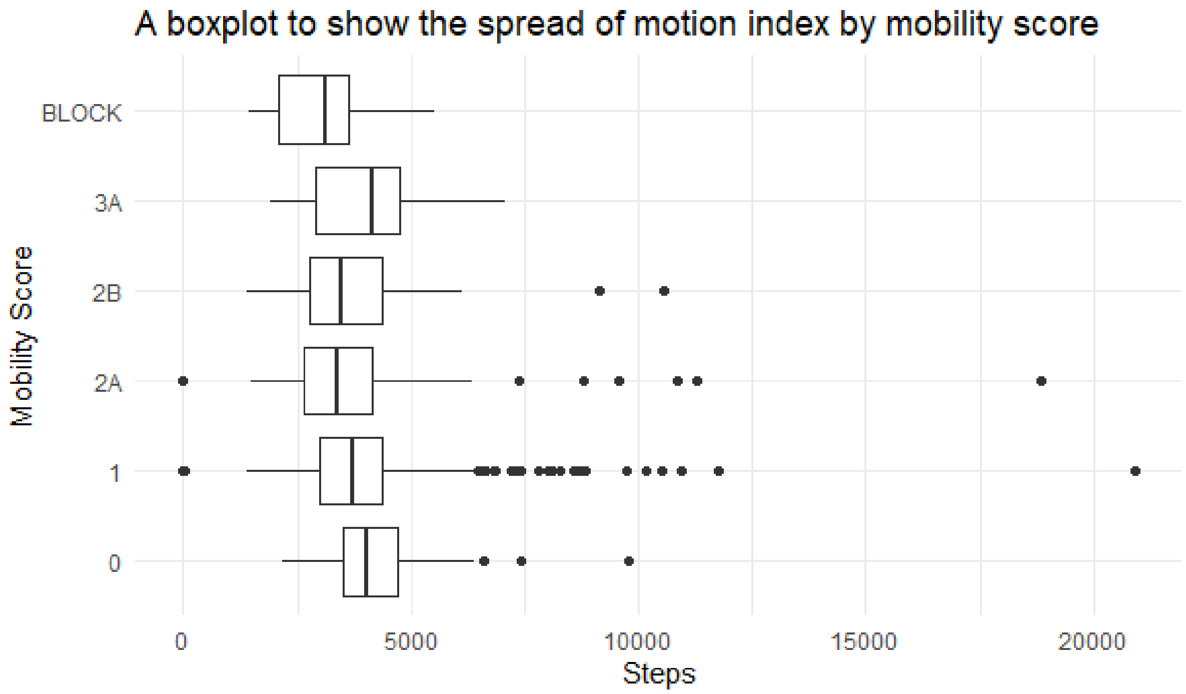

Figure 29 A graph to show the spread of motion index by mobility score. Motion index is collected by the ICEQube sensors. Since it is heavily correlated with steps from the ICEQube sensors, we have left it out of our analysis for the most part although have included it in our statistical models in favour of total steps. Motion index is a measure of animal's activity which considers the absolute value of the 3-D acceleration and is related to the total amount of energy used by the animal over a given period (ICERobotics, Edinburgh, Scotland, UK).

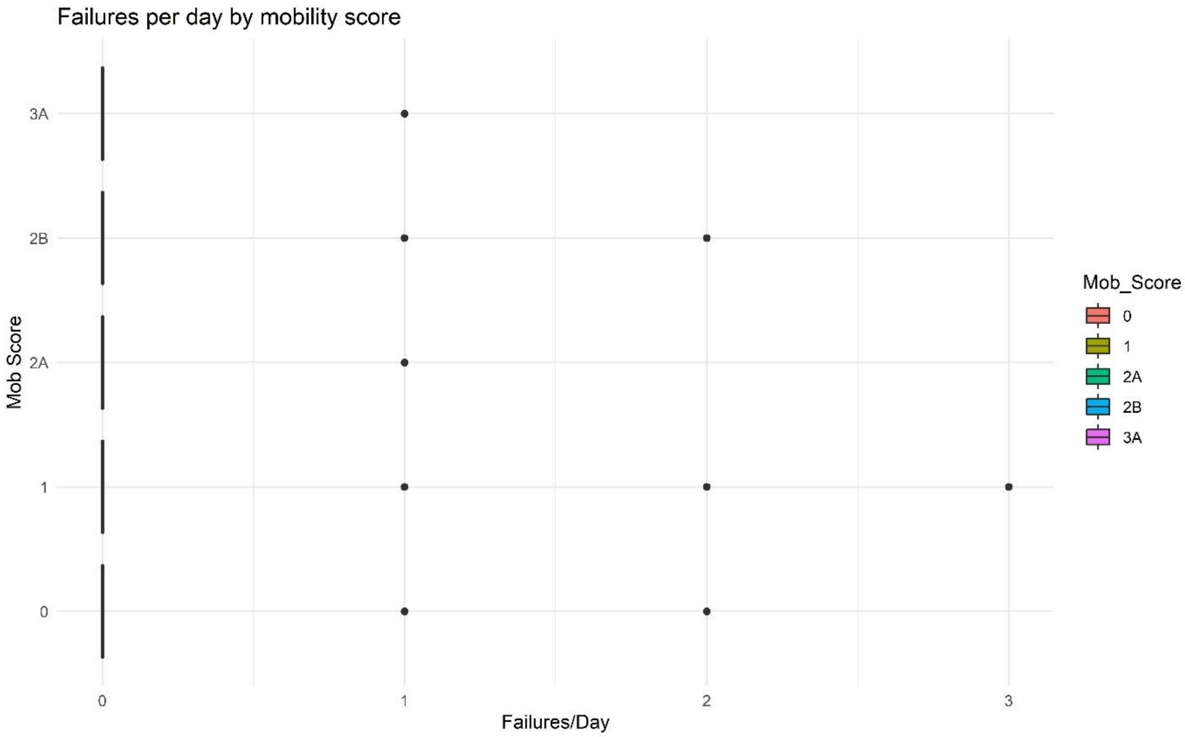

Figure 30 This graph displays the number of failures per day per mobility score. The number of failures per day is computed from the AMS Robot Milker (Lely Astronaut A3). This depicts what happens when the robot fails to connect to a cow for some reason, for example the system has broken down. This is plotted on the x axis with mobility score plotted on the y axis.

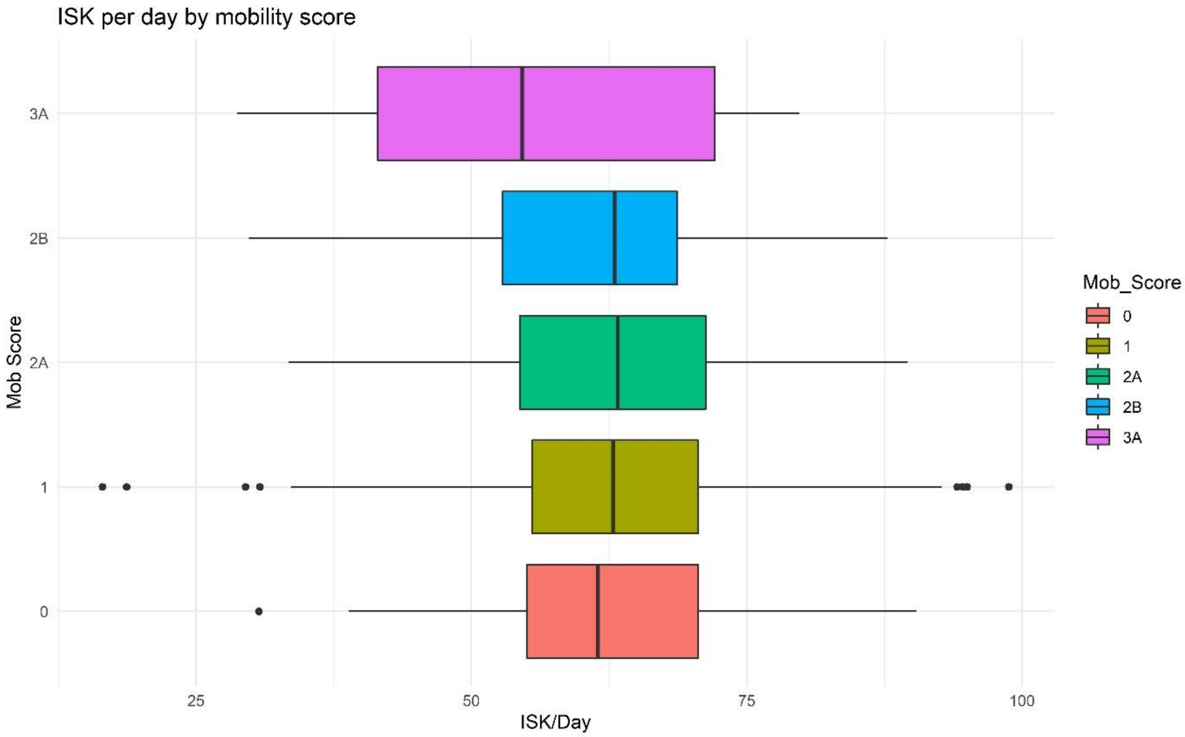

Figure 31 Displays the ISK per cow per day by mobility score. ISK is determined by standardising milk yield, and is a variable already calculated by the Lely T4C system. This is plotted on the x axis, with mobility score on the y axis. This was included as part of the production dataset in the production and combined models.

Appendix 10 Correlation matrix detailing correlation between the variables

The graph below shows a correlation matrix created between all variables. Average weight has a strong positive correlation with parity. Motion index as a strong positive correlation with total steps. There is strong negative correlation between daily yield and fat percentage, DIM and protein percentage. There is a positive correlation between fat percentage and DIM and lactose percentage. Lying time has a strong negative correlation with total steps from both the ICEQube dataset and the Lely Qwes-H dataset.

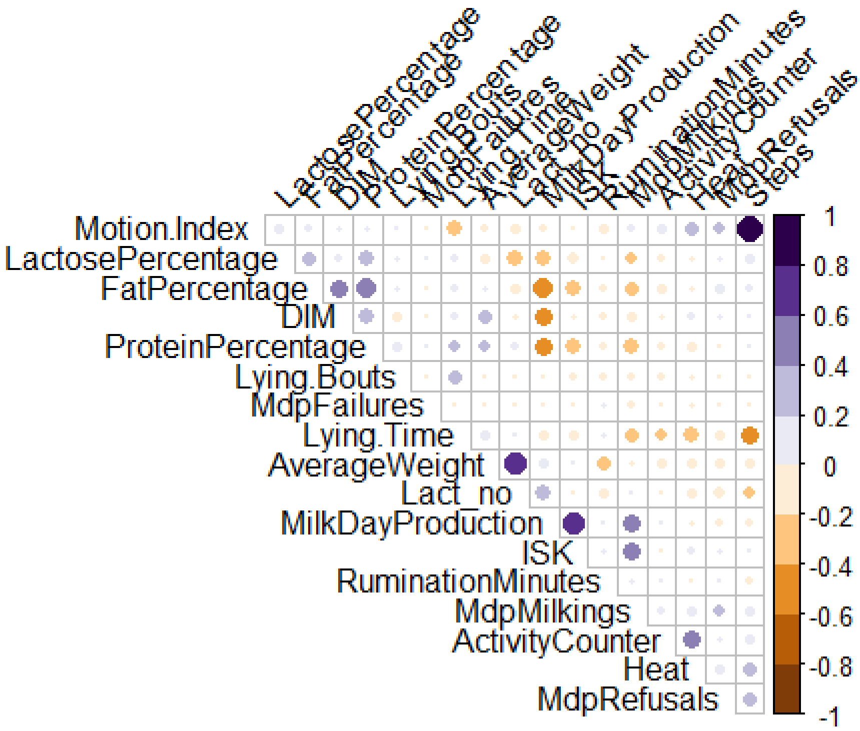

Figure 32 A correlation matrix between all of the main variables used in the combined model. This graph gives information on how correlated each possible combination of pairs of features and how well they relate to each other. Purple coloured circles depict features that are positively correlated (when one increases, the other increases) and orange coloured circles depict those that are negatively correlated (when one increases, the other decreases). The size and intensity of dot colour shows how strongly correlated each variable pair is, as displayed in the legend on the right hand side.

Appendix 11 Variable importance table for the ICEQube data

Explanation of features:

Lying time – total lying time per day

Steps – total number of steps per day

Motion index – taking into account number of steps and acceleration, very highly correlated with steps

Lying bouts – number of lying bouts for that cow for that day

Diff- difference of that cow's daily lying time/steps/motion index from the mean daily lying time/steps/motion index of all the cows in the group

Lag(x) – the value for the daily lying time/steps/motion index x number of days ago

_7da – 7 day average for that feature

_7sd – 7 day standard deviation for that feature

Table 14 A table detailing the relative importance values for the ICEQube data model. The importance of each feature is given in the right hand column, and the feature on the left hand column. The feature importance is a measure of how much that feature contributes to the model and is calculated by tracking the changes in model statistics for each predictor, by measuring the reduction in the model metric when each feature is added to the model. This is then scaled to be out of 100. Features with high variable importance have high contribution to the model.

Feature	Importance
Lying.Time_7da	100
Lying.Time.Diff	56.65433
Lying.Timelag5	50.97291
Lying.Timelag3	49.51279
Lying.Timelag6	39.29143
Lying.Time_7sd	38.64616
Steps.Diff	36.97024
Lying.Time	35.87716

Lying.Timelag1	34.3751
Lying.Timelag2	34.12681
Lying.Timelag4	33.55492
Motion.Index.Diff	30.4148
Stepslag4	28.38442
Motion.Index	26.14321
Stepslag3	25.11656
Lying.Bouts.Diff	24.37879
Motion.Indexlag1	24.37877
Motion.Indexlag6	23.99358
Motion.Indexlag5	23.81396
Stepslag1	23.60025
Lying.Bouts_7da	22.85914
Stepslag5	21.50885
Steps	21.10682
Motion.Indexlag3	20.94648
Stepslag6	20.19492
Lying.Bouts_7sd	19.72249
Stepslag2	19.27072
Motion.Indexlag2	18.51702
Motion.Index_7da	18.13796

Steps_7sd	16.79178
Motion.Indexlag4	16.44733
Motion.Index_7sd	14.42757
Steps_7da	14.2558
Lying.Boutslag6	8.812445
Lying.Boutslag3	6.120716
Lying.Boutslag5	5.319288
Lying.Boutslag4	2.899221
Lying.Boutslag2	1.82476
Lying.Boutslag1	0.107802
Lying.Bouts	0

Appendix 12 Variable importance table for the Lely Qwes-H data

Table 15 A table detailing the relative importance values for the Lely model. The importance of each feature is given in the right hand column, and the feature on the left hand column. The feature importance is a measure of how much that feature contributes to the model and is calculated by tracking the changes in model statistics for each predictor, by measuring the reduction in the model metric when each feature is added to the model. This is then scaled to be out of 100. Features with high variable importance have high contribution to the model.

Variable	Importance
ActivityCounterlag2	100
ActivityCounterlag5	95.92664
ActivityCounterlag3	94.46002
ActivityCounterlag1	91.55454
ActivityCounterlag4	90.86605
ActivityCounter	73.66088
RumDiff	68.49465
RumMinutes	67.81552
ActivityDifflag2	67.33526
RumMinuteslag1	64.6159
ActivityCounterlag6	64.40502
RumDifflag5	64.03917
RumMinuteslag5	63.59101
RumDifflag1	62.50892
RumMinuteslag3	61.4098
RumDifflag6	61.14811

ActivityDifflag1	60.83245
ActivityDiff	60.1393
RumDifflag4	58.89499
RumMinuteslag6	58.11356
RumMinuteslag4	57.29172
RumDifflag3	56.02952
RumDifflag2	55.8453
ActivityDifflag4	55.74714
ActivityDifflag5	54.65684
RumMinuteslag2	54.40449
ActivityDifflag3	53.09277
ActivityDifflag6	52.88687
Heatlag31	17.41501
Heatlag41	7.656153
Heatlag61	6.311884
Heatlag11	1.100552
Heat1	1.027254
Heatlag51	0.301634
Heatlag21	0

Appendix 13 Variable importance graph for production data

Table 16 A table detailing the relative importance values for the production data model. The importance of each feature is given in the right hand column, and the feature on the left hand column. The feature importance is a measure of how much that feature contributes to the model and is calculated by tracking the changes in model statistics for each predictor, by measuring the reduction in the model metric when each feature is added to the model. This is then scaled to be out of 100. Features with high variable importance have high contribution to the model.

Feature	Importance
AverageWeight_7da	100
AverageWeightlag1	94.9
AverageWeightlag3	93.22
AverageWeight	92.85
AverageWeightlag4	91.21
AverageWeightlag6	90.48
MilkDayProduction_7sd	89.97
AverageWeightlag2	88.82
AverageWeightlag5	88.05
Parity	88.00
MdpRefusals_7da	87.75
ISK_7sd	86.58
FatPercentagelag3	84.13
ISKlag2	82.67
ISK	82.41
AverageWeight_7sd	82.18

ProteinPercentage	79.14
ISK_7da	79.04
ISKlag4	78.3
ISKlag5	77.44
FatPercentagelag2	77.32
MilkDayProductionAveragelag1	76.31
ISKlag3	76.29
ProteinPercentagelag1	76.03
MilkDayProduction_7da	75.53
ProteinPercentagelag3	75.26
ProteinPercentagelag4	75.13
MilkDayProductionAveragelag4	75.01
MilkDayProductionAverage	74.79
FatPercentagelag6	74.59
LactosePercentagelag6	74.56
ISKlag1	74.35
FatPercentagelag5	74.09
DIM	73.86
ProteinPercentagelag5	73.37
FatPercentagelag4	73.25
MdpMilkings_7da	73.24

FatPercentage	73.17
ISKlag6	73.16
MilkDayProductionlag5	72.97
FatPercentagelag1	72.89
MilkDayProductionlag1	72.74
ProteinPercentagelag6	72.73
MilkDayProductionlag2	72.63
MilkDayProductionAveragelag5	72.25
MilkDayProductionAveragelag3	72.21
LactosePercentagelag5	72.09
MdpRefusals_7sd	71.68
LactosePercentagelag3	71.01
MilkDayProductionlag3	70.82
MilkDayProductionAveragelag6	70.12
LactosePercentagelag2	69.81
ProteinPercentagelag2	69.74
MilkDayProductionlag4	68.59
MilkDayProductionAveragelag2	68.03
LactosePercentage	67.35
MilkDayProductionlag6	67.24
MilkDayProduction	66.96

LactosePercentagelag4	66.9
LactosePercentagelag1	66.43
MdpRefusalslag6	53.34
MdpMilkings_7sd	52.84
MdpRefusalslag1	46.92
MdpRefusalslag4	43.8
MdpRefusalslag3	43.75
MdpRefusals	41.01
MdpRefusalslag5	40.62
MdpRefusalslag2	37.68
MdpMilkingslag1	32.15
MdpMilkingslag5	30.12
MdpMilkingslag3	29.81
MdpMilkingslag4	29.61
MdpMilkings	29.3
MdpMilkingslag2	27.91
MdpMilkingslag6	27.51
MdpFailures_7da	15.6
MdpFailures_7ds	15.38
MdpFailureslag4	3.467
MdpFailureslag5	2.752

MdpFailureslag1	2.555
MdpFailureslag6	2.319
MdpFailureslag3	1.666
MdpFailureslag2	0.732
MdpFailures	0

Appendix 14 The variable importance for the combined model

Table 17 A table detailing the relative importance values for the combined model. The importance of each feature is given in the right hand column, and the feature on the left hand column. The feature importance is a measure of how much that feature contributes to the model and is calculated by tracking the changes in model statistics for each predictor, by measuring the reduction in the model metric when each feature is added to the model. This is then scaled to be out of 100. Features with high variable importance have high contribution to the model.

Variable	Importance
AverageWeight	100
AverageWeightlag2	98.99479
AverageWeight	94.08855
AverageWeightlag1	93.96193
AverageWeightlag4	91.13025
Lying.Time	72.43053
AverageWeightlag5	71.15554
Lying.Time.Diff	70.24122
Lying.Time_7da	68.75977
AverageWeightlag6	68.66804
Stepslag1	64.81143
Lying.Timelag2	63.2698
Lying.Timelag6	61.40676
Steps	60.14998
Lying.Timelag5	59.65695
Lying.Timelag3	58.54609

Lying.Timelag1	57.98328
Steps.Diff	57.43284
Stepslag2	55.04249
Lying.Timelag4	50.19618
Motion.Index.Diff	47.37983
Stepslag4	46.83463
Parity	45.48141
Stepslag3	44.72832
Steps_7sd	44.62783
Stepslag5	44.16781
ProteinPercentage	43.79961
ISKlag2	43.54615
Motion.Index_7sd	42.70235
ProteinPercentagelag5	42.56325
Lying.Time_7sd	40.72175
Motion.Indexlag1	39.83004
DIM	39.3545
Lying.Bouts_7da	38.88439
ISKlag3	38.7357
ProteinPercentagelag4	38.71495
Steps_7da	38.38439

ISKlag5	38.32748
Motion.Index	37.90424
Lying.Bouts_7sd	37.61086
LactosePercentagelag3	37.33087
ProteinPercentagelag1	36.62545
Lying.Bouts.Diff	36.18853
Motion.Indexlag6	36.17053
ISKlag4	36.07695
Motion.Index_7da	36.07551
Motion.Indexlag2	35.83211
ProteinPercentagelag3	35.54775
ISKlag6	35.364
LactosePercentagelag6	35.24765
FatPercentagelag3	35.08924
ProteinPercentagelag6	35.00723
ISKlag1	33.88464
ProteinPercentagelag2	33.09312
Stepslag6	32.54833
LactosePercentagelag2	32.5349
Motion.Indexlag4	30.90861
FatPercentagelag2	30.74185

MilkDayProductionlag2	30.59999
LactosePercentagelag5	30.42071
Motion.Indexlag5	30.31343
Lying.Boutslag6	29.79329
MilkDayProductionlag3	29.56407
MilkDayProductionAveragelag5	29.50632
Motion.Indexlag3	29.30813
LactosePercentagelag1	29.24175
MilkDayProductionlag1	29.18364
FatPercentagelag4	29.14463
Lying.Boutslag5	29.14054
FatPercentagelag1	28.43176
LactosePercentagelag4	28.37204
MilkDayProductionAveragelag1	28.19166
MilkDayProductionlag6	28.14588
FatPercentagelag5	27.85483
FatPercentagelag6	27.60848
ISK	27.59241
MilkDayProductionAveragelag4	27.25614
MilkDayProductionlag4	26.94583
Lying.Boutslag3	26.42955

LactosePercentage	26.32649
MilkDayProductionAverage	26.0573
MilkDayProductionAveragelag6	25.87137
MilkDayProductionlag5	25.18455
MilkDayProductionAveragelag2	25.00943
MilkDayProduction	25.00207
MilkDayProductionAveragelag3	24.02359
MdpRefusalslag6	22.4586
FatPercentage	22.0126
Lying.Boutslag4	20.99051
MdpRefusalslag3	20.0694
MdpRefusalslag4	19.58215
Lying.Bouts	19.49132
Lying.Boutslag1	19.38202
Lying.Boutslag2	18.70768
MdpRefusalslag1	15.90286
MdpRefusals	13.28812
MdpMilkingslag3	12.71931
MdpMilkingslag2	12.22487
MdpRefusalslag5	11.94419
MdpRefusalslag2	11.22833

MdpMilkingslag5	10.84766
MdpMilkingslag1	9.705383
MdpMilkingslag6	8.379139
MdpMilkingslag4	7.411607
MdpMilkings	7.037667
MdpFailureslag6	3.60013
MdpFailureslag5	2.663971
MdpFailureslag2	2.462404
MdpFailureslag3	2.068846
MdpFailureslag1	0.934397
MdpFailureslag4	0.036225
MdpFailures	0

Appendix 15 Spread of predicted probabilities by lameness group

	group	mean	sd	median	min	max	range
Combined model	0	0.17	0.11	0.16	0.01	0.46	0.46
	1	0.29	0.17	0.28	0.03	0.85	0.82
	2A	0.41	0.19	0.42	0.03	0.84	0.81
	2B	0.59	0.16	0.61	0.18	0.88	0.71
	3A	0.68	0.19	0.75	0.33	0.91	0.58
Production model	0	0.19	0.09	0.18	0.05	0.43	0.38
	1	0.30	0.13	0.29	0.05	0.67	0.63
	2A	0.40	0.14	0.39	0.11	0.70	0.59
	2B	0.54	0.14	0.54	0.28	0.86	0.58
	3A	0.63	0.11	0.65	0.45	0.78	0.33
Lely Qwes-H model	0	0.20	0.11	0.20	0.03	0.54	0.51
	1	0.27	0.15	0.25	0.04	0.75	0.70
	2A	0.33	0.17	0.29	0.02	0.78	0.76
	2B	0.49	0.21	0.51	0.09	0.86	0.77
	3A	0.43	0.23	0.45	0.11	0.79	0.68
ICEQube model	0	0.19	0.17	0.11	0.02	0.76	0.73
	1	0.30	0.17	0.27	0.03	0.80	0.77
	2A	0.41	0.18	0.41	0.05	0.92	0.87
	2B	0.50	0.20	0.47	0.06	0.94	0.88
	3A	0.57	0.13	0.59	0.26	0.72	0.46

Table 18 A table to show the spread of predicted probabilities for each mobility score group. The predicted probabilities were generated from each model (listed on the left) by testing each random forest model on the training data. Here, they have been organised by mobility score and a mean, standard deviation, median, minimum, maximum and range predicted probability has been determined.

Appendix 16 Spread of predicted probabilities by lactation number for the combined model

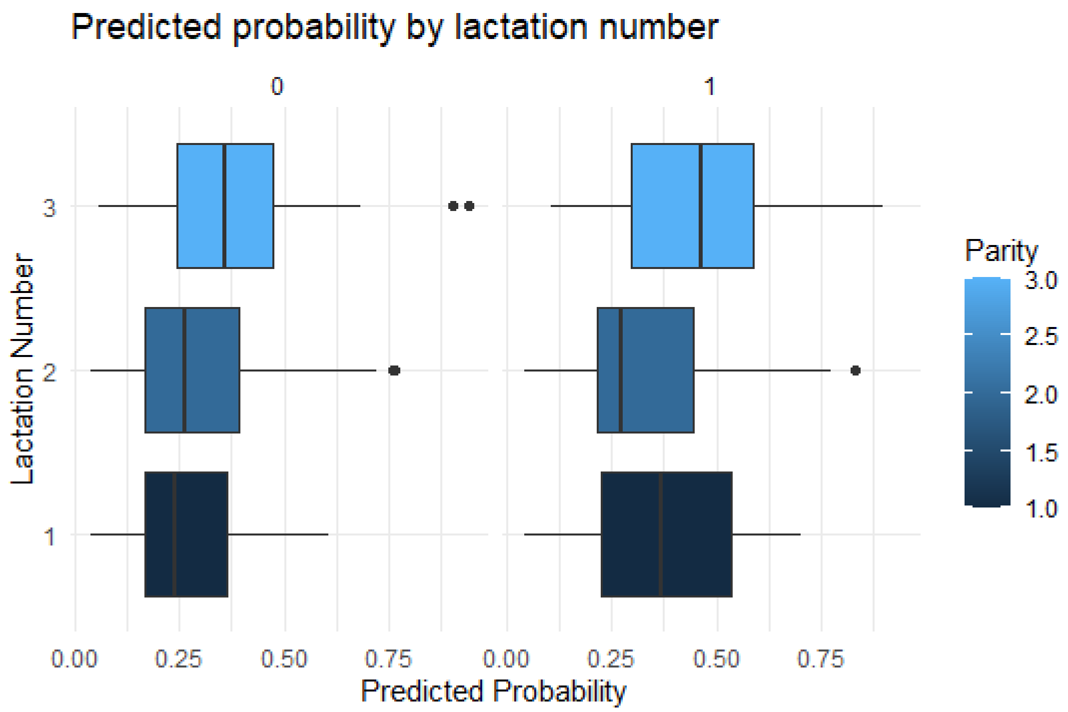

Figure 33 A graph to show the difference in distribution between lame and non-lame cows in each parity in the combined model. Parity is displayed across the x axis, and lactation number/parity on the y axis. The graph on the left represents the distribution in predicted probability for each parity (1, 2 and 3+) for non-lame cows. The graph on the right displays the distribution in predicted probability for each parity (1, 2, 3+) for lame cows. Predicted probabilities have been generated for the combined model from the test dataset, including all of the features from each of the datasets.

7 References

Abdul Jabbar, K., Hansen, M.F., Smith, M.L., Smith, L.N., 2017. Early and non-intrusive lameness detection in dairy cows using 3-dimensional video. Biosystems Engineering 153, 63–69. https://doi.org/10.1016/j.biosystemseng.2016.09.017

Adams, A.E., Lombard, J.E., Fossler, C.P., Román-Muñiz, I.N., Kopral, C.A., 2017. Associations between housing and management practices and the prevalence of lameness, hock lesions, and thin cows on US dairy operations. J Dairy Sci 100, 2119–2136. https://doi.org/10.3168/JDS.2016-11517

Alawneh, J.I., Laven, R.A., Stevenson, M.A., 2012a. Interval between detection of lameness by locomotion scoring and treatment for lameness: A survival analysis. The Veterinary Journal 193, 622–625. https://doi.org/10.1016/J.TVJL.2012.06.042

Alawneh, J.I., Stevenson, M.A., Williamson, N.B., Lopez-Villalobos, N., Otley, T., 2012b. The effect of clinical lameness on liveweight in a seasonally calving, pasture-fed dairy herd. Journal of Dairy Science 95, 663–669. https://doi.org/10.3168/JDS.2011-4505

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

Alsaaod, M., Büscher, W., 2012. Detection of hoof lesions using digital infrared thermography in dairy cows. Journal of Dairy Science 95, 735–742. https://doi.org/10.3168/JDS.2011-4762

Alsaaod, M., Fadul, M., Steiner, A., 2019a. Automatic lameness detection in cattle. The Veterinary Journal 246, 35–44. https://doi.org/10.1016/J.TVJL.2019.01.005

Alsaaod, M., Kredel, R., Hofer, B., Steiner, A., 2017a. Technical note: Validation of a semi-automated software tool to determine gait-cycle variables in dairy cows. Journal of Dairy Science 100, 4897–4902. https://doi.org/10.3168/jds.2016-12235

Alsaaod, M., Römer, C., Kleinmanns, J., Hendriksen, K., Rose-Meierhöfer, S., Plümer, L., Büscher, W., 2012. Electronic detection of lameness in dairy cows through measuring pedometric activity and lying behavior. Applied Animal Behaviour Science 142, 134–141. https://doi.org/10.1016/j.applanim.2012.10.001

Alsaaod, M., Schaefer, A.L., Büscher, W., Steiner, A., 2015a. The role of infrared thermography as a non-invasive tool for the detection of lameness in cattle. Sensors (Switzerland). https://doi.org/10.3390/s150614513

Alsaaod, M., Syring, C., Dietrich, J., Doherr, M.G., Gujan, T., Steiner, A., 2014. A field trial of infrared thermography as a non-invasive diagnostic tool for early detection of digital dermatitis in dairy cows. Veterinary Journal 199, 281–285. https://doi.org/10.1016/j.tvjl.2013.11.028

Amory, J.R., Barker, Z.E., Wright, J.L., Mason, S.A., Blowey, R.W., Green, L.E., 2008. Associations between sole ulcer, white line disease and digital dermatitis and the milk yield of 1824 dairy cows on 30 dairy cow farms in England and Wales from February 2003-November 2004. Preventive Veterinary Medicine 83, 381–391. https://doi.org/10.1016/j.prevetmed.2007.09.007

Amory, J.R., Kloosterman, P., Barker, Z.E., Wright, J.L., Blowey, R.W., Green, L.E., 2006. Risk factors for reduced locomotion in dairy cattle on nineteen farms in the Netherlands. Journal of Dairy Science 89, 1509–1515. https://doi.org/10.3168/JDS.S0022-0302(06)72218-4

Antanaitis, R.; Juozaitiene,˙ V.; Urbonaviˇcius, G.; Malašauskiene,˙ D.; Televiˇcius, M.; Urbutis, M.; Džermeikaite, K.; Baumgartner, W. ˙ Identification of Risk Factors for Lameness Detection with Help of Biosensors. Agriculture 2021, 11, 610. https://doi.org/10.3390/agriculture 11070610

Archer, S.C., Green, M.J., Huxley, J.N., 2010a. Association between milk yield and serial locomotion score assessments in UK dairy cows. Journal of Dairy Science 93, 4045–4053. https://doi.org/10.3168/JDS.2010-3062

Aungier, S.P.M., Roche, J.F., Diskin, M.G., Crowe, M.A., 2014. Risk factors that affect reproductive target achievement in fertile dairy cows. J Dairy Sci 97, 3472–3487. https://doi.org/10.3168/JDS.2013-7404

Bach, Alex, Dinarés, M., Devant, M., Carré, X., 2007. Associations between lameness and production, feeding and milking attendance of Holstein cows milked with an automatic milking system. Journal of Dairy Research 74, 40–46. https://doi.org/10.1017/S0022029906002184

Bach, A., Iglesias, C., Calsamiglia, S., Devantf, M., 2007. Effect of amount of concentrate offered in automatic milking systems on milking frequency, feeding behavior, and milk production of dairy cattle consuming high amounts of corn silage. Journal of Dairy Science 90, 5049–5055. https://doi.org/10.3168/jds.2007-0347

Balmford, A., Amano, T., Bartlett, H., Chadwick, D., Collins, A., Edwards, D., Field, R., Garnsworthy, P., Green, R., Smith, P., Waters, H., Whitmore, A., Broom, D.M., Chara, J., Finch, T., Garnett, E., Gathorne-Hardy, A., Hernandez-Medrano, J., Herrero, M., Hua, F., Latawiec, A., Misselbrook, T., Phalan, B., Simmons, B.I., Takahashi, T., Vause, J., zu Ermgassen, E., Eisner, R., 2018. The environmental costs and benefits of high-yield farming. Nat Sustain 1, 477. https://doi.org/10.1038/s41893-018-0138-5

Barker, Z.E., Amory, J.R., Wright, J.L., Blowey, R.W., Green, L.E., 2007. Management factors associated with impaired locomotion in dairy cows in England and Wales. Journal of Dairy Science 90, 3270–3277. https://doi.org/10.3168/JDS.2006-176

Barker, Z.E., Leach, K.A., Whay, H.R., Bell, N.J., Main, D.C.J., 2010a. Assessment of lameness prevalence and associated risk factors in dairy herds in England and Wales. Journal of Dairy Science 93, 932–941. https://doi.org/10.3168/JDS.2009-2309

Barker, Z.E., Vázquez Diosdado, J.A., Codling, E.A., Bell, N.J., Hodges, H.R., Croft, D.P., Amory, J.R., 2018a. Use of novel sensors combining local positioning and acceleration to measure feeding behavior differences associated with lameness in dairy cattle. Journal of Dairy Science 101, 6310–6321. https://doi.org/10.3168/JDS.2016-12172

Beer, G., Alsaaod, M., Starke, A., Schuepbach-Regula, G., Müller, H., Kohler, P., Steiner, A., 2016a. Use of extended characteristics of locomotion and feeding behavior for automated identification of lame dairy cows. PLoS ONE 11. https://doi.org/10.1371/JOURNAL.PONE.0155796

Bell, N., Huxley, J., 2009. Letters: Locomotion, lameness and mobility in dairy cows. Veterinary Record 164, 726. https://doi.org/10.1136/VR.164.23.726

Bicalho, R.C., Cheong, S.H., Cramer, G., Guard, C.L., 2007a. Association between a visual and an automated locomotion score in lactating Holstein cows. Journal of Dairy Science 90, 3294–3300. https://doi.org/10.3168/jds.2007-0076

Bicalho, R.C., Machado, V.S., Caixeta, L.S., 2009a. Lameness in dairy cattle: A debilitating disease or a disease of debilitated cattle? A cross-sectional study of lameness prevalence and thickness of the digital cushion. J Dairy Sci 92, 3175–3184. https://doi.org/10.3168/JDS.2008-1827

Blackie, N., Maclaurin, L., 2019. Influence of lameness on the lying behaviour of zero-grazed lactating jersey dairy cattle housed in straw yards. Animals 9. https://doi.org/10.3390/ANI9100829

Booth, C.J., Warnick, L.D., Gröhn, Y.T., Maizon, D.O., Guard, C.L., Janssen, D., 2004a. Effect of Lameness on Culling in Dairy Cows. Journal of Dairy Science 87, 4115–4122. https://doi.org/10.3168/JDS.S0022-0302(04)73554-7

Borderas, T.F., Fournier, A., Rushen, J., de Passillé, A.M.B., 2011. Effect of lameness on dairy cows' visits to automatic milking systems. https://doi.org/10.4141/CJAS07014 88, 1–8. https://doi.org/10.4141/CJAS07014

Borghart, G.M., O'Grady, L.E., Somers, J.R., 2021. Prediction of lameness using automatically recorded activity, behavior and production data in post-parturient Irish dairy cows. Ir Vet J 74. https://doi.org/10.1186/S13620-021-00182-6

Byabazaire, J., Olariu, C., Taneja, M., Davy, A., 2019. Lameness Detection as a Service: Application of Machine Learning to an Internet of Cattle. 2019 16th IEEE Annual Consumer Communications and Networking Conference, CCNC 2019. https://doi.org/10.1109/CCNC.2019.8651681

Cattaneo, L., Baudracco, J., Lazzarini, B., Ortega, H., 2015. Methodology to estimate the cost of delayed pregnancy for dairy cows: An example for

Argentina. Revista Brasileira de Zootecnia 44, 226–229. https://doi.org/10.1590/S1806-92902015000600005

Cha, E., Hertl, J.A., Bar, D., Gröhn, Y.T., 2010. The cost of different types of lameness in dairy cows calculated by dynamic programming. Prev Vet Med 97, 1–8. https://doi.org/10.1016/J.PREVETMED.2010.07.011

Channon, A.J., Walker, A.M., Pfau, T., Sheldon, I.M., Wilson, A.M., 2009. Variability of Manson and Leaver locomotion scores assigned to dairy cows by different observers. Veterinary Record 164, 388–392. https://doi.org/10.1136/VR.164.13.388

Chapa, J.M., Maschat, K., Iwersen, M., Baumgartner, J., Drillich, M., 2020a. Accelerometer systems as tools for health and welfare assessment in cattle and pigs – A review. Behavioural Processes 181. https://doi.org/10.1016/J.BEPROC.2020.104262

Chapinal, N., de Passillé, A.M., Pastell, M., Hänninen, L., Munksgaard, L., Rushen, J., 2011. Measurement of acceleration while walking as an automated method for gait assessment in dairy cattle. J Dairy Sci 94, 2895–2901. https://doi.org/10.3168/JDS.2010-3882

Chapinal, N., Liang, Y., Weary, D.M., Wang, Y., von Keyserlingk, M.A.G., 2014. Risk factors for lameness and hock injuries in Holstein herds in China. Journal of Dairy Science 97, 4309–4316. https://doi.org/10.3168/jds.2014-8089

Chapinal, N., Tucker, C.B., 2012. Validation of an automated method to count steps while cows stand on a weighing platform and its application as a measure to detect lameness. J Dairy Sci 95, 6523–6528. https://doi.org/10.3168/JDS.2012-5742

Charlton, G., Gauld, C., Veronesi, F., Rutter, S.M., Bleach, E., 2022. Assessing the Accuracy of Leg Mounted Sensors for Recording Dairy Cow Behavioural Activity at Pasture, in Cubicle Housing and a Straw Yard. https://doi.org/10.3390/ani12050638

Charlton, G.L., Bouffard, V., Gibbons, J., Vasseur, E., Haley, D.B., Pellerin, D., Rushen, J., de Passillé, A.M., 2016. Can automated measures of lying time help assess lameness and leg lesions on tie-stall dairy farms? Applied Animal Behaviour Science 175, 14–22. https://doi.org/10.1016/j.applanim.2015.02.011

Clarkson, M.J., Downham, D.Y., Faull, W.B., Hughes, J.W., Manson, F.J., Merritt, J.B., Murray, R.D., Russell, W.B., Sutherst, J.E., Ward, W.R., 1996. Incidence and prevalence of lameness in dairy cattle. Veterinary Record 138, 563–567. https://doi.org/10.1136/VR.138.23.563

Cocco, R., Canozzi, M.E.A., Fischer, V., 2021. Rumination time as an early predictor of metritis and subclinical ketosis in dairy cows at the beginning of lactation: Systematic review-meta-analysis. Preventive Veterinary Medicine 189, 105309. https://doi.org/10.1016/J.PREVETMED.2021.105309

Cockcroft, P.D., Henson, F.M.D., Parker, C., 2000. Thermography of a septic metatarsophalangeal joint in a heifer. Veterinary Record 146, 258–260. https://doi.org/10.1136/VR.146.9.258

Cramer, G., Lissemore, K.D., Guard, C.L., Leslie, K.E., Kelton, D.F., 2009a. The association between foot lesions and culling risk in Ontario Holstein cows. J Dairy Sci 92, 2572–2579. https://doi.org/10.3168/JDS.2008-1532

de Mol, R.M., André, G., Bleumer, E.J.B., van der Werf, J.T.N., de Haas, Y., van Reenen, C.G., 2013. Applicability of day-to-day variation in behavior for the automated detection of lameness in dairy cows. Journal of Dairy Science 96, 3703–3712. https://doi.org/10.3168/JDS.2012-6305

Delagarde, R., Lamberton, P., 2015. Daily grazing time of dairy cows is recorded accurately using the Lifecorder Plus device. Applied Animal Behaviour Science 165, 25–32. https://doi.org/10.1016/J.APPLANIM.2015.01.014

Deming, J.A., Bergeron, R., Leslie, K.E., DeVries, T.J., 2013. Associations of housing, management, milking activity, and standing and lying behavior of dairy cows milked in automatic systems. Journal of Dairy Science 96, 344–351. https://doi.org/10.3168/JDS.2012-5985

Dippel, S., Dolezal, M., Brenninkmeyer, C., Brinkmann, J., March, S., Knierim, U., Winckler, C., 2009. Risk factors for lameness in freestall-housed dairy cows across two breeds, farming systems, and countries. Journal of Dairy Science 92, 5476–5486. https://doi.org/10.3168/JDS.2009-2288

Dolecheck, K., Bewley, J., 2018. Animal board invited review: Dairy cow lameness expenditures, losses and total cost. Animal 12, 1462–1474. https://doi.org/10.1017/S1751731118000575

Dolecheck, K.A., Overton, M.W., Mark, T.B., Bewley, J.M., 2019. Use of a stochastic simulation model to estimate the cost per case of digital dermatitis, sole ulcer, and white line disease by parity group and incidence timing. Journal of Dairy Science 102, 715–730. https://doi.org/10.3168/JDS.2018-14901

Dunthorn, J., Dyer, R.M., Neerchal, N.K., Mchenry, J.S., Rajkondawar, P.G., Steingraber, G., Tasch, U., 2015. Predictive models of lameness in dairy cows achieve high sensitivity and specificity with force measurements in three dimensions. Journal of Dairy Research 82, 391–399. https://doi.org/10.1017/S002202991500028X

Dutton-Regester, K.J., Wright, J.D., Rabiee, A.R., Barnes, T.S., 2019. Understanding dairy farmer intentions to make improvements to their management practices of foot lesions causing lameness in dairy cows. Preventive Veterinary Medicine 171, 104767. https://doi.org/10.1016/J.PREVETMED.2019.104767

Engel, B., Bruin, G., Andre, G., Buist, W., 2003. Assessment of observer performance in a subjective scoring system: Visual classification of the gait of cows. Journal of Agricultural Science 140, 317–333. https://doi.org/10.1017/S0021859603002983

Erbe, M., Gredler, B., Seefried, F.R., Bapst, B., Simianer, H., 2013. A function accounting for training set size and marker density to model the average accuracy of genomic prediction. PLoS One 8. https://doi.org/10.1371/JOURNAL.PONE.0081046

Espejo, L.A., Endres, M.I., Salfer, J.A., 2006. Prevalence of lameness in high-producing holstein cows housed in freestall barns in Minnesota. J Dairy Sci 89, 3052–3058. https://doi.org/10.3168/JDS.S0022-0302(06)72579-6

Fabian, J., Laven, R.A., Whay, H.R., 2014a. The prevalence of lameness on New Zealand dairy farms: A comparison of farmer estimate and locomotion scoring. Veterinary Journal 201, 31–38. https://doi.org/10.1016/j.tvjl.2014.05.011

Flower, F.C., Sanderson, D.J., Weary, D.M., 2005. Hoof pathologies influence kinematic measures of dairy cow gait. Journal of Dairy Science 88, 3166–3173. https://doi.org/10.3168/jds.S0022-0302(05)73000-9

Flower, F.C., Weary, D.M., 2006. Effect of hoof pathologies on subjective assessments of dairy cow gait. Journal of Dairy Science 89, 139–146. https://doi.org/10.3168/JDS.S0022-0302(06)72077-X

Freeman, E.A., Moisen, G., 2008. PresenceAbsence: An R Package for Presence Absence Analysis. Journal of Statistical Software 23, 1–31. https://doi.org/10.18637/JSS.V023.I11

Fürst-Waltl, B., Egger-Danner, C., Guggenbichler, S., Kofler, J., 2021. Auswirkung von Lahmheit auf Fruchtbarkeitsmerkmale bei Fleckvieh-Kühen in Österreich – Ergebnisse aus dem Efficient-Cow-Projekt. Schweizer Archiv fur Tierheilkunde 164, 721–736. https://doi.org/10.17236/SAT00323

Galindo, F., Broom, D.M., 2002a. Effects of lameness of dairy cows. J Appl Anim Welf Sci 5, 193–201. https://doi.org/10.1207/S15327604JAWS0503_03

Galindo, F., Broom, D.M., 2002b. The effects of lameness on social and individual behavior of dairy cows. Journal of Applied Animal Welfare Science 5, 193–201. https://doi.org/10.1207/S15327604JAWS0503_03

Galindo, F., Broom, D.M., 2000. The relationships between social behaviour of dairy cows and the occurrence of lameness in three herds. Research in Veterinary Science 69, 75–79. https://doi.org/10.1053/RVSC.2000.0391

Garcia, E., Klaas, I., Amigo, J.M., Bro, R., Enevoldsen, C., 2014. Lameness detection challenges in automated milking systems addressed with partial least squares discriminant analysis. Journal of Dairy Science 97, 7476–7486. https://doi.org/10.3168/JDS.2014-7982

Garcia, E., König, K., Allesen-Holm, B.H., Klaas, I.C., Amigo, J.M., Bro, R., Enevoldsen, C., 2015. Experienced and inexperienced observers achieved relatively high within-observer agreement on video mobility scoring of dairy cows. Journal of Dairy Science 98, 4560–4571. https://doi.org/10.3168/JDS.2014-9266

Green, L.E., Hedges, V.J., Schukken, Y.H., Blowey, R.W., Packington, A.J., 2002a. The Impact of Clinical Lameness on the Milk Yield of Dairy Cows. Journal of Dairy Science 85, 2250–2256. https://doi.org/10.3168/JDS.S0022-0302(02)74304-X

Green, L.E., Huxley, J.N., Banks, C., Green, M.J., 2014a. Temporal associations between low body condition, lameness and milk yield in a UK dairy herd. Preventive Veterinary Medicine 113, 63–71. https://doi.org/10.1016/j.prevetmed.2013.10.009

Gregorini, P., dela Rue, B., Pourau, M., Glassey, C., Jago, J., 2013. A note on rumination behavior of dairy cows under intensive grazing systems. Livestock Science 158, 151–156. https://doi.org/10.1016/J.LIVSCI.2013.10.012

Gregorini, P., DelaRue, B., McLeod, K., Clark, C.E.F., Glassey, C.B., Jago, J., 2012. Rumination behavior of grazing dairy cows in response to restricted time at pasture. Livestock Science 146, 95–98. https://doi.org/10.1016/J.LIVSCI.2012.02.020

Griffiths, B.E., White, D.G., Oikonomou, G., 2018a. A cross-sectional study into the prevalence of dairy cattle lameness and associated herd-level risk factors in England and Wales. Frontiers in Veterinary Science 5, 1. https://doi.org/10.3389/FVETS.2018.00065/FULL

Grimm, K., Haidn, B., Erhard, M., Tremblay, M., Döpfer, D., 2019. New insights into the association between lameness, behavior, and performance in Simmental cows. Journal of Dairy Science 102, 2453–2468. https://doi.org/10.3168/JDS.2018-15035

Haladjian, J., Haug, J., Nüske, S., Bruegge, B., 2018. A wearable sensor system for lameness detection in dairy cattle. Multimodal Technologies and Interaction 2. https://doi.org/10.3390/MTI2020027

Haskell, M.J., Rennie, L.J., Bowell, V.A., Bell, M.J., Lawrence, A.B., 2006. Housing system, milk production, and zero-grazing effects on lameness and leg injury in dairy cows. Journal of Dairy Science 89, 4259–4266. https://doi.org/10.3168/JDS.S0022-0302(06)72472-9

Hernandez, J., Shearer, J.K., Webb, D.W., 2002. Effect of lameness on milk yield in dairy cows. J Am Vet Med Assoc 220, 640–644. https://doi.org/10.2460/JAVMA.2002.220.640

Hoffman, A.C., Moore, D.A., Vanegas, J., Wenz, J.R., 2014. Association of abnormal hind-limb postures and back arch with gait abnormality in dairy cattle. J Dairy Sci 97, 2178–2185. https://doi.org/10.3168/JDS.2013-7528

Horseman, S. v., Roe, E.J., Huxley, J.N., Bell, N.J., Mason, C.S., Whay, H.R., 2014. The use of in-depth interviews to understand the process of treating lame dairy cows from the farmers' perspective. Animal Welfare 23, 157–165. https://doi.org/10.7120/09627286.23.2.157

Hultgren, J., Manske, T., Bergsten, C., 2004. Associations of sole ulcer at claw trimming with reproductive performance, udder health, milk yield, and culling in Swedish dairy cattle. Preventive Veterinary Medicine 62, 233–251. https://doi.org/10.1016/J.PREVETMED.2004.01.002

Hut, P.R., Hostens, M.M., Beijaard, M.J., van Eerdenburg, F.J.C.M., Hulsen, J.H.J.L., Hooijer, G.A., Stassen, E.N., Nielen, M., 2021. Associations between body condition score, locomotion score, and sensor-based time budgets of dairy cattle during the dry period and early lactation. J Dairy Sci 104, 4746–4763. https://doi.org/10.3168/JDS.2020-19200

Huxley, J.N., 2013. Impact of lameness and claw lesions in cows on health and production. Livestock Science 156, 64–70. https://doi.org/10.1016/J.LIVSCI.2013.06.012

Huxley, J.N., 2012. Lameness in cattle: An ongoing concern. Veterinary Journal 193, 610–611. https://doi.org/10.1016/j.tvjl.2012.06.039

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Ito, K., von Keyserlingk, M.A.G., LeBlanc, S.J., Weary, D.M., 2010. Lying behavior as an indicator of lameness in dairy cows. Journal of Dairy Science 93, 3553–3560. https://doi.org/10.3168/jds.2009-2951

K A O'Callaghan, P J Cripps, D Y Downham, R D Murray, 2003. SUBJECTIVE AND OBJECTIVE ASSESSMENT OF PAIN AND DISCOMFORT DUE TO LAMENESS IN DAIRY CATTLE . Animal Welfare 605–610.

Kamphuis, C., Frank, E., Burke, J.K., Verkerk, G.A., Jago, J.G., 2013. Applying additive logistic regression to data derived from sensors monitoring behavioral and physiological characteristics of dairy cows to detect lameness. Journal of Dairy Science 96, 7043–7053. https://doi.org/10.3168/jds.2013-6993

Kaniyamattam, K., Hertl, J., Lhermie, G., Tasch, U., Dyer, R., Gröhn, Y.T., 2020. Cost benefit analysis of automatic lameness detection systems in dairy herds: A dynamic programming approach. Preventive Veterinary Medicine 178. https://doi.org/10.1016/J.PREVETMED.2020.104993

Kilgour, R.J., 2012. In pursuit of "normal": A review of the behaviour of cattle at pasture. Applied Animal Behaviour Science 138, 1–11. https://doi.org/10.1016/J.APPLANIM.2011.12.002

King, M.T.M., LeBlanc, S.J., Pajor, E.A., DeVries, T.J., 2017a. Cow-level associations of lameness, behavior, and milk yield of cows milked in automated systems. Journal of Dairy Science 100, 4818–4828. https://doi.org/10.3168/JDS.2016-12281

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B.J., Hróbjartsson, A., Roberts, C., Shoukri, M., Streiner, D.L., 2011. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. Journal of Clinical Epidemiology 64, 96–106. https://doi.org/10.1016/J.JCLINEPI.2010.03.002

Kramer, E., Cavero, D., Stamer, E., Krieter, J., 2009. Mastitis and lameness detection in dairy cows by application of fuzzy logic. Livestock Science 125, 92–96. https://doi.org/10.1016/J.LIVSCI.2009.02.020

Leach, K.A., Dippel, S., Huber, J., March, S., Winckler, C., Whay, H.R., 2009. Assessing lameness in cows kept in tie-stalls. Journal of Dairy Science 92, 1567–1574. https://doi.org/10.3168/JDS.2008-1648

Leach, K.A., Tisdall, D.A., Bell, N.J., Main, D.C.J., Green, L.E., 2012. The effects of early treatment for hindlimb lameness in dairy cows on four commercial UK farms. Veterinary Journal 193, 626–632. https://doi.org/10.1016/j.tvjl.2012.06.043

Leach, K.A., Whay, H.R., Maggs, C.M., Barker, Z.E., Paul, E.S., Bell, A.K., Main, D.C.J., 2010a. Working towards a reduction in cattle lameness: 1. Understanding barriers to lameness control on dairy farms. Research in Veterinary Science 89, 311–317. https://doi.org/10.1016/J.RVSC.2010.02.014

Lim, P.Y., Huxley, J.N., Willshire, J.A., Green, M.J., Othman, A.R., Kaler, J., 2015. Unravelling the temporal association between lameness and body condition score in dairy cattle using a multistate modelling approach. Prev Vet Med 118, 370–377. https://doi.org/10.1016/J.PREVETMED.2014.12.015

Liu, H., Wu, T., 2003. Estimating the Area under a Receiver Operating Characteristic Curve For Repeated Measures Design. Journal of Statistical Software 8, 1–18. https://doi.org/10.18637/JSS.V008.I12

Liu, J., Dyer, R.M., Neerchal, N.K., Tasch, U., Rajkondawar, P.G., 2011. Diversity in the magnitude of hind limb unloading occurs with similar forms of lameness in dairy cows. The Journal of Dairy Research 78, 168–177. https://doi.org/10.1017/s0022029911000057

LokeshBabu, D.S., Vasant, P.J., Jeyakumar, S., Manimaran, A., Kumaresan, A., Pushpadass, H.A., Sivaram, M., Ramesha, K.P., Kataktalware, M.A., Siddaramanna, Sathiyabarathi, M., 2018. Monitoring foot surface temperature using infrared thermal imaging for assessment of hoof health status in cattle: A review. Journal of Thermal Biology 78, 10–21. https://doi.org/10.1016/J.JTHERBIO.2018.08.021

López-Ratón, M., Rodríguez-Álvarez, M.X., Cadarso-Suárez, C., Gude-Sampedro, F., 2014. OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. Journal of Statistical Software 61, 1–36. https://doi.org/10.18637/JSS.V061.I08

Lovarelli, D., Bacenetti, J., Guarino, M., 2020. A review on dairy cattle farming: Is precision livestock farming the compromise for an environmental, economic and social sustainable production? Journal of Cleaner Production 262, 121409. https://doi.org/10.1016/J.JCLEPRO.2020.121409

Lucey, S., Rowlands, G.J., Russell, A.M., 1986. The association between lameness and fertility in dairy cows. Vet Rec 118, 628–631. https://doi.org/10.1136/VR.118.23.628

Maertens, W., Vangeyte, J., Baert, J., Jantuan, A., Mertens, K.C., de Campeneere, S., Pluk, A., Opsomer, G., van Weyenberg, S., van Nuffel, A., 2011. Development of a real time cow gait tracking and analyzing tool to assess lameness using a pressure sensitive walkway: the GAITWISE system. Biosystems Engineering 110, 29–39. https://doi.org/10.1016/j.biosystemseng.2011.06.003

Main, D.C.J., Leach, K.A., Barker, Z.E., Sedgwick, A.K., Maggs, C.M., Bell, N.J., Whay, H.R., 2012a. Evaluating an intervention to reduce lameness in dairy cattle. Journal of Dairy Science 95, 2946–2954. https://doi.org/10.3168/JDS.2011-4678

Main, D.C.J., Stokes, J.E., Reader, J.D., Whay, H.R., 2012b. Detecting hoof lesions in dairy cattle using a hand-held thermometer. Vet Rec 171, 504. https://doi.org/10.1136/VR.100533

Mangweth, G., Schramel, J.P., Peham, C., Gasser, C., Tichy, A., Altenhofer, C., Weber, A., Kofler, J., 2012. Lahmheitserkennung bei kühen durch messung der bewegung im schritt mittels accelerometer. Berliner und Munchener Tierarztliche Wochenschrift 125, 386–396. https://doi.org/10.2376/0005-9366-125-386

Marino, L., Allen, K., 2017. The psychology of cows. Animal Behavior and Cognition 2017, 474–498. https://doi.org/10.26451/abc.04.04.06.2017

Martinez-Ortiz, A., 2013. Video tracking of dairy cows for assessing mobility scores.

Martinez-Ortiz, C.A., Everson, R.M., Mottram, T., 2013. Video tracking of dairy cows for assessing mobility scores. ORE.

Maxwell, O.J.R., Hudson, C.D., Huxley, J.N., 2015. Effect of early lactation foot trimming in lame and non-lame dairy heifers: A randomised controlled trial. Veterinary Record 177, 100. https://doi.org/10.1136/VR.103155

Microsoft Corporation, 2018. Microsoft Excel, Available at https://office.microsoft.com/excel.

Miekley, B., Stamer, E., Traulsen, I., Krieter, J., 2013. Implementation of multivariate cumulative sum control charts in mastitis and lameness monitoring. Journal of Dairy Science 96, 5723–5733. https://doi.org/10.3168/JDS.2012-6460

Miguel-Pacheco, G.G., Kaler, J., Remnant, J., Cheyne, L., Abbott, C., French, A.P., Pridmore, T.P., Huxley, J.N., 2014. Behavioural changes in dairy cows with lameness in an automatic milking system. Applied Animal Behaviour Science 150, 1–8. https://doi.org/10.1016/J.APPLANIM.2013.11.003

Miguel-Pacheco, G.G., Thomas, H.J., Huxley, J.N., Newsome, R.F., Kaler, J., 2017. Effect of claw horn lesion type and severity at the time of treatment on outcome of lameness in dairy cows. The Veterinary Journal 225, 16–22. https://doi.org/10.1016/J.TVJL.2017.04.015

Mobility scoring – A simple step to stamp out lameness - Promar International [WWW Document], n.d. URL https://promar-international.com/mobility-scoring-a-simple-step-to-stamp-out-lameness/ (accessed 5.30.22).

Monrad, J., Kassuku, A.A., Nansen, P., Willeberg, P., 1983. An Epidemiological Study of Foot Rot in Pastured Cattle. Acta Veterinaria Scandinavica 24, 403. https://doi.org/10.1186/BF03546714

Murray, R.D., Downham, D.Y., Clarkson, M.J., Faull, W.B., Hughes, J.W., Manson, F.J., Merritt, J.B., Russell, W.B., Sutherst, J.E., Ward, W.R., 1996. Epidemiology of lameness in dairy cattle: Description and analysis of foot lesions. Veterinary Record 138, 586–591. https://doi.org/10.1136/vr.138.24.586

Müschner-Siemens, T., Hoffmann, G., Ammon, C., Amon, T., 2020. Daily rumination time of lactating dairy cows under heat stress conditions. J Therm Biol 88. https://doi.org/10.1016/J.JTHERBIO.2019.102484

Navarro, G., Green, L.E., Tadich, N., 2013. Effect of lameness and lesion specific causes of lameness on time budgets of dairy cows at pasture and when housed. Veterinary Journal 197, 788–793. https://doi.org/10.1016/J.TVJL.2013.05.012

Nechanitzky, K., Starke, A., Vidondo, B., Müller, H., Reckardt, M., Friedli, K., Steiner, A., 2016. Analysis of behavioral changes in dairy cows associated with claw horn lesions. Journal of Dairy Science 99, 2904–2914. https://doi.org/10.3168/JDS.2015-10109

Newcomer, B.W., Chamorro, M.F., 2016. Distribution of lameness lesions in beef cattle: A retrospective analysis of 745 cases. The Canadian Veterinary Journal 57, 401.

Newsome, R., Green, M.J., Bell, N.J., Chagunda, M.G.G., Mason, C.S., Rutland, C.S., Sturrock, C.J., Whay, H.R., Huxley, J.N., 2016a. Linking bone development on the caudal aspect of the distal phalanx with lameness during life. Journal of Dairy Science 99, 4512–4525. https://doi.org/10.3168/JDS.2015-10202

Newsome, R.F., Green, M.J., Bell, N.J., Bollard, N.J., Mason, C.S., Whay, H.R., Huxley, J.N., 2017a. A prospective cohort study of digital cushion and corium thickness. Part 2: Does thinning of the digital cushion and corium lead to lameness and claw horn disruption lesions? Journal of Dairy Science 100, 4759–4771. https://doi.org/10.3168/JDS.2016-12013

Nikkhah, A., Plaizier, J.C., Einarson, M.S., Berry, R.J., Scott, S.L., Kennedy, A.D., 2005. Infrared thermography and visual examination of hooves of dairy cows in two stages of lactation. Journal of Dairy Science 88, 2749–2753. https://doi.org/10.3168/jds.s0022-0302(05)72954-4

Norring, M., Häggman, J., Simojoki, H., Tamminen, P., Winckler, C., Pastell, M., 2014. Short communication: Lameness impairs feeding behavior of dairy cows. Journal of Dairy Science 97, 4317–4321. https://doi.org/10.3168/JDS.2013-7512

O'Callaghan, K.A., Cripps, P.J., Downham, D.Y., Murray, R.D., 2003. Subjective and objective assessment of pain and discomfort due to lameness in dairy cattle. Animal Welfare 12, 605-610 (6).

O'Driscoll, K., Boyle, L., Hanlon, A., 2008. A brief note on the validation of a system for recording lying behaviour in dairy cows. Applied Animal Behaviour Science 111, 195–200. https://doi.org/10.1016/J.APPLANIM.2007.05.014

O'Leary, N.W., Byrne, D.T., O'Connor, A.H., Shalloo, L., 2020a. Invited review: Cattle lameness detection with accelerometers. J Dairy Sci 103, 3895–3911. https://doi.org/10.3168/JDS.2019-17123

Olechnowicz, J.; Jaskowski, J.M. Impact of clinical lameness, calving season, parity, and month of lactation on milk, fat, protein, and lactose yields during early lactation of dairy cows. Bull. Vet. Inst. Pulawy 2010, 54, 605–610.

Olechnowicz, J.; Jaskowski, M.J. Relationship between clinical lameness and somatic cell counts, and fat and protein contents in the milk of dairy cows. Med. Weter 2012, 68, 12

Omontese, B.O., Bellet-Elias, R., Molinero, A., Catandi, G.D., Casagrande, R., Rodriguez, Z., Bisinotto, R.S., Cramer, G., 2020. Association between hoof lesions and fertility in lactating Jersey cows. Journal of Dairy Science 103, 3401–3413. https://doi.org/10.3168/JDS.2019-17252

Palmer, M.A., O'Connell, N.E., 2015. Digital Dermatitis in Dairy Cows: A Review of Risk Factors and Potential Sources of Between-Animal Variation in Susceptibility. Animals : an Open Access Journal from MDPI 5, 512. https://doi.org/10.3390/ANI5030369

Pastell, M., Aisla, A.M., Hautala, M., Poikalainen, V., Praks, J., Veermäe, I., Ahokas, J., 2006. Contactless measurement of cow behavior in a milking robot. Behavior Research Methods 38, 479–486. https://doi.org/10.3758/BF03192802

Pastell, M.E., Kujalaf, M., 2007. A Probabilistic Neural Network Model for Lameness Detection. Journal of Dairy Science 90, 2283–2292. https://doi.org/10.3168/JDS.2006-267

Pavlovic, D., Davison, C., Hamilton, A., Marko, O., Atkinson, R., Michie, C., Crnojević, V., Andonovic, I., Bellekens, X., Tachtatzis, C., 2021. Classification of Cattle Behaviours Using Neck-Mounted Accelerometer-

Equipped Collars and Convolutional Neural Networks. Sensors (Basel) 21. https://doi.org/10.3390/S21124050

(PDF) A herd mobility scoring service: Practicalities, opportunities and models [WWW Document], n.d. URL https://www.researchgate.net/publication/287309019_A_herd_mobility_scoring_service_Practicalities_opportunities_and_models (accessed 5.22.22).

Pérez-Cabal, M.A., Alenda, R., 2014. Clinical lameness and risk factors in a Spanish Holstein population. Livestock Science 164, 168–174. https://doi.org/10.1016/J.LIVSCI.2014.03.012

Pluk, A., Bahr, C., Poursaberi, A., Maertens, W., van Nuffel, A., Berckmans, D., 2012. Automatic measurement of touch and release angles of the fetlock joint for lameness detection in dairy cattle using vision techniques. Journal of Dairy Science 95, 1738–1748. https://doi.org/10.3168/JDS.2011-4547

Poursaberi, A., Bahr, C., Pluk, A., Berckmans, D., Veermäe, I., Kokin, E., Pokalainen, V., 2011. Online lameness detection in dairy cattle using Body Movement Pattern (BMP). International Conference on Intelligent Systems Design and Applications, ISDA 732–736. https://doi.org/10.1109/ISDA.2011.6121743

Puerto, M.A., Shepley, E., Cue, R.I., Warner, D., Dubuc, J., Vasseur, E., 2021. The hidden cost of disease: II. Impact of the first incidence of lameness on production and economic indicators of primiparous dairy cows. J Dairy Sci 104, 7944–7955. https://doi.org/10.3168/JDS.2020-19585

Rajkondawar, P.G., Lefcourt, A.M., Neerchal, N.K., Dyer, R.M., Varner, M.A., Erez, B., Tasch, U., 2002. The development of an objective lameness scoring system for dairy herds: Pilot study. Transactions of the American Society of Agricultural Engineers 45, 1123–1125. https://doi.org/10.13031/2013.9941

Rajkondawar, P.G., Liu, M., Dyer, R.M., Neerchal, N.K., Tasch, U., Lefcourt, A.M., Erez, B., Varner, M.A., 2006. Comparison of models to identify lame cows based on gait and lesion scores, and limb movement variables. Journal of Dairy Science 89, 4267–4275. https://doi.org/10.3168/jds.s0022-0302(06)72473-0

Randall, L.V., Thomas, H.J., Remnant, J.G., Bollard, N.J., Huxley, J.N., 2019. Lameness prevalence in a random sample of UK dairy herds. Veterinary Record 184, 350. https://doi.org/10.1136/VR.105047

Randall, L. v., Green, M.J., Chagunda, M.G.G., Mason, C., Archer, S.C., Green, L.E., Huxley, J.N., 2015. Low body condition predisposes cattle to lameness: An 8-year study of one dairy herd. Journal of Dairy Science 98, 3766–3777. https://doi.org/10.3168/JDS.2014-8863

Randall, L. v., Green, M.J., Chagunda, M.G.G., Mason, C., Green, L.E., Huxley, J.N., 2016. Lameness in dairy heifers; impacts of hoof lesions present around first calving on future lameness, milk yield and culling risk.

Preventive Veterinary Medicine 133, 52. https://doi.org/10.1016/J.PREVETMED.2016.09.006

Random Forest | Introduction to Random Forest Algorithm [WWW Document], n.d. URL https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/ (accessed 5.30.22).

Liaw A, Wiener M (2002). "Classification and Regression by randomForest." R News, 2(3), 18-22. https://CRAN.R-project.org/doc/Rnews/

R Core Team (2022). R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reader, J.D., Green, M.J., Kaler, J., Mason, S.A., Green, L.E., 2011a. Effect of mobility score on milk yield and activity in dairy cattle. Journal of Dairy Science 94, 5045–5052. https://doi.org/10.3168/JDS.2011-4415

Revelle W (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.2.5, https://CRAN.R-project.org/package=psych.

Riaboff, L., Shalloo, L., Smeaton, A.F., Couvreur, S., Madouasse, A., Keane, M.T., 2022. Predicting livestock behaviour using accelerometers: A systematic review of processing techniques for ruminant behaviour prediction from raw accelerometer data. Computers and Electronics in Agriculture 192, 106610. https://doi.org/10.1016/J.COMPAG.2021.106610

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., Müller, M., 2011. pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12, 1–8. https://doi.org/10.1186/1471-2105-12-77/TABLES/3

Rutherford, K.M.D., Langford, F.M., Jack, M.C., Sherwood, L., Lawrence, A.B., Haskell, M.J., 2009. Lameness prevalence and risk factors in organic and non-organic dairy herds in the United Kingdom. Veterinary Journal 180, 95–105. https://doi.org/10.1016/J.TVJL.2008.03.015

Rutten, C.J., Steeneveld, W., Inchaisri, C., Hogeveen, H., 2014. An ex ante analysis on the use of activity meters for automated estrus detection: To invest or not to invest? Journal of Dairy Science 97, 6869–6887. https://doi.org/10.3168/JDS.2014-7948

Rutten, C.J., Velthuis, A.G.J., Steeneveld, W., Hogeveen, H., 2013. Invited review: Sensors to support health management on dairy farms. Journal of Dairy Science 96, 1928–1952. https://doi.org/10.3168/JDS.2012-6107

Schindhelm, K., Lorenzini, I., Tremblay, M., Döpfer, D., Reese, S., Haidn, B., 2017. Automatically recorded performance and behaviour parameters as risk factors for lameness in dairy cattle. Chemical Engineering Transactions 58, 583–588. https://doi.org/10.3303/CET1758098

Schlageter-Tello, A., Bokkers, E.A.M., Koerkamp, P.W.G.G., van Hertem, T., Viazzi, S., Romanini, C.E.B., Halachmi, I., Bahr, C., Berckmans, D., Lokhorst, K., 2014a. Manual and automatic locomotion scoring systems in dairy cows: a review. Prev Vet Med 116, 12–25. https://doi.org/10.1016/J.PREVETMED.2014.06.006

Schlageter-Tello, A., van Hertem, T., Bokkers, E.A.M., Viazzi, S., Bahr, C., Lokhorst, K., 2018. Performance of human observers and an automatic 3-dimensional computer-vision-based locomotion scoring method to detect lameness and hoof lesions in dairy cows. Journal of Dairy Science 101, 6322–6335. https://doi.org/10.3168/jds.2017-13768

Sepúlveda-Varas, P., Weary, D.M., von Keyserlingk, M.A.G., 2014. Lying behavior and postpartum health status in grazing dairy cows. Journal of Dairy Science 97, 6334–6343. https://doi.org/10.3168/JDS.2014-8357

Shane, D.D., White, B.J., Larson, R.L., Amrine, D.E., Kramer, J.L., 2016. Probabilities of cattle participating in eating and drinking behavior when located at feeding and watering locations by a real time location system. Computers and Electronics in Agriculture 127, 460–466. https://doi.org/10.1016/J.COMPAG.2016.07.005

Shearer, J.K., van Amstel, S.R., Brodersen, B.W., 2012a. Clinical diagnosis of foot and leg lameness in cattle. Vet Clin North Am Food Anim Pract 28, 535–556. https://doi.org/10.1016/J.CVFA.2012.07.003

Shearer, J.K., van Amstel, S.R., Brodersen, B.W., 2012b. Clinical Diagnosis of Foot and Leg Lameness in Cattle. Veterinary Clinics of North America: Food Animal Practice 28, 535–556. https://doi.org/10.1016/J.CVFA.2012.07.003

Shepley, E., Berthelot, M., Vasseur, E., 2017. Validation of the ability of a 3D pedometer to accurately determine the number of steps taken by dairy cows when housed in tie-stalls. Agriculture (Switzerland) 7. https://doi.org/10.3390/AGRICULTURE7070053

Solano, L., Barkema, H.W., Pajor, E.A., Mason, S., LeBlanc, S.J., Nash, C.G.R., Haley, D.B., Pellerin, D., Rushen, J., de Passillé, A.M., Vasseur, E., Orsel, K., 2016. Associations between lying behavior and lameness in Canadian Holstein-Friesian cows housed in freestall barns. Journal of Dairy Science 99, 2086–2101. https://doi.org/10.3168/JDS.2015-10336

Somers, J.R., Huxley, J., Lorenz, I., Doherty, M.L., O'Grady, L., 2015. The effect of Lameness before and during the breeding season on fertility in 10 pasture-based Irish dairy herds. Irish Veterinary Journal 68, 1–7. https://doi.org/10.1186/S13620-015-0043-4/TABLES/4

Sprecher, D.J., Hostetler, D.E., Kaneene, J.B., 1997. A lameness scoring system that uses posture and gait to predict dairy cattle reproductive performance. Theriogenology 47, 1179–1187. https://doi.org/10.1016/S0093-691X(97)00098-8

Tadich, N., Flor, E., Green, L., 2010. Associations between hoof lesions and locomotion score in 1098 unsound dairy cows. Veterinary Journal 184, 60–65. https://doi.org/10.1016/J.TVJL.2009.01.005

Thomas, H.J., Miguel-Pacheco, G.G., Bollard, N.J., Archer, S.C., Bell, N.J., Mason, C., Maxwell, O.J.R., Remnant, J.G., Sleeman, P., Whay, H.R., Huxley, J.N., 2015. Evaluation of treatments for claw horn lesions in dairy cows in a randomized controlled trial. Journal of Dairy Science 98, 4477–4486. https://doi.org/10.3168/jds.2014-8982

Thomas, H.J., Remnant, J.G., Bollard, N.J., Burrows, A., Whay, H.R., Bell, N.J., Mason, C., Huxley, J.N., 2016. Recovery of chronically lame dairy cows following treatment for claw horn lesions: A randomised controlled trial. Veterinary Record. https://doi.org/10.1136/vr.103394

Thompson, A.J., Weary, D.M., Bran, J.A., Daros, R.R., Hötzel, M.J., von Keyserlingk, M.A.G., 2019. Lameness and lying behavior in grazing dairy cows. Journal of Dairy Science 102, 6373–6382. https://doi.org/10.3168/JDS.2018-15717

Thorup, V.M., Munksgaard, L., Robert, P.E., Erhard, H.W., Thomsen, P.T., Friggens, N.C., 2015. Lameness detection via leg-mounted accelerometers on dairy cows on four commercial farms. Animal 9, 1704–1712. https://doi.org/10.1017/S1751731115000890

Thorup, V.M., Nielsen, B.L., Robert, P.E., Giger-Reverdin, S., Konka, J., Michie, C., Friggens, N.C., 2016. Lameness affects cow feeding but not rumination behavior as characterized from sensor data. Frontiers in Veterinary Science 3, 37. https://doi.org/10.3389/FVETS.2016.00037/BIBTEX

van Amstel, S.R., Shearer, J., 2008. Manual for Treatment and Control of Lameness in Cattle. Manual for Treatment and Control of Lameness in Cattle 1–212. https://doi.org/10.1002/9780470344576

van de Gucht, T., Saeys, W., van Meensel, J., van Nuffel, A., Vangeyte, J., Lauwers, L., 2018. Farm-specific economic value of automatic lameness detection systems in dairy cattle: From concepts to operational simulations. Journal of Dairy Science 101, 637–648. https://doi.org/10.3168/JDS.2017-12867

van de Gucht, T., Saeys, W., van Nuffel, A., Pluym, L., Piccart, K., Lauwers, L., Vangeyte, J., van Weyenberg, S., 2017a. Farmers' preferences for automatic lameness-detection systems in dairy cattle. Journal of Dairy Science 100, 5746–5757. https://doi.org/10.3168/JDS.2016-12285

van de Gucht, Tim, Saeys, W., van Weyenberg, S., Lauwers, L., Mertens, K., Vandaele, L., Vangeyte, J., van Nuffel, A., 2017. Automatically measured variables related to tenderness of hoof placement and weight distribution are valuable indicators for lameness in dairy cows. Applied Animal Behaviour Science 189, 13–22. https://doi.org/10.1016/J.APPLANIM.2017.01.011

van der Tol, P.P.J., Metz, J.H.M., Noordhuizen-Stassen, E.N., Back, W., Braam, C.R., Weijs, W.A., 2003. The vertical ground reaction force and the pressure distribution on the claws of dairy cows while walking on a flat substrate. Journal of Dairy Science 86, 2875–2883. https://doi.org/10.3168/jds.s0022-0302(03)73884-3

van der Tol, P.P.J., Metz, J.H.M., Noordhuizen-Stassen, E.N., Back, W., Braam, C.R., Weijs, W.A., 2002. The pressure distribution under the bovine claw during square standing on a flat substrate. Journal of Dairy Science 85, 1476–1481. https://doi.org/10.3168/JDS.S0022-0302(02)74216-1

van Hertem, T., Maltz, E., Antler, A., Romanini, C.E.B., Viazzi, S., Bahr, C., Schlageter-Tello, A., Lokhorst, C., Berckmans, D., Halachmi, I., 2013. Lameness detection based on multivariate continuous sensing of milk yield, rumination, and neck activity. Journal of Dairy Science 96, 4286–4298. https://doi.org/10.3168/JDS.2012-6188

van Nuffel, A., Saeys, W., Sonck, B., Vangeyte, J., Mertens, K.C., de Ketelaere, B., van Weyenberg, S., 2015. Variables of gait inconsistency outperform basic gait variables in detecting mildly lame cows. Livestock Science 177, 125–131. https://doi.org/10.1016/J.LIVSCI.2015.04.008

Variable Importance in Random Forests - Code and Stats [WWW Document], n.d. URL https://blog.hwr-berlin.de/codeandstats/variable-importance-in-random-forests/ (accessed 5.30.22).

Velez, D.R., White, B.C., Motsinger, A.A., Bush, W.S., Ritchie, M.D., Williams, S.M., Moore, J.H., 2007. A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. Genet Epidemiol 31, 306–315. https://doi.org/10.1002/GEPI.20211

Viazzi, S., Bahr, C., Schlageter-Tello, A., van Hertem, T., Romanini, C.E.B., Pluk, A., Halachmi, I., Lokhorst, C., Berckmans, D., 2013. Analysis of individual classification of lameness using automatic measurement of back posture in dairy cattle. Journal of Dairy Science 96, 257–266. https://doi.org/10.3168/JDS.2012-5806

Walker, S.L., Smith, R.F., Jones, D.N., Routly, J.E., Morris, M.J., Dobson, H., 2010. The effect of a chronic stressor, lameness, on detailed sexual behaviour and hormonal profiles in milk and plasma of dairy cattle. Reprod Domest Anim 45, 109–117. https://doi.org/10.1111/J.1439-0531.2008.01263.X

Walker, S.L., Smith, R.F., Routly, J.E., Jones, D.N., Morris, M.J., Dobson, H., 2008a. Lameness, Activity Time-Budgets, and Estrus Expression in Dairy Cattle. Journal of Dairy Science 91, 4552–4559. https://doi.org/10.3168/JDS.2008-1048

Weigele, H.C., Gygax, L., Steiner, A., Wechsler, B., Burla, J.B., 2018. Moderate lameness leads to marked behavioral changes in dairy cows. Journal of Dairy Science 101, 2370–2382. https://doi.org/10.3168/JDS.2017-13120

Wells, S.J., Garber, L.P., Wagner, B.A., 1999. Papillomatous digital dermatitis and associated risk factors in US dairy herds. Preventive Veterinary Medicine 38, 11–24. https://doi.org/10.1016/S0167-5877(98)00132-9

Westin, R., Vaughan, A., de Passillé, A.M., DeVries, T.J., Pajor, E.A., Pellerin, D., Siegford, J.M., Vasseur, E., Rushen, J., 2016. Lying times of lactating cows on dairy farms with automatic milking systems and the relation to lameness, leg lesions, and body condition score. Journal of Dairy Science 99, 551–561. https://doi.org/10.3168/JDS.2015-9737

Whay, H., 2002. Locomotion scoring and lameness detection in diary cattle. In Practice 24, 444–449. https://doi.org/10.1136/INPRACT.24.8.444

Whay, H.R., Main, D.C.J., Green, L.E., Webster, A.J.F., 2003a. Assessment of the welfare of dairy cattle using animal-based measurements: Direct observations and investigation of farm records. Veterinary Record 153, 197–202. https://doi.org/10.1136/vr.153.7.197

Whay, H.R., Shearer, J.K., 2017a. The Impact of Lameness on Welfare of the Dairy Cow. Vet Clin North Am Food Anim Pract 33, 153–164. https://doi.org/10.1016/J.CVFA.2017.02.008

Whay, H.R., Shearer, J.K., 2017b. The Impact of Lameness on Welfare of the Dairy Cow. Veterinary Clinics of North America - Food Animal Practice. https://doi.org/10.1016/j.cvfa.2017.02.008

Whay, H.R., Waterman, A.E., Webster, A.J.F., O'Brien, J.K., 1998. The influence of lesion type on the duration of hyperalgesia associated with hindlimb lameness in dairy cattle. The Veterinary Journal 156, 23–29. https://doi.org/10.1016/S1090-0233(98)80058-0

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Willshire, J., Practice, N.B.-C., 2009, undefined, n.d. An economic review of cattle lameness. cabdirect.org.

Wilson, J.P., Green, M.J., Randall, L.V., Rutland, C.S., Bell, N.J., Hemingway-Arnold, H., Thompson, J.S., Bollard, N.J., Huxley, J.N., 2022. Effects of routine treatment with nonsteroidal anti-inflammatory drugs at calving and when lame on the future probability of lameness and culling in dairy cows: A randomized controlled trial. J Dairy Sci. https://doi.org/10.3168/JDS.2021-21329

Winckler, C., Willen, S., 2001. The Reliability and Repeatability of a Lameness Scoring System for Use as an Indicator of Welfare in Dairy Cattle. Acta Agriculturae Scandinavica A: Animal Sciences 51, 103–107. https://doi.org/10.1080/090647001316923162

Wood, S., Lin, Y., Knowles, T.G., Main, D.C.J., 2015. Infrared thermometry for lesion monitoring in cattle lameness. Veterinary Record 176, 308. https://doi.org/10.1136/VR.102571

Youden, W.J., n.d. INDEX FOR RATING DIAGNOSTIC TESTS. https://doi.org/10.1002/1097-0142

Yunta, C., Guasch, I., Bach, A., 2012a. Short communication: Lying behavior of lactating dairy cows is influenced by lameness especially around feeding time. Journal of Dairy Science 95, 6546–6549. https://doi.org/10.3168/JDS.2012-5670