Sensor-based approaches to monitoring the behaviour, health, and welfare of dairy calves

Charles Carslake BA BVSc MRCVS

School of Veterinary Medicine and Science

Thesis submitted to The University of Nottingham
for the degree of Doctor of Philosophy

September 2022

Abstract

Currently, the detection of ill health in UK farmed calves is based on intermittent visual observation which is subjective and poorly accurate. Sensor-based monitoring may offer an improved alternative. For example, sensors could be used to monitor behaviour and detect signs of ill health in calves. However, substantial individual variation exists for many behaviours, the extent of which is poorly understood in calves. Here, within- and between- individual variation in calf feeding behaviours are quantified using data from computerised milk feeders. Results show that substantial, temporally stable individual differences exist. In addition, the average behavioural expression of two distinct feeding behaviours were positively and significantly correlated and the between-individual differences observed were shown to be consistent over time and context, and to be associated with weight gain. This improves our understanding of normal variation in calf feeding behaviour, which could be helpful in detecting potential behavioural changes indicative of ill health. Machine learning models were trained and tested using feeding data from computerised milk feeders to detect ill health. In a separate study, a similar methodology was used to detect ill health using reticulo-rumen temperature boluses. Results indicate low and moderate predictive performance, respectively. Study limitations and areas for future research are discussed. Finally, the development of novel technologies to enable a more holistic approach to behavioural monitoring in calves is explored. Results show that signals from a single collar-based sensor can be used to accurately detect nine different behaviours as well as to quantify rarely occurring behaviours, such as locomotor play. Quantifying play behaviour could provide a useful indicator of positive welfare in calves. It is also shown that these behaviours can be detected using computer vision, but that further work is needed to enable generalisation to new camera angles and scenes. Overall, this thesis highlights the potential of sensor-based technologies to improve our understanding of behavioural variation in calves, as well as to monitor a greatly more diverse range of behaviours than previously attempted. It is hoped that this work will contribute towards the improvement of health and welfare in calves.

Publications and contributions

Chapter 2

A close derivative of this work has been published as: Carslake, C., Occhiuto, F., Vázquez-Diosdado, J.A. and Kaler, J., 2022. Repeatability and Predictability of Calf Feeding Behaviours—Quantifying Between-and Within-Individual Variation for Precision Livestock Farming. Frontiers in Veterinary Science, p.257. https://doi.org/10.3389/fvets.2022.827124

Contributions: J.K., J.A.V.-D., FO, and CC: conceptualization and writing—review and editing. C.C.: formal analysis, data curation, and writing—original draft preparation. J.K.: resources, funding acquisition, and supervision. C.C. and F.O.: data collection. F.O., C.C., and J.K.: project administration.

Chapter 3

A close derivative of this work has been submitted for publication as: Carslake, C., Occhiuto, F., Vázquez-Diosdado, J.A. and Kaler, J. Indication of a personality trait in dairy calves and its link to weight gain through automatically collected feeding behaviours. Submitted to *Scientific Reports* 2022.

Contributions: J.K, J.A.V-D., F.O., and C.C.: conceptualization and writing—review and editing. CC: formal analysis, data curation, and writing—original draft preparation. JK: resources, funding acquisition, and supervision. C.C. and F.O.: data collection. F.O., C.C., and J.K.: project administration.

Chapter 4

Contributions: JK, J.A.V.-D., and CC: conceptualization and writing—review and editing. CC: formal analysis, data curation, and writing—original draft preparation. JK: resources, funding acquisition, and supervision. CC and FO: data collection. FO, CC, and JK: project administration.

Chapter 5

Contributions: J.K., J.A.V.-D., and C.C.: conceptualization and writing—review and editing. C.C.: formal analysis, data curation, and writing—original draft preparation. J.K.: resources, funding acquisition, and supervision. C.C. and FO: data collection. F.O., C.C., and J.K.: project administration.

Chapter 6

A close derivative of this work has been published as: Carslake, C., Vázquez-Diosdado, J.A. and Kaler, J., 2020. Machine learning algorithms to classify and quantify multiple behaviours in dairy calves using a sensor: Moving beyond classification in precision livestock. Sensors, 21(1), p.88. https://doi.org/10.3390/s21010088

Contributions: Conceptualisation, J.K., J.A.V.-D. and C.C.; software, J.A.V.-D.; formal analysis, C.C. and J.A.V.-D.; resources, J.K.; data curation, C.C.; writing—original draft preparation, C.C. and J.A.V.-D.; writing—review and editing, C.C., J.A.V.-D., J.K.; supervision, J.K.; project administration, J.K.; funding acquisition, J.K.

Chapter 7

Contributions: Conceptualisation, J.K., J.A.V.-D., R.H (Robert Hyde) and C.C.; software, R.H.; formal analysis, C.C. and R.H.; resources, J.K.; data curation, C.C. and R.H.; writing—original draft preparation, C.C.—review and editing, C.C., J.A.V.-D., J.K.; supervision, J.K.; project administration, J.K.; funding acquisition, J.K

Acknowledgements

I feel hugely grateful to have had the opportunity to undertake this PhD thesis within the Ruminant Population Heath Research Group at the University of Nottingham. I feel privileged to have worked in such an inspiring and supportive group. Thank you to my funders the Biotechnology and Biological Sciences Research Council for enabling this enriching experience.

Thank you to my supervisory team Professor Jasmeet Kaler and Dr Jorge Vázquez-Diosdado. Throughout, any difficulties - from failing technologies to the covid pandemic - have been met with optimism and a determination to find a way through. Thank you for your dedication, leadership, and mentorship. Thank you to my industry partners PrognostiX for supporting and encouraging my journey from farm animal clinician to research scientist. Thank you also to Francesca, to the many students, and to the farm staff Emma and Nigel for their invaluable support of the trials.

Finally, thank you to my family and to Catrina for all that you do.

Table of Contents

1 Chapter 1 - Introduction

The 2.7 million cattle under one year of age on UK farms suffer from a high burden of disease, which leads to poor health and welfare. One proposed solution is to use new technologies such as sensors and smart systems to monitor behavioural and physiological indicators and thereby provide tools to detect and record health and welfare problems. In doing so these systems, called precision livestock technologies, could allow interventions and management processes to be optimised and so reduce disease burden and improve welfare. However, precision livestock technologies are underutilised in calves, and none have been shown to accurately reflect health and welfare in calves. The core aim of this thesis is to explore ways in which precision livestock technologies may be used to monitor calf behaviour and physiology, and how such data could be used to improve calf health and welfare.

This first introductory chapter has four parts. Firstly, it will highlight the problem of poor health and welfare in calves in the UK. Secondly, it will discuss how monitoring behaviour and body temperature in calves could provide indicators for health and welfare. Thirdly, it will introduce the topic of individual variation and show how this is an important but neglected area for precision livestock technologies. Finally, it shall introduce the technologies and statistical techniques that could enable more successful health and welfare monitoring in calves.

1.1 The health and welfare of calves

1.1.1 The problem

The UK cattle industry produces £9.2 billion worth of milk and £2.8 billion worth of beef annually. 2.7 million of the UK's 10 million cattle are under one year of age (DEFRA, 2022). Youngstock and calves are therefore an integral part of the cattle industry. Youngstock and calf rearing contribute around 20% of whole farm costs on dairy farms and account for a quarter of whole herd related enteric methane emissions (Bell et al., 2011; Heinrichs, 1993). However, they suffer large economic losses to disease and high mortality. For example, perinatal mortality (stillbirths and within 24 hours) in the UK is around 7.9% (Brickell et al., 2009). High mortality persists throughout the calf and youngstock rearing process as evidenced by cattle registration data which shows mortality rates of 6.0% in dairy calves and 3.9% in all calves prior to 3 months of age which is in addition to perinatal mortality (Hyde et al., 2017). In dairy farms, an estimated 14.5% of dairy heifers die or are culled prior to reaching first calving (Brickell et al., 2009). These losses continue into the first lactation, for example data from 437 UK farms indicate a primiparous culling rate of 15.9% (Sherwin et al., 2016).

A major driver of mortality rates and culling is infectious disease. For example, results from post-mortems carried out on calves from Scottish dairy herds between 2014 and 2018 show that of 1,017 diagnoses made, infectious disease was responsible for 69% of all deaths (CHAWG, 2020). Studies conducted on UK farms also indicate a high prevalence of infectious disease, for example results from a longitudinal study which undertook weekly health checks on calves from 11 UK dairy farms show that 48% of all calves were diagnosed with diarrhoea and 45.9% of were diagnosed with respiratory disease which accounted for approximately two thirds of deaths recorded (Johnson et al., 2017). As well as contributing to mortality, infectious disease causes indirect losses. For example, studies in the dairy sector have shown that calves affected by infectious disease have slower growth, are at increased risk of culling, are older at first calving, are less likely to complete their first lactation, and have lower future milk production as adult cattle (Bach, 2011; Schaffer et al., 2016). In the beef sector, infectious disease has a similar detrimental effect with studies reporting lower average daily weight gain, increased time to slaughter, and the production of meat at a lower quality grade (Smith, 2000; Wittum et al., 1996). Whilst a comprehensive assessment of the economic impact of infectious disease in calves and youngstock in the UK is lacking, as evidenced by direct and indirect impacts, its burden is likely to be substantial. For example, one French study indicated that eradicating bovine respiratory disease on young bull and veal enterprises would result in an increase of productivity by 8.7% and 12.8% for these sectors respectively (Delabouglise et al., 2017).

1.1.2 Targeted detection of disease

To improve the management of infectious disease in calves and youngstock one common focus has and continues to be strategies to prevent disease in the first place. Strategies include good colostrum management, better housing and ventilation, minimising weaning stress, minimising mixing of different age groups, vaccination programs and prophylactic use of antimicrobials around periods of high risk (e.g., Gorden & Plummer, 2010). However, data from cattle birth and death registrations shows calf mortality rate between 2011 and 2019 remained essentially unchanged during this period (CHAWG, 2020) indicating that calf mortality remains persistently high. Even more strikingly, one estimate from 1952 put UK calf mortality at 5.2% (Withers, 1952), remarkably close to a more recent study which estimated it to be 5.5% (Hyde et al., 2017). These similar estimates, separated by over 60 years, indicate minimal improvement in calf mortality during this period. Given the relationship between disease and mortality, these results suggest that infectious disease has also remained stubbornly high.

One area that has yet to be adequately addressed is the detection of ill health in calves and young stock. As prey animals, cattle tend to hide clinical symptoms which can make it difficult detect sick animals from visual observation alone (Weary et al., 2009). Furthermore, most stockpersons detect ill health through unstructured observation of clinical signs such as lethargy, coughing, diarrhoea, dyspnoea, inappetence and lethargy (Glover et al. 2017). This is prone to error and complicated by a lack of labour resources for calf monitoring. Agreement and consistency of diagnosis varies between farmers but there is evidence to suggest that the detection of ill health is very poorly sensitive (27 – 62%) and moderately specific (63-100%) (Timsit et al., 2016; White & Renter, 2009; Wittum et al., 1996).

Poor diagnostic accuracy is a problem for several reasons:

- Clinical signs may not be recognised until they are severe, and treatment is less successful than if they were treated earlier (Ferran et al., 2011; Janzen et al., 1984).
- Limitations of current diagnostic strategies contribute to the decision to metaphylactically treat cattle rather than identifying sick animals (Ives & Richeson, 2015) which is an important risk factor for antimicrobial resistance.
- Untreated sick animals are a welfare problem

Better detection of ill health is also important since it could help improve record keeping and data management on farm. Capturing data related to calf health and welfare to benchmark between farms could act as a useful driver in improving calf health. One study of Welsh dairy farms shows that only 39% of farmers surveyed recorded calf disease and veterinary surgeons described records as adequate on 16% of farms for respiratory disease and 8% of farms for diarrhoea (Atkinson, 2015), indicating that record keeping is currently of a low standard. This is a problem since studies show that record keeping and benchmarking mortality and feeding practices against other comparable farms can improve engagement and motivation in calf rearing to address management problems (Atkinson et al., 2017; Sumner et al., 2018). Additionally, an increasing interest in improving decision support and data capture is highlighted by a recent UK industry report which rated these areas as in the top 5 investment priorities for the beef and dairy sectors (CHAWG, 2020).

The next section will discuss health and welfare in calves in more detail. It will focus on two of the most common causes of poor health in calves which are respiratory disease and diarrhoeal disease, before discussing calf welfare.

1.1.3 Common infectious diseases in calves

Diarrhoeal disease

Diarrhoea most commonly affects calves of just a few days to a few weeks of age and is the most important cause of disease in calves less than 30 days (McGuirk & Peek, 2014). For example, one study on UK dairy herds reported that 47% of calves were diagnosed with diarrhoea, the cause of one third of all deaths in pre-weaned calves during the study period (Johnson et al., 2017). Multiple pathogens are implicated in this disease syndrome and infections are often mixed (e.g., Bartels et al., 2010). Pathogens include viruses such as rotavirus and coronavirus, bacteria such as *E. coli* and *Salmonella* species and parasites such as C*ryptosporidium* and C*occidia.* Results from 614 calf post-mortems indicate that 21% and 11% of submissions were affected by cryptosporidium and rotavirus respectively (CHAWG, 2020). Prompt diagnosis of calf diarrhoea is important since it can be rapidly fatal due to dehydration and the resulting acidosis. Diarrhoea is routinely detected by visual inspection of calves and the environment by farm staff. There is evidence to suggest that observing calf cleanliness is less effective at detecting diarrhoea compared to faecal scores (Graham et al., 2018). However, observing faecal scores requires observing each calf defecating or taking rectal samples which is time consuming and may not be routinely undertaken on farms. Once diagnosed, treatment is focused on rehydration therapy to restore electrolyte balance (Roussel & Brumbaugh, 1991). Depending on the causative agent and accompanying clinical signs (e.g., fever), antimicrobial and anti-inflammatory treatment may also be indicated. Prevention strategies focus on reducing environmental contamination through good hygiene and improving calf immunity through colostrum management and reducing environmental stress (e.g., Cho & Yoon, 2014)

Respiratory disease

Bovine respiratory disease (BRD) is endemic in UK cattle farms and is the most common cause of disease in calves over 3 months of age (McGuirk & Peek, 2014). UK cattle industry estimates annual costs of approximately £60 million

(National Animal Disease Information Service, 2015). BRD is a disease syndrome which frequently consists of an initial viral respiratory infection which is followed by secondary bacterial involvement (Snowder et al., 2006; Taylor et al., 2010). Common viral pathogens are bovine respiratory syncytial virus, infectious bovine rhinotracheitis and parainfluenza virus. These viruses cause damage to the upper respiratory tract which predisposes to secondary bacterial infections of the lungs resulting in pneumonia. Common bacteria include *Mannheimia haemolytica*, *Pasteurella multocida* and *Mycoplasma bovis*. As previously discussed, studies suggest that approximately half of all calves may be affected on UK dairy farms, and it is a major driver of on-farm mortality (Johnson et al., 2017). For example, results from post-mortem submissions between 2014-2018 show that approximately one quarter of calves submitted were diagnosed with pneumonia (CHAWG, 2020). BRD is currently diagnosed on-farm mainly using visual assessment. Signs include coughing, nasal and ocular discharge, and depression. However, undetected BRD remains a major challenge for the sector. For example, one meta-analysis of the diagnostic accuracy of visual pen-side checks compared to lung lesions post-mortem indicates that the sensitivity of BRD diagnosis by visual appraisal may be as low as 27% (Timsit et al., 2016). One option to improve the pen side disease diagnosis is to incorporate standardised calf health scoring systems. For example. The Wisconsin Clinical Health Score (WisCHS) assigns a weight to the clinical signs nasal discharge, ocular discharge, head tilt, coughing and rectal temperature. However, this sensitivity and specificity of WisCHS are still moderate, between 46-62.4% and 74.1-91.2% respectively (Buczinski et al, 2015; Love et al, 2016), indicating that improvements to pen side diagnosis are needed. In addition, this method requires time, labour, training, and commitment to implement. For 100 calves, if one full time employee is needed for routine chores, another 0.5 full time employees may be needed for WisCHS health monitoring (McGuirk & Peek, 2014). Treatment for BRD involves the use of an appropriate antibiotic alongside an anti-inflammatory (Scott, 2013).

1.1.4 Welfare

The five freedoms are a commonly used framework through which to assess welfare. The concept of five freedoms to assess animal welfare first originated in the Brambell Report (1965) and was later formalised by the Farm Animal Welfare Council (FAWC, 1979). The five freedoms are freedom from hunger or thirst, freedom from discomfort, freedom from pain, injury or disease, freedom to express natural behaviour and freedom from fear and distress. This framework to assess animal welfare has had a major impact in thinking around animal welfare, has been translated into government policy and law, and has influenced numerous animal welfare organisations worldwide (McCulloch, 2013). In calves, requirements to ensure freedom from hunger and thirst can be achieved by good access to large amounts of milk and ad lib access to water, freedom from pain can be promoted using local anaesthetic and analgesics for husbandry procedures such as castrations and disbudding and freedom from injury and disease can be assessed by recording calf morbidity and mortality.

Criticisms of the five freedoms include their emphasis on suffering and low welfare states. Other definitions of animal welfare have positive aspects, for example good welfare has been defined as ‘fit and feeling good’ (Webster, 2005). Indeed, more recent calls for animal welfare frameworks to evolve beyond the five freedoms have placed a renewed emphasis on positive welfare states (Mellor, 2015). Animals engaged in behaviours they find rewarding could be used to indicate their perception of their external circumstances and assess if they have lives worth living (Mellor, 2016). However, current methodologies and metrics to assess positive welfare states are limited and tend to rely on single timepoint visual observations. Potential indicators of positive welfare in calves could include measures of behaviour such as synchronisation of lying, grooming and play behaviours (Mattiello et al., 2019).

This section has briefly discussed the specific diseases that affect calf health and how current thinking in animal welfare is placing increasing importance on positive experiences. The next section shall discuss how monitoring behaviours and body temperature could provide indictors for calf health and welfare.

1.2 Monitoring behaviour and physiology for calf health and welfare

1.2.1 Sickness behaviour

The relationship between fever and sickness behaviours such as anorexia, depression and inactivity were first proposed by the veterinarian Hart in the 1980s. In a review article he wrote: "…the behavior of sick animals and people is not a maladaptive response or the effect of debilitation, but rather an organized, evolved behavioral strategy to facilitate the role of fever in combating viral and bacterial infections" (Hart, 1988). The insight that sickness behaviour and fever are adaptive was in contrast with previous thinking which positioned these responses as the result of debilitation due to illness. Subsequent research has helped unravel the relationship between infection, fever, and behaviour leading to an improved understanding of how cytokines act on the brain to induce specific physiological and behavioural changes. Sickness behaviour is modulated at a cellular level by inflammatory mediators called cytokines that convey to the brain that an infection has occurred in the periphery (Kent et al., 1992). Sickness behaviours that occur in response to physiological concentrations of cytokines include fatigue, reduced appetite, increased rest, and reductions in social behaviour (Kelley et al., 2003). Alongside sickness behaviours, cytokines also induce an increase in body temperature which is hypothesised to create a hostile environment for invading pathogens and enhance the efficacy of the immune system (Johnson, 2002). Increased heat production and the associated immune response is demanding in terms of energy. By reducing those activities that are not immediately necessary for conserving homeostasis, sickness behaviour is hypothesised to conserve energy to fight infection and therefore promote survival (Dantzer & Kelley, 2007). Livestock keepers have used changes to animal behaviour as an indication of ill health for centuries. More recently, technologies such as video recordings and automated behavioural monitoring have enabled detailed observations of sickness behaviour in calves and other livestock (see review by Weary et al., 2009). The following sections shall discuss specific sickness behaviours and the febrile response in calves and youngstock.

There is some evidence to suggest that calves increase their total lying time as part of sickness behaviour. Increases in total lying time have been reported in calves that were experimentally infected, compared to control calves (Eberhart et al., 2017; Hanzlicek et al., 2010; Hixson et al., 2018; Theurer et al., 2013). However, these differences were short lived lasting for 1-2 days post inoculation (Eberhart et al., 2017; Hixson et al., 2018 Theurer et al., 2013). Another study, where clinical symptoms were mild, reported no difference in lying time (Borderas et al., 2008). There is evidence to suggest that changes to activity whilst lying (e.g., head movements to groom, ruminate, observe surroundings etc) and to activity whilst standing could be an important part of sickness behaviour. Results from studies where sickness behaviours in calves were experimentally induced through inoculation, indicate that inoculated calves increase the proportion of lying and standing time spent non-active (Borderas et al., 2008; Hixson et al., 2018; Theurer et al., 2013). In addition, changes to activity whilst lying and standing persisted for longer post-inoculation than any changes to total lying time. There is also evidence to suggest locomotion reduces in calves as part of sickness behaviour. In one study, calves that were the most severely affected post inoculation with a respiratory pathogen, as measured by examining lung lesions post-mortem, travelled the shortest distance daily (White et al., 2012).

Reduced motivation to eat is an important component of illness behaviour in adult cattle (Beauchemin, 2018). However, in pre-weaned calves, evidence for this finding is mixed with one experimental study which induced mild disease reporting no difference in milk intake between treated and control calves (Borderas et al., 2008), whilst another reported a reduction in milk intake compared to controls for the day of inoculation only (Hixson et al., 2018). One study reported that illness was associated with decreased in milk intake when calves were fed ad-lib fed milk replacer and no association between illness and milk intake when calves were fed a restricted diet (Borderas et al., 2009), indicating that factors other than illness such as hunger may influence motivation to feed during sickness. This suggests that the degree to which feed intake alters during illness behaviour will depend on the cost-benefit trade-off of reducing intake.

Social behaviour could be a useful indicator of sickness behaviour in calves. Social behaviour involves the development of relationships in which animals build up regular patterns with each other and with the environment. There is evidence to suggest social behaviour is altered in sick cows, for example cows which are sick isolate themselves (Proudfoot et al. 2014) and have fewer competitive interactions (Huzzey et al. 2007; Goldhawk et al. 2009; Sepulveda-Varas et al. 2016). Calves inoculated with *Mannhaemia haemolytica* decreased their initiation of social grooming (although they were groomed more often by control calves), decreased social lying bouts, and initiated fewer displacements at the milk feeder (Hixson et al., 2018). In another experimental study, inoculated calves reduced overall social contact frequency (Burke et al., 2022). One study has investigated changes to social behaviour during naturally occurring disease in calves. Calves affected by ill health had reduced social contacts leading to a reduction of the centrality and weight in their corresponding social network (Vasquez-Diosdado et al. 2022 – in press).

Changes to grooming behaviour could be an important indicator of sickness behaviour in calves. Results from three experimental studies indicate substantial decreases in self grooming between infected calves. In one study, control calves self-groomed for over 126 seconds/hour whereas calves infected with *Mannheimia haemolytica* self-groomed for only 3 seconds/hour on the day of inoculation, indicating a 98% reduction ($p < 0.001$) in self-grooming behaviour (Hixson et al., 2018). In a second study where calves were infected with a low dose bacterial endotoxin to induce mild disease, inoculated calves had approximately half the number of self-grooming bouts compared to controls (13.47 ± 1.75 vs. 24.07 ± 3.12; $P < 0.01$) (Borderas et al., 2008). Finally, another study monitored calves' usage of a mechanised grooming brush alongside other behaviours following experimental inoculation with a pathogen. Results indicate that the most severely affected calves (as measured by post-mortem examination) showed a significant difference in grooming and reductions in grooming post inoculation, which were amongst the most persistent changes in behaviour (Toaff-Rosenstein et al. 2016).

Overall, research in calves and youngstock indicate that a wide diversity of behaviours can change as part of sickness behaviour. Results from these studies show that there are substantial differences between behaviours in terms of the magnitude and duration of the changes reported. For example, whilst several studies report that sick calves increased their lying time and decreased feed intake these changes were not consistent across studies, especially where clinical signs were less severe. Where studies were able to measure changes to more subtle behaviours during ill health, such as activity whilst resting, self-grooming and social behaviours, results indicate that reductions to these behaviours are substantial and tend to occur earlier and persist for longer than changes to lying times or feed intake. Sickness behaviour can be viewed as the combination of the interplay between motivation to maintain behaviours that are important to the animal in the short to moderate term and the effect of pro-inflammatory cytokines on the brain that mediate a reduction energy expenditure. Indeed, the hypothesis that behavioural change in response to illness is gradual, with behaviours that are less critical to homeostasis affected first, has been proposed previously in livestock (Weary et al. 2009). Evidence around sickness behaviour in calves is in line with this hypothesis. This indicates that 'less critical' behaviours that are reduced early in the disease process such as activity whilst resting, grooming and locomotion are of particular interest when looking to detect early changes that could be indicative of ill health in calves.

1.2.2 Fever

The pro-inflammatory cytokines that induce sickness behaviour in response to infection are intrinsically related to the generation of increased body temperature known as the febrile response. The febrile response is remarkably conserved across different species. For example, in warm blooded animals, pro-inflammatory cytokines act on the hypothalamus to prevent compensatory mechanisms that regulate temperature (Kluger, 1991) whilst in reptiles, amphibians and fish, which cannot internally raise their body temperature, inflammatory cytokines promote heat seeking behaviour (Boltana et al., 2018; Kluger et al., 1998). The resulting increase in temperature is hypothesized to create a hostile environment for bacteria thereby increasing the effectiveness of the immune system (Johnson, 2002).

Experimental studies which detail sickness behaviour in calves have also monitored internal temperature changes. In one experiment, 15 pre-weaned calves were either subject to a low dose *E. Coli Lipopolysaccharide* (*LPS*) challenge to induce mild sickness behaviour or to saline as a control (Borderas et al., 2008). Rectal temperature was recorded hourly using a handheld digital thermometer. All LPS injected calves showed a significant ($p < 0.05$) increase in rectal temperature compared to control calves post inoculation. The increased temperatures lasted between 2 and 8 hours post inoculation and the maximum temperature recorded was 40.55 +/- 0.13°C. In another experiment, 22 steers were split into a treatment and a control group. To simulate naturally occurring respiratory disease, steers in the treatment group were initially inoculated with a viral pathogen followed by a bacterial pathogen 5 days later (*Histophilus Somni*). Rectal temperature was recorded continuously using an indwelling digital logger from 3 days post viral inoculation (Reuter et al., 2010). On day 3 post the viral challenge, inoculated steers had a mean increase in rectal temperature of 2.1°C ($p < 0.001$) which persisted until day 7, compared to control steers. In naturally occurring disease, increased rectal temperature (typically >39.5°C) is a widely used indicator for infectious diseases in cattle (e.g., metritis, mastitis, respiratory disease) and is a key component of any clinical exam. These results indicate that temperature is a useful indicator of infection in calves.

1.2.3 Play behaviour

Calves display a mixture of locomotor and social play which could provide a useful indicator of positive welfare (Boissy et al 2007). Locomotor play consist of a mixture of running, kicking, jumping, and bucking (Jensen et al. 1998) (Jensen and Kyhn, 2000) whereas social play involves two calves butting head against head/neck in a playful manner or one calf mounting another calf's head or body (Jenson 2000). Burghardt (2005) defined play behaviour based on a series of characteristics amongst which are the notions that play is not 'fully functional', i.e., that play behaviour does not contribute to current survival. Instead play behaviour is thought to occur because the display of play behaviour itself, rather than the consequences of the display, has rewarding psychological properties which brings welfare benefits (Held et al. 2011).

There is some evidence to suggest that play behaviour is initiated when there are no threats to the animal's fitness. For example, play behaviour may decrease and can disappear completely in injured, unhealthy animals, and re-emerge as they recover (Fagen 1981). For instance, castration depresses play behaviour for at least 3 days in 1-week-old lambs (Thornton & Waterman-Pearson 2002). In calves, reduced locomotor play has been associated with disbudding and reduced feed allowance (Krachun et al. 2010) (Rushen and de Passillé, 2012). Play behaviour is also affected by social contact, for example in one study calves housed individually played less than those housed socially (Duve et al. 2012).

However, despite evidence that positive welfare and play behaviour are related, play behaviour can reflect previous deprivations in animals. A period of social isolation in rats is known to increase rebound play activity when presented with a companion (Holloway & Suter 2004). In calves, increased locomotor play was recorded in a spatially confined group when exposed to an open area compared to unconfined controls (Jensen 1999). Conversely, play behaviour may increase in calves in some situations where welfare could be judged to be reduced, such as when there is a reduction in maternal nourishment (Held et al. 2001). In addition, play behaviour in calves has mainly been monitored at the group level and only over short periods. Further research is needed to monitor play over longer periods to investigate how consistent play is over time, how it might change as calves get older and how individuals may differ in terms of their propensity to play.

1.3 Individual variation in behaviour

Identifying behaviours that are reliable indicators of health and welfare is complicated by the fact individual animals behave remarkably differently from one another. In fact, animals have been referred to as CITD (complex, individual, time-variant and dynamic) systems (Berckmans, 2017). Whilst comparing population distributions and their means is a fundamental part of statistical testing, few individuals meet the mean of their respective group (Réale et al., 2007). Instead, a 'per animal approach' is required to monitor individuals - what is normal for one individual may be abnormal for another. Individualised monitoring is particularly important for studies that seek to infer health and welfare by use of sensor-based monitoring, since failure to correctly define such thresholds may result in animals which are sick not being detected (false negatives) or animals which are healthy being unnecessarily detected (false positives). Improving our understanding of the existence and extent of normal behavioural variation is therefore an important first step in the development of systems that can use behaviour to infer health and welfare in livestock.

1.3.1 Measures of individual variation

Whilst few studies have investigated individual variation in livestock by use of directly measured behaviours, in the related discipline of behavioural ecology, monitoring behavioural variation in populations of animals is gaining significant interest (e.g., Hertel et al., 2020, 2021). One development which is driving this trend is the use of sensor technologies to capture repeated measures of behaviour. These technologies can include wearables such as accelerometer sensors or location sensors which capture data regarding movement and behaviour (Brown et al., 2013; Cooke et al., 2004). Using technology enables researchers to study behavioural variation in potentially large numbers of multiple individuals over extended durations. Alongside technology options for monitoring behaviour, statistical techniques which quantify and decompose phenotypic variation are increasingly used (Dingemanse et al., 2010; Dingemanse & Dochtermann, 2013; Westneat et al., 2015). Concepts which quantify and qualify individual variation in behaviour are derived from linear mixed models where the identity of each individual animal is included as a random effect, observations correspond to repeated measures of individuals, and potential confounding factors such as differences in age or sex are controlled for by including them as fixed effects. Different measures of behavioural variation can be extracted from these models.

Concepts and measures of behavioural variation include:

- Behavioural type: This corresponds to an individual's value intercept in a random intercept model and quantifies an individual's average behavioural expression.
- Predictability: Residual within individual variation after controlling for differences in average behaviour.
- Plasticity: an individual's change in behaviour over an environmental gradient (e.g., time) which corresponds to the gradient of slope from a random slope model.
- Behavioural syndrome: Within-individual correlation between an individual's average expression (behavioural types) for two or more behaviours.
- Repeatability: proportion of the variation attributed to variation among individuals. This is most frequently estimated as the intra-class correlation coefficient (ICC).
- Coefficient of variation in predictability: Between individual differences in residual within-individual variation after controlling for differences in average behaviour.

Behavioural type, predictability, plasticity, and behavioural syndrome can be represented graphically as in Figure 1.1.

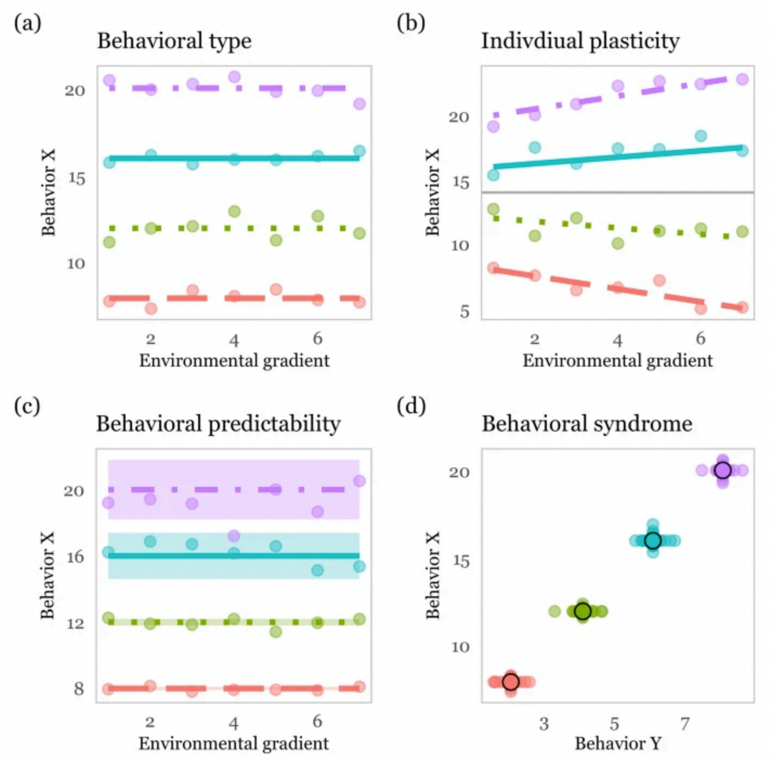


Figure 1.1. Schematic representation of concepts used to measure behavioural variation for four different individuals, each represented by a different colour, from (Hertel et al., 2020). a. Behavioural types – individuals differ in their mean behavioural expression over repeated measures. b. individuals differ in their plasticity (slope) along an environmental gradient. c. Predictability – individuals differ in their within-individual variation around their behavioural types. d. Behavioural syndrome: there is a positive correlation for individuals' behavioural types for the two distinct behaviours X and Y. Those that have higher behavioural types for behaviour X also tend to have higher behavioural types for behaviour Y.

That individuals differ is an important consideration for technologies and systems that seek to detect abnormalities. In addition to this consideration, there is increasing evidence that behavioural differences in livestock have consequences for individuals' themselves (i.e., behavioural differences have implications for production, growth, health, and welfare). This emerging field is called animal personality.

1.3.2 Personality in livestock

Personality in humans and animals can be defined as the combination of a set of personality traits. These personality traits drive temporally and contextually consistent between-individual differences in behaviour (Kaiser & Müller, 2021). In humans, the five-factor model, developed by Goldberg and others during research in the 1980s and 90s, is an extensively studied framework in psychology which combines five traits to describe and measure individual differences in personality (Goldberg, 1990; Gosling et al., 2003). The 5 traits are conscientiousness, agreeableness, extraversion, neuroticism and openness. Together these traits are thought to capture most differences in human personality and to remain relatively stable over time. These traits have been shown to be associated with a range of outcomes such as job performance, longevity, and mental health.

In farm animals, there is growing evidence that personality exits and can be described using similar traits to the five-factor model in humans (Finkemeier et al., 2018). Proposed traits include boldness, aggressiveness, activity, sociality, and exploration as represented by Figure 1.2.

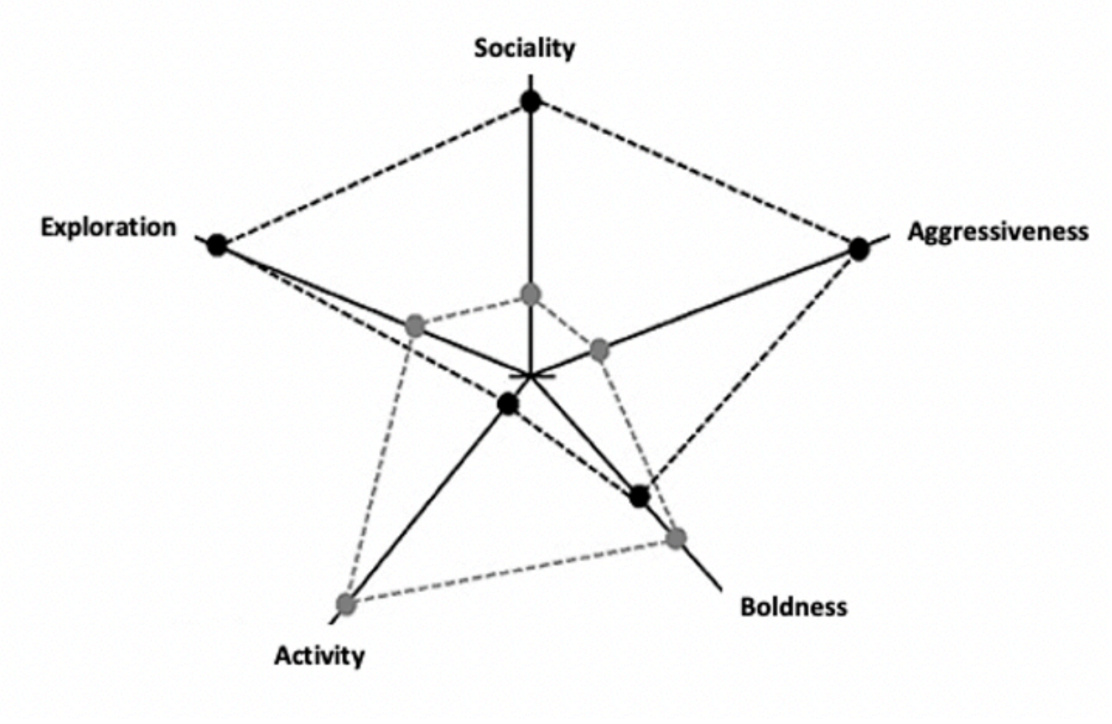


Figure 1.2. Hypothetical presentations of five personality traits in animals. Adapted from Finkemeier et al. (2018). Two individuals with different personalities are represented, one by the grey and one by the black lines. The individual represented by the grey line scores higher for activity, boldness and lower for sociality, aggressiveness and exploration compared to the individual represented by the black lines.

In addition, an increasing number of studies indicate that personality in farm animals is related with underlying physiology, health, and production outcomes. For example, one study showed that nervous cattle had significantly higher cortisol levels, differing immunological measures, lower average live weight gain and higher morbidity than calm cattle (Fell et al. 1999). Cortisol levels in lambs exposed to a new environment correlate with behavioural traits such as fearfulness (Rice et al. 2016). Bulls with more temperamental traits (aggressiveness, volatility) had the highest resting levels of cortisol and epinephrine and lowest peak rectal temperature and sickness behaviour scores following an LPS injection (Burdick et al. 2011). In calves, there is evidence to suggest that personality traits are related to the stress response. In one study, stable individual behavioural traits 'fearfulness' and 'pessimism' were positively associated with the emotional response to transportation (as measured by maximum eye temperature and number of vocalisations) (Lecorps et al. 2018). In another study, adreno-cortical indicators of stress and the personality trait 'fearfulness' were correlated (Van Reenen et al., 2005). These results suggest that prior knowledge regarding individuals' personality could be useful to predict animals' responsiveness to stress and future productivity which may have implications for welfare and management. For example, measuring personality traits could help identify individuals which may be suited to a particular environment and identify those individuals which may struggle to cope (Dawkins, 2004; Wechsler, 1995).

In livestock, personality traits are usually assessed by conducting several behavioural tests, often conducted at least twice during an animal's lifetime. Cross context correlation and consistency are then measured, and when these are consistent, they are said to measure personality (Finkemeier et al. 2018). However, these behavioural tests are time consuming to perform and impractical beyond research studies research which limits their application in a commercial setting. A different approach to measuring personality traits could be to use sensors that measure behaviour (e.g., activity) to quantify individual behavioural differences. If these sensor-based behavioural measures are consistently different between individuals, they could provide a novel methodology to measure certain personality traits in the farm environment.

1.4 Precision livestock technologies to monitor behaviour and body temperature

Sensors that can monitor behaviour and temperature in dairy cattle are increasingly adopted throughout the industry (e.g., Mottram, 2016). Areas of interest for this thesis include the use of movement sensors and computer vision to detect behaviour and temperature sensors to monitor body temperature. The following section shall firstly detail how data from movement sensors such as accelerometers can be used to monitor different behaviours. Secondly it shall describe what behaviours can currently be monitored by use of movement sensors. Thirdly it shall describe how computer vision can be used to monitor behaviours. Finally, it shall describe different options for temperature monitoring in calves.

1.4.1 Using movement sensors data to record behaviour

Accelerometers are small, robust sensors which make use of the piezoelectric effect to monitor movement. The piezoelectric effect occurs when a microscopic crystal structure generates a voltage when mechanically stressed from pressure or vibration. Accelerometers use this effect to translate the force generated by a change in motion into a voltage. The resulting voltage can be expressed as an amplitude in different dimensions, thus providing a record of the corresponding force’s direction and magnitude (Vijaya, 2012).

Over the past decade, numerous studies have used accelerometer-based sensors to monitor different behaviour in livestock. The raw data captured by accelerometers contains patterns that are characteristic of the behaviour record.

To create an algorithm capable of translating raw data into a prediction of a particular behavioural several steps are required. These steps are outlined below:

1. Choice of sensor: Most studies that monitor animal behaviour use a three-dimensional accelerometer which records the amplitude of motion in three orthogonal axes (X, Y and Z). However, studies have also used two-dimensional accelerometers (e.g., White et al., 2008) or have additionally incorporated other sensors such as location sensors (e.g., Occhiuto et al., 2022) and gyroscope sensors which record orientation and angular velocity (e.g., Walton et al., 2018).
2. Point of attachment for the sensor: Points of attachment are typically the leg, the collar, ear tag or less commonly the halter. The point of attachment can have an important impact on the behaviours that can be monitored. For example, leg-based sensors generally represent standing and walking with high accuracy (Robert et al., 2009), whilst collar-based sensors are better able to detect feeding behaviours and rumination (Vázquez Diosdado et al., 2015).
3. Sampling frequency: There is evidence to suggest that reducing sampling frequency decreases the accuracy at which behaviours are detected in livestock (Walton et al., 2018). In sheep, more complex behaviours (such as grazing) were most affected by reducing sampling frequency and less complex behaviours (such as posture) least affected (Walton et al., 2018). However, sampling frequency directly affects power consumption and therefore battery life which is an important consideration for practical applications (Khan et al., 2016).
4. Behavioural observations: once sensors are deployed on the livestock, behaviours need to be observed and recorded according to a pre-defined ethogram. An ethogram is a comprehensive list and description of the behaviour of an organism which is used to study animal behaviour over time. Typically, behaviours are recorded by use of video cameras allowing time for detailed labelling of observed behaviours. Behavioural observations are then merged with sensor data according to the identity of the animal and time stamp of the observation.

5. Window size: most studies group consecutive measurements of raw acceleration data into short segments referred to as windows (Bersch et al., 2014). Window length is typically expressed in seconds with longer windows able to capture more complex signal patterns in the data but may be less able to identify behaviours that occur quickly (Bersch et al., 2014).
6. Statistical features: from each window, a range of statistical features are calculated (Figo et al., 2010; Preece et al., 2009). These can be simple such as the mean, maximum and standard deviation of the acceleration magnitude. Data from each window can also be transformed into a frequency domain representation using the fast Fourier transformation and used to calculate frequency domain features (Dargie, 2009).
7. Algorithm training and testing: predicting behaviour from the calculated features requires the implementation of an algorithm which can be used to infer behaviour. The algorithm can be as simple as an acceleration threshold for differentiating lying from standing using a leg-based accelerometer (Darr & Epperson, 2009) but more commonly involves a supervised machine learning algorithm. Previously used classification algorithms include decision trees, support vector machines and random forests (e.g., Martiskainen et al., 2009; Vázquez Diosdado et al., 2015; White et al., 2008) .

Together, these steps are used by studies to monitor a diverse range of behaviours in livestock. The following sections summarises the literature relevant to the behavioural monitoring in livestock by use of accelerometers in order provide a comprehensive overview of current capabilities and highlight some limitations.

1.4.2 Current examples of movements sensors

Data from leg-based accelerometers can accurately identify lying and standing with excellent accuracy in calves and youngstock and is one of the most widely used ways of remotely monitoring behaviour in livestock. For example, Robert et al., (2009) developed a classification algorithm using data collected from a three-dimensional accelerometer attached to the hind limbs of 15 crossbred calves to correctly predict standing 98% of the time and lying 99.2% of the time when compared to a human observer. Similarly, Darr & Epperson (2009) reported leg based single axis accelerometer could identify 100% lying events using a simple threshold. Over the past decade, several commercial devices have been validated for monitoring lying and standing behaviour in livestock (Bonk et al., 2013; Ledgerwood et al., 2010; Swartz et al., 2016; Trénel et al., 2009). In addition to lying and standing behaviour there is evidence to suggest that accelerometers can be used to identify subtle changes in resting behaviour in livestock. For example, one study in calves correctly classified 90% of sleeping time and distinguished between non rapid eye movement and rapid eye movement using collar-based accelerometers (Hokkanen et al., 2011). In cows, one study also used a collar-based accelerometer to distinguish between resting behaviours. 'Awake' (standing, head up, alert, eyes open) was identified with 93.7% accuracy and 'sleep-like' (lying, still, head resting on ground, eyes closed) with 92.2% accuracy (Klefot et al., 2016).

Studies using accelerometers to monitor walking report variable accuracy. For example, results from one study show that a leg-based accelerometer predicted walking with 67.8% accuracy in beef cattle (Robert et al., 2009). Another study in dairy cattle reported 79% sensitivity and 79% specificity for walking behaviour (Martiskainen et al., 2009). Moderate to poor accuracy for walking has been reported for commercial sensors, results from one study indicate that a commercial leg-based accelerometer was unable to accurately represent walking behaviour in calves (Trénel et al., 2009). Another approach has been to incorporate features that are less affected by high frequency movements with one study in sheep indicating that features such as movement variation to identify walking using collar, leg, and ear-based accelerometers with high accuracy (Barwick et al., 2018; Walton et al., 2018). Identifying gaits other than walking such as locomotor play has received some attention in calves due to increasing interest in using this behaviour as an indicator of positive welfare. Results from two studies indicate that acceleration signals from accelerometers differ for walking, trotting and locomotor play (de Passillé et al., 2010; Luu et al., 2013). More recently, two studies have explored the use of accelerometers to detect locomotor play in calves. In one study, instances of play were overestimated by 200% possibly because of the low prevalence of this behaviour (Größbacher et al., 2020). Results from another study, which used a commercially available accelerometer, indicate that it is possible to detect whether play was present or absent within a 1min or 15min sampling period (Gladden et al., 2020). This method would therefore not be able to measure the duration or number of play instances within each sampling period. It is therefore unlikely to be a suitable measure to quantify play behaviour.

There is evidence to suggest that feeding and ruminating behaviour can be identified by collar and ear-based accelerometers. In dairy cows, Vázquez Diosdado et al., (2015) classified feeding with 98% sensitivity and 93.1% precision using a collar-based accelerometer. High accuracy for monitoring feeding behaviours has also been reported for commercial accelerometers. For example, commercial ear-based sensor was able to reasonably identify ruminating and eating with kappa values of 0.85 and 0.77 respectively (Bikker et al., 2014). There is also evidence to suggest that collar-based accelerometers can be used to detect milk suckling with high accuracy reported in beef calves (Kour et al., 2018) and in lambs (Kuźnicka & Gburzyński, 2017).

Whilst the number of studies which use sensors to monitor behaviour in livestock is growing, most of these only accurately record basic posture and feeding behaviours. In addition, ear or collar mounted sensors can monitor feeding behaviours and rumination adequately but tend to poorly represent posture and activity (Martiskainen et al., 2009; Vázquez Diosdado et al., 2015). Expanding the use to behavioural monitoring technologies to detect behaviours such as activity as well as resting behaviours, self-grooming and locomotor play could provide researchers with new methodologies and could potentially provide insight into more subtle behaviours, opening the door to more comprehensive and accurate methodologies for assessing health and welfare. Another limitation is that most studies have focused on classifying behaviours (i.e., predicting when a behaviour occurs). However, since different behaviours occur at different rates, there is need for robust methodologies that can quantify behaviours even when they occur rarely. Otherwise, activities that have low or very low prevalence (e.g., under 1%) may vastly overestimate activity budgets, such as has been reported for play behaviour (Größbacher et al., 2020) and walking in calves (Trénel et al., 2009). This overestimation occurs because the positive predictive value of classification algorithms reduces with prevalence. As such, there is a need to develop methodologies that can accurately quantify behaviours that occur rarely.

1.4.3 Computerised milk feeders

Computerised milk feeders are an example of a technology which is already being used on commercial farms in the UK. These milk feeders are designed to feed milk to calves housed in small groups with minimal intervention from the farmer. They can reduce labour requirements and allow milk to be delivered at higher volumes and more regularly throughout the day than might be practical with manual feeding (Kung et al. 1997; Medrano-Galarxa et al. 2017). In addition to reducing labour requirements, the data recorded by automatic feeders provides interesting information regarding calf feeding behaviour and can be used to monitor calves throughout the milk feeding period. Data recorded includes the identity of each calf, the timing and duration of each visit, the amount of milk consumed, drinking speed, the number of rewarded visits (where the calf visits the milk feeder and is entitled to a milk feed) and number unrewarded visits (where the calf visits the feeder when is it is not entitled to a milk feed). These feeding behaviours may be useful for monitoring calf health and welfare. For example, unrewarded visits have been suggested as an indicator of hunger (De Paula Vieira et al., 2008; Jensen & Holm, 2003). Some feeders are additionally equipped with partial weight scales which allows calf body weight to be recorded (Cantor et al. 2021). Data captured by these feeders has been used to inform different milk feeding regimes in calves, such as documenting the benefits of feeding larger amounts of milk which include improved weight gain and reduced unrewarded visits (Jensen & Holm, 2003; Rosenberger et al., 2017). Studies have also used data from feeders to show that gradual weaning regimes result in less of a sharp decrease in the rate of weight gain compared to abrupt weaning (M. A. Khan et al., 2007; Roth et al., 2008). Behaviour from feeders may also be used as indicators of ill health, for example there is evidence to suggest that behaviours such as rewarded visits are reduced in sick compared to healthy calves (e.g., Johnston et al., 2016; Svensson & Jensen, 2007).

1.4.4 Computer vision

Computer vision is the use of computers to extract information from an image. It involves developing artificial systems to handle diverse problems using image processing and analysis techniques. One technique of interest is deep learning which uses multiple layers of artificial neural networks. Neural networks are highly flexible computing systems inspired by the structure of the biological neurones (LeCun et al., 2015). Signals tend to travel from the first layer to the last, with each layer contributing another level of abstraction allowing high level features to be extracted from the raw input (Gu et al., 2018). For example, if the raw input is a matrix of pixels representing a calf, the first layer might detect edges around the calf's body, the second may extracts the arrangement of the edges (i.e., how they fit together), the third layer might detect a head or a tail from the edges and the final layer may recognise that the image contains a calf. With this technique, such a process is not programmed but learnt through a process of trial and error with minimal human supervision (Gu et al., 2018; Maier et al., 2019). Recently, there has been significant interest in applying computer vision techniques in livestock to detect behaviour (Oliveira et al., 2021; Chen et al., 2021). There is evidence to suggest that multiple behaviours can be classified with high accuracy. For example, Fuentes et al., (2020) detected 15 different behaviours in cattle from CCTV video footage of adult beef cattle housed in a group pen. These included detailed individual behaviours such as self-grooming, resting, sleeping, social behaviours such as social licking and aggressive interactions, as well as the recognition of individual body parts to detect tail movement and ruminating. Results show high accuracy indicating that detailed behavioural monitoring of cattle behaviour at a group level is possible. However, thus far, no studies have used computer vision to monitor detailed behaviours in group housed calves.

1.4.5 Temperature monitoring

Several technologies exist that are capable of continuously monitoring temperature in livestock. These include intra-ocular temperature probes, thermal imaging, and intra-ruminal temperature boluses. Whilst these technologies are promising they require further development and research before they can deployed on farm. These technologies and their limitations are briefly outlined below.

Ear canal temperature (within the external auditory meatus) can be monitored in cattle by use of a temperature sensor contained within an ear tag and an associated probe which is inserted into the external auditory meatus. Ear canal temperature is affected by external ambient temperature indicating that ear temperature recordings should be tailored to external temperature to improve correlations with core body temperature (McCorkell et al., 2014). However, this technology is limited by practical concerns such as reported the difficulty with placing the tags, the potential for dislodgement and an increased risk of tympanic infection (Davis et al., 2003; Mahendran et al., 2017; McCorkell et al., 2014).

Thermography has a been proposed as a promising, non-invasive method of monitoring temperature in calves. Eye temperature has been proposed as a promising proxy for core body temperature, with studies indicating that eye temperature measurements have lower variability compared to other body regions such as the flank or back, ears and nose (Schaefer et al., 2004). There is evidence to suggest that eye temperature increases in line with core body temperature in response to infection. For example, BVDV infected calves increased their mean eye temperature by 2.6 C as measured by a thermal camera compared to controls (Schaefer et al., 2004). However, most studies have used thermography to monitor body temperature over short periods only. Longer duration monitoring requires placement of a thermal camera at a location that is frequently used by an animal and integration of an animal identification protocol (for example using an RFID) at this location. In addition, there are other practical and technical considerations to overcome include automating image segmentation to detect regions of interest, correcting for camera drift and ensuring a continuously clean environment to minimise measurement error (Wirthgen et al., 2011).

Intra-reticular sensors take the form of a bolus that is administered orally (i.e., swallowed and passed through the oesophagus) and then resides in the rumen. Of all proposed methods, rumen boluses are perhaps the most practical for on farm usage since, compared to thermography and intra-ocular sensors they involve minimal infrastructure and come with minimal risk of infection or inflammation. The bolus diameter should be sufficiently narrow to pass through the oesophagus without obstruction (e.g., < 20mm for pre-weaned calves) and the specific gravity over 3 allowing it to remain in the reticulum without passage further down the digestive tract or regurgitation (Ghirardi et al., 2006). Intra-reticular temperature sensors are equipped with a transponder and antennae which enable data transmission via radio transmission to a receiver and data logger (AlZahal et al., 2011). The correlation between intra-ruminal temperature and rectal temperature varies between studies and appears to be highest when temperature is elevated. For example, Sievers et al., (2004) showed a high correlation between rumen and rectal temperatures (r = 0.92, n = 36), whereas Bewley et al. (2008) undertook a greater number of observations and observed a lower correlation (r = 0.64, n = 2042). The lower correlation reported by Bewley et al. (2008) could be explained by variation in feed and water intake as well as season, milking, housing system and parity. Feed intake may influence reticular temperature as a higher concentrate diet can lead to a lower ruminal pH which is associated with a higher rumen temperature (>40C) (AlZahal et al., 2011). However, whilst diet may affect temperature recordings, there is there some evidence to suggest that generalised pyrexia induces an increase in rumen temperature above the expected effect of diet alone. For example, results from one study which examined the effect of diet (low or moderate roughage) on LPS induced pyrexia, as measured by an intra-ruminal temperature bolus, show that differences in diet did not impede the increase in temperature (AlZahal et al., 2011). Water intake also influences rumen temperature causing sudden drops in temperature upon drinking. These drops can be used to quantify drinking in cattle (Vázquez-Diosdado et al., 2019a) and should be removed prior to calculating rumen temperature. There is evidence to suggest that reticulum and rectal temperature are highly correlated in calves that are affected by bovine respiratory disease. One study reported high correlation of (r = 0.91, n = 135) in steers diagnosed with naturally occurring bovine respiratory disease (Timsit et al., 2011a). Results from an experimental study also indicate that rectal and rumen temperatures in calves inoculated with *Mannheimia haemolytica* are highly correlated ($r = 0.80$) (Rose-Dye et al., 2011). Further work is needed to assess how temperature gathered from reticular boluses may be used to detect respiratory disease in calves and youngstock.

1.5 Machine learning for health and welfare monitoring

As previously outlined, calves are a high disease risk group where improved and more timely disease detection would have a significant beneficial effect on calf health and welfare and could provide better data to support decisions on farm. Thanks to increasingly sophisticated technologies, it is becoming possible to monitor a wide range of behaviours as well as body temperature in calves and youngstock (Costa et al., 2021). These technologies could be used to identify health and welfare problems early. Whilst some of these technologies such as accelerometers, automatic milk feeders and reticular temperature boluses are commercially available, it is not yet known which systems are the most appropriate for monitoring the health and welfare calves or whether the data they provide can be used to provide accurate predictions for health and welfare monitoring.

Machine learning can be thought of as a way creating predictive algorithm by learning from data without being explicitly programmed (Niemann, 2013). It is a flexible approach capable of combining data from multiple sources of information to make a prediction. For these reasons it is receiving increasing attention to produce insights for livestock health and welfare (García et al., 2020). In supervised machine learning, the algorithm is trained on data where the outcome class or prediction is known (James et al., 2013). The learning process consists of trial and error where a prediction is made using prior knowledge (if any) about the relationship between the predictors and the label. Next, the error is computed and used to update the prior knowledge. These steps are repeated resulting in a process where prior knowledge is updated with each iteration (Niemann, 2013).

Machine learning algorithms are flexible tools that can handle data from multiple sources and can be used in large and complex datasets. Algorithms used in machine learning, such as decision trees and K-nearest neighbour, are frequently non-linear, this means that they can exploit patterns and associations in data that are not easily analysed using conventional techniques (Rebala et al., 2019). For these reasons machine learning is very suited to solving complex challenges and could be a suitable option to make sense of data gathered from multiple precision livestock technologies (Hudson et al., 2018; Neethirajan, 2020; Slob et al., 2021). Examples in adult dairy cattle include the use of data from sensors and machine learning to predict diseases such as mastitis (Dhoble et al., 2019), lameness (Taneja et al., 2020) and to predict calving time (Keceli et al., 2020) .

Before an algorithm can be used as a decision support tool for livestock management there is a need to validate it. In the case where an algorithm is attempted to predict a class outcome (i.e., a classification problem), k-fold cross-validation is a frequently used methodology to evaluate model performance. It helps assess how good the model is at correctly predicting the outcome class based on the predictor variables (James et al., 2013). The process of cross-validation is represented by Figure 1.3. In the original dataset, the class outcome (also called label) is known for each observation. The parameter k specifies the number of subsets that the original dataset will be split in to. From these k datasets, k-1 are used to train the classifier and the last (validation set) is used to test. During the testing the predicted outcome is compared with the true outcome. This process is repeated k times until all subsets have been used for testing. The results of cross-validation can be summarised by calculating the mean of the model performance scores across each of the k datasets. Typically, 5 or 10 are chosen as the value of k (James et al., 2013).

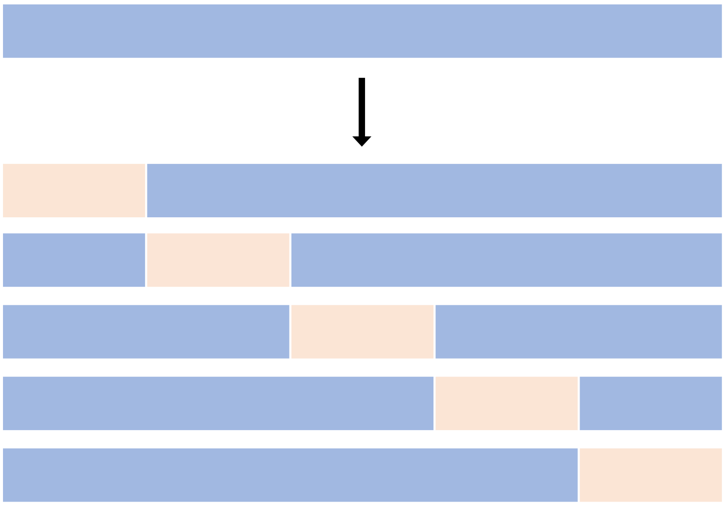


Figure.1.3. A schematic showing the process of 5-fold cross-validation. A set of observations are split into 5 non-overlapping groups. Each of these is used in turn as a validation set (shown in beige) whilst and the remainder (shown in blue) is used as a training set. The overall test error is estimated by averaging the five resulting test error estimates, adapted from James et al. (2013).

At each fold, predictions of the test dataset are compared with ground truth as shown in Table 1.1 which enable the representations of predictions and labels by use of a confusion matrix. From the confusion matrix different performance metrics can be calculated (James et al., 2013). These are represented in Table 1.2.

Table 1.1. Confusion matrix which shows the possible results when applying a classifier or diagnostic test to a population.

		True class		
		Negative	Positive	Total
Predicted class	Negative	True Negative (TN)	False Negative (FN)	N*
	Positive	False Positive (FP)	True Positive (TP)	P*
	Total	N	P	

Table 1.2. Measures to evaluate algorithm performance

Name	Definition
Specificity	TN/N
Sensitivity	TP/P
Negative predictive value	TN/N*
Positive predictive value	TP/P*

Predictive algorithms need to strike a balance between sensitivity and specificity. The trade-off is represented in a graphical way using the receiver operating characteristic (ROC) curve which shows the relationship between sensitivity and specificity for different cut-offs (fig. 1.4). In clinical testing, the choice of cut-off will depend on the scenario and the impact of falsely diagnosing negative cases as positive, the impact of missing positive cases as well as the availability of follow up checks or diagnostics to confirm or refute a prediction from a diagnostic test.

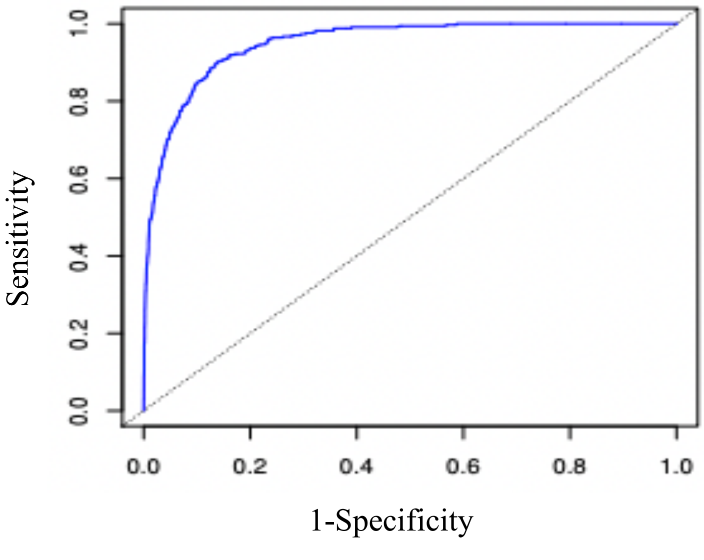


Figure 1.4. Example of a ROC curve. The perfect ROC curve hugs the top left corner of the graph, which would indicate a high sensitivity and a high specificity. The dotted line represents the “no information rate” which can be expected if there is no association in the data between features and the predicted classes. Adapted from James et al., (2013).

To be applied in the field of precision livestock, sensitivity needs to be high enough to provide additional benefits compared to visual appraisal. Specificity also needs to be high since from a practical perspective those animals that are falsely flagged as positive may need to be checked by the farmer which can result in a significantly increased workload (Eckelkamp & Bewley, 2020). For example, if an algorithm has a 90% specificity, each time it is used to predict, 10% of negative cases will be predicted as positive. High specificity has been proposed as important factor in determining farmer engagement with precision livestock technologies (Eckelkamp & Bewley, 2020). This is particularly true when algorithms are used to predict on a regular basis (e.g., daily) as a high number of false alerts can rapidly accumulate.

In conclusion, whilst machine learning algorithms have excellent potential to combine data from diverse sources and to be used on farms to detect health and welfare, their overall performance needs to be high. This highlights the need to target changes in behaviour and physiology that are the most informative of health and welfare as well as the need to control for potentially confounding sources of behavioural variation.

1.6 Summary and plan for thesis

There is an urgent need to improve the detection of infectious disease in calves and youngstock. Both diarrhoeal and respiratory disease cause substantial production losses and welfare concerns but may not be detected if visual observation alone is used to identify sick animals. Technologies exist that monitor a range of behaviours such as activity, posture, and feeding. However, there is substantial between- and within-individual variation in behaviour, the extent, and implications of which are poorly understood. If such monitoring is to be used to measure health and welfare it is vital that we first improve our understanding of this variation. For example, there is evidence to suggest that consistent between-individual variation has implications for production, health, and welfare. In addition, individual variation needs to be accounted for to enable accurate inferences regarding health and welfare. In addition, many of the most promising indicators of health and welfare in calves, such as play, grooming and resting, are not currently captured by sensors.

New research developments could help address these issues. New sensor technologies and methods could improve our ability to monitor a wider range of behavioural and physiological indicators and of health and welfare. New statistical techniques and terminology from disciplines such as behavioural ecology could improve our ability to unravel the complexities surrounding behavioural variation. Finally, methods such as machine learning have the potential to harness data from diverse indicators which could improve the detection of ill health in calves.

The overall aim of this thesis is to explore ways in which precision livestock technologies may be used to monitor behaviour, health, and welfare in calves.

To help achieve this aim, three themes are explored:

- Applying statistical methods from behavioural ecology to improve our understanding of behavioural variation and test the contextual as well as temporal consistency of these behaviours.
- Evaluating the use of machine learning to predict respiratory disease in calves using currently available technologies
- Developing novel behavioural monitoring technologies to evaluate if it is possible to monitor diverse behaviours that are indicative of calf health and welfare.

This thesis benefited from two currently available precision livestock technologies which were used to collect data; computerised milk feeders and reticulo-rumen temperature boluses. In addition, novel technologies to monitor more diverse behaviours were developed as part of this thesis. To explore these themes, this thesis is composed of six different studies. Each study corresponds to a chapter in this thesis and is introduced and discussed in turn.

Chapters 2 and 3 of the thesis focus on applying methods to quantify individual variation in behaviour using data feeding collected in calves from commercially available computerised milk feeders. Whilst numerous studies have used feeding data to explore different questions, the existence and implications of individual variation remain under researched. In Chapter 2, we firstly test if different feeding behaviours are repeatable which is indicative of consistent inter-individual differences in behaviour. Secondly, we examine the predictability of feeding behaviours and test if certain individuals have more predictable feeding behaviours than others. Finally, we test for correlations in measures of behavioural variation. In Chapter 3, we test if the between individual differences observed in chapter two are contextually and temporally consistent and correlated with weight gain.

In Chapters 4 and 5, machine learning algorithms are used to predict the health status in calves and youngstock using data collected from computerised milk feeders and from reticulo-rumen temperature boluses. In chapter 4, a machine learning approach is implemented to test if calves' feeding behaviour can be used to predict respiratory disease. We create features that quantify within individual change in our machine learning model to determine if changes in feeding behaviour are reliable indicators of health status. In Chapter 5, reticulo-rumen temperature sensors are used in steers to test if statistical features calculated from reticulo-rumen temperature records can be used to predict health status. These studies are amongst the first to use machine learning to predict health status in calves and youngstock.

Chapters 6 and 7 focus on the use of technologies to monitor complex behaviours in calves. In chapter six, the core objective is to develop and expand on the precision livestock literature by demonstrating that is possible to classify a diverse range of behaviours in dairy calves with high accuracy using a single movement sensor. Specifically, we develop an algorithm to simultaneously identify calf posture as well as seven other behaviours, chosen for their importance to calf health and welfare, using signals from a collar worn sensor. Furthermore, we implement a machine learning quantification algorithm to test if the prevalence of locomotor play behaviour can be estimated despite the fact this behaviour occurs rarely which makes it difficult to quantify using traditional classification approaches. In Chapter 7 the aim to identify many of those same behaviours as chapter six but use computer vision rather than a physical sensor. We train and test a deep learning object detection algorithm and evaluate its performance to detect multiple behaviours in calves at a group level. These studies are the first to test if such diverse behaviours indicative of calf health and welfare can be monitored using precision livestock technologies.

2 Chapter 2 - Repeatability and predictability of calf feeding behaviours – quantifying between and within individual variation for precision livestock farming

A close derivative of this chapter has been published in *Frontiers in Veterinary Science* (Carslake et al., 2022).

2.1 Introduction

Repeated measures of livestock behaviour are currently available to researchers and farmers thanks to the increasing sophistication and availability of sensor technologies (Berckmans, 2014; Brown et al., 2013; Cooke et al., 2004). This has opened opportunities to continuously observe and analyse behaviour at the level of individual farm animals. Such individualized monitoring can improve management, for example by improving heat detection in dairy cattle (Mottram, 2016). In addition, numerous technologies are being developed that may assist in detecting ill health livestock by detecting changes in behaviour that precede or occur alongside clinical disease (e.g., Barker et al., 2018; Matthews et al., 2016; Schillings et al., 2021). However, different individuals tend to behave differently. Individuals may differ in their average behavioural expression (e.g., can be more or less active) or may differ in the degree to which their behaviour varies around their respective means (Dingemanse et al., 2010; Stamps et al., 2012). Where present, this variability in behaviour may have important implications for behavioural monitoring since failure to account for normal, intrinsic variation (i.e., treating all animals the same) could result in false inferences, mislabelling and hence impede successful algorithm development. In addition, the underlying behavioural tendencies that drive contextually and temporally consistent between-individual differences is termed animal personality (Réale et al., 2007). Differences in farm animal personality have implications for health, welfare and productivity (Finkemeier et al., 2018; Koolhaas & Van Reenen, 2016; Proudfoot et al., 2012). Improving our understanding of the existence and extent of individual variation in behaviour is an essential first step to enable researchers to exploit the potential of individualized behavioural monitoring.

The advancement of methods and measures in behavioural ecology offers a valuable opportunity to assess individual differences in behavioural expression (e.g., Cleasby et al., 2015; Hertel et al., 2020, 2021; Houslay & Wilson, 2017). These methods use repeated measures of animal behaviour in a multilevel modelling framework allowing the calculation of behavioural measures that are statistically defined (Dingemanse & Dochtermann, 2013). For example, an individual's behavioural type corresponds to the value of its random intercept and respectively the individual's position upon a behavioural spectrum (Hertel et al., 2020). The measure repeatability indicates the proportion of within and between individual variation in a behaviour that can be explained by differences between individuals (Bell et al., 2009). Another measure is predictability, which quantifies the degree to which behavioural observations vary around an individual's average behaviour or behavioural type (Stamps et al., 2012). This approach is especially relevant for high-precision observations with low measurement error, where residual variation is assumed to be mainly systematic and biologically meaningful. Multivariate mixed modelling approaches build on this approach by allowing behavioural types and predictability, as well as correlations between these measures, to be estimated within a single statistical framework (Houslay & Wilson, 2017). Where correlations between these exist, these are termed behavioural syndromes the uncovering of which can improve our understanding of how different behaviours are related (Sih et al., 2004).

Monitoring feeding behaviour in livestock is important for assessing productivity and evaluating health and welfare (Weary et al., 2009). Within dairy farms, the increasing use of computerized milk feeders for pre-weaned calves means that a wealth of data detailing calf feeding behaviour is readily available. A large body of literature exists that harnesses these data to explore a variety of questions, such as the effect of different feeding regimes, changes in behaviour that occur prior to ill health and differences in feeding behaviour between personality types (e.g.,Neave et al., 2018; Rosenberger et al., 2017; Svensson & Jensen, 2007). While significant between-individual variation in feeding behaviour has been reported in calves (Appleby et al., 2001; Miller-Cushon et al., 2013), no studies have employed a quantitative approach to explore within and between individual variability. Quantifying the repeatability of different feeding behaviours in calves could assist in characterizing consistent inter-individual differences in behavioural variation. Quantifying the coefficient of variation in predictability for each behaviour could help identify behaviours for which calves differ in their residual intra-individual variation around their behavioural type. Correlating measures of behavioural variation among individuals gives us new insights into how these measures may be related. Such a quantitative analysis could form the basis of future work exploring the phenotypes of calves based on their feeding behaviour.

In this study, repeated measures of feeding behaviour are obtained from a computer-controlled automatic milk feeder and used to quantify individual differences in feeding behaviour in pre-weaned dairy calves using a multivariate multilevel modelling approach. Firstly, individual variation in the average expression of behavioural traits is quantified by calculating the behavioural type for each calf and repeatability for different feeding behaviours. Secondly, differences in predictability are quantified by calculating residual intraindividual variation (rIIV) for each individual and the coefficient of variation in predictability for each behaviour. Finally, we report the correlations between behavioural types and rIIV estimates.

2.2 Materials and methods

2.2.1 Data collection

Calf recruitment

The study was conducted at the Centre for Dairy Science Innovation at the University of Nottingham, UK. All calves enrolled in the study were born at the farm between 26/03/2021 and 29/08/2021. 64 calves were enrolled in the study; all calves were Holstein Friesian and female. This study used data collected by a computerized milk feeder (Forster-Technik Compact Smart) during the first 35 days of a period of group housing which took place as part of routine calf management. There were 16 calves per group and data was collected for four groups (cohorts) of calves.

Housing

Calf housing consisted of two stages: first, a period of pair housing (i.e., two calves per pen) followed by a period of group housing (16 calves per pen). At the first stage, calves were removed from the dam within 4 hours of birth and housed in a straw-bedded pen (3m x 2m) in pairs (the two calves closest in age were paired together). Each pair had access to a feeding station which was equipped with a teat and operated by a computerized milk feeder. The feeding stations were approximately 1m x 0.5m and equipped with sides but there was no back gate to prevent displacements at the feeder. Each computerized milk feeder operated four feeding stations.

The second stage of housing commenced once 8 pairs of calves (i.e., 16 calves) reached a minimum of 21 days old. These calves were then grouped together in a large straw bedded pen (6m x 12m). Throughout this group housing period each group of 16 calves had access to a single milk feeding station (i.e., one teat per 16 calves). The feeding station was not equipped with a back gate and there was one computerized feeder which operated two feeding stations (i.e., it fed 32 calves split between two pens). Data collection for this study took place during the first 35 days of this second stage (i.e., the period of group housing) only.

Feeding and colostrum

Within 2 hours of birth, calves were fed four litres of pasteurized colostrum as per farm protocols. All colostrum is checked for quality using a colostrometer (a hydrometer that estimates IgG density by measuring colostrum density). The colostrum protocol is routinely evaluated by screening subsets of calves for failure of passive transfer by checking total serum protein levels using a refractometer (failure of passive transfer is defined as serum total protein levels < 5.0g/dL). Post colostrum, calves are fed pasteurized transition milk (4 liters twice daily) from a bucket equipped with a teat until two days of age. At two days of age, farm staff gently guided calves towards the feeding station present in the pen where the calves were shown the teat and fed a milk allowance by the computerized feeder. From this point, the calves learnt, with occasional assistance by farm staff, to feed from the teat present in the feeding station.

Each calf was equipped with an RFID ear tag and each feeding station is equipped with an RFID reader. This allows the computerized milk feeder to recognize the identity of each calf when present at the feeder, mix a portion of milk replacer from milk powder and warm water (130g/litre), and dispense the calf a milk allowance. All calves were fed the same milk replacer (Milkivit Energizer ECM, Trouw Nutrition GB) for the entire milk-feeding period. Upon recognition of the calf's RFID ear tag at the station, if the calf is due a milk feed it is dispensed a maximum of 2 litres per feed. These feeds are spaced evenly throughout the day. If the calf does not drink the whole 2-litre entitlement, the remaining is kept available. Once the calf has drank the allowance (called entitlement), the next entitlement will not be dispensed for a minimum of two hours.

The computerized feeder was preprogramed to allocate each calf a total daily milk allowance which renewed from 00:00:00 every morning. The daily allowance fed by the computerized feeder started at 6 litres at 2 days old and increased daily in line with age, reaching 8 litres at 5 days old. From 8 days old, the daily allowance increased daily reaching a plateau of 10 litres from 40 days old. During the second stage (i.e., group housing), all calves, regardless of age, were fed 10 litres daily for the 35 days following the move to the group pen. After this, the allowance was reduced by 400mls/day. This meant that 25 days later (i.e., after 60 days in the group pen) the milk allowance was reduced to zero. Calves had ad libitum access to concentrates (FiMLAC Sweet Start Pellets), chopped straw and water throughout.

Health monitoring and vaccines

A veterinary surgeon manually inspected all calves twice weekly for any signs of ill health using the Wisconsin calf health scoring system (McGuirk & Peek, 2014). This system combines rectal temperature and weighted scores of clinical signs (i.e., nasal discharge) to detect ill health in calves. In addition, farm staff visually inspected calves twice daily. Any calves with signs of ill health were treated according to the farm protocols and advice from the farm's veterinary surgeon (e.g., anti-biotic and anti-inflammatory for respiratory disease, oral rehydration therapy for mild diarrheal disease). Calves were vaccinated with a vaccine against respiratory disease (Rispoval RS+Pi3 IntraNasal; Zoetis) at 9 days of age.

2.2.2 Data acquisition and selection

The computerized milk feeder used in this study logs each visit a calf makes to the feeder on a software program. Data recorded by the feeder include calf identity, date, time the calf entered the feeder, time the calf left the feeder, if the calf was entitled to a milk feed, feed consumption, and feeding rate for each visit. A new visit (row) was created whenever the RFID reader loses and then regains contact with a RFID tag.

Data from the computerized milk feeder for the group-housed calves were downloaded and combined. The first 2 days of group housing were excluded to allow a period of acclimatization to the new environment and feeder. The subsequent 33 days of group housing were included in this analysis to ensure that all calves were on a level feeding plane prior to any reduction in milk allowance. We excluded all calves that were categorized as sick by our health scoring from our analysis (Wisconsin score > = 5; n = 16) to ensure no clinically diseased calves were included in our study (Table 1.) (McGuirk & Peek, 2014). Due to technical problems (failure to save data onto SD card) and management procedures (e.g., cleaning of pens) a maximum of 10 days and a minimum of 3 days were excluded for each group. The remaining data corresponded to 57 196 rows.

2.2.3 Data processing

All pre-processing and analysis were undertaken in R software (version 4.1.0) (R Core Team, 2021). We grouped visits to the feeder by the same calf that were closely clustered in time (<100s) into a single meal. Meal based estimates better characterizes calf feeding behaviour compared to studying visits alone, since they allow the calculation of interesting characteristics such as whether each meal was associated with a milk feed, whether the calf was entitled to milk feed, and total feed consumption during each meal. Clustering visits into meals requires the usage of a meal criterion. A meal criterion corresponds to a maximum time interval between the end of the same calf's visit to the feeder and the start of its next visit, to consider these visits as part of the same meal (Howie et al., 2009; Tolkamp et al., 1998). We used a simple method previously described in adult cattle where, for each calf, the interval between consecutive visits to the feeder is calculated, and its log-transformed distribution plotted (Tolkamp & Kyriazakis, 1999). Visual inspection revealed three distributions with intersections of approximately 100 seconds and 1600 seconds. Since we are interested in the quantifying returns to the feeder that could occur within the longer interval period of 1600 seconds (approximately 26 min), we chose the shorter period of 100 seconds as our meal criterion. This process is detailed in appendix 10.1.

Table 2.1. Number of calves included in analysis per cohort as per the inclusion criteria detailed in the methods and age of calves at start of data collection.

	Cohort 1	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Overall**
Calves					
Calves excluded	4	6	3	3	16
Calves included	12	10	13	13	48
Age at trial start (days) – included calves only					
Mean	34.4	31.8	41.0	39.9	40.0
Min	21	23	26	26	21
Max	48	39	54	50	54

2.2.4 Feeding behaviours

For each calf and for each day of the group-housing period, variables to describe the calves’ feeding behaviour are calculated. These are detailed in Table 2.2. During this stage of the pre-processing a small number (n = 4) of non-sensical recordings (assumed measurement errors) were excluded.

Table 2.2. Definition of feeding behaviours used.

Feeding behaviour	**Definition**
Total meals (number per day)	Daily sum of all meals. This variable includes meals where the calf is entitled to a milk feed and meals where the calf is not entitled to a milk feed.
Meal size	Mean daily meal size calculated from meals where the calf is entitled to milk and consumes a milk feed within the same meal. It corresponds to the quantity of milk consumed divided by the number of these meals.
Feeding rate (ml/min)	Mean daily feeding rate. The feeding rate for each visit where the calf is entitled to a milk feed and consumes milk is calculated by the feeder. From this, the mean feeding rate is calculated.

2.2.5 Statistical analysis

All statistical analyses were carried out using R software (version 4.1.0) (R Core Team, 2021). Code for the figures was adapted from Hertel et al. (Hertel et al., 2020, 2021).

Multivariate Double Hierarchical Generalized Linear Model

We ran a single multivariate double-hierarchical generalized linear model (DHGLM) with the three feeding behaviours (feeding rate, total meals, meal size) as response variables using the "brms" package in R (Bürkner, 2018). A DHGLM was chosen as it includes two parts; a 'mean' and a 'dispersion' part. The mean part of the model is focused on the estimation of individuals' means whilst the dispersion part is concerned with modelling the residual variance (i.e., the variation around the mean). The model can be written as equation 3-4 for mean and equation 5-8 for the dispersion part of the models (Cleasby et al., 2015).

$$\boldsymbol{Y}_i = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{\varepsilon} \quad (3)$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \boldsymbol{I}_m \sigma_\alpha^2) \quad (4)$$

$$\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, Diag\{\boldsymbol{\sigma}_{\varepsilon\}}^2\right) \quad (5)$$

$\log(\boldsymbol{\sigma}_{\varepsilon}) = \boldsymbol{\eta}_{\boldsymbol{d}}$, (6)

$\boldsymbol{\eta}_{\boldsymbol{d}} = \boldsymbol{X}_{\boldsymbol{d}}\boldsymbol{\beta}_{\boldsymbol{d}} + \boldsymbol{Z}_{\boldsymbol{d}}\boldsymbol{\alpha}_{\boldsymbol{d}}$ (7)

$\boldsymbol{\alpha}_{\boldsymbol{d}} \sim N\left(\mathbf{0}, \boldsymbol{I}_m \omega_{\sigma_d}^2\right)$ (8)

In the model α represents individual-specific random effect variation, Y_i represents the response variables (feeding rate, total meals and meal size), X represents the fixed effects, Z the random effects, residual deviations from the prediction are represented by ε. Terms X_d represent the fixed effects for the dispersion part of the model, $Z_d\alpha_d$ the random effect component of the dispersion and ω_σ^2 represent the dispersion model hyperparameter. Between individual random effect of variance (α) is assumed normally distributed as well as α_d and individual-specific residual standard deviations (σ_ε^2) are assumed to follow a log-normal distribution. For both the mean and dispersion parts of the model, age at grouping, day number and their interaction term are included as fixed effects. Individual Calf ID and cohort are included as random effects. All distributions were specified as Gaussian. To capture a Gaussian posterior distribution we log transformed the variable total meals. Meal size was transformed using an ordered quantile normalization which was selected using the 'bestNormalize' package in R (Peterson & Cavanaugh, 2020). Feeding rate was normally distributed. All variables were scaled after transformation (mean = 0; SD = 1).

We used uninformative priors for both fixed and random effects. We ran four chains for 12,000 iterations, a warmup of 4,000 iterations and a thinning interval of 4. Model diagnostics indicated satisfactory convergence with $\hat{R} < 1.01$ and effective sample sizes > 400. Posterior predictive checks indicated that the underlying Gaussian distribution was satisfactorily captured.

Repeatability estimates

Repeatability $\boldsymbol{Rpt}$ was defined as the variance among group means (i.e., the variance explained by differences between individual calves) $\boldsymbol{V}_{\boldsymbol{CalfID}}$ over the sum of the variance explained by differences between individual calves and the residual variance $\boldsymbol{V}_{\boldsymbol{residual}}$ that reflects the variance within individuals. In our model $\boldsymbol{V}_{\boldsymbol{residual}}$ corresponds to the population intercept of the residual model and was converted to a variance by taking its exponent and squaring the resulting value (Hertel et al., 2021).

$$\boldsymbol{Rpt} = V_{\boldsymbol{CalfID}}/(V_{\boldsymbol{CalfID}} + V_{\boldsymbol{residual}})$$

$\boldsymbol{Rpt}$ values are between 0 and 1. Higher values of $\boldsymbol{Rpt}$ for a behaviour indicate the population is composed of individuals that behave consistently differently from each other whereas low values indicate that individuals are more similar. We describe our results for $\boldsymbol{Rpt}$ as higher or lower with reference to a meta-analysis which summarized 759 estimates of repeatability from 114 studies and indicated that the mean level of repeatability was 0.37 [0.36 – 0.38] (A. M. Bell et al., 2009).

Coefficient of variation in predictability

For each behaviour, the dispersion part is used to estimate the residual intra-individual variation (rIIV), after controlling for fixed effects, for each individual calf (Cleasby et al., 2015). Calves with higher rIIV are less predictable (i.e., have greater variation around their means) than calves with lower rIIV. For each behaviour, between calf variation in rIIV was quantified by calculating the coefficient of variation in predictability $\boldsymbol{CV_P}$. This measure quantifies the population-level variation in predictability (Cleasby et al., 2015). In the equation below, the term $\boldsymbol{\omega^2}$ represents the dispersion model hyperparameter (the estimate for individual differences in residual variance) which can be extracted from the DHGLM.

$$\boldsymbol{CV_P} = \sqrt{}(\mathbf{exp}\ (\boldsymbol{\omega^2} - \mathbf{1})$$

$\boldsymbol{CV_P}$ values are between 0 and 1. Higher values of $\boldsymbol{CV_P}$ for a behavior indicate the population is composed of individuals that vary in their rIIV (i.e., a mixture of predictable and unpredictable individuals), whereas a lower value for $\boldsymbol{CV_P}$ indicates that individuals express similar levels of behavioural variation around their respective behavioural types. We describe our results for $\boldsymbol{CV_P}$ as higher or lower with reference to a meta-analysis which summarized 64 estimates of $\boldsymbol{CV_P}$

from 39 studies indicated that behavioural traits had mean $\boldsymbol{CV_P}$ of 0.27 [0.22, 0.33] (Mitchell et al., 2021).

Correlation between variance components

In addition to calculating behavioural types (the mean behaviour after controlling for fixed effects) and the rIIV for each behaviour the multivariate DHGLM computes the correlations between these estimates. Since these behavioural types and rIIVs are estimated in the same framework, any uncertainty around estimates of the mean is carried forward into the correlations between these. This approach allows us to test for correlations between behavioural types and predictability estimates (i.e., feeding rate behavioural type and total meals behavioural type) whilst avoiding the potential pitfall of inflated p-values that can occur when uncertainty around model estimates is ignored (Houslay & Wilson, 2017).

2.3 Results

Calf feeding behaviours have different repeatability

The degree to which individuals differ from each other, as a proportion of within and between individual variation, varied by behaviour and is reported in Table 2.3. Table 2.3 reports repeatability after controlling for the effect of age, day number and cohort (adjusted repeatability) of the feeding behaviours monitored in this study. Repeatability was highest for feeding rate followed by total meals. Repeatability was considerably lower for meal size.

Table 2.3. Mean, median, inter-quartile range, repeatability, and coefficient of variation in predictability for total meals, feeding rate and meal size. IQR and CrI correspond to interquartile range and credibility interval respectively.

	Feeding rate (ml/min)	**Meal size (ml)**	**Total meals (n)**
Mean	831	1989	10.1
Median	835	2045	9
IQR	770 - 901	1995 – 2054	7 - 12
Repeatability			
Estimates	0.50	0.03	0.42
CrI	0.32 - 0.68	0.00 - 0.06	0.30 - 0.55
Coefficient of variation in predictability			
Estimates	0.27	0.07	0.13
CrI	0.21 - 0.37	0.00 - 0.13	0.06 - 0.21

Figure 2.1 illustrates the concept that different behaviours have different repeatability by plotting behavioural type estimates on a spectrum from low to high. Behaviours are Z-transformed (mean = 0, SD = 1) to facilitate comparison. As a proportion of total behavioural variation, behaviours with high repeatability (e.g., feeding rate) had greater between individual differences in behavioural type than behaviours with low repeatability (e.g., meal size).

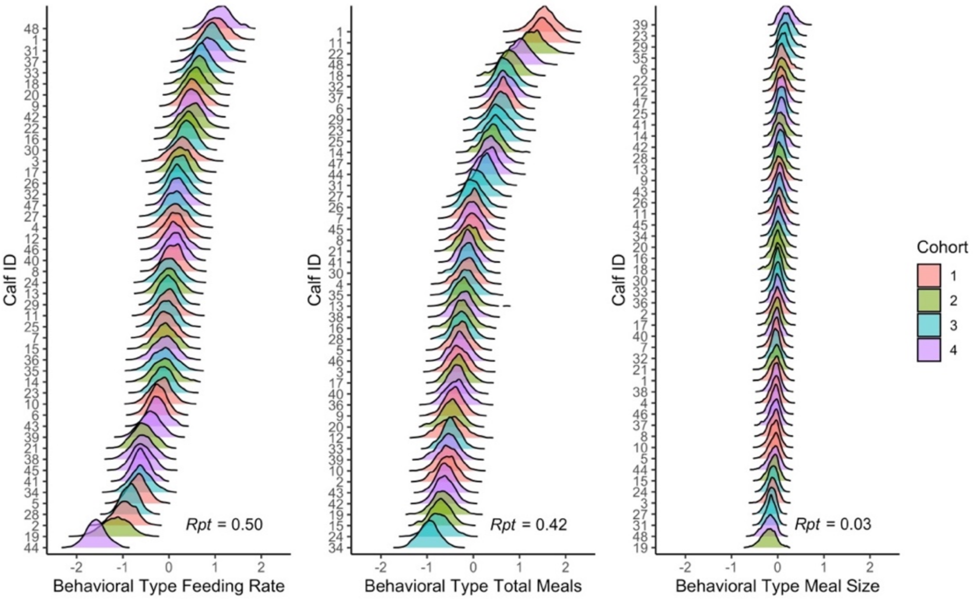


Figure 2.1. Posterior distributions of behavioural types from the double hierarchical mixed models for feeding rate, total meals and meal size for each calf. The models controlled for between individual differences in age, day number and cohort. Variables are Z-transformed (mean = 0, SD = 1). The ridges indicate the posterior 95% credible interval, and the different colours correspond to the different cohorts. Repeatability (Rpt) is reported in the bottom right corner of each panel.

Calf feeding behaviours have different coefficients of variation in predictability

For each behaviour the degree to which calves differ in terms of their predictability, i.e., the degree to which individual calves differ in their residual intra-individual variation around their respective means, is quantified by the coefficient of variation in predictability in Table 2.3. The coefficient of variation in predictability was highest for feeding rate.

Less predictable individuals have high variance around their respective behavioural types (high residual intraindividual variation), while more predictable individuals have low residual intraindividual variation (rIIV). Figure 2.2 illustrates the concept that different behaviours have different coefficients of variation in predictability by plotting rIIV estimates on a spectrum from low to high. Behaviours are Z-transformed (mean = 0, SD = 1) to facilitate comparison. Behaviours with higher coefficients of predictability (e.g., feeding rate) have greater between individual differences in rIIV than behaviours with lower coefficients of predictability (e.g., total meals and meal size).

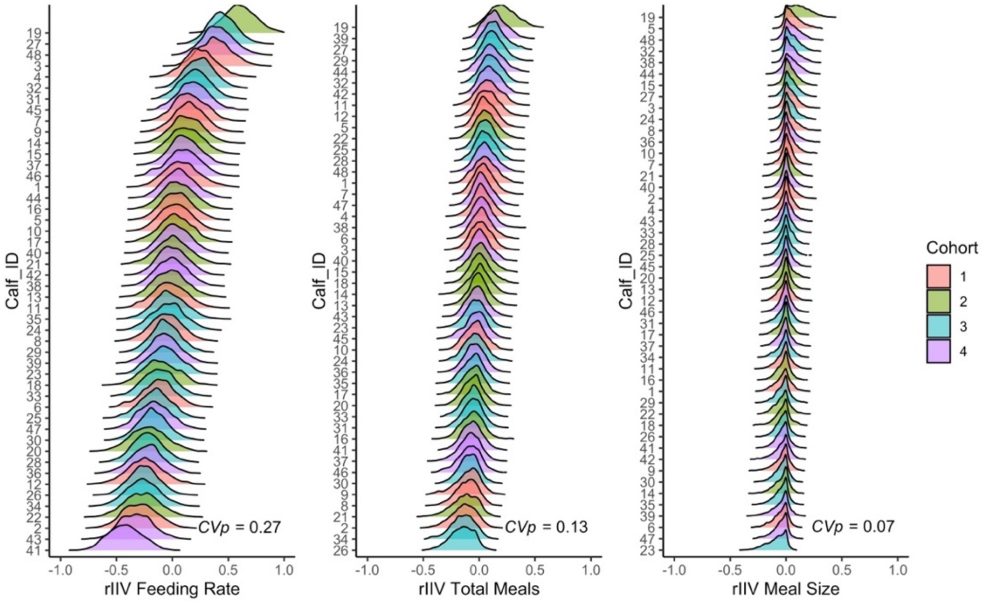


Figure 2.2. Posterior distributions of residual intraindividual variation (rIIV) from the double hierarchical mixed models for feeding rate, total meals and meal size for each calf. The models controlled for between individual differences in age, day number and cohort. Variables were Z-transformed (mean = 0, SD = 1). The ridges indicate the posterior 95% credible interval, and the different colours correspond to the different cohorts. The coefficient of predictability (CVp) is reported in the bottom right corner of each panel.

Calves' feeding rate behavioural type was correlated with their total meals behavioural type.

Estimates of our multivariable double hierarchical mixed model indicate that there was a significant positive linear correlation (r = 0.29 [0.00 – 0.54]) between individual calves' behavioural types for feeding rate and total meals as is shown in Figure 2.3. This result shows that calves that drank faster had more meals, and calves that drank slower had fewer meals. No significant correlations were present between individual calves' predictability (rIIV) and behavioural types for the other behaviours.

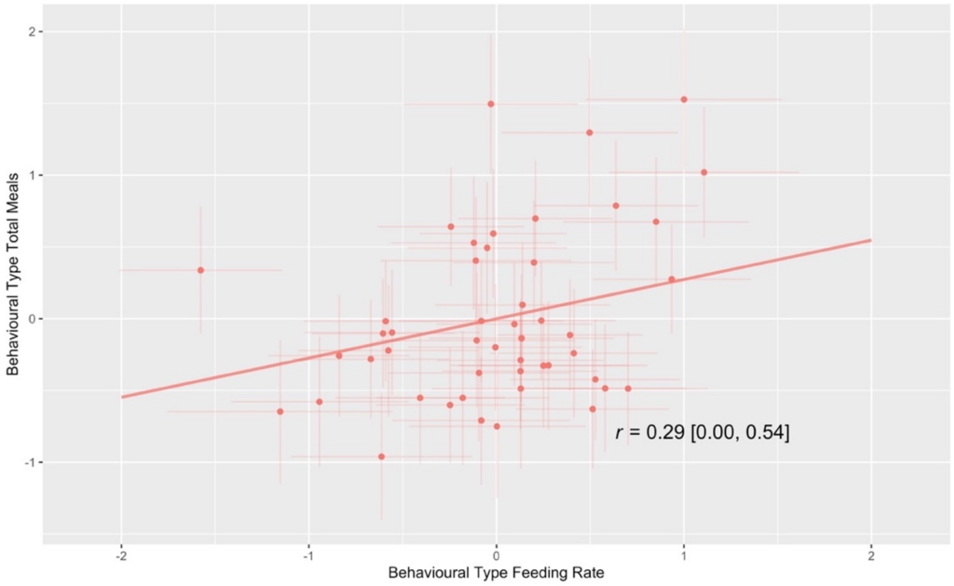


Figure 2.3. Visual representation of the line of best fit of the among individual correlation (r) between feeding rate behavioural type and total meals behavioural type. Posterior means and 95% credible intervals of point estimates for each calf are shown.

2.4 Discussion

Our study is the first to quantify individual differences in calf feeding behaviour at the between and within individual levels. Behavioural type refers to individual's average behavioural expression and predictability refers to its within-individual variation around its behavioural type. For each behaviour, individual calves can be situated both on a spectrum of behavioural types (low to high) and a spectrum of predictability (low to high). The repeatability (the degree to which these individual differences in behavioural type explained the total variation) varied by behaviour and was greatest for feeding rate and lowest for meal size. The coefficient of variation in predictability (the degree to which individuals differed from each other in their predictability) was greatest for feeding rate. Interestingly, for feeding rate and total meals, our results revealed a within-individual correlation between behavioural types, suggesting a behavioural syndrome for this behaviour. This result indicates that calves may be placed on an underlying axis with calves that drink more quickly and visit the milk feeder more frequently at one end to calves that drink more slowly and visit the feeder less frequently at the other.

Repeatability was highest for feeding rate indicating that calves had consistently different feeding rates. High levels of between-individual variation in feeding rate have been noted in calves (Appleby et al., 2001), and between-individual variation in feeding rate has been examined in rats (Whishaw et al., 1992), pigs (B. L. Nielsen et al., 1995) and goats (Cellier et al., 2021; Giger-Reverdin et al., 2020). Our results support the idea that feeding rate is a robust feature of the individual in a variety of species (B. L. Nielsen, 1999). The degree to which calves differed from each other in their predictability, quantified by the coefficient of variation in predictability, was also greatest for feeding rate. Our results suggest that calves with predictable and calves with unpredictable feeding rates coexist.

The existence of between individual variation in behavioural types and in predictability has implications for studies that aim to detect ill calves by use of their feeding behaviour. The fact that calves differ in their behavioural type for some feeding behaviours (i.e., feeding rate, total meals) means that algorithms aiming to detect abnormalities using this behaviour will need to account for different behavioural types for each calf. This can be achieved with approaches such as cumulative sum, which aims to detect abrupt change from an individualized baseline (Knauer et al., 2018; Quimby et al., 2001). However, our results also suggest that for behaviours with relatively higher coefficients of variation in predictability, such as feeding rate in our study, it may also be necessary to allow different individuals differing levels of variation around the mean. This may be necessary to avoid flagging unpredictable but otherwise healthy individuals as abnormal.

We can also consider how these results may relate to the study of animal personality. Animal personality is defined as underlying behavioural tendencies that drive contextually and temporally consistent differences between individuals' behavioural expression (Dingemanse & Réale, 2005; Réale et al., 2007). The study of animal personality is based on the characterization of between individual behavioural variation and behaviours with high repeatability, such as feeding rate in our study, are particularly useful to the study of animal personality (A. M. Bell et al., 2009). Indeed, one study in calves found that individuals which had lower feeding rates were slower to interact with a novel object (more fearful) (Neave et al., 2019). Furthermore, in other species, an individual's boldness has been linked with its predictability (Stamps et al., 2012). Bolder and more risk-taking individuals are more predictable, possibly, because they are less likely to change their behaviour in response to micro-environmental perturbations (Coppens et al., 2010; Harris et al., 2020; Jolles et al., 2019). Between-individual variation in predictability is an axis of behavioural variation has not previously been explored in calves. Recent work in adult cattle indicates that those individuals with lower within individual variation (i.e., more predictable) may be better able to cope with their environment (Poppe et al., 2020; van Dixhoorn et al., 2018). Future research could explore how feeding behaviour behavioural types and predictability estimates in calves are related to personality traits as well as medium and long-term health and production outcomes.

Repeatability was lowest for meal size. This result contrasts with one study in goats where meal size had high repeatability (Cellier et al., 2021). Calves in our study were fed a restricted feeding plan using the automatic feeder, which allocated a maximum meal size of 2 litres and restricted total daily intake to 10 litres. In addition, there were 16 calves present for a single feeder which is likely to have led to some competition between calves for access to the feeder (Von Keyserlingk et al., 2004). These constraints may have prevented calves displaying their preferred meal size and number of meals, especially if their preference exceeded the limits imposed or if they were frequently displaced. Nonetheless, there was evidence for moderate levels of repeatability for total meals. The variable total meals is the sum of meals with entitlement and meals without entitlement. Between individual variability in the number of meals with entitlement is likely to be limited by the feeder due the restrictions on total meals size and total daily intake. No such limits are present for meals without entitlement indicating that the between individual variation in total meals could be largely driven by meals without entitlement. Fewer visits to the feeder without entitlement is associated with ill health in dairy calves (Svensson & Jensen, 2007). Our results suggest that to use this behaviour as a feature to predict ill health it may be necessary to firstly account for substantial between individual variation in the number of visits the feeder.

Results from our model show a positive and significant correlation between total meals and feeding rate behavioural types. This result indicates that calves that had higher behavioural types for feeding rate also had higher behavioural types for total meals. Among individual correlations between behavioural types of distinct behaviours have been termed behavioural syndromes, a concept which is used to support the idea that there are underlying traits that drive the expression of more than one behaviour. For example, in animal personality research, behavioural syndromes are of interest as they may relates to an underlying personality trait which is responsible for the co-expression of suites of correlated behaviours (Sih et al., 2004). More broadly, the identification of behavioural syndromes is an important area of research that is helpful for understanding the (co)evolution of behaviours or behavioural specialization within a group (Sih et al., 2012). Indeed, it is possible that a behavioural syndrome exists between feeding rate and total meals because of an underlying factor that drives them both. For example, the paradigm of pace of life hypothesis developed in behavioural ecology situates individuals on an axis between fast and slow pace of life (Dammhahn et al., 2018). Benefits of a fast pace of life include higher metabolism, growth, and earlier reproduction. However, there are associated costs, such as a shorter lifespan and reduced investment in immune function (Stamps, 2007). Calves that have consistently higher feeding rates and higher meal frequencies could be those situated at the ‘fast’ end of the continuum, while those that have lower feeding rates and meal frequencies could be at the ‘slow’ end. Accordingly, using such data to phenotype individuals as a young age could enable management strategies such individualized feeding plans or identify individuals that may benefit from smaller group sizes where there is less competition around the feeder.

A variance partitioning approach such as presented here provides a relatively simple and scalable way to explore individual variation in livestock behaviour.

It is important to note that our measures of repeatability and predictability were calculated over a relatively short period (33 days) and therefore represent estimates of short-term repeatability and predictability (David et al., 2012). Both short- and long-term intervals are required to assess the temporal stability of these behavioural traits and the correlations between them. In addition, while four different cohorts in our study over different times of year were included, all calves were housed in the same environment, were fed the same milk allowance and each group had the same number of calves per feeding station. Future studies could vary these constraints to investigate the contextual consistency of the between individual differences and correlations reported here, evidence of which would further support the hypothesis that the observed differences in feeding behaviours could be driven by differences in personality (Kluen & Brommer, 2013). Differences in personality have been correlated with health and production outcomes (Koolhaas & Van Reenen, 2016; Neave, Costa, et al., 2018). Future studies could also explore the relationship between the measures reported here and outcomes such as daily live weight gain or immune function. In this study we did not include birth weight, its impact could be explored in future studies. During our study, only clinically healthy calves were included, and we manually inspected the calves twice weekly using an industry standard scoring system in addition to twice-daily visual inspections (McGuirk & Peek, 2014). However, no gold standard observational test exists for the diagnosis of ill health in calves (Buczinski et al., 2015; Timsit et al., 2016). It is possible that calves with subclinical disease could have been included in our analysis. New technologies may improve the sensitivity of disease detection by continuously monitoring physiological parameters such as core body temperature, allowing studies to better control for the potential effect of subclinical disease (Timsit et al., 2011a).

The approach presented here could be expanded further by incorporating measures from sensors that can monitor other behaviours such as general activity and/or social interactions between calves. Such research could assist in improving our understanding of between individual variation in these behaviours and how different behavioural types and predictability estimates are related to each other. Finally, whilst quantifying between individual differences and exploring the relationships between measures of behaviour allows us to hypothesize that personality traits could be driving among individual differences, it may be necessary to employ different statistical approaches to test if these traits can be measured directly using farm technologies. One such approach could be structural equation models (e.g., Araya-Ajoy & Dingemanse, 2014; Santostefano et al., 2017), which could be used to estimate relationships between such hypothesized latent traits and observed variables such as feeding behaviours.

3 Chapter 3 - Indication of a personality trait in dairy calves and its link to weight gain through automatically collected feeding behaviours

A close derivative of this chapter has been submitted for publication to *Scientific Reports*.

3.1 Introduction

In the same environment, conspecific animals often vary remarkably in their behaviour. Animal personality may contribute towards these differences (Réale et al., 2007). Animal personality is the result of personality traits which are defined as underlying dispositions that drive temporally and contextually consistent between-individual differences in animal behaviour (Kaiser & Müller, 2021). Personality traits have also been termed ‘personality axes’ (Sih et al., 2004) or ‘personality dimensions’(Gosling, 2001). Personality traits are of interest as they can help explain individual variation in life history and physiological traits (Biro & Stamps, 2008; Koolhaas, 2008). For example, the ‘pace-of-life syndrome’ hypothesis predicts that individuals with consistently faster behaviours (e.g., more active) have higher metabolism, faster growth and earlier reproduction but suffer trade-offs such as a more easily compromised immune system and a shorter lifespan (Réale et al., 2010). Indeed, there is evidence to suggest these trade-offs contribute towards the maintenance of differences in personality within a population (Stamps, 2007).

Whilst research into farm animal personality is in its infancy, studies indicate that personality traits, such as the boldness, sociality, and aggressiveness exist in livestock (M. A. Finkemeier et al., 2018). Personality traits cannot be measured directly because they are underlying (i.e., latent) traits (Kaiser & Müller, 2021; Réale et al., 2007). Instead, personality traits must be inferred from measured behaviours. In livestock, personality traits are typically inferred by use of tests that measure behaviour, often under controlled conditions (Gosling, 2001; Murphy et al., 2014a). For example, numerous studies in cattle have used behavioural tests, such as exposure to a novel object, to measure related behaviours (e.g., latency to approach novel object, time in contact with novel object) to infer a personality trait (e.g., boldness) (Foris et al., 2018; Lauber et al., 2006; Neave, Costa, et al., 2018). However, such tests are impractical beyond the research environment. A different approach is to collect repeated measures of behaviour and partition behaviour into between- and within- individual, by use of multi-level models (Dingemanse et al., 2010; Dingemanse & Dochtermann, 2013). These techniques allow between-individual variation to be quantified as a proportion of total variation (a measure termed repeatability) (Nakagawa & Schielzeth, 2010). Behaviours that are more repeatable indicate consistent between-individual differences and are therefore particularly useful to the study of animal personality (Bell et al., 2009). Where between-individual differences in behaviour are related, consistent over time, and consistent between contexts these differences indicate the presence of a personality trait (Kaiser & Müller, 2021).

In dairy calves, recent studies have combined data gathered by computerised milk feeders with suites of behavioural tests to show that differences in calf personality are associated with differences in feeding behaviour and performance. For example, studies in calves have reported positive associations between the exploration personality trait, feed intake and weight gain (Neave et al., 2018; Neave et al., 2019). However, despite the availability of suitable data gathered by precision livestock technologies, (e.g., Berckmans 2014; Carslake, et al. 2020; Hertel et al. 2020), few studies have attempted to harness these data to quantify between individual variation in behaviour directly (but see (Occhiuto et al., 2022)). One example of this approach is from our recent work which indicates that substantial, temporally consistent between-individual differences exist for calf feeding behaviour, specifically for meal frequency and feeding rate (Carslake et al., 2022). Furthermore, behavioural types for feeding rate and for meal frequency were positively and significantly correlated (i.e., calves that had higher feeding rates visited the feeder more frequently whilst those with lower feeding rates visited the feeder less frequently) suggesting the presence of a personality trait which could be driving the reported differences. However, to meet the definition of a personality trait, between individual differences in associated behaviours must be contextually as well as temporally consistent

(Kaiser & Müller, 2021). To our knowledge, no studies have tested the contextual consistency of between-individual differences in calf feeding behaviour. Furthermore, since personality traits are associated with physiological and production differences, if between-individual differences in feeding behaviours are associated with a personality trait these differences may also help explain individual variation in weight gain. The present study examined between-individual differences of two distinct feeding behaviours, feeding rate and meal frequency, over two different contexts in 76 pre-weaned dairy calves. We chose to study feeding rate and meal frequency as these behaviours are mathematically distinct (i.e., the calculation of meal frequency is independent of feeding rate) and our previous research indicates that these behaviours are repeatable and related in pre-weaned calves (Carslake et al., 2022). The two contexts studied here were pair housing (where each pair had access to a milk feeding station) and group housing (where one milk feeding station was shared between sixteen calves). The objective of this study was to test if between-individual differences in calf feeding behaviour, as measured by a computerised milk feeder, indicate the presence of a personality trait.

3.2 Materials and methods

3.2.1 Data collection

Calf recruitment

The study was conducted at the Centre for Dairy Science and Innovation at the University of Nottingham, UK. 76 female Holstein Friesian calves were enrolled in the study between 21/06/2021 and 22/01/2022. The calves followed normal management procedures for the farm. Ethical permission was obtained for the School of Veterinary Medicine and Science, University of Nottingham (unique reference number 1481 150603). All methods were performed in accordance with the relevant guidelines and regulations.

Housing contexts

This study followed calves during the pre-weaning period (i.e., prior to any reduction in milk allowance). During this time, calves were managed in two different housing contexts. The first context was pair housing where the two calves closest in age were grouped together in a small (3m x 2m) straw bedded pen from birth. Each pair had continuous access to a feeding station (i.e., one teat for two calves). The second context was group housing where 16 calves closest in age were moved simultaneously from pair housing to a larger straw bedded group pen (6m x 12m). Pairs stayed with their conspecifics throughout. Calves were moved together once the youngest calf in the group of 16 was approximately 3 weeks old. During this second context, all 16 calves shared access to a single feeding station (i.e., there was one teat per 16 calves). The feeding stations' dimensions were approximately 1m x 0.5m. Each station was equipped with sides but there was no back gate to reduce competitive behaviour around the feeder.

Feeding and colostrum

The farm management and protocols are as described in chapter 2. Briefly, during the first two days from birth, calves were fed colostrum followed by transition milk in line with farm protocols. From two days old calves transitioned onto milk replacer (Milkivit Energizer ECM, Trouw Nutrition GB) which was mixed by with warm water at 130g/L and dispensed by the computerized feeder (Forster-Technik Compact Smart) via the teat present in the feeding station. Each calf was equipped with an RFID ear tag enabling the feeder to recognise its identity by use of an RFID reader and dispense a milk allowance. The total daily allowance was distributed evenly throughout the day and the maximum amount of milk dispensed within any 2-hour period was limited to 2 litres. The daily allowance started at 6 litres at 2 days old and increased daily in line with age, reaching 8 litres at 5 days old. From 8 days old, the total daily allowance increased more gradually, reaching a plateau of 10 litres at 40 days old. During group housing, all calves, regardless of age, were fed 10 litres daily for 35 days after which the allowance was reduced daily until reduced to zero after 60 days in the group pen. Calves had ad libitum access to concentrates (FiMLAC Sweet Start Pellets), chopped straw and water throughout.

Data acquisition and selection

The computerized milk feeder used in this study logged feeding behaviour (calf identity, date, time start visit, time stop visit, milk consumption, drinking speed) during each visit a calf makes to the feeder. A new visit (row) was created whenever the RFID reader loses and then regains contact with a RFID tag. Data from the computerized milk feeder for the pair housed and the group-housed calves were downloaded and appended. The pair housing only included data where calves were a minimum of 7 days old since the daily increase in daily milk allowance was more gradual from this age (63mls/day). For each calf, 10 days of data were included. In the group housing period, the first week of data were excluded to allow a period of acclimatisation to the new pen and the subsequent 10 days were included in our analysis to ensure that an equal number of observations for each context. All calves in the group housing were on a level feeding plane of 10 litres daily. Any calves that we detected as sick during the data collection periods were excluded. Sick calves were detected during twice weekly health scores (Wisconsin score > = 5; n = 20) or during visual twice daily observations by farm staff) - (see Table 3.1 for calves excluded.)

Table 3.1. Number of calves included in analysis per cohort as per the inclusion criteria detailed in the methods and age of calves at start of data collection.

	Cohort 1	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Cohort 5**	**Cohort 6**	**Overall**
Calves							
Calves excluded	7	0	0	5	2	6	20
Calves included	9	16	16	11	14	10	76
Age at start of pair housing data collection (days) – included calves only							
Mean	16	23	12	12	12	11	15
Min	7	7	7	7	7	7	7
Max	33	42	17	16	16	17	42
Age at start of group housing data collection period (days) – Included calves only							
Mean	32	40	40	43	28	29	35
Min	23	26	26	37	24	24	23
Max	39	54	50	49	36	35	54

3.2.2 Data processing

Meal based criterion

A meal criterion was used to group visits by the individual that were close time into the same meal (Tolkamp & Kyriazakis, 1999). In this study, a meal criterion of 100 seconds was used as detailed in chapter 2 and appendix 10.1. Therefore, if two visits to the feeder by the same calf were separated by 100 seconds or less, they were grouped into the same meal. Those visits to the feeder by the same calf that separated by more 100 seconds were grouped into separate meals.

Feeding behaviours

For each calf and for each day of the group-housing period, feeding rate and meal frequency were calculated to describe the calves' feeding behaviour. These are detailed in Table 2.

Table 3.2. Definitions of feeding behaviours used

Feeding behaviour	Definition
Meal frequency (number per day)	Daily sum of all meals. This variable includes meals where the calf is entitled to a milk feed and meals where the calf is not entitled to a milk feed.
Feeding rate (ml/min)	Mean daily feeding rate. The feeding rate for each visit where the calf is entitled to a milk feed and consumes milk is calculated by the feeder. From this, mean feeding rate was calculated.

Weight data

Birth weight for each calf was manually recorded by use of an electronic weight scale prior to colostrum feeding. During the group housing period, weight data was continuously collected by use of a partial weigh scale which was attached to the front of the automatic feeder. The partial scales collected a recording every time the calf visited the feeder. The partial weigh scale had been previously validated (personal correspondence from manufacturer). Weight data was downloaded from the feeder for group housing context, processed and weight at 70 days old was calculated (see appendix 10.2 for details of data processing). Weight gain between birth and 70 days old was used to represent weight gain for this study as 35 days was the average age at which calves moved to the group pen and 70 days therefore reflects weight gain across both contexts. Weight gain between birth and 70 days was calculated by subtracting birth weight from weight at 70 days old.

3.2.3 Statistical analysis

A multivariate multilevel linear model was used to quantify behavioural variation and test for relationships within and between contexts. A multivariate model was also used to test for a relationship between feeding behaviour in the pair housing context and weight gain. Multivariate approaches were chosen as these models carry forward the uncertainty around point estimates into correlations between them thereby generating estimates with valid estimates of uncertainty (Houslay & Wilson, 2017). This is important since failing to account for uncertainty around point estimates (for example around estimates for individuals' average behaviour or behavioural type) by simply carrying forward central estimates from individual linear mixed models into a further analysis, can lead to spurious p values when correlations are computed between these (Houslay & Wilson, 2017). All statistical analyses were carried out using R software v3.5.1 (R Core Team, 2021). Code for the figures was adapted from Hertel et al. (2020).

Multivariate model to quantify behavioural variation within and across contexts

We ran a multi-level multivariate linear model with repeated measures of each feeding behaviour in each context (feeding rate pair housing, meal frequency pair housing, feeding rate group housing and meal frequency group housing) as response variables. This used "brms" package in R (Bürkner, 2018). The model can be written as per equation (1).

$$\boldsymbol{Y}_i = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{\varepsilon} \quad (1)$$

In the model α represents individual-specific random effect variation, Y_i represents the four response variables, X represents the fixed effects, Z the random effects. We controlled for the effects of age and birthweight by including them as fixed effects for feeding rate and meal frequency in the pair housing period. For the group housing age at grouping, number of days since grouping and their interaction term as well as weight at grouping were included as fixed effects. For all four response variables, individual Calf ID and cohort were included as random effects.

All distributions were specified as Gaussian. To capture a gaussian posterior distribution the variables meal frequency pair housing and meal frequency group housing were log transformed and feeding rate pair housing and feeding rate group housing were boxcox transformed. All variables were scaled after transformation (mean = 0; SD = 1). We used uninformative priors for both fixed and random effects. We ran four chains for 12,000 iterations, a warmup of 4,000 iterations and a thinning interval of 4. Model diagnostics indicated satisfactory convergence with $\hat{R} < 1.01$ and effective sample sizes > 400. Posterior predictive checks indicated that the underlying Gaussian distribution was satisfactorily captured.

Quantifying individual differences in feeding behaviours

Each calf had an individual specific intercept. For each response variable in our model, repeatability was calculated which represents the proportion of total variance that is explained by consistent differences between individuals. Rpt was defined according to the equation below:

$$Rpt = \frac{V_{ind_0}}{(V_{ind_0} + V_{eo})}$$

Where V_{ind_0} denotes the variance explained by differences between individual calves and V_{eo} denotes the residual (within individual) variance.

Rpt values are between 0 and 1. Higher values of Rpt for a behaviour indicate the population is composed of individuals that behave consistently differently from each other whereas low values indicate that individuals are more similar. Values for repeatability are described as high, moderate, or low in relation to a meta-analysis which summarized 759 estimates of repeatability from 114 studies and indicated that the mean level of repeatability was 0.37 [CI: 0.36 – 0.38](A. M. Bell et al., 2009). Values of repeatability from 0 to 0.25 are described as low, those from 0.25 – 0.50 as moderate, those above 0.5 as high.

Correlations within and between contexts

Our multivariate model estimates the correlations and credibility intervals between each of the response variables at Calf ID level. This allows us to test for correlations between the adjusted average behaviours (i.e., the behavioural types) for meal frequency and feeding rate within each context as well as test for correlations between contexts. Statistically, a calf's behavioural type corresponds to the individual level intercept from random intercept model for that behaviour.

To test if meal frequency behavioural types were correlated between contexts, the between-individual correlation and credibility interval for meal frequency in the pair housing and meal frequency in the group housing were extracted. To test if feeding rate behavioural types were correlated between contexts, the between-individual correlation and credibility interval for feeding rate in the pair housing and feeding rate in the group housing were extracted. To test if behavioural types for feeding rate and meal frequency are related within each context, we extracted the mean between-individual correlation and credibility interval for meal frequency and feeding rate in the pair housing context and the between-individual correlation and credibility interval for meal frequency and feeding rate in the group housing context. The absolute values of correlation coefficients from 0 to 0.4 are described as weak, from 0.4 to 0.7 as moderate and above 0.7 as strong (Dancey & Reidy, 2007).

Relationships between behavioural type and weight gain

A second multivariate model was used to test if behavioural types from the pair housing period were associated with weight gain. This model had three response variables: weight gain between birth and 70 days, meal frequency in the pair housing and feeding rate in the pair housing. The model can also be written as per equation (1). In this model α represents individual-specific random effect variation, Y_i represents the response variables, X represents the fixed effects, Z the random effects. For the response variables feeding rate and meal frequency, fixed effects included age and birthweight. For the weight gain response variable fixed effects included birthweight only. For all three response variables, we included individual Calf ID and cohort as random effects. Since there was only a single observation per calf for the weight gain response variable, its residual variance was fixed to 0.002 in our prior specification (Houslay & Wilson, 2017).

3.3 Results

Repeatability estimates by context

Repeatability estimates are reported in Table 3.3. In both contexts and for both behaviours, the degree to which individuals differed from each other as a proportion of total variation, as quantified by repeatability estimates, were moderate. These results imply that substantial between-individual differences exist for meal frequency and feeding rate in both the pair and group housing contexts.

Correlations between contexts

For feeding rate, the correlation coefficient for individual calves' behavioural types in the pair housing with individual calves' behavioural types in the group housing was 0.44 (CI: 0.20 – 0.64). For meal frequency, the correlation coefficient for individual calves' behavioural types in the pair housing with those from the group housing was 0.38 (CI: 0.10 – 0.63). These correlations are reported in Figure 3.1.

Table 3.3. Mean, median, inter-quartile range, repeatability, and coefficient of variation in predictability for meal frequency and feeding rate for the same calves by context (pair housing and group housing). IQR and CI correspond to interquartile range and credibility interval respectively

	Pair Housing		**Group Housing**	
	Feeding rate (ml/min)	**Meal frequency (number per day)**	**Feeding Rate (ml/min)**	**Meal frequency (number per day)**
Mean	609	18	836	10
Median	621	15	840	9
IQR	487 - 735	10 – 22	775 - 902	6 – 12
Repeatability				
Estimates	0.48	0.49	0.46	0.32
CI	0.39 - 0.58	0.29 - 0.49	0.37-0.56	0.22 – 0.41

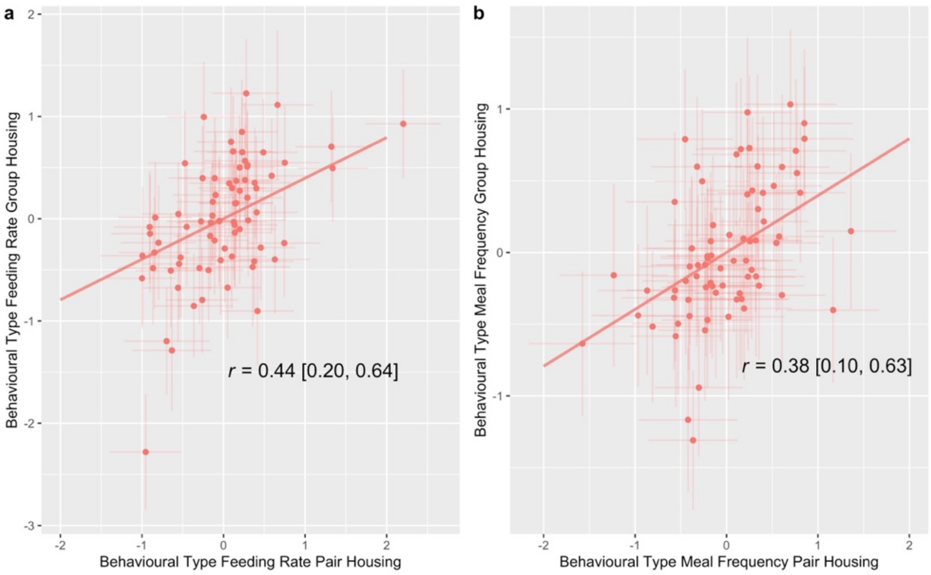


Figure 3.1. Visual representation of the line of best fit of the individual correlation (r) for feeding rate and meal frequency behavioural types between contexts. The credibility intervals are shown in square brackets. Panel a reports the correlation for feeding rate behavioural types between contexts. Panel b reports the correlation for meal frequency behavioural types between contexts. Posterior means and 95% credible intervals of point estimates for behavioural types are shown. The values on the x and y axes represent estimates for behavioural types which were calculated from the scaled response variables in the model.

Calves' feeding rate in the pair housing context was positively and significantly correlated with feeding rate in the group housing context and calves' meal frequency in the pair housing context was positively and significantly correlated their meal frequency in the group housing context. These results show that between-individual differences in feeding rate and meal frequency were positively and significantly correlated between contexts.

Correlations within contexts

The correlation coefficient between individual calves' behavioural types for meal frequency in the pair housing and feeding rate in the pair housing was 0.50 (CI: 0.28 – 0.68). The correlation coefficient between individual calves' the feeding rate behavioural types in group housing and the meal frequency behavioural types in the group housing was 0.28 (CI: 0.01 – 0.51). These correlations are reported in Figure 3.2.

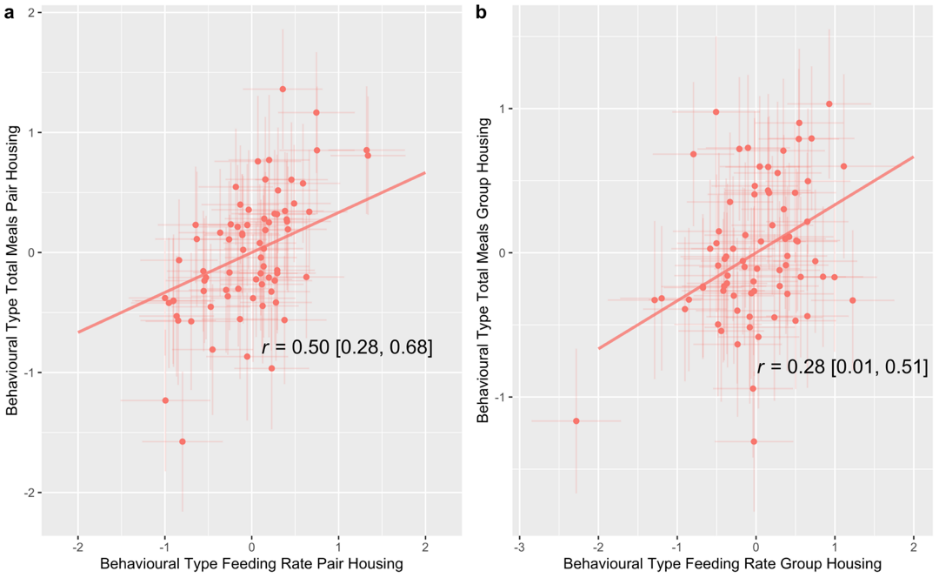


Figure 3.2. Visual representation of the line of best fit of the individual correlation (r) for feeding rate and meal frequency behavioural types within contexts. The credibility intervals are shown in square brackets. Panel a reports the correlation for feeding rate and meal frequency behavioural types in the pair housing and panel b reports the correlation for feeding rate and meal frequency behavioural types in the group housing. Posterior means and 95% credible intervals of estimates behavioural types are shown. The values on the x and y axes represent estimates for behavioural types which were calculated from the scaled response variables in the model.

These results show that, within each context, between-individual differences are positively and significantly correlated at the intra-individual level (i.e., calves that had higher feeding rates tended to have higher meal frequency whilst those with lower feeding rates tended to have lower meal frequency.)

Correlations with weight gain

Results from the multivariate model which included weight gain alongside feeding rate and meal frequency in the pair housing indicate that the calves' behavioural types for these behaviours were significantly and positively associated with weight gain. Correlation estimates were 0.42 (CI: 0.26 – 0.63) and 0.32 (CI: 0.12 – 0.52) for meal frequency and feeding rate respectively. This result shows that calves that had higher meal frequency and/or feeding rate in the pair housing had greater weight gain.

3.4 Discussion

This study is amongst the first in livestock to use automatically collected data to test if between-individual differences in behaviour are consistent with the existence of a personality trait. Our results show that for meal frequency and feeding rate in calves:

I. Repeatability estimates were moderate and between-individual differences were positively and significantly correlated across contexts, indicating that substantial, temporally, and contextually stable between-individual differences exist.
II. Within each context, between-individual differences were positively and significantly correlated at the intra-individual level indicating that the expression of meal frequency and feeding rate are related.
III. Between-individual differences in young calves were positively and significantly correlated with future weight gain.

Together, these results support our proposal that a personality trait exists which is driving contextually and temporally consistent between individual differences in calf feeding rate and meal frequency. The proposed personality trait situates calves on an underlying axis with faster drinking, high meal frequency calves at one end and slower drinking, low meal frequency calves at the other. This approach offers substantial practical advantages compared to traditional behavioural tests and indicates that data which is already available on many commercial farms could be used to automatically phenotype calves.

Between contexts, individual differences in feeding rate and meal frequency were positively and significantly correlated. This study is the first to report contextual consistency of between-individual differences in feeding behaviour non-laboratory animals. It is worth highlighting that the physical and social environment between the two contexts (e.g., group size, access to the feeder etc.,) deviated substantially. For example, there was one feed station for every two calves in the pair housing context whereas in the group housing context, a single feed station was shared between sixteen calves. It is reasonable to assume that the calves were required to adjust their behavioural strategy to each context. Indeed, the mean meal frequency decreased from 17 in the pair housing to 11 in group housing, suggesting reduced access to the feed station in the group housing which could be due to increased competition around the feeder (Von Keyserlingk et al., 2004). It is therefore particularly striking that, despite the substantial differences between contexts, between-individual differences were positively and significantly correlated between contexts indicating that the calves were highly motivated to display a preferred meal frequency and feeding rate. Whilst the correlations between contexts for feeding rate and meal frequency were weak to moderate (r = 0.39 and 0.45 respectively) there is evidence to suggest that these are comparable with the reported stability of more established personality traits in cattle. For example, one longitudinal study in cattle reported that exploration and boldness, as inferred from two behavioural tests taken 6 months apart, had correlations of 0.33 and 0.49 respectively indicating temporal stability (Foris et al., 2018). Our reported correlation between contexts is also in line with the magnitude of correlation which has been used to support trait continuity in children, with one meta-analysis of longitudinal studies of established personality traits indicating correlations of 0.41 within childhood (Fraley & Roberts, 2005). In humans there is evidence to suggest individual continuity in eating behaviours (such as eating speed) from the age of four years. This could indicate that between-individual differences in eating behaviour may also remain relatively stable in non-human animals (Ashcroft et al., 2008). If between individual differences in calf feeding behaviour persist into later life, these differences could inform individual management strategies in older cattle. One recent study in cattle reported that boldness and exploration personality traits had poor stability across puberty (i.e., from pre- and post-weaning to lactation) but were consistent from pre- and post-weaning to puberty and from puberty to lactation indicating long term consistency within these developmental periods (Neave et al., 2020). Further research is needed to explore the longer-term stability of between individual differences in calf feeding behaviour. It is worth noting that our study explored between individual differences over two different contexts which mainly differed in terms of their group size and access to the feeder. Further research is needed to test if changes to the milk feeding regime (e.g., milk allowance, meal size, meal timings etc) as well as if reductions in feed allowance during weaning affect the relative stability of the between individual differences in feeding behaviour reported here.

Our results show that calves that drank more quickly and those that had higher meal frequency gained more weight. This result indicates that calves that are situated on the higher meal frequency and faster feeding rate end of the proposed personality trait have higher growth rates. Indeed, there is evidence that personality traits are related to production in livestock (Müller & von Keyserlingk, 2006; Neave, Costa, et al., 2018; Neja et al., 2015). For example, several studies have reported that increased 'fearfulness' or increased 'reactivity to handling' is significantly associated with reduced weight gain in beef cattle and reduced milk production in dairy cattle (Haskell et al., 2014). Exploration, as measured by behavioural tests, has been associated with performance with one study in calves reporting that more exploratory individuals gained more weight (Neave, Costa, et al., 2018). However, a similar study in calves reported no difference (Whalin et al., 2022). More broadly, our results are in line with the 'pace-of-life syndrome' which predicts that individuals on the 'faster' end of the fast/slow continuum tend to have higher metabolism and grow more quickly (Dammhahn et al., 2018). The pace of life hypothesis draws associations between range of physiological characteristics and personality traits. Future research could explore whether calves at the faster and slower ends of the feeding personality trait continuum differ in terms physiological processes, such as metabolism or appetite regulation. Future research is also needed to investigate if these 'faster' calves may also suffer the trade-offs predicted by the 'pace of life syndrome' such as later reproduction and increased susceptibility to infectious disease.

Despite the marked differences between the pair housing and the group housing contexts, repeatability estimates for feeding rate and meal frequency remained moderate. This result indicates that, within each context, substantial between-individual variation exists. Between-individual variation in feeding behaviours has been reported in cattle (DeVries et al., 2003; Kelly et al., 2020; Neave, Weary, et al., 2018), as well as our previous work in calves (Carslake et al., 2022). Indeed, there is some evidence to suggest that between individual differences in feeding behaviours in cattle have a genetic component with heritability estimates of 0.44 for feeding rate and 0.51 for number of visits to the feed bunk (Kelly et al., 2021). The repeatability estimates in this study build upon previous results since this is the first to study to estimate repeatability in such young animals (mean age of 15 days in the pair housing context) indicating that between individual differences in feeding behaviour exist from a young age. Furthermore, since between individual differences persisted despite the substantial changes in context (from pair to group housing) our results suggest that these differences may be relatively robust to contextual change. Our results also show that within each context, a positive and significant correlation between calves' feeding rate and meal frequency exists. This relationship was present in both contexts suggesting that it too is relatively robust to contextual change. Feeding rate and meal frequency are mathematically distinct (i.e., the calculation of feeding rate is independent to that of meal frequency) indicating that this relationship is not the result of a mathematical artefact. Instead, this correlation could result from an underlying personality trait responsible for driving the co-expression of these behaviours. This is supported by some evidence in growing beef cattle which reported that feeding rate and meal frequency were positively though weakly correlated ($r = 0.20$) (Kelly et al., 2020), indicating that this relationship may persist into later life. It is worth highlighting that meal frequency in our study was a combination of entitled and non-entitled meals (where the calf visits a feeder but is not allocated a proportion). It has been suggested that non-entitled visits to the feeder may have an exploratory component (i.e., the calf is testing if any milk is available) (Weary et al., 2009). Further research is needed to determine how between-individual differences in calf feeding behaviour at an automatic milk feeder relate to feeding behaviours post-weaning.

One limitation of our study is that despite the differences reported meeting the definition for a personality trait, it is not yet possible to position this proposed personality trait in relation to those already established in the literature. However, there is some evidence to suggest that the proposed trait could be linked to 'exploration' and/or 'boldness'. For example, in one study, exploratory/active calves, as inferred from behavioural tests, had more visits to feeder and greater liveweight gain (Neave, Costa, et al., 2018). Another similar study reported that more exploratory calves tended to drink more quickly (Whalin et al., 2022). Furthermore, bolder, and more exploratory individuals tend to grow more quickly and are therefore situated on that faster end of the 'pace of life syndrome' slow/fast axis (Réale et al., 2010). Since calves on the faster end of the trait also gained more weight, this could also indicate a relationship with our proposed personality trait and boldness/exploration. Whilst future research could focus on the correlating between individual differences in measured behaviours with the results of suites of behavioural tests to infer personality, it is worth highlighting that these tests also have limitations (Carter et al., 2013). Criticisms include the fact the individuals are generally tested alone therefore their behaviour therefore may not generalise well to a social or group situation (Biro, 2012). A different avenue to improving the interpretability of the proposed trait could be to use sensors to monitor other behaviours (e.g., location, activity etc.) alongside feeding behaviour. Data driven approaches, as we present here, could assist in uncovering other personality traits (e.g., activity, exploration, sociality, aggressiveness etc.), enabling our proposed feeding personality trait to be interpreted within the context of a multi-dimensional personality framework.

4 Chapter 4 - Harnessing feeding data to detect bovine respiratory disease in calves

4.1 Introduction

Respiratory disease in calves is recognised as a major driver of production losses and poor welfare on cattle farms. It has been estimated that almost half of all UK dairy calves are affected (Johnson et al., 2017) and infection results in reduced growth rates, later age at first calving, and reduced milk production (e.g., Bach, 2011; Schaffer et al., 2016). Improved calf health can be achieved through better recording of health data and benchmarking between farms (Atkinson et al., 2017; Sumner et al., 2018). However, recording respiratory disease is notoriously difficult. Manually inspecting calves for signs of respiratory disease (e.g., taking rectal temperatures) is time consuming (McGuirk & Peek, 2014) and may not be routinely undertaken. In addition, the visual appraisal of calves for signs of respiratory disease is very poorly sensitive (Sivula et al., 1996; White & Renter, 2009; Wittum et al., 1996) meaning that numerous sick calves are unidentified and therefore are not treated or appropriately managed, compounding production losses and poor welfare. One approach to improving respiratory disease diagnosis could be to employ new technologies. For example, sensors that monitor behaviour could be used to detect deviations from behavioural norms which are indicative of ill health in livestock (Berckmans, 2017). An area of interest in calves is computerised milk feeders (Costa et al., 2021). These feeders are increasingly used in commercial farms to feed group housed calves. They automatically record behaviours such as visit frequency, milk consumption and feeding rate. There is an increasing body of evidence to suggest that changes to these behaviours are indicative of ill health. For example, studies have reported that sick calves have fewer visits to the feeder, lower milk consumption. and lower feeding rate (Borderas et al., 2009; Knauer et al., 2018; Svensson & Jensen, 2007).

A small number of studies have attempted to predict calf health status based on feeding behaviour. For example, one study implemented a time series approach which detects deviations in specific feeding behaviours from an individualised baseline based on past behaviour (Knauer et al., 2018). However, this method only had moderate (60%) accuracy for bovine respiratory disease detection. A different approach is to employ machine learning algorithms. Machine learning enables the non-linear combination of numerous features potentially providing excellent prediction accuracies (James et al., 2013). For example, results from an increasing number of studies show that machine learning can be useful in disease detection in adult dairy cattle (e.g., Keceli et al., 2020; Slob et al., 2021; Taneja et al., 2020). However, only two studies have employed machine learning to detect respiratory disease in calves. Both studies incorporated activity (as measured by an accelerometer) as well as feeding behaviour to predict health status (Bowen et al., 2021; Cantor et al., 2022). There is evidence to suggest that the accuracy of algorithms that use feeding behaviour alone may be low, for example results from (Bowen et al., 2021) indicate using feeding behaviours alone resulted in a low (45%) classification accuracy. However, this study only used a limited number of features (e.g., feeding behaviour on the day diagnosed) to predict ill health.

One difficulty that could impede the development of accurate predictions using feeding behaviour is the existence of substantial individual differences for these behaviours (Carslake et al., 2022). For example, repeatability estimates of 0.48 for feeding rate and 0.42 for meal frequency indicate that nearly half of all the variation for these behaviours is explained by differences between individuals. Calves also differ in their within-individual variation (predictability), e.g., calves with predictable feeding rates (low within-individual variation) and calves with unpredictable feeding rate (high within-individual variation) have been shown to coexist. Such variability in behaviour indicates that approaches are needed that can account for between- and within- individual differences in order prevent misclassification. One approach could be to incorporate features that quantify individuals' behaviour change around the time of disease into a machine learning model. Evaluating if different features calculated from computerised feeders can be used to predict of ill health, without the additional need to measure activity, would be useful to producers since computerised feeders are already available on many farms whereas implementing other technologies involves additional investment.

The aim of this study is to evaluate if using features that quantify individual behavioural change in a machine learning algorithm could enable the accurate prediction of bovine respiratory disease in calves by use of feeding behaviours.

4.2 Materials and methods

4.2.1 Data collection

Study population

160 Holstein Friesian female calves born between 22/12/2020 and 04/04/2022 at the Centre for Dairy Science and Innovation, University of Nottingham were recruited onto the trial. The 160 calves were pair housed at birth and then moved to a larger pen in groups of 16 with other calves closest in age. A total of 10 different groups (cohorts) were enrolled during the study period. Calf housing and feeding are briefly outlined below and are described in more detail in chapters 2 and 3.

Calf housing

From birth, calves were housed in pairs with calves closest in age being grouped into the same pen. The pair housing consisted of small, straw bedded pens (3m x 2m). Each pair had continuous access to one feeding station. Once 16 calves (i.e., 8 pairs) had reached a minimum of 3 weeks old, all 16 were moved together and housed in a large straw bedded group pen (6m x 12m) for a minimum of 70 days. This pen was also bedded with straw. All 16 calves shared access to a single milk feeding station. The data collection period for this study took place during the first 40 days for the period of group housing.

Feeding and Colostrum

Within four hours of birth, calves were fed 4 litres of pasteurised colostrum followed by transition milk (4 litres per feed fed by a bucket equipped with an artificial teat) fed twice daily for the subsequent two daily. From two days of age, calves were fed milk replacer by a computerised milk feeder on an increasing plane of nutrition which increased from 6 litres daily to 8 litres from 2 to 5 days old followed by a more gradual daily increase from 8 litres to 10 litres daily over the next 45 days. From the first day of group housing, all calves were 10 litres daily. The daily milk allowance stayed the same for the following 35 days and was gradually reduced thereafter such that all calves were weaned 25 days later. All calves had ad lib access to concentrates and chopped straw throughout both the pair and group housing periods.

Computerised milk feeder

The computerised milk feeder used in study were as described in chapters 2 and 3. Briefly, the milk feeder was Forster-Technik Compact Smart. During the pair housing one computerised feeder operated 4 milk stations whilst in the group housing one feeder operated two different stations. Each station is equipped with an RFID reader allowing it to recognise the ID of each calf from the calves' RFID ear tag. The feeder records a new row each the tag comes into close contact with the RFID reader. Thus, calf identity, visit time, date, duration, entitlement, feed consumption and feeding rate are recorded by the feeder software.

Health scoring

Calves were health scored twice weekly (e.g., Tuesday and Friday) according to an industry standard health scoring system that we modified to focus on respiratory disease (Table 4.1) (McGuirk & Peek, 2014). This system assigned a score of 0-3 for clinical signs indicative of bovine respiratory disease according to 5 different categories.

Table 4.1. Health scoring methodology used (McGuirk & Peek, 2014)

Category	0	1	2	3
Rectal temperature	< 38.3	38.4 – 38.8	38.9-39.4	> 39.5
Cough	None	Induce single cough	Induce repeated coughs or single spontaneous cough	Repeated spontaneous coughs
Nasal discharge	Normal serous discharge	Small amount of unilateral cloudy discharge	Bilateral, cloudy or excessive mucus discharge	Copious bilateral mucopurulent discharge
Eye scores	Normal	Small amount of ocular discharge	Moderate amount of bilateral discharge	Heavy ocular discharge
Ear score	Normal	Ear flick or head shake	Slight unilateral droop	Head tilt or bilateral droop

From these, the total health score was categorised as low, moderate, or high based on the summed total of these signs as well as their rectal temperature (Table 4.2).

Table 4.2. Health score categories

Low	Moderate	High
Total health score < 3 and temperature < 39.2	Total health score > 3 and < 5 and/or temperature >=39.2 and <= 39.5	Total health score >= 5 and temperature >= 39.5

4.2.2 Data processing

As this study focused on detecting changes in behaviour that occur when calves transition from ‘healthy’ to ‘diseased’, we only selected calves that had been scored as ‘low’ on the three previous health scores. Any calf with a moderate or high score during this time was excluded from the analysis. From these calves with three healthy scores, we categorised calves as ‘healthy’ or ‘diseased’ based on whether their subsequent health score was ‘high’ or ‘low’. Again, we excluded all moderate scores. The day of this subsequent (4th) health score is referred to as day zero. Days prior to day zero are referred with a minus sign (e.g., day -3 refers to the day 3 days prior to the fourth health score). The data was filtered to limit the number of times the same calf could be scored as either healthy or sick to two. This was done to prevent the over-representation of specific calves (e.g., calves scored repeatedly as healthy). Calves that had been in the group pen for more than 40 days prior to day zero were excluded to ensure all calves were on a similar plane of nutrition throughout as step down weaning occurred at this time. This data selection process left 48 calves in the 'healthy' category and 19 calves in the 'sick' (Table 4.3). Finally, for each calf, feeding data corresponding to the 15 days prior to day zero, day zero and the 3 subsequent days (i.e., days -15 to +3) were selected.

Table 4.3. Number of sick and healthy calves after data selection process by cohort

Cohort	4	5	6	7	8	9	10	11	12	14	Total
Sick calves	1	2	2	2	1	1	1	2	3	4	19
Healthy calves	2	3	3	10	15	3	12	0	0	0	48

From the selected dataset, 9 different feeding behaviours to describe feeding behaviour were created (Table 4.4). These were inspired by studies in the literature which indicates that visits without entitlement, feeding rate, and milk consumption could be useful to detect ill health in calves by use of their feeding behaviour (reviewed by Morrison et al., 2021). All behaviours were summarised per calf, per day.

Table 4.4. Feeding behaviours calculated from feeding data

Feeding behaviour	**Definition**
Time feeder	Total time at the feeder (s)
Time drinking prop	Time at the feeder spent drinking as a percentage of total a total time at the feeder (%)
Time interval max	Maximum duration between consecutive visits to the feeder (s)
Visits without entitlement	Number of visits without entitlement (n)
Visits with entitlement	Number of visits with entitlement (n)
Hit intensity	Number of reads of the RFID ear tag by the RFID reader. Each ‘hit’ corresponds to a new row in the data (n)
Feeding rate	Mean feeding rate recorded (ml/min)
Milk consumption	Total milk consumption (ml)
Meal size	Mean meal size (ml)

4.2.3 Feature engineering

From the feeding behaviour calculated, multiple features to describe each feeding behaviour were computed. These included behaviour around the day of diagnosis as well as features that describe within-individual changes around the time of diagnosis. To describe within individual change, features that quantify the difference in behaviour on the day of diagnosis relative to previous values (delta features) were calculated. We also used linear models to predict individuals’ future behaviour and quantified the residual deviation between these expected values and the observed values around the time of diagnosis (model-based features). Figure 4.1 shows the proportion of time at the automatic feeder spent drinking (Time drinking prop) for one healthy calf and for one sick calf.

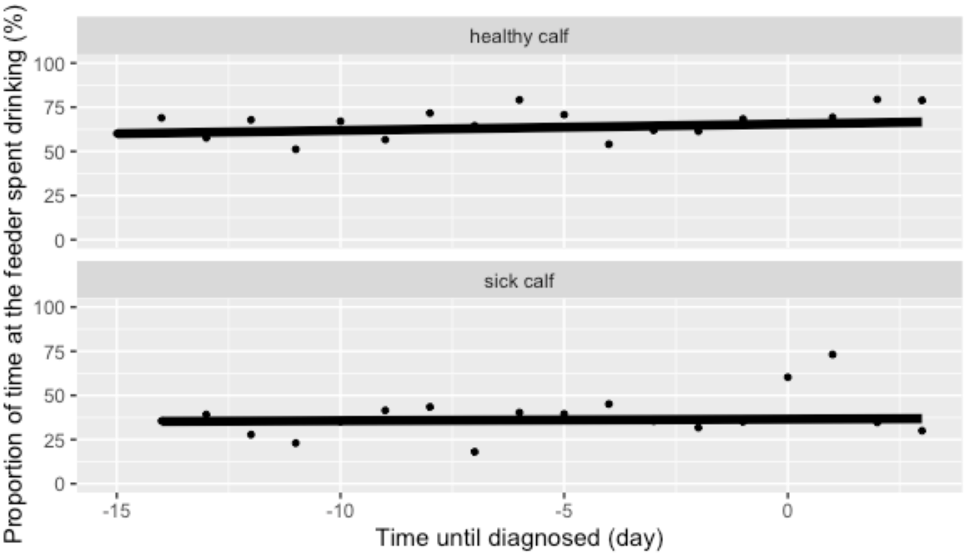


Figure 4.1. Plot representing the proportion of time spent drinking for one healthy and one sick calf. The line represents the model predictions, the points the observations. Calves are categorised as healthy or sick based on their health score on day zero.

For each behaviour, the following features were calculated:

- **Behaviour on the day of diagnosis:** Value of each behaviour on day zero.
- **Mean and standard deviation behaviour 3-day window:** For each behaviour, the mean value and standard deviation between day -2 and day 0 (inclusive) were calculated.
- **Delta features:** For each behaviour and each calf, we calculated the difference in behaviour on day 0 compared to the 10 previous observations. High values for delta features indicate a greater difference between the behaviour on day 0 and previous days.
- **Model based features:** For each behaviour we created a random intercept linear model with Calf ID and Cohort as random effects. Fixed effects in the model were age at day 0, number of days since grouping and weight at grouping. We log transformed the variables time interval max and hit intensity. We did not fit models for visits with entitlement and visits without entitlement as these had highly skewed distributions

and our models failed to converge. To create these models, we only used observations from day -14 to day -5. We then used these models to predict each behaviour over all observations for each calf (i.e., from day -14 to day 0) (Figure 4.1). To quantify the differences between model predictions and the observed values around the time calves were diagnosed as either healthy or sick, we calculated the residual deviation on day zero as well as the mean and the residual standard deviation for the window between days -3 and day 0. Calves with higher values for these residual features differ to a greater extent from their expected behaviour around day zero.

Features that are highly correlated provide the same information to the machine learning model. To reduce the number of highly correlated features we calculated the absolute value of the correlation between all features by computing their correlation matrix. For those pairs that were correlated more than 0.8, one feature was eliminated from each pair at random. This process removed 34 features leaving 105 to be carried forward into the final model. Finally, we standardised all features (mean = 0, SD = 1) to ensure that each feature had similar distributions (this was done to prevent certain features over-contributing in the classification model based on their distribution alone). Finally, since our dataset was unbalanced with more healthy calves than sick, we randomly selected a sub-sample of 19 calves from those that were labelled as healthy. This provided us with a balanced dataset containing 19 calves in each category.

4.2.4 Statistical analysis

To evaluate the usefulness of feeding behaviours to detect ill health, a random forest classification algorithm was implemented using the caret package in R (Kuhn et al., 2022; R Core Team, 2021). Here, classification performance was evaluated using fivefold cross-validation. This technique splits the original dataset into five different subsets. Four subsets are used to train the classifier and the last subset is used to test. This process is then repeated for five iterations. Prior to dividing the data into five subsets, we held out 20% of the dataset (hold out dataset) and tested the final performance of our model on this dataset. To evaluate performance on the hold out data set, we ran this process of cross validation and prediction on the hold out dataset 100 times. We then averaged predictions and observed outcomes on hold out dataset allowing us to estimate performance metrics of the model. Finally, feature ranking for the classification of health status was performed using the varImp function in the “caret” package in R (Kuhn et al., 2022; R Core Team, 2021). To calculate feature ranking, this function firstly calculates the contribution of each variable to the model by recording the prediction accuracy of the algorithm on the validation dataset for each tree. This process is then repeated after permuting each predictor variable. Finally, the difference between the two accuracies is averaged over all trees and then normalised by dividing by the standard error. Those variables with the highest difference between the two accuracies are considered the most important in the feature ranking (Kuhn et al., 2022).

4.3 Results

Classification results

Mean model performance on the hold out datasets was accuracy of 0.52 (SD 0.12), a sensitivity of 0.52 (SD 0.12) and a positive predictive value of 0.52 (SD 0.15). The confusion matrix from one of the hold out dataset is reported in Table 4.5.

Table 4.5. Example of Confusion matrix from one of the hold out datasets.

	Sick	Healthy
Sick	1	2
Healthy	0	3

Feature ranking

The top 10 features are reported in Table 4.6 by order of importance.

Table 4.6. Example of the relative feature importance from one of the models

Rank	Feature
1	Mean time spent at the feeder without entitlement
2	Hit intensity residual standard deviation
3	Time at the feeder spent drinking as a proportion of total time
4	Maximum daily feeding rate mean
5	Mean residual deviation of the time interval between visits
6	Time without entitlement residual standard deviation change
7	Mean allowance consumed
8	Time without entitlement 1
9	Standard deviation feeding rate
10	Difference between maximum time interval between visits 10 day previously and day of diagnosis

4.4 Discussion

This study is amongst the first to use machine learning to predict respiratory disease in pre-weaned calves from feeding behaviours and the first to incorporate features that explicitly quantify within-individual change. The overall classification performance was only 52%, indicating that monitoring feeding behaviours alone are not sufficient to detect bovine respiratory disease in pre-weaned calves.

The performance of our model on the hold out datasets was very low with a mean accuracy of 0.52 and a mean sensitivity 0.52. These results indicate that monitoring feeding behaviour alone is insufficient to predict respiratory disease in calves. Our results reflect those from two other studies that also reported low to moderate accuracy of automatically collected feeding behaviours to predict respiratory disease in calves. Bowen et al., (2021) reported an accuracy of 75% using a combination of activity and feeding features which decreased to approximately 45% accuracy when only feeding behaviours were included in the model. Another study in pre-weaned calves used statistical control processes to quantify deviations from previous behaviours and reported low accuracy (e.g., 60% when using a combination of feeding rate, milk consumption and unrewarded visits) (Knauer et al., 2018). It has been hypothesized that sickness behaviour in livestock follows a 'hierarchy of needs' where behaviours that are essential to maintain homeostasis in the short term are conserved, whereas non-essential behaviours are reduced to conserve energy (Weary et al., 2009). The poor predictive performance of studies that use feeding behaviour in calves to predict ill health could indicate that feeding behaviours are a core behaviour in pre-weaned calves and are maintained unless disease is severe. Our results contrast with publications in older cattle which indicate that changes in feeding behaviours are indicative of ill health (e.g., González et al., 2008). For example, Quimby et al. (2001) describes a sensitivity of 90%, positive predictive value of 91% and accuracy of 87% using feeding behaviour to detect sick steers. The discrepancy between studies in older cattle and those in calves could be explained by differences in diet composition. Milk is highly calorific and consumed in short time scale requiring low energy investment, whereas feedstuffs in adult cattle are high in roughage meaning that adult cattle must invest time and energy into consuming and digesting feed. This mismatch could help explain why studies report deviations in feeding behaviour (such as time at the feed bunk) enable the early detection of ill health in adult cattle but appear to be less useful in calves.

It worth highlighting that computerised feeders record several behaviours that are 'peripheral' to milk consumption (i.e., where the calf visits or spends time at the feeder but not consume any milk). These include behaviours such visits without entitlement, which results from several studies indicate are reduced in sick calves (Svensson & Jensen, 2007). Results from our feature ranking show that these peripheral behaviours (e.g., time spent at the feeder without entitlement) were more important in the model. However, the low performance of our algorithm indicates that even when combined with other behaviours they are insufficient. The inability of the model to predict health status, even when these peripheral behaviours are included, could be explained by the substantial between-individual variation that for these behaviours. For example, certain calves are more likely than others to engage in unrewarded visits to the feeder

(Carslake et al. 2022). Therefore, a reduction in unrewarded visits to the feeder may only be useful as an indicator of ill health in calves that tend to engage in this behaviour. Future research could explore if ill health is easier to detect by use of feeding data in those calves that tend have more unrewarded visits.

Of the 10 features that were most important in our model, 5 of these quantified within-individual change around the time of diagnosis. These included features such as the residual mean and residual standard deviation that quantify the departure of a behaviour from the expected value. However, the low model performance indicates that these within individual features were not sufficient to predict disease status in calves. Other studies have used time series-based methods to deviations in behaviour, these include statistical control process where approaches such as the cumulative sum are used to detect sudden changes. Whilst these have been used to accurately detect BRD related changes to feeding time in beef cattle (Quimby et al., 2001) and rumen temperature (Timsit et al., 2011a) such techniques had low accuracy using calf feeding behaviour (Knauer et al. 2018). Potential modifications to our methods that could be explored in future studies include adding a random slope which would allow individuals to differ in the degree to which they change their behaviour over time. Whilst our features quantified within individual change, it did not include any features at a resolution greater than daily measurements. Future studies could explore if features that quantify within day changes in behaviour could be useful in the detection of BRD in calves by use of their feeding behaviour.

One limitation of our study is the use of health scoring to detect clinical respiratory disease. Health scoring has low to moderate sensitivity which can result in some clinically sick calves or being misclassified as healthy. We attempted to improve the accuracy of our labels by only including calves that had consistently low health scores (=< 2) and lower temperatures (<39.2) as healthy. However, further research is needed to determine the if this methodology is sufficiently sensitive to prevent the inclusion of calves with clinical disease in the health category (i.e., false negatives.) Whilst the sensitivity of health scoring is low to moderate, specificity is greater which indicates that those calves labelled as sick are likely to be truly BRD positive (i.e., true positives). There is some evidence to suggest that combining lung ultrasonography with Wisconsin health scoring improves the sensitivity of bovine respiratory disease diagnosis which could provide future disease prediction studies with a different methodology of labelling sick and healthy calves. However, credibility estimates around the sensitivity of lung ultrasonography are wide (Berman et al., 2019) indicating that that is still substantial uncertainty around this method (Reynolds & Brennan, 2021). Nonetheless, one recent study reported high classification performance for BRD detection (accuracy = 0.94) where calves were labelled as 'clinical BRD' if diagnosed with an area of consolidated lung (>3cm^2) and a high Wisconsin score (Cantor et al., 2022). This study used a combination of activity as well as feeding behaviours and 'clinical BRD' calves were then classified against themselves using data collected prior to their diagnosis. It is worth noting clinical BRD calves had a substantial area of consolidated lung >3cm^2. It is possible therefore that these calves would have been suffering relatively severe symptoms associated with BRD. In addition, it is possible that lung lesions that are detected by lung ultrasonography only appear at a later stage, potentially resulting in more persistent changes in behaviour in this study compared to those that used a high health score alone (Reynolds & Brennan, 2021). Unfortunately, this study does not detail the predictive performance of feeding behaviours alone (i.e., without features calculated from activity) limiting comparisons with our results.

Other options for that could improve accurate labelling individuals according to BRD status include regularly taking biological samples to screen for indicators of ill health. These could include monitoring the immune response (Schaefer et al., 2012), taking samples from the respiratory tract to detect and quantify different pathogens (Fulton & Confer, 2012), or new approaches such as using blood metabolites (Blakebrough-Hall et al., 2020). However, none of these can serve as a gold standard. A less invasive approach which has yet to be explored by any study is combining behavioural and continuous temperature monitoring. This could be achieved using currently available technologies such as rumen temperature boluses (Timsit et al., 2011a). Pyrexia is indicative of infection in a large variety of species (Wingfield, 2003) and is associated with clinical BRD in calves (Grigor et al., 2001). Monitoring temperature alongside feeding in calves could allow the accurate labelling of behavioural changes in feeding behaviour that occur alongside pyrexia, potentially enabling future studies to conduct a more robust assessment of the use of feeding behaviour to predict ill health in calves.

To conclude, our results indicate that monitoring feeding behaviours in calves is insufficient for the detection of respiratory disease even when models include features that quantify within-individual change at the daily resolution carried out in this study. However, conclusions are limited by our use of health scoring to detect ill calves, which may have resulted in the mislabelling of some calves due to its low sensitivity. Future studies should explore different methodologies to label BRD positive calves more accurately such as lung ultrasonography, biological sampling, and continuous temperature monitoring. Finally, future research should focus on comparing the predictive performance of different features from a diversity of behaviours (e.g., activity, social interactions, feeding) to determine those that are most useful for the prediction of respiratory disease in calves.

5 Chapter 5 – Continuous temperature monitoring to detect bovine respiratory disease in cattle youngstock

This chapter was conducted with the studentship's industrial partners PrognostiX.ltd who provided the bolus temperature sensors, hardware, and technical support.

5.1 Introduction

Fever occurs in all vertebrates to help fight infection (Wingfield, 2003). It is mediated by pro-inflammatory cytokines which act on the hypothalamus to prevent compensatory mechanisms that regulate temperature (Kluger, 1991). The resulting increase in temperature creates a hostile environment for pathogens thereby increasing the effectiveness of the immune system (Johnson, 2002). In livestock, increased body temperature is an important clinical sign of infectious diseases such as metritis, mastitis, and respiratory disease (Schutz & Bewley, 2009). For example, in cattle body temperature increases post-inoculation with respiratory pathogens (e.g., Theurer et al., 2013) and is an important clinical sign for the diagnosis of respiratory disease (Grigor et al., 2001; McGuirk & Peek, 2014)

Traditionally, measuring body temperature in livestock requires the capture and restraint of the animal to take a rectal temperature using a handheld thermometer. Due to handling and labour constraints, temperature is not routinely monitored in livestock. Instead, stockpersons tend to rely on visual appraisal to detect sick animals which is very poorly sensitive, especially for bovine respiratory disease (Timsit et al., 2016). For example, one study showed that 68% of untreated feedlot steers had lung lesions which were detected at slaughter (Wittum et al., 1996). New technologies that permit continuous temperature monitoring could enable the improved detection of cattle experiencing a fever. One example is intra-ruminal sensors; these take the form of a bolus that is administered orally, swallowed, and passed through the oesophagus to the rumen (more specifically the reticulum). Studies have reported variable correlations between rectal and rumen temperature. For example, Sievers et al., (2004) showed a high correlation, ($r = 0.92$) in a single group of adult cattle, whereas (Bewley et al., 2008) undertook a greater number of observations ($n = 2042$) over different times of year in different groups of animals and observed a lower correlation ($r = 0.64$) indicating that rumen temperature is affected by dynamics such as feed and water intake, season, and environment factors. However, there is evidence to suggest that rumeno-reticulum boluses could assist in detecting cattle experiencing pyrexia. For example, Timsit et al., (2011) reported a high correlation of ($r = 0.91$) from 24 beef bulls with pyrexia indicating that rumen temperature and rectal temperature are well correlated when temperature is elevated. One experimental study monitored rumen temperature in 24 beef cattle that were exposed to *bovine viral diarrhoea virus* and then experimentally infected with the respiratory pathogen *Mannheimia haemolytica* (Rose-Dye et al., 2011). Results from this study show that maximum rumen temperature was increased ($p < 0.01$) by 1.2 C in infected calves compared to controls indicating that maximum rumen temperature could indicate respiratory disease in cattle.

Two studies have evaluated if algorithms that could continuously monitor temperature as measured by intra-ruminal boluses are predictive of bovine respiratory disease (Timsit, et al., 2011; Voss et al., 2016). The algorithms evaluated were based on the cumulative sum methods where an alert is triggered based on the accumulation of the differences between consecutive measurements and a reference threshold. Using this method, Timsit et al., (2011a) reported that sick steers were detected with a positive predictive value of 73%, whilst Voss et al., (2016) reported a sensitivity of 71% and specificity of 98%. These results indicate that temperature monitoring by use of intra-ruminal boluses shows promise for the detection of respiratory disease in cattle. However, these studies did not evaluate the use of other features can be computed from rumen temperature data. For example, consuming water leads to sudden drops in rumen temperature and can be used to measure drinking behaviour (Vázquez-Diosdado et al., 2019a). Other features that could assist in the identification of sick individuals include maximum temperature, minimum temperature, and the total area above different temperature thresholds. It is not known if incorporating these features into a single algorithm could improve predictions, nor which features are the most helpful for the identification of sick individuals. Furthermore, algorithms that are developed on a single population may be specific to the data distribution in that underlying population. This can affect their ability to generalise and result in reduced performance when tested on a new dataset. As remarked by Hoadley (2001) “high performance on test sample does not guarantee high performance on futures samples, things do change”. Testing an algorithm on a new population could help provide insight into its generalisability.

This study has two objectives, firstly develop, evaluate, and rank the features of a machine learning algorithm that uses diverse features calculated from temperature boluses data to predict respiratory disease in steers. Secondly, to test the generalisability of the algorithm by comparing predictions and observations in a new population of cattle.

5.2 Materials and methods

This study consisted of two experiments. In experiment one, we developed machine learning algorithms to predict health status using rumen bolus temperature data from single population of cattle using different health score thresholds to define healthy and sick animals. In experiment two, these algorithms were tested on a new population and their performance evaluated.

Temperature monitoring system

The rumen temperature bolus used for this study was manufactured by *PTS technologies, Singapore* (PTS Technologies Ltd). The bolus provides continuous measurement of the temperature in the rumen with a relative measurement accuracy of +/- 0.1 Celsius. The temperature and bolus ID data were transmitted wirelessly from the bolus located in the rumen to a central collector (base station) from which data were relayed to cloud storage. The bolus has a maximum range of 30 metres and has memory capacity allowing data to be stored and transmitted at a later stage if the bolus is out of range of the base station. The bolus dimensions were 95mm in length and 35mm in diameter. The same temperature monitoring system was used for both experiments 1 and 2.

5.2.1 Experiment 1

Study population and management

106 Holstein Friesian steers aged between 19 and 31 weeks were bolused with the rumen temperature monitoring sensor upon arrival on a commercial feedlot on the 11/11/2019. A picture of the bolus is shown in Figure 5.1. In line with industry standards for this type of system, the cattle been transported to the farm from several calf rearing units and had previously originated from twenty-two different dairy farms. The steers had received two doses of a vaccination against multiple respiratory pathogens (*Rispoval 4- Zoetis*) prior to arrival. They were housed in two identical covered corrals (approximately 50 steers in each corral) which was bedded on straw. Each group had access to a water trough and ad lib access to a total mixed ration which was delivered twice daily to the feed face. The ingredients for the ration were whole crop maise, crimped maise, dried bakery product, potato cream and rolled oats. Protein content was 15.5%.

Fig. 5.1. Picture of the temperature bolus

Health scoring

The steers were health scored by a veterinary surgeon once weekly for 5 weeks post arrival at the farm. Health scoring was completed using a modified version of the Wisconsin Calf Health Score (McGuirk & Peek, 2014) where parameters for nose, eyes, ear, and rectal temperature were recorded (Table 5.1). Due to a rising incidence of respiratory disease, all steers were treated with an anti-inflammatory and a long-acting antibiotic on the 27/11/2019.

Table 5.1. Health scoring methodology used

Category	0	1	2	3
Rectal temperature	=< 38.3	38.4 – 38.8	38.9 - 39.4	> 39.5
Cough	None	Induce single cough	Induce repeated coughs or single spontaneous cough	Repeated spontaneous coughs
Nasal discharge	Normal serous discharge	Small amount of unilateral cloudy discharge	Bilateral, cloudy or excessive mucus discharge	Copious bilateral mucopurulent discharge
Eye scores	Normal	Small amount of ocular discharge	Moderate amount of bilateral discharge	Heavy ocular discharge
Ear score	Normal	Ear flick or head shake	Slight unilateral droop	Head tilt or bilateral droop

Data pre-processing

Data from three different sources (temperature bolus, health scores and steer characteristics) were integrated. The data file contained temperature information at a frequency of 15 minutes from the 14 of November 2019 to the 18 December 2019.

Feature exploration and generation

Firstly, we explored different statistical features that could be used in a detection algorithm. Initially, we explored information from sick and health steers. Figures 5.2 and 5.3 show an increase of temperature above 40°C for the sick compared to the healthy steer.

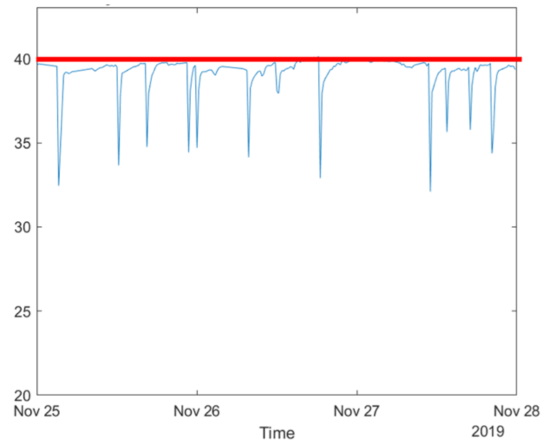


Figure 5.2. Temperature profile for a healthy steer (score 0) and threshold (in red) at 40ºC.

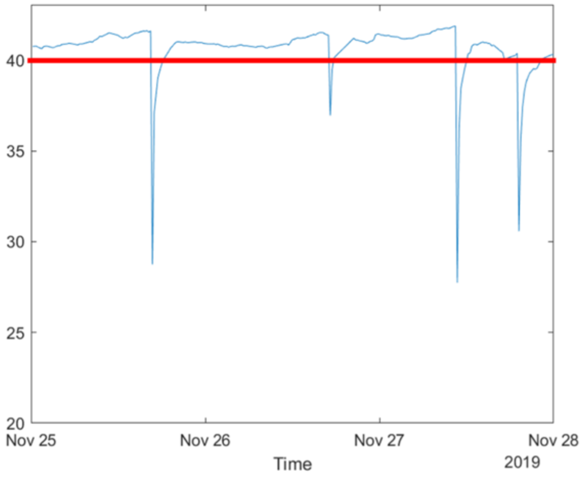


Figure 5.3. Temperature profile for a sick steer (score 7) and threshold (in red) at 40ºC.

Another set of characteristics that we explored was the number of drinking events, these were based on rumen temperature drops that are characteristic of drinking. We used two different algorithms that have been previously validated to determine drinking behaviour (Vázquez-Diosdado et al. 2019a). These are a general threshold and a cow-day specific algorithm. The threshold chosen for the general algorithm was 38.1ºC and the cow-day specific threshold is defined as µday – σday where µday is the mean cow temperature of the specific day and σday is the standard deviation of cow temperature of the specific day. From these two algorithms, a drinking event is detected if temperature is below 38.1C for the general threshold and below µday – σday for the cow day specific threshold.

Figure 5.4 is taken from Vázquez-Diosdado et al., (2019a) and shows an example taken from of the cow-day specific drinking algorithm for one day. The drinking events are denoted by the red circles.

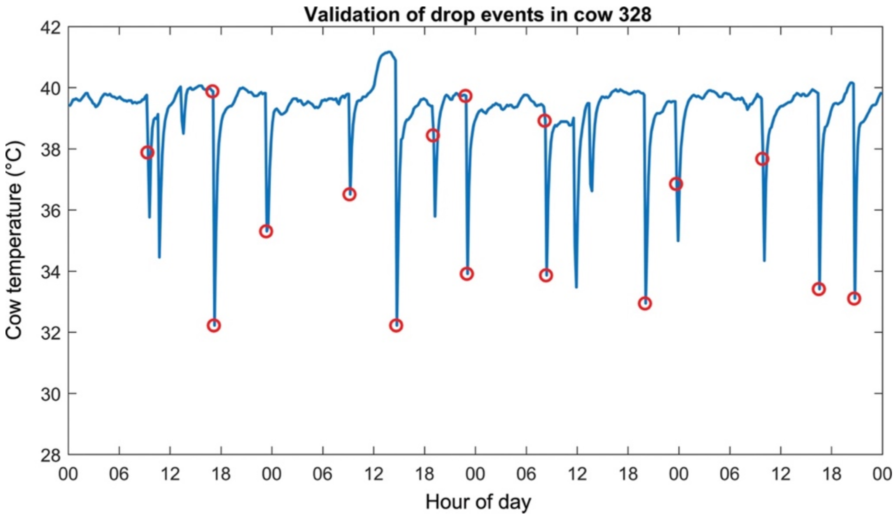


Figure 5.4. Example of the detection of drinking events using the cow-day specific threshold algorithm. The red circles denote the detected drinking events. From Vázquez-Diosdado et al. (2019a).

Based on the feature exploration, feature characteristics based on the area under the curve for different thresholds were computed as well as feature characteristics for drinking behaviour. The feature calculated are detailed below.

Temperature features computed:

- Overall area above 39.5 °C,
- Overall area above 40 °C,
- Overall area above 40.5 °C
- Overall area above and 41 °C
- Overall area above 41.5 °C
- Overall area above 42 °C
- Maximum temperature after removing drinking with cow-day specific algorithm
- Maximum temperature after removing drinking with general algorithm

Drinking features:

- Number of drinking events cow-day specific algorithm
- Maximum of drinking events cow-day specific algorithm
- Minimum of drinking events cow-day specific algorithm
- Number of drinking events general algorithm
- Maximum of drinking events general algorithm
- Minimum of drinking events general algorithm
- Mean number of drinking events general algorithm
- Mean duration between drinking events general algorithm
- Standard deviation of the number of drinking events from the general algorithm
- Standard deviation of the number between drinking events from the cow-day specific algorithm

Data selection

The data selection process is detailed in Figure 5.5. During this process we created two labelled datasets from our original dataset. Firstly, to test the impact of including different health scores on model performance, we created one dataset (dataset A) where we only included those steers with the highest (> 6) or lowest (< 3) health scores. This meant all steers with a health score between 3 and 6 (i.e., edge cases) were removed from this dataset. Steers with a health score < 3 were labelled as healthy, those with a health score > 6 were labelled as sick. In our second dataset (dataset B) we included all steers. In line with the clinical recommendations for the health scoring system we used (McGuirk & Peek, 2014), steers with a health score >= 5 or two or more clinical parameters which scored 2 or above were labelled as sick, the rest were labelled as healthy. The total number of observations in dataset A was 101 (72 healthy and 29 sick), the total number of observations in dataset B was 514 (342 healthy and 172 sick).

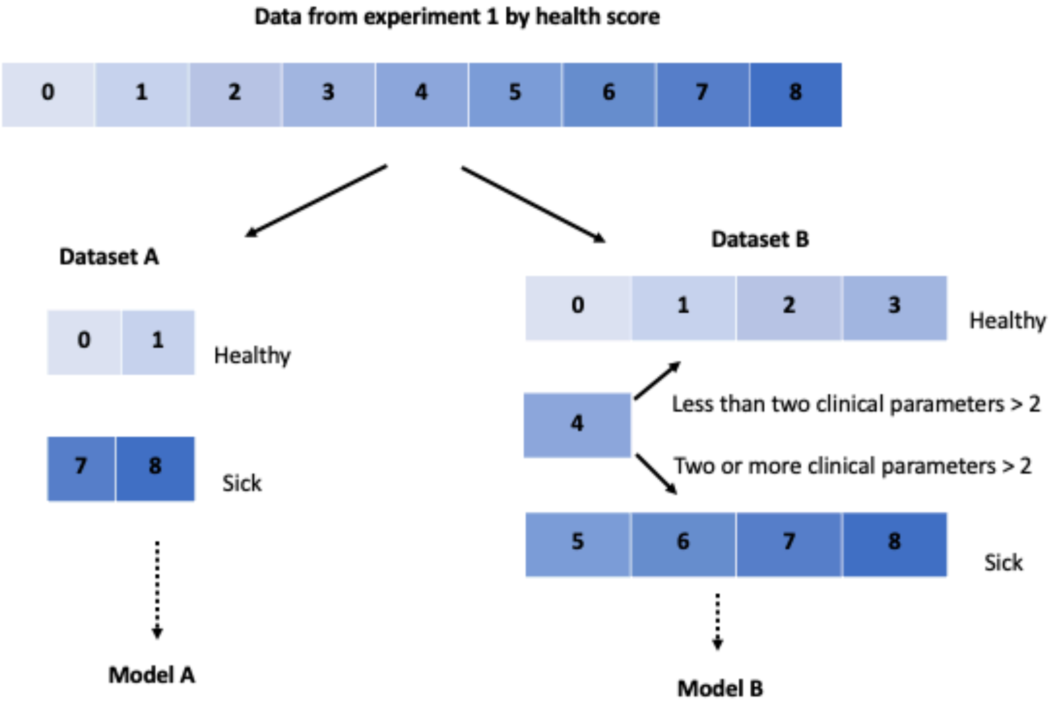


Figure 5.5. Data selection for datasets A and B

Model training and evaluation

For each dataset A and dataset B we trained and tested a classification models, we refer to these as model A and model B respectively (see Figure 5.5). The models used feature characteristics computed over a 3-day window which included the day the health score was recorded and the 2 previous days as predictor variables. The response variable was the label ‘healthy’ or ‘sick’. The algorithm used to create each model was a random forest algorithm with 5-fold cross validation. We chose area under the curve as the performance metric to be maximised. Model performance was evaluated using a holdout dataset (30% of the original dataset), which was removed prior to training the classifier. To obtain estimates of model performance, we compared our model predictions on the holdout dataset with the labelled ground truth. We ran this process of model training and cross validation followed by testing on the hold out dataset 100 times and report the mean and standard deviation for the model performance metrics on the hold out datasets. This process is detailed in Figure 5.6. All model training and evaluation was computed using the “caret” package (Kuhn et al., 2022) in R (R Core Team, 2021).

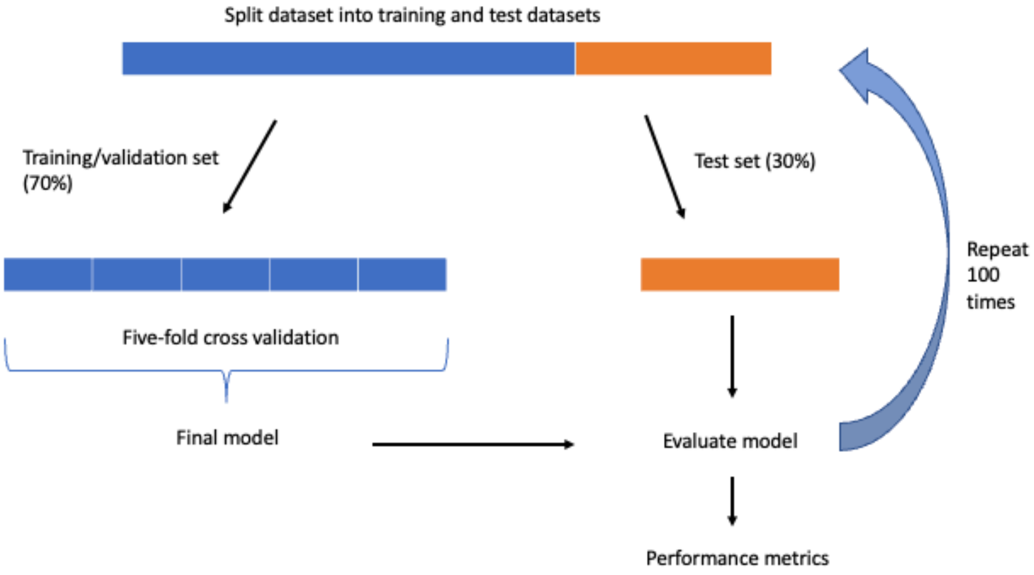


Figure 5.6. Process for model evaluation

Feature ranking

For model A, we computed the top 10 feature characteristics in order of importance in the classifier. These were computed using the varImp function in the caret package (Kuhn et al., 2022).

5.2.2 Experiment two

Study population

In experiment two, the two models that were trained and evaluated in experiment one were used to predict health status in a new group of cattle. The new group of cattle consisted of 108 Holstein Frisian steers which were sourced through the same rearing system. They arrived at the farm on the 27/11/2020 and were bolused on the 05/12/2020. Ages were comparable with experiment 1 (20 weeks – 35 weeks) and steers were housed in the same building as experiment 1 and fed the same total mixed ration.

Health scoring

Two separate visits were undertaken to evaluate health scores on the 10th and the 15th of December. Due to handling constraints, we were unable to inspect all the steers and inspected 43 on the first visit and 17 on the second. The health scoring system used was as described in experiment 1.

Model evaluation

To assess the performance of the two detection algorithms on different cut-offs for the definition of sick and healthy steers, we used two different methodologies to categorise steers as either healthy or as sick. Firstly, we only included steers with the highest scores (>5) which we categorised sick, and those with the lowest scores (<3) which we categorised as healthy. We refer to this dataset as “dataset extremes”. Secondly, we included all steers and labelled those with a health score >= 5 or two or more clinical parameters which scored 2 or above as sick, the rest were labelled as healthy. We refer to this dataset as “dataset all” (see Figure 5.7). The total number of datapoints in dataset extremes was 27 (19 healthy and 8 sick), the total number of datapoints in dataset all was 56 (37 healthy and 19 sick).

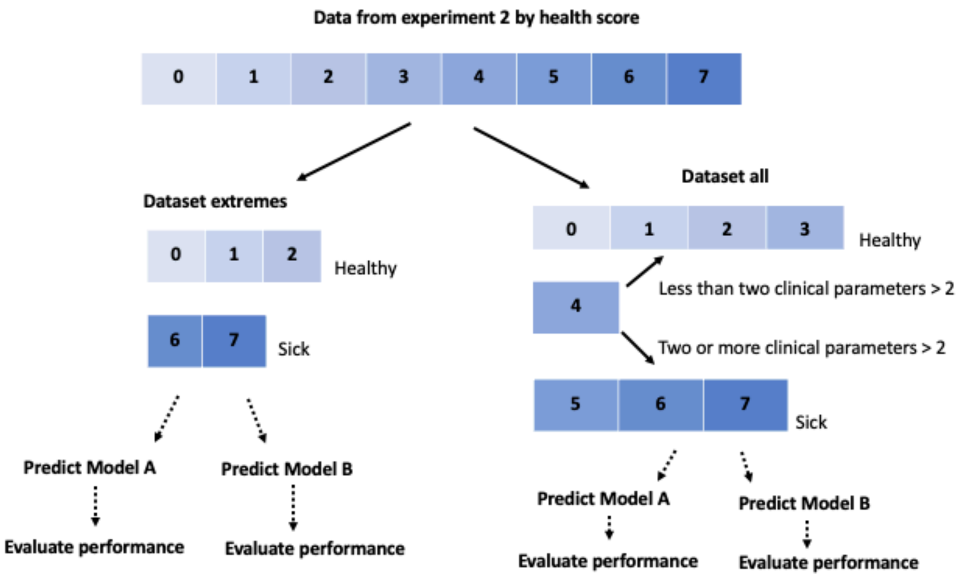


Figure 5.7 Model evaluation using data from a new group of steers

As per experiment one, we used features calculated from the bolus temperature data collected on the day of the visit and the two previous days as predictor variables for the labelled outcome variable (sick for healthy). We predicted each steer as either sick or healthy using model A and model B and compared the predictions from these models with our observations. From these, we assessed model performance on each dataset using the standard definitions of accuracy, specificity, recall, precision, and F-score.

5.3 Results

5.3.1 Experiment 1

Health inspections

The number of steers by health score and date are detailed in Table 5.2.

Table 5.2. Number of steers by health score and by date

Health Score **Date Scored**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
20/11/2019	1	10	14	19	20	18	13	2	7
27/11/2019	0	3	7	19	25	14	21	11	6
04/12/2019	2	32	36	19	7	4	1	1	0
11/12/2019	1	18	46	16	9	7	0	0	0
18/12/2019	1	4	21	25	22	20	10	2	0
Total	5	67	124	98	83	63	45	16	13

Classification performance

The mean and standard deviation of the classification performance of the random forest models A and B on the hold out datasets are reported in Table 5.3. The performance of the algorithm used to predict a sick steer as a positive case is presented.

Table 5.3. Mean and standard deviations of classifier performance metrics computed over 100 iterations on holdout datasets.

Performance	**Accuracy**	**Specificity**	**Sensitivity**	**Precision**	**F-Score**
Model A (mean +/- SD)	82.5% (+/- 7.2)	92.9% (+/- 5.8)	72.1% (+/- 14)	81.5% (+/- 13)	75.3% (+/- 10)
Model B (mean +/- SD)	62.6% (+/- 3.0)	87.0% (+/- 3.6)	38.1% (+/- 5.8)	64.8% (+/- 13)	46.5% (+/- 5.5)

Table 5.4 details a confusion matrix computed over one of the holdout datasets used to evaluate model A

Table 5.4. One example of a confusion matrix for model A from one of the hold out datasets which was computed over the 100 iterations. Healthy denotes steers with a health score < 2, sick a health score > 6.

Confusion Matrix		**Predicted**	
		Healthy	**Sick**
Observed	**Healthy**	21	2
	Sick	0	6

Table 5.5 details a confusion matrix computed over one of the holdout datasets used to evaluate model B.

Table 5.5. One example of a confusion matrix for model B from one of the hold out datasets which was computed over the 100 iterations. Sick denotes steers with a health score > 5 or two or more clinical parameters >= 2.

Confusion Matrix		**Predicted**	
		Healthy	**Sick**
Observed	**Healthy**	93	33
	Sick	9	18

Feature ranking

The total areas over specific thresholds had the highest performance in the classifier followed by the maximum temperature and various drinking related features. The 10 most important features are detailed in Table 5.6.

Table 5.6. Feature ranking of the classifier in order of importance

Rank	**Variable**
1	Overall area above 39.5 °C
2	Overall area above 40.0 °C
3	Overall area above 40.5 °C

4	Maximum temperature cow-day specific
5	Maximum temperature after removing drinking with general algorithm
6	Minimum of drinking events general algorithm
7	Standard deviation number of drinking events general algorithm
8	Standard deviation of number between drinking events cow-day specific algorithm
9	Maximum of drinking events cow-day specific algorithm
10	Number of drinking events general algorithm

5.3.2 Experiment 2

Health assessment

The number of steers by health scores by date are detailed in Table 5.7.

Table 5.7. Number of steers by health score and date

Health Score / **Date Scored**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
09/12/2020	3	4	8	7	7	5	4	1
12/12/2020	0	0	4	2	2	6	3	0
Total	3	4	12	9	9	11	7	1

Model performance

The performance metrics for models A and B over the two datasets from experiment 2 are detailed in Table 5.8. In the Table below, ‘extremes’ refers to the dataset where all steers had either a health score >= 6 (sick) or <= 3 (healthy). ‘All values’ refers to the dataset containing all steers where those with a health score > 5 or two or more clinical parameters >= 2 were categorised as sick.

Table 5.8. Performance metrics for the two models A and B developed in experiment 1 to predict health status a new group of steers

Performance	Accuracy	Specificity	Sensitivity	Precision	F-score
Model A - extremes	74%	79%	62%	55%	59%
Model A – all values	61%	72%	37%	41%	39%
Model B - extremes	70%	73%	63%	50%	55%
Model B – all values	60%	70%	42%	42%	42%

5.4 Discussion

Our results show that temperature boluses have potential to assist in the detection bovine respiratory disease and that features such as area over the curve and drinking behaviours are useful in the detection of respiratory disease. However, sensitivity was moderate to low, and performance was reduced when the algorithms were tested on a new group of steers indicating that further work is needed.

In experiment one, model A was able to differentiate between steers with low health score (<3) and high health scores (>6) with high specificity (92.9%) and moderate sensitivity (72.1%). These results show that most steers that had a health score low health were correctly predicted as healthy and that close to three quarters of steers with a high health score were correctly predicted as sick. These results indicate that continuously monitoring by use of temperature boluses could be a promising tool to identify steers suffering from bovine respiratory disease. However, in experiment 1, model A was trained and tested on datasets which only included steers with the most extreme scores. When we used the same methodology to repeatably train and test another algorithm (model B) using a dataset that contained all health scores (dataset B), model performance was substantially reduced (from 82.5% for model A to 62.6% for model B). Specificity decreased from 92.9% for model A to 87.0% for model B and sensitivity dropped to a mean of 43.9% for model B compared to a mean of 72.1% for model A. These results indicate that algorithms that use rumen temperature data are relatively specific indicating that they may be useful to identify negative individuals whilst the low to moderate sensitivity indicates that these algorithms were less able to identify those positive individuals. Similar studies have reported that other algorithms such as the cumulative sum method can monitor reticulo-rumen temperature recordings to predict respiratory disease in cattle with high specificity. For example, (Timsit, et al., 2011a) reported a positive predictive value of 86% and Voss et al., (2016) reported a 97% specificity. However, results from these studies also indicate that sensitivity is lower. Voss et al., (2016) reported 71% sensitivity at identifying a forthcoming respiratory disease (an alert was triggered a median of 3.5days prior to a high health score being observed) whilst Timsit et al., (2011a) did not report model sensitivity. Cattle in these studies were scored daily limiting comparisons with our study. However, the high specificity but moderate sensitivity reported by these studies supports our proposal that further work is needed to improve the sensitivity of algorithms that aim to predict respiratory disease by use of temperature monitoring.

Results from experiment 2 indicate that further work is needed to develop algorithms that can generalise to new populations. Both models had low to moderate sensitivity (62% for model A and 63% for model B) and low to moderate specificity (79% and 63%) when tested on the extreme dataset. When tested on the ‘all values’ dataset, the sensitivity of model A decreased to 37% and the sensitivity of B decreased from to 42%. The decrease in sensitivity that occurred when steers with more moderate scores were included could indicate that the algorithms developed in this study struggled to identify those steers that are BRD positive but are not amongst the highest scores. This could have implications for the wider use of rumen temperature monitoring, since if the algorithms used are not sufficiently sensitive to identify more mildly affected individuals, this could limit the usefulness of deploying this technology in field where the greatest benefits may come from identifying moderate cases that have been missed by visual observation. It is also worth noting that we only predicted health status once. However, this is not realistic of how precision technologies are used in commercial situations where algorithms are used generate predictions at regular intervals such as once daily (Eckelkamp & Bewley, 2020). Thus, a low sensitivity may be less problematic since repeat testing (i.e., using the algorithm to predict sick steers daily) could enable the improved detection of BRD positive animals. Likewise, algorithms with low or moderate specificity have drawbacks in the context of repeat testing since the number of disease negative animals that are incorrectly detected as positive can rapidly accumulate which may have a negative impact on the user's trust of the technology (Eckelkamp & Bewley, 2020). Further studies are needed to assess the impact of repeat testing on the algorithm's performance.

One limitation of our study is that due to handling restraints we were not able to health score steers at daily intervals (we health scored once a week for experiment 1 and on two separate occasions only for experiment 2). Therefore, we are unable to test if rumen temperature was elevated prior to the appearance of other clinical signs or if steers that were falsely predicted as positive went on to develop respiratory disease. Indeed, it is possible that misclassification occurred around individuals transitioning between disease states. Scoring individuals daily could help assess if those individuals that were falsely predicted as positive may have become positive a few days later and vice versa. Another limitation is that health scoring has low sensitivity and moderate specificity. The low sensitivity of health scoring is particularly problematic since it indicates that there may have been BRD positive steers which were not identified during our health scoring resulting in mislabelling in our datasets. Other studies that have used temperature boluses to detect BRD in cattle indicate that numerous hyperthermic episodes occur in steers that were not diagnosed with BRD. For example, Voss et al., (2016) reported that of the steers that were not diagnosed with BRD experienced reticulo-rumen hyperthermic episodes, 25% of which lasted longer than 31 hours. In addition, there is evidence to suggest that hyperthermic episodes have a negative impact on production, even when these are not associated with any clinical signs. For example, one study reported that beef cattle that experience high rumen temperatures and but were not identified as sick by farm staff experienced reduced weight gain (Timsit et al., 2011b), indicating that these individuals affected by increased temperature may benefit from treatment. Indeed, in the absence of a gold standard for BRD, an elevated temperature alone could be useful indicator of malaise indicating that the animal could benefit from an intervention. Studies are needed that explore the relationship between rumen hyperthermic episodes and a greater range of diagnostic indicators such as lung lesions, immune indicators such as neutrophile lymphocyte ratios, cortisol, and tests to identify and quantify respiratory pathogens. Such research could improve our understanding of the relationship between bovine respiratory disease and reticulo-rumen temperature potentially providing an indication of the temperature thresholds at which a medical intervention may be indicated.

This is the first study to rank temperature-based features for the prediction of BRD in cattle. This is important since it helps detail which features could be useful for future studies. The features with the highest importance in were those that quantified the total area over specific thresholds, indicating that features which quantify the magnitude as well as the duration of increases in reticulo-rumen temperature could be useful for the identification of sick animals. Interestingly, our feature ranking shows that drinking events had high importance in our classifier suggesting that differences in drinking behaviour are predictive of ill health. Since drinking behaviour can be accurately quantified using rumen boluses (Vázquez-Diosdado et al., 2019a), our results suggests that temperature boluses may have advantages over other temperature monitoring methods (such as ocular probes) which are not able to also capture drinking behaviour. One limitation is that whilst we used an individualised baseline to detect deviations specific to drinking behaviour, we did not use this to detect pyrexia. Since individuals differ slightly in their normal baseline temperature (e.g., Burdick et al., 2011), adapting the temperature to the individual (i.e., an animal specific approach) could enable improved classification performance for the prediction of ill health. In addition, there is evidence to suggest that in rumen temperature changes with age (e.g., during the development of rumination in young steers) as well as according to time of day in steers (Voss et al., 2016). Rumen temperature is also affected by diet, with feed high in concentrates resulting in higher rumen temperatures (AlZahal et al., 2008) as well as other factors such as milk production, breed, and days in milk (Bewley et al., 2008;

Liang et al., 2013; Stone et al., 2017). A flexible approach that tailors temperature thresholds to individuals could allow for differences between individuals and changes to rumen temperature over time.

To conclude, our results indicate that reticulo-rumen temperature boluses show some promise for the detection of respiratory disease in cattle, but that further work is needed to improve sensitivity, particularly when clinical symptoms are more moderate. Future studies could focus on the use of pyrexia alone as an indicator of malaise and on improving our understanding of the relationship between reticulo-rumen temperature and the pathogenesis of respiratory disease.

6 Chapter 6 - Machine learning algorithms to classify and quantify multiple behaviours in dairy calves using a sensor: Moving beyond classification in precision livestock

This chapter details the use of machine learning algorithms to monitor behaviour in calves using a sensor. A close derivative of this chapter has been published in *Sensors* (Carslake et al., 2020).

6.1 Introduction

Our ability to assess animal behaviour is a key component of our ability to recognise ill-health and evaluate welfare in domestic livestock (Broom, 1988; Dawkins, 2004). Deviations from normal behaviour can be reflective of pathology, an adaptive response to a health problem, a signal of vigor or of need (Weary et al., 2009). However, the visual assessment of animal behaviour has numerous limitations such as the time, labour and expense needed to observe individual animals. New technologies such as wearable sensors and expert systems are transforming our ability to monitor animal behaviour including livestock (Brown et al., 2013; Cooke et al., 2004). Behavioural data gathered by sensors can be continuously processed by expert systems capable of detecting abnormalities and warning the farmer where interventions are necessary (Berckmans, 2017). In calves, sensors that monitor lying behaviour and step count have been developed (Robert et al., 2009; Trénel et al., 2009), data which can be useful in identifying early signs of ill health in calves (Hanzlicek et al., 2010; Toaff-Rosenstein et al., 2016). However, the accuracies of sensors evaluating other behaviours such as rumination in calves are mixed (Costa et al., 2021) and the simultaneous identification of multiple behaviours requires further research. For example, one commercial sensor reasonably identified lying behaviours in calves compared to visual observations but failed to accurately identify feeding and drinking behaviours (Roland et al., 2018).

Monitoring a wider set of behaviours has been hypothesised to be of greater predictive value for detecting ill health in livestock than a more restricted set of behaviours (Matthews et al., 2016). In calves, behaviours such as activity whilst lying, self-grooming, feeding and locomotor play are likely to be promising indicators for health and welfare. For example, calves inoculated with bacterial lipopolysaccharides have been shown to decrease time spent lying active, increase time spent lying inactive whilst total lying time was not affected (Borderas et al., 2008). Another behaviour of interest is self-grooming. Rats injected with an inflammatory cytokine show a dose dependant reduction in grooming behaviour (Crestani et al., 1991) and grooming behaviour decreases in sick calves (Borderas et al., 2008; Hixson et al., 2018; Toaff-Rosenstein et al., 2016). Alongside rumination and nutritive suckling (i.e., suckling milk at an automatic feeder), feeding behaviours monitored in calves could include non-nutritive suckling at the milk feeder. Non-nutritive visits to the milk feeder have been shown to decrease in sick calves prior to any reduction in overall feed intake (Svensson & Jensen, 2007). Finally, changes in play behaviour could serve as an early indicator of ill health (Fagen, 1981). In calves, painful procedures such as disbudding as well as reduced feed allowance have been associated with reductions in locomotor play behaviour (Krachun et al., 2010; Rushen & de Passillé, 2012).

If an increase or decrease in specific behaviours identified by a sensor is to be used as an indicator of ill health in calves, this approach must be capable of estimating the distribution of each behaviour in an unlabelled dataset. This quantification task seems almost trivial, and researchers have mostly either ignored it (as most studies present mainly a behaviour identification task) or have tried to solve it by simply counting the number of samples predicted as positive by the algorithm i.e., Classify and Count Method (Forman, 2008). However, such an approach fails to consider the fact that positive predictive value decreases with prevalence (Brenner & Gefeller, 1997) and possible differences in behaviour prevalence between the training/test dataset and a new unlabelled dataset. This can result in vast overestimation of low prevalence behaviours. For example, an algorithm developed to identify play behaviour in calves overestimated occurrence by some 200% despite pre-processing to increase the prevalence of positive samples (Größbacher et al., 2020).

Overestimation has also been reported in other low-prevalence behaviours such as movement activity (Trénel et al., 2009) and rumination (Rodrigues et al., 2019) Estimates were improved when the prevalence of these behaviours increased (Rodrigues et al., 2019). The importance and need for quantification methods has been discussed widely in human machine learning tasks (Forman, 2008) but to our knowledge, no such methods have been presented in the precision livestock literature.

Finally, different behaviours may require different sampling frequencies and statistical features in order to be identified (Walton et al., 2018). Sensor sampling rate and feature significantly impact battery life and should therefore be optimised for the behaviours monitored and battery life required (Walton et al., 2018). However, many studies fail to establish which signal features and sampling rate are most appropriate for the behaviours classified.

To address the limitations outlined above we propose a novel approach to sensor-based behaviour monitoring in calves with the following aims:

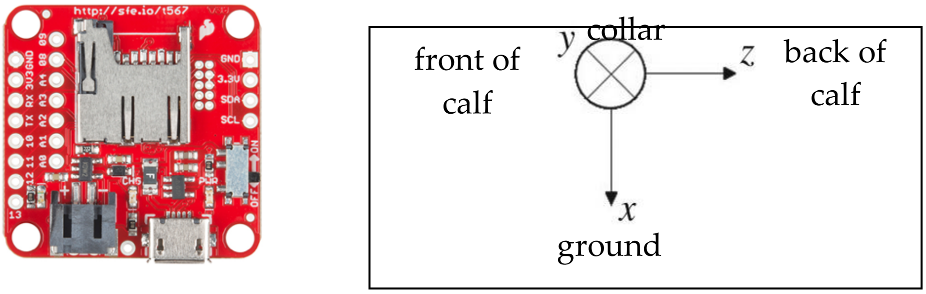
- Create machine learning algorithms to classify two postures (standing and lying) and seven behaviours (locomotor play, self-grooming, active lying, non-active lying, non-nutritive sucking at the automatic feeder, nutritive sucking at the feeder and ruminating) using a single sensor.
- Explore signal feature importance and impact of sampling frequency on classification performance.
- Implement a quantification algorithm to accurately estimate the number of samples of locomotor play behaviour in test dataset with a low prevalence of positive samples.

6.2 Materials and methods

6.2.1 Raw data collection

Thirteen Holstein dairy calves were selected by random number generator from a pen of 20. The calves selected were between 5-7 weeks old and housed in straw-bedded pen (6m X 12m) along with the 7 other (non-trial) calves. An automatic feeder (Forster Technik COMPACT smart) fed calves milk replacer based on an individualised feeding plan and calves had ad-lib access to concentrates, chopped straw and water. The study was conducted at the Centre for Dairy Science Innovation at the University of Nottingham, UK. Ethical permission was obtained for the School of Veterinary Medicine and Science, University of Nottingham (unique reference number 1481 150603).

The study duration was 12 days of which 2 days were used as a pilot for troubleshooting and preparation and 10 days for data collection. During week 1 (21/01/2019 - 25/01/2019) six calves were enrolled in the study. During week 2 (28/01/2019 - 02/02/2019) seven different calves were enrolled. Enrolled calves were caught daily by trained handlers. Each calf was equipped with a neck-worn collar onto which we had previously attached a sensor. The sensor was firstly placed in a lightweight plastic bag before being wrapped in tape and then attached to a collar using plastic cable ties and tape. Each sensor was fixed at the same orientation and location on the collar for consistency (Figure 6.1). Sensors recorded continuously from approximately 16.00 until 13.00hrs the following day when they were removed and replaced with new sensors. Specific sensors used were SparkFun 9 degrees of freedom razor IMU MO sensor (https://www.sparkfun.com/) which combines a SAMD21 microprocessor with an MPU-9250 9DoF sensor. The device was set to record data from a 3-axis accelerometer and a 3-axis gyroscope. Sampling rate was set to 100Hz with a range of ±8g and gyroscope range was 2000°/s.



(a) (b)

Figure 6.1. Photo of sensor used (a) sensor orientation on the collar (b).

6.2.2 Behavioural observations

Calf behaviour was recorded using four video cameras (5Mp, 30m IR. Hikvision Digital Technology Co., Ltd., California, USA). Three cameras were mounted on the walls of the pen at 3-4m of height and the fourth was mounted on a tripod overlooking the automatic milk feeder. The cameras were oriented to ensure maximum cover of the pen. The cameras were set to record at high quality video (HEVC, H.265; and at 2944x1656 pixels quality) and 30 frames/s. Cameras were connected to a 4MB video recorder (Hikvision Digital Co., Ltd., California, USA) from which data was retrieved using an external hard drive. Definitions for postures and behaviours for enrolled calves were recorded by three trained observers using the video recordings according to the ethogram shown in Table 6.1. Precise time stamp (start and stop) of postures and behaviours were recorded manually. A reliability test showed on average good to high agreement between observers (kappa > 0.7-0.9). Video footage was labelled between 16.00 – 20.00hrs daily and only behaviours with a duration of more than 3 seconds were recorded. Each behaviour was labelled for a maximum of one hour per calf. The exception to this was locomotor play behaviour where all instances were labelled.

Table 6.1. Definition of the calf postures and behaviours used for the classification.

Posture states	**Description**
Lying	Calf lying down on sternum or side, body to the floor
Standing	Calf is standing and may be moving one or more limbs in a forward or backwards motion.
Behaviour states	**Description**
Non-active lying	Calf lying down on the sternum or side, body on the floor with head not moving for more than 3s.
Active lying	Calf is lying down and with the head lifted from the ground, supported by the neck and moving.
Ruminating	Calf is lying down and show regular jaw movements interrupted by regurgitation and swallow cycles with the head remaining in a constant position.
Self-grooming	All self-grooming movements where tongue is visible across body surface.
Nutritive suckling	Calf is standing in milk feeder, holds teat in his/her mouth and makes swallowing movements. The automatic feeder dispenses milk (milk flows through tube visible on video).
Non-nutritive suckling	Calf is standing in milk feeder, regularly (< every 3s) holds teat in his/her mouth. The automatic feeder does not dispense any milk (milk does not flow through tube visible on video).
Locomotor play	Rapid forward movement that lasts 3 s or longer (in real time) and could include instances of jumping or bucking. It includes all instances of trotting (two beat leg movements synchronized diagonally), cantering (three-beat gait in between a trot and a gallop) and galloping (four-beat gait with a phase where all legs are off the ground).

6.2.3 Data processing

Merging of the behavioural observations and raw sensor data according to timestamp was performed using custom made scripts written in Python 3.5. Visualization of recorded accelerometer magnitude alongside the associated video recording was performed for each sensor recording to check any possible delays between video camera and sensor data due to sensor desynchronization. Any delays due to time stamp desynchronization were corrected. Data for the first four hours of sensor recording was used for the analysis.

Individual data files of both sensor and labelled data were discretised into windows of equal length. In this study windows sizes of 1s-10s with a 50% overlap were explored (Bulling et al., 2014). The set of feature characteristics was extracted from the magnitude of the acceleration and the magnitude of the gyroscope which are defined by $\bar{A} = \sqrt{A_x^2 + A_y^2 + A_z^2}$ and $\bar{G} = \sqrt{G_x^2 + G_y^2 + G_z^2}$, respectively where Ax, Ay, Az, Gx, Gy, Gz represent the acceleration and gyroscope signals at the axes x, y, z, respectively. Forty-four feature characteristics were computed using a set of previously defined statistical features (Walton et al., 2018). Full details of the definition and formula of the feature characteristics can be found in (Walton et al., 2018).

For the classification algorithm, merged data contained all labelled behaviours and the sensor features (Dataset 1). For the quantification algorithm, we merged sensor features characteristics from both labelled play and non-labelled play behaviour data (Dataset 2).

6.2.4 Classification algorithm

An AdaBoost ensemble learning algorithm (Rokach, 2010) was implemented using the fitcensemble function in Matlab 2019. The AdaBoost algorithm learner was set to have a minimum leaf size of 5 and maximum number of splits per tree. Classification performance for postures and activities was evaluated using a 5-fold cross validation which is a commonly used technique for robust evaluation of performance in classification (Kohavi, 1995). Within this technique the original dataset was split into 5 subsets of equal size, and a total of 5 iterations are performed. At each iteration, 4 subsets are used to train the classification algorithm and the remaining one is held back to test. At each fold, performance values are computed using the test set and the average of these are used to represent the performance of the cross validation. As Dataset 1 was relatively well balanced for posture (i.e., standing and lying), there were 81492 samples at 3s for lying and 31951 samples with 3second (3s) windows for standing, no further processing was required.

However, for the different behaviours an under-sampling balancing technique was applied to address the problem of inter-class imbalance (He & Garcia, 2009) i.e. behavioural classes are not equally represented due to the nature of the different activities. For example, locomotor play behaviour occurred rarely compared to non-active lying. The technique used the total number of samples of locomotor play behaviour as a measure to select samples for each individual behaviour across individual files. More precisely, let $\alpha_{i,b}^{im}$ represent the number of samples of behaviour b collected for individual dataset i over the original imbalanced dataset (im). Hence, the original imbalanced dataset $\{\alpha_{i,b}^{im}\}$ was balanced according to:

$$\alpha_{i,b}^{bal} = \begin{cases} randomly\ selected\ a\ total\ of\ {}^{s_{min}}/_{k_b}\ samples\ from\ \{\alpha_{i,b}^{im}\}\ if\ \alpha_{i,b}^{im} > {}^{s_{min}}/_{k_b} \\ \qquad\qquad |\alpha_{i,b}^{im}| \quad if \qquad \alpha_{i,b}^{im} \leq {}^{s_{min}}/_{k_b} \end{cases}$$

Where $|\alpha_{i,b}^{im}|$ represents the cardinality (number of samples) of behaviour b collected from the individual datafile i in the original dataset and k_b ($0 < k_b \leq N$, where N is the total number of datafiles) that contain samples on behaviour b and s_{min} is the minimum number of samples to balance the data.

This method ensures a balanced data for each individual across all the different behaviours, since the number of data samples per individual per class will be equals to or ${}^{s_{play}}/_{k_b}$ or $|\alpha_{i,b}^{im}|$. After balancing the data, the total number of 3s window samples was 396 for active-lying, 396 for non-active lying, 396 for ruminating, 392 for non-nutritive suckling, 396 for nutritive suckling, 272 for self-grooming and 393 for locomotor play. Within this evaluation, performance was assessed using metrics which included overall accuracy, precision, recall, F-score and Cohen’s Kappa (Ben-David, 2008) as defined in (Dohoo et al., 2003).

6.2.5 Quantification algorithm

An adjusted count (AC) method with a maximum selection threshold as described in (Forman 2008) was implemented using Dataset 2. All instances of locomotor play behaviour (P) were labelled for Dataset 2 and hence any sample that does not have a label can be consider as non-play (NP).

The AC algorithm is a two-step algorithm that corrects the estimate provided by a binary classifier using its true positive rate (tpr) and false positive rate (fpr). AC algorithm first trains a binary classifier and then estimates tpr = true positives(tp)/(true positives(tp) +false negatives(fn)) and fpr = false positives(fp)/(true negative(tn) +false positives(fp)) by means of a cross-validation over the training set. In the second step, AC corrects the prevalence of an unknown sample using the following formula:

$$p' = \frac{p_0' - fpr}{tpr - fpr}$$

where p_0' is the initial estimate of prevalence from the prediction of the classifier, p' is the adjusted prevalence and fpr and tpr are as previously described. When applied to highly imbalanced datasets the performance of the AC method degrades quickly. Class imbalance can be solved by selecting a threshold that maximises the $tpr - fpr$ (denominator of formula 1) over a varying range of training conditions.

When applying the AC method for the quantification of play behaviour we first split *Dataset 2* into two subsets of equal size: a training subset and a test subset. From the training subset a varying range of training conditions was generated by randomly selecting P = 10, 20, 30,…, 190 positive play instances and NP = 10000 non play instances for the training subset. The total number of positive play cases was 196 and negative cases was 70983.

For each value of the varying range of training condition we estimated tpr and fpr characteristics via a 5-fold cross validation on a binary ensemble classifier algorithm for play and non-play. The threshold was selected using the training conditions that maximised $tpr - fpr$. Afterwards, initial training conditions that maximised the threshold were used to train a binary ensemble classification algorithm. The algorithm was then used to predict the number of play behaviour samples over the test subset. Predictions made by the classification algorithm were adjusted according to the above formula.

Since all instances of play behaviour were labelled in *Dataset 2* it was possible to compare the adjusted number of window samples predicted as play with the total of samples observed as play behaviour via non-parametric correlation.

Additionally, the total number of over or under estimation of instances of play behaviour were computed. Similarly, the number of instances of over/underestimation was computed for each individual data file.

6.2.6 Feature ranking and down-sampling

Ranking of the 44 feature characteristics was obtained using ReliefF feature selection (Kononenko, 1994) for the classification of posture and behaviours.
We investigated the effect that different sampling frequencies can have on the performance of the classification. This was achieved by down-sampling the data originally sampled at 100Hz to sample frequencies of 50Hz, 20Hz, 10Hz and 4Hz. Down-sampling was performed by selecting a subset of the original raw dataset as follows:

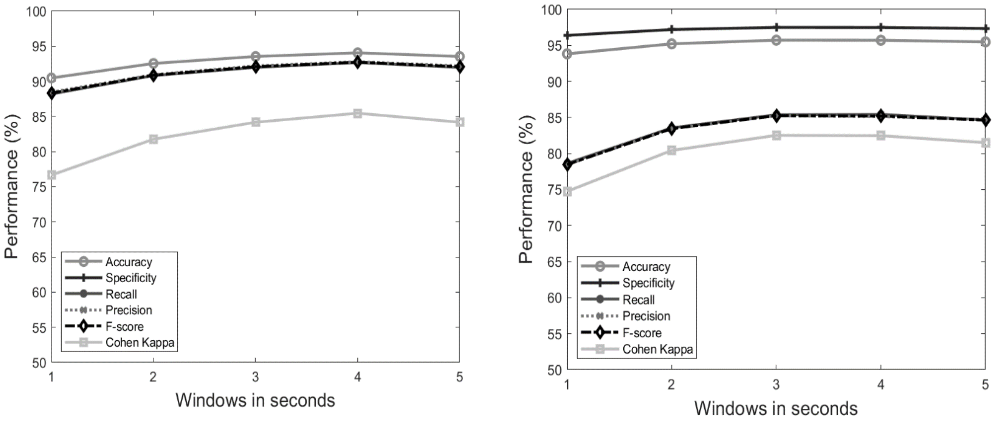
$\{\boldsymbol{a}_{2*i+1}\}_{i=0}^{i=N/2}$ for down-sampling to 50Hz.
$\{\boldsymbol{a}_{5*i+1}\}_{i=0}^{i=N/5}$ for down-sampling to 20Hz.
$\{\boldsymbol{a}_{10*i+1}\}_{i=0}^{i=N/10}$ for down-sampling to 10Hz.
$\{\boldsymbol{a}_{25*i+1}\}_{i=0}^{i=N/25}$ for down-sampling to 4Hz.

Where N is the total number of samples at 100Hz. After data was down-sampled an ensemble classification algorithm for postures and behaviours was generated and assessed using a 5-fold cross validation in the same manner as previously described previously.

6.3 Results

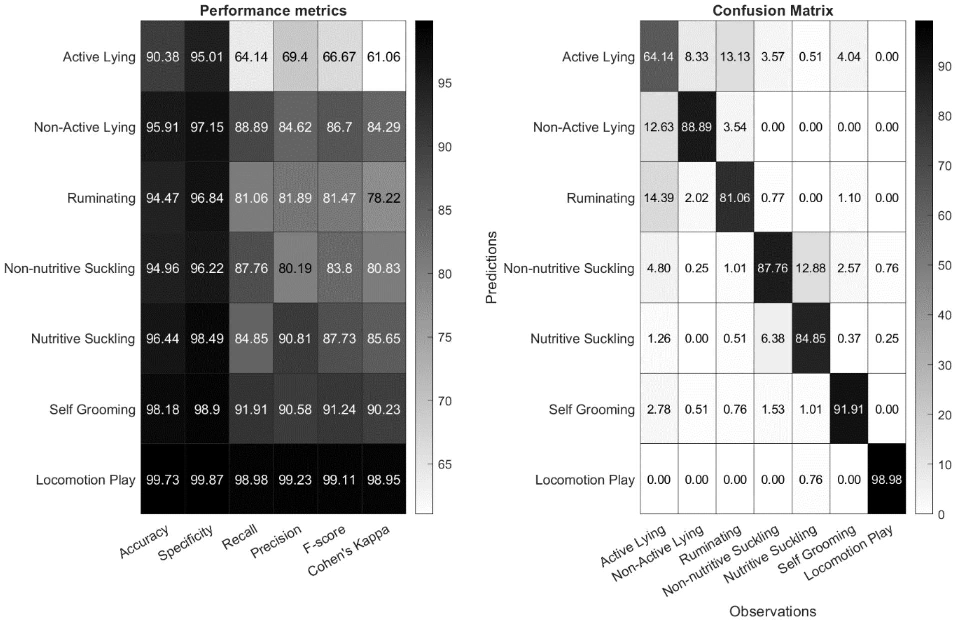
6.3.1 Classification results

An initial comparison of the performance of the classification across windows sizes 1-5s were investigated for both posture and behaviours. Results using the different performance metrics are shown in Figure 6.2. The best results for posture were found when using a 4s window as observed from Figure 6.2, with a 94.38% overall accuracy, 92.99% specificity, 92.99% recall, 93.11% precision, 93.05% F-score and a Cohen’s Kappa of 0.8611. The best results for behaviour were using a window of 3s as shown in Figure 6.2, providing an 95.72% overall accuracy, 97.46% specificity, 85.36% recall, 85.24% precision, 85.24% F-score and Cohen's Kappa of 0.8247. Detailed results for classification performance and the confusion matrix using a 3s window for the behaviours are presented in Figure 6.2 and Figure 6.3 respectively.



(a) (b)

Figure 6.2. Performance (%) of the classifier for postures (a) and behaviours (b) as described in Table 1 across different window sizes (1-5s). Metrics are computed as the mean of the postures/behaviours.



Performance metrics

	Accuracy	Specificity	Recall	Precision	F-score	Cohen's Kappa
Active Lying	90.38	95.01	64.14	69.4	66.67	61.06
Non-Active Lying	95.91	97.15	88.89	84.62	86.7	84.29
Ruminating	94.47	96.84	81.06	81.89	81.47	78.22
Non-nutritive Suckling	94.96	96.22	87.76	80.19	83.8	80.83
Nutritive Suckling	96.44	98.49	84.85	90.81	87.73	85.65
Self Grooming	98.18	98.9	91.91	90.58	91.24	90.23
Locomotion Play	99.73	99.87	98.98	99.23	99.11	98.95

Confusion Matrix

Predictions \ Observations	Active Lying	Non-Active Lying	Ruminating	Non-nutritive Suckling	Nutritive Suckling	Self Grooming	Locomotion Play
Active Lying	64.14	8.33	13.13	3.57	0.51	4.04	0.00
Non-Active Lying	12.63	88.89	3.54	0.00	0.00	0.00	0.00
Ruminating	14.39	2.02	81.06	0.77	0.00	1.10	0.00
Non-nutritive Suckling	4.80	0.25	1.01	87.76	12.88	2.57	0.76
Nutritive Suckling	1.26	0.00	0.51	6.38	84.85	0.37	0.25
Self Grooming	2.78	0.51	0.76	1.53	1.01	91.91	0.00
Locomotion Play	0.00	0.00	0.00	0.00	0.76	0.00	98.98

(a) (b)

Figure 6.3. Classification performance metrics for behaviours (a) and confusion matrix (b) shown as a percentage, the darker the shading the higher the performance. Results shown in these figures were computed using a 3s window size.

6.3.2 Feature ranking and down-sampling

Results of the top 10 feature ranking are shown in Table 6.2.

Postures		Behaviours
Rank	Feature characteristics	Feature characteristic
1	Minimum	Difference Zero crossing
2	First quantile*	Zero Crossings
3	Minimum*	Kurtosis
4	Difference Kurtosis	Difference Zero Crossing*
5	Difference Spectral Entropy*	Zero Crossing*
6	Mean	Min
7	Signal Area	Difference Spectral Entropy*
8	Difference Zero Crossing	Kurtosis
9	Difference Zero Crossing	Difference Kurtosis
10	Spectral Entropy*	Signal Area*

Table 6.2. Top 10 ranked features using ReliefF algorithm for both postures and behaviours. Grey shading and no grey shading represent acceleration magnitude difference-based features and gyroscope difference-based features respectively. Frequency domain-based features can be differentiated from time domain features by the presence of an asterisk (*).

The effect of down-sampling to the frequencies 50Hz, 10Hz, 10Hz and 4HZ is shown in Figure 6.4 and which exhibits the average decrease in performance (accuracy, specificity, recall, precision, and f-score).

A decrease in performance was obtained with a decrease in sample frequency (i.e., a decrease of 0.34% (± 0.50%sd) in accuracy from 100Hz to 50Hz and a decrease of 3.94% ± 2.34% s.d) in accuracy from 100Hz to 4Hz. However, the largest decreases were obtained in recall (13.92% ± 9% s.d), in precision (14.27% ±7.62% s.d) and in F-score (14.18% ± 8.15% s.d) when down-sampled to 4Hz. The percentage decrease in performance when sampling is detailed by behaviour in Table 6.3. Play behaviour was the least affected by down-sampling (decrease in F_score by only 1% when down-sampled to 4Hz) whilst performance statistics for active lying, ruminating, non-nutritive suckling, nutritive suckling, and self-grooming were more severely affected (decrease in F_score by 19.89%, 17.21%, 19.73%, 21.41% and 15.64%) when downsampled from 100Hz to 4Hz. Non active lying was only moderately affected when down sampling to 4Hz (4.79% decrease in F_score).

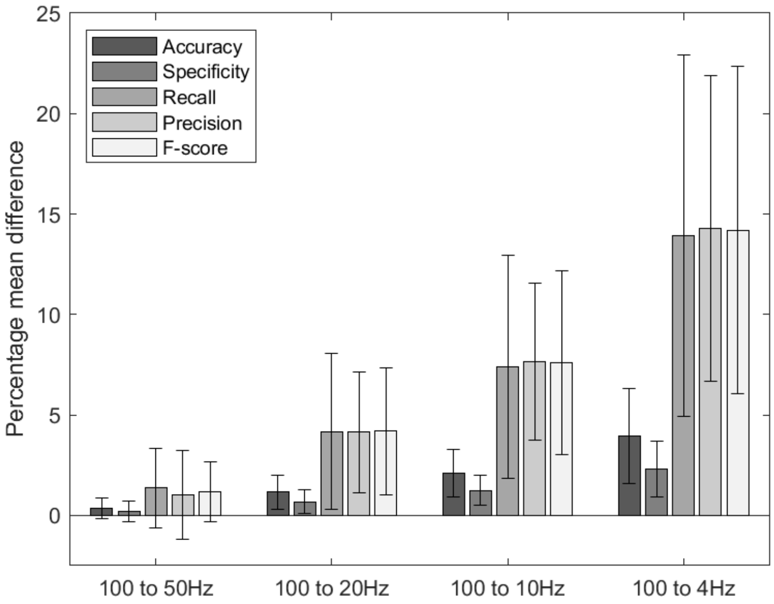


Figure 6.4. Decrease of the performance when down sampling from 100Hz to 50Hz, from 100Hz to 20Hz, from 100Hz to 10Hz and from 100Hz to 4Hz. The bars show the average decrease across all the different behaviours, the black error bars show the variation in reduction (s.d) across the different behaviours.

6.3.3 Quantification results

Quantification of play behaviour using Dataset 2 and the AC method is shown in Figure 6.5 where the number of window samples predicted as play vs the number of observed number of play windows is shown. A significant positive correlation of 0.97 ($p < 0.001$) was obtained between observed and predicted play behaviour. The total number windows 72377 of which 195 were play (0.27%) versus 232 predicted, providing a total overestimation of 37 (18.97%).

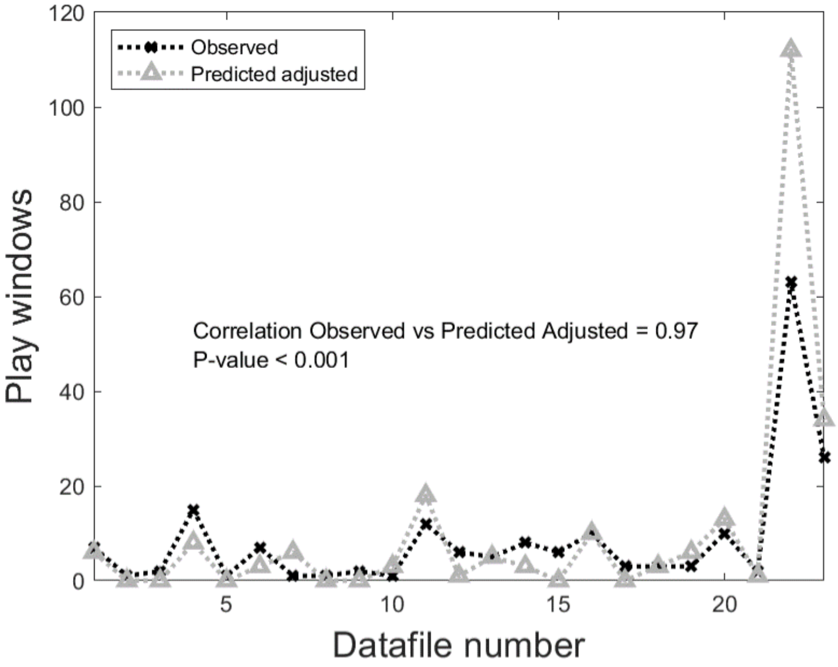


Figure 6.5. Comparison of the number of window samples predicted and the number of window samples observed for each individual dataset.

Sampling frequency (Hz)	50	20	10	4
	% decrease in Accuracy from 100Hz			
Active Lying	1.33	2.16	3.37	4.78
Non-Active Lying	0.04	0.49	1.63	1.67
Ruminating	0.49	2.16	2.20	5.35
Non-nutritive Suckling	0.42	1.55	2.96	6.14
Nutritive Suckling	0.00	0.87	2.99	6.33
Self-Grooming	0.42	0.98	1.55	3.15
Locomotor Play	0.00	0.00	0.00	0.19
	% decrease in Specificity from 100Hz			
Active Lying	1.34	0.67	0.89	1.74
Non-Active Lying	0.13	0.53	1.16	1.61
Ruminating	0.22	1.83	1.92	3.48
Non-nutritive Suckling	0.09	0.84	1.96	3.61
Nutritive Suckling	0.09	0.67	1.92	3.92
Self-Grooming	0.00	0.30	0.76	1.60
Locomotor Play	0.00	0.00	0.00	0.00
	% decrease in Recall from 100Hz			
Active Lying	1.26	10.61	17.42	21.97
Non-Active Lying	1.01	0.25	4.29	2.02
Ruminating	2.02	4.04	3.79	15.91
Non-nutritive Suckling	2.30	5.61	8.67	20.66
Nutritive Suckling	0.00	2.02	9.09	19.95
Self-Grooming	4.78	6.99	8.52	16.73
Locomotor Play	0.00	0.00	0.00	0.25
	% decrease in Precision from 100Hz			
Active Lying	5.72	6.86	11.04	16.88
Non-Active Lying	0.00	2.41	5.79	7.14
Ruminating	1.43	8.75	9.03	18.50
Non-nutritive Suckling	0.80	4.60	9.57	18.88
Nutritive Suckling	0.39	3.81	11.24	23.00
Self-Grooming	0.00	3.08	6.88	14.55
Locomotor Play	0.00	0.00	0.00	1.01
	% decrease in F-score from 100Hz			
Active Lying	3.39	8.98	14.77	19.89
Non-Active Lying	0.24	1.40	5.09	4.79
Ruminating	1.73	6.44	6.47	17.21
Non-nutritive Suckling	1.49	5.07	9.19	19.73
Nutritive Suckling	0	2.86	10.11	21.41
Self-Grooming	2.31	5.05	7.69	15.64
Locomotor Play	0	0	0	0.63

Table 6.3. Percentage decrease in algorithm performance by behaviour when down-sampling from 100Hz to 50Hz, from 100Hz to 20Hz, from 100Hz to 10Hz and from 100Hz to 4Hz.

6.4 Discussion

To the authors' knowledge this the first study to develop an algorithm that can successfully identify such a diverse range of behaviours in calves using a sensor. In addition, this is also first attempt in precision livestock research to develop and present a machine learning quantification algorithm. Our results demonstrate that signal data from a collar-based movement sensor can be used to accurately classify lying and standing posture whilst simultaneously identifying locomotor play, self-grooming, active lying, inactive lying and different feeding behaviours. Several of these behaviours have not been previously identified using a sensor such as self-grooming and non-nutritive suckling. Our algorithm also reported high level of accuracies for most behaviours. Previous research using collar-based accelerometers has reported high levels of misclassification between feeding behaviours and posture in dairy cattle (Martiskainen et al., 2009; Vázquez Diosdado et al., 2015) and in sheep (Barwick et al., 2018). One explanation for the high performance of our algorithm is the inclusion of both gyroscope and accelerometer signal based features, which were evenly ranked in the top ten features (Table 2). This finding reflects research in sheep showing that improved behavioural classification performance can be achieved when gyroscope features are included (Walton et al., 2018). Active lying was the behaviour with the overall worst performance (90% accuracy, 64% sensitivity and 69% precision). It was more regularly confused with similar behaviours such as non-active lying and ruminating (Figure 6.4). Misclassification between different resting behaviours (i.e., lying awake and NREM sleep) has previously been reported in calves (Hokkanen et al., 2011). It is possible that the sensor may not be sufficiently sensitive to consistently detect the slight movements of active lying or differentiate active lying from ruminating. Additionally, a different behaviour was only assigned if it lasted longer than 3s, it is therefore possible that short periods of non-movement during episodes of active lying were sufficient allow misclassification as non-active lying. Another problem could be that the definition of active lying was broad (lying with head moving – see Table 1). Further differentiation of activities during active lying (i.e. chewing forage and social-grooming) could aid in reducing potential misclassification with similar behaviours (i.e. ruminating and self-grooming) and identify other interesting behaviours.

One behaviour of interest is locomotor play which was identified with 98.98% sensitivity, 99.73% specificity and 99.23% precision by our classification algorithm. Recent studies have explored the potential of commercially available leg worn sensors, to identify play behaviour in calves. One study used a summary acceleration data feature (motion index) to predict whether play was present or absent within predefined sampling periods (1min or 15min) (Gladden et al., 2020). The optimised threshold was an accurate predictor of whether play occurred or not in each 15min. However, this approach does not allow metrics such as the behaviour's duration and frequency to be calculated as multiple behaviours can occur within the same sampling period. Another approach has been to use raw accelerometer recordings of play behaviour in order to train and test a classifier (Größbacher et al., 2020). Whilst the classifier's predictions for play behaviour were correlated with observations (Pearson $r = 0.87$) predictions overestimated play occurrence by approximately 200% when predicting on a subset of the full dataset where locomotor play behaviour had low prevalence (6.5%). To address this problem, we implemented an adjusted count quantification algorithm on a binary classifier (Forman, 2008) using locomotor play behaviour as an example. Our quantification algorithm provided a high level of correlation with real observations (0.97; $p < 0.0001$). Our results show promise since play behaviour was only overestimated by 19% despite its very low prevalence (0.27%). This is significant improvement from any published work in the field so far. Classifiers are not perfect and the test class distributions (i.e., behaviour distributions in real world) are not generally representative of training dataset. We believe this offers a method for researchers to apply quantification for behavioural monitoring in livestock. In our current work we were limited by the complete labelled data being only available for locomotor play behaviour. To deliver a complete framework that can accurately monitor a larger number of low prevalence behaviours a multi-class quantification algorithm needs to be implemented.

We should highlight that our neither our proposed quantification algorithm nor classification algorithm addresses changes that could occur due to ‘concept drift’, this is where the feature space may change and algorithm performance may be altered (Fawcett & Flach, 2005). Concept drift occurs where a sensor which has been developed in a particular environment under-performs in a new environment. In precision livestock this could be due to differences in the animals (age, breed, etc.) and environment characteristics (elevation, soil, particular farm constraints, etc.) Further studies could explore algorithm verification in new environments to confirm the accuracies reported.

Our study also explored which sampling frequencies were most appropriate for our classification algorithm. A small decrease in the performance metrics occurred when we down sampled to 50Hz and a more significant decrease when down sampling to 4Hz (i.e., 14.18% decrease in F-score) (Figure 6.4). Interestingly there was minimal decrease in classification performance for locomotor play and non-active lying even when downsampled to 4Hz. This is likely due to the characteristic high amplitude acceleration pattern of locomotor play behaviour and the low amplitude pattern for non-active lying indicating that these can be differentiated from other behaviours even at lower sampling frequencies. Active lying, ruminating, self-grooming, nutritive suckling, and non-nutritive suckling suffered more significant decreases in performance when down sampled to 4Hz. This indicates the necessity to identify more complex signal patterns to classify these behaviours. If sampled at a low frequency one approach to improve classification performance could be to increase window size and thereby increase the likelihood of identifying a characteristic signal pattern. Sampling between 20Hz and 10Hz is a good compromise between algorithm performance and battery life dependent on the intended application. This confirms previous findings using a similar sensor in sheep which recommended a sampling frequency of 16Hz (Walton et al., 2018).

Finally, behaviours identified in this study were chosen for their relevance to calf health and welfare as well as for regular occurrence in pre-weaned calves and relative ease of labelling. Behaviours that could be incorporated by future studies include further drinking (Vázquez-Diosdado et al., 2019a) and feeding behaviours (Mansbridge et al., 2018) (e.g., drinking from water trough, eating roughage, eating concentrate) other resting behaviours (i.e. REM and NREM sleep, non-active standing), walking, social grooming, stereotypical behaviours such as navel sucking as well as possible health indicators such as coughing or laboured breathing. Considerations include sampling frequency required, the need to carefully define behaviours, the labour requirements to label behaviours and need to implement multi-class quantification methods for those behaviours that occur at low prevalence.

7 Chapter 7 – Machine vision for monitoring calf behaviour

7.1 Introduction

Sensors which are based on movement detection to monitor behaviour offer the opportunity to monitor a diverse range of behaviours in calves which could provide insights for health and welfare (Carslake et al., 2020). However, sensors worn by livestock have some important limitations such as the cost of each unit and the need to strike a trade-off between sampling frequency, computational power, and battery life. Limited battery life can make it difficult to monitor more complex behaviours since these behaviours necessitate higher sampling frequencies. For example, one study in calves showed that complex behaviours such as self-grooming and different resting behaviours can be identified at 16Hz but that there are high levels of misclassification between these behaviours when sampling frequency is reduced (Carslake et al., 2020). Similar trade-offs between behaviour complexity and sampling rate have been reported in sheep (Walton et al., 2018). In addition, many interesting behaviours such as interactions with objects in the environment or interactions between individuals require contextual information from the scene, meaning it may not be possible to capture them using movement sensors alone.

Computer vision has been proposed as a different approach to monitoring livestock behaviour. Computer vision aims to extract information from a scene by processing corresponding images or videos. For example, processing techniques such as segmentation can be used to separate the animal from its background. From these, features such as measures of movements of a segmented animal across a scene can be used to predict a particular behaviour such as walking or running using machine learning techniques. In cattle, one of the first examples of computer vision was able to identify standing, lying, and feeding with 87%, 92% and 86% accuracy respectively by using image segmentation and feature extractions (Porto et al., 2013, 2015.) However, the process of segmentation and feature extractions requires choices to be made regarding which features to extract. By defining features manually this process runs the risk that selected features may be specific to the task in hand which can result in low accuracy when tested in a new situation where the feature space may change (e.g., different background, illumination, camera angle…) (Chen et al., 2021). Generalisation is an important consideration livestock behavioural monitoring as farm environments can be very diverse, e.g., they can be indoors, outdoors, have different pen designs, group sizes, bedding substrate etc. Livestock also differ in appearance according to breed and life stage. As such, there is a need to for computer vision to work in variety of situations further increasing the difficulty of applying computer vision techniques to monitor livestock.

A different approach that has garnered much attention in computer vision for livestock behaviour detection is deep learning (Borges Oliveira et al., 2021). Deep learning does not require the computation of specific features but instead uses the entire dataset to create patterns and inferences replacing the need to manually define features. One example of a deep learning vision algorithm that has been used for behaviour detection in livestock is YOLO (You Only Look Once). YOLO is an object detection deep learning algorithm which uses a single end-to-end neural network to predict and localise an object in an image (Redmon et al., 2015). YOLO's basic approach consists of splitting each image into grids composed of cells (Figure 7.1). Each cell is used to predict bounding boxes that could contain an object of interest. Each box is assigned a confidence score which corresponds to the model's probability estimate that the cell contains the object of interest. Each cell also predicts class probabilities for those boxes (i.e., the class of the object and the probability of that the prediction is correct). The confidence scores for the box and the class probability are combined giving bounding boxes that are weighted by the combined probability. Only those bounding boxes with the highest combined probability are taken forward leaving the final decision. This process takes place in a single, connected layer of neurons and results in algorithms that can make inferences from videos in real time (i.e., up to 45 frames per second for YOLO version 5 using conventional hardware).

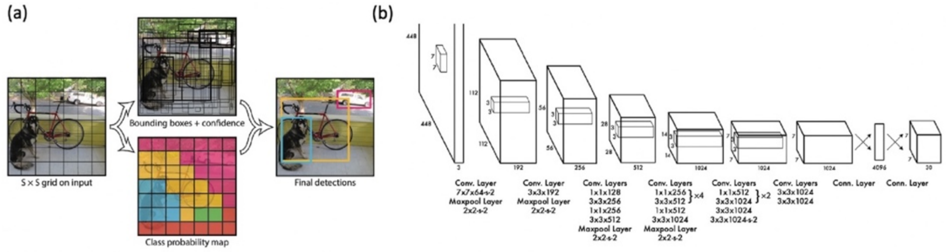


Fig.7.1. YOLO architecture. Images from Redmon et al. (2015).

The fast inference time of YOLO indicates that it could be useful applications in livestock behaviour monitoring where continuous monitoring may be required and computational power may be limited. YOLO has been used to detect different behaviours in different livestock species. For example, Alameer et al., (2020) detected 5 different behaviours in pigs, Jiang et al., (2020) accurately classified eating, drinking and resting behaviours in goats. Fuentes et al., (2020) detected 15 different behaviours, including self-grooming, feeding and different social behaviours in cattle. One study has implemented machine vison to detect calf behaviour (Y. Guo et al., 2020). However, this study only included one calf per scene and did not identify interesting behaviours such as locomotor play, rest-quality, grooming or social interactions. In addition, most vision behaviour detection studies in livestock have only used data where training and test data have similar characteristics to the test data (e.g., Alameer et al., 2020; Fuentes et al., 2020; Jiang et al., 2020). This is an important consideration for object detections algorithms as when training and test data have different characteristics, model performance for object detection tends to be reduced. Calves are housed in diverse environments (group housing, pair housing, straw bedded, indoors, out-doors) and the same scenario used for training is unlikely to be replicated across farms. Research is needed to explore how changes to the test data (such as a different camera angle) may impact model performance to gain insight into the generalisability of these algorithms to new situations.

In this study we aim to evaluate the applicability of the YOLOv5 object detection algorithm to monitor diverse behaviours in group housed dairy calves. To achieve this objective, we collected video data from a single pen of 20 group housed dairy calves which was filmed from three different angles and labelled 17 different behaviours. Next, we trained our model to detect the behaviours of interest and validated it using footage filmed from the same angles as the training dataset. Finally, to test how well the model was able to generalise, we trained and our model on footage taken from two different angles and then tested on a footage taken from the same scene but from a new, previously unseen angle. Our discussion also includes some of the management opportunities that could arise from the successful deployment machine vision algorithms in the field.

7.2 Materials and methods

Data collection

Data collection for this experiment took place at the Centre for Dairy Science and Innovation, University of Nottingham, UK. 20 Holstein dairy calves between 5 and 7 weeks old were housed in a single straw bedded pen (6m x 12m). Calves had continuous access to a feeding station which was operated by a computerised feeder (Forster Technik COMPACT smart), chopped straw, concentrates and water. Ethical permission was obtained for the School of Veterinary Medicine and Science, University of Nottingham (unique reference number 1481 150603).

Calf behaviour in the group pen was recorded using 3 video cameras, these were positioned to give an overview of the pen (Figure 7.2). The cameras used were 5 MP fixed bullet cameras (Hikvision Digital Technology Co., Ltd., Los Angeles, CA, USA) which were connected via ethernet cables to a 4TB video recorder (Hikvision Digital Co., Ltd., CA, Los Angeles, USA). The 3 cameras that were focused on the pen recorded from different locations, thus the 3 cameras gave three different angles (angles 1, 2 and 3) of the same scene. Each camera was secured at approximately 3.5m from the ground and orientated downwards towards the pen to give an overview of the pen. The cameras recorded continuously in high efficiency video coding (HEVC), 2944 x 1656 resolution, at 30frames/second for two weeks between 21/01/2019 and 02/02/2019. Video footage from this period was exported from the CCTV's internal hard drive to a USB storage device.

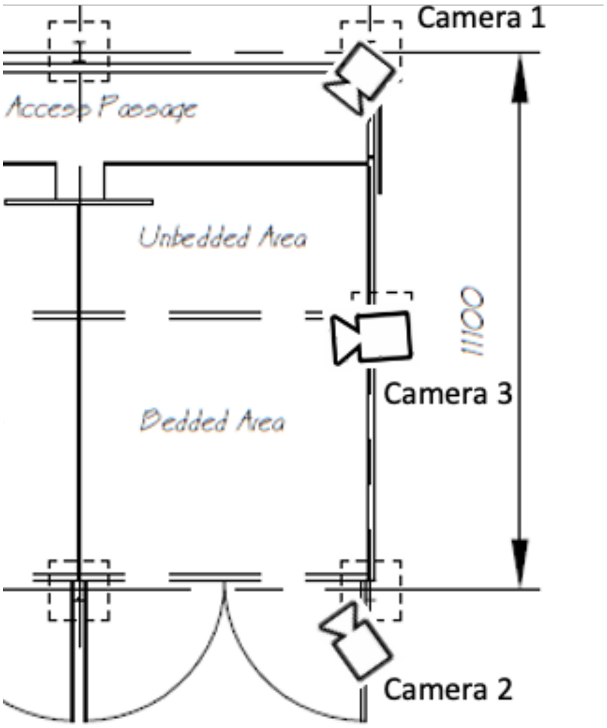


Figure 7.2. Calf group pen showing the location of the three cameras.

Dataset construction

Instances where calves were displaying behaviours of interest such as locomotor play, and grooming behaviours as well as different resting behaviours were selected from the footage using observations of the video with the software VSPlayer_x64 (Hikvision Digital Technology). Short videos (between 30 seconds and 3mins; n = 30) were created by trimming the original video to focus on these time periods where calves were displaying behaviours of interest. Videos were chosen from daylight hours only. The trimmed videos were from 3 different angles (angles 1-3), from each video, 10 frames were extracted which resulting in a total of 300 frames.

Behaviour labelling

An ethogram was constructed to represent 17 behaviours of interest (Table 7.1). These were chosen to represent postures (lying, standing, moving) as well as more detailed behaviours (i.e., active lying versus non active lying). The behaviours also included interactions between calves (e.g., allo-grooming) and interactions between calves and their environment (e.g., standing at the milk feeding station) which require information from a scene (e.g., the calf's proximity to the feeding station). For each frame in the training data, labelling involves a human observer placing a box around the area of interest and assigning it according to the behaviour of interest. We used the software labelImg (Tzutalin, 2015) to label our frames. All calves and behaviours visible in the 300 frames were labelled.

Data augmentation was applied to the dataset within the YOLO framework (Jocher, 2022). Data augmentation for YOLO involves flipping each image such that new frames are images are a mirror reflection of the original frame (e.g., flipping each image vertically, horizontally, and vertically and horizontally), rotating images and adding blur (replacing each pixel with the median value of its neighbouring pixels) (Redmon et al., 2015). This process enables the generation a larger number of frames from those that have already been labelled, with the objective of making the training data more diverse and improving the generalisability of the resulting algorithm.

Model training and testing

The specific algorithm used was YOLOv5. YOLO algorithms are pre-trained using the COCO dataset (Lin et al., 2014) and have been widely used for object detection problems in machine vision. YOLOv5 was chosen as there is evidence to suggest that it is computationally efficient, allowing for real-time inferences, and that version 5 is both faster and more accurate compared to previous versions (Jocher, 2022).

The data selection process is detailed in Figure 7.3. To create training and validation dataset we split labelled behaviours that were gathered collected from angles 1 and 2 into a training and validation sets using an 80/20 split. We trained the algorithm on the training set and tested it on the validation set.

Table 7.1. Definitions of behaviours labelled

Behaviour class	Abbreviation	Description
Active lying	L_AL	Lying down on the sternum or side, body to the floor. Head lifted from the ground supported by the neck.
Lying, self-grooming	L_SG	Calf is grooming itself (movements with tongue over body surface)
Lying, allo-grooming	L_AG	Calf is grooming another calf (movements with tongue over body surface)
Lying, oral Behaviours calf	L_OBC	Calf's mouth is in direct contact with another calf. Behaviour to include biting, sucking other calves.
Lying, oral behaviour object	L_OBO	Behaviour to include biting, sucking, licking at objects in its surrounding.
Non-active lying	L_NAL	Lying with head resting against body, object, or floor.
Standing, self-grooming	S_SG	Calf is grooming itself (movements with tongue over body surface)
Standing, allo-grooming	S_AG	Calf is grooming another calf (movements with tongue over body surface)
Standing, oral behaviour calf	S_OBC	Calf's mouth is in direct contact with another calf's body. Behaviour to include biting, sucking other calves.
Standing, oral behaviour object	S_OBO	Behaviour to include biting, sucking, licking at objects in its surrounding.
Standing, eating	S_E	Calf is at straw or concentrate trough. Head in the trough/mouth in contact with feed.
Standing at milk feeder	S_MF	Calf is standing in the milk feeder. Standing within side gates present at milk feeder.
Standing, drinking	S_D	Calf is drinking at water station.
Standing	S	Standing, not interacting.
Locomotor Play	LP	Include instances of jumping or bucking. Includes all instances of trotting (two beat leg movements synchronized diagonally), cantering (three-beat gait in between a trot and a gallop) and galloping (four-beat gait with a phase where all legs are off the ground).
Social play	SP	Two calves are standing front to front, butting head against head/neck in a playful manner. or A calf mounts another calf's head or body from front, side, or back.
Walking	W	A slow four-beat gait with continuous forward movement with two or three of the hooves touching the ground at any time.

Next, we tested the algorithm labelled frames which were taken of the same scene but from a different angle (angle 3). Since footage from angles 1 and 2 only were used to train the dataset, comparing predictions with the labelled ground truth provides with performance estimates from an angle the algorithm has not previously seen allowing us to test the generalisability of the model.

The model training and testing was carried out using a virtual machine with a 6 core CPU, 56GB RAM and an 8GB NVIDIA Tesla M60 GPU, run time was approximately 2 hours.

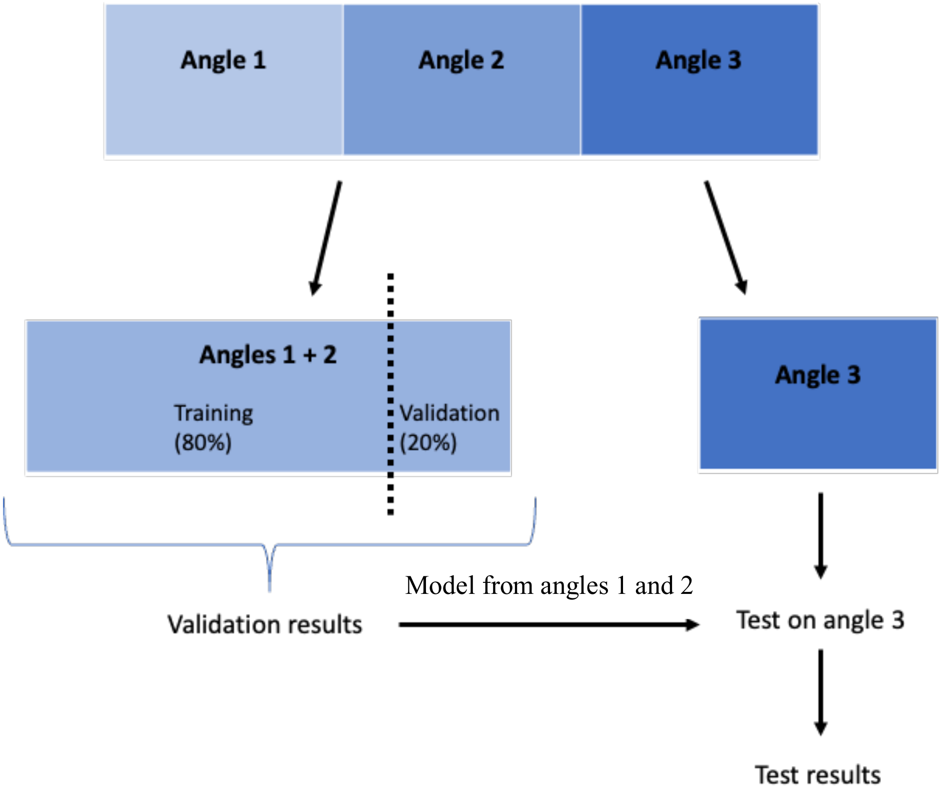


Figure 7.3. Diagram showing data selection from 3 camera angles for model training, validation, and testing

Model performance metrics

The model output includes a predicted bounding box, its class prediction (i.e., the behaviour detected) and the confidence score. To assess model performance, it is necessary to decide if a prediction and ground truth agree. For the purposes of assessing object detection, the methodology used was to measure the overlap between the predicted bounding box and the ground truth. To be considered as a correct detection, the area of overlap (α) between the predicted box (Bp) and the ground truth (B_{gt}) must exceed 50%. α is defined according to equation (1).

$$\alpha = \frac{area(Bp \cap Bgt)}{area(Bp \cup Bgt)} \qquad (1)$$

Failure to detect an object which is present counts as a false negative. If a single object is detected multiple times, whereas it is only present once, one instance will count as a true positive whilst the surplus will count as a false positive. This definition was used to compute the confusion matrix.

7.3 Results

7.3.1 Data labelling

For the purposes of the illustration, Figure 7.4 shows an example of a labelled image. Up to 20 calves can be seen in each frame therefore multiple instances of different behaviours can be captured from a single frame.

Figure. 7.4. Example of a labelled frame taken from angle 2. The annotations correspond to the different behaviours.

Table 7.2 details the number of instances labelled by behaviour. Four behaviours were not taken forward into the model training due to the small number of labelled instances (< 20 total) these behaviours. These were lying allo-grooming

(L_AG), lying oral behaviour calf (L_OBC), lying oral behaviour object (L_OBO), and standing drinking (S_D). The behaviours with the highest numbers of labels were active lying, non-active lying and standing.

Table 7.2. Number of instances labelled by behaviour and by dataset

Dataset	**L_AG**	**L_AL**	**L_NAL**	**L_OBC**	**L_OBO**	**L_SG**	**LP**	**S**	**S_AG**	**S_E**	**S_MF**	**S_OBC**	**S_OBO**	**S_SG**	**S_D**	**SP**	**W**
Train + validation (angles 1 + 2)	7	987	858	10	0	35	88	610	33	73	19	40	119	67	5	10	124
Test (angle 3)	3	77	109	1	1	4	10	108	7	8	7	9	24	17	4	5	24
Total	10	1064	967	11	1	39	98	718	40	81	26	49	143	84	9	15	148

7.3.2 Validation set results

The validation dataset consisted of images taken from the same two angle (angles 1 and 2) as the training dataset. We report the precision recall curve for the different behaviours (Figure 7.5). The mean average precision for all classes was 0.808 using a minimum area of overlap (α) of 0.5. The behaviours with the highest average precision were non-active lying (L_NAL), lying self-grooming (L_SG), standing eating (S_E), and standing at the milk feeder (S_MF) all had an average precision > 0.99. The behaviour with the lowest average precision was walking (W).

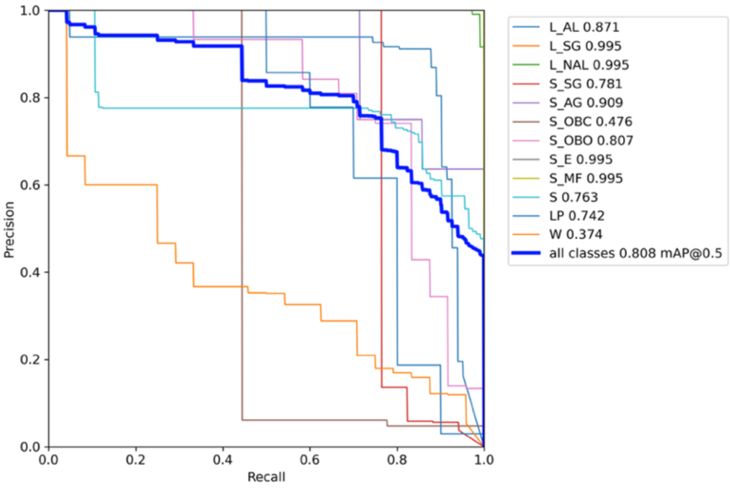


Figure 7.5. Precision recall curve for the different behaviours for the validation dataset

Results from the confusion matrix in presented in Figure 7.6 show variable levels of misclassification. There is some misclassification between standing and other standing behaviours, between standing and walking, and between locomotor play and walking. Background false positives (where the algorithm predicts a behaviour which isn't present) represented 1% of predictions.

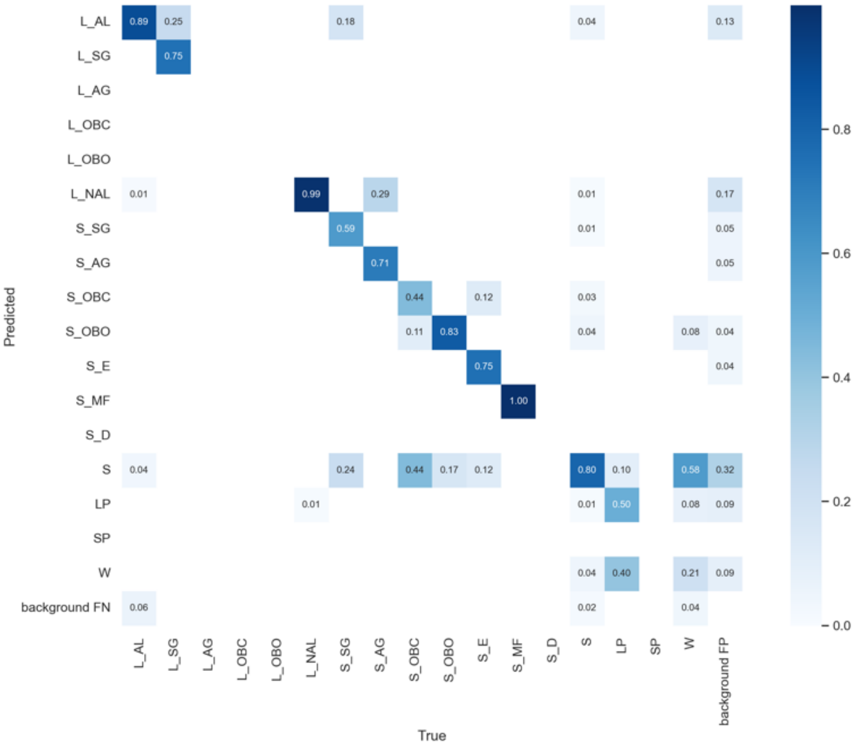


Figure 7.6. Confusion matrix from validation dataset

7.3.3 Test set results

Performance metrics on the test set are substantially reduced compared to results from the validation set. The test dataset consisted of images taken from a different angle (angle 3) compared to the training dataset (angles 1 and 2). The mean average precision for all classes was 0.234 using a minimum area of overlap (α) of 0.5. Figure 7.7 reports the Precision Recall AUC and average precision for each behaviour. The behaviour with the highest average precision was locomotor play (LP) with 0.819. The next highest were active lying (L_AL) and non-active lying (L_NAL) with a mean average precision of 0.449 and 0.406 respectively.

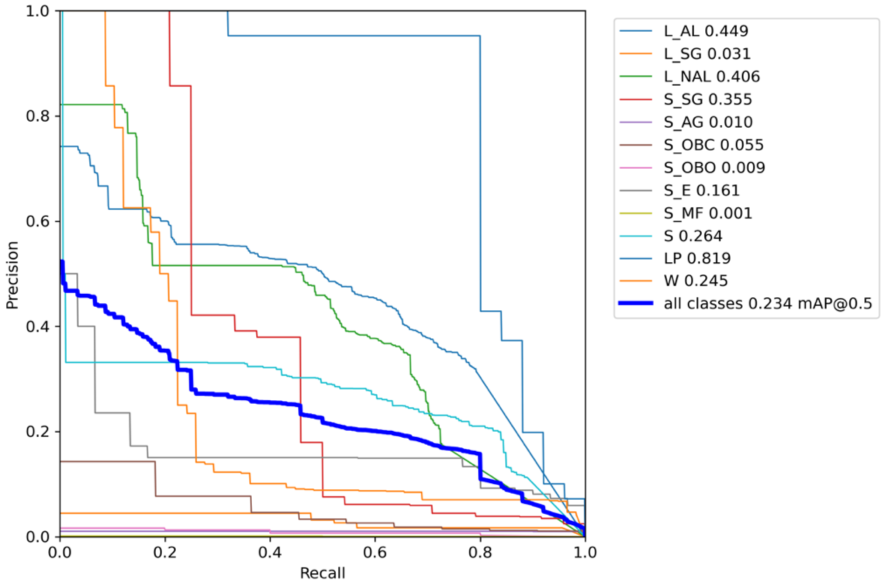
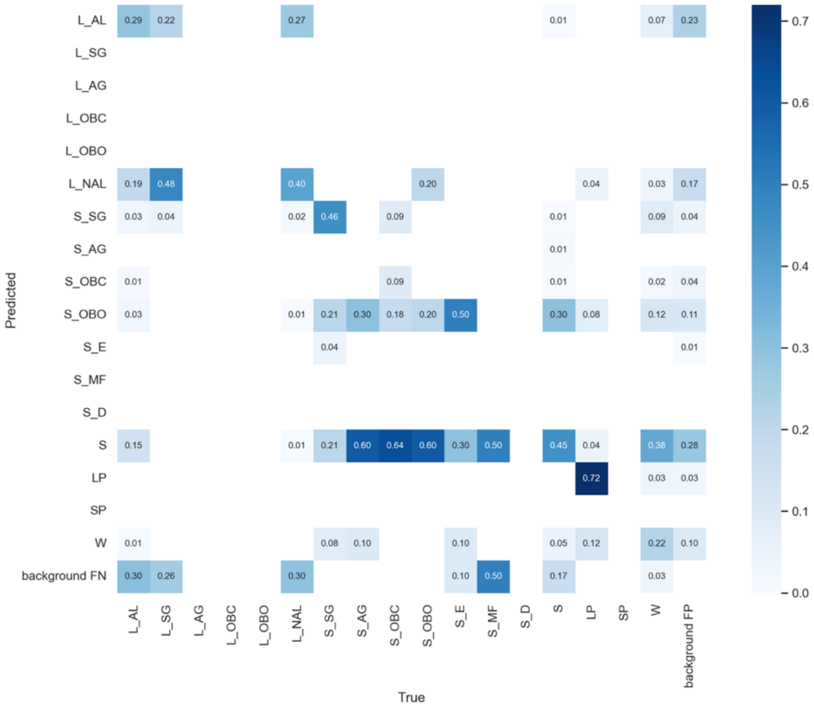


Figure 7.7. Precision recall curve for the different behaviours (a) and precision as a function of confidence score (b) for the test dataset

The confusion matrix for the validation dataset is detailed in Figure 7.8. Results show substantial levels of misclassification between behaviours, especially between standing behaviours and between lying behaviours. For example, standing self-grooming (S_SG), standing allo-grooming (S_AG), standing oral behaviour calf (S_OBC) and standing oral behaviour object (S_OBO) were frequently misclassified as standing. Lying self-grooming (L_SG) was also frequently misclassified as active lying (L_AL) or non-active lying (L_NAL) and there was moderate to high levels of misclassification between active lying (L_AL) and non-active lying (L_NAL).



	L_AL	L_SG	L_AG	L_OBC	L_OBO	L_NAL	S_SG	S_AG	S_OBC	S_OBO	S_E	S_MF	S_D	S	LP	SP	W	background FP
L_AL	0.29	0.22				0.27								0.01			0.07	0.23
L_SG																		
L_AG																		
L_OBC																		
L_OBO																		
L_NAL	0.19	0.48				0.40				0.20					0.04		0.03	0.17
S_SG	0.03	0.04				0.02	0.46		0.09					0.01			0.09	0.04
S_AG														0.01				
S_OBC	0.01								0.09					0.01			0.02	0.04
S_OBO	0.03					0.01	0.21	0.30	0.18	0.20	0.50			0.30	0.08		0.12	0.11
S_E							0.04											0.01
S_MF																		
S_D																		
S	0.15					0.01	0.21	0.60	0.64	0.60	0.30	0.50		0.45	0.04		0.38	0.28
LP															0.72		0.03	0.03
SP																		
W	0.01						0.08	0.10			0.10			0.05	0.12		0.22	0.10
background FN	0.30	0.26				0.30					0.10	0.50		0.17			0.03	

Figure 7.8. Confusion matrix from test dataset

7.4 Discussion

This study is the first to train and test a computer vision object detection algorithm to detect a wide variety of behaviours in group housed calves. Results from the validation set show high performance indicating that the YOLOv5 object detection algorithm is a promising approach for the group level monitoring of diverse behaviours in calves. However, results from the test dataset indicate that model performance is reduced when test and training datasets have different characteristics.

Our model was able to identify a diversity of behaviours in calves, some of which with high accuracy. In the validation set, of the 12 behaviours that were well represented 10 were identified with an average performance over 0.75. Four behaviours (standing eating, standing milk feeding, lying self-grooming and non-active lying) were identified with average precision over 0.99. The performance reported here reflects those from similar studies in other livestock species which also show high performance for different behaviours. For example, Fuentes et al., (2020) used YOLOv3 to detect 15 different behaviours in adult cattle with a mean average precision of 0.788. Similarly, to our study, detailed behaviours such as allo-grooming and aggressive interactions were accurately identified (average precision of 0.851 and 0.791 respectively). Results from a study in pigs show that YOLO outperformed other deep learning algorithms achieving a mean average precision of 0.98 for 5 different behaviours (Alameer et al., 2020). Together with the results from our study, these results indicate object detection vision algorithms can detect a diversity of behaviours in livestock. Further work is needed to assess the suitability of these algorithms to be used in the farm environment for longer term behavioural monitoring.

The ability to monitoring a diversity of behaviours as at a group level has important implications for calf health and welfare monitoring. Monitoring lying and resting behaviour could provide insights into calf comfort and the suitability of the bedding material provided. For example, results from studies in adult cattle show that increased lying time occurs in more comfortable conditions such as larger pens, deeper substrate, with rubber mats and with sand substrate versus concrete floor (Drissler et al., 2005; Haley et al., 2000; Norring et al., 2010; Sahu et al., 2021). Monitoring the interplay between non-active lying and active lying could enable insights into sleep behaviour in calves (Hokkanen et al., 2011) which may have implications for health and welfare (Ruckerbusch, 1975). One interesting behaviour that was identified by our algorithm was locomotor play behaviour which had an average performance of 0.742. This is of interest since play behaviour is most likely to be expressed in animals when threats to an animal's fitness are minimised and has been proposed as a useful indicator of positive welfare state (Held et al. 2011). For example, in calves reduced locomotor play has been associated with disbudding and reduced feed allowance and is positively associated with energy intake (Krachun et al. 2010) (Rushen and de Passillé, 2012.). Monitoring play behaviour at a group level could help assess if the calves are able to express positive behaviour, experience positive emotions and potentially flag situations where welfare is compromised. Similarly, changes to other behaviours such as social interactions or resting behaviours could help indicate a change in the environment that is disturbing calf behaviour. Monitoring social interactions such as social contact between calves may provide insight into group cohesion and sociality. In pigs, machine has been used to detect aggressive interactions such as tail biting and fighting (Chen et al., 2018; Liu et al., 2020). In calves, future studies could include agonistic interactions such as displacements from the automatic feeder. The group level vision based behavioural monitoring could also provide insights into the synchronicity of behaviour (where individuals tend to perform the same behaviour as that same time). Synchronicity in behaviour has been observed to increase in cattle with increased space allowance (L. H. Nielsen et al., 1997) and cattle have been observed to have more synchronised at pasture compared to tie stalls (Krohn et al., 1992). These studies have led to synchronicity in lying behaviour being proposed as an indicator of positive welfare in ruminants (Mattiello et al., 2019). Future studies could monitor the synchronicity of different behaviours in calves captured using machine vision techniques.

Results from our study show that performance was greatly reduced on the test dataset, which was filmed at a different angle compared to the train dataset, with a mean average precision 0.234. Testing a previously trained algorithm on a new angle or dataset is important since if machine vision systems are to be deployed in the field they will need generalise to new situations (e.g., new pens, different camera angles). Despite this, most similar studies have trained and test models on data with similar characteristics. Whilst performance was reduced for all behaviours in the test dataset, it is worth noting that confusion tends to occur between classes that represent similar behaviours (e.g., behaviours within the same posture) (Figure 7.8). For example, behaviours such as standing allo-grooming and standing eating were misclassified as standing and there was frequent misclassification between active lying and non-active lying. Misclassification between behaviours where the calf was lying and behaviours where the calf was standing occurred less frequently. This could indicate that an algorithm that only seeks only to classify posture (i.e., standing and lying) may have higher performance on the test dataset. The behaviour that was the best identified on the test dataset was locomotor play with an average precision of 0.819. It is possible that locomotor play was the had the highest performance since it is a high velocity behaviour and therefore distinctive. For example, studies using movement sensors have noted locomotor play's characteristic acceleration pattern (de Passillé et al., 2010) and locomotor play can be identified with >98% sensitivity and specificity using accelerometer data (Carslake et al. 2021). Further studies are needed to gather data more training and testing data and test if behaviours such as play can be reliably identified across a variety of environments.

One limitation of our study is that certain behaviours that we were we were interested in detecting were poorly represented in the training, validation, and test datasets. Increasing the number of labelled instances could increase the algorithm's performance. For example, one study trained the YOLOv5 algorithm to detect four different behaviours in sheep using different training and testing conditions. When the train and test data had different characteristics (filmed from different angles), the mean average performance was 0.258 when the algorithm was based off 1500 labelled images and 0.746 based of 9000 images (Cheng et al., 2022). These results could indicate that including more images can enable improved generalisation. It is also worth noting that this study only used two different camera angles. Further research is needed to test if including a greater diversity of images taken from different angles, hights and locations in the training dataset could create deep learning-based livestock behaviour detection algorithms capable of generalising across different environments and situations.

To conclude, our results show that the YOLOv5 object detection algorithm has potential to enable group level monitoring of diverse behaviours in calves in scenarios where training and test data are similar. This approach could be used to infer group level metrics which could enable improved health and welfare monitoring in calves. Future studies using video data from multiple environments and angles are needed to evaluate if this approach can be expanded to create a model capable of detecting diverse calf behaviours in new environments.

8 Chapter 8 – Key results and general discussion

There are approximately 2.7 million calves born in the UK each year as part of dairy and beef industries (DEFRA, 2022). Despite efforts to improve health and welfare across the industry, calves born into faming systems in the UK suffer from high mortality and morbidity (Hyde et al., 2017; K. F. Johnson et al., 2017). There are many factors which influence the outcome of the range of initiatives which have been proposed to potentiate the improvement of calf health and welfare, however, as evidenced by high morbidity and mortality in calves, none have had sufficient penetration and impact.

Over the past decade, the burgeoning field of precision livestock has been proposed as an alternative solution to the complex challenge of improving farm animal health and welfare (Berckmans, 2014, 2017). Precision livestock encompasses the use of new technologies to monitor indicators of health and welfare which could be used to inform management. However, research into how to use these new technologies to monitor calf health and welfare is in its infancy (Costa et al., 2021). The core aim of this thesis was to explore ways in which precision livestock technologies may be used to inform behaviour, health, and welfare in calves.

8.1 Key results

The first theme of this thesis was to apply novel methods from behavioural ecology research to data collected by precision livestock technologies to improve our understanding of the existence and extent of individual differences in calf behaviour. Previous studies in calves have shown that there is large variation in calf feeding behaviour (e.g., de Passillé et al., 2016; Sweeney et al., 2010; Webb et al., 2014). However, these studies use summary measures such as the mean and standard deviation of each behaviour over the study period. They do not estimate the extent to which between-individual variation explains total variation observed, quantify within-individual variation, or examine how between-individual differences may be related across suites of behaviours. In Chapter 2, individual variation was quantified for feeding rate, meal frequency and meal size by calculating repeatability and the coefficient of variation in predictability. Calves differed from each other in their average behavioural expression (behavioural type) and in their residual, within-individual variation around their behavioural type (predictability). Feeding rate and total meals had the highest repeatability (>0.4) indicating that substantial, temporally stable between-individual differences exist for these behaviours. Additionally, for feeding rate calves varied from more to less predictable whereas for meal size calves were more homogenous in their within-individual variation around their behavioural types. Finally, behavioural types for feeding rate and total meals were positively and significantly correlated which could suggests the existence of an underlying personality trait responsible for driving the (co)expression of these two behaviours. These results highlight how the application of methods from the behavioural ecology literature can improve our understanding of the existence and extent individual differences in calf feeding behaviour. Furthermore, by uncovering consistencies in between-individual behavioural differences, results indicate that animal personality may play a role in driving variability in calf feeding behaviour.

In Chapter 3, to test if between-individual differences in feeding behaviour, as captured by on farm technologies, were contextually and temporally consistent, the feeding behaviour of 76 calves was monitored over two contexts using computerised feeders. These contexts were a period of pair housing with two calves per feeder which was followed by a period group housing with 16 calves per feeder. Results show that for individual calves, between-individual differences in feeding rate and meal frequency were related, and, for each behaviour, between-individual differences were positively and significantly correlated across contexts. In addition, feeding rate and meal frequency in the pair housing were positively and significantly associated with weight gain. To meet the definition of a personality trait, between-individual differences in related behaviours must be temporally as well as contextually stable (Kaiser & Müller, 2021). Together, these results could indicate the existence of a personality trait which positions high meal frequency, fast drinking, fast growing calves at one end of the trait spectrum and low meal frequency, slow drinking, and slow growing calves at the other. The results from this study suggest the intriguing prospect that data already available on commercial farms could be harnessed to phenotype calves according to a personality trait. Future studies are needed to test the generalisability of these results and how such phenotyping could be used to inform management strategies.

The second theme of this thesis was to explore the use of machine learning models to predict ill health in calves by use of sensor data. In Chapter 4, insights regarding individual differences in feeding behaviour that were identified in Chapters 2 and 3 were used to create statistical features that quantify within-individual change for different feeding behaviours. These features were used alongside health scores to build a machine learning model to classify sick versus healthy calves by use of their feeding behaviours. Results suggest machine learning algorithms that monitor feeding behaviour in calves may be insufficiently sensitive and specific to detect the early signs of disease. This could indicate that feeding behaviours is a core behaviour in pre-weaned calves and is not a reliable early indicator of ill health in these animals. To detect respiratory disease in calves by use of remote monitoring it may be necessary to incorporate other behaviours and physiological indicators, such as core body temperature. However, results from this chapter are limited by the relatively small number of calves included and the use of health scoring to identify sick calves.

As part of the physiological response to infection, affected animals increase their core body temperature (Kluger et al., 1998; Wingfield, 2003). Elevated body temperature is indicator of ill health in cattle (Grigor et al., 2001; McGuirk & Peek, 2014) which suggests remote temperature monitoring could be used to detect bovine respiratory disease. In Chapter 5, reticulo-rumen temperature data by use of an indwelling temperature bolus was collected. Different machine learning algorithms were created to test the predictive value of statistical features calculated from temperature recordings for respiratory disease in cattle. Whilst specificity was moderate to high, sensitivity was moderate to low and reduced further when the algorithm was tested on a new population. As with Chapter 4, limitations include the lack of a gold standard from which to label calves.

The third theme was to develop novel technologies and test if novel technologies could be used to monitor continuously and simultaneously a wider variety of behaviours in calves. In calves, monitoring multiple behaviours such as locomotor play, self-grooming, feeding and activity whilst lying are likely to be informative of health and welfare. However, the sensors that are currently used to monitor behaviour in calves are only able to detect a limited diversity and number of behaviours. In addition, many interesting behaviours, such as locomotor play, occur rarely, which means simply counting behaviours based on the prediction of a classifier can lead to vast overestimation of their occurrence.

Chapter 6 addressed two key challenges that are encountered by precision livestock research a) develop a classification algorithm for a breadth of behaviours using a single sensor and b) quantify the distribution of the behaviour in a real-world scenario where the prevalence of the behaviour is different than it was in the training set. The classification algorithm was able to identify accurately a breadth of behaviours in calves e.g., posture, self-grooming, feeding behaviours, resting and locomotor play. Furthermore, results demonstrate the accuracy of a quantification algorithm for predicting the occurrence of locomotor play behaviour; predicting the occurrence of low rarely occurring behaviours is an area mostly ignored in precision livestock. Results showed high accuracy with relatively low overestimation of locomotor play behaviour in unseen real-world data despite its very low (0.27%) prevalence. This is of particular importance for research where change in behaviour distribution is of interest for disease and welfare prediction and quantification. Further work should involve further developing multiclass quantification algorithms and using the developed methods to improve our knowledge of the relationship between an individual's behaviour, its health, and its welfare.

Physical sensors face numerous practical challenges when used to monitor livestock continuously such as the requirement to find a balance between computational requirements, processing power and battery life (Carslake et al., 2020; Walton et al., 2018). For example, results from Chapter 6 also showed that accurately monitoring more complex behaviours may require higher sampling frequency which may place constraints on battery life. Chapter 7 aimed to test if an object detection computer vision algorithm (YOLOv5) could be trained to successfully identify a wide diversity of behaviours in calves at a group level using CCTV footage of calf behaviour in a group pen. Results show that a wide variety of behaviours in pre-weaned calves can be successfully identified using CCTV footage alone. These behaviours include posture, different resting behaviours (active lying, non-active lying), feeding behaviour, grooming and social grooming as well as locomotor play. However, results show that performance was greatly reduced when the algorithm was tested on video footage filmed from a different camera angle indicating that further work is needed to create a generalisable vision algorithm capable of accurately detecting a range of calf behaviours.

8.2 General discussion

Results from this thesis reveal two major themes. Firstly, that substantial between and within-individual differences in calf feeding behaviour exist. Secondly, that technologies such as cameras and movement sensors can be used to capture a wide diversity of behaviours in calves which may be useful to measure health and welfare in calves.

When considering the first theme, results from this thesis show that precision livestock technologies can be used to quantify within- and between-individual variation in different behaviours. This knowledge can be used to improve our understanding of farm animals as individuals. For example, results show that individual calves have patterns of behavioural variation that are characteristic of each individual animal. Results indicate that feeding behaviour in young calves is related with future weight gain and is consistent between contexts, indicating a personality trait. Personality traits in animals are associated with a suite of behavioural and physiological indicators (Biro & Stamps, 2008; Dingemanse & Réale, 2005) which have implications for health, welfare, and productivity in farm animals (Finkemeier et al., 2018). Quantifying behavioural variation is made possible by the combination of technological advances that can record individuals' behaviour over time with conceptual advances that use repeated measures to quantify individual variation (e.g., Dingemanse et al., 2010; Hertel et al., 2020). This approach offers a break with the previous methodology of behavioural tests used to measure variation in behaviour and personality in farm animals (Forkman et al., 2007; Murphy et al., 2014b). Whilst the personality tests have the advantage that they can be undertaken in controlled conditions thereby reducing bias introduced through confounding factors such as the external environment, it is possible that these same controlled conditions may not accurately reflect natural behaviour, especially if the test is conducted in a stressful context (Carter et al., 2013). For example, in calves, exploration is frequently tested using a novel object test where animals are tested individually, often in a new environment (e.g., Neave et al., 2018). In this situation, calves may be stressed by separation from the group and exposure to a new environment, meaning their exploratory behaviour may not reflect their natural tendency. Moreover, for logistical reasons these tests are not practical beyond research settings. The statistical partitioning of behavioural variation as captured by precision livestock technologies offers a practical solution to quantify behavioural variation in the farm environment with minimal need to handle or interfere with the livestock. Moreover, where these technologies are available on commercial farms, they offer the opportunities to provide data which could be used to integrate findings from this area of research into management strategies.

Another advantage of measuring behavioural variation using repeated measures from sensors is that terminology used to described behavioural variation are statistically defined (Dingemanse et al., 2010). These measures are therefore unambiguous which helps facilitate communication between similar studies. For each behaviour monitored, repeated measures could be harnessed to measure between-individual differences in average behaviour (repeatability) and the degree to which individuals differ in their within-individual variation (coefficient of variation in predictability). Multivariate approaches could be used to test if the behavioural responses of individuals are correlated across behaviours or with other outcomes such as production or health. Since technologies that measure behaviours such as posture and activity are increasingly available on farms, especially in adult dairy cattle (Frost et al., 1997; Mottram, 2016), opportunities exist to harness these measures and compare differences in behavioural variation across farms and management systems. For example, one recent study showed that cows with increased living space had a significantly increased lying time of approximately 1hour per day compared to a control group (Thompson et al., 2022). Future studies could explore how changes in living space might impact within- and between-individual variation of behaviours such as lying times. More broadly, a greater awareness and wider adoption of this terminology within livestock behaviour and precision livestock research would allow for improved cross-fertilisation of ideas with related disciplines such as behavioural ecology which are already using this terminology (Hertel et al., 2020, 2021).

A second theme revealed by the results of this thesis is that precision livestock technologies can be used to monitor a diversity of behaviours in calves. This could help improve the monitoring of health and welfare in calves by use of precision technologies. For example, in Chapter 4, the model which used feeding behaviours to predict respiratory disease, had poor performance and was unable to accurately distinguish between calves that remained healthy and those that became sick. It has been previously suggested that changes in different behaviours around illness are sequential and those behaviours that are less critical are reduced first (Weary et al., 2009). Results from chapter 4 could indicate that feeding behaviours are core behaviours and monitoring other behaviours is important to develop accurate disease detection models. This is also evidenced by the earlier review of sickness behaviours in calves which indicates that changes to behaviours such as grooming, different resting behaviours and play may be useful in detecting the early signs of ill health. These behaviours were identified by the movement sensor in Chapter 6 and computer vision algorithm in Chapter 7 indicating that there is potential to monitor these behaviours using precision livestock technologies. Future studies that aim to test the predictive value of diverse behaviours to improve disease detection in calves may benefit from improved methodologies to identify and label sick calves. Limitations to Chapters 4 and 5 include the use of health scoring to label animals as sick or as healthy in the training data. Health scoring to detect respiratory disease has moderate sensitivity and specificity of 62.4% and 74.1%

respectively (Buczinski et al., 2015) whereas there is some evidence to suggest that ultrasonography has a higher sensitivity of 76%-94% and specificity of 93% to 100% (Berman et al., 2019). Future studies could incorporate the regular scoring of calves using ultrasonography in addition to health scoring to improve label accuracy. As well as being only moderately accurate, manual health scoring limits the quantity of training data that can be collected as it is time consuming to undertake (McGuirk & Peek, 2014). Another currently available technology that could assist in the labelling of sick calves which was explored in this thesis is temperature boluses. As discussed in Chapter 5, elevated temperature is an important indicator of ill health and elevated rumen temperature is related to negative production outcomes in cattle even when no other clinical signs have been recorded (Timsit et al., 2011b). However, no studies have explored the relationship between increases in rumen temperature and other outcomes such as changes in behaviour. As temperature boluses and activity monitors become more widespread, future studies could use temperature increases as response variables in different modelling approaches to identify and evaluate which behavioural features are most predictive of elevated body temperature in cattle.

In addition to identifying behaviours that may help improve the detection of ill health, results from chapters 6 and 7 also indicate that precision livestock technologies could help gather data that may be useful for assessing calf welfare. Assessing welfare involves physical components such as health but an increasing emphasis has been put on emotional components such as mental state and naturalness (Mellor, 2016). Behaviours such as play, grooming and exploration could provide insight into an animal's positive experiences, which along with other indicators, could help enable a more holistic assessment of calf welfare (Mattiello et al., 2019). It has previously been suggested that the low prevalence of behaviours such as play means that they are not possible to measure on farm (Napolitano et al., 2009). However, results from this thesis show that by combining sensor data with a quantification machine learning algorithm it was possible to estimate locomotor play behaviour with good accuracy despite it composing only 0.27% of behaviours measured. These results indicate that multiclass quantification algorithms, as reviewed by Forman

2008), could be used to measure the occurrence of other behaviours that may indicate positive welfare states, such as grooming or affiliative social behaviours but also occur at low prevalence (Mattiello et al., 2019; Napolitano et al., 2009). Results from Chapter 7 show computer vision algorithms can be used to monitor a diversity of behaviours calves. However, this was at the group level only (i.e., the algorithm did not attempt to identify individual calves). Algorithms that monitor behaviour at the group level can give insights into the welfare of the group. For example, monitoring lying, and resting behaviours can help inform how comfortable the environment is. One example is the cow comfort index which has been used to assess comfort in adult dairy cattle (e.g., Cook et al., 2005; Haley et al., 2000). The index corresponds to the number of cows lying in free stalls divided by the number of cows touching a stall. However, this metric is reliant on single time point observations. Computer vision techniques could provide one way of continuously monitoring this index. Whilst metrics captured at a group level may be informative, there is still a need to monitor behaviour at the level of individuals. As detailed in Chapters 2 and 3, mean and variance are not stable across individuals therefore group metrics conceal patterns and processes at the individual level (Broom 1986). As such, computer vision systems are being developed that identify individual animals and track their behaviour over time (e.g., Wang et al., 2022). Combined with the identification of diverse behaviours such solutions could enable practical solutions for long term, more comprehensive monitoring of health and welfare in calves.

This thesis used a range of different technologies to explore the application of precision livestock technologies in on-farm environments to monitor behaviour, health, and welfare in calves. These technologies are at different stages of development. For example, the movement sensors and vision technologies used in this thesis were developed specifically for these studies. One area that will require development is the creation of algorithms that can accurately detect behaviours in new situations (e.g., new populations). Generalisation to new situations can be challenging due to concept drift, a term used to describe when systems are presented with a new scenario where the new data does not reflect the training data (i.e., the assumption of stationary training data does not hold) (Gama et al., 2014). These changes in the feature space can result in reduced performance. For example, in sheep differences in sward height resulting in changing performance for sensor based grazing behaviour detection (Guo et al., 2018). Changes in the feature space can be a significant problem in precision livestock where there is large variation amongst farms and management systems (breed, housing system, age, space allowance etc) and it may not be possible to generate training data that covers all possible testing scenarios. To maintain performance, algorithms need to adapt (Pan & Yang, 2010). Adaptive learning relates to updating predictive models, through the transfer of knowledge from a previous task to data collected in the new scenario, allowing models to perform well in new environments (Pan & Yang, 2010). In precision livestock, one study has tested an approach to update an algorithm which predicts behaviour in sheep by use of a data from a movement sensor (Vázquez-Diosdado et al. 2019b). This approach composes of an algorithm which is trained using historically collected data, and a new component which uses a comparatively smaller amount of data from the new scenario. Results from this study showed an improvement in accuracy from 78.35% to 92.93% when the historic algorithm was updated using newly collected data. Using such techniques may provide a methodology to reduce the discrepancy between learning and testing datasets and improve algorithm performance when behaviour detection systems are faced with new scenarios.

To conclude, the increasing diversity of behaviours that can be captured with improving accuracy, the development of statistical methodologies that can harness these data and an improved understanding and appreciation of behavioural variation are promising developments for the future of precision livestock farming. The development of these areas could improve management in calves, by supporting the detection of ill health and by providing insights into animal's quality of life as expressed by physical as well as emotional perspectives. Improving the uptake of these technologies will require further technological developments to improve their reliability, decreases in costs and the integration of clear management advice (Buller et al., 2020). In conjunction, new business models will be needed to support the use and distributions of precision livestock metrics between food actors and animal health professionals (Lokhorst et al., 2019). Finally, animal behaviour and welfare research will need to stay at the forefront technological advances, ensuring that new developments offer creative, evidence-based, and holistic improvements to the health and welfare of calves.

9 References

Alameer, A., Kyriazakis, I., & Bacardit, J. (2020). Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs. *Scientific Reports*, *10*(1), 13665. https://doi.org/10.1038/s41598-020-70688-6

AlZahal, O., AlZahal, H., Steele, M. A., Van Schaik, M., Kyriazakis, I., Duffield, T. F., & McBride, B. W. (2011). The use of a radiotelemetric ruminal bolus to detect body temperature changes in lactating dairy cattle. *Journal of Dairy Science*, *94*(7), 3568–3574. https://doi.org/10.3168/jds.2010-3944

AlZahal, O., Kebreab, E., France, J., Froetschel, M., & McBride, B. W. (2008). Ruminal Temperature May Aid in the Detection of Subacute Ruminal Acidosis. *Journal of Dairy Science*, *91*(1), 202–207. https://doi.org/10.3168/jds.2007-0535

Appleby, M. C., Weary, D. M., & Chua, B. (2001). Performance and feeding behaviour of calves on ad libitum milk from artificial teats. *Applied Animal Behaviour Science*, *74*(3), 191–201. https://doi.org/10.1016/S0168-1591(01)00171-X

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2014). Characterizing behavioural ‘characters’: an evolutionary framework. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1776), 20132645. https://doi.org/10.1098/rspb.2013.2645

Ashcroft, J., Semmler, C., Carnell, S., van Jaarsveld, C. H. M., & Wardle, J. (2008). Continuity and stability of eating behaviour traits in children. *European Journal of Clinical Nutrition*, *62*(8), 985–990. https://doi.org/10.1038/sj.ejcn.1602855

Atkinson, D. J., von Keyserlingk, M. A. G., & Weary, D. M. (2017). Benchmarking passive transfer of immunity and growth in dairy calves. *Journal of Dairy Science*, *100*(5), 3773–3782. https://doi.org/10.3168/jds.2016-11800

Atkinson, O. (2015). *To survey current practices and performance and to determine success factors for rearing replacement dairy heifers in Wales. Dairy Youngstock Project – Wales Report.* .

Bach, A. (2011). Associations between several aspects of heifer development and dairy cow survivability to second lactation. *Journal of Dairy Science*, *94*(2), 1052–1057. https://doi.org/10.3168/jds.2010-3633

Barker, Z. E., Vázquez Diosdado, J. A., Codling, E. A., Bell, N. J., Hodges, H. R., Croft, D. P., & Amory, J. R. (2018). 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*(7), 6310–6321. https://doi.org/10.3168/JDS.2016-12172

Bartels, C. J. M., Holzhauer, M., Jorritsma, R., Swart, W. A. J. M., & Lam, T. J. G. M. (2010). Prevalence, prediction and risk factors of enteropathogens in normal and non-normal faeces of young Dutch dairy calves. *Preventive Veterinary Medicine*, *93*(2–3), 162–169. https://doi.org/10.1016/J.PREVETMED.2009.09.020

Barwick, J., Lamb, D., Dobos, R., Schneider, D., Welch, M., Trotter, M., Barwick, J., Lamb, D., Dobos, R., Schneider, D., Welch, M., & Trotter, M. (2018). Predicting Lameness in Sheep Activity Using Tri-Axial

Acceleration Signals. *Animals*, *8*(1), 12. https://doi.org/10.3390/ani8010012

Barwick, J., Lamb, D. W., Dobos, R., Welch, M., & Trotter, M. (2018). Categorising sheep activity using a tri-axial accelerometer. *Computers and Electronics in Agriculture*, *145*, 289–297. https://doi.org/10.1016/j.compag.2018.01.007

Beauchemin, K. (2018). Invited review: Current perspectives on eating and rumination activity in dairy cows. *Journal of Dairy Science*, *101*, 4762–4784. https://doi.org/10.3168/jds.2017-13706

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, *77*(4), 771–783. https://doi.org/10.1016/J.ANBEHAV.2008.12.022

Bell, M. J., Wall, E., Russell, G., Simm, G., & Stott, A. W. (2011). The effect of improving cow productivity, fertility, and longevity on the global warming potential of dairy systems. *Journal of Dairy Science*, *94*(7), 3662–3678. https://doi.org/10.3168/jds.2010-4023

Ben-David, A. (2008). Comparison of classification accuracy using Cohen's Weighted Kappa. *Expert Systems with Applications*, *34*(2), 825–832. https://doi.org/10.1016/j.eswa.2006.10.022

Berckmans, D. (2014). Precision livestock farming technologies for welfare management in intensive livestock systems. *OIE Revue Scientifique et Technique*, *33*(1), 189–196. https://doi.org/10.20506/rst.33.1.2273

Berckmans, D. (2017). General introduction to precision livestock farming. *Animal Frontiers*, *7*(1), 6–11. https://doi.org/10.2527/af.2017.0102

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

Bersch, S., Azzi, D., Khusainov, R., Achumba, I., & Ries, J. (2014). Sensor Data Acquisition and Processing Parameters for Human Activity Classification. *Sensors*, *14*(3), 4239–4270. https://doi.org/10.3390/s140304239

Bewley, J. M., Einstein, M. E., Grott, M. W., & Schutz, M. M. (2008). Comparison of Reticular and Rectal Core Body Temperatures in Lactating Dairy Cows. *Journal of Dairy Science*, *91*(12), 4661–4672. https://doi.org/10.3168/jds.2007-0835

Bikker, J. P., van Laar, H., Rump, P., Doorenbos, J., van Meurs, K., Griffioen, G. M., & Dijkstra, J. (2014). Technical note: Evaluation of an ear-attached movement sensor to record cow feeding behavior and activity. *Journal of Dairy Science*, *97*(5), 2974–2979. https://doi.org/10.3168/jds.2013-7560

Biro, P. A. (2012). Do rapid assays predict repeatability in labile (behavioural) traits? *Animal Behaviour*, *83*(5), 1295–1300. https://doi.org/10.1016/j.anbehav.2012.01.036

Biro, P. A., & Stamps, J. A. (2008). Are animal personality traits linked to life-history productivity? *Trends in Ecology & Evolution*, *23*(7), 361–368. https://doi.org/10.1016/j.tree.2008.04.003

Blakebrough-Hall, C., Dona, A., D'occhio, M. J., McMeniman, J., & González, L. A. (2020). Diagnosis of Bovine Respiratory Disease in feedlot cattle using blood 1H NMR metabolomics. *Scientific Reports*, *10*(1), 115. https://doi.org/10.1038/s41598-019-56809-w

Boltana, S., Sanhueza, N., Donoso, A., Aguilar, A., Crespo, D., Vergara, D., Arriagada, G., Morales-Lange, B., Mercado, L., Rey, S., Tort, L., & Mackenzie, S. (2018). The expression of TRPV channels, prostaglandin E2 and pro-inflammatory cytokines during behavioural fever in fish. *Brain, Behavior, and Immunity*, *71*, 169–181. https://doi.org/10.1016/j.bbi.2018.03.023

Bonk, S., Burfeind, O., Suthar, V. S., & Heuwieser, W. (2013). Technical note: Evaluation of data loggers for measuring lying behavior in dairy calves. *Journal of Dairy Science*, *96*(5), 3265–3271. https://doi.org/10.3168/JDS.2012-6003

Borderas, T. F., de Passillé, A. M., & Rushen, J. (2008). Behavior of dairy calves after a low dose of bacterial endotoxin. *Journal of Animal Science*, *86*(11), 2920–2927. https://doi.org/10.2527/jas.2008-0926

Borderas, T. F., Rushen, J., von Keyserlingk, M. A. G., & de Passillé, A. M. B. (2009). Automated measurement of changes in feeding behavior of milk-fed calves associated with illness. *Journal of Dairy Science*, *92*(9), 4549–4554. https://doi.org/10.3168/jds.2009-2109

Borges Oliveira, D. A., Ribeiro Pereira, L. G., Bresolin, T., Pontes Ferreira, R. E., & Reboucas Dorea, J. R. (2021). A review of deep learning algorithms for computer vision systems in livestock. *Livestock Science*, *253*, 104700. https://doi.org/10.1016/j.livsci.2021.104700

Bowen, J. M., Haskell, M. J., Miller, G. A., Mason, C. S., Bell, D. J., & Duthie, C.-A. (2021). Early prediction of respiratory disease in preweaning dairy calves using feeding and activity behaviors. *Journal of Dairy Science*, *104*(11), 12009–12018. https://doi.org/10.3168/jds.2021-20373

Brambell. (1965). *Report of the Technical Committee to enquire into the welfare of animals kept under intensive livestock husbandry systems*.

Brenner, H., & Gefeller, O. (1997). Variation of sensitivity, specificity, likelihood rations and predictive values with disease prevalence. *Statistics in Medicine*, *16*(9), 981–991. https://doi.org/10.1002/(SICI)1097-0258(19970515)16:9<981::AID-SIM510>3.0.CO;2-N

Brickell, J. S., McGowan, M. M., Pfeiffer, D. U., & Wathes, D. C. (2009). Mortality in Holstein-Friesian calves and replacement heifers, in relation to body weight and IGF-I concentration, on 19 farms in England. *Animal*, *3*(8), 1175–1182. https://doi.org/10.1017/S175173110900456X

Broom, D. M. (1988). The scientific assessment of animal welfare. *Applied Animal Behaviour Science*, *20*(1–2), 5–19. https://doi.org/10.1016/0168-1591(88)90122-0

Brown, D. D., Kays, R., Wikelski, M., Wilson, R., & Klimley, A. P. (2013). Observing the unwatchable through acceleration logging of animal behavior. In *Animal Biotelemetry* (Vol. 1, Issue 1, p. 20). BioMed Central Ltd. https://doi.org/10.1186/2050-3385-1-20

Buczinski, S., L Ollivett, T., & Dendukuri, N. (2015). Bayesian estimation of the accuracy of the calf respiratory scoring chart and ultrasonography for the diagnosis of bovine respiratory disease in pre-weaned dairy calves. *Preventive Veterinary Medicine*, *119*(3–4), 227–231. https://doi.org/10.1016/J.PREVETMED.2015.02.018

Buller, H., Blokhuis, H., Lokhorst, K., Silberberg, M., & Veissier, I. (2020). Animal Welfare Management in a Digital World. *Animals*, *10*(10), 1779. https://doi.org/10.3390/ani10101779

Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*, *46*(3), 1–33. https://doi.org/10.1145/2499621

Burdick, N. C., Carroll, J. A., Hulbert, L. E., Dailey, J. W., Ballou, M. A., Randel, R. D., Willard, S. T., Vann, R. C., & Welsh, T. H. (2011). Temperament influences endotoxin-induced changes in rectal temperature, sickness behavior, and plasma epinephrine concentrations in bulls. *Innate Immunity*, *17*(4), 355–364. https://doi.org/10.1177/1753425910379144

Burke, K. C., do Nascimento-Emond, S., Hixson, C. L., & Miller-Cushon, E. K. (2022). Social networks respond to a disease challenge in calves. *Scientific Reports*, *12*(1), 9119. https://doi.org/10.1038/s41598-022-13088-2

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Cantor, M. C., Casella, E., Silvestri, S., Renaud, D. L., & Costa, J. H. C. (2022). Using Machine Learning and Behavioral Patterns Observed by Automated Feeders and Accelerometers for the Early Indication of Clinical Bovine Respiratory Disease Status in Preweaned Dairy Calves. *Frontiers in Animal Science*, *3*. https://doi.org/10.3389/fanim.2022.852359

Carslake, C., Occhiuto, F., Vázquez-Diosdado, J. A., & Kaler, J. (2022). Repeatability and Predictability of Calf Feeding Behaviors—Quantifying Between- and Within-Individual Variation for Precision Livestock Farming. *Frontiers in Veterinary Science*, *9*. https://doi.org/10.3389/fvets.2022.827124

Carslake, C., Vázquez-Diosdado, J. A., & Kaler, J. (2020). Machine Learning Algorithms to Classify and Quantify Multiple Behaviours in Dairy Calves Using a Sensor: Moving beyond Classification in Precision Livestock. *Sensors 2021, Vol. 21, Page 88*, *21*(1), 88. https://doi.org/10.3390/S21010088

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? *Biological Reviews*, *88*(2), 465–475. https://doi.org/10.1111/brv.12007

Cellier, M., Duvaux-Ponter, C., & Nielsen, B. L. (2021). Inter- and intra-individual variability of feeding behaviour in group housed dairy goats. *Applied Animal Behaviour Science*, *234*, 105167. https://doi.org/10.1016/J.APPLANIM.2020.105167

CHAWG. (2020). *Fifth Report - GB Cattle Health & Welfare Group*.

Chen, C., Zhu, W., Guo, Y., Ma, C., Huang, W., & Ruan, C. (2018). A kinetic energy model based on machine vision for recognition of aggressive behaviours among group-housed pigs. *Livestock Science*, *218*, 70–78. https://doi.org/10.1016/j.livsci.2018.10.013

Chen, C., Zhu, W., & Norton, T. (2021). Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning. *Computers and*

Electronics in Agriculture, *187*, 106255. https://doi.org/10.1016/j.compag.2021.106255

Cheng, M., Yuan, H., Wang, Q., Cai, Z., Liu, Y., & Zhang, Y. (2022). Application of deep learning in sheep behaviors recognition and influence analysis of training data characteristics on the recognition effect. *Computers and Electronics in Agriculture*, *198*, 107010. https://doi.org/10.1016/j.compag.2022.107010

Cho, Y.-I., & Yoon, K.-J. (2014). An overview of calf diarrhea - infectious etiology, diagnosis, and intervention. *Journal of Veterinary Science*, *15*(1), 1–17. https://doi.org/10.4142/jvs.2014.15.1.1

Cleasby, I. R., Nakagawa, S., & Schielzeth, H. (2015). Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, *6*(1), 27–37. https://doi.org/10.1111/2041-210X.12281

Cook, N. B., Bennett, T. B., & Nordlund, K. V. (2005). Monitoring Indices of Cow Comfort in Free-Stall-Housed Dairy Herds. *Journal of Dairy Science*, *88*(11), 3876–3885. https://doi.org/10.3168/jds.S0022-0302(05)73073-3

Cooke, S. J., Hinch, S. G., Wikelski, M., Andrews, R. D., Kuchel, L. J., Wolcott, T. G., & Butler, P. J. (2004). Biotelemetry: A mechanistic approach to ecology. In *Trends in Ecology and Evolution* (Vol. 19, Issue 6, pp. 334–343). Elsevier Current Trends. https://doi.org/10.1016/j.tree.2004.04.003

Coppens, C. M., Boer, S. F. de, & Koolhaas, J. M. (2010). Coping styles and behavioural flexibility: towards underlying mechanisms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4021–4028. https://doi.org/10.1098/RSTB.2010.0217

Costa, J. H. C., Cantor, M. C., & Neave, H. W. (2021). Symposium review: Precision technologies for dairy calves and management applications. *Journal of Dairy Science*, *104*(1), 1203–1219. https://doi.org/10.3168/jds.2019-17885

Crestani, F., Seguy, F., & Dantzer, R. (1991). Behavioural effects of peripherally injected interleukin-1: role of prostaglandins. *Brain Research*, *542*(2), 330–335. https://doi.org/10.1016/0006-8993(91)91587-Q

Dammhahn, M., Dingemanse, N. J., Niemelä, P. T., & Réale, D. (2018). Pace-of-life syndromes: a framework for the adaptive integration of behaviour, physiology and life history. *Behavioral Ecology and Sociobiology 2018 72:3*, *72*(3), 1–8. https://doi.org/10.1007/S00265-018-2473-Y

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*. Pearson education.

Dantzer, R., & Kelley, K. W. (2007). Twenty years of research on cytokine-induced sickness behavior. *Brain, Behavior, and Immunity*, *21*(2), 153–160. https://doi.org/10.1016/j.bbi.2006.09.006

Dargie, W. (2009). Analysis of Time and Frequency Domain Features of Accelerometer Measurements. *2009 Proceedings of 18th International Conference on Computer Communications and Networks*, 1–6. https://doi.org/10.1109/ICCCN.2009.5235366

Darr, M., & Epperson, W. (2009). Embedded sensor technology for real time determination of animal lying time. *Computers and Electronics in Agriculture*, *66*(1), 106–111. https://doi.org/10.1016/J.COMPAG.2009.01.004

David, M., Auclair, Y., & Cézilly, F. (2012). Assessing Short- and Long-Term Repeatability and Stability of Personality in Captive Zebra Finches Using Longitudinal Data. *Ethology*, *118*(10), 932–942. https://doi.org/10.1111/J.1439-0310.2012.02085.X

Dawkins, M. S. (2004). Using behaviour to assess animal welfare. *Animal Welfare*, *13*(1), 3–7. https://www.ingentaconnect.com/content/ufaw/aw/2004/00000013/A00101s1/art00002

de Passillé, A. M., Jensen, M. B., Chapinal, N., & Rushen, J. (2010). Technical note: Use of accelerometers to describe gait patterns in dairy calves. In *Journal of Dairy Science* (Vol. 93, Issue 7). https://doi.org/10.3168/jds.2009-2758

de Passillé, A. M., Rabeyrin, M., & Rushen, J. (2016). Associations between milk intake and activity in the first days of a calf's life and later growth and health. *Applied Animal Behaviour Science*, *175*, 2–7. https://doi.org/10.1016/j.applanim.2014.10.002

De Paula Vieira, A., Guesdon, V., de Passillé, A. M., von Keyserlingk, M. A. G., & Weary, D. M. (2008). Behavioural indicators of hunger in dairy calves. *Applied Animal Behaviour Science*, *109*(2–4), 180–189. https://doi.org/10.1016/j.applanim.2007.03.006

DEFRA. (2022). *Livestock numbers in England and the UK, Available at: https://www.gov.uk/government/statistical-data-sets/structure-of-the-livestock-industry-in-england-at-december, accessed 19-08-2022* .

Delabouglise, A., James, A., Valarcher, J.-F., Hagglünd, S., Raboisson, D., & Rushton, J. (2017). Linking disease epidemiology and livestock productivity: The case of bovine respiratory disease in France. *PLOS ONE*, *12*(12), e0189090. https://doi.org/10.1371/journal.pone.0189090

DeVries, T. J., von Keyserlingk, M. A. G., Weary, D. M., & Beauchemin, K. A. (2003). Measuring the Feeding Behavior of Lactating Dairy Cows in Early to Peak Lactation. *Journal of Dairy Science*, *86*(10), 3354–3361. https://doi.org/10.3168/jds.S0022-0302(03)73938-1

Dhoble, A. S., Ryan, K. T., Lahiri, P., Chen, M., Pang, X., Cardoso, F. C., & Bhalerao, K. D. (2019). Cytometric fingerprinting and machine learning (CFML): A novel label-free, objective method for routine mastitis screening. *Computers and Electronics in Agriculture*, *162*, 505–513. https://doi.org/10.1016/j.compag.2019.04.029

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39–54. https://doi.org/10.1111/1365-2656.12013

Dingemanse, N. J., Kazem, A. J. N., Réale, D., & Wright, J. (2010). Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, *25*(2), 81–89. https://doi.org/10.1016/J.TREE.2009.07.013

Dingemanse, N. J., & Réale, D. (2005). Natural selection and animal personality. *Behaviour*, *142*(9), 1159–1184. https://doi.org/10.1163/156853905774539445

Dohoo, I., Martin, W., & Stryhn, H. (2003). *Veterinary epidemiologic research: Vol. No V413* (2nd edition). https://www.islandscholar.ca/islandora/object/ir:ir-batch6-2657/DC/Default%20Dublin%20Core%20Record

Drissler, M., Gaworski, M., Tucker, C. B., & Weary, D. M. (2005). Freestall Maintenance: Effects on Lying Behavior of Dairy Cattle. *Journal of Dairy Science*, *88*(7), 2381–2387. https://doi.org/10.3168/jds.S0022-0302(05)72916-7

Eberhart, N. L., Storer, J. M., Caldwell, M., Saxton, A. M., & Krawczel, P. D. (2017). Behavioral and physiologic changes in Holstein steers experimentally infected with *Mannheimia haemolytica*. *American Journal of Veterinary Research*, *78*(9), 1056–1064. https://doi.org/10.2460/ajvr.78.9.1056

Eckelkamp, E. A., & Bewley, J. M. (2020). On-farm use of disease alerts generated by precision dairy technology. *Journal of Dairy Science*, *103*(2), 1566–1582. https://doi.org/10.3168/jds.2019-16888

Fagen, R. (1981). *Animal Play Behavior*. Oxford University Press. http://agris.fao.org/agris-search/search.do?recordID=US19820758466

FAWC. (1979). *Five Freedoms - Farm Animal Welfare Council*. National Archives FAWC.

Fawcett, T., & Flach, P. A. (2005). A response to Webb and Ting's on the application of ROC analysis to predict classification performance under varying class distributions. *Machine Learning*, *58*(1), 33–38. https://doi.org/10.1007/s10994-005-5256-4

Ferran, A. A., Toutain, P.-L., & Bousquet-Mélou, A. (2011). Impact of early versus later fluoroquinolone treatment on the clinical; microbiological and resistance outcomes in a mouse-lung model of Pasteurella multocida infection. *Veterinary Microbiology*, *148*(2–4), 292–297. https://doi.org/10.1016/j.vetmic.2010.09.005

Figo, D., Diniz, P. C., Ferreira, D. R., & Cardoso, J. M. P. (2010). Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, *14*(7), 645–662. https://doi.org/10.1007/s00779-010-0293-9

Finkemeier, M. A., Langbein, J., & Puppe, B. (2018). Personality research in mammalian farm animals: Concepts, measures, and relationship to welfare. In *Frontiers in Veterinary Science* (Vol. 5, Issue JUN, p. 131). Frontiers Media S.A. https://doi.org/10.3389/fvets.2018.00131

Foris, B., Zebunke, M., Langbein, J., & Melzer, N. (2018). Evaluating the temporal and situational consistency of personality traits in adult dairy cattle. *PLOS ONE*, *13*(10), e0204619. https://doi.org/10.1371/journal.pone.0204619

Forkman, B., Boissy, A., Meunier-Salaün, M.-C., Canali, E., & Jones, R. B. (2007). A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. *Physiology & Behavior*, *92*(3), 340–374. https://doi.org/10.1016/j.physbeh.2007.03.016

Forman, G. (2008). Quantifying counts and costs via classification. *Data Mining and Knowledge Discovery*, *17*(2), 164–206. https://doi.org/10.1007/s10618-008-0097-y

Fraley, R. C., & Roberts, B. W. (2005). Patterns of Continuity: A Dynamic Model for Conceptualizing the Stability of Individual Differences in

Psychological Constructs Across the Life Course. *Psychological Review*, *112*(1), 60–74. https://doi.org/10.1037/0033-295X.112.1.60

Frost, A. R., Schofield, C. P., Beaulah, S. A., Mottram, T. T., Lines, J. A., & Wathes, C. M. (1997). A review of livestock monitoring and the need for integrated systems. In *Computers and Electronics in Agriculture* (Vol. 17, Issue 2, pp. 139–159). Elsevier. https://doi.org/10.1016/s0168-1699(96)01301-4

Fuentes, A., Yoon, S., Park, J., & Park, D. S. (2020). Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. *Computers and Electronics in Agriculture*, *177*, 105627. https://doi.org/10.1016/j.compag.2020.105627

Fulton, R. W., & Confer, A. W. (2012). Laboratory test descriptions for bovine respiratory disease diagnosis and their strengths and weaknesses: gold standards for diagnosis, do they exist? *The Canadian Veterinary Journal = La Revue Veterinaire Canadienne*, *53*(7), 754–761.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, *46*(4), 1–37. https://doi.org/10.1145/2523813

García, R., Aguilar, J., Toro, M., Pinto, A., & Rodríguez, P. (2020). A systematic literature review on the use of machine learning in precision livestock farming. *Computers and Electronics in Agriculture*, *179*, 105826. https://doi.org/10.1016/j.compag.2020.105826

Ghirardi, J. J., Caja, G., Garín, D., Casellas, J., & Hernández-Jover, M. (2006). Evaluation of the retention of electronic identification boluses in the forestomachs of cattle. *Journal of Animal Science*, *84*(8), 2260–2268. https://doi.org/10.2527/jas.2005-758

Giger-Reverdin, S., Duvaux-Ponter, C., Sauvant, D., & Friggens, N. C. (2020). Repeatability of traits for characterizing feed intake patterns in dairy goats: a basis for phenotyping in the precision farming context. *Animal*, *14*(5), 1083. https://doi.org/10.1017/S1751731119002817

Gladden, N., Cuthbert, E., Ellis, K., & McKeegan, D. (2020). Use of a Tri-Axial Accelerometer Can Reliably Detect Play Behaviour in Newborn Calves. *Animals*, *10*(7), 1137. https://doi.org/10.3390/ani10071137

Goldberg, L. R. (1990). An alternative “description of personality”: The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

González, L. A., Tolkamp, B. J., Coffey, M. P., Ferret, A., & Kyriazakis, I. (2008). Changes in Feeding Behavior as Possible Indicators for the Automatic Monitoring of Health Disorders in Dairy Cows. *Journal of Dairy Science*, *91*(3), 1017–1028. https://doi.org/10.3168/jds.2007-0530

Gorden, P. J., & Plummer, P. (2010). Control, Management, and Prevention of Bovine Respiratory Disease in Dairy Calves and Cows. *Veterinary Clinics of North America: Food Animal Practice*, *26*(2), 243–259. https://doi.org/10.1016/j.cvfa.2010.03.004

Gosling, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin*, *127*(1), 45–86. https://doi.org/10.1037/0033-2909.127.1.45

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Graham, A. N., Renaud, D. L., Duffield, T. F., & Kelton, D. F. (2018). Short communication: Calf cleanliness does not predict diarrhea upon arrival at a veal calf facility. *Journal of Dairy Science*, *101*(4), 3363–3366. https://doi.org/10.3168/jds.2017-14113

Grigor, P. N., Cockram, M. S., Steele, W. B., Le Sueur, C. J., Forsyth, R. E., Guthrie, J. A., Johnson, A. K., Sandilands, V., Reid, H. W., Sinclair, C., & Brown, H. K. (2001). Effects of space allowance during transport and duration of mid-journey lairage period on the physiological, behavioural and immunological responses of young calves during and after transport. *Animal Science*, *73*(2), 341–360. https://doi.org/10.1017/S135772980005832X

Größbacher, V., Bučková, K., Lawrence, A. B., Špinka, M., & Winckler, C. (2020). Discriminating spontaneous locomotor play of dairy calves using accelerometers. *Journal of Dairy Science*, *103*(2), 1866–1873. https://doi.org/10.3168/jds.2019-17005

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guo, L., Welch, M., Dobos, R., Kwan, P., & Wang, W. (2018). Comparison of grazing behaviour of sheep on pasture with different sward surface heights using an inertial measurement unit sensor. *Computers and Electronics in Agriculture*, *150*, 394–401. https://doi.org/10.1016/j.compag.2018.05.004

Guo, Y., He, D., & Chai, L. (2020). A Machine Vision-Based Method for Monitoring Scene-Interactive Behaviors of Dairy Calf. *Animals*, *10*(2), 190. https://doi.org/10.3390/ani10020190

Haley, D. B., Rushen, J., & Passillé, A. M. de. (2000). Behavioural indicators of cow comfort: activity and resting behaviour of dairy cows in two types of housing. *Canadian Journal of Animal Science*, *80*(2), 257–263. https://doi.org/10.4141/A99-084

Hanzlicek, G. A., White, B. J., Mosier, D., Renter, D. G., & Anderson, D. E. (2010). Serial evaluation of physiologic, pathological, and behavioral changes related to disease progression of experimentally induced *Mannheimia haemolytica* pneumonia in postweaned calves. *American Journal of Veterinary Research*, *71*(3), 359–369. https://doi.org/10.2460/ajvr.71.3.359

Harris, S. M., Descamps, S., Sneddon, L. U., Bertrand, P., Chastel, O., & Patrick, S. C. (2020). Personality predicts foraging site fidelity and trip repeatability in a marine predator. *Journal of Animal Ecology*, *89*(1), 68–79. https://doi.org/10.1111/1365-2656.13106

Hart, B. L. (1988). Biological basis of the behavior of sick animals. *Neuroscience & Biobehavioral Reviews*, *12*(2), 123–137. https://doi.org/10.1016/S0149-7634(88)80004-6

Haskell, M. J., Simm, G., & Turner, S. P. (2014). Genetic selection for temperament traits in dairy and beef cattle. *Frontiers in Genetics*, *5*. https://doi.org/10.3389/fgene.2014.00368

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Heinrichs, A. J. (1993). Raising Dairy Replacements to Meet the Needs of the 21st Century. *Journal of Dairy Science*, *76*(10), 3179–3187. https://doi.org/10.3168/jds.S0022-0302(93)77656-0

Hertel, A. G., Niemelä, P. T., Dingemanse, N. J., & Mueller, T. (2020). A guide for studying among-individual behavioral variation from movement data in the wild. *Movement Ecology 2020 8:1*, *8*(1), 1–18. https://doi.org/10.1186/S40462-020-00216-8

Hertel, A. G., Royauté, R., Zedrosser, A., & Mueller, T. (2021). Biologging reveals individual variation in behavioural predictability in the wild. *Journal of Animal Ecology*, *90*(3), 723–737. https://doi.org/10.1111/1365-2656.13406

Hixson, C. L., Krawczel, P. D., Caldwell, J. M., & Miller-Cushon, E. K. (2018). Behavioral changes in group-housed dairy calves infected with Mannheimia haemolytica. *Journal of Dairy Science*, *101*(11), 10351–10360. https://doi.org/10.3168/JDS.2018-14832

Hoadley, B. (2001). Comment on “Statistical modeling: The two cultures,” by L. Breiman. *Statistical Science*, *16*(3), 199–231.

Hokkanen, A.-H., Hänninen, L., Tiusanen, J., & Pastell, M. (2011). Predicting sleep and lying time of calves with a support vector machine classifier using accelerometer data. *Applied Animal Behaviour Science*, *134*(1–2), 10–15. https://doi.org/10.1016/J.APPLANIM.2011.06.016

Houslay, T. M., & Wilson, A. J. (2017). Avoiding the misuse of BLUP in behavioural ecology. *Behavioral Ecology*, *28*(4), 948. https://doi.org/10.1093/BEHECO/ARX023

Howie, J. A., Tolkamp, B. J., Avendaño, S., & Kyriazakis, I. (2009). A novel flexible method to split feeding behaviour into bouts. *Applied Animal Behaviour Science*, *116*(2–4), 101–109. https://doi.org/10.1016/j.applanim.2008.09.005

Hudson, C., Kaler, J., & Down, P. (2018). Using big data in cattle practice. *In Practice*, *40*(9), 396–410. https://doi.org/10.1136/inp.k4328

Hyde, R. M., Remnant, J. G., Bradley, A. J., Breen, J. E., Hudson, C. D., Davies, P. L., Clarke, T., Critchell, Y., Hylands, M., Linton, E., Wood, E., & Green, M. J. (2017). Quantitative analysis of antimicrobial use on British dairy farms. *Veterinary Record*, *181*(25), 683–683. https://doi.org/10.1136/vr.104614

Ives, S. E., & Richeson, J. T. (2015). Use of Antimicrobial Metaphylaxis for the Control of Bovine Respiratory Disease in High-Risk Cattle. *Veterinary Clinics of North America: Food Animal Practice*, *31*(3), 341–350. https://doi.org/10.1016/j.cvfa.2015.05.008

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (Vol. 103). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Janzen, E. D., Stockdale, P. H., Acres, S. D., & Babiuk, L. A. (1984). Therapeutic and prophylactic effects of some antibiotics on experimental pneumonic pasteurellosis. *The Canadian Veterinary Journal = La Revue Veterinaire Canadienne*, *25*(2), 78–81. http://www.ncbi.nlm.nih.gov/pubmed/17422363

J.D. Davis, E.S. Vanzant, J.L. Purswell, A.R. Green, J.R. Bicudo, R.S. Gates, L.E. Holloway, & W.T. Smith. (2003). Methods of Remote, Continuous

Temperature Detection in Beef Cattle. *2003, Las Vegas, NV July 27-30, 2003*. https://doi.org/10.13031/2013.15646

Jensen, M. B., & Holm, L. (2003). The effect of milk flow rate and milk allowance on feeding related behaviour in dairy calves fed by computer controlled milk feeders. *Applied Animal Behaviour Science*, *82*(2), 87–100. https://doi.org/10.1016/S0168-1591(03)00054-6

Jiang, M., Rao, Y., Zhang, J., & Shen, Y. (2020). Automatic behavior recognition of group-housed goats using deep learning. *Computers and Electronics in Agriculture*, *177*, 105706. https://doi.org/10.1016/j.compag.2020.105706

Jocher, G. (2022). YOLOv5. *GitHub Repository*.

Johnson, K. F., Chancellor, N., Burn, C. C., & Wathes, D. C. (2017). Prospective cohort study to assess rates of contagious disease in pre-weaned UK dairy heifers: management practices, passive transfer of immunity and associated calf health. *Veterinary Record Open*, *4*(1), e000226. https://doi.org/10.1136/vetreco-2017-000226

Johnson, R. W. (2002). The concept of sickness behavior: a brief chronological account of four key discoveries. *Veterinary Immunology and Immunopathology*, *87*(3–4), 443–450. https://doi.org/10.1016/S0165-2427(02)00069-7

Johnston, D., Kenny, D. A., McGee, M., Waters, S. M., Kelly, A. K., & Earley, B. (2016). Electronic feeding behavioural data as indicators of health status in dairy calves. *Irish Journal of Agricultural and Food Research*, *55*(2), 159–168. https://doi.org/10.1515/ijafr-2016-0016

Jolles, J. W., Briggs, H. D., Araya-Ajoy, Y. G., & Boogert, N. J. (2019). Personality, plasticity and predictability in sticklebacks: bold fish are less plastic and more predictable than shy fish. *Animal Behaviour*, *154*, 193–202. https://doi.org/10.1016/J.ANBEHAV.2019.06.022

Kaiser, M. I., & Müller, C. (2021). What is an animal personality? *Biology & Philosophy*, *36*(1), 1. https://doi.org/10.1007/s10539-020-09776-w

Keceli, A. S., Catal, C., Kaya, A., & Tekinerdogan, B. (2020). Development of a recurrent neural networks-based calving prediction model using activity and behavioral data. *Computers and Electronics in Agriculture*, *170*, 105285. https://doi.org/10.1016/j.compag.2020.105285

Kelley, K. W., Bluthé, R.-M., Dantzer, R., Zhou, J.-H., Shen, W.-H., Johnson, R. W., & Broussard, S. R. (2003). Cytokine-induced sickness behavior. *Brain, Behavior, and Immunity*, *17*(1), 112–118. https://doi.org/10.1016/S0889-1591(02)00077-6

Kelly, D. N., Sleator, R. D., Murphy, C. P., Conroy, S. B., & Berry, D. P. (2021). Phenotypic and genetic associations between feeding behavior and carcass merit in crossbred growing cattle. *Journal of Animal Science*, *99*(12). https://doi.org/10.1093/jas/skab285

Kelly, D. N., Sleator, R. D., Murphy, C. P., Conroy, S. B., Judge, M. M., & Berry, D. P. (2020). Large variability in feeding behavior among crossbred growing cattle. *Journal of Animal Science*, *98*(7), 1–10. https://doi.org/10.1093/JAS/SKAA216

Kent, S., Bluthé, R.-M., Kelley, K. W., & Dantzer, R. (1992). Sickness behavior as a new target for drug development. *Trends in Pharmacological Sciences*, *13*, 24–28. https://doi.org/10.1016/0165-6147(92)90012-U

Khan, A., Hammerla, N., Mellor, S., & Plötz, T. (2016). Optimising sampling rates for accelerometer-based human activity recognition. *Pattern Recognition Letters*, *73*, 33–40. https://doi.org/10.1016/j.patrec.2016.01.001

Khan, M. A., Lee, H. J., Lee, W. S., Kim, H. S., Ki, K. S., Hur, T. Y., Suh, G. H., Kang, S. J., & Choi, Y. J. (2007). Structural Growth, Rumen Development, and Metabolic and Immune Responses of Holstein Male Calves Fed Milk Through Step-Down and Conventional Methods. *Journal of Dairy Science*, *90*(7), 3376–3387. https://doi.org/10.3168/jds.2007-0104

Klefot, J. M., Murphy, J. L., Donohue, K. D., O'Hara, B. F., Lhamon, M. E., & Bewley, J. M. (2016). Development of a noninvasive system for monitoring dairy cattle sleep. *Journal of Dairy Science*, *99*(10), 8477–8485. https://doi.org/10.3168/JDS.2015-10695

Kluen, E., & Brommer, J. E. (2013). Context-specific repeatability of personality traits in a wild bird: a reaction-norm perspective. *Behavioral Ecology*, *24*(3), 650–658. https://doi.org/10.1093/BEHECO/ARS221

Kluger, M. J. (1991). Fever: Role of Pyrogens and Cryogens. *Physiological Reviews*, *71*(1), 93–127. https://doi.org/10.1152/physrev.1991.71.1.93

Kluger, M. J., Kozak, W., Conn, C. A., Leon, L. R., & Soszynski, D. (1998). Role of Fever in Disease. *Annals of the New York Academy of Sciences*, *856*(1 MOLECULAR MEC), 224–233. https://doi.org/10.1111/j.1749-6632.1998.tb08329.x

Knauer, W., Godden, S., Dietrich, A., Hawkins, D., & James, R. (2018). Evaluation of applying statistical process control techniques to daily average feeding behaviors to detect disease in automatically fed group-housed preweaned dairy calves. *Journal of Dairy Science*, *101*, 8135–8145. https://doi.org/10.3168/jds.2017-13947

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *In Proceedings of the 14th International Joint Conference on Artificial Intelligence*, *2*, 1137–1143. http://robotics.stanford.edu/~ronnyk

Kononenko, I. (1994). Estimating attributes: Analysis and extensions of RELIEF. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *784 LNCS*, 171–182. https://doi.org/10.1007/3-540-57868-4_57

Koolhaas, J. M. (2008). Coping style and immunity in animals: Making sense of individual variation. *Brain, Behavior, and Immunity*, *22*(5), 662–667. https://doi.org/10.1016/J.BBI.2007.11.006

Koolhaas, J. M., & Van Reenen, C. G. (2016). ANIMAL BEHAVIOR AND WELL-BEING SYMPOSIUM: Interaction between coping style/personality, stress, and welfare: Relevance for domestic farm animals1. *Journal of Animal Science*, *94*(6), 2284–2296. https://doi.org/10.2527/jas.2015-0125

Kour, H., Patison, K. P., Corbet, N. J., & Swain, D. L. (2018). Validation of accelerometer use to measure suckling behaviour in Northern Australian beef calves. *Applied Animal Behaviour Science*, *202*, 1–6. https://doi.org/10.1016/j.applanim.2018.01.012

Krachun, C., Rushen, J., & de Passillé, A. M. (2010). Play behaviour in dairy calves is reduced by weaning and by a low energy intake. *Applied Animal Behaviour Science*, *122*(2–4), 71–76. https://doi.org/10.1016/j.applanim.2009.12.002

Krohn, C. C., Munksgaard, L., & Jonasen, B. (1992). Behaviour of dairy cows kept in extensive (loose housing/pasture) or intensive (tie stall) environments I. Experimental procedure, facilities, time budgets — diurnal and seasonal conditions. *Applied Animal Behaviour Science*, *34*(1–2), 37–47. https://doi.org/10.1016/S0168-1591(05)80055-3

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., & Kenkel, B. (2022). *Caret: Classification and Regression Training* (6.0-86). CRAN.

Kuźnicka, E., & Gburzyński, P. (2017). Automatic detection of suckling events in lamb through accelerometer data classification. *Computers and Electronics in Agriculture*, *138*, 137–147. https://doi.org/10.1016/J.COMPAG.2017.04.009

Lauber, M. C. Y., Hemsworth, P. H., & Barnett, J. L. (2006). The effects of age and experience on behavioural development in dairy calves. *Applied Animal Behaviour Science*, *99*(1–2), 41–52. https://doi.org/10.1016/j.applanim.2005.10.009

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Ledgerwood, D. N., Winckler, C., & Tucker, C. B. (2010). Evaluation of data loggers, sampling intervals, and editing techniques for measuring the lying behavior of dairy cattle. *Journal of Dairy Science*, *93*(11), 5129–5139. https://doi.org/10.3168/jds.2009-2945

Liang, D., Wood, C. L., McQuerry, K. J., Ray, D. L., Clark, J. D., & Bewley, J. M. (2013). Influence of breed, milk production, season, and ambient temperature on dairy cow reticulorumen temperature. *Journal of Dairy Science*, *96*(8), 5072–5081. https://doi.org/10.3168/jds.2012-6537

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2014). *Microsoft COCO: Common Objects in Context*.

Liu, D., Oczak, M., Maschat, K., Baumgartner, J., Pletzer, B., He, D., & Norton, T. (2020). A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs. *Biosystems Engineering*, *195*, 27–41. https://doi.org/10.1016/j.biosystemseng.2020.04.007

Lokhorst, C., de Mol, R. M., & Kamphuis, C. (2019). Invited review: Big Data in precision dairy farming. *Animal*, *13*(7), 1519–1528. https://doi.org/10.1017/S1751731118003439

Luu, J., Johnsen, J. F., Passillé, A. M. de, & Rushen, J. (2013). Which measures of acceleration best estimate the duration of locomotor play by dairy calves? *Applied Animal Behaviour Science*, *148*(1–2), 21–27. https://doi.org/10.1016/J.APPLANIM.2013.07.004

Mahendran, S. A., Booth, R., Bell, N. J., & Burge, M. (2017). Randomised positive control trial of NSAID and antimicrobial treatment for calf fever caused by pneumonia. *The Veterinary Record*, *181*(2), 45. https://doi.org/10.1136/vr.104057

Maier, A., Syben, C., Lasser, T., & Riess, C. (2019). A gentle introduction to deep learning in medical image processing. *Zeitschrift Für Medizinische Physik*, *29*(2), 86–101. https://doi.org/10.1016/j.zemedi.2018.12.003

Mansbridge, N., Mitsch, J., Bollard, N., Ellis, K., Miguel-Pacheco, G., Dottorini, T., & Kaler, J. (2018). Feature Selection and Comparison of Machine Learning Algorithms in Classification of Grazing and Rumination Behaviour in Sheep. *Sensors*, *18*(10), 3532. https://doi.org/10.3390/s18103532

Martiskainen, P., Järvinen, M., Skön, J. P., Tiirikainen, J., Kolehmainen, M., & Mononen, J. (2009). Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied Animal Behaviour Science*, *119*(1–2), 32–38. https://doi.org/10.1016/j.applanim.2009.03.005

Matthews, S. G., Miller, A. L., Clapp, J., Plötz, T., & Kyriazakis, I. (2016). Early detection of health and welfare compromises through automated detection of behavioural changes in pigs. *The Veterinary Journal*, *217*, 43–51. https://doi.org/10.1016/J.TVJL.2016.09.005

Mattiello, Battini, De Rosa, Napolitano, & Dwyer. (2019). How Can We Assess Positive Welfare in Ruminants? *Animals*, *9*(10), 758. https://doi.org/10.3390/ani9100758

McCorkell, R., Wynne-Edwards, K., Windeyer, C., Schaefer, A., & UCVM Class of 2013, the U. C. of. (2014). Limited efficacy of Fever Tag(®) temperature sensing ear tags in calves with naturally occurring bovine respiratory disease or induced bovine viral diarrhea virus infection. *The Canadian Veterinary Journal = La Revue Veterinaire Canadienne*, *55*(7), 688–690. http://www.ncbi.nlm.nih.gov/pubmed/24982523

McCulloch, S. P. (2013). A Critique of FAWC's Five Freedoms as a Framework for the Analysis of Animal Welfare. *Journal of Agricultural and Environmental Ethics*, *26*(5), 959–975. https://doi.org/10.1007/s10806-012-9434-7

McGuirk, S. M., & Peek, S. F. (2014). Timely diagnosis of dairy calf respiratory disease using a standardized scoring system. *Animal Health Research Reviews*, *15*(2), 145–147. https://doi.org/10.1017/S1466252314000267

Mellor, D. (2015). Enhancing animal welfare by creating opportunities for positive affective engagement. *New Zealand Veterinary Journal*, *63*(1), 3–8. https://doi.org/10.1080/00480169.2014.926799

Mellor, D. (2016). Updating Animal Welfare Thinking: Moving beyond the "Five Freedoms" towards "A Life Worth Living." *Animals*, *6*(3), 21. https://doi.org/10.3390/ani6030021

Miller-Cushon, E. K., Bergeron, R., Leslie, K. E., & DeVries, T. J. (2013). Effect of milk feeding level on development of feeding behavior in dairy calves. *Journal of Dairy Science*, *96*(1), 551–564. https://doi.org/10.3168/JDS.2012-5937

Mitchell, D. J., Beckmann, C., & Biro, P. A. (2021). Understanding the unexplained: The magnitude and correlates of individual differences in residual variance. *Ecology and Evolution*, *11*(12), 7201–7210. https://doi.org/10.1002/ece3.7603

Morrison, J., Renaud, D. L., Churchill, K. J., Costa, J. H. C., Steele, M. A., & Winder, C. B. (2021). Predicting morbidity and mortality using automated milk feeders: A scoping review. *Journal of Dairy Science*, *104*(6), 7177–7194. https://doi.org/10.3168/jds.2020-19645

Mottram, T. (2016). Animal board invited review: precision livestock farming for dairy cows with a focus on oestrus detection. *Animal*, *10*(10), 1575–1584. https://doi.org/10.1017/S1751731115002517

Müller, R., & von Keyserlingk, M. A. G. (2006). Consistency of flight speed and its correlation to productivity and to personality in Bos taurus beef cattle. *Applied Animal Behaviour Science*, *99*(3–4), 193–204. https://doi.org/10.1016/j.applanim.2005.05.012

Murphy, E., Nordquist, R. E., & van der Staay, F. J. (2014a). A review of behavioural methods to study emotion and mood in pigs, Sus scrofa. *Applied Animal Behaviour Science*, *159*, 9–28. https://doi.org/10.1016/j.applanim.2014.08.002

Murphy, E., Nordquist, R. E., & van der Staay, F. J. (2014b). A review of behavioural methods to study emotion and mood in pigs, Sus scrofa. *Applied Animal Behaviour Science*, *159*, 9–28. https://doi.org/10.1016/j.applanim.2014.08.002

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, no-no. https://doi.org/10.1111/j.1469-185X.2010.00141.x

Napolitano, F., Knierim, U., Grass, F., & De Rosa, G. (2009). Positive indicators of cattle welfare and their applicability to on-farm protocols. *Italian Journal of Animal Science*, *8*(sup1), 355–365. https://doi.org/10.4081/ijas.2009.s1.355

National Animal Disease Information Service. (2015). *Respiratory disease in dairy and beef rearer units*. National Animal Disease Information Service Retrieved on 20 August from Https://Www.Nadis.Org.Uk/Disease-a-z/Cattle/Calf-Management/Respiratory-Disease-in-Dairy-and-Beef-Rearer-Units/ 2022.

Neave, H. W., Costa, J. H. C., Benetton, J. B., Weary, D. M., & von Keyserlingk, M. A. G. (2019). Individual characteristics in early life relate to variability in weaning age, feeding behavior, and weight gain of dairy calves automatically weaned based on solid feed intake. *Journal of Dairy Science*, *102*(11), 10250–10265. https://doi.org/10.3168/jds.2019-16438

Neave, H. W., Costa, J. H. C., Weary, D. M., & von Keyserlingk, M. A. G. (2018). Personality is associated with feeding behavior and performance in dairy calves. *Journal of Dairy Science*, *101*(8), 7437–7449. https://doi.org/10.3168/jds.2017-14248

Neave, H. W., Costa, J. H. C., Weary, D. M., & von Keyserlingk, M. A. G. (2020). Long-term consistency of personality traits of cattle. *Royal Society Open Science*, *7*(2), 191849. https://doi.org/10.1098/rsos.191849

Neave, H. W., Weary, D. M., & Von Keyserlingk, M. A. G. (2018). *Review: Individual variability in feeding behaviour of domesticated ruminants*. https://doi.org/10.1017/S1751731118001325

Neethirajan, S. (2020). The role of sensors, big data and machine learning in modern animal farming. *Sensing and Bio-Sensing Research*, *29*, 100367. https://doi.org/10.1016/j.sbsr.2020.100367

Neja, W., Sawa, A., Jankowska, M., Bogucki, M., & Krężel-Czopek, S. (2015). Effect of the temperament of dairy cows on lifetime production efficiency. *Archives Animal Breeding*, *58*(1), 193–197. https://doi.org/10.5194/aab-58-193-2015

Nielsen, B. L. (1999). On the interpretation of feeding behaviour measures and the use of feeding rate as an indicator of social constraint. *Applied Animal Behaviour Science*, *63*(1), 79–91. https://doi.org/10.1016/S0168-1591(99)00003-9

Nielsen, B. L., Lawrence, A. B., & Whittemore, C. T. (1995). Effect of group size on feeding behaviour, social behaviour, and performance of growing pigs using single-space feeders. *Livestock Production Science*, *44*(1), 73–85. https://doi.org/10.1016/0301-6226(95)00060-X

Nielsen, L. H., Mogensen, L., Krohn, C., Hindhede, J., & Sørensen, J. T. (1997). Resting and social behaviour of dairy heifers housed in slatted floor pens with different sized bedded lying areas. *Applied Animal Behaviour Science*, *54*(4), 307–316. https://doi.org/10.1016/S0168-1591(96)01211-7

Niemann, H. (2013). *Pattern analysis and understanding* (Vol. 4). Springer Science & Business Media.

Norring, M., Manninen, E., de Passillé, A. M., Rushen, J., & Saloniemi, H. (2010). Preferences of dairy cows for three stall surface materials with small amounts of bedding. *Journal of Dairy Science*, *93*(1), 70–74. https://doi.org/10.3168/jds.2009-2164

Occhiuto, F., Vázquez-Diosdado, J. A., Carslake, C., & Kaler, J. (2022). Personality and predictability in farmed calves using movement and space-use behaviours quantified by ultra-wideband sensors. *Royal Society Open Science*, *9*(6). https://doi.org/10.1098/rsos.212019

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Peterson, R. A., & Cavanaugh, J. E. (2020). Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, *47*(13–15), 2312–2327. https://doi.org/10.1080/02664763.2019.1630372

Poppe, M., Veerkamp, R. F., van Pelt, M. L., & Mulder, H. A. (2020). Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. *Journal of Dairy Science*, *103*(2), 1667–1684. https://doi.org/10.3168/JDS.2019-17290

Porto, S. M. C., Arcidiacono, C., Anguzza, U., & Cascone, G. (2013). A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns. *Biosystems Engineering*, *115*(2), 184–194. https://doi.org/10.1016/j.biosystemseng.2013.03.002

Porto, S. M. C., Arcidiacono, C., Anguzza, U., & Cascone, G. (2015). The automatic detection of dairy cow feeding and standing behaviours in free-stall barns by a computer vision-based system. *Biosystems Engineering*, *133*, 46–55. https://doi.org/10.1016/j.biosystemseng.2015.02.012

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., & Howard, D. (2009). A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities From Accelerometer Data. *IEEE Transactions on Biomedical Engineering*, *56*(3), 871–879. https://doi.org/10.1109/TBME.2008.2006190

Proudfoot, K. L., Weary, D. M., & von Keyserlingk, M. A. G. (2012). Linking the social environment to illness in farm animals. *Applied Animal Behaviour Science*. https://doi.org/10.1016/j.applanim.2012.02.008

Quimby, W. F., Sowell, B. F., Bowman, J. G. P., Branine, M. E., Hubbert, M. E., & Sherwood, H. W. (2001). Application of feeding behaviour to predict morbidity of newly received calves in a commercial feedlot. *Canadian Journal of Animal Science*, *81*(3), 315–320. https://doi.org/10.4141/A00-098

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Réale, D., Garant, D., Humphries, M. M., Bergeron, P., Careau, V., & Montiglio, P.-O. (2010). Personality and the emergence of the pace-of-life syndrome concept at the population level. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4051–4063. https://doi.org/10.1098/rstb.2010.0208

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318. https://doi.org/10.1111/J.1469-185X.2007.00010.X

Rebala, G., Ajay, R., & Sanjay, C. (2019). *An introduction to machine learning*. Springer.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). *You Only Look Once: Unified, Real-Time Object Detection*.

Reuter, R. R., Carroll, J. A., Hulbert, L. E., Dailey, J. W., & Galyean, M. L. (2010). Technical note: Development of a self-contained, indwelling rectal temperature probe for cattle research12. *Journal of Animal Science*, *88*(10), 3291–3295. https://doi.org/10.2527/jas.2010-3093

Reynolds, J., & Brennan, M. (2021). Is thoracic ultrasound more efficient than the Wisconsin calf scoring system for the detection of pneumonia in calves? *Veterinary Record*, *189*(2), 73–75. https://doi.org/10.1002/vetr.752

Robert, B., White, B. J., Renter, D. G., & Larson, R. L. (2009). Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. *Computers and Electronics in Agriculture*, *67*(1–2), 80–84. https://doi.org/10.1016/j.compag.2009.03.002

Rodrigues, J. P. P., Pereira, L. G. R., Diniz Neto, H. do C., Lombardi, M. C., Lage, C. F. de A., Coelho, S. G., Sacramento, J. P., Machado, F. S., Tomich, T. R., Maurício, R. M., & Campos, M. M. (2019). Technical note: Evaluation of an automatic system for monitoring rumination time in weaning calves. *Livestock Science*, *219*, 86–90. https://doi.org/10.1016/j.livsci.2018.11.017

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, *33*(1–2), 1–39. https://doi.org/10.1007/s10462-009-9124-7

Roland, L., Schweinzer, V., Kanz, P., Sattlecker, G., Kickinger, F., Lidauer, L., Berger, A., Auer, W., Mayer, J., Sturm, V., Efrosinin, D., Breitenberger, S., Drillich, M., & Iwersen, M. (2018). Technical note: Evaluation of a triaxial accelerometer for monitoring selected behaviors in dairy calves. *Journal of Dairy Science*, *101*(11), 10421–10427. https://doi.org/10.3168/JDS.2018-14720

Rose-Dye, T. K., Burciaga-Robles, L. O., Krehbiel, C. R., Step, D. L., Fulton, R. W., Confer, A. W., & Richards, C. J. (2011). Rumen temperature change monitored with remote rumen temperature boluses after challenges with bovine viral diarrhea virus and Mannheimia haemolytica12. *Journal of Animal Science*, *89*(4), 1193–1200. https://doi.org/10.2527/jas.2010-3051

Rosenberger, K., Costa, J. H. C., Neave, H. W., von Keyserlingk, M. A. G., & Weary, D. M. (2017). The effect of milk allowance on behavior and weight gains in dairy calves. *Journal of Dairy Science*, *100*(1), 504–512. https://doi.org/10.3168/JDS.2016-11195

Roth, B. A., Hillmann, E., Stauffacher, M., & Keil, N. M. (2008). Improved weaning reduces cross-sucking and may improve weight gain in dairy calves. *Applied Animal Behaviour Science*, *111*(3–4), 251–261. https://doi.org/10.1016/j.applanim.2007.06.007

Roussel, A. J., & Brumbaugh, G. W. (1991). Treatment of Diarrhea of Neonatal Calves. *Veterinary Clinics of North America: Food Animal Practice*, *7*(3), 713–728. https://doi.org/10.1016/S0749-0720(15)31081-1

Ruckerbusch, Y. (1975). The hypnogram as an index of adaptation of farm animals to changes in their environment. *Applied Animal Ethology*, *2*(1), 3–18. https://doi.org/10.1016/0304-3762(75)90061-9

Rushen, J., & de Passillé, A. M. (2012). Automated measurement of acceleration can detect effects of age, dehorning and weaning on locomotor play of calves. *Applied Animal Behaviour Science*, *139*(3–4), 169–174. https://doi.org/10.1016/j.applanim.2012.04.011

Sahu, D., Mandal, D. K., Hussain Dar, A., Podder, M., & Gupta, A. (2021). Modification in housing system affects the behavior and welfare of dairy Jersey crossbred cows in different seasons. *Biological Rhythm Research*, *52*(9), 1303–1312. https://doi.org/10.1080/09291016.2019.1619130

Santostefano, F., Wilson, A. J., Niemelä, P. T., & Dingemanse, N. J. (2017). Behavioural mediators of genetic life-history trade-offs: a test of the pace-of-life syndrome hypothesis in field crickets. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1864), 20171567. https://doi.org/10.1098/rspb.2017.1567

Schaefer, A. L., Cook, N. J., Bench, C., Chabot, J. B., Colyn, J., Liu, T., Okine, E. K., Stewart, M., & Webster, J. R. (2012). The non-invasive and automated detection of bovine respiratory disease onset in receiver calves using infrared thermography. *Research in Veterinary Science*, *93*(2), 928–935. https://doi.org/10.1016/j.rvsc.2011.09.021

Schaefer, A. L., Cook, N., Tessaro, S. V., Deregt, D., Desroches, G., Dubeski, P. L., Tong, A. K. W., & Godson, D. L. (2004). Early detection and prediction of infection using infrared thermography. *Canadian Journal of Animal Science*, *84*(1), 73–80. https://doi.org/10.4141/A02-104

Schaffer, A. P., Larson, R. L., Cernicchiaro, N., Hanzlicek, G. A., Bartle, S. J., & Thomson, D. U. (2016). The association between calfhood bovine respiratory disease complex and subsequent departure from the herd, milk production, and reproduction in dairy cattle. *Journal of the American Veterinary Medical Association*, *248*(10), 1157–1164. https://doi.org/10.2460/javma.248.10.1157

Schillings, J., Bennett, R., & Rose, D. C. (2021). Exploring the Potential of Precision Livestock Farming Technologies to Help Address Farm Animal

Welfare. *Frontiers in Animal Science*, *0*, 13. https://doi.org/10.3389/FANIM.2021.639678

Schutz, M. M., & Bewley, J. M. (2009). *Implications of Changes in Core Body Temperature Compost bedded loose housing dairy barn*. https://www.researchgate.net/publication/228683896

Scott, P. (2013). Responsible use of antimicrobial drugs in bovine respiratory disease. *Livestock*, *18*(4), 109–113. https://doi.org/10.12968/live.2013.18.4.109

Sherwin, V. E., Hudson, C. D., Henderson, A., & Green, M. J. (2016). The association between age at first calving and survival of first lactation heifers within dairy herds. *Animal*, *10*(11), 1877–1882. https://doi.org/10.1017/S1751731116000689

Sievers, A., Kristensen, N., Laue, H.-J., & Wolffram, S. (2004). Development of an intraruminal device for data sampling and transmission. *Journal of Animal and Feed Sciences*, *13*(Suppl. 1), 207–210. https://doi.org/10.22358/jafs/73840/2004

Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution*, *19*(7), 372–378. https://doi.org/10.1016/j.tree.2004.04.009

Sih, A., Cote, J., Evans, M., Fogarty, S., & Pruitt, J. (2012). Ecological implications of behavioural syndromes. *Ecology Letters*, *15*(3), 278–289. https://doi.org/10.1111/J.1461-0248.2011.01731.X

Sivula, N. J., Ames, T. R., Marsh, W. E., & Werdin, R. E. (1996). Descriptive epidemiology of morbidity and mortality in Minnesota dairy heifer calves. *Preventive Veterinary Medicine*, *27*(3–4), 155–171. https://doi.org/10.1016/0167-5877(95)01000-9

Slob, N., Catal, C., & Kassahun, A. (2021). Application of machine learning to improve dairy farm management: A systematic literature review. *Preventive Veterinary Medicine*, *187*, 105237. https://doi.org/10.1016/j.prevetmed.2020.105237

Smith RA. (2000). Effects of feedlot disease on economics, production and carcass value. *Proceedings of the Thirty-Third Annual Conference, American Association of Bovine Practitioners Rapid City, SD* . https://www.cabdirect.org/cabdirect/abstract/20003012115

Snowder, G. D., Van Vleck, L. D., Cundiff, L. V., & Bennett, G. L. (2006). Bovine respiratory disease in feedlot cattle: Environmental, genetic, and economic factors. *Journal of Animal Science*, *84*(8), 1999–2008. https://doi.org/10.2527/jas.2006-046

Stamps, J. A. (2007). Growth-mortality tradeoffs and "personality traits" in animals. *Ecology Letters*, *10*(5), 355–363. https://doi.org/10.1111/j.1461-0248.2007.01034.x

Stamps, J. A., Briffa, M., & Biro, P. A. (2012). Unpredictable animals: individual differences in intraindividual variability (IIV). *Animal Behaviour*, *83*(6), 1325–1334. https://doi.org/10.1016/J.ANBEHAV.2012.02.017

Stone, A. E., Jones, B. W., Becker, C. A., & Bewley, J. M. (2017). Influence of breed, milk yield, and temperature-humidity index on dairy cow lying time, neck activity, reticulorumen temperature, and rumination behavior. *Journal of Dairy Science*, *100*(3), 2395–2403. https://doi.org/10.3168/jds.2016-11607

Sumner, C. L., von Keyserlingk, M. A. G., & Weary, D. M. (2018). Perspectives of farmers and veterinarians concerning dairy cattle welfare. *Animal Frontiers*, *8*(1), 8–13. https://doi.org/10.1093/af/vfx006

Svensson, C., & Jensen, M. B. (2007). Short Communication: Identification of Diseased Calves by Use of Data from Automatic Milk Feeders. *Journal of Dairy Science*, *90*(2), 994–997. https://doi.org/10.3168/JDS.S0022-0302(07)71584-9

Swartz, T. H., McGilliard, M. L., & Petersson-Wolfe, C. S. (2016). Technical note: The use of an accelerometer for measuring step activity and lying behaviors in dairy calves. *Journal of Dairy Science*, *99*(11), 9109–9113. https://doi.org/10.3168/jds.2016-11297

Sweeney, B. C., Rushen, J., Weary, D. M., & de Passillé, A. M. (2010). Duration of weaning, starter intake, and weight gain of dairy calves fed large amounts of milk. *Journal of Dairy Science*, *93*(1), 148–152. https://doi.org/10.3168/JDS.2009-2427

Taneja, M., Byabazaire, J., Jalodia, N., Davy, A., Olariu, C., & Malone, P. (2020). Machine learning based fog computing assisted data-driven approach for early lameness detection in dairy cattle. *Computers and Electronics in Agriculture*, *171*, 105286. https://doi.org/10.1016/j.compag.2020.105286

Taylor, J. D., Fulton, R. W., Lehenbauer, T. W., Step, D. L., & Confer, A. W. (2010). The epidemiology of bovine respiratory disease: what is the evidence for preventive measures? *The Canadian Veterinary Journal = La Revue Veterinaire Canadienne*, *51*(12), 1351–1359. http://www.ncbi.nlm.nih.gov/pubmed/21358927

Theurer, M. E., Anderson, D. E., White, B. J., Miesner, M. D., Mosier, D. A., Coetzee, J. F., Lakritz, J., & Amrine, D. E. (2013). Effect of Mannheimia Haemolytica pneumonia on behavior and physiologic responses of calves during high ambient environmental temperatures. *Journal of Animal Science*, *91*(8), 3917–3929. https://doi.org/10.2527/jas.2012-5823

Thompson, J. S., Hudson, C. D., Huxley, J. N., Kaler, J., Robinson, R. S., Woad, K. J., Bollard, N., Gibbons, J., & Green, M. J. (2022). A randomised controlled trial to evaluate the impact of indoor living space on dairy cow production, reproduction and behaviour. *Scientific Reports*, *12*(1), 3849. https://doi.org/10.1038/s41598-022-07826-9

Timsit, E., Assié, S., Quiniou, R., Seegers, H., & Bareille, N. (2011a). Early detection of bovine respiratory disease in young bulls using reticulo-rumen temperature boluses. *The Veterinary Journal*, *190*(1), 136–142. https://doi.org/10.1016/j.tvjl.2010.09.012

Timsit, E., Assié, S., Quiniou, R., Seegers, H., & Bareille, N. (2011b). Early detection of bovine respiratory disease in young bulls using reticulo-rumen temperature boluses. *The Veterinary Journal*, *190*(1), 136–142. https://doi.org/10.1016/J.TVJL.2010.09.012

Timsit, E., Bareille, N., Seegers, H., Lehebel, A., & Assié, S. (2011). Visually undetected fever episodes in newly received beef bulls at a fattening operation: Occurrence, duration, and impact on performance1. *Journal of Animal Science*, *89*(12), 4272–4280. https://doi.org/10.2527/jas.2011-3892

Timsit, E., Dendukuri, N., Schiller, I., & Buczinski, S. (2016). Diagnostic accuracy of clinical illness for bovine respiratory disease (BRD) diagnosis in beef cattle placed in feedlots: A systematic literature review and hierarchical Bayesian latent-class meta-analysis. *Preventive Veterinary Medicine*, *135*, 67–73. https://doi.org/10.1016/J.PREVETMED.2016.11.006

Toaff-Rosenstein, R. L., Gershwin, L. J., Zanella, A. J., & Tucker, C. B. (2016). The sickness response in steers with induced bovine respiratory disease before and after treatment with a non-steroidal anti-inflammatory drug. *Applied Animal Behaviour Science*, *181*, 49–62. https://doi.org/10.1016/J.APPLANIM.2016.05.016

Tolkamp, B. J., Allcroft, D. J., Austin, E. J., Nielsen, B. L., & Kyriazakis, I. (1998). Satiety Splits Feeding Behaviour into Bouts. *Journal of Theoretical Biology*, *194*(2). https://doi.org/10.1006/jtbi.1998.0759

Tolkamp, B. J., & Kyriazakis, I. (1999). To split behaviour into bouts, log-transform the intervals. *Animal Behaviour*, *57*(4), 807–817. https://doi.org/10.1006/ANBE.1998.1022

Trénel, P., Jensen, M. B., Decker, E. L., & Skjøth, F. (2009). Technical note: Quantifying and characterizing behavior in dairy calves using the IceTag automatic recording device. *Journal of Dairy Science*, *92*(7), 3397–3401. https://doi.org/10.3168/JDS.2009-2040

Types of Sensor | PTS Technologies Pte Ltd. (n.d.). Retrieved August 22, 2022, from http://ptstechnologies.com/types-of-sensor/

Tzutalin. (2015). *LabelImg*. Git code.

van Dixhoorn, I. D. E., de Mol, R. M., van der Werf, J. T. N., van Mourik, S., & van Reenen, C. G. (2018). Indicators of resilience during the transition period in dairy cows: A case study. *Journal of Dairy Science*, *101*(11), 10271–10282. https://doi.org/10.3168/JDS.2018-14779

Van Reenen, C. G., O'Connell, N. E., Van der Werf, J. T. N., Korte, S. M., Hopster, H., Jones, R. B., & Blokhuis, H. J. (2005). Responses of calves to acute stress: Individual consistency and relations between behavioral and physiological measures. *Physiology & Behavior*, *85*(5), 557–570. https://doi.org/10.1016/j.physbeh.2005.06.015

Vázquez Diosdado, J. A., Barker, Z. E., Hodges, H. R., Amory, J. R., Croft, D. P., Bell, N. J., & Codling, E. A. (2015). Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry*, *3*(1), 15. https://doi.org/10.1186/s40317-015-0045-8

Vázquez-Diosdado, J. A., Miguel-Pacheco, G. G., Plant, B., Dottorini, T., Green, M., & Kaler, J. (2019). Developing and evaluating threshold-based algorithms to detect drinking behavior in dairy cows using reticulorumen temperature. *Journal of Dairy Science*, *102*(11), 10471–10482. https://doi.org/10.3168/jds.2019-16442

Vázquez-Diosdado, J. A., Paul, V., Ellis, K. A., Coates, D., Loomba, R., & Kaler, J. (2019). A Combined Offline and Online Algorithm for Real-Time and Long-Term Classification of Sheep Behaviour: Novel Approach for Precision Livestock Farming. *Sensors*, *19*(14), 3201. https://doi.org/10.3390/s19143201

Vijaya, M. S. (2012). *Piezoelectric materials and devices: applications in engineering and medical sciences*. CRC press.

Von Keyserlingk, M. A. G., Brusius, L., & Weary, D. M. (2004). Competition for Teats and Feeding Behavior by Group-Housed Dairy Calves. *Journal* *of Dairy Science*, *87*(12), 4190–4194. https://doi.org/10.3168/JDS.S0022-0302(04)73563-8

Voss, B., Laue, Hans.-Joachim., Hoedemaker, M., & Wiedemann, S. (2016). Field-trial evaluation of an automatic temperature measurement device placed in the reticulo-rumen of pre-weaned male calves. *Livestock Science*, *189*, 78–81. https://doi.org/10.1016/J.LIVSCI.2016.05.005

Walton, E., Casey, C., Mitsch, J., Vázquez-Diosdado, J. A., Yan, J., Dottorini, T., Ellis, K. A., Winterlich, A., & Kaler, J. (2018). Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour. *Royal Society Open Science*, *5*(2), 171442. https://doi.org/10.1098/rsos.171442

Wang, M., Larsen, M. L. V., Liu, D., Winters, J. F. M., Rault, J.-L., & Norton, T. (2022). Towards re-identification for long-term tracking of group housed pigs. *Biosystems Engineering*, *222*, 71–81. https://doi.org/10.1016/j.biosystemseng.2022.07.017

Weary, D. M., Huzzey, J. M., & Von Keyserlingk, M. A. G. (2009). Board-invited Review: Using behavior to predict and identify ill health in animals. *Journal of Animal Science*, *87*(2), 770–777. https://doi.org/10.2527/jas.2008-1297

Webb, L. E., Engel, B., Berends, H., van Reenen, C. G., Gerrits, W. J. J., de Boer, I. J. M., & Bokkers, E. A. M. (2014). What do calves choose to eat and how do preferences affect behaviour? *Applied Animal Behaviour Science*, *161*, 7–19. https://doi.org/10.1016/j.applanim.2014.09.016

Webster, A. J. F. (2005). *Animal welfare - Limping towards Eden*. Blackwell Publishing.

Wechsler, B. (1995). Coping and coping strategies: a behavioural view. *Applied Animal Behaviour Science*, *43*(2), 123–134. https://doi.org/10.1016/0168-1591(95)00557-9

Westneat, D. F., Wright, J., & Dingemanse, N. J. (2015). The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, *90*(3), 729–743. https://doi.org/10.1111/brv.12131

Whalin, L., Neave, H. W., Føske Johnsen, J., Mejdell, C. M., & Ellingsen-Dalskau, K. (2022). The influence of personality and weaning method on early feeding behavior and growth of Norwegian Red calves. *Journal of Dairy Science*, *105*(2), 1369–1386. https://doi.org/10.3168/jds.2021-20871

Whishaw, I. Q., Dringenberg, H. C., & Comery, T. A. (1992). Rats (Rattus norvegicus) modulate eating speed and vigilance to optimize food consumption: effects of cover, circadian rhythm, food deprivation, and individual differences. *Journal of Comparative Psychology*, *106*(4), 411–419. https://doi.org/10.1037/0735-7036.106.4.411

White, B. J., Anderson, D. E., Renter, D. G., Larson, R. L., Mosier, D. A., Kelly, L. L., Theurer, M. E., Robért, B. D., & Walz, M. L. (2012). Clinical, behavioral, and pulmonary changes in calves following inoculation with Mycoplasma bovis. *American Journal of Veterinary Research*, *73*(4), 490–497. https://doi.org/10.2460/ajvr.73.4.490

White, B. J., Coetzee, J. F., Renter, D. G., Babcock, A. H., Thomson, D. U., & Andresen, D. (2008). Evaluation of two-dimensional accelerometers to monitor behavior of beef calves after castration. *American Journal of*

Veterinary Research, *69*(8), 1005–1012. https://doi.org/10.2460/ajvr.69.8.1005

White, B. J., & Renter, D. G. (2009). Bayesian Estimation of the Performance of Using Clinical Observations and Harvest Lung Lesions for Diagnosing Bovine Respiratory Disease in Post-weaned Beef Calves. *Journal of Veterinary Diagnostic Investigation*, *21*(4), 446–453. https://doi.org/10.1177/104063870902100405

Wingfield, J. C. (2003). Control of behavioural strategies for capricious environments. *Animal Behaviour*, *66*(5), 807–816. https://doi.org/10.1006/ANBE.2003.2298

Wirthgen, T., Zipser, S., Franze, U., Geidel, S., & Lempe, G. (2011). Automatic infrared based temperature measuring system for health monitoring in veterinary applications. *2011 IEEE SENSORS Proceedings*, 1800–1803. https://doi.org/10.1109/ICSENS.2011.6127107

Withers, F. W. (1952). Mortality Rates and Disease Incidence in Calves in Relation to Feeding, Management and Other Environmental Factors. *British Veterinary Journal*, *108*(10), 382–405. https://doi.org/10.1016/S0007-1935(17)51348-2

Wittum, T. E., Woollen, N. E., Perino, L. J., & Littledike, E. T. (1996). Relationships among treatment for respiratory tract disease, pulmonary lesions evident at slaughter, and rate of weight gain in feedlot cattle. *Journal of the American Veterinary Medical Association*, *209*(4), 814–818. http://www.ncbi.nlm.nih.gov/pubmed/8756886

10 Appendices

10.1 Appendix 1. Meal based criterion

10.1.1 Method

The meal criterion corresponds to the maximum time interval between two visits to the feeder by the same calf for those visits to be grouped into the same meal. Thus, those visits that are clustered in time are counted as the same meal. To determine the meal criterion, we calculated the interval in seconds between consecutive visits to the feeder for each calf. We plotted its log10-transformed distribution.

Three log10 distributions with intersections of 10^2 (approximately 100 seconds) and 10^3.2 (approximately 1600 seconds) can be seen in the Figure 10.1. Where the same calf had visits to the feeder that were separated by a duration equal to or less than 100 seconds these visits were considered as the same meal. Visits that were separated by more than 100 seconds were considered as separate meals.

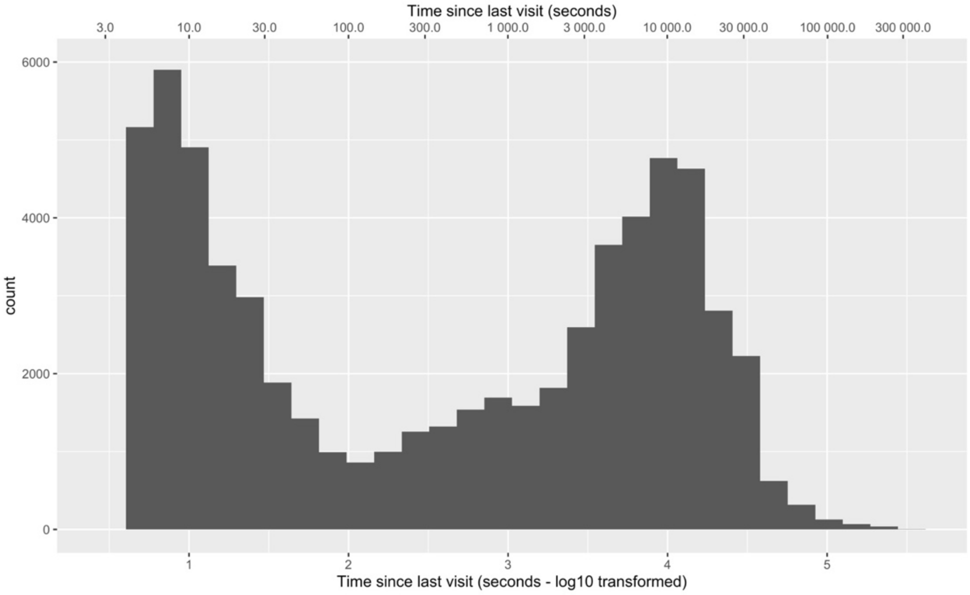


Figure 10.1. Log 10 transformed time interval between consecutive visits

10.2 Appendix 2. Data processing and code for measuring weight using automatically collected data

10.2.1 Weight data pre-processing

Weight data was collected by a partial weight scale that was situated in front of the milk feeding station. Each time the calf visits the feeder a weight is recorded by the computerised feeder. These data were downloaded and appended alongside the feeding data.

After filtering null values, weight values were visually inspected which revealed outlier observations that occurred alongside observations that were consistent with an expected daily liveweight gain of between 0.7 – 1.2kg (see Figure 10.2) Outlier observations were assumed to occur due to misplacement of the calves' front hooves on the partial scale (e.g., the calf places one hoof rather than both hooves on the weight scale) resulting in a drop in weight when compared with previous recordings. Erroneous measurements greater than the calf's true bodyweight were assumed less likely to occur since they would require the calf to assert downward pressure on the partial weight scale.

To process the weight data, we firstly fitted a quantile regression model (tau = 0.9) of weight on day number for each individual calf. Quantile regression allows us to assign greater influence to observations in the quantile of interest (here the top 10%). To assess the deviation of each observation, we calculated the residual mean and standard deviation. We excluded those observations where the residual estimate was less than residual mean or where its absolute value was greater than the residual standard deviation. Secondly, for each calf we ran a linear regression of weight on day number and excluded those observation where the cook's standard deviation exceeded 0.2 or where the absolute value of the residual exceeded the standard deviation for all residuals. Finally, using this filtered data, we fitted a linear regression of bodyweight on day number for each calf. Weight at the start of the group housing period was estimated by extracting the intercept from the linear model. Weight at 70 days old was calculated from the model prediction for the corresponding day number.

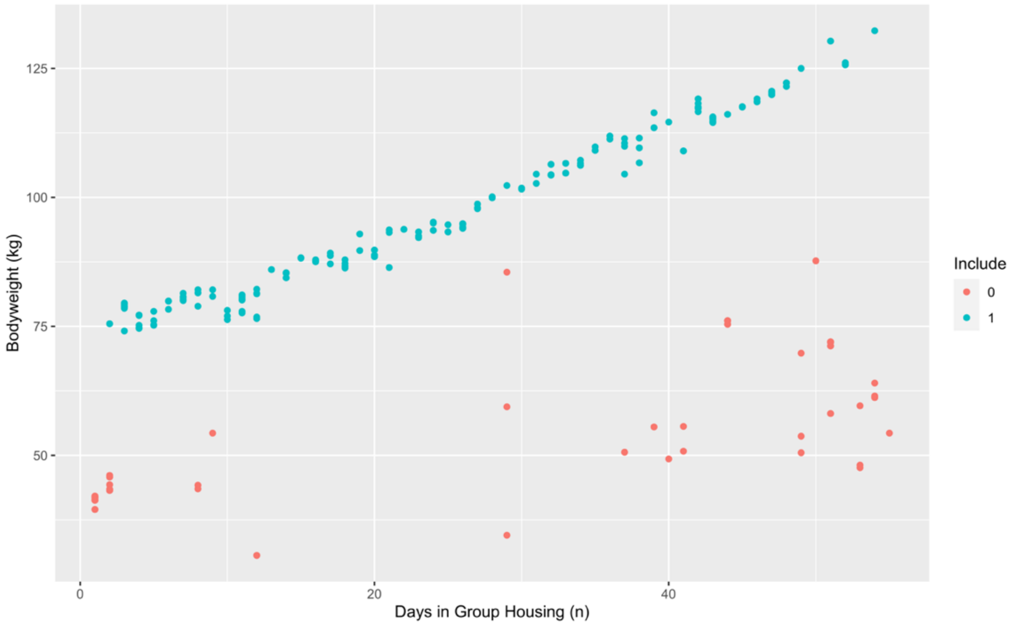


Figure 10.2. Example of raw data collected by the partial weight scale for one calf over the group housing period. The red observations correspond to those that were excluded after the pre-processing whereas the blue observations correspond to those that were included.

10.2.2 Example code

```
# Quantile regression and filtering

weights_filtered <- data %>%
  filter(housing.weight > 0) %>%
  nest(-Calf_ID) %>%
  mutate(fit = map(data, ~ rq(housing.weight ~ day_number
                             ,tau = 0.9, data = .)),
         results = map(fit, augment)) %>%
  unnest(results) %>%
  dplyr::select(Calf_ID, day_number, housing.weight, .resid, .tau) %>%
  mutate(residual_cut_off = sd(.resid)) %>%
  mutate(include = ifelse((abs(.resid) < residual_cut_off | .resid > mean(.resid)), 1, 0))

# Linear regression

weights_w_residuals_all <- weights_filtered %>%
  filter(include == 1) %>%
  nest(-Calf_ID) %>%
  mutate(fit = map(data, ~ lm(median.weight ~ day_number
                             , data = .)),
         results = map(fit, augment)) %>%
  unnest(results) %>%
```

```
  dplyr::select(Calf_ID, day_number, median.weight, .cooksd, .resid) %>%
  mutate(sd.residual_cut_off = 2*sd(.resid),
         cooks_cut_off = 0.2) %>%
  mutate(cooks_include = ifelse((.cooksd  < cooks_cut_off), 1, 0),
         resid_include = ifelse((.resid  > (0-sd.residual_cut_off) & .resid  < sd.residual_cut_off),
1, 0)) %>%
  mutate(include = ifelse((cooks_include == 1 & resid_include == 1), 1, 0))
```

10.3 Appendix 3. Example code for processing feeding data

10.3.1 Example code feed data processing

```
# Load required libraries

library(tidyverse)
library(purr)
library(lubridate)

data <- read.csv("feeding_data.csv") %>%
  mutate(Calf_ID = as.factor(Calf_ID),
         start.visit = ymd_hms(start.visit),
         end.visit = ymd_hms(end.visit),
         next.entitlement.at = ymd_hms(next.entitlement.at),
         age_move_date = as.double(age_move_date),
         housing.weight = as.double(housing.weight),
         feed.speed.ml.min = as.double(feed.speed.ml.min),
         Cohort = as.factor(Cohort),
         DoB = ymd(DoB),
         move_date = ymd(move_date),
         date = ymd(date),
         day_number = as.double(day_number),
         age = as.double(age))

glimpse(data)

table(data$date)

#create time since last visit for each calf

visit_interval <- data %>%
  distinct() %>%
  dplyr::group_by(Calf_ID) %>%
  dplyr::arrange(as.numeric(Calf_ID), start.visit) %>%
  mutate(diff = ymd_hms(start.visit) - ymd_hms(lag(end.visit)),
         time_since_last_visit = as.numeric(diff, units = 'secs'))

#plot time since last visit (log 10)

visit_interval %>%
  ggplot(aes(x = log10(time_since_last_visit))) +
  geom_histogram() +
  ylim(0, 20000)

#assign the same meal number to visits seperated by less than 10^2 (100s)

meal_number <- visit_interval %>%
  filter(as.numeric(end.visit - start.visit, units = 'secs') > 0) %>%
```

```
  arrange(as.numeric(Calf_ID), start.visit) %>%
  dplyr::group_by(as.numeric(Calf_ID, date)) %>%
  mutate(meal_number = cumsum(ifelse(difftime(start.visit,
                                              shift(end.visit, fill = end.visit[1]),
                                              units = "secs") >= 100
                                     ,1, 0)) + 1)


#For each meal, was the calf entitled to a feed? - entit varaible says which meal numbers are with entitlement

entit_w_wo <- function(x){
  entitlement.act. <- x
  if(max(entitlement.act.) > 0){x <- 1
  } else {x <- 0
  }
}

#and map onto nested data

myvars <- c("Calf_ID", "entitlement.act.", "meal_number")
newdata <- meal_number[myvars]

nested <- newdata %>%
  group_by(Calf_ID, meal_number) %>%
  nest()

test2 <- nested %>% mutate(entit = map(data, entit_w_wo))
entit_nested <- test2 %>% dplyr::select(-data)
entit <- unnest(entit_nested)


#For each meal, did the calf drink any milk - feed variable says which meal numbers are with a milk feed

meal_number <- meal_number %>%
  arrange(as.numeric(Calf_ID), start.visit) %>%
  mutate(cons_previous = lag(total.MP.g.))

cons_per_meal <- meal_number %>%
  dplyr::group_by(Calf_ID, meal_number) %>%
  dplyr::summarise(meal_cons = max(total.MP.g.) - min(cons_previous)) %>%
  mutate(consumed = ifelse(meal_cons > 0, 1, 0))

cons_and_entit <- full_join(entit, cons_per_meal)

table(cons_and_entit$entit, cons_and_entit$consumed)

meal_numbers_w_cons_and_entitlement <- full_join(meal_number, cons_and_entit) %>%
  droplevels()

MWOE_daily <- meal_numbers_w_cons_and_entitlement %>%
  filter(entit == 0) %>%
  dplyr::group_by(Calf_ID, day_number, meal_number) %>%
  dplyr::summarise(Meal_number = max(meal_number)) %>%
  mutate(MWOE = ifelse(Meal_number > 0, 1, 0)) %>%
  dplyr::group_by(Calf_ID, day_number) %>%
  dplyr::summarise(MWOE = sum(MWOE))

MWE_daily <- meal_numbers_w_cons_and_entitlement %>%
```

```
  filter(entit == 1) %>%
  dplyr::group_by(Calf_ID, day_number, meal_number) %>%
  dplyr::summarise(Meal_number = max(meal_number)) %>%
  mutate(MWE = ifelse(Meal_number > 0, 1, 0)) %>%
  dplyr::group_by(Calf_ID, day_number) %>%
  dplyr::summarise(MWE = sum(MWE))


meals_daily <- full_join(MWOE_daily, MWE_daily) %>%


# feeding rate - only use rows where calf is newly entitled

FR_daily <- meal_number %>%
  filter(feed.speed.ml.min > 0 & entitlement.act. > 0) %>%
  dplyr::group_by(Calf_ID, day_number) %>%
  dplyr::summarise(FR = mean(feed.speed.ml.min))


#meal_size_daily

meal_size_daily <- meal_number %>%
  dplyr::group_by(Calf_ID, day_number, meal_number) %>%
  dplyr::summarise(meal_cons = max(total.MP.g.) - min(cons_previous)) %>%
  mutate(consumed = ifelse(meal_cons > 0, 1, 0)) %>%
  filter(consumed == 1) %>%
  dplyr::group_by(Calf_ID, day_number) %>%
  dplyr::summarise(MS = mean(meal_cons)/130)

##Time at feeder

DFT_daily <- meal_number %>%
  mutate(visit_duration = ymd_hms(end.visit) - (start.visit),
         visit_duration_secs = as.numeric(visit_duration, units = 'secs')) %>%
  dplyr::group_by(Calf_ID, day_number) %>%
  dplyr::summarise(DFT = sum(visit_duration_secs))


# daily feed intake

DFI_daily <- meal_number %>%
  dplyr::group_by(Calf_ID, date, day_number) %>%
  dplyr::summarise(DFI = (max(total.MP.g.) - min(cons_previous))/130) %>%
  filter(DFI > 0)

data_daily <- full_join(meals_daily, FR_daily) %>%
  full_join(meal_size_daily) %>%
  full_join(DFT_daily) %>%
  full_join(DFI_daily) %>%
  full_join(FR_daily)
```

10.4 Appendix 4. Example R Code for double hierarchical models and plots used in chapter 2 and chapter 3

R code has been adapted adapted from the code made available by Hertel et al. (2021, 2022)

10.4.1 Example code for model, repeatability, coefficient of variation in predictability estimates and plots chapter 2

```
# Load required libraries

library(tidyverse)
library(nadiv)
library(brms)
library(rptR)
library(ggpubr)
library(broom)
library(ggpubr)
library(tidyr)
library(coda)
library(RVAideMemoire)
library(tidybayes)
library(parallel)
library(stringr)
library(bayestestR)

# Key for variables:

    # FR - Feeding rate

    # MTOT - Total Meals (log transformed)

    # MS - Meal size (Order norm transformed)


#### Multivariate DHGLM - Feeding Rate and Total Meals


CompDHGLM_calves <- brm(bf_MTOT + bf_FR + bf_MS + set_rescor(FALSE),
  data = data,
  iter = 12000, warmup = 4000, thin = 4,
  chains = 4, cores = detectCores(), seed = 12345
)


CompDHGLM <- CompDHGLM_calves

# Check mixing of chains

plot(CompDHGLM_calves)

summary(CompDHGLM_calves)

# Model validation - Feeding Rate

pp_check(CompDHGLM, resp = "scaleFR") +
  theme_bw(base_size = 20)

pp_check(CompDHGLM_calves,
  resp = "scaleFR",
  nsamples = 1e3,
  type = "stat_2d"
```

```
) +
  theme_bw(base_size = 20)

# Model validation - MTOT

pp_check(CompDHGLM, resp = "scaleMTOT")

pp_check(CompDHGLM,
  resp = "scaleMTOT",
  nsamples = 1e3,
  type = "stat_2d"
) +
  theme_bw(base_size = 20)


# See model results

summary(CompDHGLM_calves)


#Repeatability calculations (rpt)

#rpt FR

var.CalfID.FR <-
posterior_samples(CompDHGLM_calves)$"sd_Calf_ID__scaleFR_Intercept"^2
var.res.FR <-
exp(posterior_samples(CompDHGLM_calves)$"b_sigma_scaleFR_Intercept")^2

Rep.FR <- (var.CalfID.FR) /
  (var.CalfID.FR + var.res.FR)

mean(Rep.FR);HPDinterval(as.mcmc(Rep.FR),0.95)


#rpt MTOT

var.CalfID.MTOT <-
posterior_samples(CompDHGLM_calves)$"sd_Calf_ID__scaleMTOT_Intercept"^2
var.res.MTOT <-
exp(posterior_samples(CompDHGLM_calves)$"b_sigma_scaleMTOT_Intercept")^2

Rep.MTOT <- (var.CalfID.MTOT) /
  (var.CalfID.MTOT + var.res.MTOT)

mean(Rep.MTOT);HPDinterval(as.mcmc(Rep.MTOT),0.95)

#rpt MS

var.CalfID.MS <-
posterior_samples(CompDHGLM_calves)$"sd_Calf_ID__scaleMS_Intercept"^2
var.res.MS <-
exp(posterior_samples(CompDHGLM_calves)$"b_sigma_scaleMS_Intercept")^2

Rep.MS <- (var.CalfID.MS) /
  (var.CalfID.MS + var.res.MS)

mean(Rep.MS);HPDinterval(as.mcmc(Rep.MS),0.95)
```

```
#Coefficient of variation in predictability (CVP)

#CVP FR

log.norm.res.FR <-
exp(posterior_samples(CompDHGLM_calves)$"sd_Calf_ID__sigma_scaleFR_Intercept"^2)
CVP.FR <- sqrt(log.norm.res.FR - 1)

mean(CVP.FR);HPDinterval(as.mcmc(CVP.FR),0.95)


#CVP Mtot

log.norm.res.MTOT <-
exp(posterior_samples(CompDHGLM_calves)$"sd_Calf_ID__sigma_scaleMTOT_Intercept"^
2)
CVP.MTOT <- sqrt(log.norm.res.MTOT - 1)

mean(CVP.MTOT);HPDinterval(as.mcmc(CVP.MTOT),0.95)


#CVP MS

log.norm.res.MS <-
exp(posterior_samples(CompDHGLM_calves)$"sd_Calf_ID__sigma_scaleMS_Intercept"^2)
CVP.MS <- sqrt(log.norm.res.MS - 1)

mean(CVP.MS);HPDinterval(as.mcmc(CVP.MS),0.95)



# Code for Figure 2.1 - Individual variation in behavioural types (BT)

# Extract estimates

BT_FR_plot_data <- posterior_samples(CompDHGLM, pars = "^r_Calf_ID__scaleFR") %>%
  tidyr::gather(
    Calf_ID, value,
    "r_Calf_ID__scaleFR[1,Intercept]":
      "r_Calf_ID__scaleFR[48,Intercept]"
  ) %>%
  select(Calf_ID, value) %>%
  mutate(Calf_ID = substr(Calf_ID, start = 20, stop = 21)) %>%
  mutate(Calf_ID = as.factor(as.numeric(gsub(",", "", Calf_ID)))) %>%
  group_by(Calf_ID) %>%
  mutate(Mean_FR = mean(value)) %>%
  left_join(Cohorts)


BT_MTOT_plot_data <- posterior_samples(CompDHGLM, pars =
"^r_Calf_ID__scaleMTOT") %>%
  tidyr::gather(
    Calf_ID, value,
    "r_Calf_ID__scaleMTOT[1,Intercept]":
      "r_Calf_ID__scaleMTOT[48,Intercept]"
  ) %>%
  select(Calf_ID, value) %>%
  mutate(Calf_ID = substr(Calf_ID, start = 22, stop = 23)) %>%
  mutate(Calf_ID = as.factor(as.numeric(gsub(",", "", Calf_ID)))) %>%
  group_by(Calf_ID) %>%
```

```
  mutate(Mean_MTOT = mean(value)) %>%
  left_join(Cohorts)

BT_MS_plot_data <- posterior_samples(CompDHGLM, pars = "^r_Calf_ID__scaleMS")
%>%
  tidyr::gather(
    Calf_ID, value,
    "r_Calf_ID__scaleMS[1,Intercept]":
    "r_Calf_ID__scaleMS[48,Intercept]"
  ) %>%
  select(Calf_ID, value) %>%
  mutate(Calf_ID = substr(Calf_ID, start = 20, stop = 21)) %>%
  mutate(Calf_ID = as.factor(as.numeric(gsub(",", "", Calf_ID)))) %>%
  group_by(Calf_ID) %>%
  mutate(Mean_MS = mean(value)) %>%
  left_join(Cohorts)


# Plot estimates

BT_FR_plot <- ggplot() +
  ggridges::geom_density_ridges(
    data = BT_FR_plot_data,
    aes(
      x = value,
      y = reorder(as.double(Calf_ID), Mean_FR),
      height = ..density.., fill = Cohort, ,
      scale = 3
    ), alpha = 0.6, rel_min_height = 0.005
  ) +
  labs(
    y = "Calf ID",(size = 12),
    x = "Behavioral Type Feeding Rate", (size = 12),
    fill = "Cohort"
  ) +
  theme_classic(base_size = 12) +
  xlim(-2.4, 2.4) +
  annotate("text",
    x = 1.2, y = 3,
    label = expression(paste(italic("Rpt")[] * " = 0.50")),
    size = 4
  )


BT_MTOT_plot <- ggplot() +
  ggridges::geom_density_ridges(
    data = BT_MTOT_plot_data,
    aes(
      x = value,
      y = reorder(as.double(Calf_ID), Mean_MTOT),
      height = ..density.., fill = Cohort, ,
      scale = 3
    ), alpha = 0.6, rel_min_height = 0.005
  ) +
  labs(
    y = "Calf ID",(size = 12),
    x = "Behavioral Type Total Meals", (size = 12),
    fill = "Cohort"
  ) +
  theme_classic(base_size = 12) +
```

```
  xlim(-2.4, 2.4) +
  annotate("text",
           x = 1.2, y = 3,
           label = expression(paste(italic("Rpt")[] * " = 0.42")),
           size = 4
  )


BT_MS_plot <- ggplot() +
  ggridges::geom_density_ridges(
    data = BT_MS_plot_data,
    aes(
      x = value,
      y = reorder(as.double(Calf_ID), Mean_MS),
      height = ..density.., fill = Cohort, ,
      scale = 3
    ), alpha = 0.6, rel_min_height = 0.005
  ) +
  labs(
    y = "Calf ID",(size = 12),
    x = "Behavioral Type Meal Size",size = 12,
    fill = "Cohort"
  ) +
  theme_classic(base_size = 12) +
  xlim(-2.4, 2.4) +
  annotate("text",
    x = 1.2, y = 3,
    label = expression(paste(italic("Rpt")[] * " = 0.03")),
    size = 4
  )


figure_BT <- ggarrange(BT_FR_plot, BT_MS_plot,
  common.legend = TRUE, legend = "right",
  ncol = 2, nrow = 1
)


# Code for Figure 2.3 Correlation behavioural types Feeding Rate / Total Meals

# Extract estimates

BT_FR <- posterior_samples(CompDHGLM_calves, pars = "^r_Calf_ID__scaleFR") %>%
  tidyr::gather(
    Calf_ID, value,
    "r_Calf_ID__scaleFR[1,Intercept]":
    "r_Calf_ID__scaleFR[48,Intercept]"
  ) %>%
  dplyr::select(Calf_ID, value) %>%
  dplyr::mutate(Calf_ID = substr(Calf_ID, start = 20, stop = 21)) %>%
  mutate(Calf_ID = as.factor(as.numeric(gsub(",", "", Calf_ID)))) %>%
  dplyr::group_by(Calf_ID) %>%
  dplyr::mutate(
    Mean_FR = mean(value),
    Up_FR = mean(value) + 1.96 * sd(value),
    Lo_FR = mean(value) - 1.96 * sd(value)
  ) %>%
  dplyr::select(-value) %>%
  filter(!duplicated(Calf_ID))
```

```
BT_MTOT <- posterior_samples(CompDHGLM_calves, pars = "^r_Calf_ID__scaleMTOT")
%>%
  tidyr::gather(
    Calf_ID, value,
    "r_Calf_ID__scaleMTOT[1,Intercept]":
    "r_Calf_ID__scaleMTOT[48,Intercept]"
  ) %>%
  dplyr::select(Calf_ID, value) %>%
  dplyr::mutate(Calf_ID = substr(Calf_ID, start = 22, stop = 23)) %>%
  mutate(Calf_ID = as.factor(as.numeric(gsub(",", "", Calf_ID)))) %>%
  dplyr::group_by(Calf_ID) %>%
  dplyr::mutate(
    Mean_MTOT = mean(value),
    Up_MTOT = mean(value) + 1.96 * sd(value),
    Lo_MTOT = mean(value) - 1.96 * sd(value)
  ) %>%
  dplyr::select(-value) %>%
  filter(!duplicated(Calf_ID))

outputs_BTs_calves <- full_join(BT_MTOT, BT_FR)

glimpse(outputs_BTs_calves)


# Estimate correlations

cov.Trav <-

posterior_samples(CompDHGLM_calves)$"cor_Calf_ID__scaleMTOT_Intercept__scaleFR_I
ntercept" *
    sqrt((posterior_samples(CompDHGLM_calves)$"sd_Calf_ID__scaleMTOT_Intercept")^2)
*
    sqrt((posterior_samples(CompDHGLM_calves)$"sd_Calf_ID__scaleFR_Intercept")^2)

var.Trav <-
(posterior_samples(CompDHGLM_calves)$"sd_Calf_ID__scaleMTOT_Intercept"^2)

LT_slope_Trav <- cov.Trav / var.Trav

mean(LT_slope_Trav)


# Create plot

BS_MTOT_FR_calves <- ggplot() +
  geom_segment(
    data = outputs_BTs_calves[!duplicated(outputs_BTs_calves$Calf_ID), ],
    aes(
      x = Lo_FR,
      xend = Up_FR,
      y = Mean_MTOT,
      yend = Mean_MTOT
    ),
    color = "#F8766D", alpha = 0.2
  ) +
  geom_segment(
    data = outputs_BTs_calves[!duplicated(outputs_BTs_calves$Calf_ID), ],
    aes(
      x = Mean_FR,
```

```
    xend = Mean_FR,
    y = Lo_MTOT,
    yend = Up_MTOT
  ),
  color = "#F8766D", alpha = 0.2
) +
geom_point(
  data = outputs_BTs_calves[!duplicated(outputs_BTs_calves$Calf_ID), ],
  aes(x = Mean_FR, y = Mean_MTOT), color = "#F8766D"
) +
geom_segment(aes(
  x = -2,
  xend = 2,
  y = 0 + mean(LT_slope_Trav) * -2,
  yend = 0 + mean(LT_slope_Trav) * 2
),
color = "#F8766D", size = 1, alpha = 0.8
) +
ylab("Behavioural Type Total Meals") +
xlab("Behavioural Type Feeding Rate") +
guides(
  fill = guide_legend(
    order = 2,
    keywidth = 0.08,
    keyheight = 0.1,
    default.unit = "inch"
  ),
  color = guide_legend(
    order = 1,
    keywidth = 0.08,
    keyheight = 0.1,
    default.unit = "inch"
  )
) +
annotate("text",
  x = 1.1, y = -0.8,
  label = expression(paste(italic("r")[] * " = 0.29 [0.00, 0.54]")),
  size = 5
) +
xlim(-2.2, 2.2)
```

10.4.2 Example R code for model and repeatability estimates chapter 3

```
# Model specification

bf_FR_pairs <- bf(scale(FR_pairs) ~ scale(birth_weight) + scale(age_pair_housing) +
(1|Cohort) + (1|b|Calf_ID),
                 family = gaussian
)

bf_MTOT_pairs <- bf(scale(Mtot_pairs) ~ scale(birth_weight) + scale(age_pair_housing) +
(1|Cohort) + (1|b|Calf_ID),
                   family = gaussian
)
```

```
bf_FR_group <- bf(scale(FR_group) ~ scale(weight_enter_pen_model) +
scale(day_number_group_housing)*scale(age_move_date) + (1|Cohort) + (1|b|Calf_ID),
                 family = gaussian
)

bf_MTOT_group <- bf(scale(Mtot_group) ~ scale(weight_enter_pen_model) +
scale(day_number_group_housing)*scale(age_move_date) + (1|Cohort) + (1|b|Calf_ID),
                   family = gaussian
)


CompDHGLM <- brm(bf_MTOT_group + bf_FR_group + bf_FR_pairs + bf_MTOT_pairs +
set_rescor(FALSE),
                      data = data,
                      iter = 12000, warmup = 4000, thin = 4,
                      chains = 4, cores = 4, seed = 12345
)

summary(CompDHGLM)

plot(CompDHGLM)

# repeatability estimates (rpt)

# rpt FR

var.CalfID.FR <-
posterior_samples(CompDHGLM)$"sd_Calf_ID__scaleFRpairs_Intercept"^2
var.res.FR <- (posterior_samples(CompDHGLM)$"sigma_scaleFRpairs")^2

Rep.FR <- (var.CalfID.FR) /
  (var.CalfID.FR + var.res.FR)

mean(Rep.FR);HPDinterval(as.mcmc(Rep.FR),0.95)


# rpt Mtot

var.CalfID.Mtot <-
posterior_samples(CompDHGLM)$"sd_Calf_ID__scaleMtotpairs_Intercept"^2
var.res.Mtot <- (posterior_samples(CompDHGLM)$"sigma_scaleMtotpairs")^2

Rep.Mtot <- (var.CalfID.Mtot) /
  (var.CalfID.Mtot + var.res.Mtot)

mean(Rep.Mtot);HPDinterval(as.mcmc(Rep.Mtot),0.95)

# rpt FR

var.CalfID.FR <-
posterior_samples(CompDHGLM)$"sd_Calf_ID__scaleFRgroup_Intercept"^2
var.res.FR <- (posterior_samples(CompDHGLM)$"sigma_scaleFRgroup")^2

Rep.FR <- (var.CalfID.FR) /
  (var.CalfID.FR + var.res.FR)

mean(Rep.FR);HPDinterval(as.mcmc(Rep.FR),0.95)


# rpt Mtot
```

```
var.CalfID.Mtot <-
posterior_samples(CompDHGLM)$"sd_Calf_ID__scaleMtotgroup_Intercept"^2
var.res.Mtot <- (posterior_samples(CompDHGLM)$"sigma_scaleMtotgroup")^2

Rep.Mtot <- (var.CalfID.Mtot) /
  (var.CalfID.Mtot + var.res.Mtot)

mean(Rep.Mtot);HPDinterval(as.mcmc(Rep.Mtot),0.95)
```

10.5 Appendix 5. Example code for machine learning algorithms

Example R code used in Chapters 4 and 5 to partition data into test and train, create, cross validate, and test random forest classifier.

Looped 100 times to obtain performance estimates.

This code was used as part of chapters 4 and 5 to obtain estimates of machine learning model performance.

```
library(caret)
library(tidyverse)

features <- data %>%
  mutate(label = as.factor(label))

table(features$label)

num_iter = 100;

accuracy<- matrix(ncol=1, nrow=num_iter)
macroPrecision<- matrix(ncol=1, nrow=num_iter)
macroRecall<- matrix(ncol=1, nrow=num_iter)
macroF1<-matrix(ncol=1, nrow=num_iter)
macroSpecificity<-matrix(ncol=1, nrow=num_iter)
set.seed(100)


for(i in 1:num_iter){
  # Create the training and test datasets
  print(i)
  # Step 1: Get row numbers for the training data

  trainRowNumbers <- createDataPartition(features$label, p=0.7, list=FALSE)

  # Step 2: Create the training dataset

  trainData <- features[trainRowNumbers,]


  # Step 3: Create the test dataset

  testData <- features[-trainRowNumbers,]
```

```
  y2 <- testData$label

  # Define the training control
  fitControl <- trainControl(
    method = 'CV',                  # k-fold cross validation
    number = 5,                     # number of folds
    savePredictions = 'final',      # saves predictions for optimal tuning parameter
    classProbs = T,                 # should class probabilities be returned
    summaryFunction=twoClassSummary  # results summary function
  )

  # Train the model using rf train data only

  model_rf = train(label ~ ., data=trainData, method='rf', tuneLength=5, trControl = fitControl,
metric = "ROC")

  y_pred_test =predict(model_rf, newdata= testData)

  cm <- confusionMatrix(reference = y2, data = y_pred_test,
mode='everything', positive = "sick")

 # obtain performance metrics

  accuracy[i,] = as.double(cm$byClass[11])
  macroRecall[i,] = as.double(cm$byClass[1])
  macroSpecificity[i,] = as.double(cm$byClass[2])
  macroPrecision[i,] = as.double(cm$byClass[5])
  macroF1[i,] = as.double(cm$byClass[7])

}

DF_performance <- data.frame(matrix(ncol = 5, nrow = num_iter))
x <-c("Accuracy","macroPrecision","macroRecall", "macroF1", "macroSpecificity")


# colnames(DF_daily) <- x

# DF_daily$Date <-Exclude_date_percentage$Date[1:num_days]
DF_performance$Accuracy<-accuracy
DF_performance$mmacroPrecision<-macroPrecision
DF_performance$mmacroRecall<-macroRecall
DF_performance$mmacroF1<-macroF1
DF_performance$mmacroSpecificity<-macroSpecificity

A <- data.frame(accuracy,macroPrecision, macroRecall, macroF1, macroSpecificity)

mean(A$accuracy)
sd(A$accuracy)

mean(A$macroSpecificity)
sd(A$macroSpecificity)

mean(A$macroRecall)
sd(A$macroRecall)

mean(A$macroPrecision)
sd(A$macroPrecision)

mean(A$macroF1)
sd(A$macroF1)
```

10.6 Appendix 6. Example MATLAB code for chapter 6

Example code showing feature generation for different sampling frequencies

```
%%% Set the frequency for the feature creation

fqr = 4;

[c d] =size(DATA);

    num_points = win*fqr;
    num_win = floor(c/(num_points/2))+1;

 DATA.DiffAccMag = zeros(c,1);
 DATA.DiffGyrMag= zeros(c,1);
 DATA.DiffAccMag(2:end,1) = diff(DATA.AccMag);
 DATA.DiffGyrMag(2:end,1) = diff(DATA.GyrMag);

 tic
num_fea  = 50;
FeaMat = zeros(num_win,num_fea);

  i=1;
   while (i<num_win-2)
   %disp(i)
       if(i==1)
         e = 1;
          f = num_points;
       else
%  a = a+block_size*(j-1)+(num_points/2);
%  b = b+block_size*(j-1)+(num_points/2);

           e = e+(num_points/2);
           f = f+(num_points/2);

       end

      WinLen = f-e;

isempty_flagAcc = sum(isnan(DATA.AccMag(e:f,1)));
isempty_flagGyr = sum(isnan(DATA.GyrMag(e:f,1)));
isempty_flagAccDiff = sum(isnan(DATA.DiffAccMag(e:f,1)));
isempty_flagGyrDiff = sum(isnan(DATA.DiffGyrMag(e:f,1)));

if((isempty_flagAcc==0)&&(isempty_flagGyr==0)&&(isempty_flagAccDiff==0)&&(isempty
_flagGyrDiff==0))


%%% For Acceleration Magnitude

 FeaMat(i,1) = mean(DATA.AccMag(e:f,1));                  %%% Mean   %%
 FeaMat(i,2) = std(DATA.AccMag(e:f,1));       %%% Standard deviation %%
 FeaMat(i,3) = iqr(DATA.AccMag(e:f,1));        %%% Interquantile range %%
 FeaMat(i,4) = max(DATA.AccMag(e:f,1));          %%% Maximum %%
 FeaMat(i,5) = min(DATA.AccMag(e:f,1));          %%% Minimum %%
 FeaMat(i,6) = kurtosis(DATA.AccMag(e:f,1));  %%% Kurtosis %%
```

```
 PeriodAcc = periodogram(DATA.AccMag(e:f,1));
  FeaMat(i,7 )   = computed_zero_crossing(DATA.AccMag(e:f,1));
  FeaMat(i,8)    = entropy(PeriodAcc);   %%% Spectral entropy %%
  FeaMat(i,9)    = max(PeriodAcc)*fqr;   %%%  Dominant frequency %%
  FeaMat(i,10)   = sum(sqrt((DATA.AccMag(e:f,1)/fqr).^2));       %%%   Signal Area%%
  FeaMat(i,11)   =    sum(PeriodAcc*(fqr/num_points));   %%%  Spectral Area%%
%%% Spectral Area %%


  %%% For Gyro Magnitude

 FeaMat(i,12) = mean(DATA.GyrMag(e:f,1));                   %%% Mean   %%
 FeaMat(i,13) = std(DATA.GyrMag(e:f,1));       %%% Standard deviation %%
 FeaMat(i,14) = iqr(DATA.GyrMag(e:f,1));        %%% Interquantile range %%
 FeaMat(i,15) = max(DATA.GyrMag(e:f,1));          %%% Maximum %%
 FeaMat(i,16) = min(DATA.GyrMag(e:f,1));          %%% Minimum %%
 FeaMat(i,17) = kurtosis(DATA.GyrMag(e:f,1));  %%% Kurtosis %%


 PeriodGyr = periodogram(DATA.GyrMag(e:f,1));
  FeaMat(i,18 )   = computed_zero_crossing(DATA.GyrMag(e:f,1));                %%% Zero
crossing  %%
 %DATA_FEA.AccMagZeroCrossing(i,1) = zcd(DATA.AccMag(e:f,1));
  FeaMat(i,19)    = entropy(PeriodGyr);   %%% Spectral entropy %%
  FeaMat(i,20)    = max( PeriodGyr)*fqr;   %%%  Dominant frequency %%
  FeaMat(i,21)   = sum(sqrt((DATA.GyrMag(e:f,1)/fqr).^2));       %%%   Signal Area%%
  FeaMat(i,22)   =    sum(PeriodGyr*(fqr/num_points));   %%%  Spectral Area%%


  %%% For Acceleration Magnitude Difference


 FeaMat(i,23) = mean(DATA.DiffAccMag(e:f,1));                   %%% Mean   %%
 FeaMat(i,24) = std(DATA.DiffAccMag(e:f,1));       %%% Standard deviation %%
 FeaMat(i,25) = iqr(DATA.DiffAccMag(e:f,1));        %%% Interquantile range %%
 FeaMat(i,26) = max(DATA.DiffAccMag(e:f,1));          %%% Maximum %%
 FeaMat(i,27) = min(DATA.DiffAccMag(e:f,1));          %%% Minimum %%
 FeaMat(i,28) = kurtosis(DATA.DiffAccMag(e:f,1));  %%% Kurtosis %%

 PeriodAccDiff = periodogram(DATA.DiffAccMag(e:f,1));
  FeaMat(i,29 )   = computed_zero_crossing(DATA.DiffAccMag(e:f,1));                %%%
Zero crossing  %%
 %DATA_FEA.AccMagZeroCrossing(i,1) = zcd(DATA.AccMag(e:f,1));
  FeaMat(i,30)    = entropy(PeriodAccDiff);   %%% Spectral entropy %%
  FeaMat(i,31)    = max(PeriodAccDiff)*fqr;   %%%  Dominant frequency %%
  FeaMat(i,32)   = sum(sqrt((DATA.DiffAccMag(e:f,1)/fqr).^2));       %%%   Signal Area%%
  FeaMat(i,33)   =    sum(PeriodAccDiff*(fqr/num_points));   %%%  Spectral Area%%
%%% Spectral Area %%


   %%% For Acceleration Magnitude Difference

 FeaMat(i,34) = mean(DATA.DiffGyrMag(e:f,1));                   %%% Mean   %%
 FeaMat(i,35) = std(DATA.DiffGyrMag(e:f,1));       %%% Standard deviation %%
 FeaMat(i,36) = iqr(DATA.DiffGyrMag(e:f,1));        %%% Interquantile range %%
 FeaMat(i,37) = max(DATA.DiffGyrMag(e:f,1));          %%% Maximum %%
 FeaMat(i,38) = min(DATA.DiffGyrMag(e:f,1));          %%% Minimum %%
 FeaMat(i,39) = kurtosis(DATA.DiffGyrMag(e:f,1));  %%% Kurtosis %%
```

```
 PeriodGyrDiff =periodogram(DATA.DiffGyrMag(e:f,1));
  FeaMat(i,40 )   = computed_zero_crossing(DATA.DiffGyrMag(e:f,1));
  FeaMat(i,41)    = entropy(PeriodGyrDiff);   %%% Spectral entropy %%
  FeaMat(i,42)    = max(PeriodGyrDiff)*fqr;   %%%  Dominant frequency %%
  FeaMat(i,43)   = sum(sqrt((DATA.DiffGyrMag(e:f,1)/fqr).^2));       %%%   Signal Area%%
  FeaMat(i,44)   =   sum( PeriodGyrDiff*(fqr/num_points));   %%%  Spectral Area%%

 FeaMat(i,45)  = mode(DATA.Posture(e:f,1));
 FeaMat(i,46) = mode(DATA.Activity(e:f,1));
 FeaMat(i,47)  = mode(DATA.Behaviour(e:f,1));
 FeaMat(i,48)  = mode(DATA.Behaviour_2(e:f,1));

 FeaMat(i,49)  = mode(DATA.Activity_2(e:f,1));
 FeaMat(i,50)  = mode(DATA.Posture_2(e:f,1));

end

 clear PeriodAcc PeriodGyr PeriodAccDiff PeriodGyrDiff
 %disp(i)
 i= i+1;

   end

 toc
   varNames=
{'AccMagMean','AccMagStd','AccMagIqr','AccMagMax','AccMagMin','AccMagKurtosis','Ac
cMagZeroCrossing',...
        'AccMagSpectralEntropy','AccMagDominantFreq','AccMagSignalArea',
'AccMagSpectralSignalArea',...

'GyrMagMean','GyrMagStd','GyrMagIqr','GyrMagMax','GyrMagMin','GyrMagKurtosis','Gyr
MagZeroCrossing',...
        'GyrMagSpectralEntropy','GyrMagDominantFreq','GyrMagSignalArea',
'GyrMagSpectralSignalArea',...

'DiffAccMagMean','DiffAccMagStd','DiffAccMagIqr','DiffAccMagMax','DiffAccMagMin','Di
ffAccMagKurtosis','DiffAccMagZeroCrossing',...
        'DiffAccMagSpectralEntropy','DiffAccMagDominantFreq','DiffAccMagSignalArea',
'DiffAccMagSpectralSignalArea',...

'DiffGyrMagMean','DiffGyrMagStd','DiffGyrMagIqr','DiffGyrMagMax','DiffGyrMagMin','Dif
fGyrMagKurtosis','DiffGyrMagZeroCrossing',...
        'DiffGyrMagSpectralEntropy','DiffGyrMagDominantFreq','DiffGyrMagSignalArea',
'DiffGyrMagSpectralSignalArea','Posture', 'Activity','Behaviour', 'Behaviour2',
'Activity2','Posture2'};

   DATA_FEA  = array2table(FeaMat(:,1:num_fea), 'VariableNames',varNames);

end
```