Validation of tri-axial accelerometers and the impact of environmental enrichment on behaviour and welfare of domestic dogs (*Canis familiaris*)

A thesis presented
in partial fulfilment of the requirements
for the degree of

Master of Science
in
Animal Science

Massey University
Palmerston North, New Zealand

Cushla Redmond

2024

Synopsis

Monitoring and quantifying behaviour in dogs can provide valuable insight into their overall health and welfare. Behavioural observation in dogs can often be labour-intensive and time-consuming. However, recent advances in remote sensing technologies, such as accelerometery, offer an automated method for continuously measuring behaviour without an observer present. The first aim of this thesis was to investigate the use of remote sensing technology, particularly tri-axial ActiGraph® WGT3X-BT accelerometers, along with machine learning (ML) algorithms to automatically classify behaviour in six colony-housed domestic dogs (**Chapter 2**). A total of 132,295 seconds (~36.7 hours; ~6.1 hours per dog) of video footage of behaviour were recorded. Five modelling rounds were created using ML techniques, with model 4 achieving the highest overall accuracy and kappa coefficient while still capturing a wide range of behaviours. The study found that the ActiGraph® WGT3X-BT accelerometer can accurately classify behaviour in domestic dogs. However, it also revealed challenges in differentiating behaviours with similar acceleration profiles, particularly in classifying the "standing" behaviour. As a result, behaviours were grouped during the model-building process to improve overall accuracy. The refined models significantly improved over time, indicating a promising method for detailed and remote assessment of domestic dog behaviour. The secondary aim of the thesis was to evaluate the potential use of tri-axial accelerometery and a validated random forests model for determining the efficacy of environmental enrichment treatments and to assess the effect of food, olfactory, and tactile enrichment treatments on the behaviour and activity of six colony-housed domestic dogs (**Chapter 3**). Significant differences were observed among enrichment treatments regarding active/inactive behaviour, ODBA levels, individual behaviours, and interaction durations.

The use of ActiGraph® devices were demonstrated to be an accurate and objective method for measuring the success of enrichment activities, particularly when used alongside other observational methods such as interaction duration. The study emphasised the impact of environmental factors, individual differences among dogs, and seasonal variations on the effectiveness of enrichment. It highlighted the need for personalised enrichment programs to enhance the overall effectiveness of treatments.

Acknowledgements

Firstly, I would like to thank my excellent supervisors, Assoc. Prof David Thomas, Assoc. Prof Rene Corner-Thomas, Dr Chris Andrews and Dr Ina Draganova, for your continuous encouragement and guidance throughout each step of my Master's program over the last two years. I have learnt so much through your knowledge and expertise.

I would also like to thank Colin Naftel, the Massey University Canine Nutrition Unit manager, for taking such good care of the dogs in my study and giving me many great tips and tricks for making data collection as seamless as possible throughout my research. I would also like to thank Colin for all our great conversations during coffee breaks and his valuable insights during my time at the Canine Unit.

I want to express my gratitude to PhD student Michelle Smit, who took time out of her busy schedule to share her knowledge about the coding systems required for my study. Learning statistical coding systems was new to me, and I am grateful for Michelle's guidance and patience during this time.

I want to give lots of thanks and pats to the wonderful dogs who were part of my Master's research. I loved getting to know each dog's personality, and they always brightened my days.

Lastly, I would like to thank my family and friends who have supported me throughout my time at Massey University. I could not have done it without the constant encouragement and laughs with you all, especially Layla Conoglen, for keeping me sane as another Master's student. Special thanks to my parents, Siobhan and Aidan, for their unwavering support throughout my studies. Their constant presence and encouragement have been invaluable.

Contents

List of Figures

List of Tables

Chapter 1

Introduction- A review of the assessment of behaviour in domestic dogs (*Canis familiaris*)

Chapter 1: A review of the assessment of behaviour in domestic dogs

1.1 Domestication and behaviour overview

1.1.1 Domestication of dogs

It is widely accepted that the grey wolf (*Canis lupus*) is the progenitor of all domestic dog breeds today (*Canis familiaris*) and the first animal to be domesticated (Morey, 1994; Serpell, 2017). Domestication of dogs occurred at least 15,000 years ago (Morey, 1994; Vilà et al., 1997; Feddersen-Petersen, 2007; Thalmann et al., 2013; Ostrander et al., 2017; Wynne, 2021). Domestication is described as an evolutionary process whereby an animal develops an adaptation to man and a captive environment (Price, 1999). Purugganan (2022) further expands on this process, explaining that it arises from a mutualistic association between two species where one species (a domesticator) manages another species' survival and reproductive aspects in exchange for providing services and/or resources. This leads to a stable evolutionary relationship between the two species over generations (Purugganan, 2022).

During domestication, dogs learned to live in close contact with humans, leading them to communicate and cohabitate almost harmoniously (Thalmann et al., 2013). Consequently, it has been suggested that the domestication of dogs has significantly influenced the behavioural repertoire of dogs observed today (Feddersen-Petersen, 2007). Over many periods of domestication, selection pressures have resulted in significant changes in behaviour and morphology compared to their wild progenitors, thus resulting in dogs developing a dependence on humans for survival (Clutton-Brock and Museum, 1999; Wayne and Ostrander, 1999; Lord et al., 2016). Dogs have become less responsive to their environment and are notably different in their behaviour from their wild conspecifics, grey wolves. However, the grey wolf population, from which dogs were domesticated, exhibits notable distinctions from the modern-day grey wolves encountered today (Leonard, 2015). Furthermore, humans have selected for dogs that show traits of docility, adaptability to different environments, fixed paedomorphic features, and reduction of undesirable wild characteristics such as aggressive and antagonistic behaviour (Fox, 1978).

1.1.2 What is behaviour?

Grier (1984) described behaviour as 'an animal's action or reaction to a particular stimulus. However, in recent years, behaviour has been described as an animal's action displayed through its body plan (organisation of body parts and features) in response to the environment where internal/external stimuli are present (Miklósi, 2014). An animal's behavioural repertoire is characterised as fixed patterns of behaviour observed amongst a particular species and often a direct reflection of a species' environment (Huntingford, 2012). The 'normal' behaviour of domestic dogs is thought to be primarily influenced by human-dog coexistence. Additional factors affecting the domestic dog's behavioural development and how dogs respond to their environment include genetics, environmental factors, training, and socialisation (Svartberg and Forkman, 2002).

Behavioural tests have been used to evaluate and categorise an animal's behaviour and welfare state (Diederich and Giffroy, 2006; Dare and Strasser, 2023). The domesticated dog has become an integral part of human society, serving many roles such as companionship, working, hunting, herding, research, and military service (Menache, 1998; Svartberg and Forkman, 2002). Therefore, assessing behaviour can provide information on the temperament of an individual dog/breed and can be used to predict the suitability of dogs for specific societal roles (King et al., 2012). Understanding dog behaviour is also essential when selecting dogs for breeding, as breeding may be based on targeting particular temperament characteristics required for specific roles (King et al., 2012). Not only can behavioural assessments allow for the selection of desirable behavioural traits, but they can also allow us to understand the welfare state of dogs (Protopopova, 2016). Behavioural assessments are one of the best indications of emotional states, as they indirectly measure an animal's stress level by their behaviour (Dare and Strasser, 2023). Protopopova (2016) also stated that understanding abnormal behaviours or stress indicators can help determine the dog's welfare status. Consequently, behavioural assessments of dogs can guide improvements that need to be made to a dog's environment or daily life to enhance its welfare. However, accurate and reliable behavioural testing can be complex due to being labour-intensive and time-consuming (King et al., 2012).

1.2 Common behavioural assessments

1.2.1 Test batteries

Jones and Gosling (2005) reported that the most common dog behavioural assessment method was the 'Test Battery'. This method is described as a test in a controlled environment where the dog's behaviour and reactions are systematically evaluated in response to the presentation or withdrawal of a particular stimulus (Marcato et al., 2022). The measurement of the Test Battery can be conducted in multiple ways. Behavioural coding is typically recorded by a behaviour's absence, presence, frequency, or duration. These behavioural rating scales aim to group behaviours and create intensity levels for categories of behaviour (Dowling-Guyer et al., 2011). A significant advantage of the Test Battery is that it can be conducted in a controlled setting by trained staff over a short period of time while still being able to address the main behaviours (Dowling-Guyer et al., 2011). Another advantage is that this method appears more objective than other behavioural assessments, such as observational tests, making it more robust (Dowling-Guyer et al., 2011). Rayment et al. (2015) stated that an advantage of Test Batteries are that they are readily observable. However, they often lack detail and rely on groups of behaviours to represent the whole behavioural repertoire of a dog. Therefore, this can lead to inaccuracy when compared to the accurate behavioural representation of the dog (Rayment et al., 2015).

1.2.2 Observational methods

Observational behavioural assessments are tests that take place in non-controlled environments and allow for more natural surveillance of a dog's behaviour (Muñana et al., 2020). Thus, an ethogram is recommended to be developed as it can precisely identify specific behaviours (Fugazza and Miklósi, 2014). An ethogram is a listed catalogue of an animal's behavioural repertoire (Donát, 1991). It is important to note that an ethogram is only designed to approximate an entire behavioural repertoire, and therefore, only a portion of an animal's total behaviour can be included as many behaviours may be grouped (Donát, 1991). If a study is designed to reflect every behaviour of an animal, it would likely have complications with statistical analysis and become too explorative (Beerda et al., 1998).

Observational tests are limited by the risk of observer bias or differences among observers, which mainly occur if behaviours are not clearly defined (Donát, 1991). An observational behavioural assessment relies on human observers or video recordings (for retrospective evaluations). The role of human observers is to identify specific behaviours that occur in a particular environment. Observers typically focus on a minimal number of behaviours that are specifically relevant to that study (Fugazza and Miklósi, 2014). This reductionist approach is beneficial for the simplicity of statistical analysis and data collection, but it can result in a significant loss of information and produce results with high variability between observers (Fugazza and Miklósi, 2014). Observational tests aim to assess and describe relatively broad traits of behaviour that can be observed in a naturalistic environment (Jones and Gosling, 2005). This allows broader conclusions about the dog's behaviour that would not necessarily be explored when utilising behavioural tests such as Test Batteries. This is mainly because observational tests are usually performed in carefully selected but not controlled environments and incorporate naturally occurring stimuli.

1.2.3 Questionnaire

Questionnaire-based behavioural assessments are other standard methods professionals and researchers use for measuring behaviour (Hsu and Serpell, 2003; Temesi et al., 2014; Wiener and Haskell, 2016). An interview or questionnaire with a dog owner or professional is often targeted at assessing a dog's personality traits and indicating how a dog would behave in specific situations (Wiener and Haskell, 2016). This is typically measured on a point scale (Hsu and Serpell, 2003; Rayment et al., 2015). Rayment et al. (2015) added that it is essential to understand what 'personality traits' are being measured as they can be directly observable measures, behaviours, or more abstract constructs used to explain behaviour patterns between individuals. Often, questionnaire-based assessments are implemented in animal shelters when dogs are being relinquished, with staff interviewing owners to obtain behavioural information on the dog(s) (Segurson et al., 2005).

An example of a widely used questionnaire assessment is 'The Canine Behavioural Assessment and Research Questionnaire (C-BARQ)', an assessment designed to be a reliable, standardised method for testing behavioural problems in dogs (Hsu and Serpell, 2003; Duffy et al., 2014; Wauthier and Williams, 2018; Clay et al., 2020). The C-BARQ method measures 14 significant behaviours assessed on a 1-5 point scale alongside other additional measures. Questionnaire-based behavioural assessments depend on two critical assumptions for their success (Hsu and Serpell, 2003). The first assumption is that the person who answers the questionnaire knows the temperament and behaviour of the dog (Hsu and Serpell, 2003). Although questionnaires reduce the need for time-consuming behavioural tests and can identify a complex range of behaviours, it is not always possible to source an individual with adequate knowledge to record reliable information (Brady et al., 2018). The second assumption is that the appropriate questions will be answered candidly to ensure accurate and quantitative data (Hsu and Serpell, 2003). This assumption has been tested and contrasting studies exist on whether questionnaire-based behavioural assessments are biased (Segurson et al., 2005; Duffy et al., 2014). Wiener and Haskell (2016) concluded that although questionnaire-based assessments could provide many practical and financial advantages, they did not offer the same objectiveness and accuracy as standardised assessments conducted by knowledgeable observers.

1.3 Automated assessment of behaviour

Despite behavioural assessments being traditionally achieved through direct visual observation, recent technological advances have provided researchers with more automated methods of measuring behaviour (Hansen et al., 2007). Traditional methods of measuring behaviour can also be subject to bias, caused by either overestimation or underestimation of the behaviour and activity of animals, mainly if untrained observers or owners conduct the assessment (Lascelles et al., 2008; Dow et al., 2009; Morrison et al., 2013). Thus, there is a need for a method of behavioural assessment that is objective, practical and does not require extensive input time from a human observer (Jones et al., 2014). Activity counts (AC) are numerical measurements recorded by accelerometer devices that can be used to predict overall activity in dogs (Michel and Brown, 2011). While activity counts have provided valuable insights into general activity level differences between groups or treatments, they offer limited information regarding specific behaviours and activity patterns. Despite some limitations in the current models, technological approaches to measuring behaviour in dogs and other animals can revolutionize behavioural measurement in the veterinary field, agriculture industry, and future animal research studies (Rushen et al., 2012; den Uijl et al., 2017).

1.3.1 Pedometers

Researchers and practitioners have faced the challenge of identifying valid quantitative activity measurements for both animals and humans (Tudor-Locke et al., 2002). Pedometers were an early technology used to measure physical activity objectively (Chan et al., 2005). Pedometers are inexpensive and basic devices designed to measure simple physical activity (Tudor-Locke et al., 2002). Many pedometers count steps and estimate the distance travelled based on gait modelling (Ladha et al., 2018). In most pedometers, physical activity is detected by a horizontal, spring-suspended lever arm that responds to vertical acceleration (Tudor-Locke et al., 2002). Therefore, pedometers are predominantly designed to identify vertical accelerations, restricting their specific ability to detect ambulatory activities (Tudor-Locke and Myers, 2001). Unlike more complicated technological tools like tri-axial accelerometers, pedometers cannot measure patterns and intensity or be used to differentiate between various behaviours (Chan et al., 2005). Although pedometers may not be suitable for research that demands high precision in behavioural studies, they may still be helpful for studies focused on identifying daily variations in overall animal activity (Andrews et al., 2015). They have also been successfully utilised as an accurate activity measurement tool in dog obesity studies (Chan et al., 2005). Additionally, in terms of practicality, pedometers have been previously shown to be a favoured option for measuring activity as they are low-cost and do not require highly technical expertise to process data (Tudor-Locke et al., 2002).

1.3.2 Accelerometers

Over the past decade, research-grade and consumer-based accelerometer devices have gained popularity as activity measures in animal populations, including domestic dogs. They can provide a far more detailed behavioural assessment than pedometers, as they can record a subject's movement's intensity, duration, and frequency (Hansen et al., 2007; Lascelles et al., 2008; Yashari et al., 2015; Ladha and Hoffman, 2018). Accelerometers are small devices comparable to the size of a watch capable of monitoring activity counts and detecting changes in the intensities of certain behaviours over extended periods (Hansen et al., 2007; Dow et al., 2009; Brown et al., 2010). Due to recent advances in accelerometer technology, the reduced size of the devices has allowed them to monitor the activity of companion animals (Barthélémy et al., 2009; Smit et al., 2023).

Numerous studies have demonstrated that accelerometers can objectively and accurately measure behaviour and activity levels in companion animals (Hansen et al., 2007; Dow et al., 2009; Moreau et al., 2009; Brown et al., 2010; Yam et al., 2011; Preston et al., 2012; Chambers et al., 2021; Smit et al., 2023). It has been noted that the functionality of accelerometer models can vary widely. Most accelerometer devices contain bi-axial or tri-axial accelerometers, enabling omnidirectional movement detection (Watanabe et al., 2005; Lascelles et al., 2008; Yam et al., 2011; Jones et al., 2014). Consequently, the advantage of these activity monitors is that they can detect dynamic and static behaviours (Watanabe et al., 2005).

Accelerometer devices such as the Actical® (Phillips Respironics, Netherlands) and ActiGraph® GT3X (ActiGraph Inc, USA) have commonly been used in both human and animal research as they provide software that enables data to be easily exported in a format suitable for post-processing data analysis (Michel and Brown, 2014; Morrison et al., 2014). However, devices such as these have been developed and manufactured to assess human activity. Thus, users interested in other species' behaviour may have to validate their own species-specific algorithm, which can be complicated and time-consuming (Ladha et al., 2018). Additionally, using accelerometers requires complex data analysis. It can incur high costs and additional resources such as software programs and technical expertise, which has prevented their use in wider-scale applications (Tudor-Locke and Myers, 2001). Therefore, it is critical to validate these devices before using them in companion animal research or on a commercial scale (Ladha et al., 2018).

1.3.3 Benefits of the ActiGraph® device

ActiGraph® devices are tri-axial, meaning they can measure movement in three directions: ‘x’, ‘y’, and ‘z’ (Morrison et al., 2013). Additionally, ActiGraph® devices are hard and durable whilst remaining small and lightweight, making them practical for pets to wear, including those who spend a lot of time outdoors (Yam et al., 2011). Additionally, dogs that have worn the

ActiGraph® devices for extended periods do not seem affected by wearing them (Yam et al., 2011). Previous studies demonstrate that ActiGraph® devices have successfully measured activity levels in domestic dogs (Yam et al., 2011; Morrison et al., 2013; Ortmeyer et al., 2018). For instance, the ActiGraph® GT3-X was validated by monitoring different activity intensities, demonstrating that the accelerometers provided a valid and reliable measure of activity in free-living dogs (Yam et al., 2011). Additionally, Morrison et al. (2013) employed the ActiGraph® GT3-X to assess physical activity intensity in domestic dogs, focusing on investigating the relationship between obesity and physical activity. A similar study examined the correlation between post-chemotherapy physical activity in domestic dogs (Helm et al., 2016). While these studies have investigated activity levels using accelerometers, there is a noticeable gap in research targeting dogs' specific behaviours. Recently, the ActiGraph® wGT3X-BT was successfully validated for behaviour classification in domestic cats (Smit et al., 2023). This suggests potential applications are possible in behavioural studies in dogs.

1.3.4 How does the ActiGraph® device work?

Tri-axial accelerometers such as the ActiGraph® use a piezoelectric bi-morph plate to generate a voltage proportional to the change in velocity per unit of time (Dow et al., 2009). The piezoelectric bi-morph plate and seismic mass detect movement in three axes (x, y, and z) (Lascelles et al., 2008). Essentially, a voltage is generated by the piezoelectric sensor in response to the subject's acceleration. The piezoelectric sensor is sensitive to movement in all directions, but is most sensitive in the direction parallel to the accelerometer case's longest side (Hansen et al., 2007). The acceleration force causes the piezoelectric bi-morph plate to bend, generating voltage output amplified and converted into a quantifiable digital value by an inbuilt microprocessor (Lascelles et al., 2008). To ensure the data's accuracy, the voltage generated by the sensor is amplified and then filtered by analogue circuitry (Hansen et al., 2007; Preston et al., 2012). The filtered and amplified voltage is then passed through an analogue converter to create a digital value (Hansen et al., 2007; Yam et al., 2011). Further, this digital value can then be used to adjust a running baseline value so that constant accelerations such as gravity can be filtered out (constant acceleration of 9.8 ms^{-2}) (Hansen et al., 2007). The digital value can then be compared to the baseline value, and raw activity values will be created for a summary period, also known as an epoch (Hansen et al., 2007).

Most accelerometers record data using user-defined epochs with proprietary software (Preston et al., 2012). Epochs are described as a set of sampling intervals that indicate how many activity counts will be recorded over a period of time. They additionally compress raw activity and convert it into an unsigned integer using computer software, which is then reported as an activity count, following a calibration constant. This raw activity data can be downloaded and used to produce an Actogram of raw activity counts. The raw activity counts for each epoch over the total sampling period can also be exported for statistical analysis.

1.4 Applications of accelerometer devices

1.4.1 Behavioural assessment

Accelerometer devices have enabled the objective quantification of various aspects of dog behaviour, including overall daily activity, resting periods, changes in behaviour, postural changes, and even movement patterns associated with each posture using machine learning analytical methods (Yashari et al., 2015; Kumpulainen et al., 2021).

Accelerometer technology has been used in previous dog behaviour studies, facilitating the detection of fundamental aspects of dog behaviour, such as the intensity of activity that a dog is exhibiting (Yam et al., 2011; Morrison et al., 2013; Ortmeyer et al., 2018). There are clear advantages of utilising accelerometers to measure the behaviour and activity levels of dogs, including the objective and accurate quantification of behaviour without a human observer being present or labour-intensive analysis of video recordings (Yam et al., 2011; Kumpulainen et al., 2021).

The advancement of accelerometer technology has also extended beyond behavioural research. From an ethical perspective, activity monitors are non-invasive and lightweight devices that can be easily attached to a dog collar or harness without causing disturbance to the expected behaviour of the animals (Olsen et al., 2016). From a pet owner's perspective, further advancement in accelerometer technology could allow a better understanding of their dog's behaviour, aiding in the early detection of issues such as separation anxiety (Kumpulainen et al., 2021). Furthermore, this data can indicate the dog's overall well-being and stress levels, offering valuable insights for pet owners (Kumpulainen et al., 2021).

In terms of precision, accelerometer behavioural classification algorithms have made significant progress. Ladha et al. (2013) found an overall accuracy of 68.6% when distinguishing 16 different dog behaviours in their home environment. However, in another study, a much higher accuracy of 95% was reported when classifying behaviours such as walking, trotting, cantering, eating, drinking and headshaking (den Uijl et al., 2017). Resting behaviours have also been successfully distinguished, with an 86% accuracy rate (Ladha and Hoffman, 2018).

1.4.2 Identification of stress and expression of stereotypical behaviour

Accelerometers have the potential to be used to differentiate between normal and excessive stress-related behaviours in dogs and other companion animals, especially shelter dogs. Notably, Jones et al. (2014) reported that an accelerometer device, 'SNIF tag' (SNIF labs, LLC), could determine the activity levels of the shelter dogs. The SNIF tag was embedded in a radiofrequency identification tag and could capture and transmit motion data and direct it to a central online database (Jones et al., 2014). Interestingly, this study showed that abnormally high and low activity levels, as recorded by the accelerometers, served as stress indicators among the shelter dogs (Jones et al., 2014). Additionally, a study by Hoffman et al. (2019) used the 'VetSens' (VetSens, Newcastle, United Kingdom) triaxial accelerometer sensor to compare the intensity and consistency of shelter dogs activity levels compared to pet dogs as an indicator of stress. The activity data collected in the study suggested that the environment of shelter dogs may prevent them from having as much rest as pet dogs (Hoffman et al., 2019). Collectively, these studies indicate that accelerometer technology could help study sleep quality and the overall psychological well-being of dogs. Accelerometer data could also provide essential insights into how well dogs handle the transition between the shelter and their new adoptive home, as this transition can be difficult for some dogs, resulting in a negative welfare state (Hoffman et al., 2019).

1.4.3 Gait analysis

Previous research has commonly utilised accelerometers to measure activity in movement intensity, activity counts, and arbitrary measures (Brown et al., 2010; Knazovicky et al., 2015). Using tri-axial accelerometers that can measure acceleration in three different planes allows the detection of parameters such as stride length, stride frequency, and power output in different movements (Pillard et al., 2012).

The categorisation of dog locomotion by gait or speed into well-defined variables holds substantial promise. This approach can permit the early identification of adverse health outcomes, such as lameness, but also facilitate the discernment of treatment effects of therapeutic compounds (Pillard et al., 2012; Bolton et al., 2021). Barthélémy et al. (2009) used accelerometers to analyse the gait of dogs affected with Golden Retriever Muscular Dystrophy (GRMD). The study found that dogs suffering from GRMD tended to have a shortened stride, slower locomotion and unsteady gait (Barthélémy et al., 2009). This information provided valuable insight for assessing the fall risk and functional capacity of dogs that suffered from GRMD (Barthélémy et al., 2009).

Pillard et al. (2012) later investigated predicting lameness in dogs using accelerometers. They found that symmetry index (SI) and regularity index (RI) were reliable predictors of walking gait for detecting even slight lameness (Pillard et al., 2012). Bolton et al. (2021) later explored this theory in a study where dogs wore tri-axial accelerometers while running on a treadmill. Gait analysis showed that total delta-G, a sum of acceleration across the three planes recorded by the accelerometer, could be used to predict the speed or gait of these dogs (Bolton et al., 2021). However, the study revealed that the delta-G sums could not distinguish between faster gaits such as trotting and cantering, so it was recommended that a more complex delta-G analysis be carried out to gather a more detailed categorisation of gait (Bolton et al., 2021). They noted, however, that the distinction between faster gaits may not be necessary when investigating whether the dog is experiencing injury, illness or the effects of a therapeutic substance (Bolton et al., 2021).

1.4.4 Health monitoring

Accelerometers are not only capable of detecting changes in behaviour or activity, but they can also recognise health risks in dogs (den Uijl et al., 2017). To improve the overall welfare state of dogs, improve training success, and help humans interpret the emotional responses of dogs, there is an increasing interest in methods to continuously monitor dog health from pet owners, dog handlers and veterinarians (Brugarolas et al., 2015). This research has the potential to be applied to other areas of dog health, such as predicting seizures, post-surgery activity, and evaluation of medical interventions where activity monitoring of dogs is required (Yam et al., 2011). Muñana et al. (2020) investigated whether accelerometers could predict seizures. They utilised a combination of accelerometery and questionnaire assessments as the basis of behaviour measurement. The study showed that seizures could be detected through accelerometery, although further device development was needed for accurate predictions (Muñana et al., 2020).

Accelerometer technology has also been used to detect and prevent animal dysfunction or disease (Michel and Brown, 2011; Gerencsér et al., 2013; Cheung et al., 2014; Clark et al., 2014; Clarke and Fraser, 2016). Many diseases and dysfunctions show exercise-related outcome measures, so overall activity counts and behavioural classification can indicate health status (Ladha et al., 2018). In a study by Helm et al. (2016), the effect of different chemotherapy treatments on healthy domesticated dog physical activity was assessed using ActiGraph® accelerometer devices. The study successfully utilised these devices to collect objective data on the physical activity levels of dogs receiving chemotherapy treatment (Helm et al., 2016). It stated that this technology could be applied to determining patients' quality of life (QOL) in combination with other activity measures, such as questionnaires (Helm et al., 2016).

1.4.4.1 Obesity

Obesity is one of the most significant health problems for companion animals and can lead to numerous other health issues, including orthopaedic disease, diabetes mellitus, lipid profile abnormalities, cardiorespiratory disease, urinary disorders, neoplasia, reproductive disorders, dermatological diseases and anaesthetic complications (German, 2006). These conditions reduce the quality of life of affected animals and cause suffering, often resulting in shortened life expectancy (Sandøe et al., 2014). Accelerometers can contribute to an improved understanding of how physical activity is related to dog obesity (Kienzle et al., 1998; Courcier et al., 2010). Evidence suggests that habitual physical activity and the weight status of dogs and owners are related (Chan et al., 2005; Cutt et al., 2008). Therefore, accelerometers could be used as a tool to prevent obesity in both dogs and humans. The relationship between diet and activity has been examined in adult domestic cats successfully using Actical® accelerometers to assess the impact of increased meal frequency and dietary water content on physical activity (Deng et al., 2014).

1.4.4.2 Degenerative joint disease

Degenerative joint disease (DJD), also known as Osteoarthritis (OA), is a slowly progressing, chronic, degenerative disease that commonly causes pain, swelling of the joint, and lameness (Aragon et al., 2007; Ksenija, 2019). Researchers have quantified an objective outcome measure for DJD/OA using accelerometer activity counts (AC) (Brown et al., 2010; Rialland et al., 2012; Belshaw et al., 2016). In a study by Brown et al. (2010), dogs with clinical signs of OA wore Actical® activity monitors and were treated with either a placebo or carprofen, a known pain relief designed for dogs with OA. It was found that dogs treated with carprofen had a 20% greater increase in activity counts than those in the placebo group (Brown et al., 2010). Another study used motor activity (MA) recorded using telemetered accelerometric counts to monitor how diet affected adult dogs with pain caused by OA (Rialland et al., 2012). The results from this study were varied, and methods would require some refinement to be used as an accurate measure of OA pain in the future (Rialland et al., 2012). Therefore, further research is needed to explore the relationship between diet and OA, using accelerometers.

1.4.5 Environmental enrichment

1.4.5.1 Review of environmental enrichment in domestic dogs

Environmental enrichment is a strategy used to enhance the quality of life of domestic and captive animals (Hunt et al., 2022). This is achieved by providing additional temporary stimuli and activities to an animal's external environment (Desforges, 2021; Hunt et al., 2022). For environmental stimuli to be considered enriching, it must improve an animal's overall welfare state (Hunt et al., 2022). The benefits of environmental enrichment include increased frequency and diversity of species-specific behavioural repertoires, reduced stress and anxiety, reduced abnormal behaviours, and improved cognitive abilities (Hubrecht, 1993; Shepherdson, 2003; Wells, 2004; Schipper et al., 2008; Herron et al., 2014; Hunt et al., 2022; Kang, 2022).

Environmental enrichment is a tool that has been used successfully in previous research for shelter dogs, kennelled dogs, research dogs, companion dogs, and working dogs (Amaya et al., 2020; Desforges, 2021; Hunt et al., 2022; Kang, 2022). Currently, environmental enrichment strategies are mainly focused on dogs living in confined environments such as kennels, shelters, and research facilities, as these environments can increase stress levels and problem behaviours in dogs (Hiby et al., 2006; Herron et al., 2014). This is portrayed in a variety of behavioural disorders, including excessive licking, stiff posture, sighing, and even depression (Hiby et al., 2004). This can be a problem for dogs in shelter environments as these traits can decrease their chance of adoptability (Hennessy et al., 2002). Dogs will be provided with different environmental enrichments based on their individual behaviour and needs (Hunt et al., 2022). Other factors determining the type of enrichment supplied include space availability, cost, practicality, and equipment durability (Desforges, 2021).

Food-based enrichments are popular for domestic dogs; however, previous research indicates varied results (Gaines et al., 2008; Schipper et al., 2008; Herron et al., 2014; Hunt et al., 2022). Food-based enrichment typically involves scattering food around an animal's environment or hiding food in toys or puzzle feeders (Markowitz, 1982; Newberry, 1995; Young, 1997). Food-based enrichment methods aim to promote natural appetitive feeding behaviours that they don't typically have the opportunity to display in captive environments (Schipper et al., 2008). Schipper et al. (2008) investigated the effect that feeding enrichment toys such as the 'kong' had on activity levels, appetitive levels and display of abnormal behaviour in laboratory dogs in kennelled environments. The study revealed that the presence of these feeding toys did improve activity levels and frequency of appetitive behaviours, and it was found that barking levels decreased (Schipper et al., 2008). However, it was also noted that these benefits of the food-based enrichment were short-lasting, and these positive behaviours were no longer observed after the kong toy was removed (Schipper et al., 2008).

Similarly, another study reported the effects of providing Kong's to kennelled military working dogs (Gaines et al., 2008). It was reported that there was no noticeable improvement in the dog's overall behaviour or working ability after providing this feeding enrichment (Gaines et al., 2008). In contrast, Herron et al. (2014) found that a complex enrichment protocol involving multiple feeding enrichment strategies resulted in lower stress levels and better overall welfare of kennelled dogs.

Sensory enrichment is a practical strategy for environmental enrichment due to its low cost, simplicity, and effectiveness in stimulating domestic dogs (Desforges, 2021). Sensory enrichment includes auditory, olfactory, and visual methods of stimulation. The application of sensory enrichment methods can be adjusted to indoor and outdoor environments, and these methods have previously been linked to reducing stress levels in domestic dogs (Wells, 2009; Amaya et al., 2020). Typical approaches to olfactory enrichment include scent work and scent play, where dogs learn how to use their olfactory system to find designated toys, food or people (Desforges, 2021). The reported benefit of this strategy is that scent work is tiring for dogs due to their highly developed olfactory system (Desforges, 2021). Therefore, it can be a relaxing and mentally stimulating activity for anxious and excitable dogs while promoting natural behaviour (Desforges, 2021). Odours such as essential oils, prey urine, and dog-appealing pheromones have all been used as effective olfactory enrichment treatments, resulting in decreased stress levels, increased overall activity, and reduced barking (Wells, 2004; Tod et al., 2005; Wells, 2006; 2009).

There has been growing interest in the veterinary field regarding musical therapy as an auditory enrichment tool due to its low cost and practicality (Lindig et al., 2020). It has been reported that classical music can improve dogs' mental state and increase relaxation, therefore improving the long-term welfare of dogs, particularly in kennelled environments (Wells et al., 2002; Kogan et al., 2012; Lindig et al., 2020). Visual and auditory enrichment strategies such as DOGTV® have been reported to be effective when used alongside a complex enrichment plan in kennelled environments where social contact is limited (Epstein et al., 2021).

Tactile enrichment, such as toys and blankets, is one of the most frequently employed enrichment methods for dogs and other domestic and exotic captive animals (Wells, 2004; Dare and Strasser, 2023). The reported benefits of providing toys for captive and domestic animals include reduced boredom, encouraging natural play behaviours, promoting exploration, and reducing abnormal behaviours (Wells, 2004). However, the effectiveness of toy enrichment for domestic dogs has not been well documented (Pullen et al., 2010). Garvey et al. (2016) noted that it is essential for toy enrichment to be carefully catered towards each dog's needs and rotated frequently to ensure that the novelty of the item does not wane over time. Krichbaum et al. (2023) investigated chew toys' effect on cognition in dogs. It was reported that dogs with high fearfulness were provided with chew toys, which resulted in improved cognition levels, whilst the opposite effect was observed for dogs with low fearfulness (Krichbaum et al., 2023). However, there have been mixed results on the effectiveness of food-based toys and games in domestic dogs (Gaines et al., 2008; Schipper et al., 2008; Hunt et al., 2022).

1.4.5.2 Application of accelerometery

The variability in efficacy seen in enrichment studies may be attributed to the absence of objective and continuous measures for quantifying the effectiveness of enrichment in domestic dogs. Recent research has increasingly utilised accelerometers as a valuable tool for behavioural assessment in enrichment studies. Tri-axial accelerometers have effectively measured enrichment success in various animal species, including pigs, rats, cattle, and chickens (Sudo et al., 2018; Liu, 2019; Bruno et al., 2020; Pullin et al., 2020; Veldkamp et al., 2023). Animal welfare researchers within zoo settings have also begun utilising automated technologies such as accelerometers to evaluate animals' responses to their surroundings and assess positive welfare status (Whitham and Miller, 2016). There has been minimal research carried out to determine the effects of environmental enrichment, specifically on dogs. However, accelerometers have previously been used to measure dogs' stress levels, behaviour, and welfare states (Jones et al., 2014; den Uijl et al., 2017; Kumpulainen et al., 2021; Hussain et al., 2022). Therefore, there is a clear gap in research on using accelerometers to measure environmental enrichment's effects on domestic dogs.

The first aim of this thesis is to investigate the use of remote sensing technology, particularly tri-axial ActiGraph® WGT3X-BT accelerometers, along with machine learning (ML) algorithms to automatically classify behaviour in six colony-housed domestic dogs (**Chapter 2**). The secondary aims of the thesis are to evaluate the potential use of tri-axial accelerometery and a validated random forests model for determining the efficacy of environmental enrichment treatments and to assess the effect of food, olfactory, and tactile enrichment treatments on the behaviour and activity of six colony-housed domestic dogs (**Chapter 3**).

Chapter 2

The use of tri-axial accelerometers and machine-learning algorithms for behavioural identification in domestic dogs (*Canis familiaris*): a validation study

Part of this chapter is currently under review for publishing as:
Redmond, C., Smit, M., Draganova, I., Corner-Thomas, R., Thomas, D., Andrews, C., 2024. The use of triaxial accelerometers and machine learning algorithms for behavioural identification in domestic dogs (*Canis familiaris*): a validation study. Sensors. Status: Under review.

Chapter 2: The use of tri-axial accelerometers and machine learning algorithms for behavioural identification in domestic dogs (*Canis familiaris*): a validation study

2.0 Abstract

Monitoring and quantifying domestic dog (Canis familiaris) behaviour is crucial for understanding their health and well-being. Traditional behavioural observations for domestic dogs can be labour-intensive, time-consuming, and subjective. This study aimed to investigate the use of remote sensing technology (tri-axial ActiGraph® WGT3X-BT accelerometers), along with machine learning (ML) algorithms to automatically classify behaviour in six domestic dogs. A total of 132,295 seconds (~36.7 hours; ~6.1 hours per dog) were recorded for behavioural observation. Five modelling rounds were created using ML techniques, with model 4 achieving the highest overall accuracy and kappa coefficient while still capturing a wide range of behaviours. The study found that the ActiGraph® WGT3X-BT accelerometer can accurately classify behaviour in domestic dogs. However, it also revealed challenges in differentiating behaviours with similar acceleration profiles, particularly in classifying the "standing" behaviour. As a result, behaviours were grouped during the model-building process to improve overall accuracy. The refined models showed significant improvement over time, indicating a promising method for detailed and remote assessment of domestic dog behaviour. Future research should validate ActiGraph® devices for larger groups of domestic dogs, diverse breeds, and various environments. Additionally, it would be useful to investigate whether any predictor variables used to develop the models are redundant and improve these to increase the overall accuracy of predictive models in the future.

2.1 Introduction

Domestic dogs (*Canis familiaris*) have become integral to human society, serving various purposes such as companionship, working, hunting, research, and military service roles (Menache, 1998; Svartberg and Forkman, 2002). Assessing their behaviour is critical for predicting suitability and targeting specific temperament characteristics required for these particular roles (King et al., 2012). Importantly, behavioural assessments not only facilitate the selection of desirable traits but also enhance our understanding of the overall welfare state of dogs (Diederich and Giffroy, 2006; Protopopova, 2016; Dare and Strasser, 2023). Standard behavioural assessment methods include test batteries, observational methods, and questionnaires (Jones and Gosling, 2005; Duffy et al., 2014; Brady et al., 2018). However, these methods can be time-consuming, labour-intensive, and subject to bias, making large-scale implementation challenging (Donát, 1991; Lascelles et al., 2008; Dow et al., 2009; Morrison et al., 2013; Rayment et al., 2015; Wiener and Haskell, 2016).

Thus, there is a pressing need for objective, practical behavioural assessment methods that do not require extensive time or training from human observers (Jones et al., 2014). Recent technological advances have paved the way for remote and automated behavioural measurement. Pedometers were among the earliest technologies used to measure physical activity objectively (Chan et al., 2005). Although relatively inexpensive and easy to use, pedometers' main limitation is that they cannot measure the intensity or differentiate movement patterns due to their simple design (Tudor-Locke et al., 2002; Chan et al., 2005). On the other hand, accelerometer devices offer a far more detailed behavioural assessment by recording the intensity, frequency, and duration of every movement across three axes: x, y, and z (Hansen et al., 2007; Lascelles et al., 2008; Yashari et al., 2015).

Recent advances in accelerometer technology have seen a reduction in the sizes of devices, enabling their use in monitoring the activity of companion animals (Barthélémy et al., 2009;

Smit et al., 2023). Numerous studies have demonstrated that accelerometers can provide objective and accurate measurements of behaviour and activity levels in companion animals (Hansen et al., 2007; Dow et al., 2009; Moreau et al., 2009; Brown et al., 2010; Yam et al., 2011; Preston et al., 2012; Smit et al., 2023). ActiGraph® devices have been previously validated in dogs and are tri-axial accelerometers, allowing for observation of omnidirectional movement (Yam et al., 2011; Ortmeyer et al., 2018; Hoffman et al., 2019). This enables them to detect dynamic and static behaviours and provide a comprehensive view of the dog's overall activity.

The application of accelerometers, such as the ActiGraph®, can enable objective quantification of various aspects of dog behaviour, including overall daily activity, resting periods, changes in behaviour, postural changes, and even movement patterns associated with each behaviour using machine learning analytical methods (Yashari et al., 2015; Kumpulainen et al., 2021). Machine learning (ML) allows complex data sets, such as those produced by accelerometers, to be analysed using complementary data modelling techniques (Valletta et al., 2017). This allows for accurate predictions to be generated from previously unobserved data (Valletta et al., 2017).

Modern technology, such as accelerometers, could also be used to identify stress-related behaviour, perform gait analysis, and monitor health on a broader scale (Barthélémy et al., 2009; Michel and Brown, 2011; Pillard et al., 2012; Gerencsér et al., 2013; Cheung et al., 2014; Clark et al., 2014; Jones et al., 2014; Clarke and Fraser, 2016; Hoffman et al., 2019; Muñana et al., 2020; Bolton et al., 2021). Therefore, this study aims to validate the ActiGraph® WGT3X-BT accelerometer as a tool for the behavioural classification of domestic dogs.

2.2 Methods

This study was conducted at Massey University Canine Nutrition Unit (CNU), Palmerston North, New Zealand (latitude 40°230'S, longitude 175°365'E) from August to September 2023. All research was conducted in accordance with Massey University Animal Ethics Committee (MUAEC) protocol number 23/27. All husbandry of the dogs complied with MUAEC protocol number 21/25 and the Animal Welfare Code of Welfare: Dogs (Ministry for Primary Industries, 2018).

2.2.1 Animal husbandry

Six healthy domesticated dogs from the CNU were enrolled in the study (**Table 2.1**). Two dogs were female, and four were male (all de-sexed), aged 3.9-7.5 years (mean ± SD, 6.02 ± 1.59 years). The dogs participating in the study were all healthy and had weights that ranged between 22.7-32.8 kgs (mean ± SD, 25.98 ± 4.27 kgs).

The CNU, a purpose-built colony facility, housed 29 domestic dogs (10 female and 19 male) at the time of the study. Dogs were managed in outdoor exercise paddocks (**Figure 2.1**) during daylight hours (07:00 h to 16:00 h) and subsequently spent the night in central heated indoor runs (16:00 h to 07:00 h). Dogs were normally kept in specific pairings, and they continued to be housed in these pairs throughout the study. The standard dietary regimen of the dogs used in the study consisted of a complete and balanced adult maintenance diet (Black Hawk Working Dog, Masterpet Corporation Ltd., Lower Hutt, New Zealand) fed daily in the morning. In the current study, the typical daily food allocation of the dogs was fed across three meals throughout the day, twice in the morning and once in the afternoon, to increase the frequency of feeding behaviours during the data collection period. The morning feeds usually occurred at 09:00 h and 10:00 h, and the afternoon feed usually occurred between 15:00 h –

16:00 h. This did not affect the normal amount of food consumed by each dog participating in the study. There was *ad libitum* access to water in their outdoor paddocks and indoor pens.

Table 2.1 The name, sex, age, breed, reproductive status and body weight of the six domesticated dogs enrolled in the study.

Name	Sex	Pair	Age (years)	Breed	Desexed	Weight (kg)
Belvedere	Female	3	7.5	Huntaway	Yes	22.4
Blacky	Male	3	3.9	Huntaway/Heading	Yes	23.9
Chevelle	Female	2	7.5	Huntaway	Yes	23.0
Gizmo	Male	1	5.7	Harrier Hound	Yes	31.1
Gus	Male	2	4.0	Huntaway/Smithfield Terrier	Yes	22.7
Monaro	Male	1	7.5	Huntaway	Yes	32.8

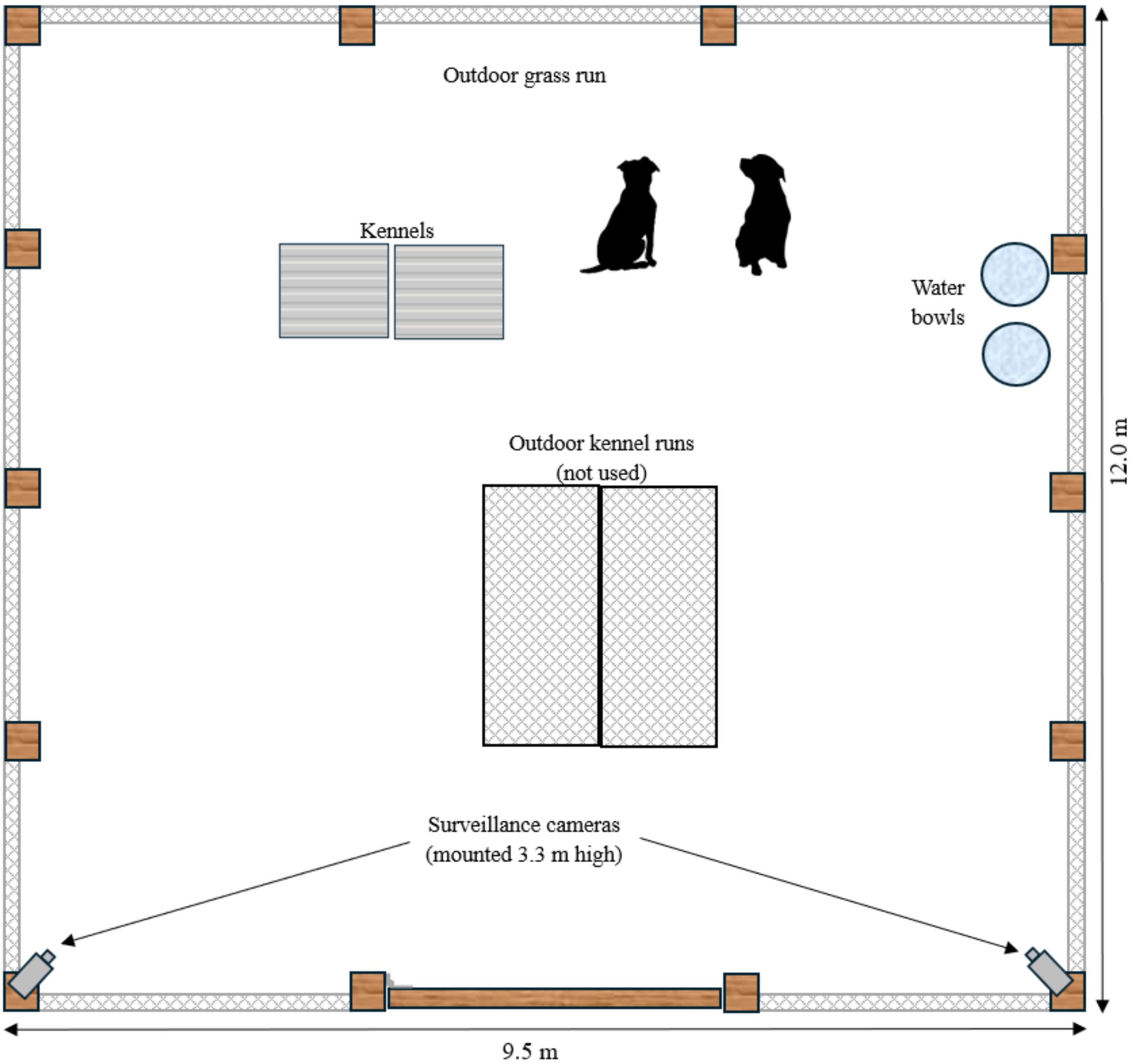

Figure 2.1 Diagram of the outdoor observation paddock showing paddock dimensions and features, including the positioning of the two surveillance cameras used to monitor the animals. Note that the image has not been drawn to scale.

2.2.2 Study design

The study comprised a multi-day experimental design, encompassing an initial one-day habituation phase followed by a three-day data collection (validation) phase for each pair of dogs. All the data collection for this study occurred between August and September 2023. A single observation paddock was used for the study (**Figure 2.1**). The data collection phase took place over three weeks, and one pair of dogs was monitored per week. Data collection was conducted sequentially for each pair of dogs. The first pair (Gizmo and Monaro) underwent a habituation day on August 29th, followed by a three-day validation phase from August 30th to September 1st. Subsequently, the second pair (Chevelle and Gus) underwent a habituation day on September 6th and participated in a three-day validation phase from September 7th to September 9th. The third pair of dogs (Belvedere and Blacky) underwent their habituation day on September 12th and participated in the three-day validation phase between September 13th and September 15th. After each recording week, the next pair of dogs was rotated into the observational paddock accordingly, and the methods were repeated.

2.2.3 Habituation phase

For each pair of dogs, activity monitors were attached to the dog’s collar on the day preceding the validation phase to habituate the dog's to wearing the device. During this habituation phase, dogs were placed in the observation paddock and observed to ensure there were no noticeable adverse effects on their behaviour or well-being. If any of the dogs exhibited signs of reactivity for extended periods while wearing the accelerometer device during the habituation period, the habituation period was to be extended by one day. No dogs in this study needed the habituation phase to be extended. Furthermore, during the habituation phase, the dogs were also acclimated to a new feeding schedule, where they were fed three times daily within their observation paddock instead of once a day in their typical indoor run. During each feeding period, the dogs were monitored to ensure that this adjustment did not cause any alterations to their feeding behaviour or increase aggression.

2.2.4 Activity devices

The ActiGraph® wGT3X-BT (ActiGraph®, Pensacola, FL, USA) activity monitors were used to measure the activity of dogs in this study. These activity monitors contained a tri-axial accelerometer, which measured acceleration in three independent dimensions: x (up/down), y (forwards/backwards) and z (left/right). The piezoelectric sensor of the ActiGraph® senses an acceleration force in response to varying accelerations ranging in magnitude from approximately 0.05 to 2.5 grams (Yam et al., 2011). The voltage output produced by the acceleration force is then amplified and converted into a digital value at a rate of 30 times per second (Lascelles et al., 2008; Yam et al., 2011). Once digitised, the signal is filtered to a frequency range of 0.25 to 2.5 Hz, and epochs are formed due to the sum of samples over specific time intervals (Yam et al., 2011). The ActiGraph® wGT3X-BT activity monitors weighing 19 grams and measuring 33 × 46 × 15 mm (**Figure 2.2a**) were placed inside a protective casing. The casing was then attached to an existing collar worn by the dogs and positioned ventrally (**Figure 2.2b**).

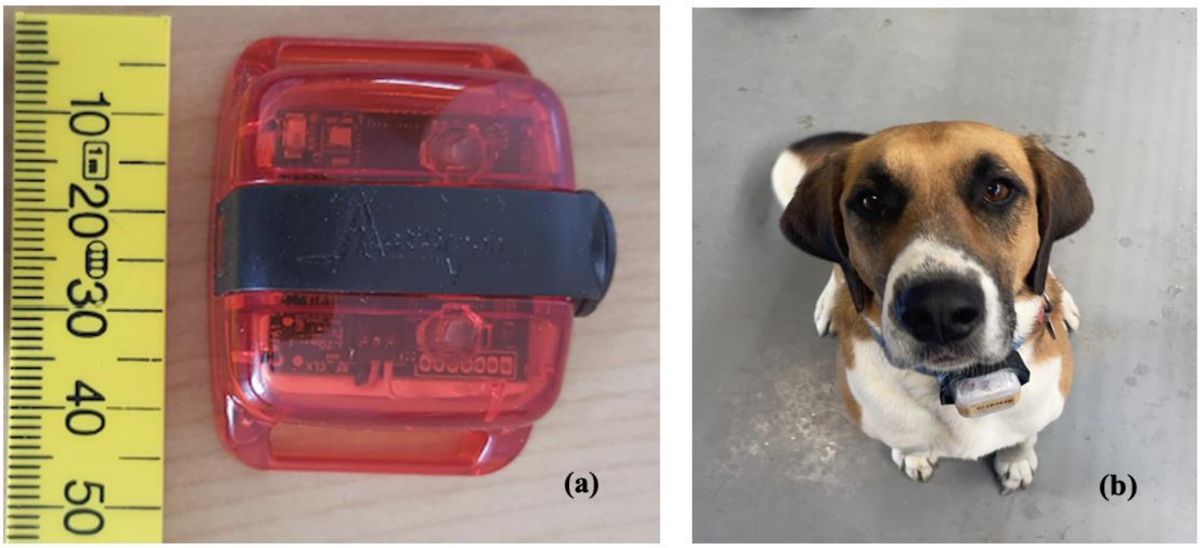

Figure 2.2 (a) Photo of ActiGraph wGT3X-BT device with relative measurements **(b)** Ventral view of the ActiGraph wGT3X-BT accelerometer, which was consistently orientated, placed within a protective housing and fitted ventrally to the collars of the dogs.

The activity monitors were encased in a plastic container and filled with bubble wrap to protect the devices from damage and avoid movement in the plastic casing (**Figure 2.2b**). Waterproof tape was wrapped around the ActiGraph® devices to further protect them from water damage. The accelerometer devices were fitted to ensure a snug fit, and the specific hole in the collar used to secure it to the dog was noted to ensure that the tension on the collar was consistent throughout the study period (Martin et al., 2016; Olsen et al., 2016). All the devices were oriented in the same direction and fitted to ensure that accelerometers were all exposed to similar forces (Hansen et al., 2007; Yam et al., 2011; Olsen et al., 2016). The devices were inserted into the plastic casing with the screw-lid of the device oriented downward and the broader side of the device positioned towards the outer edge of the casing, away from the collar (**Figure 2.3**). The collar was positioned ventrally under the neck to collect an acceleration profile for each dog for the duration of the validation trial (Olsen et al., 2016). The plastic casing surrounding the devices was wrapped in insulation tape to protect the clips from damage and to avoid discomfort or rubbing on the dog's skin. The total weight of the collar and device inside the plastic casing was 115.2 grams, which equated to ~ 0.4 % of the dog's body weight.

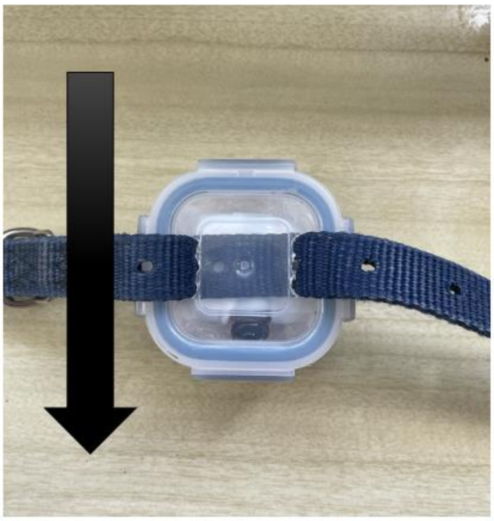

Figure 2.3 Orientation of ActiGraph device with screw lid facing downwards.

2.2.5 Collection and assessment of the video footage

The observation paddock was under constant video surveillance during the day using a 4K security camera system (Swann® Communications USA, Santa Fe Springs, CA, USA). Two cameras were mounted at elevated positions (3.3 meters above the ground) on adjacent corners of the observation paddock (**Figure 2.1**), allowing for an almost continuously unobstructed view of the dogs throughout the paddock. Video footage was recorded at 15 fps with a resolution of 1920 x 1080 and a bit rate of 2048 Kbps.

The behaviour of each of the six dogs was scored continuously (1 s intervals) from the recorded video footage using BORIS® version 7.10.2 (Friard et al., 2016). A total of 18 behaviours were scored by a single observer using the ethogram presented in **Table 2.2**. These behaviours were categorised as either active (walking, trotting, running, jumping, digging, barking, standing and sniffing), inactive (resting head up, resting head down, sitting, and lateral recumbency) or maintenance (defecating, urinating, eating, drinking and scratching), after Yam et al. (2011). In addition, an 'Other' category was created to account for behaviours not listed in Table 2, and an 'Out of sight' category was used when the dogs could not be seen clearly in the observation paddock.

2.2.6 Building and assessing the behavioural models

All statistical analyses were conducted using R statistical software (version 1.4.1, R Foundation for Statistical Computing, Vienna, Austria). Statistical outcomes were considered significant at $P \leq 0.05$, and values are presented as mean ± SEM unless otherwise stated. The predictive behavioural models were built from the raw (i.e., 30 Hz) triaxial acceleration data (x, y, and z axes) using machine learning, as previously described by Smit et al. (2023). The three axes created unique activity patterns that could then be associated with each behaviour. The accuracy of the predicted behaviours was then compared against the observed behaviours.

From the triaxial acceleration data (30 Hz), 32 identifier variables were calculated and summarised into 1 s epochs (**Table 2.3**). The correlation between the three accelerometer axes (XY, XZ, YZ; three identifier variables), overall dynamic body acceleration (ODBA; one identifier variable), and vector magnitude (VM) was calculated as described in **Table 2.3**. For each acceleration axis (X, Y, and Z) and VM, the mean, sum, minimum (min), maximum (max), standard deviation (SD), skewness (skew), and kurtosis (Kurt) were calculated.

Table 2.2 Ethogram of defined dog behaviours categorised as either active, inactive, or maintenance

Category	Behaviour	Description
Active	Walking	The slowest upright gait where the body is moving forward, each paw lifting from the ground one at a time in a regular sequence (Koler-Matznick et al., 2005).
	Trotting	A rhythmic two-beat gait where diagonally opposite paws strike the ground at the same time as the subject moves forward. This gait is faster than walking (Koler-Matznick et al., 2005)
	Running	Can also be defined as a 'canter'. This is a three-beat gait in which two legs move separately and two as a diagonal pair. This gait is faster than a walk and trot (Koler-Matznick et al., 2005).
	Jumping	Subject has both hindlegs on the floor and rears in a manner that results in both forelegs in contact with the fencing of paddock, kennel, or person (Walker et al., 2016).
	Barking	Barking is defined as the mouth being opened and closed quickly in a snapping motion, releasing a low frequency vocalisation (Walker et al., 2016).
	Sniffing	Nose directed to a point of interest and sniffs (Lee et al., 2022).
	Digging	The dog uses its forepaws to repeatedly scratch the ground surface (Walker et al., 2016).
	Scratching	Grooming behaviour directed towards subjects' own body, using paw (Walker et al., 2016).
	Standing	All four paws planted on ground and legs extended so they are upright in stationary position (Walker et al., 2016).
Inactive	Lying (alert)	Lying on stomach with forelegs extended to the front, hind legs bent and resting close to the body on each side, or with the body twisted and both hind legs on one side. Head is held up off the ground or surface (Koler-Matznick et al., 2005).
	Lying (asleep)	Lying on stomach with forelegs extended to the front, hind legs bent and resting close to the body on each side, or with the body twisted and both hind legs on one side. Head is lowered to rest on either forelegs or the ground between them (Koler-Matznick et al., 2005).
	Lateral recumbency	Lying down flat on one side with head resting on surface in sideways position (Fukuzawa and Nakazato, 2015).
	Sitting	Hind quarters on ground with front legs standing up straight and being used for support (Walker et al., 2016).
Maintenance	Defecating	Excretion of faeces from the subject's body (Walker et al., 2016).
	Urinating	Excretion of urine from the subject's body (Walker et al., 2016).
	Eating	Subject chews and ingests food from bowl provided by human (Walker et al., 2016).
	Drinking	Subject drinks from water bowl in paddock by lapping up the water with their tongue (Koler-Matznick et al., 2005; Walker et al., 2016).
	Auto grooming	Grooming behaviour directed towards the subjects own body including licking, self-biting, and scratching (Walker et al., 2016).
Other	Other	Any behaviour that does not fit into one of the behaviours included in this ethogram.
	Out of sight	Subject is out of view and behaviour cannot be observed.

Table 2.3 Description of identifier variables

Identifier Variable	Description
Mean acceleration	Mean which is calculated for every second using the raw acceleration data (30 measures per second)
Sum acceleration	$\text{Sum}_{(\text{Axis})} = \sum \text{Axis}_i$
Minimum (min)	Minimum value of every 30 measures per second
Maximum (max)	Maximum value of every 30 measures per second
Standard deviation (SD)	Quantifies the amount of variability within a dataset
Skewness	Measures the asymmetry of the probability distribution of a dataset
Kurtosis	Measures the weight of the tails in relation to normal distribution
Vector magnitude (VM)	$VM = \sqrt{X^2 + Y^2 + Z^2}$
Overall dynamic body acceleration (ODBA)	$ODBA = \sum_{i=1}^{n} \lvert DBA_X \rvert + \lvert DBA_Y \rvert + \lvert DBA_Z \rvert$
Dynamic body acceleration (DBA)	$DBA = \text{Sum}_{\text{axis}} - \text{moving average}$

Before building the models, the behaviour categories ‘Other’ and ‘Out of sight’ were removed from the dataset as they did not represent a known behaviour. A form of machine learning (ML), called Random Forests (RF), was used to develop an algorithm to predict the behaviours of the dogs using the identifier variables calculated from the raw triaxial acceleration data. This method has been successfully used to build predictive behavioural models in other species (Kleanthous et al., 2020; Eyre et al., 2023; Smit et al., 2023). In short, RF is an ML technique that builds many decision trees and aggregates them to provide an accurate model for given classifications (Shaik and Srinivasan, 2019).

The RF models were built using the packages ‘caret’ and ‘randomForest’ (Smit et al., 2023). The default settings of 500 decision trees and the number of variables randomly sampled as candidates at each split ($n\sqrt{n_{variables}}$) were used. Using this method, five models were built from a subset of 70% of the complete dataset with varying levels of behavioural complexity (i.e., number of behaviours assessed) (**Figure 2.4**). Model performances were then tested using the remaining 30% of the complete data set.

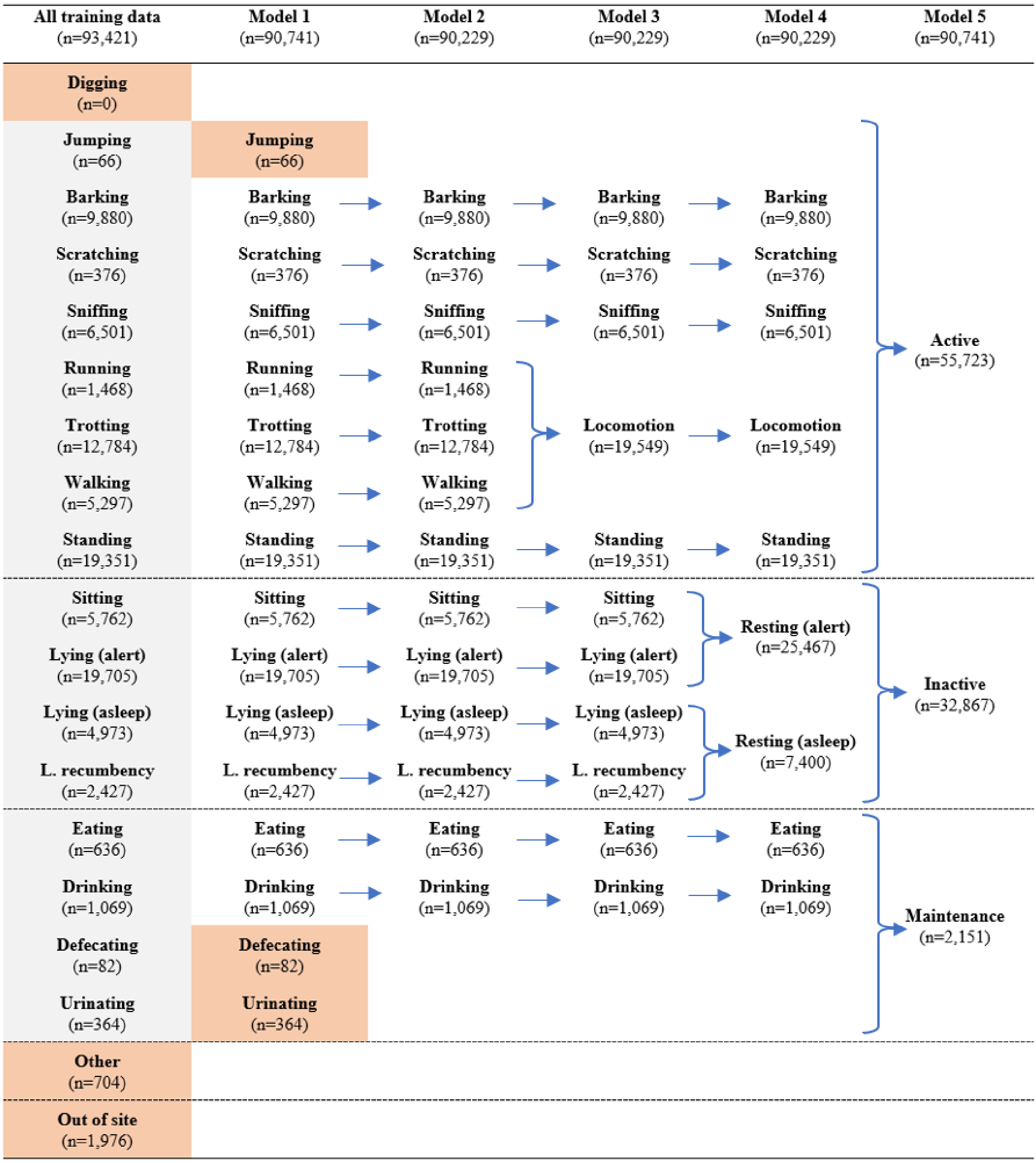

Figure 2.4 A description of the five modelling rounds and the total number of observation(s) of each behaviour used for the training data set. The total data set comprised 129,615 observations, excluding other and out-of-site categories. This data set used 90,741 (70%) and 38,874 observations (30%) to train and test the models, respectively. Abbreviations: Lateral (L.) Cells highlighted orange have been removed from the subsequent models due to low accuracy/precision and/or sample size.

2.2.7 Model evaluation

The performances of the five models were compared by constructing and comparing confusion matrices for each model (**Appendix 1**). From these confusion matrices, the number of observations(s) that were classified as true positive (TP, correctly identified by the model), true negative (TN, not observed and not identified by the model), false positive (FP, identified by the model but not observed) and false negative (FN, observed but not identified by the model) were determined for each behaviour. These data were then used to calculate the sensitivity/recall (the ability of the model to identify TP values), specificity (the ability of the model to identify TN), balanced accuracy positive predictive value or precision (accuracy of positive predictions), precision-recall/F1 score (an accuracy test that is particularly useful for imbalanced data sets; weighted mean of precision and recall), observed prevalence (actual rate of positive observations in the data set), detected prevalence (proportion of observations predicted to be positive) for all behaviours within each of the models (**Table 2.4**). The overall accuracy of each model was determined by calculating the overall accuracy and the Kappa coefficient (κ; **Table 2.4**). In addition, for each model, the average coefficient of variance (CV%) between the observed and detected prevalence of each behaviour was calculated.

Table 2.4 Calculations for the parameters used to assess the performance of the identifier variables.

Parameter	Calculation
Sensitivity/recall	= TP/(TP + FN)
Specificity	= TP/(TN + FP)
Balanced accuracy	= (*sensitivity + specificity*)/2
Precision	= TP/(TP+ FP)
Precision recall (F1 Score)	$= 2 \times ((Precision \times Sensitivity)/(Precision + Sensitivity))$
Observed prevalence	= (TP+FN)/(TP+TN+FP+FN)
Detected prevalence	= (TP+FP)/(TP+TN+FP+FN)
Overall accuracy	= (TP + TN)/(TP + TN + FP + FN)
Kappa coefficient (κ)	$= (N\ x\ \sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k} x_i(x_{i+} \times x_{+i}))/N^2 - \sum_{i=1}^{k}(x_{i+} \times x_{+i})$ where, N = Total number of observations (all behaviours) k = Number of behaviour categories i = Behaviour category i x_{ii} = Number of observations that both that both visual observation and the predictor model classified into the i-th category x_{i+} = Number of observations that were visually classified into the i-th category X_{+i} = Number of observations that the predictor model classified into the i-th category

Abbreviations: true positive (TP), true negative (TN), false positive (FP), false negative (FN).

The ActiGraph® accelerometers were validated to quantify overall physical activity by comparing the time spent active per hour (as determined by Model 4) with the corresponding sum of overall dynamic body acceleration (ODBA). These data were compared using a polynomial regression (2nd order). For each behaviour in Model 4 (barking, drinking, eating, locomotion, resting-alert, resting-asleep, scratching, sniffing, and standing), the average ODBA per s was determined. This was then compared, using CV%, against the overage ODBA per s for the same behavioural categories based on the observed behaviour.

2.3 Results

During the study period, all dogs maintained their body weight, and no adverse reactions were observed when they wore the devices. From the three days (24 h) of video footage collected from the six dogs, a total of 132,295 s (~36.7 h; ~6.1 h per dog) were scored for observed behaviour. Of the 18 behaviours included in the ethogram (**Table 2.2**), 'digging' was not observed in the video recordings and was removed from the model-building process (**Figure 2.4**). Additionally, periods over which the dog's behaviour was visually classified as 'other' (n=704 s) and 'out of sight' (n=1,976 s) were also removed from the data set. Thus, 90,741 and 38,874 s of scored behavioural data were used as the model training and testing, respectively. It has been previously shown that behaviours with large numbers of observations (e.g., >20,000 s) can result in overfitting of the models to these behaviours, thus reducing the model's overall performance (Smit et al., 2023). However, in the present study, model performance behaviours decreased when behaviours with large numbers of data points were randomly subsampled and limited to 7,000 s (data not presented). As such, all models were built using the complete training data set. The total amount of data used, and the number of behavioural categories differed depending on the model: Model 1 (16 behavioural categories; 90,741 s), Model 2 (13 behavioural categories; 90,229 s), Model 3 (12 behavioural categories; 90,229 s), Model 4 (nine behavioural categories; 90,229 s), and Model 5 (three behavioural categories; 90,741 s).

2.3.1 Modelling round 1 (16 behavioural categories)

Model 1 included 16 behavioural categories (**Figure 2.4**) and exhibited an overall accuracy of 0.69 and a κ coefficient of 0.64. While the average specificity was high (0.97 ± 0.01, range 0.86 – 1.00), the average sensitivity of this model was low compared to other models (0.60 ± 0.07, range 0.12 - 0.94). Model 1 had the greatest sensitivity and precision-recall for inactive behaviours (lateral recumbency, lying-alert, and lying-asleep) and sniffing (**Table 2.5**). However, the sensitivity, balanced accuracy, and/or precision-recall values were lower for jumping, running, sitting, standing, trotting, urinating, and walking (**Table 2.5**).

From the confusion matrix (**Appendix 1.1**), it was evident that the model often misclassified behaviours as standing, with standing being the main source of error for eight of the 15 other behaviours assessed. Indeed, the model misclassified 10.5% of drinking, 29.6% of jumping, 54.8% of running, 12.5% of scratching, 24.9% of sitting, 28.5% of trotting, 18.1% of urinating, and 33.6% of walking as standing behaviour. The model also misclassified jumping behaviour as barking (14.8% of observations) or trotting (31.2% of observations). Running was also miscategorised as barking by model 1 (17.4% of observations). In addition to standing, sitting behaviour was frequently misclassified as lying-alert (32% of observations). Standing and trotting behaviours were often confused. Urinating was misclassified as both sniffing (21.9% of observations) and standing (18.1% of observations). Lastly, walking behaviour was incorrectly categorised as sniffing (11.9% of observations), standing (33.6% of observations), and trotting (21.6% of observations). While there were other misclassifications of behaviour, these were <10% of observations and will not be discussed here (for more information, see **Appendix 1.1**).

Table 2.5 The performance characteristics of the five Random Forest models. The sensitivity reflects the proportion of true positives scored by the mode. The sensitivity reflects the proportion of true negatives correctly scored by the model. Balanced accuracy is the average of sensitivity and specificity. Observed and detected prevalence is the time the animals were observed (visually) or detected (by the model) exhibiting a given behaviour. *The coefficient of variance (CV%) between the observed and detected prevalence was calculated as CV%=SD/mean*100. Behaviours highlighted in red have a CV% >20. Cells highlighted orange reflect performance characteristics that scored <0.70.

Behaviour	Sensitivity	Specificity	Balanced accuracy	Precision	Precision-Recall	Prevalence (observed)	Detection prevalence	CV%*
Model 1								
Barking	0.84	0.97	0.91	0.80	0.82	0.109	0.115	3.77
Defecating	0.71	1.00	0.85	0.96	0.81	0.001	0.001	21.57
Drinking	0.71	1.00	0.85	0.92	0.80	0.012	0.009	18.75
Eating	0.78	1.00	0.89	0.96	0.86	0.007	0.006	14.95
Jumping	0.00	1.00	0.50	-	-	0.001	0.000	141.42
L. recumbency	0.94	1.00	0.97	0.98	0.96	0.027	0.026	2.83
Lying-asleep	0.85	1.00	0.92	0.91	0.88	0.055	0.051	4.67
Lying-alert	0.85	0.94	0.89	0.79	0.82	0.217	0.234	5.11
Running	0.12	1.00	0.56	0.54	0.20	0.016	0.004	88.67
Scratching	0.61	1.00	0.80	0.97	0.75	0.004	0.003	32.65
Sitting	0.37	0.99	0.68	0.63	0.47	0.064	0.037	36.54
Sniffing	0.93	0.98	0.95	0.75	0.83	0.072	0.089	15.06
Standing	0.64	0.86	0.75	0.55	0.59	0.213	0.251	11.57
Trotting	0.58	0.92	0.75	0.56	0.57	0.141	0.147	3.04
Urinating	0.48	1.00	0.74	1.00	0.65	0.004	0.002	49.20
Walking	0.25	0.99	0.62	0.54	0.34	0.058	0.027	52.88
Average	*0.60 ± 0.07*	*0.97 ± 0.01*	*0.79 ± 0.04*	*0.34 ± 0.08*	*0.79 ± 0.05*	*∑ = 1.000*	*∑ = 1.000*	*31.4 ± 9.4*
Model 2								
Barking	0.84	0.97	0.90	0.80	0.82	0.110	0.115	3.25
Drinking	0.71	1.00	0.86	0.91	0.80	0.012	0.009	17.18
Eating	0.79	1.00	0.90	0.93	0.86	0.007	0.006	11.23
L. Recumbency	0.94	1.00	0.97	0.98	0.96	0.027	0.026	3.12
Lying-asleep	0.85	1.00	0.92	0.92	0.88	0.055	0.051	5.27
Lying-alert	0.85	0.93	0.89	0.78	0.81	0.218	0.238	6.07
Running	0.12	1.00	0.56	0.56	0.20	0.016	0.003	91.69
Scratching	0.72	1.00	0.86	0.95	0.82	0.004	0.003	19.63
Sitting	0.33	0.98	0.66	0.58	0.42	0.064	0.037	37.87
Sniffing	0.94	0.98	0.96	0.77	0.85	0.072	0.087	13.42
Standing	0.64	0.86	0.75	0.55	0.59	0.215	0.248	10.13
Trotting	0.59	0.92	0.76	0.55	0.57	0.142	0.152	5.02
Walking	0.24	0.99	0.61	0.56	0.34	0.059	0.025	57.21
Average	*0.66 ± 0.07*	*0.97 ± 0.01*	*0.81 ± 0.04*	*0.76 ± 0.05*	*0.69 ± 0.07*	*∑ = 1.000*	*∑ = 1.000*	*21.6 ± 7.3*
Model 3								
Barking	0.82	0.98	0.90	0.83	0.83	0.110	0.109	0.49
Drinking	0.71	1.00	0.86	0.93	0.81	0.012	0.009	18.95
Eating	0.79	1.00	0.90	0.97	0.87	0.007	0.006	14.32
L. Recumbency	0.94	1.00	0.97	0.99	0.96	0.027	0.025	4.11
Locomotion	0.67	0.88	0.77	0.61	0.64	0.217	0.236	6.09
Lying-asleep	0.87	1.00	0.93	0.92	0.89	0.055	0.052	4.13
Lying-alert	0.84	0.94	0.89	0.79	0.81	0.218	0.233	4.57
Scratching	0.64	1.00	0.82	0.94	0.77	0.004	0.003	26.83
Sitting	0.35	0.98	0.67	0.60	0.44	0.064	0.037	37.22
Sniffing	0.93	0.98	0.95	0.79	0.85	0.072	0.084	11.20
Standing	0.57	0.89	0.73	0.59	0.58	0.215	0.205	3.06
Average	*0.74 ± 0.05*	*0.96 ± 0.01*	*0.85 ± 0.03*	*0.82 ± 0.05*	*0.77 ± 0.05*	*∑ = 1.000*	*∑ = 1.000*	*11.9 ± 3.5*
Model 4								
Barking	0.82	0.98	0.90	0.83	0.82	0.110	0.108	0.70
Drinking	0.73	1.00	0.86	0.93	0.82	0.012	0.009	17.37
Eating	0.72	1.00	0.86	0.96	0.82	0.007	0.005	19.87
Locomotion	0.66	0.88	0.77	0.61	0.64	0.217	0.234	5.37
Rest-asleep	0.89	1.00	0.94	0.94	0.92	0.082	0.077	4.11
Rest-alert	0.85	0.94	0.89	0.85	0.85	0.282	0.283	0.08
Scratching	0.66	1.00	0.83	0.96	0.79	0.004	0.003	26.19
Sniffing	0.93	0.98	0.95	0.79	0.85	0.072	0.085	11.51
Standing	0.54	0.90	0.72	0.59	0.57	0.215	0.196	6.52
Average	*0.76 ± 0.04*	*0.96 ± 0.02*	*0.86 ± 0.03*	*0.83 ± 0.05*	*0.78 ± 0.04*	*∑ = 1.000*	*∑ = 1.000*	*10.2 ± 3.0*
Model 5								
Active	0.95	0.86	0.9053	0.91	0.93	0.613	0.638	2.9
Inactive	0.87	0.95	0.9095	0.91	0.89	0.364	0.345	3.8
Maintenance	0.71	1.00	0.85618	0.98	0.83	0.023	0.017	22.4
Average	*0.84 ± 0.07*	*0.93 ± 0.04*	*0.86 ± 0.04*	*0.98 ± 0.02*	*0.88 ± 0.03*	*∑ = 1.000*	*∑ = 1.000*	*9.7 ± 6.4*

In general, the model struggled to accurately categorise behaviours that were less frequently expressed by the dogs (i.e., prevalence <0.5) (**Table 2.5**). The average coefficient of variance between the observed prevalence and the detected prevalence of each of the behaviours was 31.4 ± 9.4% (3.0 – 141.4%). The following behaviours had >20% variance between observed and detected prevalence: defecating, jumping, running, scratching, sitting, urinating, and walking (**Table 2.5**).

2.3.2 Modelling round 2 (13 behavioural categories)

For Model 2, three behaviours that were exhibited infrequently and/or had a poor sensitivity in Model 1 were removed: defecating, jumping, and urinating. Thus, Model 2 assessed a total of 13 behavioural categories. In general, the performance of Model 2 was similar to Model 1, with an overall accuracy of 0.69 and a κ coefficient of 0.64. The average sensitivity and specificity for Model 2 were 0.66 ± 0.07 (range, 0.24 – 0.94) and 0.97 ± 0.01 (range, 0.86 - 1.00), respectively (**Table 2.5**). This led to a slightly higher average balanced accuracy (0.81 ± 0.04) for Model 2 than Model 1. Surprisingly, the average precision-recall of Model 2 (0.69 ± 0.07) as similar to that of Model 1 (0.69 ± 0.05). Model 2 had a precision-recall of <0.70 for running behaviour (0.20), sitting (0.42), standing (0.59), trotting (0.57), and walking (0.34).

The confusion matrix for Model 2 (**Appendix 1.2**) showed that standing behaviour was once again the leading cause for the misclassification of behaviour. Model 2 misclassified 55.1%, 13.8%, 23.9%, 27.9%, and 33.2% of observed running, scratching, sitting, trotting, and walking behaviour as standing, respectively. Sitting behaviour was also misclassified as lying-alert; 35.7% of observations) and drinking and standing were frequently misclassified as trotting by the model (12.7% and 19.0% of observations, respectively). In addition, running was often miscategorised as barking (15.0% of observations) and sitting was miscategorised as lying-alert (35.7% of observations; **Appendix 1.2**). Despite this, Model 2 had a much higher balanced accuracy for the less frequent behaviours than Model 1 (**Table 2.5**). The average variance between the observed and detected prevalence of the behaviours was lower than in Model 1, with a mean CV% of 21.6 ± 7.3% (range, 3.1 – 91.7%) for Model 2. The difference between the observed and detected prevalence was greatest for running, sitting, and walking behaviour (**Table 2.5**).

2.3.3 Modelling round 3 (11 behavioural categories)

As locomotive behaviours (walking, trotting, running) were one of the leading sources of misclassification in Model 2, these behaviours were combined and categorised as locomotion for Model 3. Therefore, Model 3 assessed a total of 11 behavioural categories. Model 3 had an overall accuracy of 0.72 and a κ coefficient of 0.66, which was better than Models 1 and 2. The average sensitivity was 0.74 ± 0.05 (range, 0.35 – 0.94) and the average specificity was 0.96 ± 0.01 (range, 0.89 – 1.00). While the average specificity for Model 3 was similar to that of Models 1 and 2, the average sensitivity was much higher (**Table 2.5**). Thus, the average balanced accuracy of Model 3 was also higher (0.85 ± 0.05, range 0.67 – 0.97) than both Models 1 and Model 2 (**Table 2.5**). In fact, sitting was the only behaviour for which the model had poor sensitivity (0.35) and balanced accuracy of less than 0.7 (**Table 2.5**). The precision-recall of Model 3, however, was less than 0.7 for the locomotion (0.64), sitting (0.44), and standing (0.58) categories.

It was evident from the confusion matrix (**Appendix 1.3**) that Model 3 tended to miscategorise sitting behaviour as lying alert or standing (33.5% and 23.6% of observations, respectively). Interestingly, Model 3 incorrectly classified behaviours as standing less frequently than Models 1 and 2. However, observations of locomotion and sitting behaviours were sometimes recorded as standing by the model (21.0% and 23.6% of observations, respectively). The model also miscategorised observations of standing behaviour as locomotion (31.1% of observations).

Despite this, this model's observed and detected prevalences for each of the behaviours were in agreement, with an average CV% of 11.9 ± 3.5% (range, 0.5 – 37.2%).

2.3.4 Modelling round 4 (nine behavioural categories)

Since Model 3 had the lowest accuracy for sitting behaviour and was often misclassified as lying-alert, these behaviours were combined and categorised as "resting-alert" for Model 4. In addition, lateral recumbency and lying-resting were combined to form the category "resting-asleep". Model 4, therefore, considered a total of nine behavioural categories. The overall accuracy of Model 4 was 0.74, and the κ coefficient was 0.68; these were higher than all previous models. The average sensitivity for Model 4 (0.76 ± 0.04, range 0.54 – 0.93) was similar to Model 1 but higher than Models 2 and 3 (**Table 2.5**). The specificity for Model 4 (0.96 ± 0.02, range 0.88 - 1.00) was similar to Models 1, 2, and 3. This ultimately meant that Model 4 had a balanced accuracy and precision-recall comparable to Model 3 but higher than Models 1 and 2 (**Table 2.5**). In terms of the variation between the observed and detected prevalences, Model 4 had a lower average CV% (10.2 ± 3.0%, range 0 – 26.6%) than all previous models, with only scratching behaviour having greater than 20% variation between observed and detected prevalences.

The confusion matrix for Model 4 (**Table 2.6**) showed that the standing and locomotion behaviours were often misclassified, with observed standing behaviour being classified as locomotion (31.1% of observations) and vice versa (20.7% of observations). Barking, drinking, and scratching behaviour were also occasionally misclassified as locomotion (9.5%, 16.9%, and 11.9% of observations, respectively; **Table 2.6**). The model also recorded eating behaviour as sniffing 12.5% of the time (**Table 2.6**). Overall, however, this model accurately distinguished between the assessed behaviours, reflected by the low CV% and **Figure 2.5**, the raw triaxial acceleration data for each behaviour identified using this model.

Table 2.6 The confusion matrix of predicted and observed observations (s) from Model 4 is presented as percentages (%). Correct categorisations by the model are indicated in cells highlighted green, and incorrect categorisations >10% are in cells that have been highlighted red.

Model prediction	Observed behaviour								
	Barking	**Drinking**	**Eating**	**Locomotion**	**Resting-asleep**	**Resting-alert**	**Scratching**	**Sniffing**	**Standing**
Barking	81.77	0.44	2.94	4.20	0.19	1.24	0.00	0.07	2.71
Drinking	0.00	72.65	0.00	0.18	0.00	0.00	0.00	0.00	0.12
Eating	0.00	0.00	72.06	0.02	0.00	0.01	0.00	0.14	0.02
Locomotion	9.59	16.85	5.88	66.22	0.41	2.59	11.88	3.59	31.08
Resting-asleep	0.09	0.00	0.00	0.20	89.06	1.11	1.88	0.22	0.21
Resting-alert	4.53	0.44	2.21	4.81	8.01	84.89	8.75	0.72	9.29
Scratching	0.00	0.00	0.00	0.01	0.03	0.02	66.25	0.00	0.00
Sniffing	0.09	4.81	12.50	3.69	0.57	0.82	2.50	92.89	2.54
Standing	3.92	4.81	4.41	20.66	1.73	9.33	8.75	2.37	54.03
Total observations (s)	**4,234**	**457**	**272**	**8,377**	**3,171**	**10,914**	**160**	**2,785**	**8,292**

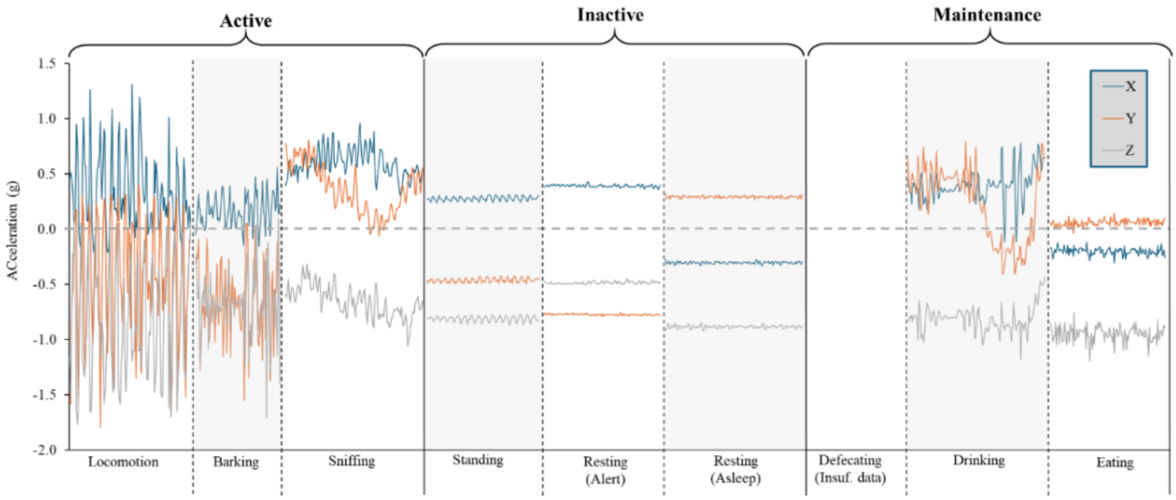

Figure 2.5 Raw (30 Hz) triaxial (x-axis = blue line, y-axis = orange line, z-axis = grey line) acceleration profiles for each of the behaviours classified by Model 4. A total of 3 s to 5 s were present per behaviour, although insufficient continuous acceleration data (Insuf. data) were available for defecation behaviour. These behaviours have been grouped according to the categories used for Model 5: Active, inactive, and maintenance.

2.3.5 Modelling round 5 (three behavioural categories)

The final model (Model 5) was simplified with three behavioural categories: active, inactive, and maintenance. This model had by far the highest overall accuracy (0.92) and a κ coefficient (0.82) of all the tested Models. While the average sensitivity (0.84 ± 0.07, range 0.71 – 0.95) was the highest of all the models, the average specificity (0.93 ± 0.04, range 0.86 – 1.0) was the lowest. As a result, the balanced accuracies of Models 5, 4, and 3 were similar but better than Models 1 and 2 (**Table 2.5**). Interestingly, the average precision and precision-recall were much higher for Model 5 (0.98 ± 0.02 and 0.88 ± 0.03, respectively) than all other models (Table 4). Despite this, the average CV% between observed and detected prevalences (9.7 ± 6.4%, 2.9 – 22.4) was similar to Model 4 (**Table 2.5**). This was probably because Model 5 often miscategorised inactive and maintenance behaviours as active (13.4% and 26.3% of observations, respectively). However, Model 5 could accurately distinguish between the maintenance and inactive categories (**Table 2.7**).

Table 2.7 The confusion matrix of predicted and observed behaviours for Model 5 is presented as percentages (%). Correct categorisations by the model are indicated in cells highlighted green, and incorrect categorisations >10% are in cells that have been highlighted red.

Model Predictions	Observed behaviour		
	Active	Inactive	Maintenance
Active	95.22	13.40	26.26
Inactive	4.75	86.57	2.47
Maintenance	0.03	0.04	71.27
Total Observations (s)	23,690	14,085	891

2.3.6 Overall physical activity/overall dynamic body acceleration

The hourly ODBA amount of time spent on active behaviours (barking, locomotion, scratching, and sniffing), as determined by Model 4, were strongly correlated (R^2 = 0.91, *P*<0.001: **Figure 2.6**). Overall, total ODBA was a significant predictor of the amount of time spent active per hour (*P*<0.001) and, thus, overall physical activity. Indeed, active behaviours such as barking and locomotion were associated with the highest average ODBA per second based on both observed behavioural data and the detected outputs of Model 4 (**Table 2.8**). It was interesting to note that the maintenance behaviours of drinking and eating had higher overall ODBA counts per second than sniffing and standing.

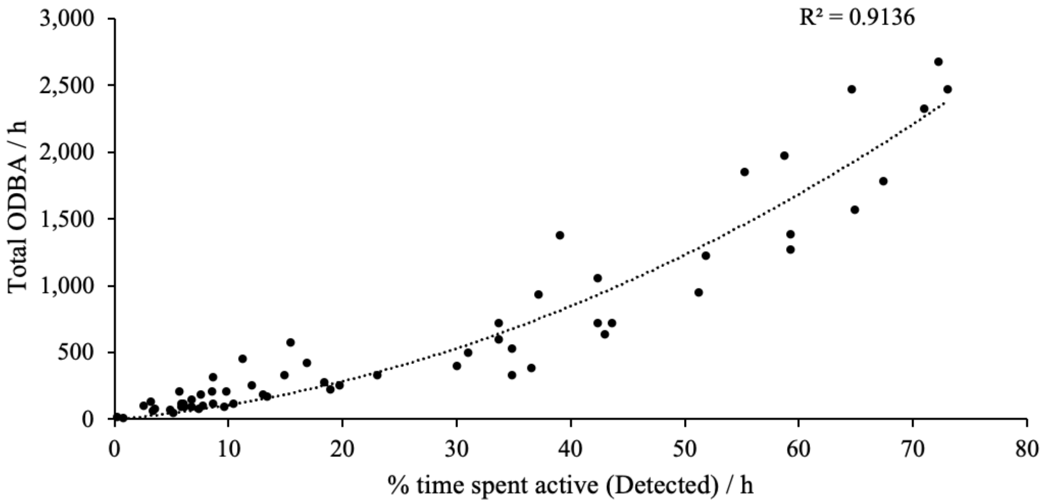

Figure 2.6 Graph showing the correlation between the time spent active (%) per hour (h) and the total ODBA per hour (h).

Table 2.8 shows the observed and detected average ODBA/s and the coefficient of variance (CV) for each behaviour. Behaviours are selected from Model 4 (9 behavioural categories).

Behaviour	Average ODBA / s (Observed)	Average ODBA / s (Detected)	CV%
Barking	0.726 ± 0.015	0.707 ± 0.014	1.9
Drinking	0.573 ± 0.032	0.659 ± 0.047	9.9
Eating	0.514 ± 0.031	0.427 ± 0.031	13.1
Locomotion	0.819 ± 0.007	0.996 ± 0.006	13.8
Resting-alert	0.069 ± 0.001	0.047 ± 0.001	26.8
Resting-asleep	0.020 ± 0.001	0.014 ± 0.000	25.3
Scratching	0.372 ± 0.057	0.465 ± 0.075	15.7
Sniffing	0.364 ± 0.008	0.357 ± 0.006	1.4
Standing	0.454 ± 0.006	0.277 ± 0.004	34.2
Average	0.435 ± 0.089	0.439 ± 0.105	15.7 ± 3.7

2.4 Discussion

This study aimed to validate the use of ActiGraph® accelerometers as a tool for remotely classifying behaviour in domestic dogs. Five RF models were built, and their performance characteristics were compared. Based on the performance characteristics, confusion matrix, and comparatively low variance between the observed and detected prevalences, Model 4 was determined to be the optimal model with an overall accuracy of 74%. This model evaluated nine behavioural categories: barking, defecating, drinking, eating, locomotion, resting-asleep, resting-alert, sniffing, and standing. These findings align with other collar-mounted behavioural classification models derived from accelerometer data, where 75% overall accuracy was observed across seven behavioural states (Kumpulainen et al., 2022).

Additionally, Ladha et al. (2013) reported that their overall model accuracy was approximately 70% when investigating 17 behaviours in domestic dogs. A study by den Uijl et al. (2017) developed a model for 8 behavioural states with far higher overall accuracy of the model of over 95%. However, the model developed by den Uijl et al. (2017) was created using accelerometer data from 51 dogs compared to just six dogs in the current study. This may have contributed to the higher overall accuracy of the model performance (den Uijl et al., 2017).

ActiGraph® devices have previously been validated to assess overall physical activity (Yam et al., 2011; Morrison et al., 2013; Helm et al., 2016; Ortmeyer et al., 2018). The present study also validated the use of ActiGraph® for monitoring dogs' overall physical activity, in the form of ODBA. While using accelerometery data to assess the expression of specific behaviours has yielded more variable results, the availability of ML techniques has greatly enhanced the potential of building behavioural models for acceleration data (Hounslow et al., 2019; Kumpulainen et al., 2021).

However, modelling a wide range of behaviours remains challenging, especially if they have a similar acceleration profile (Martiskainen et al., 2009). In this context, increasing the number of behavioural classes in the model has generally corresponded with decreasing overall accuracy in companion animals (den Uijl et al., 2017; Smit et al., 2023). In the present study, there was a progressive decrease in the overall accuracies and κ coefficients going from Models 4 to 1 (i.e., from nine to 17 behavioural categories). At the same time, the variability between the observed and detected prevalences progressively increased. Consolidating similar behaviours during model development has also improved overall accuracy for domestic cat behaviour (Smit et al., 2023). For my study, merging various locomotor gaits (walking, trotting, and running) into a single locomotion category improved the performance of Model 3 compared to Model 2. Similarly, consolidating observed resting behaviours (lateral recumbency, resting–asleep, sitting, and resting–alert) as resting-alert or resting-asleep further improved model performance from Model 3 to Model 4. den Uijl et al. (2017) also found that model sensitivity and specificity increased when resting behaviours were broadly categorised as either ‘sleep’ or ‘static’. While the combination of locomotor gaits and resting behaviours into broader categories improved model performance.

Standing was a consistently challenging behaviour to model accurately, with the misclassification of observations into standing behaviour being one of the leading causes of error in Models 1 to 4 (**Appendix 1.1-1.3**; **Table 2.5**). Standing had the lowest sensitivity and balanced accuracy of the behaviour categories classified by the optimal model (Model 4), followed by locomotion. Differentiating between the locomotion gaits and standing was a problem for Models 1 and 2. This was most likely due to the similarities in the device's position with respect to gravity for all of the active behaviours (Tatler et al., 2018). In addition, standing is often an intermittent behaviour interspersed with locomotion. Thus, these behaviours are frequently seen within close chronological proximity (Walker et al., 2016). Panting, which usually occurs with standing behaviour, may also cause movement in the dog's body, especially in the head area, resulting in unwanted motion (Gerencsér et al., 2013).

In model 1, many observed non-locomotor behaviours were frequently misclassified as standing, including barking, defecating, drinking, eating, sniffing, and urinating. While this is not ideal from a model-building perspective, it is logical as these behaviours are exhibited while the dog is standing (Walker et al., 2016). This perhaps illustrates the main limitation of the RF approach for modelling behaviour, that is, all modelled behaviours are considered to be independent and mutually exclusive (Martiskainen et al., 2009). From a modelling perspective, this means that the RF model can only assign a single behavioural classification for a given time point (Martiskainen et al., 2009). In reality, many of the behaviours that dogs exhibit, occur simultaneously (e.g., barking and walking, barking and standing, drinking and standing, or eating and standing) which would adversely affect the performance of the model for these behaviours. Thus, we are confronted with a trade-off: either we create an excessive number of behaviour categories, which lowers the model's accuracy and misclassifies closely related behaviours, or we need more categories to represent the full range of the species behaviours adequately.

In the initial models (Models 1 to 3), many static or resting behaviours (e.g., standing, lying–alert, and lying-asleep) were also misclassified despite being considered mutually exclusive. For example, sitting often faced misclassifications with other static behaviours such as lying-alert and standing. Kumpulainen et al. (2021) also noted challenges differentiating static postures such as lying down, sitting, and standing from triaxial acceleration data. It has been hypothesised that the minimal neck and back orientation changes during these behaviours might be too subtle for the activity device to distinguish (den Uijl et al., 2017; Kumpulainen et al., 2021). Alternatively, unwanted rotation of the device and collar around the neck could also be problematic for distinguishing between static behaviours.

A previous study in cats has shown that RF and self-organising map (SOM) models built for collar-mounted devices generally performed worse than those fitted to a harness. However, both the final collar and harness models were considered satisfactory by the authors (Smit et al., 2023). Many authors have attributed this lower model performance to an increase in the residual movement of the devices (i.e., continued movement after a behaviour has stopped) and/or the rotation of the devices around the collar, leading to changes in device orientation (Martin et al., 2016; Westgarth and Ladha, 2017; Kumpulainen et al., 2021; Smit et al., 2023). Westgarth and Ladha (2017) and Smit et al. (2023) stated that harness-attached accelerometers might be advantageous due to their inability to rotate. However, many dogs are unaccustomed to or unwilling to wear harnesses over a long period (Westgarth and Ladha, 2017). Given this and the practical simplicity of collar attachment, many pet owners would likely prefer a collar-attached device. Indeed, a study on cats gave owners the choice of harness or collar attachment of ActiGraph® devices; the majority selected collar attachment (Smit et al., 2024). Whether this remains to be the case for dogs remains to be investigated.

Regardless of the attachment method, RF can accurately assess the behaviour of animals from continuous triaxial acceleration data, but caution is needed when constructing the models to determine an appropriate balance between detail (number of behaviours) and performance (Smit et al., 2023; Smit, 2024). Thus, progressively simplifying the RF model over several modelling rounds, as done in the present study, is an essential aspect of building behavioural algorithms for acceleration data. While many behaviours were misclassified in the initial models of the present study (especially Model 1), the optimal model (Model 4) showed a high level of accuracy and precision.

The performance characteristics obtained from the confusion matrices provide useful insight during the model-building process. However, from a practical perspective, the critical feature of a good model is that it accurately predicts the percentage of time an animal spends exhibiting each behaviour over a given time point. Indeed, previous studies utilising these models to assess factors affecting animal behaviour have concentrated on the percentage of time spent exhibiting each behaviour per hour, day, or week (Smit et al., 2024; Liu et al., 2024). My final model, Model 4, showed a high agreement between the observed prevalence (proportion of time spent exhibiting each behaviour based on actual observations) and detection prevalence (the proportion of time spent exhibiting each behaviour based on the model classifications) of most behaviours. Thus, this model provided the most accurate and reliable dog behaviour assessment from the triaxial acceleration data.

2.5 Conclusions

This study successfully validated the use of ActiGraph® wGT3X-BT accelerometers as tools for remotely classifying behaviours in domestic dogs. After using machine learning (ML) to build a number of Randomforest (RF) models, the optimal model encompassed nine behavioural classification categories (barking, defecating, drinking, eating, locomotion, resting-asleep, resting-alert, sniffing, and standing) and demonstrated a high overall accuracy whilst maintaining a sufficiently large behavioural repertoire.

The results support previous validations of ActiGraph® devices for measuring overall physical activity in animals. This study also expanded the validation to specific behaviours using machine learning techniques to improve model accuracy. Despite the difficulty of modelling a variety of behaviours with similar acceleration patterns, grouping similar behaviours was found to be a practical approach.

The most challenging behaviour to classify was standing, which frequently led to inaccuracies in the model. This problem is commonly seen in studies using accelerometers and highlights the difficulty of distinguishing between similar behaviours. However, my models showed significant improvement over time, and Model 4 achieved a high level of accuracy and precision.

The use of ActiGraph® accelerometers combined with refined RF models offers a promising method for detailed and remote assessment of canine behaviour. Although challenges remain in accurately classifying similar behaviours, advancement holds the potential to improve our understanding of animal behaviour and enhance the welfare of domestic dogs through better monitoring and analysis.

Chapter 3

Evaluating the efficacy and duration of enrichment types on domestic dogs (*Canis familiaris*): A behavioural analysis using tri-axial accelerometers

Chapter 3: Evaluating the efficacy and duration of enrichment types on domestic dogs (*Canis familiaris*): A behavioural analysis using tri-axial accelerometers

3.0 Abstract

It is known that environmental enrichment is associated with improving welfare, reducing stress, and encouraging natural behaviours in domestic dogs (Canis familiaris), especially in confined environments such as kennels, shelters, and research facilities. However, the current methods for measuring enrichment success can be subjective and time-consuming, providing variable results. Therefore, this study aimed to evaluate the potential for tri-axial accelerometery and a validated random forests (RF) model as a method for determining the efficacy and duration of different environmental enrichment treatments for colony-housed dogs. The study also aimed to assess the effect of the various enrichment treatments (food, olfactory, and tactile) on the behaviour and activity of colony-housed dogs. Using accelerometer data and a validated machine learning model, behaviours including barking, defecating, drinking, eating, locomotion, resting-asleep, and resting-alert were quantified for six domestic dogs exposed to food, olfactory and tactile enrichment treatments. Daily behavioural analysis of the accelerometer data revealed differences in behaviour and interaction duration between enrichment treatment groups. Dogs were more active in the food enrichment treatment than in the olfactory enrichment treatment ($P<0.05$). Proportional ODBA was significantly lower in dogs exposed to the olfactory enrichment treatment compared to the food ($P<0.001$) and tactile ($P<0.001$) enrichment treatments. Dogs displayed less locomotory behaviour after exposure to the olfactory enrichment treatment compared to the food ($P<0.05$), tactile enrichment treatment ($P<0.05$), and baseline treatment ($P<0.001$). Dogs exhibited higher amounts of standing behaviour in the baseline treatment compared to the food enrichment treatment ($P<0.001$) and the olfactory enrichment treatment ($P<0.05$).

Dogs spent less time resting (asleep) in the baseline treatment compared to the food ($P<0.05$), tactile ($P<0.05$) and olfactory ($P<0.05$) enrichment treatments. Daily barking and resting (alert) behaviours were not significantly different by the treatment group. The time that the dogs spent interacting with enrichment items (seconds) was significantly different between treatments ($P<0.001$), with the food treatment having the highest daily mean interaction time (1068.98s), followed by the olfactory treatment (204.59s) and then the tactile treatment (21.52s). Other variables, such as 'dog', 'hour', and 'day', also were investigated between treatments, showing significant effects. The study indicates that implementing environmental enrichment treatments significantly impacts dog behaviour, providing valuable insights for future research and practical applications.

3.1 Introduction

The purpose of environmental enrichment (EE) is to enhance the quality of life of domestic and captive animals by providing additional temporary stimuli and activities to an animal's external environment (Desforges, 2021; Hunt et al., 2022). For EE to be considered enriching, it must enhance an animal's overall welfare state (Hunt et al., 2022). The benefits of EE include an increase in the frequency and diversity of species-specific behavioural repertoires, as well as a reduction in stress and anxiety, abnormal behaviours, and improved cognitive abilities (Hubrecht, 1993; Shepherdson, 2003; Wells, 2004; Schipper et al., 2008; Herron et al., 2014; Hunt et al., 2022; Kang, 2022).

Food-based enrichments are a popular way to engage domestic dogs, but their effectiveness can vary (Gaines et al., 2008; Schipper et al., 2008; Herron et al., 2014; Hunt et al., 2022). Food-based enrichment usually involves scattering food around the animal's environment or hiding food in toys or puzzle feeders (Markowitz, 1982; Newberry, 1995; Young, 1997). This encourages natural feeding behaviours that they may not be typically displayed in captive environments (Schipper et al., 2008).

Sensory enrichment, including auditory, olfactory, and visual stimulation, is another practical strategy due to its low cost, simplicity, and effectiveness in enriching domestic dogs (Desforges, 2021). Typical approaches to olfactory enrichment include scent work and scent play, where dogs use their olfactory system to find designated toys, food, or people (Desforges, 2021). The reported benefit of this strategy is that scent work is tiring for dogs due to their highly developed olfactory system, making it a relaxing and mentally stimulating activity for anxious and excitable dogs while promoting natural behaviour (Desforges, 2021). Odours such as essential oils, prey urine, and dog-appealing pheromones have all been used as effective olfactory enrichment treatments, resulting in decreased stress levels, increased overall activity, and reduced barking (Wells, 2004; Tod et al., 2005; Wells, 2006; 2009).

Tactile enrichment, such as toys and blankets, is one of the most frequently used enrichment methods for both dogs and other domestic and exotic captive animals (Wells, 2004; Dare and Strasser, 2023). Providing toys for captive and domestic animals has been reported to reduce boredom, encourage natural play behaviours, promote exploration, and reduce abnormal behaviours (Wells, 2004). However, the effectiveness of food-based toys and games in domestic dogs has also yielded mixed results (Gaines et al., 2008; Schipper et al., 2008; Hunt et al., 2022).

It appears that the variability in efficacy seen in domestic dog enrichment studies may be attributed to the absence of objective and continuous measures for quantifying the effectiveness of the enrichment. In recent years, accelerometers have become increasingly important for behavioural assessment in enrichment studies. Tri-axial accelerometers have effectively measured the success of enrichment activities in various animal species, including pigs, rats, cattle, and chickens (Sudo et al., 2018; Liu, 2019; Bruno et al., 2020; Pullin et al., 2020; Veldkamp et al., 2023). In zoo settings, researchers have also started to use automated technologies such as accelerometers to evaluate animal responses to their surroundings and assess their welfare (Whitham and Miller, 2016). However, there has been limited research on the effects of environmental enrichment specifically on dogs. Nevertheless, accelerometers have previously been used to measure stress levels, behaviour, and welfare of dogs (Jones et al., 2014; den Uijl et al., 2017; Kumpulainen et al., 2021; Hussain et al., 2022).

Therefore, the first aim of this study was to evaluate the potential for tri-axial accelerometery and a validated random forests model for determining the efficacy of different environmental enrichment strategies in colony-housed dogs. The second aim of the study was to determine the effect of food, olfactory, and tactile enrichment on the behaviour and activity of colony-housed dogs.

3.2 Methods

This study was conducted at Massey University Canine Nutrition Unit (CNU), Palmerston North, New Zealand (latitude 40°230'S, longitude 175°365'E) from January to February 2024. The Canine Research Unit, a purpose-built colony facility, housed 29 domestic dogs (10 female and 19 male) at the time of the study. All research was approved and conducted in accordance with Massey University Animal Ethics Committee (MUAEC) protocol number 23/63. All husbandry of the dogs complied with the Code of Welfare: Dogs (Ministry for Primary Industries, 2018).

3.2.1 Animal husbandry

Six healthy desexed domestic dogs (*Canis familiaris;* two female and four male) were used for this study (**Table 3.1)**. The dogs were aged from 4.3 - 7.9 years (mean ± SD, 6.02 ± 1.59 years) and weighed 21.5 – 32.1 kg (mean ± SD, 25.6 ± 4.19 kg). Dogs were routinely managed in outdoor exercise paddocks during daylight hours (07:00 h to 16:00 h); therefore, they were familiar with the paddocks used for the research and spent the night in centrally heated indoor runs (16:00 h – 07:00 h). Dogs were normally kept in specific pairings and they continued to be housed in these pairs during the study. The dogs were fed a complete and balanced diet of Black Hawk Working Dog Adult formula (Black Hawk, Lower Hutt, New Zealand) once a day in the morning according to maintenance energy requirements (MER), which was calculated as: MER (kj) = 552 x kg BW$^{0.75}$) (AAFCO, 2024). The dogs had *ad libitum* access to water in their outdoor paddocks and indoor pens.

Table 3.1 The name, sex, age, breed, weight and pair number of the six domesticated dogs enrolled in the study.

Name	Sex	Age (years)	Breed	Weight (kg)	Pair
Belvedere	Female	7.9	Huntaway	23.4	3
Blacky	Male	4.3	Huntaway/Heading	23.6	3
Chevelle	Female	7.9	Huntaway	21.5	2
Gizmo	Male	6.1	Harrier Hound	30.8	1
Gus	Male	4.4	Huntaway/Smithfield Terrier	22.4	2
Monaro	Male	7.9	Huntaway	32.1	1

3.2.2 Enrichment treatments

The dogs in the study were exposed to three different types of enrichment: food, tactile and olfactory. For both the food and tactile enrichment treatments, two of each enrichment items were provided simultaneously to reduce inter-animal aggression over them. The food enrichment type provided in the current study was an ice block comprised of 180 grams of Pedigree Beef and Gravy Casserole (Pedigree, MARS, Auckland, New Zealand) frozen in one litre of water (**Figure 3.1**). The ice blocks were frozen in silicon moulds (27.6 x 14 x 6.5cm) and stored in a freezer at -18 °C before being offered to the dogs. Two ice blocks were placed on opposite sides of the designated enrichment paddock for each enrichment day. The tactile enrichment (**Figure 3.1**) utilised a long rope tug toy with dimensions 189 cm x 8 cm x 10.5 cm (Pet Toy Knotted Mega Rope, Anko®, Kmart Ltd. Australia). Two rope tug toys were placed on opposite sides in the designated enrichment paddock during each day of data collection. The olfactory enrichment **(Figure 3.1)** consisted of chicken litter and shavings sourced from the indoor housing facilities at the Massey University Poultry Research Unit. During each day of data collection, 400 grams of chicken litter was placed in each of the four corners of the designated enrichment paddock. This remained in the paddock for the day before being removed at the end of each day of data collection, and replaced with fresh litter the following day.

Figure 3. 1 (a) The food enrichment item was an ice block containing beef casserole, **(b)** A rope tug toy was used for the tactile enrichment, and **(c)** chicken litter was used for olfactory enrichment.

3.2.3 Study design

The study comprised a 23-day experimental design, with an initial one-day habituation phase followed by three consecutive five-day blocks of data collection (15 days of total data collection) for each pair of dogs. Each block of data collection was separated by a four-day rest period where no data were collected (see Figure 3.2). All data collected for this study occurred within this designated timeframe from January 22nd to February 12th, 2024. Three individual observation paddocks were utilised in the study, all maintained under the same experimental conditions. Each pair of dogs was subjected to three different enrichment types: food (diet frozen in water), tactile (rope pull toy) and olfactory (poultry litter) offered in a random block design (Figure 3.2). After each week of data collection and a four-day rest period, dogs were rotated to the next observational paddock with a new enrichment type.

Each pair of dogs was offered one of the enrichment items for five consecutive days, lasting seven hours daily (35 hours total per enrichment). During the data collection phase, the designated pair of dogs were exposed to the olfactory enrichment between 09:20 h and 16:20 h each day. The olfactory enrichment was placed in the paddock approximately five minutes before exposure and removed after the dogs were removed from the observation paddock each day. During the data collection phase, the designated dogs were exposed to tactile enrichment between 09:30 h and 16.30 h each day. The tactile enrichment was placed in the paddock before exposure on the first day of each data collection period and remained in the paddock for the five-day block. The tactile enrichment was removed after each five-day data collection period to avoid exposure to other pairs of dogs which used the paddock during the four-day rest period. The designated pair of dogs were exposed to food enrichment between 10:00 h and 17:00 h each day. The food enrichment was placed in the designated observation paddock approximately five minutes before exposure to the enrichment. The timing of enrichment exposure for each enrichment type remained the same for the entirety of the study, so conditions remained constant.

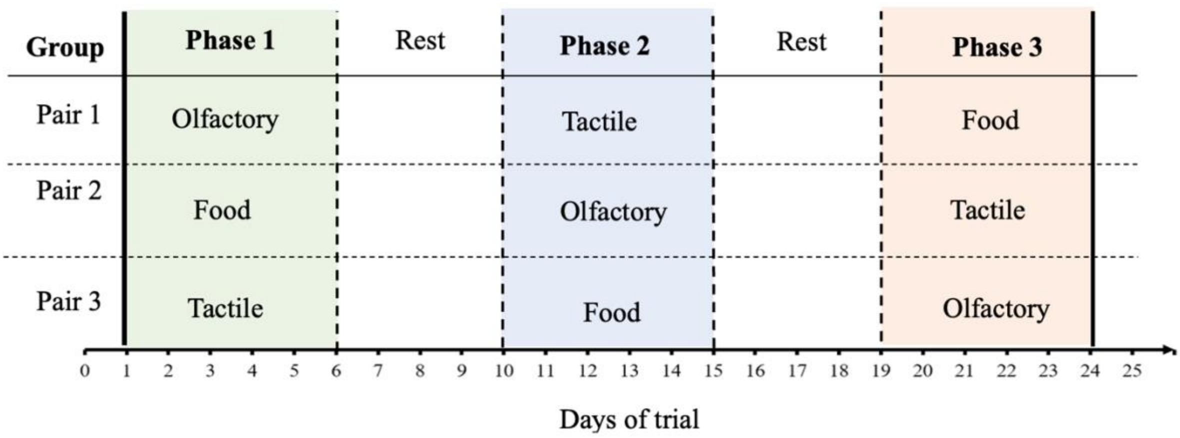

Figure 3.2 Experimental study design displaying data collection phases and rest periods for each pair of dogs across different enrichment types.

3.2.4 Habituation phase

During the week preceding the experimental phase, activity monitors were attached to each pair of dog’s existing collars to assess the response of the dogs to wearing the accelerometer devices. The dogs in this experiment had worn the ActiGraph® wGT3X-BT devices before, and there were no adverse effects, therefore, the devices were only reattached to the collars for four hours for each dog. During this habituation phase, dogs were placed in a paddock in their regular pairs and observed to ensure no noticeable effects on their behaviour or well-being. No dogs in this study showed any response to wearing the devices and, therefore, proceeded to the main study.

3.2.5 Acceleration data and video footage

The ActiGraph® wGT3X-BT (ActiGraph®, Pensacola, FL, USA) activity monitors were used to measure the activity and behaviour of dogs throughout this trial (**Figure 2.2a**). These devices were fitted to the existing collars of the dogs and oriented according to **Chapter 2**. Additionally, as stated in the previous study (**Chapter 2**), the plastic casing was filled with bubble wrap and the devices were wrapped in waterproof casing to prevent any damage or movement in the plastic cases (**Figure 2.2b**).

The six dogs participating in the enrichment study remained in regular pairs and were monitored using video surveillance in three familiar outdoor observation paddocks. Each pair of dogs was exposed to one of three forms of enrichment in these paddocks (**Figure 3.3).** The food enrichment paddock and the tactile enrichment paddock had the same dimensions (9.5-meters wide, 12 meters long), whilst the olfactory enrichment paddock was smaller (6.9 meters wide, 9.2 meters long) (**Figure 3.3**). The olfactory enrichment paddock was purposefully selected to be further away from the other paddocks, to avoid exposure to the olfactory enrichment before the data collection phase.

Video recordings were initiated approximately 15 minutes before the onset of data collection each day at 09:05 and ended approximately 15 minutes post data collection at 17:15. The ActiGraph® devices were initialised to begin recording at 08:00 on the first day of each five-day data collection period and finish recording at 20:00 on the fifth day. Acceleration data were not collected while the dogs were housed indoors overnight, so collars were removed as a precaution to prevent any damage. Thus, video footage and acceleration data were collected from each pair of dogs for seven hours each day, totalling 35 hours of data per collection period and 105 hours of data collected over the entire study duration.

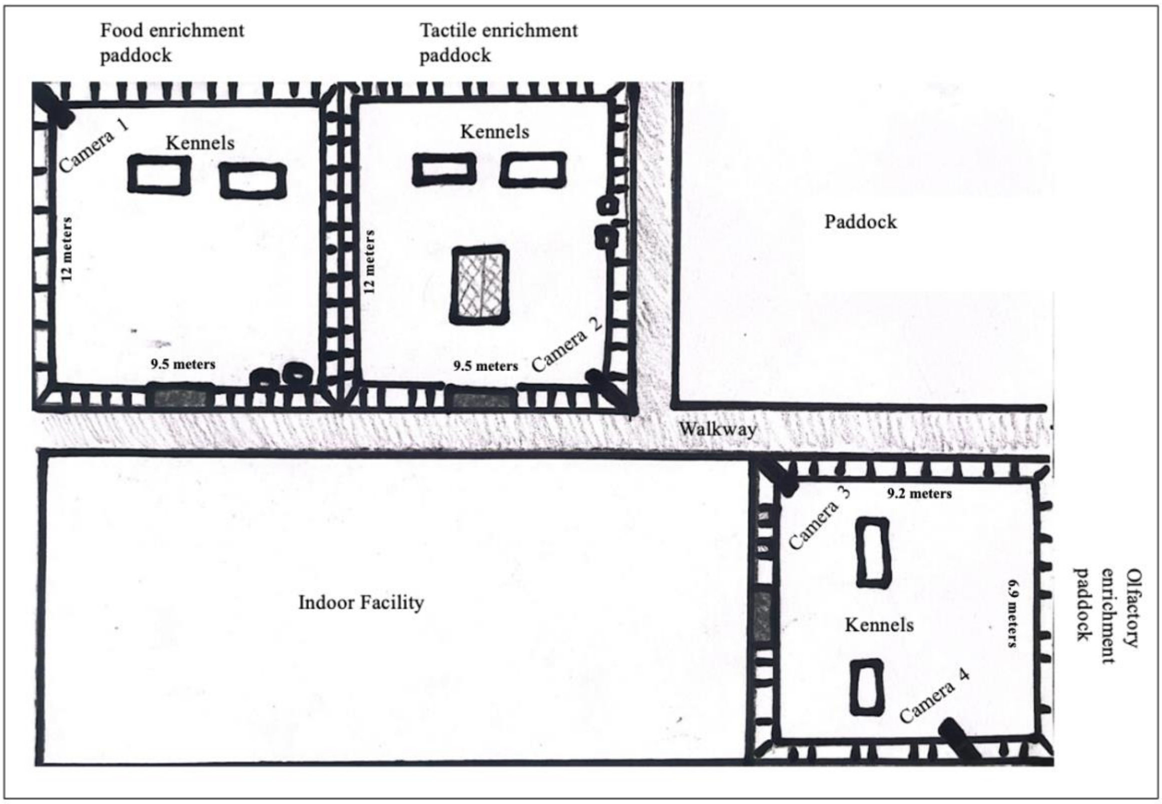

Figure 3.3 Diagram of the layout of the three enrichment paddocks in the kennelling facility: food (12 meter length x 9.5 meter width) , tactile (12 meter length x 9.5 meter width) and olfactory (9.2 meter length x 6.9 meter width).

3.2.6 Data evaluation and statistical analysis

Observational data were used to analyse the time spent engaging with each enrichment item during the first two hours of each day. This measurement was assessed through video recordings and interpreted using BORIS® version 7.10.2 (Friard et al., 2016). During this analysis, whether or not dogs had direct interactions with each enrichment type was recorded, marking ‘yes’ or ‘no’ in the BORIS programme in response to observed interactions with enrichment. An interaction was classified as licking or sniffing the ice block for the food enrichment treatment. Tactile enrichment interactions included sniffing or play behaviours with the tug rope, while olfactory enrichment was classified by sniffing behaviours in the four designated areas of the observation paddock. Each dog's total interactions were summed daily for each enrichment treatment. Statistical analyses were conducted using R version 4.3.0 (R Foundation for Statistical Computing, Vienna, Austria). A Kruskal-Wallis test was employed to analyse daily interactions between treatments due to the non-normal distribution of the data.

All data processing and statistical analyses were conducted using R version 4.3.0 (R Foundation for Statistical Computing, Vienna, Austria). The raw acceleration data (30 Hz) were downloaded from the Actigraph® devices using proprietary software ActiLife® (version 6.13.4; ActiGraph®, Pensacola, FL, USA). These data were then exported as Excel '.CSV' files. Initially, 32 identifier variables (**Table 2.3**) were calculated according to **Chapter 2**. These data were then analysed at a 1-second epochs using a modified version of a validated random forests model (**Appendix 1.4; Appendix 1.5**).

In this study, a slightly adjusted model compared to the validation study (**Chapter 2**) was utilised for data analysis. This model **(Appendix 1.4; Appendix 1.5)** exhibited an overall accuracy of 0.74 and a kappa coefficient value 0.68. This model was built using the same modelling techniques and identifier variables (**Table 2.3**) as the previous models in the validation study (**Chapter 2**). In **Chapter 2**, we developed and validated a machine learning algorithm to accurately classify a wide range of dog behaviours from triaxial acceleration data. In this study, we modified this model to better capture specific activities associated with enrichment treatments. Therefore, the 'scratching' behaviour was removed, and behaviours in the new model included 'barking', 'defecating', 'drinking', 'eating', 'locomotion', 'resting', 'resting-alert' and 'standing'. Behaviours 'sitting' and 'lying-alert from previous models were simplified into the behavioural category 'resting-alert', and behaviours 'lateral recumbency' and 'lying-asleep were simplified into the behavioural category 'resting'.

The acceleration data were filtered to exclude the periods over which the dogs were not wearing the collars (e.g., overnight), which resulted in 25,200 data points collected daily and 378,000 data points collected for the entire study period (15 days) per dog. Additionally, due to device malfunction, day 14 (04/02/2024) for Belle and Blacky, and day 10 for Gus and Chevelle (31/01/2024) of data were removed from the dataset. This resulted in a total of 2,167,287 data points collected for the enrichment study across all dogs. The modified predictive model (**Appendix 1.4; Appendix 1.5)** was then utilised to create daily, hourly, and minute predictions of different behavioural variables (Overall physical activity, total active%, total inactive%, total maintenance%, locomotion%, barking%, sniffing%, eating%, drinking%, defecating%, standing%, resting-alert% and resting%.

Further analysis of the various behavioural variables was performed using R version 4.3.0 (R Foundation for Statistical Computing, Vienna, Austria). Differences between behavioural variables were tested for significance at P=0.05. The selection of statistical tests was based on the data's normality. The data for both active and inactive behaviour (**Figure 3.4**) was normally distributed, so an ANOVA test and Tukey's HSD test were used. For the daily behavioural analysis (**Figure 3.5**), a Kruskal-Wallis test was used due to non-normal distribution. Proportional ODBA data (**Figure 3.6**) used ANOVA and Tukey's HSD test. Interaction data (**Figure 3.7**) and hourly behaviour data (**Figure 3.8**) were analysed using Kruskal-Wallis and Pairwise Wilcox tests due to non-normal distribution. Additionally, to investigate the levels of correlation between hourly behavioural variables and hourly ODBA, a Spearman Rank correlation test was conducted.

Additionally, this adjusted model (**Appendix 1.4; Appendix 1.5**) based on the validated model in **Chapter 2** was used to establish a baseline dataset, utilising ActiGraph® data previously collected from the validation study (**Chapter 2**). The baseline data consisted of a total data set of 453,618 data points collected between August 30th- September 15th, providing 7 hours of data collected for three days (21 hours total) per dog. This data was compared to the current enrichment study data (**Chapter 3**) and was used as a baseline for the dog's behaviour when not exposed to enrichment. Subsequently, this dataset was also used to utilised to create daily, hourly, and minute predictions of different behavioural variables for each dog (Overall physical activity, total active%, total inactive%, total maintenance%, locomotion%, barking%, sniffing%, eating%, drinking%, defecating%, standing%, resting alert% and resting%.

3.3 Results

3.3.1 Daily active and inactive behaviour

A total of 15 days of activity data were collected from six dogs, resulting in 2,167,287 data points. Mean percentages of active and inactive behaviours were calculated for the study duration (**Figure 3.4**). Active behaviour included 'barking', 'locomotion', and 'sniffing'. inactive behaviours included 'resting', 'resting-alert', and 'standing', maintenance behaviours were not included due to their low prevalence. Percentage time per day spent exhibiting active behaviours, ranged between 12.48 and 47.88 (mean ± SEM, 27.77 ± 1.54), during tactile enrichment, 12.40 and 40.48 (mean ± SEM, 22.92 ± 1.40), during olfactory enrichment and 17.73 and 64.02 (mean ± SEM, 35.19 ± 1.94), during the food enrichment. Percentage time per day spent exhibiting inactive behaviours, ranged between 52.08 and 87.36 (mean ± SEM, 72.19

± 1.54), tactile enrichment, 59.44 and 87.47 (mean ± SEM, 77.00 ± 1.40) for olfactory enrichment and 35.85 and 82.27 (mean ± SEM, 64.67 ± 1.95) for food enrichment.

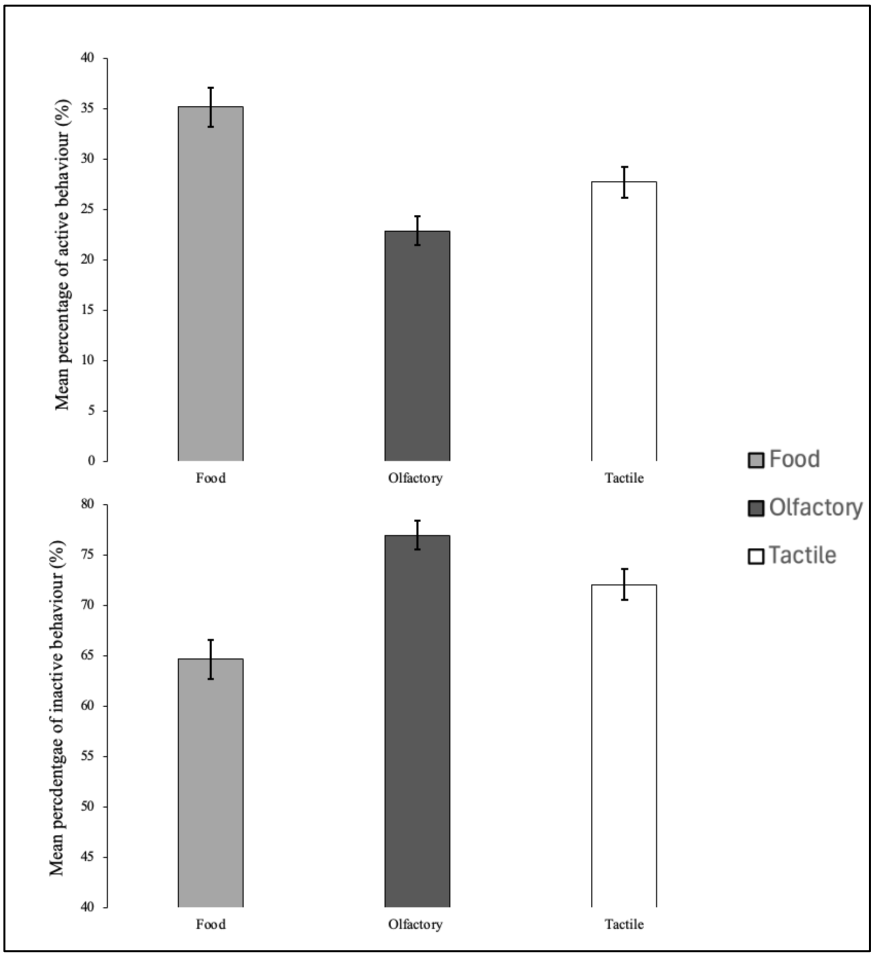

Figure 3.4 The mean active and inactive behaviours (%) among dogs, categorised by enrichment strategy (food, olfactory and tactile). Y axis on the graphs have been rescaled for better visualisation of observed data.

Both Active behaviour ($P = 0.04$) and Inactive behaviour ($P = 0.04$) demonstrated significant differences between enrichment treatments. Further analysis (Tukey's HSD test) revealed that the food enrichment treatment resulted in significantly higher active daily behaviour than the olfactory enrichment treatment ($P < 0.05$; **Figure 3.4**).

3.3.2 Behaviour profiles

Behaviour profiles (mean percentage) were calculated for baseline, food, olfactory and tactile treatment groups (**Figure 3.5**). The behavioural categories included resting (alert), locomotion, standing, resting (asleep), sniffing, and barking (defecating, drinking, and eating were removed from the figure (**Figure 3.5**) due to very low prevalence; **Appendix 1.5**). The percentage of time (%) spent exhibiting 'locomotion' (P<0.001), 'standing' (P<0.05), 'resting' (P<0.05), and 'sniffing' (P<0.001) behaviours differed significantly among the treatment groups (Baseline, Food, Olfactory and Tactile).

The percentage of locomotive behaviour was significantly lower in the olfactory treatment (mean ± SEM, 11.56 ± 1.21) compared to the food treatment (mean ± SEM, 18.56 ± 1.70, P<0.05), tactile treatment (mean ± SEM, 16.99 ± 1.57, P<0.05), and baseline treatment (mean ± SEM, 22.78 ± 2.71, P<0.001; **Figure 3.5**). There was no statistical difference in the daily locomotion behaviour among the baseline, food, and tactile treatment groups (**Figure 3.5**).

There were significant differences observed with standing behaviour being higher in the baseline treatment (mean ± SEM, 23.49 ± 2.41) compared to the food treatment (mean ± SEM, 12.68 ± 1.16, P<0.001) and the olfactory treatment (mean ± SEM, 13.90 ± 1.17, P<0.05). Additionally, baseline treatment tended to be higher than tactile treatment (mean ± SEM, 15.70 ± 1.42, P=0.077).

More specifically, resting was significantly lower in the baseline treatment (mean ± SEM, 8.04 ± 1.89) compared to the tactile treatment (mean ± SEM, 17.93 ± 2.02, P<0.05) and olfactory treatment (mean ± SEM, 17.53 ± 2.38, P<0.05). No difference was observed in daily resting behaviour among the olfactory, food, and tactile treatments.

The food treatment (mean ± SEM, 10.94 ± 1.00) showed significantly higher sniffing behaviour than the baseline treatment (mean ± SEM, 4.84 ± 0.62, P<0.001), tactile treatment (mean ± SEM, 5.48 ± 0.54, P<0.001), and olfactory treatment (mean ± SEM, 6.05 ± 0.62, P<0.05). However, barking and resting-alert behaviour were not affected by enrichment treatment.

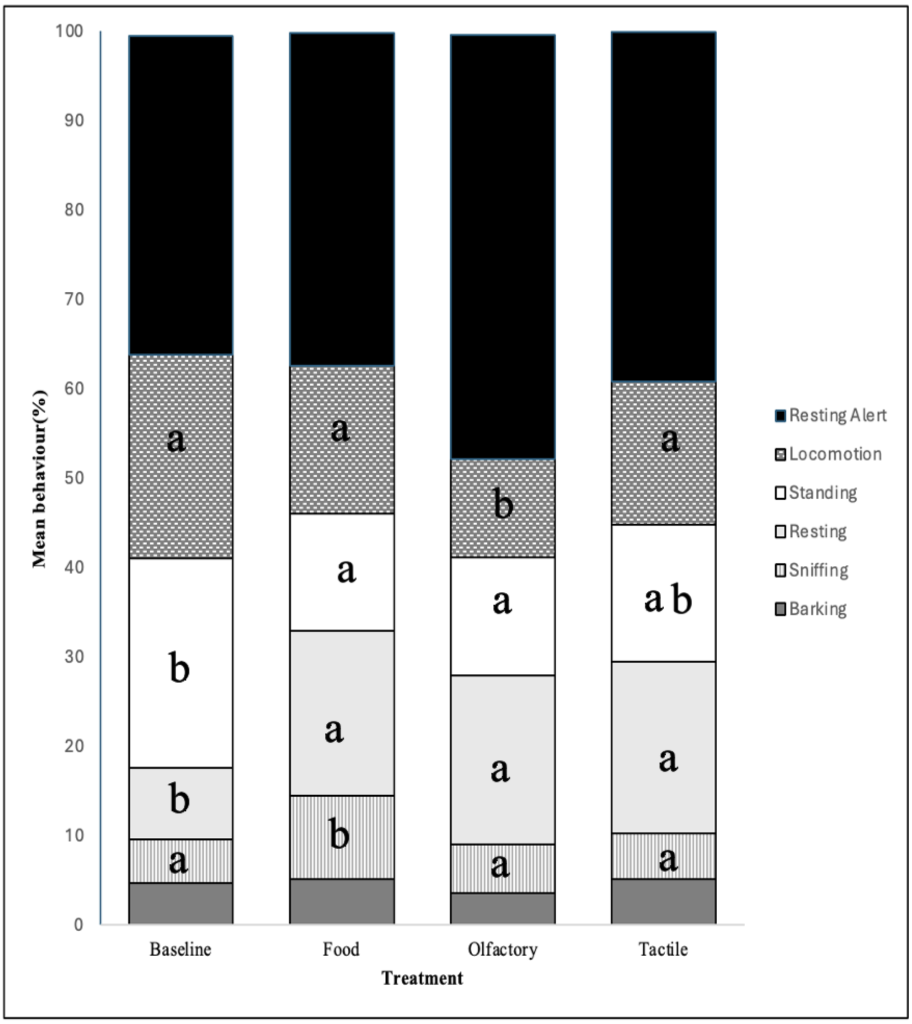

Figure 3.5 Average behaviour profiles from baseline data and each enrichment type for all dogs participating in the study for the entire day. Statistical significance of each behaviour between treatments is also demonstrated with subscripts ‘a’ and ‘b’ demonstrating the statistical difference between treatments and subscript ‘ab’ demonstrating no statistical difference between treatments.

3.3.3 Proportional ODBA

There was some significant differences in ODBA measurements between individual dogs (e.g. ‘Gizmo’ and ‘Chevelle’; $P < 0.05$). Therefore, proportional ODBA was calculated to present the data more clearly. The mean ODBA was recorded for each dog for each study day. Using this, an overall average ODBA was calculated for all dogs for all 15 days of the study period, so that the change relative to this could be plotted for each enrichment (**Figure 3.6**).

Food enrichment had the highest ODBA levels, followed by tactile enrichment, with olfactory enrichment showing a notably lower ODBA levels. There was a significant effect (P<0.001) of enrichment treatment on the ODBA of the dogs (**Figure 3.6**). The ODBA of the dogs as a proportional change from baseline (**Figure 3.6**) was lower following olfactory enrichment (mean ± SEM, 0.81 ± 0.05) compared to both food enrichment (mean ± SEM, 1.16 ± 0.06, P<0.001) and tactile enrichment (mean ± SEM, 1.10 ± 0.05, P<0.001). An ANOVA test examined whether the food treatment (P=0.38), olfactory treatment (P=0.28), and tactile treatment (P=0.27) showed any significant differences in proportional ODBA by day. Day appeared to have no significant effect on proportional ODBA for all enrichment treatments.

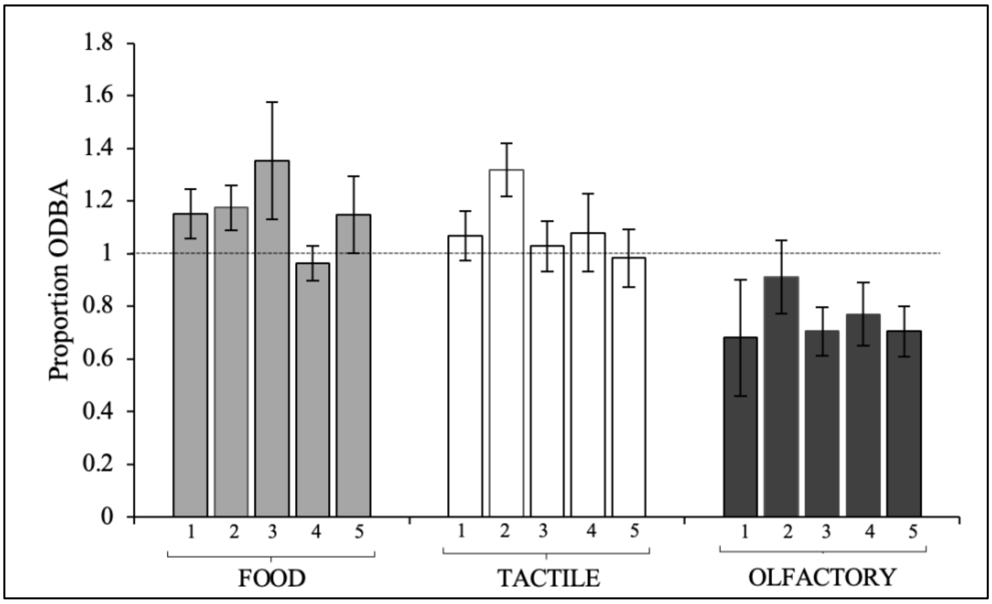

Figure 3.6 Trend lines depict all dogs' average ODBA for days 1-5 for each enrichment type (tactile, olfactory, and scent). The graph is scaled proportionally from 0 to 1.8, with 1 representing all six dogs' average activity level (ODBA) over the entire period. Error bars representing the standard error of 6 dogs are included.

3.3.4 Interaction with enrichment

Interactions with food, olfactory, and tactile enrichment types were recorded daily during the first two hours of enrichment treatment (**Figure 3.7**). The number of interactions with the enrichment item differed between the treatment groups (*P* < 0.001; **Figure 3.7**). All treatment groups differed from one another (*P* < 0.001), with the average number of interactions being highest for the food treatment (mean ± SEM, 1068.98 ± 121.25), then the olfactory enrichment (mean ± SEM, 204.59 ± 40.42) and tactile enrichment (mean ± SEM, 21.52 ± 9.16, **Figure 3.7**). For food enrichment, a slightly decreasing trend in interaction was observed from day 1 to day 5. For olfactory enrichment, a slightly increasing trend in interaction was observed from day 1 to day 5. Peak interaction occurred on day 5; the lowest interaction was observed on day 2. For tactile enrichment, an overall decreasing trend in interaction was observed from day 1 to day 5.

Additionally, time spent interacting with the enrichment item differed between dogs in the food enrichment treatment ($P<0.05$). Interestingly, the day variable did not significantly affect food or olfactory treatment. However, there was a significant effect on tactile treatment ($P <0.05$; **Figure 3.7)**. Between dogs, there were significant differences in interaction time for food enrichment treatment ($P<0.05$) and olfactory enrichment treatment ($P<0.05$).

3.3.5 Hourly ODBA and behaviour profiles

Hourly average ODBA (%) and behavioural comparisons between days one and five are provided in **Figure 3.8.** Across the enrichment types, diverse patterns of ODBA were observed over the 7 hours of behavioural observation. The food enrichment treatment showed highly significant differences ($P <0.001$), and tactile enrichment treatment also showed significant differences ($P<0.05$) in ODBA by hour. A significant difference was observed in ODBA between days 1 and 5 for olfactory enrichment ($P < 0.05$). Noticeably, ODBA was lower on day 5 compared to day 1. Day 1 saw a decline in ODBA after the first hour until a sharp increase at hour 7, while day 5 showed a smooth U-shaped trend, peaking in hour 1 and rising again between hours 5-7. Food enrichment showed highly significant differences in ‘sniffing’ behaviour by hour ($P <0.001$), and tactile enrichment also showed significant differences in ‘sniffing’ behaviour by hour ($P<0.05$). Additionally, there were significant differences in sniffing behaviour observed between day one and day 5 for food enrichment treatment ($P <$

0.05). Notably, in hour 1 of day 1, 'sniffing' behaviour was nearly double that observed in hour 1 of day 5. On days 1 and 5, 'sniffing' was highest in the first hour of the day. Locomotion behaviour showed significant differences ($P<0.05$) hourly for food and tactile enrichment treatments and significantly different for olfactory enrichment ($P<0.001$). Barking did not show a significant difference by the hour for any treatments. Resting showed significant ($P<0.001$) differences per hour for food and tactile enrichments and no significant difference for olfactory enrichment. However, there were highly significant differences in resting behaviour between day one and day 5 for olfactory enrichment ($P<0.001$). The hour variable for any enrichment treatment did not significantly affect standing behaviour. Although, there were significant differences in standing between day one and day 5 for both olfactory and tactile treatments ($P < 0.05$). Resting alert was only significantly affected ($P < 0.05$) by the hour variable in the tactile treatment. It appears that ODBA patterns are associated with locomotion behaviour observed throughout the day. Spearman Rank tests were conducted to test the relationship between ODBA and the other behavioural variables. Hourly 'locomotion' and 'ODBA' variables had a correlation coefficient R=0.677, indicating a moderately strong relationship between variables

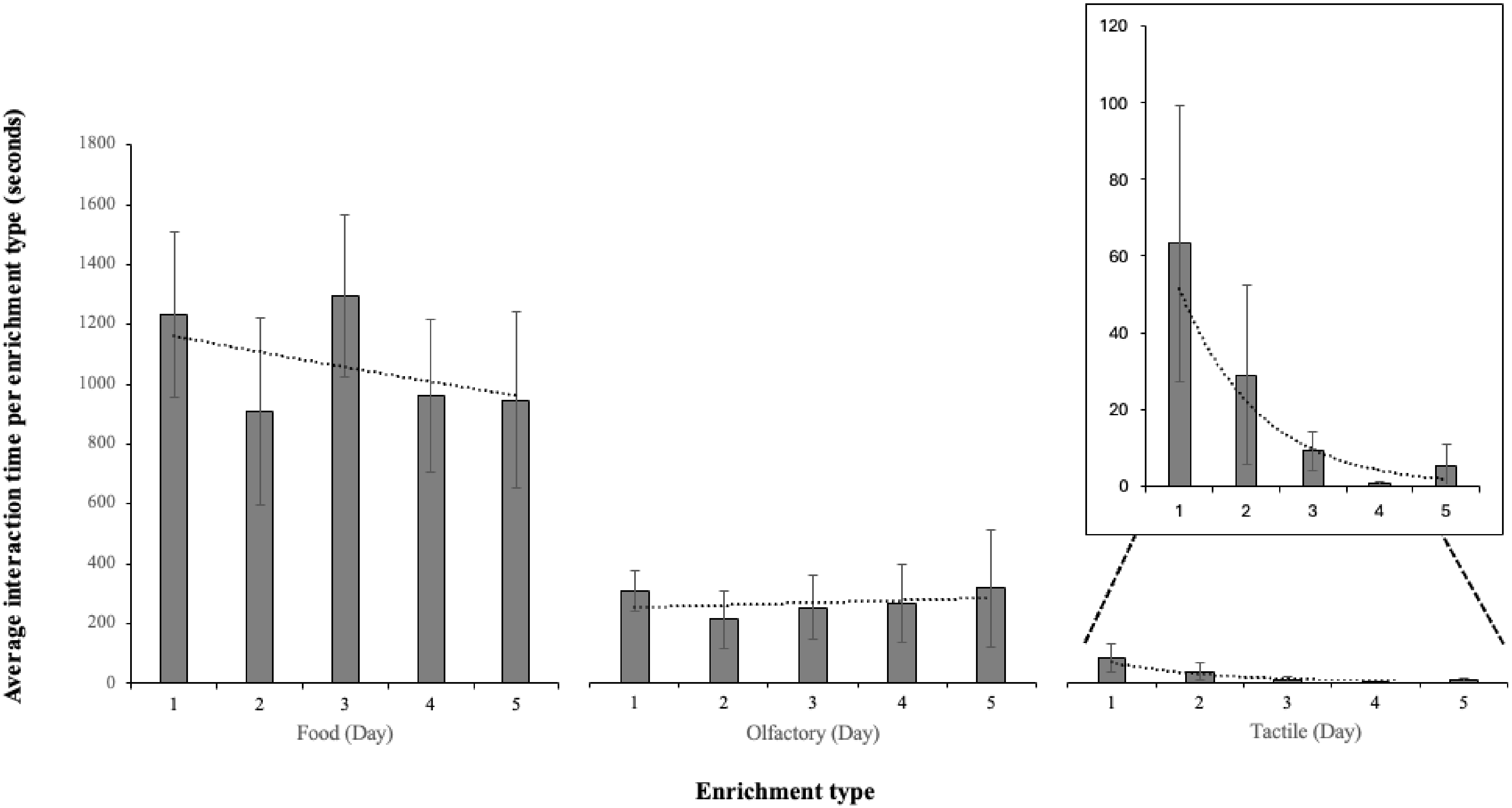

Figure 3.7 Average interaction time (seconds) for each enrichment type (food, tactile, and olfactory). A scaled-in version of tactile enrichment was included. The trend line shows the mean interaction time per day for all dogs. Standard error bars included.

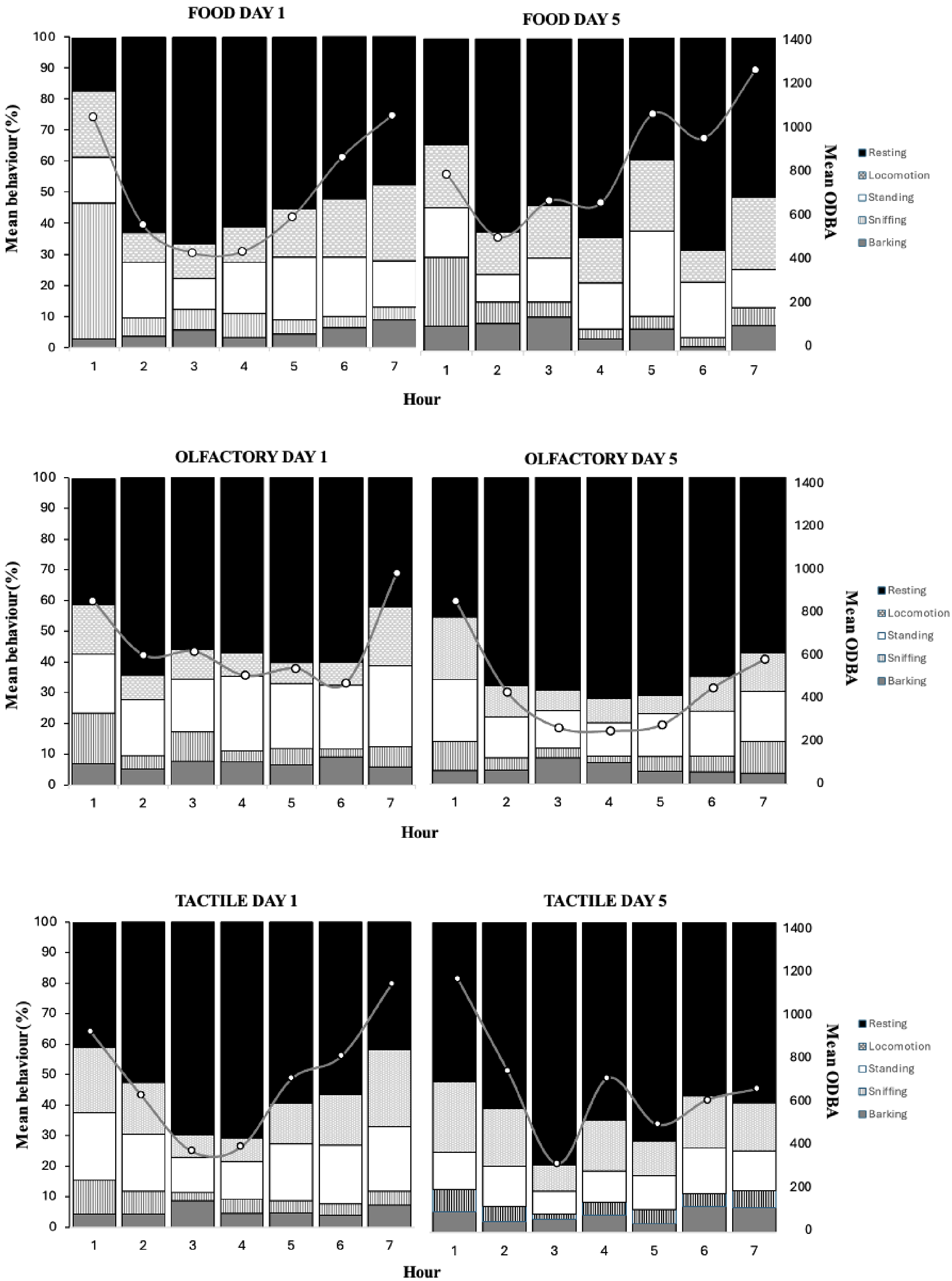

Figure 3.8 Graphs showing hourly mean behaviour (left side of X-axis) and hourly mean ODBA (right side of X-axis) for day one and day five of each enrichment type (food, olfactory, and tactile).

3.4 Discussion

Environmental enrichment plays an essential role in the overall welfare status of animals; therefore, studying the effectiveness of different enrichment types is crucial for optimising animal care practices, particularly for companion animals living in confined environments with limited human interaction. However, assessing the efficacy of enrichment items is not standardised, leading to variable results (Gaines et al., 2008; Schipper et al., 2008; Herron et al., 2014; Hunt et al., 2022). The first aim of the study was to evaluate the potential for tri-axial accelerometery and a previously validated random forests model (**Appendix 1.4; Appendix 1.5**) for determining the efficacy of different environmental enrichment treatments. The second aim of this study was to assess the effect of food, olfactory, and tactile enrichment treatments on the behaviour and activity of colony-housed dogs. This involved analysing active/inactive behaviour, ODBA levels, behavioural patterns and time spent interacting with the enrichment. Significant differences were observed among the enrichment treatments regarding active/inactive behaviour, ODBA levels, individual behaviour and interaction durations.

All enrichment treatments (food, olfactory, and tactile) showed significant hourly differences in locomotory behaviour, indicating that the hour of the day affects locomotion. Additionally, a correlation coefficient 0.677 between hourly ODBA and locomotion indicated a moderate relationship between these variables. In the validation chapter of this thesis (**Chapter 2**), it was also observed that total ODBA was a significant predictor of the amount of time spent actively per hour (P<0.001) and, thus, overall physical activity (**Figure 2.6**). These results align with a prior study on dingoes (*Canis dingo*) utilising accelerometers, which indicated a positive relationship between mean active behaviour and ODBA levels, supporting the idea that higher activity levels coincide with increased ODBA levels (Tatler et al., 2018). Consequently, the observed relationship between ODBA and active behaviour in this study suggests that tri-axial accelerometery is a dependable method for accurately assessing environmental enrichment's effects on behavioural indicators such as locomotion and ODBA.

Introducing novel items such as enrichment to an animal typically encourages them to attend to these objects quickly. Thus, analysing the initial two hours of interaction for each enrichment treatment was crucial for assessing enrichment efficacy in this study, along with the behavioural observations from the accelerometer data. The dogs spent significantly more time interacting with the food treatment daily than the other enrichment treatments. However, past research has highlighted that interaction with feeding enrichment does not always correlate to significant changes in overall behaviour (Gaines et al., 2008; Hunt et al., 2022). However, in the current study, the food enrichment treatment showed significantly higher active behaviour, ODBA and locomotion than the other treatment groups. These results align with a study by Schipper et al. (2008), where activity significantly increased for kennelled dogs exposed to a feeding enrichment device. Overall, activity and behavioural variability stimulated by engaging in appetitive behaviour have been reported as a benefit of feeding enrichment (Schipper et al., 2008). Therefore, both the increased activity and interaction with the enrichment suggest that the food enrichment treatment had a positive impact on the dogs welfare in the study.

However, Polgár et al. (2019) provided a comprehensive review of welfare measures in kennelled dogs, cautioning that activity levels are complex when measuring welfare outcomes as both increased and decreased activity levels have been suggested as potential indicators of stress in animals. Hiby et al. (2006) also noted that stress could be indicated by both high and low activity levels in dogs, with some dogs exhibiting high activity levels when stressed or anxious and others showing low activity, becoming less engaged with their environment. Therefore, it is suggested that using variables such as active behaviours and ODBA alone may not be suitable for investigating dog welfare.

Dogs exposed to the olfactory enrichment treatment showed an overall positive trend in time spent interacting with the enrichment over the study week and spent significantly more time interacting with the enrichment than the tactile enrichment treatment. The data shows a positive trend from day one to day five, suggesting that the dogs did not habituate to the treatment over time (Kuczaj et al., 2002). While the use of prey species for olfactory enrichment in domestic dogs has yet to be extensively documented, it has demonstrated the potential to be a valuable tool for olfactory enrichment in the future. However, more in depth research is needed to understand its impact on dog welfare. Previous studies have also reported success in using the faeces of prey species as enrichment. For example, placing Grant's gazelle dung outside enclosures of African wild dogs resulted in higher activity (10.6%) and increased social behaviour (Cloutier and Packard, 2014). However, dogs in the olfactory treatment exhibited the lowest daily ODBA levels and locomotion (%) among the three enrichment treatments. With these results, one factor to consider is that the paddock chosen for olfactory enrichment was 44.3% smaller than those used for food and tactile enrichment treatments, which has likely contributed to the lower activity observed in the olfactory enrichment compared to the other treatments. Siwak et al. (2003) reported that smaller housing areas limited movement in laboratory dogs, which could explain the lower ODBA levels and locomotion (%) observed in the olfactory than the tactile and food enrichment treatments. Therefore, the significantly lower ODBA and locomotion observed in the olfactory treatment suggests that other factors, such as smaller paddock size, may influence the observed results.

Although the dogs initially engaged with the tactile enrichment items, their interaction sharply decreased over the treatment period, and the interaction was very low, indicating a rapid loss of interest. Similar findings were observed in a study on kennelled dogs, showing that they spent relatively little time (<8%) playing with tactile enrichment and that they quickly became accustomed to the item (Wells, 2004). This decline may have also been influenced by the dogs' lack of prior exposure to tactile enrichment items (Sampaio et al., 2019).

Sniffing behaviour is directly linked to a dog’s overall behaviour, cognition, and welfare (Nielsen et al., 2015; Rooney and Parr-Cortes, 2023). In the current study, dogs spent significantly more time sniffing in the food enrichment treatment compared to the baseline, olfactory, and tactile treatments. While, to some extent, these results are expected due to the ice block encouraging sniffing behaviour, it seems that misclassification of the model may have occurred because no specific licking behaviour was included in the RF model’s validation process. Daily sniffing behaviour was almost double all of the other treatments, suggesting that licking the ice block may have exhibited similar motion and head placement to the ‘sniffing’ behaviour, potentially affecting the overall accuracy of predicting this particular behaviour. Additionally, there were highly significant differences in sniffing behaviour in the food enrichment treatment between the first hour of day one and day five, which was reduced on day five. This further indicates that the model misclassified licking the ice block as ‘sniffing’ behaviour, as the first hour is when we would expect the dogs to interact with the enrichment (Domjan., 2018). Therefore, the reduction in sniffing behaviour for food enrichment treatment from day one to day five indicates possible habituation to the enrichment over the week, assuming that sniffing is misclassified (Kuczaj et al., 2002; Tarou and Bashaw, 2007).

Understanding the resting patterns of animals has the potential to contribute to identifying negative welfare concerns such as stress, discomfort and other health issues in animals (Schork et al., 2024). Interestingly, dogs spent significantly more time resting (asleep) in the food, olfactory, and tactile enrichment treatments compared to the baseline treatment. This raises the question of whether more time spent sleeping during the day is associated with positive welfare. It has been reported that lack of rest may be a welfare issue in dog shelters, where dogs are typically exposed to more stress (Owczarc-Garstecka and Burman, 2016). Therefore, this may indicate that dogs exposed to the enrichment treatment exhibited less stress. However, comparing the baseline data to enrichment data raised another question of whether seasonal variations influenced dog behaviour with the baseline data collected in August/September and the enrichment data collected in January/February. Studies have suggested seasonal fluctuations in cortisol levels, with higher cortisol in winter leading to increased activity (Roth et al., 2016; Wirobski et al., 2021). These fluctuations could be due to changes in environmental conditions, such as temperature and daylight hours, affecting the dogs' physiological and behavioural responses. Therefore, future research should consider the timing of enrichment interventions and how seasonal changes might influence their effectiveness, ensuring that dogs receive optimal welfare practices throughout the year.

Barking is a behaviour that can indicate negative welfare or distress (Fox, 1971; Hetts et al., 1992; Beerda et al., 1999). Since environmental enrichment aims to improve welfare and relieve stress in animals, it was surprising that no significant daily differences were observed between treatments for barking as it would be expected that barking would decrease between baseline and enrichment treatments. The daily behavioural percentages of barking were higher for all treatment groups (baseline, food, tactile, olfactory) than those previously reported in suburban dogs (0.45%); however, this data was collected on dogs that are housed individually (Flint et al., 2013). In the current study, dogs are housed in outdoor paddocks in a colony, which may lead to more social stressors in the environment and promote barking behaviour, as observed in this study (Fox, 1971).

The ‘dog’ variable appeared to play a significant role in the differences observed in enrichment interaction between treatment groups. Although the variability in the data may be considered a limitation, assessing a range of breeds and individual temperaments is essential when designing personalised enrichment programs (Kumpulainen et al., 2021). In this study, five out of the six dogs were huntaway variants (**Table 3.1**). However, dogs with different temperaments were chosen to increase diversity in the behavioural response to the enrichment treatments. Future studies should continue to test enrichments on various breeds and individual temperaments when creating personalised enrichment programs. Therefore, more accurate and tailored enrichment treatment programs can be provided to the specific needs of each dog. This personalised approach could enhance the overall effectiveness of enrichment programs, ensuring that each dog receives the most appropriate and beneficial enrichment.

3.5 Conclusion

This study emphasised the significance of environmental enrichment in enhancing the well-being of dogs in confined spaces. It used tri-axial accelerometery and a previously validated random forests model to assess behavioural differences between enrichment treatments and compared them to a baseline dataset.

The findings demonstrated significant differences in interaction, active/inactive behaviour, activity levels, and behavioural patterns across food, olfactory and tactile enrichment treatments. The food enrichment treatment significantly increased interaction duration with enrichment active behaviour, ODBA levels, and locomotion in the dogs. Food enrichment notably increased sniffing behaviour, but the model's potential misclassification issues indicate that further refinement may be necessary to accurately predict the success of this enrichment.

The olfactory enrichment maintained interaction engagement over the five-day treatment period despite lower active behaviour, ODBA and locomotion, likely due to the smaller paddock size. The tactile enrichment interaction duration was very low compared to the other enrichment treatments and quickly lost its novelty over the treatment period, exhibiting a significant reduction in interaction by day over the treatment period.

In conclusion, the study provides valuable insights into how ActiGraph® accelerometers can be used to measure the efficacy of different environmental enrichment treatments. It emphasises the need for comprehensive and individualised approaches to environmental enrichment, and factors such as environment, season, and study design should also be carefully considered for future studies.

=

Chapter 4

Overall discussion and future directions

Chapter 4: Overall discussion and future directions

4.1 Key findings of this thesis

Firstly, this thesis aimed to investigate the use of remote sensing technology, particularly tri-axial ActiGraph® WGT3X-BT accelerometers, along with machine learning (ML) algorithms to automatically classify behaviour in six domestic dogs (**Chapter 2**). The following aims of the thesis were to evaluate the potential for tri-axial accelerometery and a validated random forests model for determining the efficacy of environmental enrichment treatments and to assess the effect of food, olfactory, and tactile enrichment treatments on the behaviour and activity of colony-housed dogs (**Chapter 3**). The research was separated into two main research chapters (**Chapter 2 and Chapter 3**) that examined each objective of the thesis independently.

Chapter 2 showed that ActiGraph® devices were a valid and objective method of measuring behaviour and activity (ODBA) in domestic dogs, thus supporting using these devices for behavioural identification and activity monitoring in future studies. Five random forest (RF) models were built using machine learning techniques, and their performance characteristics were compared. Model 4 was identified as the optimal model, achieving the highest overall accuracy (0.74) and kappa coefficient (0.68) while still capturing a relatively wide range of behaviours. This model evaluated nine behavioural categories: 'barking', 'defecating', 'drinking', 'eating', 'locomotion', 'resting-asleep', 'resting-alert', 'sniffing' and 'standing'. The model-building process revealed challenges in differentiating behaviours with similar or overlapping acceleration profiles, particularly in classifying 'standing' behaviour. As a result, behaviours were grouped during the model-building process to improve overall accuracy. The refined models showed significant improvement over time, indicating a promising method for detailed and remote assessment of domestic dog behaviour.

In **Chapter 3**, six colony-housed dogs were exposed to three environmental enrichment treatments (food, olfactory, and tactile). The study evaluated the efficacy of each enrichment type and the effectiveness of ActiGraph® as a tool for behavioural monitoring. Significant differences were observed among enrichment treatments regarding active/inactive behaviour, ODBA levels, individual behaviours, and interaction durations. Additionally, ActiGraph® devices have been shown to be an accurate and objective method for measuring the success of enrichment activities, particularly when used alongside other observational methods, such as assessing the duration of interaction with each enrichment treatment. The study highlighted the impact of environmental factors, individual dog differences, and seasonal changes on enrichment efficacy, emphasising the necessity of personalised enrichment programs to enhance the overall effectiveness of enrichment treatments.

4.2 Validation of ActiGraph® accelerometers

Accelerometers allow for the remote study of large quantities of behavioural data, offering advantages such as objective and reliable data collection for long periods without an observer present (Lascelles et al., 2008; Dow et al., 2009; Morrison et al., 2013; Hounslow et al., 2019). With the emergence of technological tools such as accelerometers, machine learning has become a powerful tool for automatically classifying animal behaviours (Hounslow et al., 2019). While accelerometers have been successfully validated for measuring activity levels in domestic dogs (Yam et al., 2011; Morrison et al., 2013; Ortmeyer et al., 2018), studies are recently beginning to focus on specific behaviour classification, as this may reveal more about the overall welfare state of a dog than activity levels alone (den Uijl et al., 2017; Chambers et al., 2021; Kumpulainen et al., 2021; Eerdekens et al., 2022). Therefore, the research presented in Chapter 2 aimed to address this gap in research by validating tri-axial accelerometers (ActiGraph®) and training unique machine learning algorithms to predict behaviour automatically in domestic dogs. This validation of tri-axial accelerometers, such as the

ActiGraph® WGT3X-BT device, can be applied to future research, as demonstrated in **Chapter 3**.

The overall accuracy increased as the model included fewer behaviours, which aligns with other studies on companion animals that used accelerometers for behavioural classification (den Uijl et al., 2017; Smit et al., 2023). Additionally, grouping behaviours with similar acceleration profiles also contributed to higher overall accuracy observed in the models, as this approach reduced the complexity of the behaviour classification task and improved the model's performance. Consolidating similar behaviours during model development has also improved overall accuracy for domestic cat behaviour (Smit et al., 2023) and domestic dog resting behaviours (den Uijl et al., 2017). In future studies, the main objective will be to continue validating these ActiGraph® devices to encompass a broader range of behaviours and to achieve higher overall accuracies and kappa coefficients. This will lead to a more objective and accurate measure of behaviour and activity in domestic dogs. One recognised limitation of the study (**Chapter 2**) was the relatively small sample size of dogs (n=6). To increase the accuracy and versatility of accelerometers as a behavioural measurement tool for dogs, it is recommended to further validate the ActiGraph® accelerometers across a broader range of dog breeds, temperaments, and environmental settings to confirm their utility and accuracy in different contexts (Kumpulainen et al., 2021). For example, den Uijl et al. (2017) conducted a study where 51 domestic dogs participated from 8 different breed categories where tri-axial accelerometers were proved to have over 95% accuracy in behaviour classification of 'walking', 'trotting', 'galloping', 'eating', 'drinking', and 'headshaking' and over 90% accuracy in identifying inactive behaviours. Additionally, Chambers et al. (2021) validated numerous tri-axial accelerometers using a large data set consisting of 2500 dogs of different breeds and ran the algorithms on more than 11 million days of device data where high sensitivities and specificities were observed for behaviours such as drinking (sensitivity: 0.949, specificity: 0.999) and eating (sensitivity: 0.988, specificity: 0.983). Therefore, continuing to validate ActiGraph® devices across a broader spectrum of behaviours in diverse dog populations will be critical for achieving high accuracy and reliability in monitoring dog behaviour, ultimately leading to a more precise assessment of domestic dog activity and behaviour.

The current validation study (**Chapter 2**) used 32 identifier variables to create a behaviour classification model derived from a study investigating domestic cat behaviour (Smit et al., 2023). It has been reported in other animal studies that the importance of predictor variables may change depending on what the species of interest target specific behaviours (Graf et al., 2015; Alvarenga et al., 2016; Tatler et al., 2018). For future studies, it would be noteworthy to investigate which predictor variables are most important for the acceleration signature of domestic dogs and whether the model's accuracy can be improved by adding or removing any of the predictor variables.

4.3 Measuring behavioural response to enrichment using ActiGraph® accelerometer

Environmental enrichment is vital in improving an animal's overall welfare state. It has numerous reported benefits in dogs, including increased behavioural repertoire, reduced stress and anxiety, reduced abnormal behaviours, and improved cognitive abilities (Hubrecht, 1993; Shepherdson, 2003; Wells, 2004; Schipper et al., 2008; Herron et al., 2014; Hunt et al., 2022; Kang, 2022). However, previous studies have focused on more subjective observational measures of behaviour when considering environmental enrichment efficacy in domestic dogs (Gaines et al., 2008; Schipper et al., 2008; Pullen et al., 2010; Herron et al., 2014; Amaya et al., 2020; Hunt et al., 2022). Therefore, this study (**Chapter 3**) highlighted a significant gap in research for using objective measures like accelerometers to evaluate the behavioural impacts of environmental enrichment in dogs. In the current study (**Chapter 3**), ActiGraph® devices revealed significant differences in daily active/inactive behaviour, ODBA, and individual behaviours across various environmental enrichment treatments. These differences were further supported by observational data, including the assessment of time spent engaging with enrichment items during the first two hours of each day, a measurement analysed through video recordings and interpreted using BORIS® version 7.10.2 (Friard et al., 2016). This provided a holistic approach to measuring enrichment success by assessing behavioural differences and interaction with the enrichment item.

Furthermore, to get a more accurate overall measure of enrichment success, future studies should incorporate physiological indicators such as cortisol concentration as a measure of success alongside behavioural data obtained from accelerometers to evaluate environmental enrichment's effects on stress levels and overall welfare in domestic dogs. This would allow for a better overall view of welfare. This approach would provide a more comprehensive view of welfare and a more accurate assessment of whether stress levels increased or decreased, rather than solely relying on behavioural observations. Additionally, this approach could provide more information on individual dogs within a population and how the enrichment affects them. Studies have previously reported that glucocorticoid cortisol concentrations increase long after dogs enter stressful, confined environments such as kennelled facilities (Hennessy et al., 1997; Hennessy et al., 2002; Stephen and Ledger, 2006; Part et al., 2014). Additionally, studies have combined tri-axial accelerometer devices alongside physiological indicators such as cortisol concentrations to create a more holistic approach to measuring dog welfare (Jones et al., 2014; van der Laan et al., 2023). However, in Jones et al. (2014) the study, the relationship between accelerometer activity and cortisol concentrations was inconclusive as they proved to be complicated measures of stress.

It would also be valuable to validate specific behaviours associated with the enrichment being tested to prevent the suggested misclassification of "sniffing" behaviour with licking the ice block, as observed in **Chapter 3**. This would provide a more accurate understanding of the effectiveness of the enrichment. Chambers et al. (2021) has previously utilised tri-axial accelerometers as a behavioural classification tool, demonstrating high sensitivity (0.77) and specificity (0.99) for 'licking' behaviour; therefore, this behaviour could also be validated for the current model. Additionally, it would be beneficial to include more positive behaviours, such as variations of 'play' behaviour, as this would allow for direct observation of behavioural responses to tactile enrichment (Polgár et al., 2019). Classifying play behaviour, in general, would also allow for better overall welfare understanding in the other treatments.

The lack of prior exposure to tactile enrichment among the dogs may have impacted their ability to engage effectively with the tactile enrichment items. Sampaio et al. (2019) stated that experience with tactile enrichment, such as toys, can impact their effectiveness. In the current study, the dogs' minimal exposure to toys, especially the novel rope-tug toys, might have led to low interaction levels. Additionally, this aligns with other studies stating that motivation for toy use may be low compared to other enrichment types, such as food-based enrichments (Wells and Hepper, 1992; Döring et al., 2016; Murtagh et al., 2020). These findings contrast with a study by Hunt et al. (2022), which reported lower stress behaviour scores for conspecific play, playhouse, and tug-rope play than food-based enrichment. However, this study involved a human handler encouraging interaction, whereas the current study did not. Future studies should consider human facilitation or instruction for such tactile enrichments to enhance their effectiveness. Social contact between humans and dogs can also promote a positive welfare state by decreasing stereotypical behaviours, increasing activity, and decreasing cortisol concentrations in dogs (Coppola et al., 2006; Menor-Campos et al., 2011).

4.4 Summary of future directions

- Conducting a validation study with a more extensive and diverse group of domestic dogs across various settings would provide a more robust validation of the ActiGraph® device. This would help confirm these devices' utility and accuracy across different breeds, temperaments, and environmental contexts.
- Investigating the 32 predictor variables and determining which ones are redundant could enhance the model's accuracy. Future research could also explore adding or removing predictor variables specific to domestic dogs to improve the model’s accuracy and versatility.
- Integrating physiological indicators such as glucocorticoid cortisol concentration or other blood-borne markers alongside behavioural measures when investigating environmental enrichment success would benefit future studies. Measuring physiological indicators such as cortisol concentration could allow for a more holistic and accurate approach to measuring welfare and stress in domestic dogs.
- Validating more behaviours for the RF model specific to the enrichment treatment, such as ‘licking’ and ‘playing’, would allow for a more precise behavioural analysis of the enrichment protocol.
- Future studies should consider the human facilitation of tactile enrichment to test whether this improves the interaction with enrichment. Human facilitation of tactile enrichment may encourage play behaviour or teach the dogs to use this enrichment. Additionally, human-to-dog interaction may also improve the dogs' overall welfare.

4.5 Conclusion

Accelerometery is an exciting tool that is transforming the future of animal behaviour assessment. It provides an objective and less labour-intensive method for common observational assessment. This thesis successfully validated an RF model for behaviour classification using machine learning techniques, proving that accelerometery can be applied to the behavioural assessment of domestic dogs. An adjusted version of this model was then used to investigate the efficacy and duration of different enrichment treatments (food, olfactory, and tactile), indicating that there were significant differences in behavioural observation between enrichment treatments. Although limitations of both studies have been reported, this thesis has proven that accelerometery is an exciting and promising method of behavioural assessment for domestic dogs and other animal species.

5.0 References

Alvarenga, F., Borges, I., Palkovič, L., Rodina, J., Oddy, V., Dobos, R., 2016. Using a three-axis accelerometer to identify and classify sheep behaviour at pasture. Applied Animal Behaviour Science 181, 91-99.

Amaya, V., Paterson, M., Descovich, K., Phillips, C.J., 2020. Effects of olfactory and auditory enrichment on heart rate variability in shelter dogs. Animals 10, 1385.

Andrews, C.J., Potter, M.A., Thomas, D.G., 2015. Quantification of activity in domestic cats (Felis catus) by accelerometry. Applied Animal Behaviour Science 173, 17-21.

Aragon, C.L., Hofmeister, E.H., Budsberg, S.C., 2007. Systematic review of clinical trials of treatments for osteoarthritis in dogs. Journal of the American veterinary medical association 230, 514-521.

Barthélémy, I., Barrey, E., Thibaud, J.-L., Uriarte, A., Voit, T., Blot, S., Hogrel, J.-Y., 2009. Gait analysis using accelerometry in dystrophin-deficient dogs. Neuromuscular disorders 19, 788-796.

Beerda, B., Schilder, M.B., Van Hooff, J.A., De Vries, H.W., Mol, J.A., 1998. Behavioural, saliva cortisol and heart rate responses to different types of stimuli in dogs. Applied animal behaviour science 58, 365-381.

Beerda, B., Schilder, M.B., Van Hooff, J.A., De Vries, H.W., Mol, J.A., 1999. Chronic stress in dogs subjected to social and spatial restriction. I. Behavioral responses. Physiology & behavior 66, 233-242.

Belshaw, Z., Asher, L., Dean, R.S., 2016. Systematic review of outcome measures reported in clinical canine osteoarthritis research. Veterinary Surgery 45, 480-487.

Bolton, S., Cave, N., Cogger, N., Colborne, G., 2021. Use of a collar-mounted triaxial accelerometer to predict speed and gait in dogs. Animals 11, 1262.

Brady, K., Cracknell, N., Zulch, H., Mills, D.S., 2018. A systematic review of the reliability and validity of behavioural tests used to assess behavioural characteristics important in working dogs. Frontiers in veterinary science 5, 103.

Brown, D.C., Boston, R.C., Farrar, J.T., 2010. Use of an activity monitor to detect response to treatment in dogs with osteoarthritis. Journal of the American Veterinary Medical Association 237, 66-70.

Brugarolas, R., Latif, T., Dieffenderfer, J., Walker, K., Yuschak, S., Sherman, B.L., Roberts, D.L., Bozkurt, A., 2015. Wearable heart rate sensor systems for wireless canine health monitoring. IEEE Sensors Journal 16, 3454-3464.

Bruno, K., DeSocio, E., White, J., Wilson, B.K., 2020. Effect of environmental enrichment devices on behavior of individually housed beef heifers. Translational Animal Science 4, txaa220.

Chambers, R.D., Yoder, N.C., Carson, A.B., Junge, C., Allen, D.E., Prescott, L.M., Bradley, S., Wymore, G., Lloyd, K., Lyle, S., 2021. Deep learning classification of canine behavior using a single collar-mounted accelerometer: Real-world validation. Animals 11, 1549.

Chan, C.B., Spierenburg, M., Ihle, S.L., Tudor-Locke, C., 2005. Use of pedometers to measure physical activity in dogs. Journal of the American Veterinary Medical Association 226, 2010-2015.

Cheung, K.W., Starling, M.J., McGreevy, P.D., 2014. A comparison of uniaxial and triaxial accelerometers for the assessment of physical activity in dogs. Journal of Veterinary Behavior 9, 66-71.

Clark, K., Caraguel, C., Leahey, L., Béraud, R., 2014. Evaluation of a novel accelerometer for kinetic gait analysis in dogs. Canadian Journal of Veterinary Research 78, 226-232.

Clarke, N., Fraser, D., 2016. Automated monitoring of resting in dogs. Applied Animal Behaviour Science 174, 99-102.

Clay, L., Paterson, M.B., Bennett, P., Perry, G., Phillips, C.C., 2020. Comparison of canine behaviour scored using a shelter behaviour assessment and an owner completed questionnaire, C-BARQ. Animals 10, 1797.

Cloutier, T.L., Packard, J.M., 2014. Enrichment options for African painted dogs (Lycaon pictus). Zoo biology 33, 475-480.

Clutton-Brock, J., Museum, N.H., 1999. A Natural History of Domesticated Mammals. Cambridge University Press.

Coppola, C.L., Grandin, T., Enns, R.M., 2006. Human interaction and cortisol: can human contact reduce stress for shelter dogs? Physiology & behavior 87, 537-541.

Courcier, E., Thomson, R., Mellor, D., Yam, P., 2010. An epidemiological study of environmental factors associated with canine obesity. Journal of small animal practice 51, 362-367.

Cutt, H., Giles-Corti, B., Knuiman, M., Timperio, A., Bull, F., 2008. Understanding dog owners' increased levels of physical activity: results from RESIDE. American journal of public health 98, 66-69.

Dare, P., Strasser, R., 2023. Ruff Morning? The Use of Environmental Enrichment during an Acute Stressor in Kenneled Shelter Dogs. Animals 13, 1506.

den Uijl, I., Gómez Álvarez, C.B., Bartram, D., Dror, Y., Holland, R., Cook, A., 2017. External validation of a collar-mounted triaxial accelerometer for second-by-second monitoring of eight behavioural states in dogs. Plos One 12, e0188481.

Deng, P., Iwazaki, E., Suchy, S., Pallotto, M., Swanson, K., 2014. Effects of feeding frequency and dietary water content on voluntary physical activity in healthy adult cats. Journal of animal science 92, 1271-1277.

Desforges, E., 2021. Challenges and Solutions Surrounding Environmental Enrichment for Dogs and Cats in a Scientific Environment. Animals 11, 2980.

Diederich, C., Giffroy, J.-M., 2006. Behavioural testing in dogs: A review of methodology in search for standardisation. Applied Animal Behaviour Science 97, 51-72.

Domjan, M., 2018. The essentials of conditioning and learning. American Psychological Association.

Donát, P., 1991. Measuring behaviour: the tools and the strategies. Neuroscience & Biobehavioral Reviews 15, 447-454.

Döring, D., Haberland, B.E., Bauer, A., Dobenecker, B., Hack, R., Schmidt, J., Erhard, M.H., 2016. Behavioral observations in dogs in 4 research facilities: Do they use their enrichment? Journal of Veterinary Behavior 13, 55-62.

Dow, C., Michel, K.E., Love, M., Brown, D.C., 2009. Evaluation of optimal sampling interval for activity monitoring in companion dogs. American journal of veterinary research 70, 444-448.

Dowling-Guyer, S., Marder, A., D'arpino, S., 2011. Behavioral traits detected in shelter dogs by a behavior evaluation. Applied animal behaviour science 130, 107-114.

Duffy, D.L., Kruger, K.A., Serpell, J.A., 2014. Evaluation of a behavioral assessment tool for dogs relinquished to shelters. Preventive veterinary medicine 117, 601-609.

Eerdekens, A., Callaert, A., Deruyck, M., Martens, L., Joseph, W., 2022. Dog's Behaviour Classification Based on Wearable Sensor Accelerometer Data, 2022 5th Conference on Cloud and Internet of Things (CIoT), IEEE, pp. 226-231.

Epstein, J., Dowling-Guyer, S., McCobb, E., Glotzer, C., Dodman, N.H., 2021. Addressing stress in dogs in shelters through a novel visual and auditory enrichment device. Applied Animal Behaviour Science 236, 105215.

Eyre, A.W., Zapata, I., Hare, E., Serpell, J.A., Otto, C.M., Alvarez, C.E., 2023. Machine learning prediction and classification of behavioral selection in a canine olfactory detection program. Scientific Reports 13, 12489.

Feddersen-Petersen, D.U., 2007. Social behaviour of dogs and related canids. The behavioural biology of dogs, 105-119.

Flint, E.L., Minot, E.O., Stevenson, M., Perry, P.E., Stafford, K.J., 2013. Barking in home alone suburban dogs (Canis familiaris) in New Zealand. Journal of Veterinary Behavior 8, 302-305.

Fox, M., 1971. Behaviour of wolves dogs and related canids. Dogwise Publishing.

Fox, M., 1978. The dog. Dogwise Publishing.

Friard, O., Gamba, M., 2016. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. Methods in ecology and evolution 7, 1325-1330.

Fugazza, C., Miklósi, Á., 2014. Measuring the behaviour of dogs: An ethological approach, Domestic Dog Cognition and Behavior: The Scientific Study of Canis familiaris, Springer, pp. 177-200.

Fukuzawa, M., Nakazato, I., 2015. Influence of changes in luminous emittance before bedtime on sleep in companion dogs. Journal of Veterinary Behavior 10, 12-16.

Gaines, S.A., Rooney, N.J., Bradshaw, J.W., 2008. The effect of feeding enrichment upon reported working ability and behavior of kenneled working dogs. Journal of forensic sciences 53, 1400-1404.

Garvey, M., Stella, J., Croney, C., 2016. Implementing environmental enrichment for dogs. Purdue Ext, 1-3.

Gerencsér, L., Vásárhelyi, G., Nagy, M., Vicsek, T., Miklósi, A., 2013. Identification of behaviour in freely moving dogs (Canis familiaris) using inertial sensors. PloS one 8, e77814.

German, A.J., 2006. The growing problem of obesity in dogs and cats. The Journal of nutrition 136, 1940S-1946S.

Graf, P.M., Wilson, R.P., Qasem, L., Hackländer, K., Rosell, F., 2015. The use of acceleration to code for animal behaviours; a case study in free-ranging Eurasian beavers Castor fiber. PloS one 10, e0136751.

Grier, J.W., 1984. Biology of Animal Behavior. Times Mirror/Mosby College Pub.

Hansen, B.D., Lascelles, B.D.X., Keene, B.W., Adams, A.K., Thomson, A.E., 2007. Evaluation of an accelerometer for at-home monitoring of spontaneous activity in dogs. American journal of veterinary research 68, 468-475.

Helm, J., McBrearty, A., Fontaine, S., Morrison, R., Yam, P., 2016. Use of accelerometry to investigate physical activity in dogs receiving chemotherapy. Journal of Small Animal Practice 57, 600-609.

Hennessy, M.B., Davis, H.N., Williams, M.T., Mellott, C., Douglas, C.W., 1997. Plasma cortisol levels of dogs at a county animal shelter. Physiology & behavior 62, 485-490.

Hennessy, M.B., Voith, V.L., Hawke, J.L., Young, T.L., Centrone, J., McDowell, A.L., Linden, F., Davenport, G.M., 2002. Effects of a program of human interaction and alterations in diet composition on activity of the hypothalamic-pituitary-adrenal axis in dogs housed in a public animal shelter. Journal of the American Veterinary Medical Association 221, 65-91.

Herron, M.E., Kirby-Madden, T.M., Lord, L.K., 2014. Effects of environmental enrichment on the behavior of shelter dogs. Journal of the American Veterinary Medical Association 244, 687-692.

Hetts, S., Clark, J.D., Calpin, J.P., Arnold, C.E., Mateo, J.M., 1992. Influence of housing conditions on beagle behaviour. Applied Animal Behaviour Science 34, 137-155.

Hiby, E., Rooney, N., Bradshaw, J., 2004. Dog training methods: their use, effectiveness and interaction with behaviour and welfare. Animal welfare 13, 63-69.

Hiby, E.F., Rooney, N.J., Bradshaw, J.W., 2006. Behavioural and physiological responses of dogs entering re-homing kennels. Physiology & behavior 89, 385-391.

Hoffman, C.L., Ladha, C., Wilcox, S., 2019. An actigraphy-based comparison of shelter dog and owned dog activity patterns. Journal of Veterinary Behavior 34, 30-36.

Hounslow, J., Brewster, L., Lear, K., Guttridge, T., Daly, R., Whitney, N., Gleiss, A., 2019. Assessing the effects of sampling frequency on behavioural classification of accelerometer data. Journal of experimental marine biology and ecology 512, 22-30.

Hsu, Y., Serpell, J.A., 2003. Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. Journal of the American Veterinary Medical Association 223, 1293-1300.

Hubrecht, R.C., 1993. A comparison of social and environmental enrichment methods for laboratory housed dogs. Applied Animal Behaviour Science 37, 345-361.

Hunt, R.L., Whiteside, H., Prankel, S., 2022. Effects of environmental enrichment on dog behaviour: Pilot study. Animals 12, 141.

Huntingford, F., 2012. The study of animal behaviour. Springer Science & Business Media.

Hussain, A., Ali, S., Kim, H.-C., 2022. Activity detection for the wellbeing of dogs using wearable sensors based on deep learning. IEEE Access 10, 53153-53163.

Industries, M.f.P., 2008. Code of Welfare: Dogs, In: Industries, M.f.P. (Ed.).

Jones, A.C., Gosling, S.D., 2005. Temperament and personality in dogs (Canis familiaris): A review and evaluation of past research. Applied Animal Behaviour Science 95, 1-53.

Jones, S., Dowling-Guyer, S., Patronek, G.J., Marder, A.R., Segurson D'Arpino, S., McCobb, E., 2014. Use of accelerometers to measure stress levels in shelter dogs. Journal of Applied Animal Welfare Science 17, 18-28.

Kang, O.-D., 2022. Effects of environment enrichment on behavioral problems in dogs with separation anxiety. Journal of Environmental Science International 31, 131-139.

Kienzle, E., Bergler, R., Mandernach, A., 1998. A comparison of the feeding behavior and the human–animal relationship in owners of normal and obese dogs. The Journal of nutrition 128, 2779S-2782S.

King, T., Marston, L.C., Bennett, P.C., 2012. Breeding dogs for beauty and behaviour: Why scientists need to do more to develop valid and reliable behaviour assessments for dogs kept as companions. Applied Animal Behaviour Science 137, 1-12.

Kleanthous, N., Hussain, A., Khan, W., Sneddon, J., Mason, A., 2020. Feature extraction and random forest to identify sheep behavior from accelerometer data, Intelligent Computing Methodologies: 16th International Conference, ICIC 2020, Bari, Italy, October 2–5, 2020, Proceedings, Part III 16, Springer, pp. 408-419.

Knazovicky, D., Tomas, A., Motsinger-Reif, A., Lascelles, B.D.X., 2015. Initial evaluation of nighttime restlessness in a naturally occurring canine model of osteoarthritis pain. PeerJ 3, e772.

Kogan, L.R., Schoenfeld-Tacher, R., Simon, A.A., 2012. Behavioral effects of auditory stimulation on kenneled dogs. Journal of Veterinary Behavior 7, 268-275.

Koler-Matznick, J., Brisbin, I., Feinstein, M., 2005. An Ethogram for the New Guinea Singing (Wild) Dog (Canis hallstromi). The New Guinea Singing Dog Conservation Society, USA.

Krichbaum, S., Ramey, C., Cox, E., Lazarowski, L., 2023. No bones about it: The effect of chewing on cognition in dogs. Applied Animal Behaviour Science 268, 106078.

Ksenija, I., 2019. Osteoarthritis: Degenerative Joint Disease In Dogs and Current Concept of Treatment.

Kuczaj, S., Lacinak, T., Fad, O., Trone, M., Solangi, M., Ramos, J., 2002. Keeping environmental enrichment enriching. International Journal of Comparative Psychology 15.

Kumpulainen, P., Cardó, A.V., Somppi, S., Törnqvist, H., Väätäjä, H., Majaranta, P., Gizatdinova, Y., Antink, C.H., Surakka, V., Kujala, M.V., 2021. Dog behaviour classification with movement sensors placed on the harness and the collar. Applied Animal Behaviour Science 241, 105393.

Ladha, C., Belshaw, Z., O'Sullivan, J., Asher, L., 2018. A step in the right direction: an open-design pedometer algorithm for dogs. BMC veterinary research 14, 1-10.

Ladha, C., Hammerla, N., Hughes, E., Olivier, P., Ploetz, T., 2013. Dog's life: wearable activity recognition for dogs, Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, pp. 415-418.

Ladha, C., Hoffman, C.L., 2018. A combined approach to predicting rest in dogs using accelerometers. Sensors 18, 2649.

Lascelles, B.D.X., Hansen, B.D., Thomson, A., Pierce, C.C., Boland, E., Smith, E.S., 2008. Evaluation of a digitally integrated accelerometer-based activity monitor for the measurement of activity in cats. Veterinary anaesthesia and analgesia 35, 173-183.

Lee, C.Y., Ngai, J.T.K., Chau, K.K.Y., Yu, R.W.M., Wong, P.W.C., 2022. Development of a pilot human-canine ethogram for an animal-assisted education programme in primary schools–A case study. Applied Animal Behaviour Science 255, 105725.

Leonard, J.A., 2015. Ecology drives evolution in grey wolves. Evolutionary Ecology Research 16, 461-473.

Lindig, A.M., McGreevy, P.D., Crean, A.J., 2020. Musical dogs: A review of the influence of auditory enrichment on canine health and behavior. Animals 10, 127.

Liu, Y.S., M.; Andrews, C.; Corner-Thomas, R.; Draganova, I.; Thomas, D. , 2024. Use of triaxial accelerometers and a machine learning algorithm for behavioural identification to assess the efficacy of a joint supplement in old domestic cats (Felis catus). In Proceedings of the American Society of Animal Science, Calgary.

Liu, Z., 2019. The effects of environmental enrichment on the health, behaviour, and welfare of fast-growing broilers, University of Guelph.

Lord, K., Schneider, R.A., Coppinger, R., 2016. Evolution of working dogs. The domestic dog: Its evolution, behavior and interactions with people 4, 42-66.

Marcato, M., Kenny, J., O'Riordan, R., O'Mahony, C., O'Flynn, B., Galvin, P., 2022. Assistance dog selection and performance assessment methods using behavioural and physiological tools and devices. Applied Animal Behaviour Science 254, 105691.

Markowitz, H., 1982. Behavioral enrichment in the zoo. (No Title).

Martin, K.W., Olsen, A.M., Duncan, C.G., Duerr, F.M., 2016. The method of attachment influences accelerometer-based activity data in dogs. BMC veterinary research 13, 1-6.

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, 32-38.

Menache, S., 1998. Dogs and human beings: A story of friendship. Society & Animals 6, 67-86.

Menor-Campos, D., Molleda-Carbonell, J., López-Rodríguez, R., 2011. Effects of exercise and human contact on animal welfare in a dog shelter. Veterinary Record 169, 388-388.

Michel, K.E., Brown, D.C., 2011. Determination and application of cut points for accelerometer-based activity counts of activities with differing intensity in pet dogs. American Journal of Veterinary Research 72, 866-870.

Michel, K.E., Brown, D.C., 2014. Association of signalment parameters with activity of pet dogs. Journal of nutritional science 3, e28.

Miklósi, Á., 2014. Dog behaviour, evolution, and cognition. oUp Oxford.

Moreau, M., Siebert, S., Buerkert, A., Schlecht, E., 2009. Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behaviour. Applied animal behaviour science 119, 158-170.

Morey, D.F., 1994. The early evolution of the domestic dog. American scientist 82, 336-347.

Morrison, R., Penpraze, V., Beber, A., Reilly, J., Yam, P., 2013. Associations between obesity and physical activity in dogs: a preliminary investigation. Journal of Small Animal Practice 54, 570-574.

Morrison, R., Reilly, J., Penpraze, V., Pendlebury, E., Yam, P., 2014. A 6-month observational study of changes in objectively measured physical activity during weight loss in dogs. Journal of Small Animal Practice 55, 566-570.

Muñana, K.R., Nettifee, J.A., Griffith, E.H., Early, P.J., Yoder, N.C., 2020. Evaluation of a collar-mounted accelerometer for detecting seizure activity in dogs. Journal of Veterinary Internal Medicine 34, 1239-1247.

Murtagh, K., Farnworth, M.J., Brilot, B.O., 2020. The scent of enrichment: Exploring the effect of odour and biological salience on behaviour during enrichment of kennelled dogs. Applied animal behaviour science 223, 104917.

Newberry, R.C., 1995. Environmental enrichment: Increasing the biological relevance of captive environments. Applied Animal Behaviour Science 44, 229-243.

Nielsen, B.L., Jezierski, T., Bolhuis, J.E., Amo, L., Rosell, F., Oostindjer, M., Christensen, J.W., McKeegan, D., Wells, D.L., Hepper, P., 2015. Olfaction: an overlooked sensory modality in applied ethology and animal welfare. Frontiers in Veterinary Science 2, 69.

Olsen, A.M., Evans, R.B., Duerr, F.M., 2016. Evaluation of accelerometer inter-device variability and collar placement in dogs. Veterinary Evidence 1.

Ortmeyer, H.K., Robey, L., McDonald, T., 2018. Combining actigraph link and PetPace collar data to measure activity, proximity, and physiological responses in freely moving dogs in a natural environment. Animals 8, 230.

Ostrander, E.A., Wayne, R.K., Freedman, A.H., Davis, B.W., 2017. Demographic history, selection and functional diversity of the canine genome. Nature Reviews Genetics 18, 705-720.

Owczarczak-Garstecka, S.C., Burman, O.H., 2016. Can sleep and resting behaviours be used as indicators of welfare in shelter dogs (Canis lupus familiaris)? PloS one 11, e0163620.

Part, C., Kiddie, J., Hayes, W.A., Mills, D., Neville, R.F., Morton, D., Collins, L., 2014. Physiological, physical and behavioural changes in dogs (Canis familiaris) when kennelled: Testing the validity of stress parameters. Physiology & Behavior 133, 260-271.

Pillard, P., Gibert, S., Viguier, E., 2012. Development of a 3D accelerometric device for gait analysis in dogs. Computer methods in biomechanics and biomedical engineering 15, 246-249.

Polgár, Z., Blackwell, E.J., Rooney, N.J., 2019. Assessing the welfare of kennelled dogs—A review of animal-based measures. Applied animal behaviour science 213, 1-13.

Preston, T., Baltzer, W., Trost, S., 2012. Accelerometer validity and placement for detection of changes in physical activity in dogs under controlled conditions on a treadmill. Research in Veterinary Science 93, 412-416.

Price, E.O., 1999. Behavioral development in animals undergoing domestication. Applied Animal Behaviour Science 65, 245-271.

Protopopova, A., 2016. Effects of sheltering on physiology, immune function, behavior, and the welfare of dogs. Physiology & Behavior 159, 95-103.

Pullen, A.J., Merrill, R.J.N., Bradshaw, J.W.S., 2010. Preferences for toy types and presentations in kennel housed dogs. Applied Animal Behaviour Science 125, 151-156.

Pullin, A.N., Temple, S.M., Bennett, D.C., Rufener, C.B., Blatchford, R.A., Makagon, M.M., 2020. Pullet rearing affects collisions and perch use in enriched colony cage layer housing. Animals 10, 1269.

Purugganan, M.D., 2022. What is domestication? Trends in Ecology & Evolution 37, 663-671.

Rayment, D.J., De Groef, B., Peters, R.A., Marston, L.C., 2015. Applied personality assessment in domestic dogs: Limitations and caveats. Applied Animal Behaviour Science 163, 1-18.

Rialland, P., Bichot, S., Moreau, M., Guillot, M., Lussier, B., Gauvin, D., Martel-Pelletier, J., Pelletier, J.-P., Troncy, E., 2012. Clinical validity of outcome pain measures in naturally occurring canine osteoarthritis. BMC Veterinary Research 8, 1-12.

Rooney, N., Gaines, S., Hiby, E., 2009. A practitioner's guide to working dog welfare. Journal of Veterinary Behavior 4, 127-134.

Rooney, N.J., Parr-Cortes, Z., 2023. Olfaction and Dog Welfare, Olfactory Research in Dogs, Springer, pp. 205-234.

Rushen, J., Chapinal, N., de Passilé, A., 2012. Automated monitoring of behavioural-based animal welfare indicators. Animal Welfare 21, 339-350.

Sampaio, R.A.G., Martins, Y.N.d.F., Barbosa, F.M.S., Franco, C.I.Q., Kobayashi, M.D., Talieri, I.C., 2019. Behavioral assessment of shelter dogs submitted to different methods of environmental enrichment. Ciência Rural 49, e20180181.

Sandøe, P., Palmer, C., Corr, S., Astrup, A., Bjørnvad, C.R., 2014. Canine and feline obesity: a One Health perspective. Veterinary record 175, 610-616.

Schipper, L.L., Vinke, C.M., Schilder, M.B., Spruijt, B.M., 2008. The effect of feeding enrichment toys on the behaviour of kennelled dogs (Canis familiaris). Applied Animal Behaviour Science 114, 182-195.

Schork, I., Zamansky, A., Farhat, N., de Azevedo, C.S., Young, R.J., 2024. Automated Observations of Dogs' Resting Behaviour Patterns Using Artificial Intelligence and Their Similarity to Behavioural Observations. Animals 14, 1109.

Segurson, S.A., Serpell, J.A., Hart, B.L., 2005. Evaluation of a behavioral assessment questionnaire for use in the characterization of behavioral problems of dogs relinquished to animal shelters. Journal of the American Veterinary Medical Association 227, 1755-1761.

Serpell, J., 2017. The domestic dog. Cambridge University Press.

Shaik, A.B., Srinivasan, S., 2019. A brief survey on random forest ensembles in classification model, International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, Volume 2, Springer, pp. 253-260.

Shepherdson, D.J., 2003. Environmental enrichment: past, present and future. International Zoo Yearbook 38, 118-124.

Siwak, C.T., Tapp, P.D., Zicker, S.C., Murphey, H.L., Muggenburg, B.A., Head, E., Cotman, C.W., Milgram, N.W., 2003. Locomotor activity rhythms in dogs vary with age and cognitive status. Behavioral neuroscience 117, 813.

Smit, M., Corner-Thomas, R.A., Draganova, I., Andrews, C.J., Thomas, D.G., 2024. How Lazy Are Pet Cats Really? Using Machine Learning and Accelerometry to Get a Glimpse into the Behaviour of Privately Owned Cats in Different Households. Sensors 24, 2623.

Smit, M., Ikurior, S.J., Corner-Thomas, R.A., Andrews, C.J., Draganova, I., Thomas, D.G., 2023. The Use of Triaxial Accelerometers and Machine Learning Algorithms for Behavioural Identification in Domestic Cats (Felis catus): A Validation Study. Sensors 23, 7165.

Stephen, J.M., Ledger, R.A., 2006. A longitudinal evaluation of urinary cortisol in kennelled dogs, Canis familiaris. Physiology & Behavior 87, 911-916.

Sudo, M., Nagamatsu, T., Ando, S., 2018. Does environmental enrichment increase locomotor activity in rats? Evidence from an implanted sensor device. Bulletin of the Physical Fitness Research Institute 116, 29-32.

Svartberg, K., Forkman, B., 2002. Personality traits in the domestic dog (Canis familiaris). Applied animal behaviour science 79, 133-155.

Tarou, L.R., Bashaw, M.J., 2007. Maximizing the effectiveness of environmental enrichment: Suggestions from the experimental analysis of behavior. Applied Animal Behaviour Science 102, 189-204.

Tatler, J., Cassey, P., Prowse, T.A., 2018. High accuracy at low frequency: detailed behavioural classification from accelerometer data. Journal of Experimental Biology 221, jeb184085.

Temesi, A., Turcsán, B., Miklósi, Á., 2014. Measuring fear in dogs by questionnaires: An exploratory study toward a standardized inventory. Applied Animal Behaviour Science 161, 121-130.

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V.J., Sawyer, S.K., Greenfield, D.L., Germonpré, M.B., Sablin, M.V., López-Giráldez, F., Domingo-Roura, X., 2013. Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science 342, 871-874.

Tod, E., Brander, D., Waran, N., 2005. Efficacy of dog appeasing pheromone in reducing stress and fear related behaviour in shelter dogs. Applied Animal Behaviour Science 93, 295-308.

Tudor-Locke, C., Williams, J.E., Reis, J.P., Pluto, D., 2002. Utility of pedometers for assessing physical activity: convergent validity. Sports medicine 32, 795-808.

Tudor-Locke, C.E., Myers, A.M., 2001. Challenges and opportunities for measuring physical activity in sedentary adults. Sports medicine 31, 91-100.

Valletta, J.J., Torney, C., Kings, M., Thornton, A., Madden, J., 2017. Applications of machine learning in animal behaviour studies. Animal Behaviour 124, 203-220.

van der Laan, J.E., Vinke, C.M., Arndt, S.S., 2022. Evaluation of hair cortisol as an indicator of long-term stress responses in dogs in an animal shelter and after subsequent adoption. Scientific Reports 12, 5117.

van der Laan, J.E., Vinke, C.M., Arndt, S.S., 2023. Nocturnal activity as a useful indicator of adaptability of dogs in an animal shelter and after subsequent adoption. Scientific Reports 13, 19014.

Veldkamp, F., Garcia-Faria, T.I., Witjes, V.L., Rebel, J.M., de Jong, I.C., 2023. Validation of non-invasive sensor technologies to measure interaction with enrichment material in weaned fattening pigs. Applied Animal Behaviour Science 263, 105923.

Vilà, C., Savolainen, P., Maldonado, J.E., Amorim, I.R., Rice, J.E., Honeycutt, R.L., Crandall, K.A., Lundeberg, J., Wayne, R.K., 1997. Multiple and ancient origins of the domestic dog. Science 276, 1687-1689.

Walker, J.K., Dale, A.R., D'Eath, R.B., Wemelsfelder, F., 2016. Qualitative Behaviour Assessment of dogs in the shelter and home environment and relationship with quantitative behaviour assessment and physiological responses. Applied Animal Behaviour Science 184, 97-108.

Watanabe, S., Izawa, M., Kato, A., Ropert-Coudert, Y., Naito, Y., 2005. A new technique for monitoring the detailed behaviour of terrestrial animals: a case study with the domestic cat. Applied Animal Behaviour Science 94, 117-131.

Wauthier, L.M., Williams, J.M., 2018. Using the mini C-BARQ to investigate the effects of puppy farming on dog behaviour. Applied Animal Behaviour Science 206, 75-86.

Wayne, R.K., Ostrander, E.A., 1999. Origin, genetic diversity, and genome structure of the domestic dog. BioEssays 21, 247-257.

Wells, D., 2004. The influence of toys on the behaviour and welfare of kennelled dogs. Animal Welfare 13, 367-373.

Wells, D., Graham, L., Hepper, P.G., 2002. The influence of auditory stimulation on the behaviour of dogs housed in a rescue shelter. Animal Welfare 11, 385-393.

Wells, D., Hepper, P., 1992. The behaviour of dogs in a rescue shelter. Animal Welfare 1, 171-186.

Wells, D.L., 2004. A review of environmental enrichment for kennelled dogs, Canis familiaris. Applied Animal Behaviour Science 85, 307-317.

Wells, D.L., 2006. Aromatherapy for travel-induced excitement in dogs. Journal of the American Veterinary Medical Association 229, 964-967.

Wells, D.L., 2009. Sensory stimulation as environmental enrichment for captive animals: A review. Applied Animal Behaviour Science 118, 1-11.

Westgarth, C., Ladha, C., 2017. Evaluation of an open source method for calculating physical activity in dogs from harness and collar based sensors. BMC veterinary research 13, 1-7.

Whitham, J., Miller, L., 2016. Using technology to monitor and improve zoo animal welfare. Animal Welfare 25, 395-409.

Wiener, P., Haskell, M.J., 2016. Use of questionnaire-based data to assess dog personality. Journal of Veterinary Behavior 16, 81-85.

Wynne, C.D., 2021. Dogs'(Canis lupus familiaris) behavioral adaptations to a human-dominated niche: A review and novel hypothesis. Advances in the Study of Behavior 53, 97-162.

Yam, P., Penpraze, V., Young, D., Todd, M., Cloney, A., Houston-Callaghan, K., Reilly, J.J., 2011. Validity, practical utility and reliability of Actigraph accelerometry for the measurement of habitual physical activity in dogs. Journal of Small Animal Practice 52, 86-91.

Yashari, J.M., Duncan, C.G., Duerr, F.M., 2015. Evaluation of a novel canine activity monitor for at-home physical activity analysis. BMC Veterinary Research 11, 1-7.

Young, R.J., 1997. The importance of food presentation for animal welfare and conservation. Proceedings of the Nutrition Society 56, 1095-1104.

Appendix 1

Confusion matrices and performance statistics of predictive models

Appendix 1. Confusion matrices of predictive models

Appendix 1.1 Confusion matrix of predicted and observed behaviours of Model 1 presented as percentages (%). Correct/target categorisations by the model are indicated in cells highlighted green and incorrect categorisations >10% are in cells that have been highlighted orange. Abbreviation: Lateral recumbency (Lateral R.).

Model Prediction	Observed behaviour															
	Barking	**Defecating**	**Drinking**	**Eating**	**Jumping**	**Lateral R.**	**Lying (asleep)**	**Lying (alert)**	**Running**	**Scratching**	**Sitting**	**Sniffing**	**Standing**	**Trotting**	**Urinating**	**Walking**
Barking	83.89	0.00	0.00	2.94	14.81	0.10	0.19	0.96	17.36	5.00	2.84	0.00	3.46	5.39	0.00	2.07
Defecating	0.00	70.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
Drinking	0.00	0.00	70.68	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.07	0.00	0.44
Eating	0.00	0.00	0.00	77.94	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.07	0.04	0.02	0.00	0.04
Jumping	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
L. recumbency	0.00	0.00	0.00	0.00	0.00	94.13	0.19	0.13	0.00	0.00	0.08	0.00	0.02	0.00	0.00	0.04
Lying-asleep	0.05	5.88	0.00	0.00	0.00	0.48	85.40	1.34	0.16	0.00	0.16	0.00	0.28	0.13	1.29	0.71
Lying-alert	2.88	0.00	0.88	0.00	7.41	3.56	9.67	84.96	5.41	5.63	31.15	0.65	5.40	3.16	0.00	3.53
Running	0.28	0.00	0.00	0.00	3.70	0.00	0.00	0.00	12.42	0.00	0.00	0.07	0.41	0.27	1.29	0.00
Scratching	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	60.63	0.00	0.00	0.01	0.00	0.00	0.00
Sitting	0.57	0.00	0.44	0.00	3.70	0.38	0.38	2.90	2.71	1.25	37.02	0.43	1.88	0.69	0.00	1.41
Sniffing	0.09	5.88	6.35	14.71	3.70	0.58	1.08	0.86	1.59	6.88	1.46	92.85	2.79	1.70	21.94	11.90
Standing	5.43	5.88	10.50	4.41	29.63	0.67	2.68	7.79	54.78	12.50	24.91	3.02	64.32	28.48	18.06	33.58
Trotting	6.57	5.88	7.88	0.00	33.33	0.10	0.23	0.64	5.25	6.88	1.22	1.08	18.51	58.38	5.16	21.55
Urinating	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	48.39	0.00
Walking	0.24	5.88	3.28	0.00	3.70	0.00	0.14	0.39	0.32	1.25	1.17	1.83	2.71	1.72	3.87	24.72
Total observations (s)	**4,234**	**34**	**457**	**272**	**27**	**1,040**	**2,130**	**8,444**	**628**	**160**	**2,469**	**2,785**	**8,292**	**5,478**	**155**	**2,269**

Appendix 1.2 Confusion matrix of predicted and observed behaviours of Model 2 presented as percentages (%). Correct/target categorisations by the model are indicated in cells highlighted green and incorrect categorisations >10% are in cells that have been highlighted orange. Abbreviation: Lateral recumbency (Lateral R).

Model prediction	Observed behaviour												
	Barking	Drinking	Eating	Lateral R.	Lying (asleep)	Lying (alert)	Running	Scratching	Sitting	Sniffing	Standing	Trotting	Walking
Barking	83.56	0.00	2.94	0.29	0.00	0.83	14.97	3.13	2.96	0.00	3.49	5.59	2.07
Drinking	0.00	71.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.13	0.48
Eating	0.00	0.00	79.41	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.04	0.04	0.13
L. recumbency	0.00	0.00	0.00	94.13	0.05	0.08	0.00	0.00	0.16	0.00	0.01	0.00	0.13
Lying-asleep	0.00	0.00	0.00	0.38	85.31	1.30	0.16	1.25	0.20	0.07	0.19	0.11	0.62
Lying-alert	3.16	0.66	0.00	3.56	10.47	84.81	5.41	6.25	35.68	0.50	5.73	2.76	3.35
Running	0.14	0.00	0.00	0.00	0.00	0.00	11.94	0.00	0.00	0.00	0.37	0.33	0.18
Scratching	0.00	0.00	0.00	0.00	0.00	0.06	0.00	71.88	0.04	0.00	0.00	0.00	0.00
Sitting	0.61	0.88	0.00	0.38	0.14	3.19	2.39	1.25	33.50	0.22	2.34	0.64	1.81
Sniffing	0.14	6.13	13.24	0.48	0.75	0.75	1.27	0.00	1.17	93.57	2.76	1.52	11.46
Standing	5.05	6.35	3.68	0.67	3.10	7.83	55.10	13.75	23.94	2.62	63.57	27.89	33.19
Trotting	7.09	12.69	0.00	0.00	0.19	0.90	8.60	1.88	1.13	1.01	18.96	59.22	22.65
Walking	0.24	2.19	0.74	0.10	0.00	0.26	0.16	0.63	1.22	1.72	2.36	1.79	23.93
Total observations (s)	**4,234**	**457**	**272**	**1,040**	**2,130**	**8,444**	**628**	**160**	**2,469**	**2,785**	**8,292**	**5,478**	**2,269**

Appendix 1.3 Confusion matrix of predicted and observed behaviours of Model 3 presented as percentages (%). Correct/target categorisations by the model are indicated in cells highlighted green and incorrect categorisations >10% are in cells that have been highlighted orange. Abbreviation: Lateral recumbency (Lateral R.).

Model prediction	Observed behaviour										
	Barking	Drinking	Eating	Lateral R.	Locomotion	Lying (asleep)	Lying (alert)	Scratching	Sitting	Sniffing	Standing
Barking	82.43	0.44	2.21	0.19	4.03	0.19	0.68	3.13	2.55	0.29	2.77
Drinking	0.00	71.33	0.74	0.00	0.13	0.00	0.00	0.00	0.00	0.07	0.10
Eating	0.00	0.00	79.41	0.00	0.02	0.00	0.01	0.00	0.00	0.07	0.01
L. recumbency	0.00	0.00	0.00	93.56	0.00	0.05	0.07	0.00	0.00	0.00	0.01
Locomotion	9.78	16.63	8.09	0.58	66.66	0.56	2.40	14.38	3.93	4.09	31.14
Lying-asleep	0.09	0.00	0.00	0.58	0.14	86.76	1.42	1.25	0.12	0.07	0.14
Lying-alert	3.71	0.00	0.74	3.85	3.08	8.59	84.13	8.75	33.54	0.57	4.90
Scratching	0.00	0.00	0.00	0.00	0.00	0.00	0.06	64.38	0.04	0.00	0.00
Sitting	0.43	1.31	0.00	0.58	1.10	0.38	3.33	2.50	34.95	0.29	1.86
Sniffing	0.14	3.06	6.62	0.29	3.83	0.75	0.68	2.50	1.26	92.82	2.52
Standing	3.42	7.22	2.21	0.38	21.00	2.72	7.22	3.13	23.61	1.72	56.55
Total Observations(s)	**4,234**	**457**	**272**	**1,040**	**8,377**	**2,130**	**8,444**	**160**	**2,469**	**2,785**	**8,292**

Appendix 1.4 Confusion matrix of predicted and observed behaviours of the adjusted model utilised in **Chapter 3**, derived from model 4 observed in **Chapter** 2. Presented as percentages (%). Correct/target categorisations by the model are indicated in cells highlighted green and incorrect categorisations >10% are in cells that have been highlighted orange.

Model prediction	Observed behaviour								
	Barking	Defecating	Drinking	Eating	Locomotion	Resting	Resting (Alert)	Sniffing	Standing
Barking	81.95	0.00	0.00	0.00	4.05	0.06	1.18	0.07	2.50
Defecating	0.00	64.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Drinking	0.00	0.00	72.21	0.74	0.10	0.00	0.00	0.00	0.13
Eating	0.00	0.00	0.00	75.00	0.04	0.00	0.00	0.07	0.04
Locomotion	10.34	5.88	16.19	5.15	65.85	0.63	3.08	3.59	30.82
Resting	0.09	0.00	0.00	0.00	0.16	87.73	1.23	0.22	0.26
Resting Alert	4.27	0.00	0.00	0.74	5.18	9.52	84.75	0.50	9.46
Sniffing	0.05	17.65	7.88	13.97	3.72	0.50	0.85	92.96	2.47
Standing	3.30	11.76	3.72	4.41	20.91	1.55	8.91	2.58	54.31

Appendix 1.5 Overall model performance statistics for the adjusted model used in **Chapter 3**.

Model Performance	Barking	Defecating	Drinking	Eating	Locomotion	Resting	Resting Alert	Sniffing	Standing
Sensitivity	0.82	0.65	0.72	0.75	0.66	0.88	0.85	0.93	0.54
Specificity	0.98	1.00	1.00	1.00	0.88	0.99	0.94	0.98	0.90
Pos Pred value	0.84	1.00	0.94	0.96	0.61	0.94	0.84	0.79	0.60
Neg Pred value	0.98	1.00	1.00	1.00	0.90	0.99	0.94	0.99	0.88
Prevalence	0.11	0.00	0.01	0.01	0.22	0.08	0.29	0.07	0.22
Detection Rate	0.09	0.00	0.01	0.01	0.14	0.07	0.24	0.07	0.12
Detection Prevalence	0.11	0.00	0.01	0.01	0.24	0.08	0.29	0.09	0.20
Balanced Accuracy	0.90	0.82	0.86	0.87	0.77	0.94	0.89	0.95	0.72