

The Automatic Classification of Canine State

Jack O'Sullivan



A thesis presented for the degree of
Doctor of Philosophy

The Institute of Neuroscience
Faculty of Medical Science
Newcastle University

2nd October 2020

Abstract

Osteoarthritis is a prevalent disease among domestic dogs which, even when well-managed, often causes bouts of chronic pain and a lesser quality of life. Despite a lack of training dog owners are relied upon to recognise the signs of pain or illness in their animals. This often leads to treatment being sought later than would be ideal, resulting in the unnecessary and avoidable suffering of their dogs. This can be further compounded by the qualitative nature of lameness assessment performed by veterinarians. The difficulty of which is further exacerbated when symptoms are subtle, and the disease is in its early stages. This thesis investigates the use of remote, animal-borne, tri-axial accelerometers to supplement the welfare information available to both caregivers and veterinarians. Published acceleration-derived measures, of both the time and frequency domains, common within human and non-human animal accelerometer research, are assessed for their potential as daily and weekly identifiers of osteoarthritic lameness. The suitability of identified measures was evaluated using both Principal Component Analysis based feature selection and logistic linear models. The results of this process highlighted a potential link between both the level and entropy of an animals overall weekly activity with the occurrence of osteoarthritis. It also provided insight into areas of further development and established the complexity of the task of recognising lameness from acceleration data. A behaviour-based methodology was established hybridising techniques used across wildlife ecology deployments, existing veterinary assessment of lameness and, the assessment of human gait impacted by both physical illness and neurodegeneration. This led to the development of a methodology focussing on the identification of behaviours, starting with canine postural state, to provide context as to the daily activities of the subject. Two distinct approaches to postural recognition were assessed both employing machine learning techniques with a focus on the interpretability of results. The first, examined the identification of 6 postural transitions, similar to methods established in human accelerometer assessments, using linear discriminant analyses at 3 different sliding window lengths. The inclusion of an empirical cumulative distribution function representation was also assessed. The results suggested that the isolation of transitional periods from among non-transitional periods was difficult and there was high confusion between the transitions themselves. The second examined the identification of the postures themselves alongside the occurrence of locomotion during the standing posture. Linear discrimination analyses

were once again used due to the interpretability of the method and the simplicity of its implementation. The effects of pre-processing techniques and differing posture groupings were also explored. The findings suggested a binary decision tree approach was the most effective mechanism and that the application of pre-processing techniques to clean data caused a distinct negative impact that requires forethought as to the potential costs and benefits of their use. Standing was the most easily identified, perhaps due to its prominence, and the further classification of locomotion from among standing periods was ineffective. To further supplement the postural methods of identifying osteoarthritis an investigation of the remote monitoring of circadian rhythm was established. This is of interest due to prior results highlighting the potential relationship of activity entropy and level with lameness and the reports of sleep disruption by human chronic pain sufferers. Features relating to the length and frequency of both resting and active bouts were used in logistic regression models to establish their relationship to the presence or absence of osteoarthritis. Minor disruption was observed to the amplitude of activity frequencies within osteoarthritic dogs consistent with prior findings. However, further work is needed to disentangle this effect from that of advanced age, a possible confound. The potential of remote sensing technologies is shown but further development of methodologies is required. A combination of the described approaches, with the refinements highlighted within this thesis, could further improve their efficacy and should be investigated. A behaviour based, transparent and fully interpretable monitor of lameness, pain, and/or welfare could prove valuable to the early and effective treatment of canine osteoarthritis and should be pursued further.

Acknowledgements

This thesis would not exist without the hard work, and eternal patience of my supervisor Professor Lucy Asher who devised the initial idea, attained the funding, and managed to keep me from going down too many tangents (most of time). I'd also like to thank Dr Cassim Ladha for his support and advice in the early stages of the project and provision of the devices used. Finally I would like to thank the remainder of the members of the Asher Behaviour Lab who each contributed in varying ways to help me finish this thesis.

The many dogs and their owners who comprise the DogBox sample were each a joy to work with and helped to show me just how passionate people can be about the health and well-being of their animals and of improving their ability to understand each other. Without their effort and willingness to take part this project would not have been possible

Last of all I would like thank my parents and brother for their support through my continued reluctance to, in their words, "get a real job", and to Yunyi whose support has been crucial to keeping me going through any and all rough patches.

Contents

Abstract	i
1 Introduction and Literature Review	1
1.1 Welfare	1
1.2 Measurement of Welfare	3
1.2.1 Physical Welfare	5
1.2.1.1 Animal Attached Monitors	6
1.2.1.2 Vision-based Monitors	9
1.2.2 Cognitive Welfare	11
1.2.2.1 Animal Attached Monitors	11
1.2.2.2 Vision-based Monitors	13
1.3 Models of Welfare	15
1.4 Conclusions	17
2 Activity, Energy and an Assessment of Extant Measures	21
2.1 Abstract	21
2.2 Introduction	22
2.3 Materials and Methods	24
2.3.1 Data Collection	24
2.3.2 Data Processing	25
2.3.3 Feature Reduction and Model Construction	30
2.4 Results	32
2.4.1 Weekly Summarised Acceleration Models	32
2.4.2 Repeated Measures Daily Acceleration Models	35
2.4.3 Exploration of Variance in the Osteoarthritic Condition	39
2.4.4 Age Relatedness	40
2.5 Discussion	42
2.6 Conclusion	44
3 The Automatic Categorisation of Postural Transitions of the Domestic Dog	47
3.1 Abstract	47
3.2 Introduction	48

3.3	Materials and Methods	50
3.3.1	Data Collection	50
3.3.2	Video Annotation	50
3.3.3	Accelerometer Data Processing	50
3.3.4	Feature Extraction	53
3.3.5	Model Implementation and Feature Selection	54
3.4	Results	59
3.4.1	Postural Transition Identification	59
3.4.2	Lie to Stand Transition Identification	63
3.4.3	Sit to Stand Transition Identification	65
3.4.4	Stand to Lie Transition Identification	68
3.4.5	Stand to Sit Transition Identification	72
3.4.6	Lie to Stand and Stand to Lie Combined Class Identification . . .	76
3.4.7	Sit to Stand and Stand to Sit Combined Class Identification . . .	79
3.4.8	Directional Transition Identification	83
3.5	Discussion	86
3.6	Conclusion	92

4	The Automatic Categorisation of Postural States and Locomotion of the Domestic Dog	93
4.1	Abstract	93
4.2	Introduction	94
4.3	Materials and Methods	96
4.3.1	Data Collection and Annotation	96
4.3.2	Accelerometer Data Processing	97
4.3.3	Feature Extraction	98
4.3.4	Model Implementation and Feature Selection	100
4.4	Results	101
4.4.1	One VS All Postural Identification	101
4.4.1.1	Standing Detection	101
4.4.1.2	Sitting Detection	104
4.4.1.3	Lying Detection	107
4.4.2	Locomotion Detection	110
4.4.3	Paired Postural Binary Decisions	113
4.4.3.1	Standing VS Sitting	113
4.4.3.2	Standing VS Lying	116
4.4.3.3	Lying VS Sitting	119
4.5	Discussion	122
4.6	Conclusion	127

5	Circadian Rhythm and Nocturnal Activity as Canine Welfare Measures	129
5.1	Abstract	129
5.2	Introduction	130
5.3	Materials and Methods	132
5.3.1	Data Collection and Preprocessing	132
5.3.2	Statistical Analyses	134
5.4	Results	136
5.4.1	Exploration of Variance in the Osteoarthritic Condition	143
5.4.2	Age Relatedness	144
5.5	Discussion	145
5.6	Conclusion	148
6	Discussion	151
6.1	Overview of Findings	151
6.2	Attachment Method	153
6.3	The DogBox Sample	155
6.4	Osteoarthritis as a model of reduced welfare	156
6.5	The role of Technology in animal welfare	158
6.6	Future work	159
6.7	Conclusions	160
	References	161
	Appendices	183
	Appendix A Sample Details	185
	Appendix B Read.ELAN function code	187
	Appendix C Unknown annotation generation function code	189
	Appendix D Rotation correction function code	191
	Appendix E Feature calculation function code	199
	Appendix F ECDF calculation function code	213
	Appendix G Calculation of Fisher's g Statistic	217

List of Figures

2.1	Location of collar-mounted sensor and device axes in relation to the dog	26
2.2	Scree plot showing the explained variance of the first 10 principal components of the week summarised feature-set	32
2.3	Scree plot showing the explained variance of the first 10 principal components of the daily summarised feature-set resulting from an MFA analysis . . .	35
2.4	Visualisation of the variation in group LOAD score	39
2.5	Regression of LOAD Score against A) Mean Weekly ODBA and B) 7 day mean of ODBA entropy	40
2.6	Visualisation of the variation in group ages	41
2.7	The relationship between LOAD Score and Age in years	41
2.8	Visualisation of the relationships between dog age and A) the mean weekly ODBA or B) the 7 day mean of daily ODBA entropy	42
3.1	The frequency of inclusion of features within cross-validation of Postural transition classification models. The 2.5 threshold is marked.	62
3.2	The frequency of inclusion of features within cross-validation of Lie to Stand transition classification models. The 2.5 threshold is marked. . .	65
3.3	The frequency of inclusion of features within cross-validation of Sit to Stand transition classification models. The 2.5 threshold is marked. . .	68
3.4	The frequency of inclusion of features within cross-validation of Stand to Lie transition classification models. The 2.5 threshold is marked.	71
3.5	The frequency of inclusion of features within cross-validation of Stand to Sit transition classification models. The 2.5 threshold is marked.	75
3.6	The frequency of inclusion of features within cross-validation of Lie to Stand and Stand to Lie Combined Class classification models. The 2.5 threshold is marked.	79
3.7	The frequency of inclusion of features within cross-validation of Sit to Stand and Stand to Sit Combined Class classification models. The 2.5 threshold is marked.	82
3.8	The frequency of inclusion of features within the cross-validation of the classification of directional combined transitions. The 2.5 threshold is marked.	86

4.1	The frequency of inclusion of features within cross-validation of Postural state classification models. The 2.5 frequency threshold is marked. . . .	102
4.2	The frequency of inclusion of features within cross-validation of the One vs All Sitting postural state classification models. The 2.5 frequency threshold is marked.	106
4.3	The frequency of inclusion of features within cross-validation of the One vs All Lying postural state classification models. The 2.5 frequency threshold is marked.	109
4.4	The frequency of inclusion of features within cross-validation of the Locomotion period classification models. The 2.5 frequency threshold is marked.	112
4.5	The frequency of inclusion of features within cross-validation of the combined Standing and Sitting postural state class classification models. The 2.5 frequency threshold is marked.	115
4.6	The frequency of inclusion of features within cross-validation of the combined Standing and Lying postural state class classification models. The 2.5 frequency threshold is marked.	118
4.7	The frequency of inclusion of features within cross-validation of the classification models attempting to separate the Sitting and Lying postural states from each other. The 2.5 frequency threshold is marked.	121
4.8	Arrangement of binary classification hierarchy suggested by results . . .	125
5.1	The mean 24 hour ODBA of Control and Osteoarthritic groups	133
5.2	Frequency domain representations of the mean 24 hour ODBA of the A) Control and B) Osteoarthritic groups. Frequency represented as Cycles per Day (CPD).	134
5.3	A series of plots indicating the inter and intra-group variance of significant, uncorrelated features. A) The Amplitude of the mean ODBA, B) the number of identified X axis harmonics, C) the amplitude of the 12 hour period ultradian rhythm.	141
5.4	A selection of plots providing an indication of class separability when using the selected features as predictors of group. A) The amplitude of the 12 hour period ultradian rhythm against the number of identified X axis harmonics, B) the amplitude of the 12 hour period ultradian rhythm against the Amplitude of the mean ODBA, C) the number of identified X axis harmonics against the Amplitude of the mean ODBA.	142
5.5	Regression of LOAD score against A) the amplitude of the 12 hour ultradian rhythm of the mean ODBA signal, B) the number of significant X axis harmonics, C) the amplitude of the mean ODBA signal	143

5.6	The relationships between dog age and A) the amplitude of the 12 hour ultradian rhythm of the mean ODBA signal, B) the number of significant X axis harmonics, C) the amplitude of the mean ODBA signal	144
-----	---	-----

List of Tables

1.1	Summary of a selection of potential remote monitoring methods, their uses, benefits and limitations	18
2.1	List of acceleration data transformations and relevant references.	23
2.2	List of features, their descriptions and the data transformations it is appropriate to calculate them from	29
2.3	Contributions of the highest 3 contributing features for each of the first 6 principal components	33
2.4	Results of the univariate logistic models for each of the 6 selected features	34
2.5	The intercorrelations of the 6 selected features to allow for redundant feature removal	34
2.6	Results of a multivariate logistic model including all selected features . .	34
2.7	The mean contributions of the top 5% contributing features for each of the first 7 component dimensions when the MFA contribution per day output has been summarised per dog	36
2.8	Results of the mixed logistic models, each including a random term accounting for individual differences, for each of the 7 selected features	37
2.9	The intercorrelations of the 7 selected features to allow for removal of highly correlated variables	38
2.10	Results of a mixed multivariate logistic model including all selected features and a random effect accounting for the unaccounted variance between dogs	39
3.1	Ethogram of annotated postural transitions	51
3.2	Frequencies of postural transitions and the randomly sampled "unknown" class	55
3.3	The performance statistics of the identification of postural transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted.	61
3.4	The performance statistics of the identification of lie to stand transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted	64

3.5	The performance statistics of the identification of sit to stand transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted	67
3.6	The performance statistics of the identification of stand to lie transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted	70
3.7	The performance statistics of the identification of stand to sit transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted	74
3.8	The performance statistics of the identification of the combined class of lie to stand and stand to lie transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted	78
3.9	The performance statistics of the identification of the combined class of sit to stand and stand to sit transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted	81
3.10	The performance statistics of the identification of the directional grouped transition classes using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted	85
4.1	The performance statistics of the identification of the Standing postural state from among all other recorded postures using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.	103
4.2	The performance statistics of the identification of the Sitting postural state from among all other recorded postures using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.	105
4.3	The performance statistics of the identification of the Lying postural state from among all other recorded postures using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.	108
4.4	The performance statistics of the identification of periods of locomotion from among all standing posture periods using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.	111

4.5	The performance statistics of the separation of the Standing and Sitting postural states using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.	114
4.6	The performance statistics of the separation of the Standing and Lying postural states using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.	117
4.7	The performance statistics of the separation of the Lying and Sitting postural states using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.	120
5.1	The frequencies (CPD) and period length (hours) of harmonic components of the mean ODBA of Control dogs found to be significantly distinct through the use of Fisher's g statistic.	136
5.2	The frequencies (CPD) and period length (hours) of harmonic components of the mean ODBA of Osteoarthritic dogs found to be significantly distinct through the use of Fisher's g statistic.	137
5.3	The results of logistic regression models including each feature alongside a random effect accounting for the individual dog.	138
5.4	The correlation of features revealed to be significantly related to osteoarthritic group when included in logistic regression models with only a random term for individual	140
A.1	Sample details of the Osteoarthritic dog group	185
A.2	Sample details of the Control dog group	186

Chapter 1

Introduction and Literature Review

Our understanding of animal sentience and welfare continues to deepen as animals continue to play fundamental roles within our society. Within a research context practical concerns regarding the effects of poor conditions, or poor physical or mental health on subjects have led to the continual improvement of standards (Broom, 2016; Poole, 1997). For the public, the close contact with domestic animals, the effective communication of sentience research, and continuing popularity of ethical lifestyles has led to a heightened awareness of animal welfare. In research and livestock settings this has resulted in a widely shared desire for improved standards of care, and regulation within animal agriculture, where consumers are seen to pay more for products with perceived higher welfare practices, and research, where animal technologists often go beyond the minimum required by legislation in the care of their animals (Broom, 2010; Chen & Hong, 2019; Coleman, 2007; Degeling & Johnson, 2015; Greenhough & Roe, 2018; Janssen *et al.*, 2016; Knight & Barnett, 2008). In domestic settings this has led owners to seek out ways of maintaining higher standards for their own animals and understanding what it is their animals need and want. Technology could prove useful for this task and below I review the potential of novel technological solutions to the measurement of different aspects of welfare.

1.1 Welfare

The concept of animal welfare has long been related to an organisms physical health but has since evolved to also encapsulate cognitive well-being (Broom, 2011; Dawkins, 2017). This modern combined meaning, based within an ever-improving understanding of animal sentience, has developed alongside legislation regarding non-human animal treatment (Broom, 2011). Within the United Kingdom the the Brambell Report was published to address increasing concerns over, and awareness of, animal welfare (Harrison, 2013). The report led to the establishment of The Five Freedoms by the Farm Animal Welfare Council (Brambell *et al.*, 1965; Whitham & Wielebnowski, 2013). These five basic requirements; Freedom from Hunger and Thirst, Freedom from Discomfort, Freedom from Pain, Injury, or Disease, Freedom to Express Normal Behaviour, and

the Freedom from Fear, form the basis of modern legislation in the United Kingdom as well as being used by international and intergovernmental organisations such as The World Organisation for Animal Health (Mellor, 2016; Office International des Epizooties, 2015). The Five Freedoms continue to be applied however, modern definitions of welfare attempt to also incorporate the recognition, rectification, and prevention of conditions which violate these freedoms. Additionally, many modern approaches also incorporate the assessment of positive welfare measures in addition to the more established negative measures (Mellor, 2016; Webster, 2016; Yeates & Main, 2008).

The later realisation that the Five Freedoms cannot be fully addressed, and that doing so can interfere with the freedom to perform natural behaviours, has resulted in an increased focus by researchers on the provision of resources for the satisfaction of the freedoms and a renewed definition of welfare which instead concerns the interactions of the organisms with their environment (Mellor, 2016). Through this subtle pivot natural behaviours are encouraged alongside the satisfaction of animal needs. Researchers gauge an individual's state through its ability to cope with an environment, and the behaviours employed by the individual to do so (Broom, 1986; Mellor, 2016). Where an individual is unable to adapt the resultant distress and suffering would, if severe or over a long enough timespan, result in a degradation of its physical or cognitive well-being (Broom, 2016; Dawkins, 1990; Mellor, 2016; Webster, 2016).

This approach requires an acceptance of a degree of animal sentience that, although now generally accepted, has been controversial, particularly in the consideration of invertebrate species (Dawkins, 2017; Duncan, 2006). To successfully satisfy the requirements of sentient species an understanding of how organisms reason and think is required and remains difficult to address completely (Dawkins, 2015, 2017). This complication required the development of adjustments to the interpretation of animal welfare that remain useful regardless of the subjective opinion of the researcher regarding animal sentience. To achieve this Dawkins (2004, 2017) proposed a 2-pronged approach that incorporates both the physical health, similar to prior definitions, and the cognitive health, by asking if the animal has what it wants. This allows assessors to account for the effects of non-physical aspects of an environment, and for the attachment of valence, how positive or negative a stimulus is, to welfare measures. For example, mild thirst in the absence of a water source. The animal experiences discomfort and a negative cognitive state due to the unsatisfied desire for water but will not experience a negative physical state unless the want remains unsatisfied for a prolonged period.

Despite improving the generalisability of the concept of welfare regardless of where researchers stand on the topic of non-human animal sentience, this approach creates issues in how best to account for both facets, particularly the cognitive aspect, objectively. This is of note when considering the increase in interest in automated systems of whole-picture welfare monitoring which would be required to infer want alongside

signs of physical stress (Botreau *et al.*, 2007). To allow a focus on methodology rather than the presence or absence of sentience however this thesis will use Dawkins (2004) definition while acknowledging the difficulties in inferring want.

1.2 Measurement of Welfare

To measure welfare and account for the intricacies of these definitions is therefore a challenge. This is further exasperated when measurement and monitoring attempts are made on large numbers of animals as are commonly found in agricultural or research settings. Another complicating factor is the communication of these measures to vastly different audiences (*e.g.* caregivers, veterinarians, researchers, etc.) with unique needs and understanding.

Visual assessment methods focused on the development of rating scales (scores with written and/or visual descriptors) have long been employed for the clinical assessment of health, and pain (Rutherford, 2002). Such scales are often used in agricultural settings where they attempt the rapid recognition of problems without the need to fully assess or diagnose the specific issues of every individual being observed. For example, the gait of commercially farmed broiler chickens can be scored to assess leg health with a scale ranging from 0 (healthy legs) to 5 (unable to walk) (Kestin *et al.*, 1992). This allows the rapid diagnosis of problem birds in a flock. But, it achieves this by disregarding the underlying causes of the lameness, which can vary widely, in favour of assessment speed (Dawkins, 2004). The subjectivity of such methodology means their use requires the training of assessors and the cross-validation of results to ensure minimal mislabelling (Butterworth *et al.*, 2007; Dawkins *et al.*, 2009; Engel *et al.*, 2003). The additional requirement for assessors to be either present, or able to watch extended periods of video leads to a large time investment that is also impractical. The stocking densities of agricultural settings also often requires the restriction of observations to a smaller proportion of the entire flock which reduces the representativeness of the data. Similarly, gait examination of domestic animals by veterinarians is often done by visual assessment (Belshaw *et al.*, 2016; Quinn *et al.*, 2007). This situation is the inverse of the above as only one animal is being observed, and observations are not cross-validated between multiple trained observers. Additionally, the period of observation is reduced and, unless symptoms are severe, could be missed if not exhibited, or is concealed, by the individual during the examination (Belshaw *et al.*, 2016).

The visual assessment of behaviours and behavioural patterns has also been used extensively to assess welfare (Dawkins, 2004; Paul *et al.*, 2005; Titulaer *et al.*, 2013). Choice and preference measures are among the most logical when attempting to identify "what animals want" directly but are often difficult to implement in real-world

settings (Dawkins, 2004). To address this, methodologies used in free-ranging animal behavioural ecology have been adapted. Such methods involve the observation of animal distribution in relation to environmental factors and conspecifics. These data are then used to gain insight into the question of "what animals want" (Dawkins *et al.*, 2003; Keeling, 1995; Morris *et al.*, 2001). However, such observations can only be used to make adjustments to what, and how many, resources are provided to animals, which differs between individuals and is not responsive to current welfare need. Another, complementary method is the observation of behaviours which typically characterise positive or negative reactions to stimuli. The construction of an ethogram with valencies, either positive or negative, tied to each behaviour then allows the assessment of the effect of stimuli. The vocalisations of chicks undergoing isolation stress is one example of such behaviours (Feltenstein *et al.*, 2002). The limitation of this approach is the occurrence of the same or similar behaviours regardless of stimulus valency and the variation in responses by individuals to the same stimuli. In kennel dogs both increased locomotion and decreased activity have been observed as indicators of chronic stress. However, such behavioural markers show large variation between individuals (Beerda *et al.*, 2000; Part *et al.*, 2014). Furthermore, findings have indicated that behavioural responses are tied to the previous experience of the stimulus by the individual (Rooney *et al.*, 2007). This suggests that such methods are unreliable in isolation and necessitates the frequent combination of measures to gain a true estimation of welfare. Additionally, it is not possible for the observer to be blinded to the environment, conditions, or treatments when assessing welfare through observation. This may introduce bias either through knowledge of the environment or through the introduction of anthropocentric bias, such as more attention being paid to faces which may be of less importance to the species being observed (Leach *et al.*, 2011).

In response to these issues with manual, visual observation animal welfare assessment and research has turned more and more towards technological solutions. This trend has resulted in a wealth of research regarding either instrumented or invasive methods of observation of animal condition and behaviour (Brown *et al.*, 2013; Cooke *et al.*, 2004; Jukan *et al.*, 2017; Kramer & Kinter, 2003). Central to this research is the removal of subjectivity of assessments relating to health and welfare, the capture of data imperceptible using traditional assessment methods, the removal of assessor bias, and reductions in time and training requirements (Brown *et al.*, 2013; Tuytens *et al.*, 2014; Wilson *et al.*, 2014).

Invasive methods, which require minor surgery for the implantation of devices, can monitor a range of physiological features such as; heart rate, blood flow, respiratory rate, pH and temperature. These can then be used to infer physical health and have been shown to co-vary with features of cognitive well-being such as stress (Kramer & Kinter, 2003). However, such measurements are often difficult to interpret and attribute

to state, particularly as they have frequently been found to co-vary with non-welfare related features. For example, heart rate escalates in times of fear, an affective state of negative valence, but also during experiences and behaviours, such as play or sex, which have a positive valence (Dawkins, 2015; Titulaer *et al.*, 2013). The additional requirements of surgery and recovery times make invasive methods impractical for implementation on a significant scale and undesirable for owners of animals in domestic settings. Furthermore, the act of surgically implanting a device itself, often requiring the administration of anaesthesia, could cause discomfort and pain which would result in a significant drop in welfare post-surgery (Popova *et al.*, 2017).

There has been a variety of non-invasive devices developed to automatically monitor the behaviours of animals and, by extension, infer their welfare state (Rushen *et al.*, 2012; Whitham & Miller, 2016). The removal of a human observer allows the capture of large data-sets over equally large time-scales which would usually be impractical, or impossible, to manually observe. This, coupled with the use of computational analysis methods, allows the collection and rapid processing of large data-sets of increased quality and quantity. These improvements in turn open up the possibility of employing new techniques geared towards the analysis of large, longitudinal and/or high resolution data-sets. The absence of observers can lead to the recording of behaviour that would not be performed with humans present or that occurs in inaccessible areas (Whitham & Miller, 2016). The removal of the human element in this way would go some way to address accusations of subjectivity and arbitrariness which are often used in the criticism of welfare measures.

1.2.1 Physical Welfare

Understanding and assessing the physical health of an individual is critical to building a picture of its overall welfare. This has, in the past, been a time and labour intensive process involving the diagnoses of conditions through physical examination. Such methods range from the collection and analysis of samples of serum, tests of antibodies, observations of injury or swelling, and the physical manipulation of effected limbs (Belshaw *et al.*, 2016; Mosafari *et al.*, 2015; Polgár *et al.*, 2019; Rault *et al.*, 2018; Stalley *et al.*, 2018). Alternatively imaging methods, such as radiography, can be used, but are only indicative of the severity of a condition. Research has reported mixed results as to whether radiographic severity provides insight into the degree of pain experienced (Duncan, 2006; Hannan *et al.*, 2000).

In contrast to those more invasive, potentially pain-inducing, and usually costly methods, behaviour based methods forgo formal diagnoses in favour of the rapid recognition of abnormal or aberrant behaviours, such as reduced appetite or lethargy, which are thought to be indicative of an underlying issue (Dawkins, 2004). However,

these methods rely on the accuracy of reports from owners and caregivers (Dawkins, 2004). The limitations of such methods, besides their reduction in specificity, lie with their use of human observers and has meant that there has been much interest in the development of alternative, non-invasive, technology based methodologies focused on replacing the role of the observer (Brown *et al.*, 2013; Rushen *et al.*, 2012; Whitham & Miller, 2016).

The monitoring technologies that will be discussed are examples of those affordable and accessible enough to be practically implemented in agricultural, domestic and research settings. This will exclude invasive procedures and imaging solutions in favour of methods of remote monitoring that remove the human element, and enable a longitudinal monitoring period which would allow the representation of the progression of welfare degradation or improvement. These methods fall into 2 broad categories; animal attached monitoring (where devices are secured to an animal), or vision based monitoring (where cameras are used to observe animals remotely).

1.2.1.1 Animal Attached Monitors

Inertial measurement units (IMUs) have been of much interest across multiple disciplines in recent years. These devices are a collection of sensors that measure the specific force of a body. From human medicine to wild animal behavioural ecology IMUs containing a combination of three dimensional (also known as tri-axial) accelerometers, gyroscopes and, on occasion, magnetometers have been deployed and have returned promising results (Brown *et al.*, 2013; Fong & Chan, 2010; Wilson *et al.*, 2008). In the field of animal welfare the devices have shown equal promise, particularly in the detection of behavioural indicators of injury and illness, thus allowing an objective view into the physical state of focal individuals (Barthelemy *et al.*, 2009; de Passillé *et al.*, 2010; Pfau *et al.*, 2016).

The attachment of IMUs to animals allows the high frequency recording of, typically, three dimensional movement data which can be interpreted to provide information as to the movement of the device. This, in turn, can be used to infer aspects of behaviours being performed by the individual (Wilson *et al.*, 2007). Advances in battery technology, memory capacity and the ability to transmit data have all contributed to making the devices more practical for longitudinal observation of individuals, at higher frequencies, allowing the capture of both rapid behavioural events and longitudinal patterns of behaviour, which may both be difficult to observe with traditional methods.

IMUs have been employed in a variety of ways to assess the physical welfare of individuals. Primarily among these is their use in gait scoring (Dawkins, 2004). The detail of data produced by the devices is such that, when attached to the legs of individuals, each step of a walking pattern can be clearly observed (Rushen *et al.*, 2012). Even the

gait type the animal is engaging in can be defined through comparisons with previously labelled data and the calculation of timings for different phases of the steps (de Passillé *et al.*, 2010). High fidelity data allows the objective assessment, by researchers, owners and veterinarians, of the gait of individuals over time periods previously impractical for collection. Animals with severe impairments would likely be detected easily by their owners and so the value lies in the potential of such methods to detect changes in gait at earlier stages than is typical by human observers (Mccracken *et al.*, 2012).

Alternatively IMUs have been employed for the monitoring of more simplistic behavioural aspects, such as the general physical activity of individuals. Here simple thresholds are drawn that dictate activity levels, typically defined by comparisons of data with human observations (Morrison *et al.*, 2013). Physical activity schedules are then derived based on these thresholds. Any observed drops in activity levels can be indicative of poor health, injury, or behavioural changes, such as oestrus (Holman *et al.*, 2011; Kozak *et al.*, 2016; Morrison *et al.*, 2013). The occurrence and patterns of these behaviours could be indicative of welfare state and is akin to data collected using more traditional observation-based methodologies (Martiskainen *et al.*, 2009; Rushen *et al.*, 2012; Wilson *et al.*, 2008). Postural patterns, frequency of feeding, social interactions and other behaviours of interest have all been observed through the use of IMU devices in both wild and domesticated species (Fehlmann *et al.*, 2017; Martiskainen *et al.*, 2009; Rushen *et al.*, 2012). By capturing behavioural patterns IMUs could be used in lieu of human observers with the potential to capture more detailed data over longer periods, such as the daily patterns of posture exhibited by an individual (Brown *et al.*, 2013; Martiskainen *et al.*, 2009; Ringgenberg *et al.*, 2010). For example, feeding behaviour and the amount of time spent laying down have been detected with high levels of accuracy and have both been identified as relevant to welfare in cattle (Diosdado *et al.*, 2015).

These methods do not come without their drawbacks. The ability to collect high frequency data over long time spans produces data-sets of sizes that often require the attention of trained individuals to identify behaviour patterns by eye or, more commonly, the use of sophisticated machine learning techniques. These techniques attempt to identify patterns by extrapolating thresholds and definitions, provided by previously human-labelled data-sets, or by creating likely classifications based on provided rule-sets, but little or no *a priori* information (Nathan *et al.*, 2012; Valletta *et al.*, 2017). This allows the rapid analysis of the data of multiple animals which can often stretch for hours, days or months. However, the nature of some machine learning algorithms has led to some being referred to as black-box methodologies. This is due to the difficulty involved in the assessment and interpretation of the internal rules generated and used by the algorithms to arrive at their classifications.

A further complication involves the attachment of the devices to the animal and,

although modern IMUs are small and lightweight, the potential for impact on an animals normal behaviour should be considered (Casper, 2009; Ropert-Coudert *et al.*, 2004; Wilson *et al.*, 2006). This is especially true when smaller animals are being sampled, and appropriate attachment methods should be fully investigated. Time invested in habituating animals to the presence of the device(s) would be well spent but could be impractical if managing larger numbers, such as the stocking densities of commercial agriculture. The costs of devices in agricultural or research settings could also be cause for concern. The prices of individual components and even complete IMUs have fallen drastically however when outfitting an entire herd of cattle, for example, the initial investment would be significant.

Another form of animal attached monitoring is one which focuses on a much grander scale, namely the animals location opposed to the specific movement it is currently engaged in. Global Positioning Systems (GPS), and Local Positioning Systems (LPS) have been used extensively in the tracking of free-roaming animals (Barker *et al.*, 2018; Cagnacci *et al.*, 2010; Rushen *et al.*, 2012; Tomkiewicz *et al.*, 2010). What these sensors offer is a species agnostic method for tracking distances and speeds of travel as they do not require specific calibration to account for attachment method and body plan differences, something that is needed when attempting to identify behaviour from IMU outputs. This allows the detection of distances travelled, of significant changes in daily activity, patterns of movement and resource access (Handcock *et al.*, 2009; Rufener *et al.*, 2018; Ungar *et al.*, 2005). Each of these outputs can, as in IMUs, be used to monitor for changes or inconsistencies which would be difficult to capture with manual observation methods and that may be indicative of an individuals welfare.

The battery life of GPS devices is inversely related to recording frequency and so increases in the fidelity of the data significantly reduce the potential deployment time. Additionally the dependence of GPS on clear signals to satellites being available may preclude it from use in species which spend extended periods inside, or under cover, or where device longevity is required alongside high recording frequencies (Frair *et al.*, 2010; Tomkiewicz *et al.*, 2010). This has been addressed somewhat in the use of Dead-Reckoning techniques where an IMU device is deployed simultaneously and data relating to the movements and turns of an individual are used to inform its path of travel between infrequent GPS calibration-fixes (Wilson *et al.*, 2007).

Where habitats are tightly controlled, or only specific features are of interest, radio frequency identification (RFID) can be used. RFID tags are used extensively in the beef, dairy and swine industries for the recording of feeding behaviours on an individual basis (Rushen *et al.*, 2012; Whitham & Miller, 2016). These methods involve the detection, by a receiver, of the animal-attached RFID tag at the point of interest, for example a feeder, and from this it can be inferred that the individual in question is likely engaging in a specific behaviour, such as feeding (Mendes *et al.*, 2011). Such uses

could help to isolate abnormalities in feeding behaviours, or other physiological behavioural indicators of welfare which could signal the occurrence of health conditions or changes to welfare (González *et al.*, 2008; Huzzey *et al.*, 2007; Urton *et al.*, 2005). The use of RFID is simple and effective but difficult to apply in less controlled settings or where behaviours of interest are unknown. This is due to the requirement to pre-place receivers at points of interest. In such eventualities the aforementioned LPS may be a more appropriate method.

1.2.1.2 Vision-based Monitors

Where attachment of devices to animals is impractical or impossible, due to the size of the focal species, the likelihood of the destruction of devices, or the significant costs of outfitting large numbers of animals, alternatives must be sought. One of these is the use of varying kinds of video and image capture technologies. Large groups can be captured simultaneously with much less of an initial investment than outfitting some or all with individual tags. Additionally the ability to capture video from areas inaccessible to the animals means issues arising from contact or attachment of the devices are circumvented (Dawkins *et al.*, 2009; Rushen *et al.*, 2012).

One prominent example of the use of video in the investigation of physical state in captive animals uses the theory of optical flow. Here the rate of change of brightness in a video is detected and analysed for a rapid group level indication of the prevalence of physical issues such as lameness. Dawkins *et al.* (2009) used this method to effectively assess the lameness of broiler chicken flocks ranging from 3,000 to 40,000 individuals. The group-centric nature of this approach makes it appealing for use in agricultural settings where many, often near identical, individuals are housed together on a large scale and any treatments or assessments are performed at the flock-level. This method loses the ability to easily isolate individuals that are exhibiting symptoms. However, the comparative inexpensiveness, both monetarily and temporally, makes the methodology a highly appealing approach in such settings.

A more individual approach would require group housed individuals to be easily identified from video, which often requires prior marking. For some species sophisticated machine learning solutions have been developed to allow individual identification without potential disruptions, such as marking (Zhang *et al.*, 2019). The determination of physical traits or abnormal movement characteristics which predict ailments, such as lameness, could be automatically assessed and scored by image processing algorithms (Rushen *et al.*, 2012). For example, in cattle an arched back and poor hoof tracking are distinct signs of lameness and have been shown to be detectable automatically in videos (Poursaberi *et al.*, 2010; Viazzi *et al.*, 2014). Similar deviations from the norm exist in other species, and in other conditions, although there is a requirement

of extensive prior investigation and validation before such methods can be employed (Grégoire *et al.*, 2013; Poursaberi *et al.*, 2010). This approach is highly species specific and could be impractical in real world settings as it would require the movement of an animal across the cameras field of view in a relatively controlled manner.

In common with animal attached devices, image based solutions are particularly interesting due to their ability to surpass the observational capacities of human observers. High frame rate video and infrared thermography (IRT) are two areas that have, in recent years, become both more prolific and more affordable. Each surpass the natural capabilities of human observers in different ways.

The first of these, high frame rate (also known as slow motion or high speed) cinematography has seen much use within the investigation of physical welfare. However, its use appears to rarely be the focus of the research. This is often left to the features being detected and is perhaps due to the similarity between high frame rate and standard cinematography, with the most meaningful difference lying in the sampling frequency.

One application that takes advantage of the increased temporal resolution of high frame rate cinematography is in the subjective assessment of gait. Here research has shown subjective assessment to be poorly correlated with objective alternatives when assessing the presence of subtle lameness (Lane *et al.*, 2015; Quinn *et al.*, 2007). It has been suggested that this is due to the rapid rate of limb movement making assessment of subtle differences difficult (Gillette & Angle, 2008; Lane *et al.*, 2015). However, Lane *et al.* (2015) reports that within the gait assessment of canines the use of high frame rate video provided no increase in either the accuracy or consistency of assessors ratings when compared to standard observations. This did not include classifications of the lameness of the dogs assessed and so further work would be required to observe the effects of high frame rate video on the assessment of subtle lameness. High frame rate video has also been used in the computational analysis of gait which presents the possibility of identifying differences in gait phenotype that are missed by subjective assessment (Preisig *et al.*, 2016).

IRT allows observers to visibly interpret the temperature data of individuals, remotely. This method is non-invasive with no need for restraint or other stressors inherent in the use of equivalent alternatives. IRT allows the visualisation of temperature across the entire surface visible to the camera rather than just at the location of a probe. One of the most prominent uses of such technology is as a non-invasive diagnostic tool for injury with the potential to detect inflammation of joints or tendons up to two weeks prior to the presentation of other symptoms (Church *et al.*, 2009; Purohit & McCoy, 1980).

Once again such methods are impractical for continuous monitoring, due to the specific positioning required during measurement. However, when used as part of regular

check-up procedures vision-based methodologies have been useful in the detection and diagnosis of conditions prior to severe symptom presentation in species ranging from dogs to poultry (Ben Sassi *et al.*, 2016; Church *et al.*, 2009).

1.2.2 Cognitive Welfare

The physical state of an individual is no longer considered the be all and end all of their welfare. Their cognitive, or emotional, state is of equal importance. This realisation has led to the development of a range of techniques focused on the identification of changes in cognitive state and a reassessment of the treatment of animal affective state, emotion and mood within welfare science (Dawkins, 2017). Striving to ensure individuals have what they want is a good rule of thumb in maintaining good cognitive welfare. By following this rule researchers can observe and quantify behaviours attributed to the fulfilment of an individuals perceived needs (Dawkins, 2004). These wants can be strong enough that individuals will attempt to carry out behaviours to satisfy them even when doing so adversely affects physical condition.

Subjective methods of assessing cognitive state, feeling and emotion are quickly being supplanted by new technological methods. Many of these employ similar sensors to those used for the assessment of physical welfare and differ only in analysis technique or focal behaviours. One interesting point of note is that, unlike the monitoring of physical condition, the behavioural measures of mental state often involve the assessment of positive indicators, such as play, and do not merely focus on the identification of stimuli and other aspects with a negative valence (Whitham & Wielebnowski, 2013; Yeates & Main, 2008).

1.2.2.1 Animal Attached Monitors

IMU devices have not been used as prominently in the detection of non-physical aspects of welfare as they have physical ones. However, there has been some work into their ability to provide insight into their use as objective tools for the measurement of animal affective state. For example, Wilson *et al.* (2014) showed the potential for IMU devices to detect statistically significant changes in movement and posture in African elephants when they were exposed to stimuli of differing valencies.

An alternative measure for use in inferring the presence of positive or negative affective states is that of sleep periods. Humans have been found to experience fewer sleep problems when self-reporting a positive state in waking time prior to sleep (Step-toe *et al.*, 2008). The same phenomenon has been suggested to occur in other mammals (Langford & Cockram, 2010). Hokkanen *et al.* (2011) reported the ability to successfully identify 90% of calf sleeping time using collar mounted accelerometers and a

machine learning classifier. The inverse has also been reported with stressors having been shown to negatively effect the sleep of rodents (Pawlyk *et al.*, 2008).

One further employment of IMU devices embraces their ability to differentiate between forms of locomotion. Locomotor play is a common occurrence in many species and in calves it appears within a few hours of birth (Boissy *et al.*, 2007). The time spent playing per day however is typically only a few minutes and so manual observations of these behaviours are time consuming and difficult when attempted by human observers (Held & Spinka, 2011; Whitham & Miller, 2016). As play typically only occurs when other needs are met it has much potential as an indicator of good welfare (Boissy *et al.*, 2007; Whitham & Wielebnowski, 2013). In calves, locomotor play has been shown to be impacted by negative stimuli (*e.g.* pain or insufficient energy intake) and as such the monitoring of a decrease in, or the absence of, such behaviour could be an effective warning that some aspect of the animals environment is not meeting its needs or is imposing excessive stress (Boissy *et al.*, 2007; Held & Spinka, 2011; Rushen *et al.*, 2012). Rushen *et al.* (2012) showed the potential of IMUs to automatically measure locomotor play in cattle. The application of IMUs in the measurement of play would allow the detection of the rare occurrences exhibited by adult animals, a feature of behaviour that is often highly unlikely to be captured when observing in person (Held & Spinka, 2011; Whitham & Miller, 2016).

There are further inferences which can be made when collecting inertial data for use in assessing cognitive state. IMUs, as mentioned previously, are able to be used for the counting of steps and activity in relation to aspects of environment. By extension these step counts can indicate how effective changes to this environment are in improving welfare. An example is the suitability and effects of flooring surfaces on the number of steps taken by an individual (Platz *et al.*, 2008). Alternatively, investigations into the suitability of other aspects of environment, such as stall size in horses, can be performed through investigations of stepping or postural time budgets (Raabymagle & Ladewig, 2006).

The issues faced in the use of IMU devices in the detection of cognitive states are much the same as in their use regarding physical states. There is particular issue in the application of the techniques across species. Extensive, species specific, validation is required prior to the use of these techniques to avoid the misinterpretation of the effects of affective state on behaviour. This is not helped by the absence of a current gold standard alternative for the attribution of positive or negative state to non-human animals.

Both GPS and LPS have also been embraced in the measurement of cognitive state. This technology has been of particular use in the detection of environmental use and preferences in captive zoo animals and free ranging livestock (Holdgate *et al.*, 2016; Leighty *et al.*, 2010; Schoenbaum *et al.*, 2017). These methods mirror those of-

ten used for the investigation of wild animal movement and habitat preferences (Brown *et al.*, 2013; Tomkiewicz *et al.*, 2010). Leighty *et al.* (2010) employed GPS units to monitor the habitat preferences and ranging habits of captive African elephants. They found that movement rates increased, and more closely resembled distances moved by wild individuals, when captive elephants were able to live in larger enclosures and within complex social groups. Preferences and resource use were also found to be tied to the social rank of the elephant within the group. Holdgate *et al.* (2016) showed that the walking distances of elephants, based on GPS tags, were impacted by feeding variables as well as social variables. These findings were then used to inform the use of dynamic feeding schedules. Additionally, the ability to compare the locations of multiple individuals also provides information as to the potential rate of social behaviour between individuals (Hacker *et al.*, 2015). From this, inferences regarding the sociality of individuals, the effects of increased or limited potential social interactions, and the effect of environmental change on the degree of sociality can be made (Hacker *et al.*, 2015; Whitham & Miller, 2016). The potential of this technology is again tempered somewhat by the same restraints that exist when employing these devices for the assessment of physical activity.

RFID devices offer a degree of utility here that addresses some of the issues with GPS devices or IMUs in exchange for a decrease in potential resolution of data collected. These devices show particular utility in the investigation of preference and choice in animals. Ringgenberg *et al.* (2015) attached tags to hens and monitored their choice of nest where potential nesting sites differed in the colour and pattern of the outer appearance. RFID has also been used in similar ways to GPS when referring to social interactions and where animals are detected in the same location at the same time it can be inferred there is a degree of associative behaviour occurring (Krause *et al.*, 2013).

1.2.2.2 Vision-based Monitors

Standard speed cinematography has been used for many years in replacing the observation of animals by humans. Videos captured can be replayed in real time or at accelerated speeds. Human observers can then identify individuals and note their behaviours. The ability to pause, slow down and rewind can be helpful in the full transcription of behaviour but these methods can still prove highly time consuming.

The introduction of computer vision methods have meant that the manual transcribing and annotating of video can be expedited and the time investments of observers spent elsewhere. The capture of habitat preferences and space use are of particular interest when attempting to use video to capture data on what animals want. Currently most tools used detect movement in a two dimensional plane. This means that move-

ment paths around the frame of the camera and trajectory of the animal are all that are able to be tracked leading to the typical positioning of the camera above the enclosure and the resultant inability to simultaneously detect details such as postural state (Junior *et al.*, 2012). The algorithms for the automatic detection of human movement and posture are more advanced than those currently used for the same task in animals and the breadth of methods available reveals various differing schools of thought as to how to perform such a task (Torres & Solberg, 2001). However, promise has been shown in examples as complex as the automatic detection of dog postures and patterns of movement in three dimensional space (Barnard *et al.*, 2016). This method also groups similar patterns of behaviour together to potentially allow the detection of abnormalities. It, and similar algorithms, are still very much in the prototyping stages and further validation and optimisation is required.

The capture of faces has been of great use in the detection of whether an animal is in frame, therefore allowing data to be collected regarding the time spent at the cameras location. This would allow the camera data to be used similarly to location based measures, such as GPS, or even in the detection of behaviours or locomotive modes such as in Burghardt and Calic (2006).

High frame rate video capture methods may be of particular use in the observation of facial movements. Minute aspects of facial expressions can change rapidly and as such the capture of extra frames could assist in the improvement of the performance of both observer-based and computer-based analysis methods. In humans there is evidence of micro-expressions which betray the true emotional state of an individual when attempting to obscure this and instead present an alternative state to others (Iwasaki & Noguchi, 2016). Animals may engage in similar behaviours and further investigation of the presence of such expressions could yield a useful tool in earlier recognition of poor affective state. One known example of this would be the pain face, which is often assessed through the use of grimace scales, but is considered by vets to be frequently hidden or suppressed (Fenwick *et al.*, 2014; Leach *et al.*, 2011).

The capture of facial expressions in particular has been an area of much interest in both animals and humans. The Facial Action Coding System (FACS) is an objective system of describing facial behaviours by the use of action units. Action units describe the movements of specific muscles of the face and by using these, and temporal information relating to them, observers can code any anatomically possible facial expression (Ekman & Friesen, 1976). Different versions of FACS have been developed for different species, such as chimpanzees, orangutans, cats, and dogs (Caeiro *et al.*, 2017; Caeiro *et al.*, 2013; Vick *et al.*, 2007; Wathan *et al.*, 2015). Here slow motion video allows the observance of often rapid and minuscule action units that may be missed if coded in real time. The potential to miss or misinterpret is such that the use of FACS requires the videos be viewed in slow motion under the standard methodology. FACS

is often employed in the objective assessment of pain expressions which are a clear indicator of a negative cognitive state. Attempts to identify pain expressions is common in the literature across multiple species, including mice, horses, cattle and sheep (Descovich, 2017; Gleerup *et al.*, 2015a; Gleerup *et al.*, 2015b; Langford & Cockram, 2010; McLennan *et al.*, 2016). One drawback of such a methodology, in terms of real world application, is the requirement of an individual animal to position themselves before the camera when engaging in behaviours which would elicit the desired expression. Another is that the technique has an obvious relationship to the anthropocentric focus upon the face and facial expressions which may not be equally expressive across species (Leach *et al.*, 2011).

IRT has also been used in the monitoring of the occurrence of negative affective states. During exposure of cattle to fear inducing stimuli the extremities of an individual will cool relative to baseline temperatures (Stewart *et al.*, 2008). Similar extremity focused vasoconstriction is seen in other animals. For example, rats which are placed in a foot-shock chamber, having previously been fear-conditioned in such a chamber, show similar patterns with temperatures of the feet and tail dropping while the temperature of the eye increases (Vianna & Carrive, 2005). It should be noted that eye temperature in particular has displayed conflicting results (Whitham & Miller, 2016). As such, the experimenter should be wary and ensure that IRT has been validated for use with the focal species and the desired site of measurement.

1.3 Models of Welfare

The investigation of welfare and monitoring methodologies requires the selection of a suitable model species and welfare case. Dogs (*Canis lupus familiaris*) are one of the most popular companion animals in the UK with over 9 million living as pets, over 11,000 cared for by welfare organisations, and a further 2,550 used for research (Anderson *et al.*, 2020; Asher *et al.*, 2011; Clark *et al.*, 2012; Murray *et al.*, 2015). This prevalence and their use across varying contexts, from domestic to research, presents a strong case for using dogs as a model organism for the development of welfare assessment methods. Researchers, caregivers, and veterinarians are already increasingly turning towards smart technological solutions to monitor the welfare of domestic dogs. This alongside the degree to which the species is intertwined with human society make a focus on the application of modern technology to their care, monitoring and assessment timely and relevant (Jukan *et al.*, 2017). This thesis will attempt to establish a foundation from which a low-cost, low-burden solution can be refined that would output data of use across the spectrum of contexts dogs are found in.

To assess the efficacy of methodologies in monitoring welfare a suitable condition

known to have a negative impact should be chosen. Canine osteoarthritis was selected due to it being one of the major health and welfare concerns in domestic dogs. It is widespread, with an estimated 2.5% of dogs in the UK affected by the condition annually with the difficulty inherent in its diagnosis likely resulting in a much higher actual prevalence (Anderson *et al.*, 2018). The condition can vary in severity and location but is briefly described as a disease of the joints that is typically exemplified by the degradation or dysfunction of articular cartilage (Anderson *et al.*, 2020). The problematic nature of the diagnosis of the condition is resultant from this large variability in its presentation and its uncertain etiology (Anderson *et al.*, 2020; Belshaw *et al.*, 2016). A wealth of risk factors for the development of osteoarthritis in dogs have been identified and are believed to exacerbate a pre-existing genetic disposition towards the disease (Anderson *et al.*, 2020). For example, the condition is known to be age-related, occurring in older individuals more frequently. But risk factors can also be related to aspects of a dog's breed, such as the increased occurrence of joint dysplasia due to breed conformation, or their lifestyle, such as obesity or exercise (Anderson *et al.*, 2020).

Diagnosis within day to day veterinary practice relies on prior owner recognition of the disease, or the potentially delayed diagnosis by a vet during regular check-ups. The extent and severity of expected osteoarthritis is confirmed through the combination of owner reports, which may be potentially biased by the "caregiver placebo", and a number of other, often subjective, outcome measures (Belshaw & Yeates, 2018; Conzemius & Evans, 2012; Reid *et al.*, 2013). However, there does not exist a consensus on which measures to use. Additionally other more objective alternatives, such as force-plate gait analysis, are resource or cost prohibitive and are impractical for first-opinion practice (Belshaw *et al.*, 2016). These more costly objective techniques would only be applied after an initial or suspected diagnosis and, as such, would require a noticeable prior change in behaviour or welfare detected by the owner or during an unrelated veterinary examination or consultation (Belshaw *et al.*, 2016). These difficulties in the diagnosis of osteoarthritis contribute to the significant impact that the condition has on canine welfare. The chronic pain and mobility inhibition inherent in the condition are further exacerbated by the frequent delay in its diagnosis due to the difficulties in owner, and potentially veterinary, recognition (Belshaw *et al.*, 2016; Reaney *et al.*, 2017).

For this research an individual was considered to have the condition only if formally diagnosed by a veterinary professional prior to recruitment. Variation in location and severity were not constrained within the osteoarthritic group at this stage with the intent to collect a dataset of generalisable nature to the osteoarthritic population with minimal segregation along *a priori* assumptions of presentation.

1.4 Conclusions

Technological methodologies can serve as useful tools to supplement the arsenal of those tasked with caring for animals in a number of settings. The potential for the monitoring of physical state is well established within the literature and with further development and validation, particularly in technologies which are both affordable and versatile, could provide effective means to detect potential illness at stages not possible to achieve using traditional subjective methods. Additionally, the cognitive states of captive animals have become increasingly important to all parties in recent years. With increased scrutiny upon the proposition of objective, technological monitoring methods as an answer to detect deterioration in this aspect of welfare. Results from those studies attempting to measure such aspects show much potential and the ability to remotely examine internal affective states presents an opportunity to all involved in the care of animals.

The use cases for a technological solution to welfare monitoring across fields, situations and species are vastly different and as such a modular multi-method approach would be suggested. The potential to have the same basic systems and analysis usable in the home, as are used in industrial agricultural or research settings is entirely possible, and the thinking behind such an initiative would echo that of other interdisciplinary movements such as the one health initiative (Monath *et al.*, 2010). It should also be noted that there appears to exist a slight separation between those employing remote monitoring methods within a welfare context and those using them on wild, free-ranging species. This perhaps arises from the history of device use in the two fields which are only now beginning to converge. Those in the animal tracking and behavioural ecology fields spearheaded development of devices in an effort to observe species which are difficult to monitor by traditional methods. Devices used in veterinary, research and agricultural contexts tend to originate as measures of physical condition to expedite and objectify the assessment of animals in their care. Animal welfare scientists tend to bridge this separation and employ methods from both fields however the integration of methodologies does not seem to be complete. Attempts to further bridge this gap would likely result in gains for both sides. However, care should be taken in deciding the extent to which the automation of animal care should be allowed to progress. To eliminate the human element from welfare assessment completely would add further distance between carers and their animals. Which could lead to a more "laissez faire" attitude towards the welfare of animals (Balzani & Hanlon, 2020). It would perhaps be best to suggest a system wherein automated, longitudinal monitoring is supplemented by more traditional subjective methods at regular intervals and when data suggests it would be required (Donnell *et al.*, 2015).

Table 1.1: Summary of a selection of potential remote monitoring methods, their uses, benefits and limitations

Measurement Method	Welfare Measures	Benefits	Limitations
IMU (Accelerometer, Gyroscope, Magnetometer)	Gait Assessment (Accelerometer, Gyroscope), Small scale behavioural observation (Accelerometer, Gyroscope), Longitudinal patterns of behaviour (Accelerometer, Gyroscope, Magnetometer), Postural data (Accelerometer, Gyroscope)	Versatile and widely applied across species and contexts, modular with multi-sensor deployments common, typically battery and memory efficient in comparison to alternatives	Difficult to interpret manually, Requires observational or video "truthing", Requires robust and consistent attachment, Data captured dependent on position
GPS/LPS	Movement patterns, Estimated resource use	Allows spatial and environmental analysis regardless of husbandry and degree of human interaction, can be combined with existing GIS and environmental datasets for additional behavioural context, low memory requirements	Deployments often limited by high battery use, Easily obscured signal
RFID	Movement patterns, Estimated resource use	Can allow identification of individuals within data, deployments focus on points of pre-assigned interest reducing data processing requirements	Requires pre-placed receivers, short range can result in missing data points
Video	Movement patterns, Estimated resource use, Social behaviour recording, Small scale behaviour observation, Gait assessment, Postural data	Provides gold standard behavioural observation potential without potential observer effects	Limited observation area, Time consuming manual annotation or requirement of sophisticated machine learning annotation protocols
High Framerate Video	Rapid behaviour patterns Movement patterns, Estimated resource use, Social behaviour recording, Small scale behaviour observation, Gait assessment, Postural data	Potential to capture rapid behaviours usually difficult to observe	High memory requirements, Limited observation area, Sensitive to lighting and other environmental influences
IRT	Surface temperature, Vasoconstriction / Vasodilation, Extremity temperature, internal temperature via optic nerve	Non-invasive method of capturing objective physiological data inaccessible using traditional observation methods	Limited observation area, Relative scale is sensitive to environmental temperatures

The field of animal welfare stands to gain much from the implementation of the technology mentioned. There exist many other examples of alternative hardware that could be explored however those examined here tend toward the more affordable and more widely available.

Osteoarthritic and healthy domestic dogs are common and a sizeable sample is obtainable. The unique domestic context enforces a number of constraints regarding device deployment and usability that should be considered in selecting an approach to ensure the methodology transfers well across contexts. Table 1.1 provides a summary of the presented methodologies, the benefits of their use, their methodological limitations, and the welfare measures attainable. From this the decision was made to use a single tri-axial accelerometer due to the small form factor, wealth of potential welfare measures attainable, longitudinal nature of data collection, and the ease with which devices can be deployed and data retrieved.

There is still much to be done to move towards a more cohesive approach to the monitoring of animal behaviour, welfare, and condition across fields and between industries. The rate of development of such a cohesive approach would be significantly expedited through the concerted effort of the community towards collating and combining the diversity of established knowledge. Such interdisciplinary collaboration would ultimately lead to blanket improvements across many aspects of animal welfare. This thesis aims to establish a foundational framework for such an approach which is initially validated within the prolific case of osteoarthritis in domestic dogs. To this end the application of affordable and accessible technologies, tri-axial accelerometer devices, which are used across species and research areas, will be assessed. The maturity and widespread nature of the technology has resulted in a diverse selection of potential analyses employed across various species and contexts and this thesis assesses which of these are appropriate to translate the longitudinal, high-resolution acceleration data collected into meaningful representations of the physical and cognitive welfare of domestic dogs. Chapter 2 begins this assessment through the exploration of several extant measures. These measures are utilised frequently within behavioural ecology to extract meaning from acceleration data and, as such, have often been applied within welfare science without consideration of the precise biographical or biological correlates. Here the established measures are applied to the characterisation of the differences between osteoarthritic and healthy dogs, and their effectiveness, interrelatedness, and the relation of the measures to aspects of both biology and history of participating dogs will be considered and discussed. The established understanding of the data and the various calculated features are then used to further investigate the flexibility of the accelerometer devices, and the potential of incorporating and adapting established methods from across disciplines. Chapter 3 replicates a common technique for the monitoring and assessment of human gait. This application of the

devices investigates their use in the recognition and evaluation of postural transition movements, the degradation of which is often indicative of the osteoarthritic condition within human patients. This also provides the basis of Chapter 4 which, rather than the transitory events of the prior chapter, focusses on the longitudinal monitoring of behavioural modes similar to those used in the monitoring of wild and agricultural species. These behavioural modes, such as posture or locomotion, if identified reliably, would allow a level of behavioural monitoring which could be further built upon for complex analyses of behaviour and lifestyle that may be indicative of osteoarthritic health. Finally, Chapter 5 builds upon this idea of the potential uses of the longitudinal monitoring of behavioural modes by applying the assessment of acceleration to the detection of circadian and ultradian rhythms in overall activity. This chapter differentiates itself by attempting to decipher underlying activity patterns across the entire 24 hour period experienced by dogs rather than focussing on the specific identification and characterisation of behaviour. Each included chapter attempts to build out from aspects of the former and results in a wide foundation on which to further develop a platform for the longitudinal monitoring of canine behaviour. These methodologies are utilised throughout the thesis to gain potential insights into the osteoarthritic health of domestic dogs and, in doing so, their potential application within a wider health and welfare context.

Chapter 2

Activity, Energy and an Assessment of Extant Measures

2.1 Abstract

The use of accelerometer sensors in a wide variety of research has led to a wealth of potential measures devised to characterise the output. An initial assessment and comparison of these existing measures should therefore be performed to assess their suitability to the characterisation of the behavioural, and physiological aspects which may be indicative of the osteoarthritic health of dogs. By investigating the effectiveness of these extant measures in the discrimination of healthy control and osteoarthritic groups this chapter aims to better situate these commonly employed features within the context of osteoarthritic dogs and the potential biological or biographical indicators that are possibly being reflected. Features used were derived from the 3 raw orthogonal axes of the accelerometer devices and 3 common transformations of these data. Each dog therefore produced a total of 208 acceleration features to be assessed at both the daily and weekly temporal resolutions. These feature sets were then reduced, and collinear variables removed, through the application of Principal Component Analysis, for weekly feature sets, and Multi-Factor Analysis, for the daily feature set. Multi-Factor Analysis is a variation on Principal Component Analysis which accounts for the repeated measures nature of the daily features. Despite this the overall methodology used was equivalent with the number of output principal components being selected from the visual assessment of scree plots and, for each chosen component, the highest contributing feature was identified for use in a series of logistic models. Contribution tables for the daily Multi-Factor Analysis components were summarised to allow selection of the most influential contributor across the 7 days and a random effect controlling for the individual was included within the resulting logistic models. Most features selected in this way were poor indicators of osteoarthritic group when using both daily and weekly models. Only 2 ODBA related, weekly summary measures were found to have a significant relationship. These were the mean weekly ODBA and the weekly mean of daily ODBA entropy. This may suggest a strong relationship with energy expenditure while being indicative of the masking effect caused by the high levels of variance seen within dogs between days. Further exploration of these features was performed by the

investigation of their relationship with LOAD, a subjective measure of quality of life focussing on osteoarthritic health. The resulting linear regressions mirrored the logistic results and further supported the use of LOAD as a potentially more informative measure that accounts for aspects of the condition not covered by a binary diagnosis. The collinearity of age was also demonstrated and concerns highlighted as to the effect such confounds could have on this and future analyses.

2.2 Introduction

As highlighted in Chapter 1 accelerometers have been used for many years across a multitude of disciplines, species and contexts with a large degree of crossover in both aims and methodology. Their use continues to grow in prominence within both human and animal research and processing methods for the inference of health, state and behaviour from acceleration have been developed extensively. This ubiquity has resulted in the use of many different features and measures derived from the, less easily interpretable, raw acceleration. An investigation of extant features is therefore required when adapting such prolific technologies to new contexts. Here I examine a number of features, gathered from the literature, for their applicability to the detection of osteoarthritis in domestic dogs. Reviews such as Brown *et al.* (2013) describe a number of methods for extracting meaning via the processing of raw acceleration data and form the basis of the features examined here for their potential in discriminating between healthy and osteoarthritic dogs. The past applications of accelerometers in both human and animal contexts to establish biomechanical function and behavioural monitoring present a well defined case for the use of these devices.

Perhaps the most distinct divide in the nature of the features examined here is whether they concern the time or frequency domains. The former refers to how the signal changes over time, the temporal aspect of the data, while the latter relates to which frequency bands the signal lies within, the rhythmic aspect. The time domain is frequently processed further to expose different aspects of interest, such as the dynamic or static acceleration components, or to convert the raw acceleration into alternative measures, such as the Actigraph activity count metric (Actilife, Actigraph, FL, USA). The frequency spectrum of these alternative representations can often then be calculated to provide the rhythmic information of the processed component. 8 transformations of raw acceleration data are examined here (Table 2.1).

This chapter aims to explore the features used within the literature and to form a foundation based within this to guide the thesis. The extensive feature list includes many similar or correlated measures with the intention to compare and contrast the effectiveness of these, often very similar, methodologies within the context of dogs instrumented with a collar attached, tri-axial accelerometer device. This initial explor-

ation of past methods will attempt to identify the practicality of adapting those existing measures to the detection of osteoarthritis using a computationally light combination of a PCA-based feature reduction and logistic regression methodology. Through this a more developed understanding of the detection problem itself and the suitability of the measures should be attainable and help to guide future work. This is of particular importance when measures have been taken from a wide array of species and contexts.

Table 2.1: List of acceleration data transformations and relevant references.

Transformation	Domain	Description	Reference
Raw Acceleration	Time	Acceleration (g) as output by the accelerometer in each of the three axes (x, y and z)	Moreau et al 2009, Howe et al 2009
Activity Count	Time	Proprietary counts of activity over an epoch (typically a minute) provided by Actigraph devices or using unofficial solutions such as the activityCounts R package	Morrison et al 2013, Straker et al 2012, Michel et al 2011, Westerterp 2007, Papailiou et al 2007
Vector Magnitude	Time	Square root of the sum of the squares of the three raw axes or activity counts. Robust to rotation.	Ladha et al, Morrison et al, Howe et al 2009
Static Component Measures	Time	Acceleration component relating to the effect of gravity on the device and discounting device movement or acceleration from other sources. Calculated through taking the running mean of acceleration data or through the application of signal filtering techniques.	Wilson et al 2008, Shepard et al 2008, Vázquez Diosdado et al 2015
Overall Dynamic Body Acceleration (ODBA)	Time	The summed total of the dynamic acceleration component. The dynamic component is typically calculated by the subtraction of the static component from the raw acceleration data.	Bidder et al 2012, Wilson et al 2008, Shepard et al 2008, Qasem et al 2015
Vector of Dynamic Body Acceleration (VeDBA)	Time	The vector resulting from the dynamic component of acceleration. Calculated similarly to the ODBA but rather than summing the 3 axes the root sum of the squares is taken.	Bidder et al 2012, Wilson et al 2008, Shepard et al 2008, Vázquez Diosdado et al 2015, Qasem et al 2015
Orientation	Time	The orientation of the device, relative to either gravity or the subject, converted and dealt with as rotation matrices or degrees.	Moreau et al 2009, Ringgenberg et al 2010, Ladha & Hoffman 2018
Estimated Power Spectral Density / Amplitude	Frequency	The strength of constituent frequencies decomposed from an overall signal following transformation from the time to the frequency domain. The Fast Fourier Transformation (FFT) is a common method of decomposition providing the amplitude measure for frequencies up to the Nyquist frequency.	Watanabe et al. 2005, Sakamoto et al. 2009, Soltis et al 2012

2.3 Materials and Methods

2.3.1 Data Collection

A sample of 85 domestic dogs were fitted with a tri-axial accelerometer (AX3, Axivity, York, UK: 23.0 x 32.5 x 7.6mm). Participants were recruited using a volunteer and snowball based sampling methodology involving recruitment from the general public in and around the city of Newcastle-Upon-Tyne, U.K. Participants were first asked to complete a recruitment survey requiring the provision of details regarding their dogs health and disposition to allow the exclusion of dogs with potential confounds or that could show signs of stress or aggression during handling and device attachment. Appendix A includes tables summarising the collected samples with dogs divided into Osteoarthritic and Control groups. Osteoarthritic group dogs were those with a formal diagnosis of osteoarthritis given by a veterinarian, 29 of the 85 sampled dogs were in this group with the remaining 56 assigned to the healthy control group.

The age, breed, sex and whether the individual was neutered were also collected. Of the 85 dogs 46 were female and 39 were male. Within the osteoarthritic group 12 were female and 17 were male. Within the control group 34 were female and 22 were male. This reflects some prior work suggesting sex as a risk factor in the development of some variants of the condition (Anderson *et al.*, 2018). However, it should be noted that much of the literature is directly contradictory on how sex relates to increased risk of osteoarthritis and further work would be needed to establish the true nature of such a relationship (Anderson *et al.*, 2020). Additionally, only 4 dogs were not neutered within this study and all were in the control group. Dogs were accepted for inclusion if over 2 years of age to ensure full maturity regardless of breed or background (Geiger *et al.*, 2016; Salmeri *et al.*, 1991). 3 dogs were rescued and the owners were unable to provide an age or provided estimates with large margins of error. The breeds of dogs included in the study varied in conformation and size with 31 breeds included alongside a large proportion of crosses.

Details regarding the dogs home life, condition severity and any pain management or treatments currently, or recently, undertaken were also requested and provided. However, the free text nature of these answers and wide variance in treatment method, dosage and adherence required an alternative indication of such measures for use in further analyses. To assess severity, while accounting for the dogs actual quality of life, the owner was asked to complete the Liverpool Osteoarthritis in Dogs (LOAD) questionnaire (Belshaw & Yeates, 2018; Walton *et al.*, 2013). The LOAD questionnaire gives an indication of the owner-perceived quality of life of osteoarthritic dogs. Its application here, across dogs, regardless of the reported presence of osteoarthritis, aimed to capture a quantitative representation of the qualitative differences between individuals

in the effects, and their experience, of the osteoarthritic condition which accounts for the reduction of pain, symptoms or severity across management regimens, lifestyle differences, or conformational differences. It should however be noted that such a representation is subject to potential influence by owner-related confounds such as the "caregiver placebo" effect (Belshaw & Yeates, 2018; Conzemius & Evans, 2012; Reid *et al.*, 2013). LOAD scores were not obtainable for 3 participant dogs due to incorrectly completed or absent questionnaires.

Device attachment methods vary between species and context. It is critical that due consideration is paid to the methods by which devices are attached and that such decisions are made with consideration of species physiology and individual welfare, regardless of setting (Brown *et al.*, 2013). When implementing methods on non-wild populations further consideration should also be paid to compliance, how likely owners and carers are to tamper with, remove, and correctly reattach devices and the perception of owners and carers of the comfort of the attachment method. With domestic dogs this decision is typically confined to just two widely used options: collar-based tags and harness-based tags with the former selected for this thesis. Many owners may perceive a collar to be more comfortable as the majority of domestic dogs are accustomed to regular collar wear. The effects of lead attachment reported in Martin *et al.* (2017) necessitate, that when using a collar, the device is attached to a secondary collar. However, small dogs are potentially unable to comfortably wear two collars simultaneously (Westgarth & Ladha, 2017). Despite the potential to collect more consistent data, harness-based solutions require additional habituation to the wearing of a harness, either at all times or for being walked with, and these may be perceived as more obtrusive to owners. To minimise the perceived impact of the study a collar based solution was selected.

The sensor was initially aligned ventrally as suggested in Hansen *et al.* (2007), with the device Z axis corresponding to the dog's dorsoventral axis, as shown in Figure 2.1. The device laden collar was worn in addition to any usually worn collar to avoid the need for lead attachment and disturbance. Owners were instructed to not disturb the collar unless necessary. Seven days of continuous acceleration data were collected. Accelerometers were set to record at 100Hz, which has been shown to be sufficient for the collection of canine activity (Ladha *et al.*, 2013).

2.3.2 Data Processing

Once collection had been completed the acceleration data were downloaded and resampled, to correct any deviance from 100Hz that had occurred, by using the bi-cubic interpolation method available within the OMGUI software Open Lab, 2018. Resampled data were then loaded into R (R Core Team, 2018) for further processing. This first

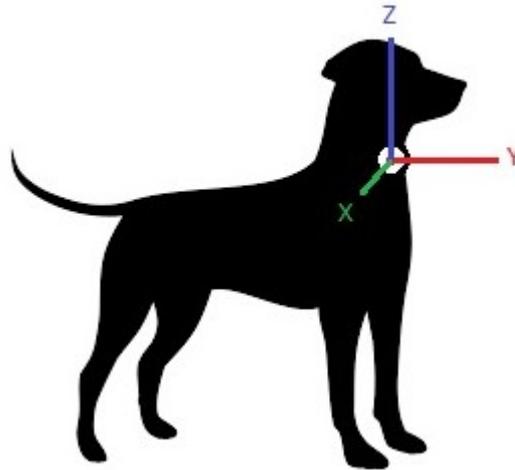


Figure 2.1: Location of collar-mounted sensor and device axes in relation to the dog

required the calculation of the data transformations described in Table 2.1.

Raw data in the three dimensions were preserved as some success had been shown in past literature with unprocessed acceleration (e.g. Moreau *et al.*, 2009). The vector magnitude (VM^3), the square root of the sum of the squared axes, was then calculated using equation 2.1, where X , Y and Z each represent the respective axis at the n^{th} reading. VM^3 has been used previously as an effective and prominent measure and provides a summary of acceleration across all three axes regardless of a devices initial and ending positions (Howe *et al.*, 2009; Robert *et al.*, 2009; Rodríguez-Martín *et al.*, 2015; Rodríguez-Martín *et al.*, 2013).

$$VM_{(n)} = \sqrt{X_{(n)}^2 + Y_{(n)}^2 + Z_{(n)}^2} \quad (2.1)$$

Activity counts are often used due to the prevalence of Actigraph accelerometers and software within the research community. Activity counts are a proprietary measure of activity given for each of the 3 dimensions and provide an abstract but accepted overall indication of activity across a set time period, typically 1 minute (Morrison *et al.*, 2013; Yam *et al.*, 2011). As Activity Counts are a proprietary metric, and I did not have access to the Actigraph software, counts used here were produced using the activityCounts R package which attempts to replicate the Actigraph activity count output (Brondeel *et al.*, 2019). The vector magnitude of activity counts was calculated here as a single summary measure of counts for the 3 dimensions (Hoffman *et al.*, 2019; Morrison *et al.*, 2013).

To calculate the static, or gravitational, component of acceleration a 4th order, zero-phase, low-pass Butterworth filter was applied to each of the 3 axes. This filter served to smooth the data and diminish the dynamic component leaving a representation of device orientation relative to gravity, and by extension the neck position of the dog (Grundy *et al.*, 2009; Shepard *et al.*, 2008b).

The dynamic component was also calculated and relates to all other, non-static acceleration removed by the above filter (Grundy *et al.*, 2009; Shepard *et al.*, 2008b). This involves the subtraction of the static component from the raw acceleration and the remaining signal relates to the bodily motion of the focal individual (Chakravarty *et al.*, 2019; Shepard *et al.*, 2008b). In addition to the dynamic component for each of the three axes two additional, combination representations were calculated, the overall dynamic body acceleration (ODBA) and the vector of dynamic body acceleration (VeDBA). These are both used extensively in the estimation of energy expenditure of wild, free-ranging animals (Bidder *et al.*, 2012; Qasem *et al.*, 2012). The ODBA is the sum of the dynamic component values across the three axes, the VeDBA is the VM³ of these values and is calculated in the same manner as other VM³ values (Bidder *et al.*, 2012; Qasem *et al.*, 2012).

A more direct estimation of orientation, in comparison to that provided by the static component, was calculated using the trigonometric equations described in 2.2 and 2.3 which provide the pitch and roll of the sensor, in degrees, relative to gravity.

$$Pitch_{(n)} = \arctan\left(\frac{-X_{(n)}}{\sqrt{Y_{(n)}^2 + Z_{(n)}^2}}\right) \quad (2.2)$$

$$Roll_{(n)} = \arctan\left(\frac{Y_{(n)}}{\sqrt{X_{(n)}^2 + Z_{(n)}^2}}\right) \quad (2.3)$$

Each of the above described transformations were then segmented into 7 24 hour epochs for the calculation of daily features. Table 2.2 describes each feature and for which transformation they were calculated.

These segments of data were also converted to the frequency domain by applying a Fast Fourier Transformation (FFT) algorithm to each of the three axes and the VM³ representation of each epoch (McClune *et al.*, 2014; Watanabe *et al.*, 2005). Data were normalised prior to the application of the FFT algorithm and, to reduce the occurrence of artefacts in the frequency spectra, a Hanning Window was applied, generated using the Signal R package (Developers, 2014). The application of a windowing function is necessary as the FFT algorithm assumes that the window assessed will be repeated perfectly and infinitely. If the window is not applied then it would result in the occurrence of a step function where the window begins and ends at differing levels and would introduce a false rhythmic pulse equal to the epoch length. The FFT algorithm is applied using the function available as part of R (R Core Team, 2018) and the R code used to calculate the frequency spectra of windows and the features calculated from these spectra is included in Appendix E. The resulting FFT spectra consist of a series of frequency bins displaying the magnitude of signal components occurring at the relevant frequencies. The resolution of the frequency bins is dictated both by sampling

rate (100Hz) and window size (24 hours). The magnitude of the signal relating to the 0Hz frequency is excluded from all feature calculations as this is equivalent to the DC component of the signal and is therefore equivalent to the mean of the time domain data.

The calculated features of the frequency domain are described in Table 2.2 and expanded upon here. They were chosen to represent the overall shape and distribution of the spectrum. The first of these is the mean power which was calculated as a representation of the central tendency of the entire spectrum. The Shannon's entropy of the frequency domain was also calculated and has been used previously in human and animal literature to estimate the stochastic nature of behaviour or the informational content of the acceleration signal (Bao & Intille, 2004; Benaissa *et al.*, 2017; Marais *et al.*, 2014; Wang *et al.*, 2005). The maximum, and second maximum values of power (calculated as the square of the absolute signal magnitude for a frequency) and the coinciding frequency are calculated to provide an indication of the dominant frequency of the spectrum. The use of dominant frequency and power is similar to methods used by Watanabe *et al.* (2005) to characterise actions. However, unlike Watanabe *et al.* (2005) the Power Spectral Density (PSD) used here is an estimation that has not been corrected for variance. This is typically addressed using Welch's method of spectral density estimation which involves the averaging of multiple FFT spectra, relating to the same action or period of interest, but here the focus is on the characterisation of an entire day rather than single actions (Welch, 1967). Variance in schedules across the weeks and across owners could be of interest, particularly as varying levels of attachment between dogs could be reflected in the data (Lund & Jørgensen, 1999). For this reason Welch's method was not used. Energy per sample was calculated as the sum of estimated PSD values divided by the total number of samples per window (8,640,000 or 24 hours at 100Hz) and is a potentially valuable counterpart to the dominant PSD values as it provides a summary of the overall energy of the signal, rather than only of those which are most prominent (Bao & Intille, 2004; Benaissa *et al.*, 2017; Ravi *et al.*, 2005; Wang *et al.*, 2005).

A total of 171 features were calculated from the time domain. This total consists of 12 features for each of the X, Y and Z axes of the raw, static and dynamic data transformations as well as a further 20 features per transformation for the combination measures (Vm^3 , VeDBA, ODBA, pitch and roll) Table 2.2 shows for which transformation the combination features were calculated. 3 inter-axes correlations (XY, XZ and, YZ) were also calculated for the raw, static and dynamic data transformations resulting in 9 further features. From the frequency spectra a total of 7 features were calculated from each of the 3 axes, and from the VM^3 of those axes, resulting in 28 total frequency features. All together 208 total features were generated per dog, per epoch.

Table 2.2: List of features, their descriptions and the data transformations it is appropriate to calculate them from

Common Features	Definition	Derived from
Mean	The arithmetic mean, typically across a set epoch or window length of the data	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Median	The central value of the data, an epoch or window which separates the upper and lower halves	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Maximum	The highest reported value within the data, an epoch, or a window.	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Minimum	The lowest reported value within the data, an epoch, or a window.	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
IQR	A measure of variability equal to the difference between the 25th and 75th percentiles of the data.	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Entropy	A measure of the information content, or unpredictability, of the data. Shannon's entropy is typically used and lower values relate to higher levels of unpredictability and more information contained in the data.	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Standard Deviation	A measure of the amount of variation within the data relative to the mean.	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Variance	A measure of the amount of variation within the data relative to the mean. Calculated as the square of the standard deviation	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Skewness	A measure of the asymmetry of a variable about the mean.	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Kurtosis	A measure of the 'tailedness', or the degree of extremity of outliers, of a variable.	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Pairwise Axis Correlations	The correlation of pairs of axes; XY, YZ, XZ.	Raw, Activity Count, Static Component, Orientation
Root Mean Square	The square root of the mean of the square of an epoch or window of data.	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Range	The difference between the highest and lowest values of an epoch or window	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Mean Absolute Deviation	A measure of variance from their mean. Calculated as the mean absolute difference between all values of an epoch and the mean value of that epoch.	Raw, Activity Count, Vector Magnitude, Static Component, ODBA, VeDBA, Orientation, Amplitude
Maximum power	The highest value of magnitude reported within a signal spectrum	Amplitude
Dominant Frequency	The frequency value which has the maximum value power value	Amplitude
Energy per sample	The sum of the squared magnitudes of the signal divided by the number of samples for normalisation.	Amplitude

2.3.3 Feature Reduction and Model Construction

Including all 208 features in a model would result in over-parametrisation of the model and a large degree of collinearity. To avoid this, and to attempt to find the features which explain the highest degree of variance between the control and osteoarthritic dogs, a dimension reduction protocol was developed. Principal component analysis (PCA) and its extension, Multi-Factor Analysis (MFA), were chosen for this. The MFA was included to better account for the repeated measures structure of the data due to the collection of acceleration across 7 24 hour epochs. Both the PCA and MFA were applied using the FactoMineR package (Lê *et al.*, 2008) in R (R Core Team, 2018).

To remove the repeated daily samplings, for the PCA, the mean of each feature per dog was taken. Once the PCA function had been applied a scree plot was generated and number of dimensions were selected. This was chosen based on the point at which the slope of the eigenvalues, the amount of variance each principle component, or dimension, describes, levelled out (Tabachnick & Fidell, 2001). It should be noted that use of the screeplot alone for this process has shown both good and poor reliability when reviewed but was deemed sufficient for this exploratory stage (Cattell & Jaspers, 1967; Crawford & Koopman, 1973; Kanyongo, 2005).

Once the number of dimensions had been selected the contributions of each feature to each of the chosen number of principle components were analysed. The highest contributor for each component was chosen. If the same feature was chosen for multiple components then the combination of components resulting in the highest summed contribution were selected. By doing this a drastically reduced subset of features was selected that explained the highest degree of variance between control and osteoarthritic dogs. The correlations within this subset of features were then assessed before generating univariate logistic regression models for each one. These features were then combined in the formula of a multivariate logistic regression with the lower contributing of any highly correlated features excluded.

The MFA methodology, performed using the FactoMineR package (Lê *et al.*, 2008), followed that described above however the separate readings per day were maintained through the component analysis stage. The mean contribution of features across the 7 days was then used for sub-setting. The logistic models were assessed as above with each also including a random effect term to control for the multiple daily recordings from each individual. Feature subset correlation was again assessed and informed model formulation.

The features found to be the most successful at differentiating between the two groups were then used to explore their potential ability to capture details as to the severity and qualitative impact of the condition upon individuals. First a t-test was performed to establish the distinction in LOAD scores between binary groupings. Once

the existence of difference was established linear regression models were assessed which used the owner-provided LOAD score for each dog as the outcome variable. An additional assessment of the correlation of LOAD with age was also performed to investigate the relationship of owner-perceived quality of life with age.

As age is a significant risk factor of osteoarthritis its relationship to the presence of the condition and to features most successful in differentiating the groups was also assessed (Anderson *et al.*, 2020). A t-test assessing the presence and significance of any difference in age between the healthy and osteoarthritic groups was performed. Additionally, the correlation of the most successful features with age was examined and the possibility of incorporating age into established linear models was explored.

2.4 Results

2.4.1 Weekly Summarised Acceleration Models

The scree plot of the week-summarised PCA, shown in Figure 2.2, was used for the selection of 6 dimensions as the minimum number of components required to explain data variance.

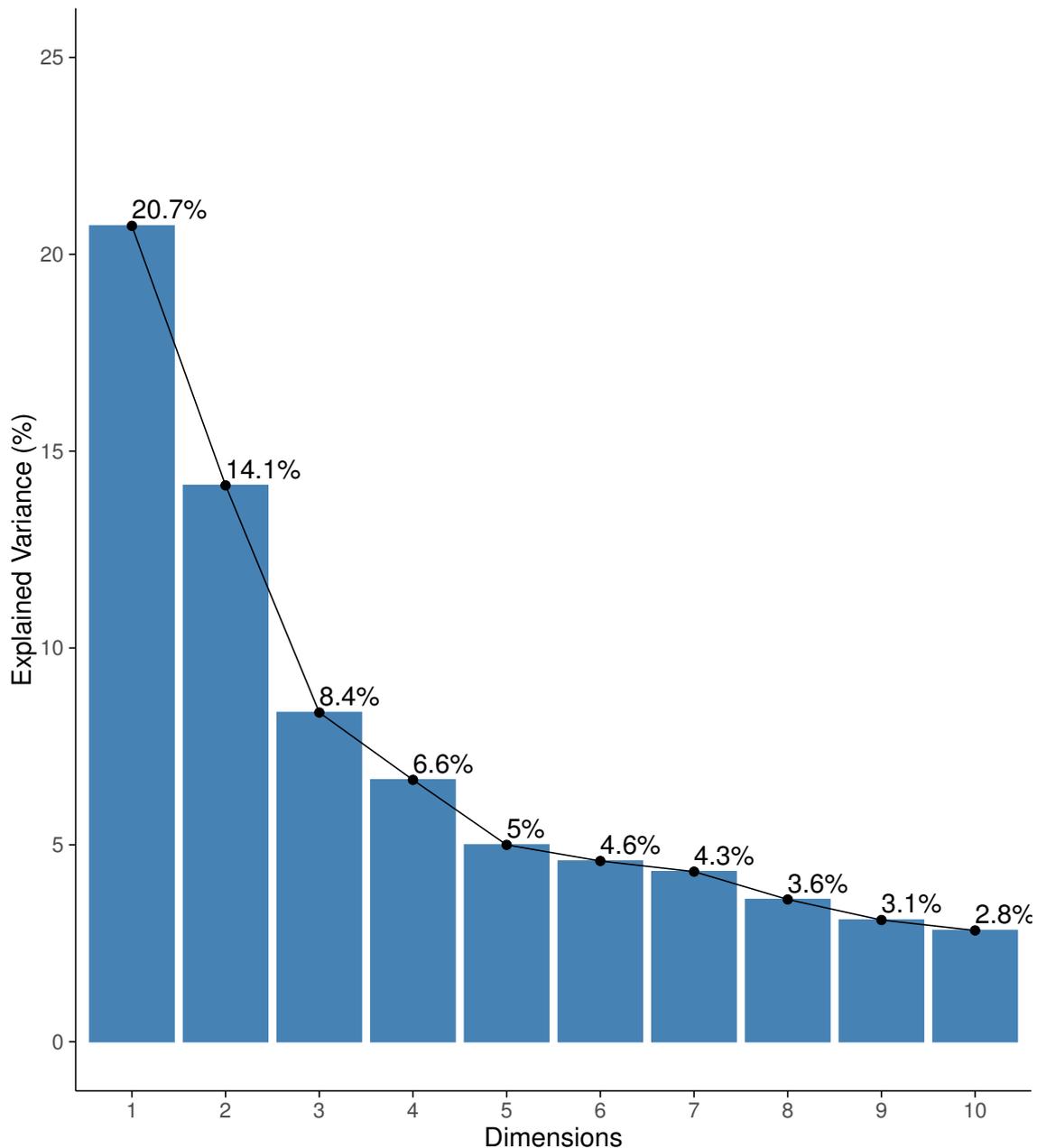


Figure 2.2: Scree plot showing the explained variance of the first 10 principal components of the week summarised feature-set

The 3 features with the highest contributions to each of the first 6 principal components are shown in Table 2.3 and were used to select the 6 features to retain. They were; Mean ODBA (2.12 to PC1), the Entropy of the ODBA (2.78 to PC6), Mean Device Pitch (4.18 to PC4), the Standard Deviation of Device Pitch (2.44 to PC2), Mean Absolute Deviation of the Y axis (3.72 to PC3), and Median Device Roll (3.27 to PC5).

Table 2.3: Contributions of the highest 3 contributing features for each of the first 6 principal components

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
Mean ODBA	2.116					
Mean VeDBA	2.108					
ODBA MAD	2.100					
Pitch SD		2.435				
MAD X Axis Static		2.427				
X Axis Static SD		2.423				
MAD Y Axis Static			3.724			
IQR Y Axis Static			3.703			
MAD Y Axis			3.639			
Mean Pitch				4.179		
Mean X Axis Static				4.146		
Mean X Axis				4.146		
Median Roll					3.274	
Median Y Axis Static					3.250	
Median Y Axis					3.228	
ODBA Entropy					2.499	2.780
VeDBA Entropy					2.254	2.735
Roll Entropy						2.164

The generated univariate logistic regression models investigate the isolated effect of each feature upon the presence or absence of osteoarthritis within the sample. An increase in the mean weekly ODBA resulted in the decreased probability of a positive osteoarthritis diagnosis, OR = 0.33 (CI: 0.13-0.70, $P = 0.01$). However, an increase in weekly ODBA entropy coincided with increased odds of a dog suffering from osteoarthritis, OR = 1.99 (CI: 1.06-4.02, $P = 0.04$). Table 2.4 also shows the results for other, non-impactful features for which univariate models were calculated.

The multicollinearity between the 6 features is shown in Table 2.5. As there were no pairs of features with correlation coefficients necessitating further feature removal all 6 features were included in a multivariate model with results given in Table 2.6. Only

the mean weekly ODBA remains significant and shows a consistent relationship with the presence of osteoarthritis, OR = 0.28 (CI: 0.09-0.66, $P = 0.01$). The pseudo R^2 of the multivariate model for weekly summarised activity was 0.464.

Table 2.4: Results of the univariate logistic models for each of the 6 selected features

Variable	Odd's Ratio	2.5% CI	97.5% CI	P Value	Pseudo R^2
Mean ODBA	0.326	0.127	0.699	0.009	0.276
Pitch SD	0.999	0.551	1.839	0.997	0.000
MAD Y Axis Static	1.080	0.582	1.916	0.795	0.002
Mean Pitch	0.897	0.492	1.635	0.718	0.004
Median Roll	1.677	0.909	3.308	0.111	0.075
ODBA Entropy	1.985	1.061	4.025	0.041	0.125

Table 2.5: The intercorrelations of the 6 selected features to allow for redundant feature removal

	Mean ODBA	Mean Pitch	Median Roll	ODBA Entropy	Pitch SD	MAD Y Axis Static
Mean ODBA	1.00	0.09	-0.02	-0.11	0.03	0.19
Mean Pitch	0.09	1.00	-0.06	0.03	-0.18	0.11
Median Roll	-0.02	-0.06	1.00	0.18	-0.06	0.14
ODBA Entropy	-0.11	0.03	0.18	1.00	-0.18	0.10
Pitch SD	0.03	-0.18	-0.06	-0.18	1.00	-0.15
MAD Y Axis Static	0.19	0.11	0.14	0.10	-0.15	1.00

Table 2.6: Results of a multivariate logistic model including all selected features

Variable	Odd's Ratio	2.5% CI	97.5% CI	P Value
Mean ODBA	0.28	0.09	0.66	0.01
Pitch SD	1.31	0.66	2.71	0.44
MAD Y Axis Static	1.28	0.62	2.57	0.48
Mean Pitch Angle	0.98	0.49	1.98	0.95
Median Roll Angle	2.04	0.92	5.01	0.09
ODBA Entropy	1.83	0.88	4.11	0.12

2.4.2 Repeated Measures Daily Acceleration Models

When each day was treated as a separate collection period, and repeated measures were controlled for, the scree plot of the MFA, shown in Figure 2.3, suggested 7 rather than 6 dimensions would explain the majority of data variance.

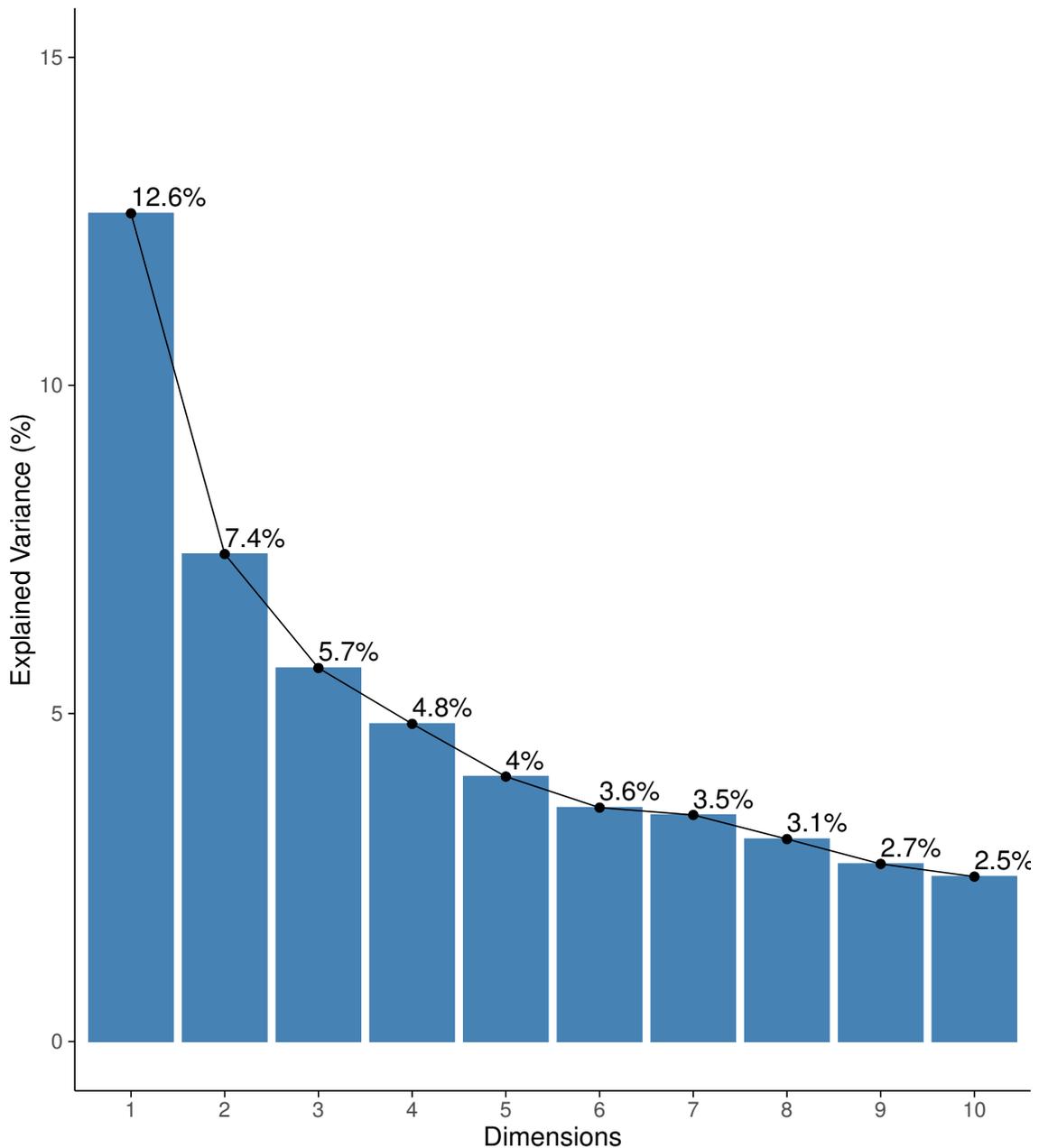


Figure 2.3: Scree plot showing the explained variance of the first 10 principal components of the daily summarised feature-set resulting from an MFA analysis

As the MFA results in an output per grouping, in this instance per day, the mean contributions per feature were taken as a summary of performance across the week, Table 2.7 shows all features which achieved contributions within the top 5% (using a threshold set at the contribution mean+2SD). The 7 features selected to be retained

were; the spectral density of the second most dominant frequency of the VM³ (1.28 to PC1), the spectral density of the second most dominant frequency of the Z axis (0.60 to PC2), the entropy of the power spectral density of the VM³ (1.01 to PC3), the power spectral density of the dominant Y axis frequency (0.83 to PC4), the entropy of the Y axis power spectral density (0.71 to PC5), the density of the second most dominant frequency of the X axis spectrum (0.63 to PC6), and the most dominant frequency of the Z axis power spectral density (0.58 to PC7).

Table 2.7: The mean contributions of the top 5% contributing features for each of the first 7 component dimensions when the MFA contribution per day output has been summarised per dog

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7
Mean Y Axis							0.32
Mean Y Axis Static							0.32
Mean Roll							0.32
Median Y Axis							0.32
Median Y Axis Static							0.32
Median Roll							0.33
X Axis Dominant PSD		0.54				0.47	0.47
Y Axis Dominant PSD			0.58	0.83	0.35		
Z Axis Dominant PSD		0.54		0.44			0.58
VM ³ Dominant PSD	1.24		0.80		0.30		0.40
X Axis 2 nd Dominant PSD		0.72				0.63	0.42
Y Axis 2 nd Dominant PSD			0.50	0.71	0.35		
Z Axis 2 nd Dominant PSD		0.60		0.40			0.55
VM ³ 2 nd Dominant PSD	1.28		0.77				0.34
Y Axis 2 nd Dominant PSD Freq.			0.31				
X Axis PSD Entropy	0.64	0.43				0.57	
Y Axis PSD Entropy	0.45			0.70	0.71	0.44	0.43
Z Axis PSD Entropy	0.73				0.43	0.30	0.39
VM ³ PSD Entropy	0.56	0.48	1.01	0.88			

The inclusion of each variable in separate mixed models, alongside only the random effect of the individual, resulted in models which did not suitably explain the presence or absence of osteoarthritis, as can be seen from the results in Table 2.8.

Table 2.8: Results of the mixed logistic models, each including a random term accounting for individual differences, for each of the 7 selected features

Variable	Odd's Ratio	2.5% CI	97.5% CI	P Value	Pseudo R ²
VM ³ 2 nd Dominant PSD	0.698	0.002	208.742	0.902	0.000
Z Axis 2 nd Dominant PSD	1.133	0.102	12.608	0.919	0.000
VM ³ PSD Entropy	1.177	0.028	48.793	0.932	0.000
Y Axis Dominant PSD	1.077	0.135	8.612	0.944	0.000
Y Axis PSD Entropy	0.605	0.056	6.485	0.678	0.000
X Axis 2 nd Dominant PSD	1.089	0.088	13.435	0.947	0.000
Z Axis Dominant PSD	1.160	0.101	13.367	0.905	0.000

An examination of multicollinearity between features revealed a high positive correlation between the dominant and second dominant densities of the Z axis power spectrum ($r = 0.85$). Notable negative correlation were also present between the dominant power spectral density of the Y axis and the entropy of the power spectrum of the Y axis ($r = -0.55$), and between the dominant power spectral density of the VM³ representation and the entropy of the VM³ power spectral density ($r = -0.66$). The contributions of each pair to the 7 principal components were compared and that with the higher mean contribution was retained. The final model including all 4 retained features performed poorly, as shown in Table 2.9, with no significant features and an overall pseudo R² value of <0.000025 .

Table 2.9: The intercorrelations of the 7 selected features to allow for removal of highly correlated variables

	Y Dom. PSD	Z Dom. PSD	X 2 nd Dom. PSD	Z 2 nd Dom. PSD	VM ³ 2 nd Dom. PSD	Y PSD Entropy	VM ³ PSD Entropy
Y Dom. PSD	1.00	0.10	0.10	-0.04	0.09	-0.03	0.06
Z Dom. PSD	0.10	1.00	1.00	0.18	0.85	-0.05	-0.11
X 2 nd Dom. PSD	-0.04	0.18	1.00	1.00	0.20	-0.08	-0.02
Z 2 nd Dom. PSD	0.09	0.20	0.20	1.00	1.00	-0.07	-0.11
VM ³ 2 nd Dom. PSD	-0.03	-0.05	-0.05	-0.07	1.00	1.00	-0.66
Y PSD Entropy	-0.55	-0.17	-0.17	0.00	-0.16	0.29	-0.11
VM ³ PSD Entropy	0.06	-0.11	-0.11	-0.02	-0.11	-0.11	1.00

Table 2.10: Results of a mixed multivariate logistic model including all selected features and a random effect accounting for the unaccounted variance between dogs

Variable	Odd's Ratio	2.5% CI	97.5% CI	P Value
VM ³ 2 nd Dominant PSD	0.72	0.00	217.88	0.91
Z Axis 2 nd Dominant PSD	1.10	0.09	13.01	0.94
Y Axis Dominant PSD	1.07	0.12	9.33	0.95
X Axis 2 nd Dominant PSD	1.07	0.08	13.68	0.96

2.4.3 Exploration of Variance in the Osteoarthritic Condition

The LOAD score of the two groups appeared to follow the expected increase corresponding to osteoarthritis grouping, albeit with a large degree of within group variance, with a mean of 5.11 years (± 5.19 SD) for the Control group and of 17.57 years (± 8.45 SD) for the Osteoarthritic group (Figure 2.4). Assessment of this relationship using a t-test confirms that there was a significant difference in LOAD between groups ($t(15) = 5.24$, $p < 0.001$).

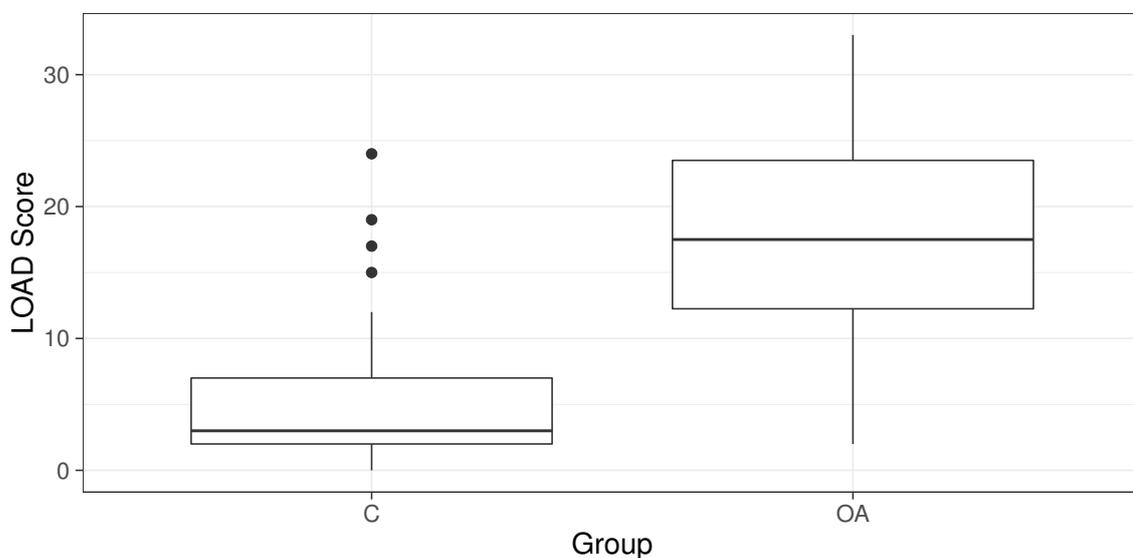


Figure 2.4: Visualisation of the variation in group LOAD score

The features most highly indicative of osteoarthritic grouping, when assessed using the logistic models above, were included in univariate linear regression models where the output variable was the LOAD score. The mean weekly ODBA was significantly associated with LOAD score ($F(1,61) = 13.46$, $p < 0.001$), with an R^2 of 0.18. This is shown in Figure 2.5 A) and reflects the relationships exposed by prior logistic regressions. Similarly the 7 day mean of daily ODBA entropy (Figure 2.5 B)) also maintained

a relationship with LOAD reminiscent of that shown in previous logistic models ($F(1,61) = 9.66, p = 0.003$), with an R^2 of 0.14.

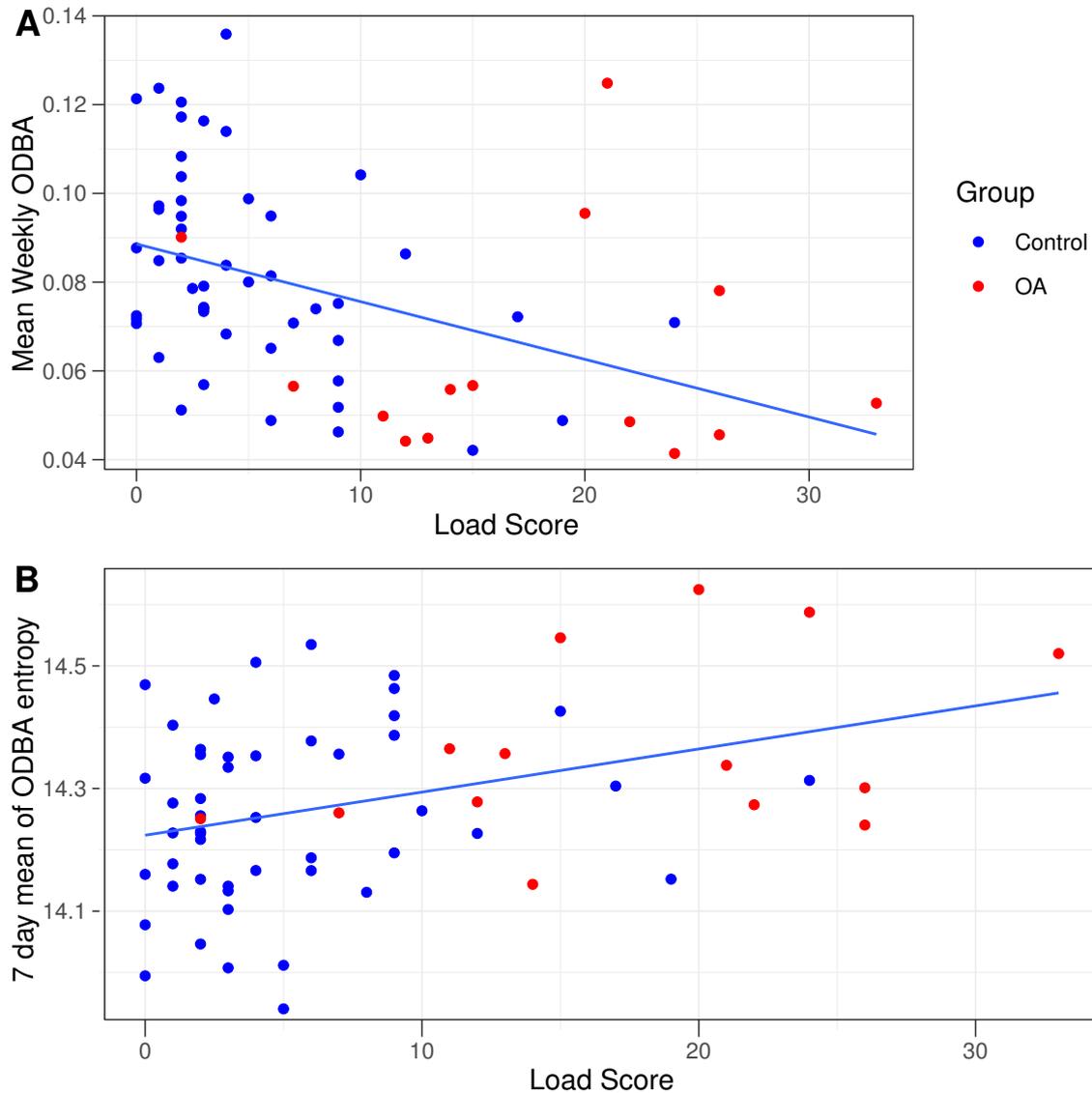


Figure 2.5: Regression of LOAD Score against A) Mean Weekly ODBA and B) 7 day mean of ODBA entropy

2.4.4 Age Relatedness

To assess the potential influence of age an assessment of the difference between the control and osteoarthritic groups was performed. Excluding the individuals whose owners had not been able to supply an accurate age, the mean age of the sample was 7.7 years (± 3.9 SD). Within individual groups the mean age was 6.1 years (± 3.5 SD) for control dogs and 11.3 years (± 2.7 SD) for osteoarthritic dogs (Figure 2.6). A t-test assessing the ages of the two groups indicated a significant difference between groups ($t(26) = 5.85, p < 0.001$) which confirmed the presence of this age-relatedness within the sample.

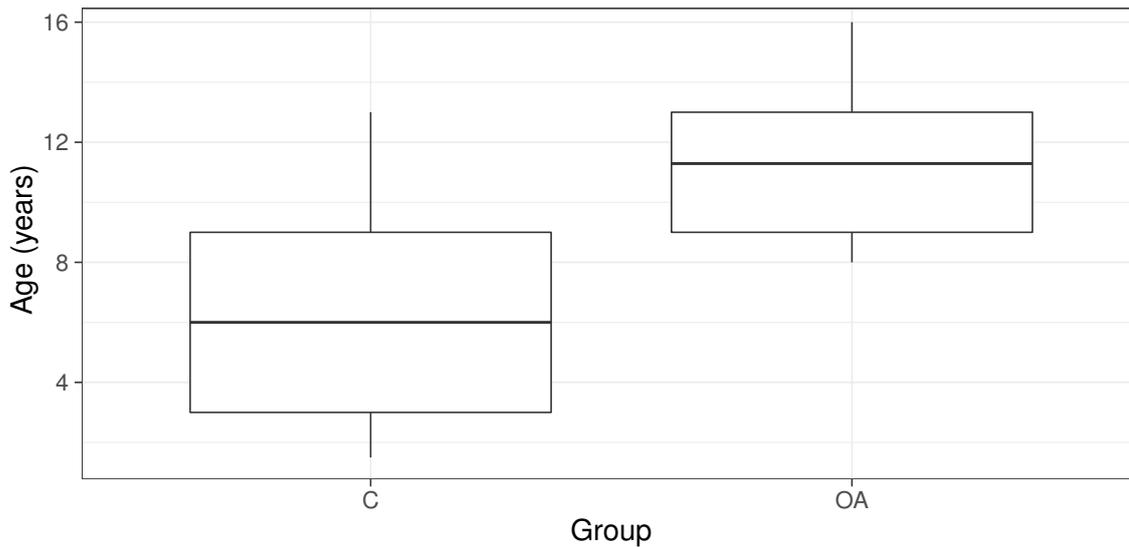


Figure 2.6: Visualisation of the variation in group ages

The use of the LOAD score in lieu of a binary diagnosis-based classification was established through the assessment of the correlation between it and age. A strong positive correlation ($r = 0.67$) is clear with an owner-perceived reduction in canine quality of life as age progresses (Figure 2.7).

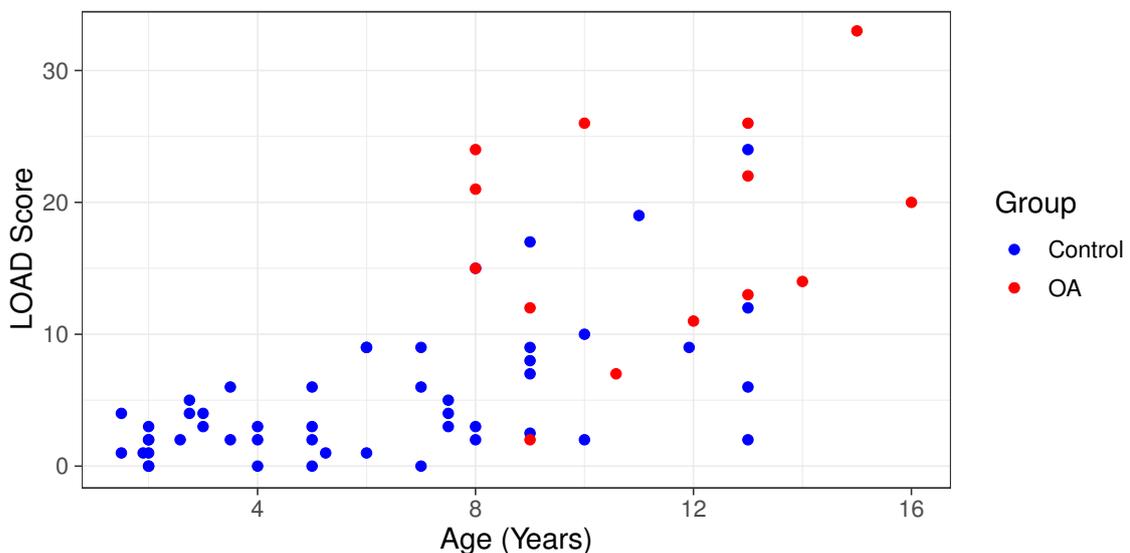


Figure 2.7: The relationship between LOAD Score and Age in years

An examination of the correlation between age and the features shown to be most related to osteoarthritic group, mean weekly ODBA and the 7 day mean of daily ODBA entropy, revealed both to have slight correlations. Figures 2.8 A) and B) show the moderate negative correlation between mean weekly ODBA and age ($r = -0.47$) and the weak positive correlation between 7 day mean of daily ODBA entropy and age ($r = 0.29$) respectively.

perceived quality of life. Additionally, an investigation of age, an established osteoarthritis risk factor, was also performed to explore the potential of features were confounds incorporated into models (Anderson *et al.*, 2020). These results have provided some insight into the possible issues with these methods and some indication of potential next steps in establishing an effective methodology.

Summarising activity measures as single, weekly values provided the best results from among both univariate and mixed model solutions. The PCA feature reduction and subsequent univariate modelling revealed 2 of the ODBA related features (Mean weekly ODBA and 7 day mean of daily ODBA entropy) to have a significant relationship to the presence or absence of osteoarthritis. ODBA has been shown to be an effective proxy for the energy expenditure of many species (Gleiss *et al.*, 2011; Miwa *et al.*, 2015; Qasem *et al.*, 2012). This result could be indicative of a relationship between the energy expenditure of dogs and osteoarthritis, with lower levels of expenditure being indicative of the condition, which has been commonly reported in humans and is consistent with anecdotal reports from dog owners (Belshaw *et al.*, 2016; Henchoz *et al.*, 2012; Mancuso *et al.*, 2007; Roubenoff *et al.*, 2002). The indication that higher levels of entropy is indicative of osteoarthritis is notable and may suggest a heightened level of behavioural fragmentation with less consistency in duration. In other words, osteoarthritic dogs appear to be less active overall, but their periods of activity are also less consistent and feature frequent breaks or changes in vigorousness. The linear models that instead use LOAD score as the output measure, and as such the indicator of osteoarthritic presence and severity, also reflect these relationships. However, as this is an owner reported measure, and does not include questions requiring the specific assessment of a dogs osteoarthritis, the measure is possibly capturing variance in the quality of life of the dogs examined, or more specifically the owners perception of their own dogs quality of life and changes in this measure may therefore not be related to the presence or severity of osteoarthritis (Belshaw *et al.*, 2016). The clearest example of this potential for confound can be seen within the exploration of age as an osteoarthritic risk factor and its inclusion in the linear models. Age is shown here to be closely correlated with the ODBA based features and LOAD score while also being significantly different between sample groups. As such, the effects of age, and other potential confounds, should be considered and further explored when attempting to use acceleration based measures for osteoarthritis prediction and gauging the severity and impact of the condition. This supports the anecdotal reports of dog owners and human osteoarthritis sufferers (Belshaw *et al.*, 2016; Henchoz *et al.*, 2012; Mancuso *et al.*, 2007; Roubenoff *et al.*, 2002). Following from these analyses we must also question the reliability of the LOAD scores gathered as each was collected by a different individual. With advancing age, as illustrated in the models correlations given above, a dogs activity levels naturally decline, this change may be perceived by the owner

in such a way as to produce artificially high LOAD scores, alternatively a dog with no known condition may be given an artificially low score due to an expectation of health by the owner.

The poor performance of daily measures and the resultant repeated measures feature selection and models may be indicative of a heightened level of variation between days that results in the obscurement of osteoarthritic state. This would result in the smoothing of these inconsistencies when taking the mean value for a week lessening their influence and revealing an overall trend towards lower, more chaotic ODBA levels. Such inconsistencies could be for several reasons. The first of these relates to husbandry with owner schedules likely playing an important role in animal activity. For example, non-anxious dogs are known to be less active when owners are absent, something that would commonly occur daily with owner work schedules (Scaglia *et al.*, 2013). Similarly the current study did not account for the day of the week at which data collection commenced leading to the occurrence of the weekend, where owners may be more likely to be home all day, at different points of the collection period for each dog. Another related potential reason is the inconsistency of exercise opportunities that may be afforded by the owner. Inconsistency in walk times, changes to walk length, or the absence of walking exercise on some days could each have impacted results in terms of weekly activity level and entropy. It should be noted that this may also be a cause of difference as an owners perception of their dogs osteoarthritis may lead to changes in husbandry and exercise schedules that are owner rather than dog-led (Belshaw *et al.*, 2016).

2.6 Conclusion

It can be seen that simple measures, as frequently used in past studies, are not sufficient to confidently identify osteoarthritic health. The two ODBA related measures identified in this study as showing potential for the identification of osteoarthritis, or for the recognition of a related reduction in physical or cognitive welfare, were insufficient for the task in isolation and using the regression based measures presented here. However, this study has provided insight on what could be required to achieve this and has successfully illustrated some of the considerations that should be made when attempting automated osteoarthritic assessment. The features both summarised the weekly energy expenditure of the animals and may also be indicative of other conditions or, as shown in this chapter, of natural aging. Additionally, these features only provide useful information regarding the welfare of an animal over a minimum of 7 days. For veterinarians and researchers such a system could still be of use but in real-world domestic applications a more immediate indication of current health and welfare would be desirable to owners for indications of immediate change. The usefulness of ODBA

shown here should be taken forward and inform future methods, as should the susceptibility of the features to osteoarthritic risk factors and confounds, but alternatives should be sought to better monitor real time health. One potential avenue would be the remote identification of behaviours themselves rather than the indications of their vigorousness. By showing how dogs act within their environment the accelerometer based system could provide information on what dogs want as well as the physical component illuminated here by ODBA related variables.

Chapter 3

The Automatic Categorisation of Postural Transitions of the Domestic Dog

3.1 Abstract

The ability of individuals to transition between postures can be informative as to the physical and cognitive condition of an individual. Within human medicine postural transitions are frequently assessed and monitored for a number of conditions using tests such as the Sit to Stand test. There has also been a wealth of research into the replacement or supplementation of the qualitative assessment methods used, or prohibitively expensive quantitative techniques, with more practical and affordable acceleration based alternatives. The similarities displayed by domestic dogs when suffering from such conditions coupled with their inability to communicate chronic pain or cognitive issues, and the difficulty owners face in noticing such problems in the early stages, presents a strong case for the development and application of equivalent methods in veterinary practice. This chapter is an initial exploration and development of such a technique focussing on the ability to identify postural transition events, between three identified posture states, using features derived from collar-based accelerometer devices. A series of linear discriminant analysis models were applied to data collected from 20, kennel housed, Labradors with approximately 10 hours of annotated video data and paired accelerometer data. A total of 51 features, derived from both human and animal based literature were assessed using 4 feature selection methods and data windows with lengths of 2,3 or 5 seconds centred on the transitional event. Additionally, the inclusion of each windows' estimated cumulative distribution function was assessed. All variations in methodology showed difficulty in the identification of transitional events from among randomly sampled, non-transitional data. Results also suggest difficulty in the differentiation of postural transitions from each other. This initial exploration therefore does not seem entirely promising however alternative methods of pattern recognition, further inspection of the data and altered methods of pre-processing, alongside a reassessment of the calculated features and their applicability should allow further insight into the ultimate practicality of such a task and these results suggest more sophisticated, or entirely reformulated, methods are required.

3.2 Introduction

The postural state of an individual constrains the potential behaviours available to be performed (Grundy *et al.*, 2009; Shepard *et al.*, 2008b). To engage in their full repertoire of behaviours domestic dogs are therefore necessitated to frequently transition between postures. These transitions could be of interest for a number of reasons but foremost would be that their occurrence or absence could be indicative of an individual's ability to perform normal behaviours, satisfy their needs, and, by extension, their ability to satisfy the five freedoms (Hart, 1988; Mellor, 2016). Extending this further the perceived normality of the transitional movement, when it occurs, may provide further insight into exactly how, and the extent to which, the health and welfare of the focal individual has been affected.

Within human medical literature the examination of postural transitions is a well-established indicator for a number of physical and cognitive conditions, such as Parkinson's disease and Osteoarthritis (Hickey *et al.*, 2016; Rodríguez-Martín *et al.*, 2013). The sit to stand (SiSt-T) and timed up and go (TUG-T) tests are used by clinicians to assess the presence or progression of such conditions and both revolve around the efficiency and quality of a postural transition (Anan *et al.*, 2015; Duncan *et al.*, 2011; Podsiadlo & Richardson, 1991). During these tests patients are asked to rise from a sitting to a standing posture and the kinematics of the transition are examined by a professional. The latter test also requires the patient to be timed as they perform the task, walk a set distance and then return to a seated position. Remote sensing technologies, and the commercial proliferation of such devices for use as fitness trackers and in medical practice and research, has led to a substantial quantity of literature attempting to either provide a more comprehensive, quantitative equivalent of such tests, to enhance the established assessments, or to capture such transitions during typically unobserved daily activity (Hickey *et al.*, 2016; Millor *et al.*, 2014; Rodríguez-Martín *et al.*, 2015; Rodríguez-Martín *et al.*, 2013; Weiss *et al.*, 2011; Yang & Hsu, 2010). Accelerometers, and other sensor-based methods are used frequently due to their small size, the ease with which they can be attached, the ability to deploy the devices longitudinally, and their relative inexpensiveness when compared to kinematic or force-plate based instrumentation which can be prohibitively expensive and inaccessible (Belshaw, 2017).

The detection of posture and postural transitions in humans is typically performed by sensors attached to the trunk or waist (Aminian & Najafi, 2004; Ganea *et al.*, 2012; Yang & Hsu, 2010). This is also often adhered to within behavioural ecological contexts and could be considered preferable for postural detection due to the direct correlation of trunk orientation and postural mode (Fourati *et al.*, 2011; Grundy *et al.*, 2009; Shepard *et al.*, 2008b). However, the unique conditions of the domestic dog result in the need to

balance the expectations and perceptions of owners along with the usual attachment constraints of comfort and habituation. As collars are required by law in the UK, when dogs are in public areas, the majority of dogs already habituated to their presence and owners are unlikely to object to the use of collars as an attachment method and such implementations have been found previously to have high compliance rates (Westgarth & Ladha, 2017). For these reasons a collar-based solution was proposed, despite the additional impact of head movement.

The inability of dogs to self-report discomfort, illness or pain, presents a further instance of utility for the longitudinal monitoring of transitions using accelerometers (Belshaw *et al.*, 2016). For example, human osteoarthritic sufferers frequently report a temporal aspect to their chronic pain which is likely present in canine sufferers and would possibly be missed during a single veterinary examination, or captured only once severe, as current metrics of companion animal quality of life and condition, outside of veterinary examination, are dependent on subjective owner report (Belshaw *et al.*, 2016; Reid *et al.*, 2013). Post-operative rehabilitation and analgesic efficacy monitoring are other potentially valuable longitudinal uses of such a sensor.

In this chapter the potential of automating the detection of postural transitions from longitudinal accelerometer data was assessed. Drawing a set of features from across both animal behavioural ecology and human medical literature an exploratory methodology was outlined and resulting models validated. This included the comparison of identical methodologies across 3 different data window lengths (2, 3, and 5 seconds). Linear Discriminant Analysis (LDA) classifiers were applied due to their ease of implementation and interpretation. LDA attempts to characterise a dependent variable as a linear combination of features. To achieve this LDA maximises the separation between classes while minimising separation within classes. During the initial LDA cross validation feature sets were reduced using one of 4 different selection methods. Once the performances of all combinations of feature selection methods and window lengths had been compared the most commonly selected features of the best performing combination was used to retrain the model and was then applied to novel data for validation. These results were then interpreted to assess the possibility and practicality of correctly identifying and differentiating between postural transition events from acceleration data when further developing the methods for real use, such as in the development of automated equivalents to the SiSt-T and TUG-T for use with free-living dogs with a single collar-based accelerometer sensor.

3.3 Materials and Methods

3.3.1 Data Collection

A pre-existing sample ($n = 20$) of kennel-housed Labrador retrievers, consisting of 9 males and 11 females was used. All dogs within this sample were aged between 3 and 6 years of age with specific ages unavailable within the provided sample data. The individuals included within this sample were known to be osteoarthritis-free.

Acceleration data were collected over a single day for each dog. Individuals were fitted with a collar and Axivity AX3 tri-axial accelerometer, set to record at 100Hz, as described in Chapter 2. The sensors were configured to run from 8am to 5pm on a single data collection day.

A total of 21 hours 18 minutes and 25 seconds of video was taken of the dogs during their daily walk and periods where they were able to access a communal paddock area. Walking periods consisted of a on-lead walk, a training session and an off-lead play session with the camera focussing on a single dog for a total duration of approximately 30 minutes. Paddock-based videos were taken from a static observation point outside of the area and focussed on multiple dogs simultaneously. Paddock periods were spread throughout the morning and afternoon and lasted between 40 and 105 minutes with 1 to 3 videos captured per kennel grouping.

3.3.2 Video Annotation

Postural transitions occurring within the videos were annotated, by one observer, as described in the ethogram provided in Table 3.1 using ELAN Version 5.7 software (Max Planck Institute for Psycho-linguistics, 2019). Transitions which were obscured or out of frame were ignored, even if posture had changed upon the dog re-entering the camera's field of view. Dogs were frequently observed using the sitting posture as a brief, transitional phase between lying and standing. In such instances transitions were considered distinct and were annotated as such where the description of sitting (a position where fore-legs are extended and hind-legs flexed as described in Table 3.1) persisted for at least 1 second before beginning a further postural transition.

3.3.3 Accelerometer Data Processing

Accelerometer data were retrieved from the devices and resampled using the OMGUI software (Open Lab, 2018) as described in Section 2.3. Resampled data were then loaded into ELAN to calculate the offset time between the accelerometer reported timestamps and the timing of the video annotations. At least three events per video, distinct when viewed in both the video and acceleration data streams, such as jumping,

Table 3.1: Ethogram of annotated postural transitions

Postural Transition	Definition
Lie to Stand	Subject moves from a ventral or lateral recumbent posture, where weight is predominantly supported by belly with sternum in contact with the floor or weight supported by the side with the shoulder in contact with the floor, to a standing position, where weight is supported exclusively by three to four paws.
Lie to Sit	Subject moves from a ventral or lateral recumbent posture, where weight is predominantly supported by belly with sternum in contact with the floor or weight supported by the side with the shoulder in contact with the floor, to a position where fore-legs are extended and hind-legs flexed.
Sit to Lie	Subject moves from a sitting position where fore-legs are extended and hind-legs flexed to a ventral or lateral recumbent posture, where weight is predominantly supported by belly with sternum in contact with the floor or weight supported by the side with the shoulder in contact with the floor.
Sit to Stand	Subject moves from a sitting position where fore-legs are extended and hind-legs flexed to a standing position, where weight is supported exclusively by three to four paws.
Stand to Sit	Subject moves from a standing position, where weight is supported exclusively by three to four paws, to a position where fore-legs are extended and hind-legs flexed.
Stand to Lie	Subject moves from a standing position, where weight is supported exclusively by three to four paws, to a ventral or lateral recumbent posture, where weight is predominantly supported by belly with sternum in contact with the floor or weight supported by the side with the shoulder in contact with the floor.

head movements and shaking were used as synchronisation events. Annotations were then loaded into R 3.6.0 (R Core Team, 2018) and adjusted using the offsets to ensure synchronisation. The function created to do this is included in Appendix B. A second function was then run to create an equal number of annotations relating to a randomly sampled "unknown" data class. This class is created to allow the assessment of future model performance against non-focal behaviours or those behaviours not accounted

for in the ethogram in Table 3.1. The code to generate these is included in Appendix C.

Attachment of devices to the collar of domestic animals is common but presents a number of idiosyncrasies that should be considered and accounted for when compared to alternative attachment methods (Brown *et al.*, 2013; Westgarth & Ladha, 2017). The most prominent of these is that the position of the device allows head and potentially jaw movements to be detected. This could prove useful and informative however neither are relevant to the current classification problem and the existence of the potential differences caused by these movements should be kept in mind as possible confounds. Additionally, the nature of using pre-existing data is such that sub-optimal or non-ideal data collection methods cannot be addressed. In this sample, when walking the dogs, leads were attached to the collar and this, as described in Martin *et al.* (2017), will result in abnormalities in the data during these periods due to any additional acceleration caused by lead tension and pulling. The exclusion of these time periods was not judged to be necessary due to the rarity of transition events during lead-walking and attachment periods as well as the frequency of transitions immediately prior to and after such periods.

The ethical attachment of the devices to individuals requires the use of sub-optimal attachment methods that prioritises the comfort of the dog over data integrity. To be comfortable for extended periods it is necessary that collars are not fit too tightly. Therefore rotation of the device, about the neck, throughout the instrumented period, was unavoidable and may have produced additional confounding acceleration profiles within the data, potentially highly disruptive due to the aim of detecting posture. As such a modified rotation correction algorithm was applied to the data adapted from that described in (Ladha *et al.*, 2017).

Acceleration data were loaded into R 3.6.0 (R Core Team, 2018) and the static component was calculated using the method described in Section 2.3. The mean of the data was taken across each 5 second window, at 2.5 second intervals. Data were then converted from Cartesian coordinates to spherical coordinates. When plotted as coordinates a 1g sphere about the origin describes the acceleration due to gravity acting upon a stationary object. It follows that when the sensor is stationary it falls near or upon the described sphere. By searching the spherical coordinates of the averaged data the windows signifying stationary periods can be identified. As this gravitational component reduces minutely with elevation, the fact a living organism will never be entirely still and that there exists a degree of signal noise, a leniency of $\pm 0.2g$ was used to signify candidate windows and all windows with a ρ coordinate greater than $0.8g$ and less than $1.2g$ were extracted. A target stationary orientation was described to coincide with the ideal ventral orientation as shown in Figure 2.1, this would be approximately equivalent to $(0,0,1)$ in 3-dimensional Cartesian coordinates. As the collar can rotate only around the neck of the dog, and therefore can only rotate about

the X axis, here equivalent to the dog's medial-lateral plane (Figure 2.1), correction was only attempted in the X axis. Rotation matrices corresponding to all possible rotations about the X axis in 1° intervals were identified and applied. The resultant coordinates with the shortest vectorial distance from the target coordinates was selected. This matrix is then applied to each subsequent reading until a new stationary period is identified and any rotation that had subsequently occurred could be accounted for and corrected. The function code is available in Appendix D.

3.3.4 Feature Extraction

Once corrected the data were used to generate features across windows of set lengths (2 seconds, 3 seconds and 5 seconds) centred on the annotation midpoints. The window lengths used included the mean annotation period rounded down, average annotation period rounded up and 5 seconds to ensure no transition is captured incompletely. Features were calculated for each axis individually as well as for a combination measure, the signal vector magnitude of the three axes. The signal vector magnitude (VM^3) is described in Section 2.3. VM^3 provides a summary of acceleration across all three axes that would be unaffected by any rotational interference that has not been fully corrected by the prior algorithm.

A collection of 13 time-domain features, common in both human and animal activity classification literature, were calculated to attempt to characterise the windows of acceleration and any inter-relationships (Martiskainen *et al.*, 2009; Nathan *et al.*, 2012; Ravi *et al.*, 2005; Rodríguez-Martín *et al.*, 2015; Shepard *et al.*, 2008b; Watanabe *et al.*, 2005). Features are calculated using the appropriate, in-built, R functions except where stated. Included were 3 measures of central tendency: mean, median and root mean square. Root mean square was calculated by taking the square root of the window mean. 6 measures of statistical dispersion were also calculated; minimum value, maximum value, range, standard deviation, mean absolute deviation and, interquartile range. Shannon's entropy was calculated per window to give an indication of the internal stochasticity, or predicability, of acceleration data within a window and was calculated using the entropy function of the entropy package (Hausser & Strimmer, 2014). Skewness and kurtosis of the windows were also calculated to provide an indication of the asymmetry and tailedness of the focal window using functions included in the e1071 package (Meyer *et al.*, 2019). The paired correlation between axes was calculated for each combination of the X, Y and Z axes as described in Ravi *et al.* (2005). Calculating each of these across the 3 axes and VM^3 results in 51 features per window. All feature calculation code can be found in Appendix E.

The above features were calculated with the aim of characterising the distribution of the acceleration data of each axis and to identify the strength or absence of any

relationships between axes. Due to the nature of accelerometer data it is often difficult to fully summarise with simplistic and computationally efficient methods such as those previously described. Hammerla *et al.* (2013) propose the Empirical Cumulative Distribution Function (ECDF) as an alternative approach to the characterisation of data that can more precisely represent the statistical characteristics of time series data when compared to the calculation of more traditional statistical features and their selection. Both Ladha *et al.* (2013) and Kumpulainen *et al.* (2018) have employed this method with domestic dogs using K-nearest neighbour and discriminant analysis based classifiers respectively. The ECDF for each window of each axis, and VM³, was calculated using the ECDF function that is part of base R. From here, the inverse ECDF is calculated through interpolation and 100 equally spaced points are selected to produce a 100 dimension data representation for further selection. The number of calculated ECDF coefficients was limited to 100 to limit the computational strain of the method and avoid problems of dimensionality when training classifiers. The function provided in Appendix F was used to achieve this.

The initial method of feature selection was a manual selection drawing on domain knowledge, and related inferences, and experience of the data from prior processing. This subset excluded all X axis variables as the movements within the medial-lateral plane have been subdued through the application of the rotation correction algorithm and because the movements involved in postural transitions were observed as being most extreme along the anterior-posterior and dorsoventral planes. Root mean square was the only measure of central tendency to be retained, as an established measure of acceleration magnitude within animal remote monitoring (Spivey & Bishop, 2013), and the minimum, maximum and standard deviation features were included as dispersion metrics. Kurtosis, skewness, entropy and the correlation between the Y and Z axis were also included.

3.3.5 Model Implementation and Feature Selection

LDA is applied here using the MASS R package implementation (Ripley, 2008; Venables & Ripley, 2002).

A k-fold cross-validation method, where $k = 5$, was used to assess the performance of the entire start-to-finish procedure of model building and feature selection. To allow for a final validation of models resulting from the cross-validation a total of 6 folds were sampled using a stratified sampling method as implemented in the caret package partitioning functions (Kuhn. *et al.*, 2019). One fold was removed from the data, so to not be used in the training of the models, and designated as the validation set. The 5 remaining folds were then used within the k fold cross-validation. The rarity of the "Lie to Sit" and "Sit to Lie" classes (see Table 3.2) results in a distinct class imbalance

Table 3.2: Frequencies of postural transitions and the randomly sampled "unknown" class

Transition Classifica- tion	Frequency
Lie -> Sit	13
Lie -> Stand	186
Sit -> Lie	56
Sit -> Stand	375
Stand -> Lie	145
Stand -> Sit	570
Unknown	1345

across many classification tasks focussing on these two transitions. This therefore necessitated the removal of these classes or of entire classification tasks to alleviate the threat of over-fitting models when subsampled for both validation and cross-validation and to reduce the adverse effects of class imbalance on the performance of the LDA classifier (Xie & Qiu, 2007).

At any point where model assessment is required during cross validation, to compare the performance of two feature subsets within a stepwise methodology for example, the area under the curve (AUC) of the receiver operating characteristics (ROC) curve, where model True Positive rate (sensitivity) is plotted against the False Positive rate (1-specificity), is used as produced by the ROCR package (Hanley & McNeil, 1982; Sing *et al.*, 2005). The AUC is used to assess performance here as it is easily comparable across differing models and feature subsets. It presents a measure of performance accounting for all possible classification thresholds, all points on the ROC curve, and indicates the probability that the classification algorithm will be able to effectively separate the positive and negative classes. The higher the AUC, the more pronounced the ROC curve, and the more effective a model is at consistently maximising the True Positive rate and minimising the False Positive rate across all classification thresholds.

A suite of 8 different classification tasks were proposed and assessed for their performance in either the detection of postural transitions or in the ability to identify the transition occurring. The first of these reclassifies all classes, except those labelled as "Unknown", into a generic "Postural Transition" class. The aim was to establish the performance of using feature sets drawn from literature, along with a simplistic classification model, for the identification of transitional events from among a random sampling of non-transitional periods. It is the only model to include the "Unknown" class as other

models operate on the assumption that transition identification has been previously performed using this, or an alternative, methodology. The 4 subsequent tasks focus on the discrimination of a single class (i.e. Lie to Stand, Sit to Stand, Stand to Lie, and Stand to Sit) from all others. This is achieved by labelling the focal transition as the positive class and all others as a generic negative class. Lie to Sit and Sit to Lie classes are not included as positive classes in this set of models due to their low frequencies within the dataset, contributing less than 1% and less than 5% of the dataset respectively, when not including the unknown class (Table 3.2). Such low frequencies of a positive class can cause difficulties in the assessment of performance using the AUC and the undersampling of data to address this would negatively impact the wider assessment of model performance (Berrar & Flach, 2011). The 6th and 7th tasks paired classes that feature inverse movements (i.e. Lie to Stand/Stand to Lie, and Sit to Stand/Stand to Sit). The Sit to Lie/Lie to Sit model was not included, this was once again due to low class frequencies. Classification task 8 paired the classes into two combinations delineated by their direction of movement, up from a lying position to that of a sit or stand posture or down to a lying position from a sit or stand posture. In this model transitions between sit and stand are excluded as it is intended to examine the ability of the calculated features to recognise the distinct forelimb extension motion common to the transition from lie to both other positions, as well as the inverse motion.

As stated previously many of the calculated features are variations upon similar statistical representations of the data (e.g. mean, median and root mean square) and as such are likely to exhibit high multicollinearity. To address this, prior to feature selection multicollinearity is assessed through the calculation of the variance inflation factor (VIF). The VIF assesses the degree to which the variance of one feature is increased due to the existence of in-feature multicollinearity (James *et al.*, 2013; Mansfield & Helms, 1982; Zuur *et al.*, 2010). It is used here to provide estimates of linear and higher order collinearity among features and implemented using a stepwise methodology that removes the feature with the highest VIF value and recalculating VIF for the remaining feature-set until all VIF values are below a threshold of 5 (James *et al.*, 2013).

Each of the above described classification tasks were run on reduced feature subsets excluding collinear variables. Further refinement of feature subsets was then performed using 4 simple and distinct methods of feature selection. Feature selection methods allow the selection of optimal subsets of features that can lower the computational overhead of future implementations of a classification model by removing the need to calculate extensive numbers of features and hold them in memory, and simultaneously allows for the improvement of model performance by identifying the most relevant features and removing those with no or negative performance contribution. The ECDF representation was included as a single feature and its dimensionality was

tuned within the cross-validation as a model hyper-parameter, with number of dimensions reported with model results. This was performed after the selection of other features and performance of the models was calculated both with and without the inclusion of the ECDF representation. Models were also produced that used the ECDF representation as the only feature, as presented in Vij *et al.* (2017).

One model was processed using the manually selected features mentioned previously. An alternative subset of features was selected through forward stepwise selection. An empty model was populated feature by feature based on the largest improvement to the AUC of the ROC curve. When no further improvement is gained by the addition of features, or the AUC remains stable despite the addition of features, the process is terminated and the posterior probabilities of the resultant model calculated. A third reduced feature set was then calculated using the inverse methodology to forward stepwise selection, backward stepwise feature selection, beginning with a full model and removing features in order of which results in the greatest improvement to the AUC. The fourth method of feature selection is correlation feature selection (CFS) as described in Hall (1999) and implemented in the FSelector R package (Romanski & Kotthoff, 2018). CFS uses a best-first search algorithm to calculate the feature subset that best maximises the correlation of each feature with the dependent variable while simultaneously minimising the correlation between features.

Once feature subsets have been calculated an additional step of adding and tuning the ECDF representation for each individual axis and combinations of axes is performed. ECDF representation based features are assessed both as the only feature set and as features included alongside feature subsets established by the previous selection methods. The 100 equally spaced inverse ECDF points that were previously calculated are repeatedly sub-setted to produce feature sets of between 1 and 100 equally spaced points. These are then assessed for the feature number and which combination of axes provides the highest AUC.

Posterior probabilities of test data produced by each LDA model, calculated within each fold of the cross-validation loop are combined and exported for interpretation alongside the coinciding observed classifications. The tuned ECDF representation numbers and model subsets are also exported for each fold and LDA model. Performance of classifiers and feature subsets produced by the differing selection methods is presented using 5 distinct metrics. AUC is calculated as in the cross-validation performance assessment from the ROC curve and provides an indication of general classifier performance at all potential decision thresholds and gives an indication of the models ability to effectively separate the positive and negative classes. The higher the AUC, the more pronounced the ROC curve, and the more effective a model is at consistently maximising the True Positive rate and minimising the False Positive rate across all classification thresholds.

The ROC curve is then used to calculate an optimal threshold to use when calculating a number of other performance measures that require the threshold to be set. This threshold was determined by finding the point at which the True Positive Rate (TPR) was maximised and False Positive Rate (FPR) was minimised through the identification of the point where subtraction of the FPR from the TPR is highest. These thresholds are then used to calculate 3 performance metrics used here.

The first of these is the TPR at the threshold selected. The TPR is also known as the Sensitivity and indicates the proportion of positive windows correctly identified. As previously stated a ROC curve is the TPR plotted against the FPR. The FPR is calculated as 1-Specificity where the Specificity indicates the proportion of correctly identified negative windows within the total number of negative windows. The 3rd of the performance measures used for assessment is the model Precision at the selected threshold. Precision is the proportion of correctly identified positive class windows from all windows identified as the positive class. Following these descriptions the best performing model is that which produces the highest values of each, therefore minimising the occurrence of misclassification of both positive and negative classes (Sensitivity and Specificity) and ensures that the selection parameters are not causing excessive misclassification by facilitating a high TPR at the expense of a high FPR (Precision). To rapidly assess each of these metrics combination measures that allow insight into the balance between Precision, Sensitivity, and Specificity are used here.

The first of the 2 combination measures is the AUC which has already been calculated and allows insight into the overall performance of the model with respect to the interplay of the Sensitivity and Specificity. The second to be calculated is the F measure which is the harmonic mean of the Sensitivity and Precision values and indicates the effectiveness of the model at correctly identifying positive classes without generalising excessively and causing a high number of False Positives. F measure was chosen over the more ubiquitous accuracy measure as accuracy often fails to control for this occurrence in what is commonly termed the "Accuracy Paradox". Both of these combination measures provide general summaries of the model performance and can be used to rapidly assess the overall performance of a model relative to alternatives. Where two different models provide the highest values of AUC and F score the sensitivity, specificity, and precision will then be compared individually for a more detailed indication of the assets and liabilities of each model.

Once a model building methodology has been selected data used within the cross validation is recombined and used to train a final model. This final model is created through a rudimentary feature voting method that draws variables used in greater than 50% of cross validation folds, taken here as a simplistic measure of model importance, from the most highly performing cross validated model. This is then applied to the previously excluded validation dataset to generate final performance metrics of a model

on data to which it is entirely naive and therefore to indicate the degree of over-fitting that may have occurred.

3.4 Results

The manually selected feature subset was consistent across the different models and was constructed to attempt to characterise both the anterior-posterior and dorsoventral acceleration based on knowledge gained during the video observation and annotation procedure. Additionally the application of the rotation correction is such that it will potentially, as well as adjusting for rotation of the sensor in either medial-lateral direction, subdue medial-lateral acceleration and the accentuate dorsoventral acceleration. The VM³ focussed features were also not included. VM³ functions as a summary of all three axes and as a representation of total acceleration magnitude and so with the inclusion of two of the three axes being summarised, coupled with the observed lack of variation in visually assessed postural transition energy noted during the annotation procedure, it was decided that the inclusion of VM³-based features would have limited impact.

The assessment of VIF revealed a varying number of features with strong relationships dependant on the classification problem posed (i.e. postural transition identification from randomly sampled miscellaneous data, Lie to Stand from all other postural transition events, etc.) and the window length.

3.4.1 Postural Transition Identification

The detection of postural transition events from among randomly sampled miscellaneous data is shown in Table 3.3. As previously described the data were divided into two parts, a combined class of postural transitions (N = 1345), and class comprised of randomly sampled non-transitional data (N = 1345). One sixth of the data was removed for validation (N = 448) resulting in a training set of 2242 cases that was further assigned into 5 cross-validation folds. A 50/50 class composition, consistent with the full data set, was maintained for training, testing and validation datasets.

When using a 2 second window length the Backward + ECDF method provides the greatest values for both AUC (0.618) and F score (0.638), and is therefore selected for further comparison. This method reports the highest value of sensitivity (0.706) relative to the other methods assessed and lower relative values of specificity (0.492) and precision (0.582).

Window lengths of 3 seconds and the use of a Backward + ECDF derived feature set again produce the highest values of AUC (0.619) and F score (0.635), and again the sensitivity (0.695) of the models devised through this method is the highest reported when using this window size. The specificity (0.506) and precision (0.584) are lower

than those reported by alternative models. As such, the Backward + ECDF method will be considered for use in the creation of the validation model.

The performance of the 5 second window methods shows both the Forward + ECDF and Backward + ECDF feature selection procedures as viable options. The Forward + ECDF method produces the highest F score (0.634), and second highest of both AUC (0.632) and sensitivity (0.700). The values of specificity (0.490) and precision (0.578) are low relative to other methods assessed using a 5 second window. In contrast, the assessment of the performance of the Backward + ECDF method reveals the highest reported value of AUC (0.646), the third highest F score (0.617), and the second highest Precision (0.605) relative to the alternative methods. Both sensitivity (0.630) and specificity (0.590) are low when compared across the assessed methods. As the Forward + ECDF performance provides the highest relative rankings for the two combined measures it is the candidate selected as the best option for the 5 second window length.

A comparison of the selected models from each of the three window sizes fails to clearly expose any one option as optimal when considering the combined measures, AUC and F score. Assessment of models focussing on metrics of sensitivity, specificity and precision reveals an imbalance between sensitivity and the other metrics in methods using either the 2 second or 5 second window sizes. As such the Backward + ECDF method using a 3 second window size is chosen to construct the validation LDA model. Figure 3.1 shows how frequently features and ECDF representations were included within the 5 cross-validation folds. 10 features were included in greater than 50% of cross-validation folds alongside the ECDF representations of the Z axis and the VM³. Both ECDF representations were comprised of a mean of 28 (\pm 33 SD) ECDF coefficients resulting in a validation model using a total of 66 features, with 56 of those being ECDF coefficients. The AUC of the resultant LDA model, when applied to the validation data, is 0.560 and the decision threshold, selected to maximise TP classifications and minimise FP classifications, is 0.511. At this threshold the sensitivity is 0.460, specificity is 0.634, precision is 0.557, and F score is 0.504.

Table 3.3: The performance statistics of the identification of postural transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted.

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
2 Seconds	Manual	0.551	0.387	0.706	0.568	0.460	0.522
	Manual + ECDF	0.588	0.484	0.664	0.590	0.532	0.514
	ECDF	0.590	0.665	0.477	0.560	0.608	0.487
	Forward	0.537	0.525	0.549	0.537	0.531	0.500
	Forward + ECDF	0.588	0.459	0.684	0.592	0.517	0.513
	Backward	0.611	0.685	0.509	0.583	0.630	0.485
	Backward + ECDF	0.618	0.706	0.492	0.582	0.638	0.484
	CFS	0.559	0.604	0.498	0.546	0.573	0.490
	CFS + ECDF	0.591	0.669	0.490	0.567	0.614	0.489
3 Seconds	Manual	0.556	0.522	0.583	0.556	0.538	0.492
	Manual + ECDF	0.603	0.633	0.529	0.574	0.602	0.483
	ECDF	0.609	0.618	0.550	0.579	0.598	0.495
	Forward	0.559	0.459	0.632	0.554	0.502	0.500
	Forward + ECDF	0.609	0.657	0.508	0.572	0.612	0.489
	Backward	0.601	0.548	0.621	0.591	0.569	0.493
	Backward + ECDF	0.619	0.695	0.506	0.584	0.635	0.482
	CFS	0.555	0.638	0.452	0.538	0.584	0.486
	CFS + ECDF	0.607	0.543	0.629	0.594	0.568	0.499
5 Seconds	Manual	0.571	0.706	0.409	0.544	0.615	0.462
	Manual + ECDF	0.623	0.616	0.576	0.593	0.604	0.490
	ECDF	0.624	0.398	0.780	0.644	0.492	0.534
	Forward	0.568	0.490	0.617	0.561	0.523	0.496
	Forward + ECDF	0.632	0.700	0.490	0.578	0.634	0.474
	Backward	0.615	0.633	0.561	0.591	0.611	0.486
	Backward + ECDF	0.646	0.630	0.590	0.605	0.617	0.492
	CFS	0.555	0.425	0.667	0.561	0.483	0.502
	CFS + ECDF	0.628	0.646	0.555	0.592	0.618	0.485

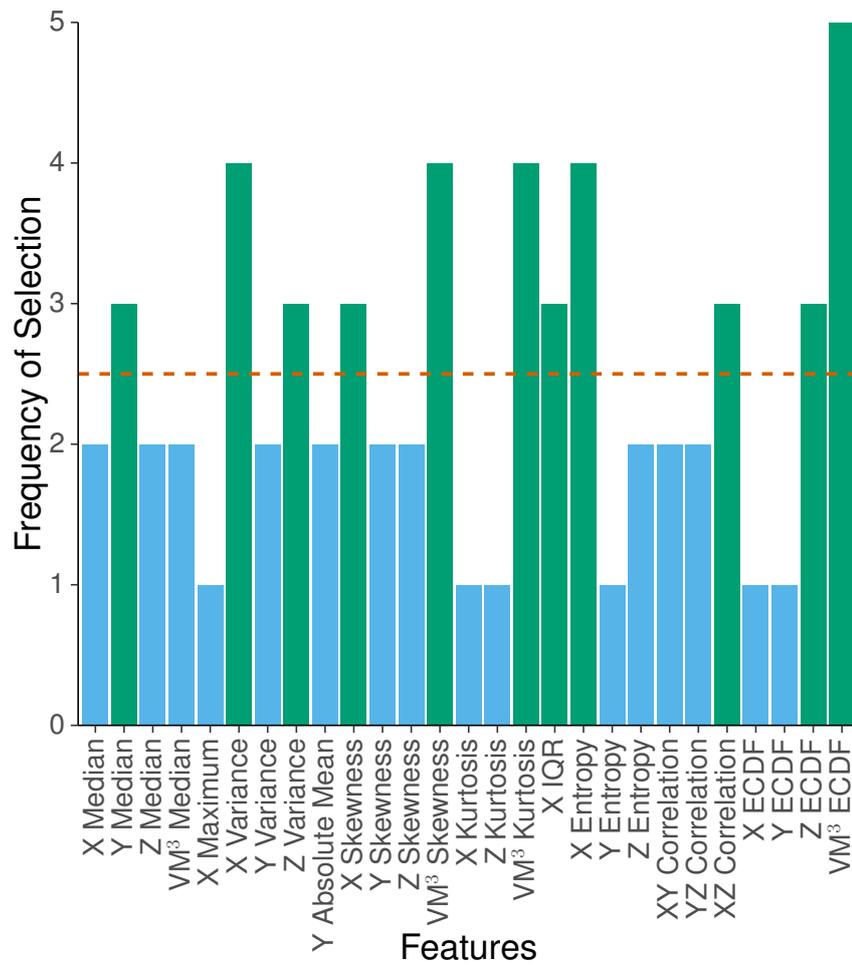


Figure 3.1: The frequency of inclusion of features within cross-validation of Postural transition classification models. The 2.5 threshold is marked.

3.4.2 Lie to Stand Transition Identification

Randomly sampled non-transitional data were excluded, all non-Lie to Stand transitions were grouped into a combined, negative class ($n = 1159$), and Lie to Stand transitions were assigned to the positive class ($n = 186$). A validation dataset comprising one sixth of the data was removed for validation ($n = 224$) resulting in a training set of 1121 cases to be further split into train and test sets through the 5 fold cross-validation. Class composition was maintained within each data subset. Performance metrics of the automatic differentiation of Lie to Stand transitions from the combined class of all other postural transitions is shown in Table 3.4.

Using a 2 second window length the Backward + ECDF is selected for further consideration. The values of AUC (0.672) and F score (0.332) are the highest reported using this window length, satisfying the selection criteria. The values of sensitivity (0.768) and precision (0.212) are the second and third highest recorded respectively but the specificity (0.541) is relatively low for the window size.

The 3 second window length presents Backward + ECDF feature selection as the optimal method choice as it gives the highest values of AUC (0.688) and F Score (0.351) as well as relatively high values of both specificity (0.654) and precision (0.237). The sensitivity (0.671) of models created with this method of feature selection is lower than alternatives but at the cost of the other metrics.

A window size of 5 seconds again suggests a Backward + ECDF method as having produced the best performing feature set due to it providing the highest values for AUC (0.719), sensitivity (0.742), precision (0.257) and F score (0.381). However, the value of specificity (0.655) given is lower than those of other methods of feature selection using a 5 second window size.

Comparison across the three window sizes reveals that the use of a Backward + ECDF feature selection method and a 5 second window size has produced the highest performing feature sets, except in terms of specificity where a 2 second window size and Backward only method perform better. As both the AUC and F score of the method, functioning as representations of the balance between the sensitivity/specificity and sensitivity/F score dichotomies, are higher than the two alternative window sizes this is the method chosen to construct the validation model.

Figure 3.2 shows the frequency of features selected using the Backward + ECDF method, for 5 second window length data, across the 5 folds of the cross-validation. 10 features occurred in greater than 50% of cross validation folds and therefore were used alongside the ECDF representation of the VM³, which also met this inclusion criteria. A mean of 9 (± 7 SD) ECDF coefficients per fold were used during cross validation. This number of coefficients were combined with the selected features to build an LDA model consisting of 19 total features for validation on the previously withheld validation

data. The AUC of the validated model was 0.463. A decision threshold of 0.062 was selected as previously described. At this cut-off the sensitivity of the validation model was reported as 1.000 but specificity, precision and F score were 0.067, 0.147, and 0.256 respectively.

Table 3.4: The performance statistics of the identification of lie to stand transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted

	AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold	
2 Seconds	Manual	0.554	0.729	0.392	0.161	0.264	0.121
	Manual + ECDF	0.620	0.594	0.638	0.209	0.309	0.137
	ECDF	0.631	0.690	0.571	0.205	0.317	0.137
	Forward	0.555	0.581	0.543	0.169	0.262	0.139
	Forward + ECDF	0.642	0.484	0.762	0.246	0.326	0.160
	Backward	0.651	0.490	0.759	0.246	0.328	0.161
	Backward + ECDF	0.672	0.768	0.541	0.212	0.332	0.130
	CFS	0.495	0.916	0.159	0.149	0.256	0.129
	CFS + ECDF	0.649	0.542	0.704	0.227	0.320	0.152
3 Seconds	Manual	0.556	0.606	0.518	0.168	0.263	0.137
	Manual + ECDF	0.641	0.626	0.588	0.196	0.298	0.130
	ECDF	0.659	0.710	0.588	0.217	0.332	0.140
	Forward	0.553	0.458	0.654	0.175	0.254	0.145
	Forward + ECDF	0.631	0.484	0.761	0.245	0.325	0.160
	Backward	0.642	0.594	0.688	0.234	0.336	0.150
	Backward + ECDF	0.688	0.671	0.654	0.237	0.351	0.147
	CFS	0.496	0.968	0.070	0.143	0.249	0.121
	CFS + ECDF	0.660	0.723	0.553	0.206	0.320	0.139
5 Seconds	Manual	0.626	0.652	0.566	0.194	0.299	0.127
	Manual + ECDF	0.684	0.613	0.681	0.236	0.341	0.141
	ECDF	0.664	0.581	0.706	0.241	0.340	0.151
	Forward	0.599	0.490	0.683	0.199	0.283	0.143
	Forward + ECDF	0.686	0.665	0.673	0.246	0.359	0.143
	Backward	0.695	0.684	0.643	0.235	0.350	0.141
	Backward + ECDF	0.719	0.742	0.655	0.257	0.381	0.143
	CFS	0.507	0.161	0.885	0.184	0.172	0.167
	CFS + ECDF	0.649	0.639	0.601	0.205	0.310	0.139

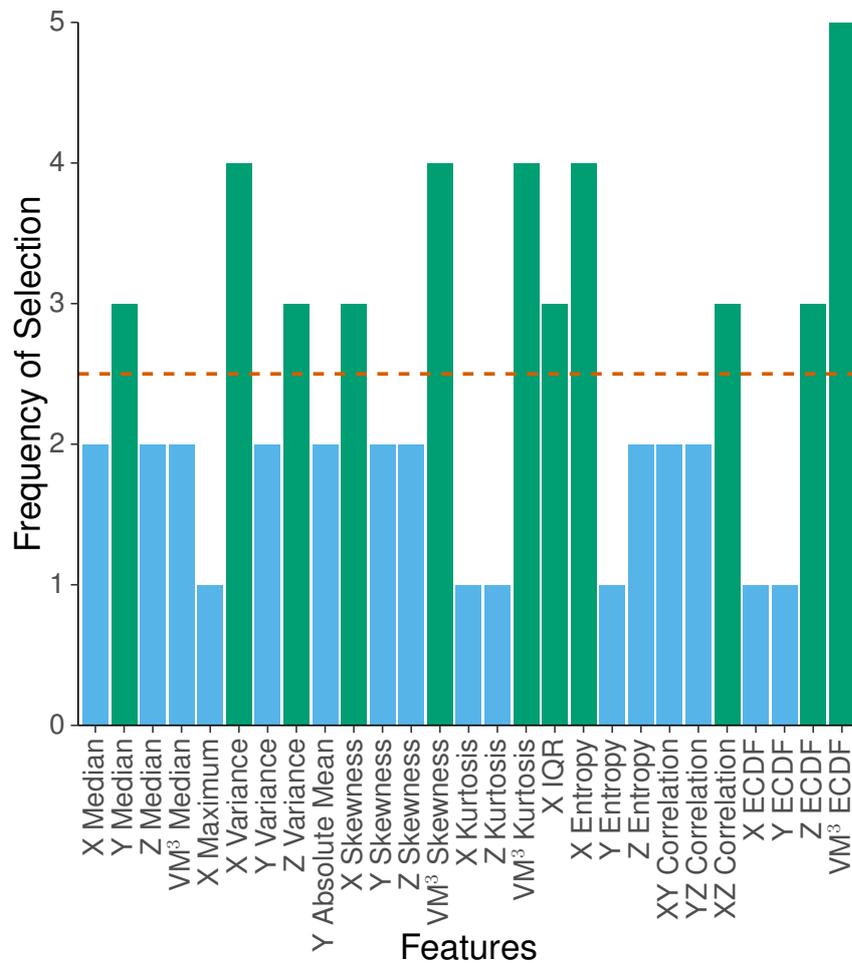


Figure 3.2: The frequency of inclusion of features within cross-validation of Lie to Stand transition classification models. The 2.5 threshold is marked.

3.4.3 Sit to Stand Transition Identification

Randomly sampled non-transitional data were excluded, all non-focal transitions were grouped into a combined, negative class ($n = 970$), and Sit to Stand transitions were assigned to the positive class ($n = 375$). A validation dataset comprising one sixth of the data was removed for validation ($n = 224$) resulting in a training set of 1121 cases to be further split into train and test sets through the 5 fold cross-validation. Class composition was approximately maintained within each data subset. The performance of feature subset selection methods for the classification of Sit to Stand postural transitions from among a combined dataset of other postural transitions is shown in Table 3.5.

When using a 2 second window size the Backward + ECDF provides the highest values of the two combined measures, AUC (0.656) and F score (0.505), relative to other methods. Using this method the sensitivity (0.728), and precision (0.386) are both the second highest among alternatives using a 2 second window length. However the specificity (0.553) is low relative to alternatives.

Increasing the window length to 3 seconds results in the Backward + ECDF methodology again being the optimal choice of model in comparison to alternatives of the same window size. The Backward + ECDF method produces the highest values of both of the combined measures, AUC (0.641) and F score (0.495). Additionally this technique provides the highest value of precision (0.384), the second highest value of specificity (0.569), and the third highest value of sensitivity (0.696) relative to other 3 second window size methods.

The 5 second window length performance metrics again suggest the Backward + ECDF method of feature selection as the technique which produces the best classifier. With the highest values of both AUC (0.686) and F Score (0.520) the Backward + ECDF methodology also presents the best value of precision (0.446) and third best of specificity (0.700). The sensitivity (0.623) of the technique is lower than alternative methods but is the highest attained that did not adversely impact specificity or precision.

A comparison of the Backward + ECDF method across each of the three window lengths reveals that a 5 second window size, using a Backward + ECDF method of feature selection produces the most effective LDA classifier with higher values for the two combined measures, specificity, and precision. The increase to both combined measures confirms that the decrease in sensitivity is offset by the higher levels of specificity and precision exhibited by model using this method. Figure 3.3 shows which features and ECDF representations were selected during cross validation of the model and the frequency of occurrence for each. 17 features occurred in greater than 50% of cross-validation folds and are included in the validation model. No ECDF representation occurred in greater than 50% of folds but, as shown in Table 3.5, the inclusion of an ECDF representation had a positive effect on performance, when compared to a Backward only approach, a finding consistent across the three window size options. As such the ECDF representation with the highest frequency, VM³ is included in the final model. A mean of 53 (\pm 19 SD) ECDF coefficients are included in ECDF representations during the 5 fold cross-validation. The validation LDA model is therefore comprised of 70 individual features, with 53 of these being the VM³ ECDF coefficients. Testing the newly trained LDA model, using this feature subset, on the validation data gives an AUC of 0.551. At an optimised decision threshold of 0.250 the Sensitivity is 0.661, Specificity is 0.481, Precision is 0.328, and F Score is 0.439.

Table 3.5: The performance statistics of the identification of sit to stand transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
2 Seconds	Manual	0.522	0.738	0.326	0.298	0.424	0.250
	Manual + ECDF	0.604	0.498	0.670	0.369	0.424	0.291
	ECDF	0.605	0.482	0.691	0.377	0.423	0.280
	Forward	0.547	0.684	0.443	0.322	0.438	0.271
	Forward + ECDF	0.624	0.629	0.590	0.372	0.468	0.257
	Backward	0.622	0.728	0.503	0.362	0.484	0.271
	Backward + ECDF	0.656	0.728	0.553	0.386	0.505	0.271
	CFS	0.490	0.112	0.928	0.376	0.172	0.284
	CFS + ECDF	0.614	0.556	0.663	0.389	0.458	0.270
3 Seconds	Manual	0.519	0.545	0.522	0.306	0.392	0.277
	Manual + ECDF	0.609	0.769	0.449	0.350	0.481	0.233
	ECDF	0.607	0.696	0.491	0.346	0.462	0.261
	Forward	0.543	0.631	0.458	0.310	0.416	0.273
	Forward + ECDF	0.621	0.702	0.520	0.361	0.477	0.262
	Backward	0.612	0.638	0.579	0.369	0.468	0.279
	Backward + ECDF	0.641	0.696	0.569	0.384	0.495	0.270
	CFS	0.512	0.606	0.448	0.298	0.399	0.278
	CFS + ECDF	0.620	0.663	0.558	0.367	0.473	0.257
5 Seconds	Manual	0.557	0.348	0.769	0.368	0.358	0.313
	Manual + ECDF	0.659	0.613	0.667	0.416	0.496	0.284
	ECDF	0.649	0.732	0.537	0.380	0.500	0.257
	Forward	0.531	0.735	0.329	0.298	0.424	0.269
	Forward + ECDF	0.657	0.687	0.579	0.387	0.495	0.256
	Backward	0.634	0.559	0.668	0.395	0.463	0.295
	Backward + ECDF	0.686	0.623	0.700	0.446	0.520	0.288
	CFS	0.537	0.760	0.329	0.305	0.435	0.267
	CFS + ECDF	0.649	0.553	0.707	0.422	0.479	0.278

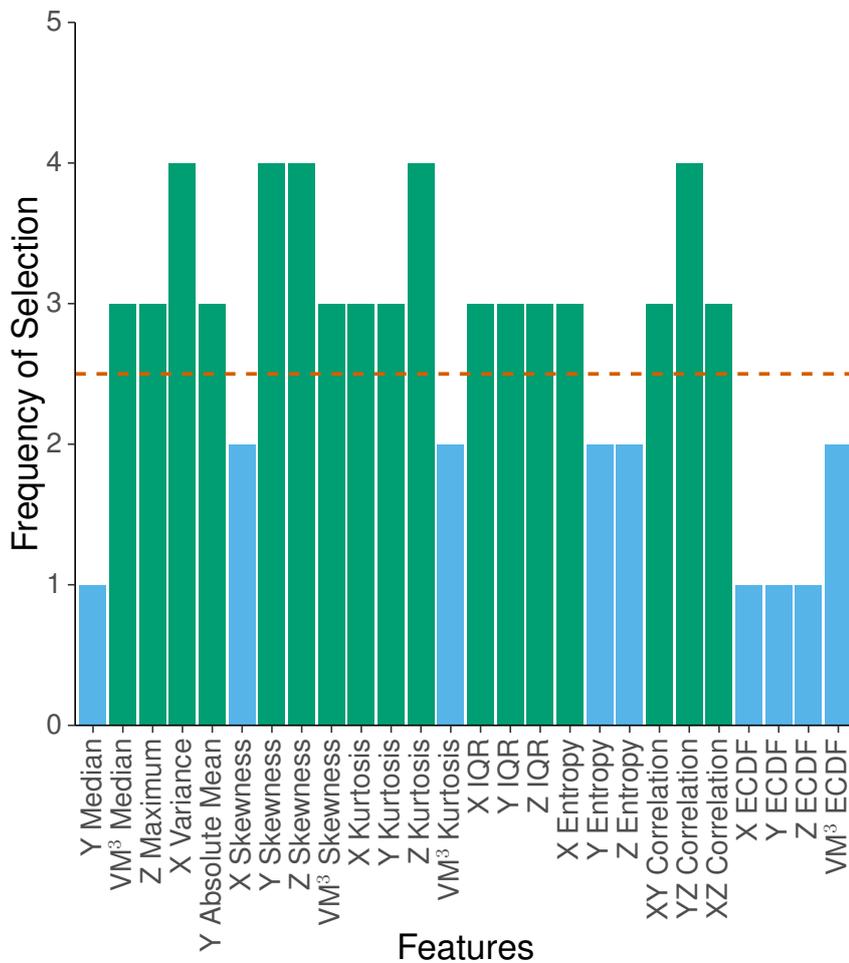


Figure 3.3: The frequency of inclusion of features within cross-validation of Sit to Stand transition classification models. The 2.5 threshold is marked.

3.4.4 Stand to Lie Transition Identification

Randomly sampled non-transitional data were excluded, all non-focal transitions were grouped into a combined, negative class ($n = 1200$), and Stand to Lie transitions were assigned to the positive class ($n = 145$). A validation dataset comprising one sixth of the data was removed for validation ($n = 224$) resulting in a training set of 1121 cases to be further split into train and test sets through the 5 fold cross-validation. Class composition was approximately maintained within each data subset. Table 3.6 shows the performance metrics of feature selection methodologies for the design of LDA classifiers of Stand to Lie transitions across 3 window lengths.

With data segmented into windows of 2 seconds the Backward + ECDF methodology is the most consistently well-performing method. The technique gives the highest values for the two combined measures, AUC (0.735), and F Score (0.348), and the highest reported value of precision (0.226). Additionally Backward + ECDF gives the second highest level of specificity (0.688), and the third highest of sensitivity (0.758) making it the optimal choice for the window size.

Using a data window length of 3 seconds results in a similar outcome. The Backward + ECDF methodology provides the best performing feature subset with the highest values of both combined methods, AUC (0.738) and F score (0.355), and of specificity (0.773), and precision (0.248). The sensitivity (0.620) is lower than alternative techniques but, as shown by the high values of the combined measures, this reduction is offset by the levels of both specificity and precision.

A 5 second window length presents Backward + ECDF as, once again, producing the best performing feature subsets. The AUC (0.732), F score (0.380), and precision (0.276) are the highest values given for the window size among all assessed feature selection methodologies. The specificity (0.806) is the second highest reported, where the higher method, the manually selected subset, features significant degradation in the other two measures and, as a result, in the combined measures. The sensitivity (0.612) is low relative to other strategies of feature selection but all those with a higher sensitivity also show degradation in specificity or precision.

Despite lower values of AUC and F score, when compared to the 3 second Backward + ECDF method, the 5 second Backward + ECDF method is selected as providing the best feature subsets for use in building a validation model. This is due to the higher levels of both specificity and precision being judged as more important to the current classification task than a slight decrease in sensitivity. Figure 3.4 shows the frequency at which features were selected across the 5 folds of the cross validation. 9 features and 1 ECDF representation, the representation of the Z axis, occurred in greater than 50% of cross validation folds. A mean of 11 (± 7 SD) ECDF coefficients were used to form the ECDF representations used in the Backward + ECDF method of each cross validation fold. The final validation model therefore consists of a total of 20 features, 11 of which are the Z axis ECDF coefficients. An AUC of 0.660 is attained when performance is assessed using the validation data. The selection of the decision threshold by maximising the TPR and minimising the FPR results in an optimal threshold of 0.111. The Sensitivity of the model is 0.625, the Specificity is 0.630, the Precision is 0.169, and the F Score is 0.265.

Table 3.6: The performance statistics of the identification of stand to lie transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
2 Seconds	Manual	0.593	0.508	0.665	0.154	0.236	0.117
	Manual + ECDF	0.655	0.700	0.558	0.160	0.260	0.098
	ECDF	0.682	0.733	0.575	0.172	0.278	0.118
	Forward	0.531	0.600	0.505	0.127	0.210	0.107
	Forward + ECDF	0.662	0.767	0.537	0.166	0.273	0.107
	Backward	0.699	0.667	0.697	0.209	0.318	0.121
	Backward + ECDF	0.735	0.758	0.688	0.226	0.348	0.121
	CFS	0.621	0.533	0.685	0.169	0.257	0.121
	CFS + ECDF	0.689	0.783	0.550	0.173	0.283	0.114
3 Seconds	Manual	0.582	0.463	0.683	0.150	0.227	0.117
	Manual + ECDF	0.649	0.777	0.497	0.157	0.262	0.085
	ECDF	0.698	0.769	0.569	0.177	0.288	0.114
	Forward	0.608	0.686	0.504	0.143	0.237	0.109
	Forward + ECDF	0.694	0.860	0.474	0.165	0.277	0.094
	Backward	0.706	0.769	0.591	0.185	0.299	0.111
	Backward + ECDF	0.738	0.620	0.773	0.248	0.355	0.133
	CFS	0.598	0.562	0.651	0.163	0.253	0.120
	CFS + ECDF	0.692	0.579	0.731	0.206	0.304	0.131
5 Seconds	Manual	0.560	0.289	0.849	0.188	0.228	0.153
	Manual + ECDF	0.644	0.645	0.599	0.163	0.260	0.094
	ECDF	0.668	0.512	0.770	0.212	0.300	0.137
	Forward	0.601	0.678	0.508	0.143	0.236	0.108
	Forward + ECDF	0.673	0.661	0.623	0.175	0.277	0.115
	Backward	0.708	0.645	0.756	0.242	0.352	0.129
	Backward + ECDF	0.732	0.612	0.806	0.276	0.380	0.140
	CFS	0.626	0.653	0.620	0.172	0.272	0.118
	CFS + ECDF	0.680	0.636	0.675	0.192	0.294	0.125

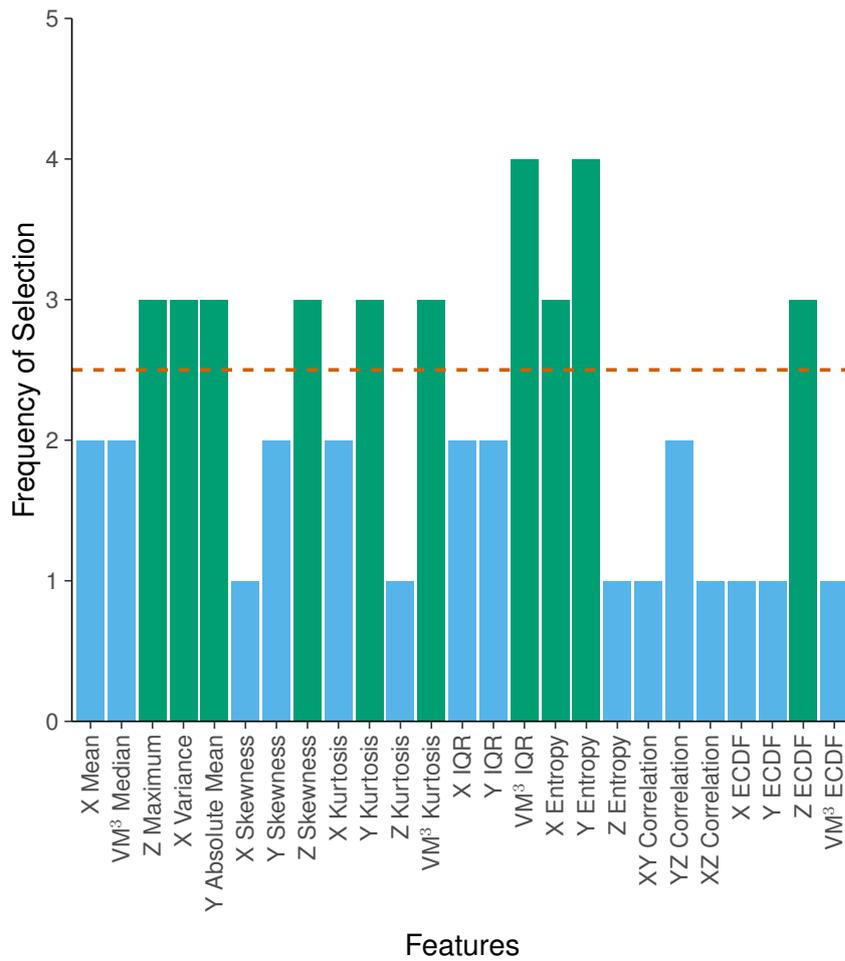


Figure 3.4: The frequency of inclusion of features within cross-validation of Stand to Lie transition classification models. The 2.5 threshold is marked.

3.4.5 Stand to Sit Transition Identification

Randomly sampled non-transitional data were excluded, all non-focal transitions were grouped into a combined, negative class ($n = 775$), and Stand to Sit transitions were assigned to the positive class ($n = 570$). A validation dataset comprising one sixth of the data was removed for validation ($n = 224$) resulting in a training set of 1121 cases to be further split into train and test sets through the 5 fold cross-validation. Class composition was approximately maintained within each data subset. The performance of LDA classifiers of Stand to Site transitions, using different feature selection methods and window sizes, is presented in Table 3.7.

When using a 2 second window length the Backward + ECDF feature selection methodology presents the most balanced performance with the highest values of AUC (0.662) and F score (0.606). The sensitivity (0.682) is the third highest of the window size and the precision (0.545) is the second highest. However, the specificity (0.582) of models produced using this method suffers relative to alternatives.

The Backward + ECDF methodology is the highest performing method when using a 3 second window length. The values of both AUC (0.670) and F score (0.615) are the highest among those using the same window length. The sensitivity (0.682) of models created using the method is the third highest of those using the same window length, and the precision (0.560) is the highest reported. A low specificity (0.605), compared to alternatives, is the highest achieved without negatively impacting other metrics of performance.

Using a 5 second window length the Backward + ECDF method once again produces the most balanced models in regards to performance across the calculated metrics. The AUC (0.669) and F score (0.586) are both the highest values reported for the window size despite all of the three contributing measures, sensitivity (0.594), specificity (0.681), and precision (0.578), being exceeded by alternative methods.

Comparing the 3 selected methods from across the 3 different sizes of windows shows 3 second windowed data with the Backward + ECDF method of feature selection to be the best choice with which to construct a validation model, as shown by the higher values of the two combination measures, from which it can be inferred that models constructed with this window size, using this method, exhibit a more balanced performance across the three individual performance factors. Using this method does require the sacrifice of some degree of specificity and precision, where the 5 second alternative excels, but the increased sensitivity is shown to offset this. Figure 3.5 shows the frequency of occurrence of each of the features selected by the method during cross validation. Of these 10 features and 1 ECDF representation were selected in greater than 50% of cross validation folds. The ECDF representation of the Y axis consisted of a mean of 13 coefficients (± 14 SD). Therefore, the resultant model consists

of 23 individual features. When tested using the validation data an AUC of 0.571 is given. 0.407 was selected as the optimal decision threshold, as described previously, and was used for the calculation of the other performance metrics. The Sensitivity of the model was 0.621, the Specificity was 0.546, the Precision was 0.500, and the F Score was 0.554.

Table 3.7: The performance statistics of the identification of stand to sit transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
2 Seconds	Manual	0.563	0.398	0.720	0.511	0.447	0.438
	Manual + ECDF	0.618	0.684	0.498	0.501	0.578	0.365
	ECDF	0.614	0.539	0.630	0.517	0.528	0.403
	Forward	0.578	0.764	0.365	0.470	0.582	0.397
	Forward + ECDF	0.633	0.434	0.763	0.574	0.494	0.450
	Backward	0.644	0.625	0.594	0.531	0.574	0.408
	Backward + ECDF	0.662	0.682	0.582	0.545	0.606	0.398
	CFS	0.575	0.566	0.565	0.489	0.525	0.409
	CFS + ECDF	0.616	0.669	0.526	0.510	0.579	0.387
3 Seconds	Manual	0.558	0.408	0.707	0.507	0.452	0.436
	Manual + ECDF	0.598	0.503	0.678	0.535	0.518	0.421
	ECDF	0.634	0.558	0.656	0.544	0.551	0.413
	Forward	0.601	0.724	0.429	0.483	0.580	0.409
	Forward + ECDF	0.632	0.545	0.660	0.542	0.544	0.411
	Backward	0.651	0.674	0.571	0.536	0.597	0.407
	Backward + ECDF	0.670	0.682	0.605	0.560	0.615	0.404
	CFS	0.513	0.817	0.239	0.441	0.573	0.382
	CFS + ECDF	0.617	0.493	0.709	0.555	0.522	0.429
5 Seconds	Manual	0.569	0.699	0.415	0.468	0.560	0.375
	Manual + ECDF	0.630	0.539	0.656	0.536	0.537	0.407
	ECDF	0.616	0.518	0.684	0.547	0.532	0.421
	Forward	0.601	0.615	0.551	0.502	0.553	0.418
	Forward + ECDF	0.647	0.383	0.841	0.639	0.479	0.468
	Backward	0.648	0.531	0.718	0.581	0.554	0.433
	Backward + ECDF	0.669	0.594	0.681	0.578	0.586	0.419
	CFS	0.547	0.423	0.666	0.482	0.451	0.426
	CFS + ECDF	0.622	0.691	0.503	0.505	0.584	0.370

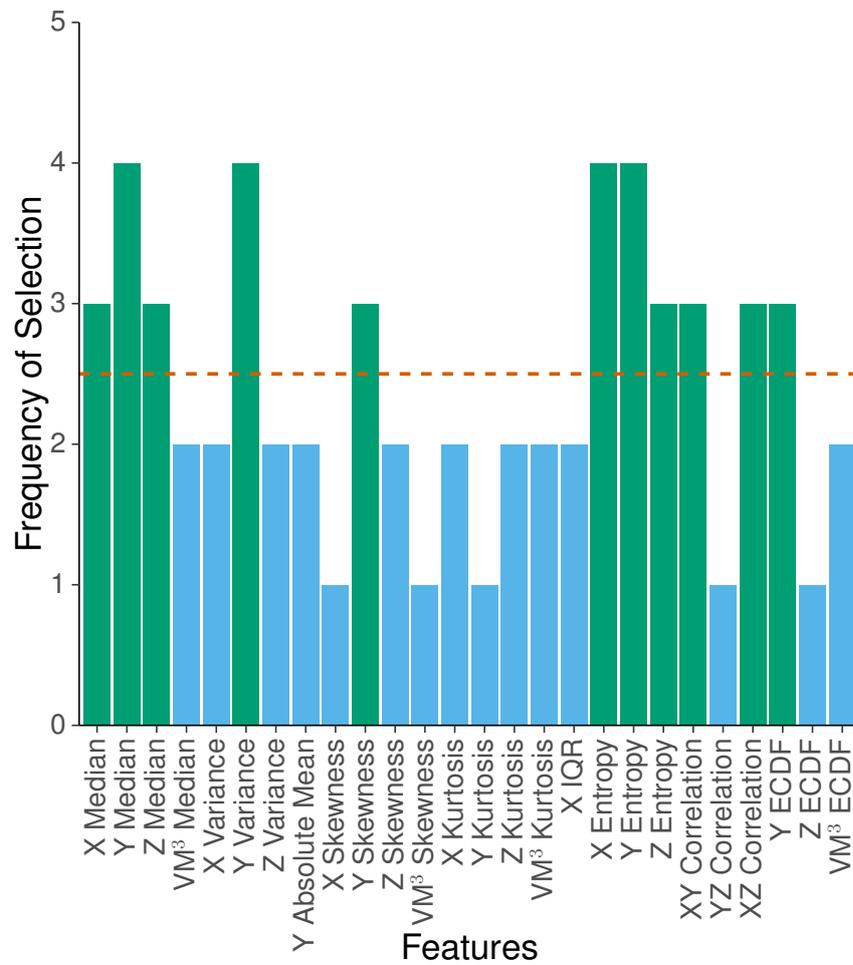


Figure 3.5: The frequency of inclusion of features within cross-validation of Stand to Sit transition classification models. The 2.5 threshold is marked.

3.4.6 Lie to Stand and Stand to Lie Combined Class Identification

Randomly sampled non-transitional data were excluded, and both Lie to Stand and Stand to Lie transitions were grouped into a combined positive class ($n = 331$). The negative class consists of the remaining transitions ($n = 1014$). A validation dataset comprising one sixth of the data was removed for validation ($n = 224$) resulting in a training set of 1121 cases to be further split into train and test sets through the 5 fold cross-validation. Class composition was approximately maintained within each data subset. The performance of the 5 fold cross validated performance across a range of feature selection methods and window sizes is presented in Table 3.8.

Both the Backward + ECDF and Backward only methods gave the best measures of performance when using data with a 2 second window length. The Backward + ECDF method gave the highest value of AUC (0.675), and the second highest value of F score (0.463), despite a low value of sensitivity (0.496). This is likely due to the method providing the highest values of both specificity (0.788), and precision (0.434). The Backward only method provides the second highest value of AUC (0.665) and highest of F score (0.470), a result of the heightened value of sensitivity (0.605) providing enough of an offset to the reduced precision (0.385) but insufficient for the reduction in specificity (0.684). The more strict, but more precise model, Backward + ECDF, is chosen for further consideration.

Performance measures for a window size of 3 seconds reveals the Backward + ECDF method as producing the best performing feature sets. The AUC (0.664) and F score (0.475) are both the highest among the alternative methods of feature selection using the 3 second window size. The sensitivity (0.725) and precision (0.353) are both the third highest reported and specificity (0.567) is relatively low in comparison to alternative methods.

The best method of feature selection for a 5 second window is once again the Backward + ECDF method. Both of the combined measures, AUC (0.696) and F score (0.492), are the highest reported. The sensitivity (0.815) is the second highest among the methods of feature selection using a 5 second window. The specificity was 0.510 and the precision was 0.352.

In a comparison of the models selected from each of the 3 window sizes the 5 second Backward + ECDF method is the best general model, when assessed by its values of the combination measures, AUC and F Score, by which it outperforms both other models. This is likely due to the high level of sensitivity given by this model that offsets sub-optimal levels of both precision and specificity. Figure 3.6 provides a list of all features included across the 5 folds of the cross validation, including which ECDF representation were used, and the frequency of feature occurrence. 10 non-ECDF based features were included in greater than 50% of cross validation folds. No ECDF

representation was included in greater than 50% of cross validation folds. As a result the most frequent ECDF representations were selected, the Y and Z axis representations, to create the ECDF portion of the Backward + ECDF methodology. The Y axis and Z ECDF representations consisted of a mean of 9 coefficients (± 13 SD). The final LDA model created from this is therefore composed of 28 individual features, 18 of which are the two included ECDF representations. When performance is assessed using the validation data the resultant LDA model gives an AUC of 0.570. The decision threshold is set at 0.208, the point at which TP results are maximised and FP results are minimised as described previously. At this threshold the sensitivity is 0.782, the specificity is 0.355, the precision is 0.283, and the F score combination measure is 0.415.

Table 3.8: The performance statistics of the identification of the combined class of lie to stand and stand to lie transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
2 Seconds	Manual	0.580	0.420	0.710	0.321	0.364	0.283
	Manual + ECDF	0.617	0.630	0.578	0.328	0.431	0.259
	ECDF	0.643	0.761	0.457	0.314	0.444	0.241
	Forward	0.612	0.688	0.522	0.320	0.437	0.247
	Forward + ECDF	0.647	0.743	0.503	0.328	0.455	0.242
	Backward	0.665	0.605	0.684	0.385	0.470	0.264
	Backward + ECDF	0.675	0.496	0.788	0.434	0.463	0.291
	CFS	0.572	0.475	0.660	0.313	0.378	0.275
	CFS + ECDF	0.633	0.529	0.685	0.354	0.424	0.283
3 Seconds	Manual	0.562	0.471	0.637	0.297	0.365	0.265
	Manual + ECDF	0.628	0.736	0.490	0.320	0.446	0.234
	ECDF	0.639	0.623	0.633	0.357	0.454	0.272
	Forward	0.582	0.601	0.544	0.301	0.401	0.249
	Forward + ECDF	0.653	0.667	0.601	0.353	0.462	0.261
	Backward	0.657	0.540	0.746	0.409	0.466	0.268
	Backward + ECDF	0.664	0.725	0.567	0.353	0.475	0.247
	CFS	0.560	0.529	0.595	0.299	0.382	0.262
	CFS + ECDF	0.633	0.739	0.501	0.326	0.452	0.250
5 Seconds	Manual	0.624	0.442	0.751	0.367	0.401	0.284
	Manual + ECDF	0.661	0.717	0.563	0.349	0.470	0.230
	ECDF	0.658	0.583	0.665	0.363	0.447	0.266
	Forward	0.604	0.696	0.509	0.316	0.435	0.245
	Forward + ECDF	0.676	0.837	0.454	0.334	0.477	0.226
	Backward	0.685	0.714	0.599	0.368	0.485	0.241
	Backward + ECDF	0.696	0.815	0.510	0.352	0.492	0.224
	CFS	0.557	0.627	0.489	0.286	0.393	0.245
	CFS + ECDF	0.639	0.757	0.475	0.320	0.450	0.226

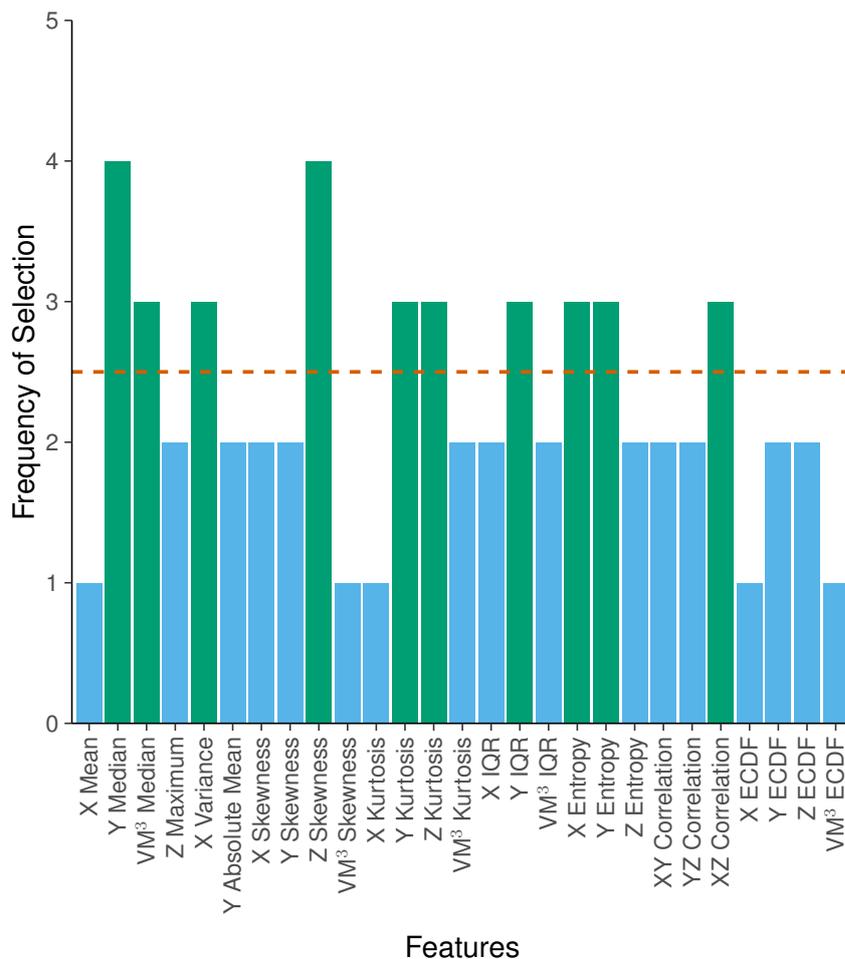


Figure 3.6: The frequency of inclusion of features within cross-validation of Lie to Stand and Stand to Lie Combined Class classification models. The 2.5 threshold is marked.

3.4.7 Sit to Stand and Stand to Sit Combined Class Identification

Randomly sampled non-transitional data were excluded, and both Sit to Stand and Stand to Sit transitions were grouped into a combined positive class ($n = 945$). The negative class consists of the remaining transitions ($n = 400$). A validation dataset comprising one sixth of the data was removed for validation ($n = 224$) resulting in a training set of 1121 cases to be further split into train and test sets through the 5 fold cross-validation. Class composition was approximately maintained within each data subset. The performance of the classification of transitions between Sit and Stand postures, regardless of the direction, is displayed in Table 3.9 and assessed to inform the construction of a validation model.

When using a 2 second window size the Backward + ECDF method of feature selection gave the best performance across the 5 folds of the cross validation. Both AUC and F score were the highest reported for the window size (0.663 and 0.730 respectively), suggesting the Backward + ECDF method exhibits the highest values of sensitivity (0.666), specificity (0.628), and precision (0.809) that are attainable while

maintaining inter-measure balance.

A 3 second window size suggests either the Backward + ECDF method or the Forward + ECDF method of feature selection. Models created using the Backward + ECDF method gave the highest value of AUC (0.667) and the second highest value of precision (0.820). The sensitivity, specificity, and F score were 0.592, 0.694 and 0.688 respectively. The Forward + ECDF method gave the highest F score (0.741), and second highest value of AUC (0.659), of methods using a 3 second window size. Both the specificity (0.571) and precision (0.793) were lower than those reported for the Backward + ECDF method. However, the sensitivity (0.695) was higher and the third highest reported for the window size. As a result of the relative ranking of combined measures the Forward + ECDF method was selected for further consideration.

Separating data into larger 5 second windows and assessing performance presents three models as potential choices for comparison. The Forward + ECDF method produces the highest F score (0.770) and sensitivity (0.753) among the alternatives assessed here. The specificity (0.523), precision (0.789), and AUC (0.655) are all low relative to values produced by alternative methods. The second method of interest, the Backward + ECDF method, gives the highest AUC (0.687) of the feature selection techniques assessed. The precision (0.812) of this method is the second highest recorded, and the specificity (0.628), while low relative to other methods, is higher than that reported by the Forward + ECDF method. This is countered by the relatively low values of sensitivity (0.680) and F score (0.740). The Backward only method is notable as a compromise between the two previously outlined methods. This is due to the combined measures, AUC (0.672) and F score (0.750), both being the second highest values reported for the window size, and values of sensitivity (0.707), specificity (0.580), and precision (0.799) which all fall between those reported by the other two methods. For these reasons it is the Backward only method that exhibits the most balanced performance across the measures of interest and will therefore be considered further.

A comparison of the methods chosen from the 3 window sizes reveals the 5 second Backward only method as having the highest performance across both of the combined measures of AUC, and F score. Figure 3.7 presents a summary of all features that were selected for inclusion in the 5 models generated by the chosen method during the 5 fold cross validation. 14 features occurred in greater than 50% of cross validation folds. When used to classify the previously excluded validation data an AUC value of 0.616 is achieved. A decision threshold, optimised so as to maximise TPs and minimise FPs, of 0.650 is used. At this threshold the sensitivity of the model is 0.834, the specificity is 0.388, the precision is 0.762, and the F score is 0.796.

Table 3.9: The performance statistics of the identification of the combined class of sit to stand and stand to sit transitions using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
2 Seconds	Manual	0.561	0.615	0.498	0.743	0.673	0.680
	Manual + ECDF	0.626	0.635	0.601	0.790	0.704	0.674
	ECDF	0.651	0.558	0.724	0.827	0.666	0.690
	Forward	0.582	0.521	0.619	0.764	0.619	0.698
	Forward + ECDF	0.660	0.632	0.640	0.806	0.708	0.669
	Backward	0.632	0.640	0.616	0.797	0.710	0.696
	Backward + ECDF	0.663	0.666	0.628	0.809	0.730	0.689
	CFS	0.544	0.628	0.480	0.741	0.680	0.683
	CFS + ECDF	0.630	0.476	0.739	0.812	0.600	0.702
3 Seconds	Manual	0.579	0.698	0.435	0.745	0.720	0.667
	Manual + ECDF	0.632	0.572	0.643	0.791	0.664	0.692
	ECDF	0.654	0.637	0.604	0.791	0.706	0.668
	Forward	0.584	0.393	0.766	0.798	0.526	0.705
	Forward + ECDF	0.659	0.695	0.571	0.793	0.741	0.662
	Backward	0.645	0.525	0.718	0.815	0.638	0.719
	Backward + ECDF	0.667	0.592	0.694	0.820	0.688	0.708
	CFS	0.589	0.704	0.453	0.753	0.728	0.673
	CFS + ECDF	0.655	0.454	0.793	0.838	0.589	0.719
5 Seconds	Manual	0.601	0.565	0.592	0.766	0.650	0.703
	Manual + ECDF	0.645	0.637	0.598	0.789	0.705	0.686
	ECDF	0.658	0.698	0.589	0.801	0.746	0.663
	Forward	0.552	0.443	0.673	0.762	0.560	0.705
	Forward + ECDF	0.655	0.753	0.523	0.789	0.770	0.654
	Backward	0.672	0.707	0.580	0.799	0.750	0.683
	Backward + ECDF	0.687	0.680	0.628	0.812	0.740	0.684
	CFS	0.586	0.334	0.823	0.817	0.474	0.722
	CFS + ECDF	0.658	0.626	0.643	0.806	0.704	0.680

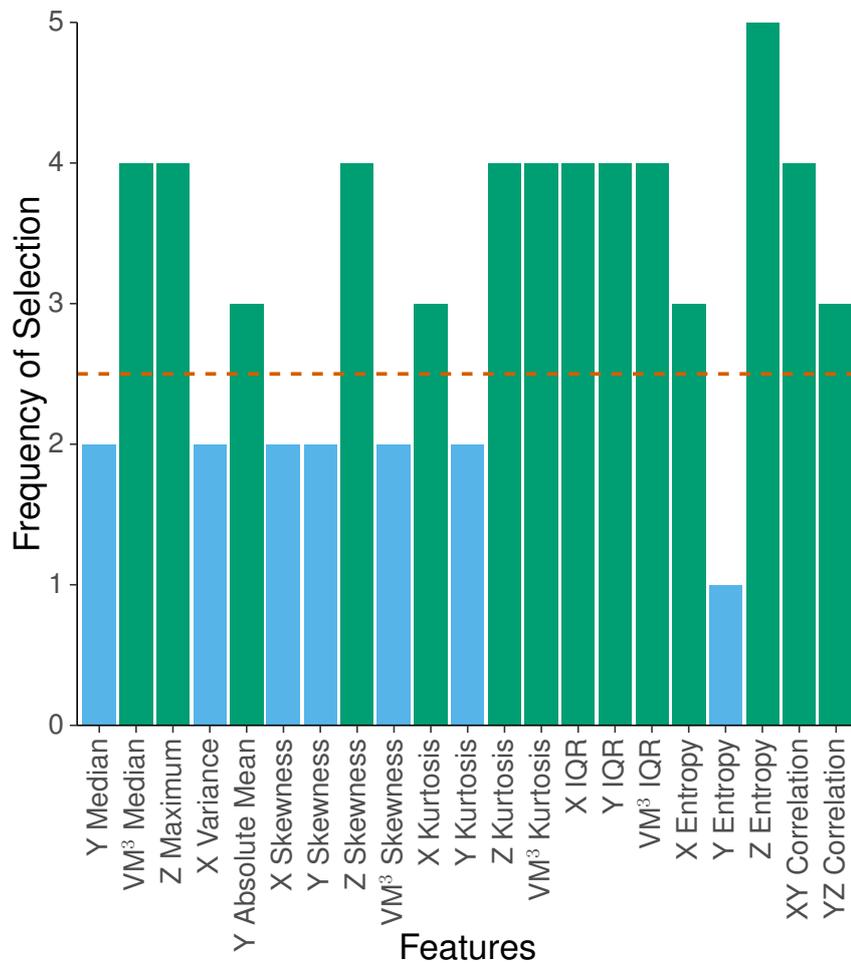


Figure 3.7: The frequency of inclusion of features within cross-validation of Sit to Stand and Stand to Sit Combined Class classification models. The 2.5 threshold is marked.

3.4.8 Directional Transition Identification

Randomly sampled non-transitional data, and transitions between sitting and standing postures were excluded. Transitions from a sit or stand posture to that of a lying posture were grouped into a combined positive class ($n = 201$). The negative class consists of the remaining transitions from lying to standing or sitting ($n = 199$). A validation dataset comprising one sixth of the data was removed for validation ($n = 66$) resulting in a training set of 334 cases to be further split into train and test sets through the 5 fold cross-validation. Class composition was approximately maintained within each data subset. The classification performance of feature selection methods attempting to discriminate between movements to and from a lying position is shown in Table 3.10.

The ECDF only and Manual + ECDF methods of feature selection each meet the selection criteria when using a 2 second window length. The Manual + ECDF method gives the highest value of F score (0.644), and the third highest of sensitivity (0.806) among the methods examined. However, the AUC (0.610), specificity (0.393), and precision (0.536) are each outperformed by alternative methods. LDA models consisting only of ECDF representations of the axes performed the best in regards to AUC (0.667) and gave the second highest value of F score (0.634). The ECDF only method gave a low sensitivity (0.658), relative to many of the other assessed methods, but the third highest values of both specificity (0.635), and precision (0.611). As a result of the relative rankings of the two combined measures the ECDF only method of model construction was chosen.

When using a 3 second window size both the Forward + ECDF and Backward + ECDF methods satisfy the selection criteria. The Forward + ECDF method gives the highest values of the AUC (0.675) and precision (0.641). The F score (0.636) of the method is the second highest recorded, but both the sensitivity (0.632) and specificity (0.691) are outperformed by other methods. The Backward + ECDF method gives the highest values of F score (0.674) and of sensitivity (0.794). The AUC (0.673) is the second highest reported for the window size. Both the specificity (0.511) and precision (0.586) using the Backward + ECDF method are lower than alternative methods. As the relative rankings of the two viable methods are equal the values of the contributing measures are assessed. The Forward + ECDF method is selected due to the more equal distribution of performance across the three metrics.

The performance of methods using a 5 second window highlights two methods which satisfy the selection criteria. The Backward stepwise feature selection method gives the second highest AUC (0.687), the highest sensitivity (0.755), the third highest value of precision (0.619), and the highest achieved F score (0.680). The specificity (0.596) did not outperform alternative methods. The addition of an ECDF representation, the Backward + ECDF method, gave the highest values of AUC (0.721), specificity

(0.725), and precision (0.671). The F score (0.658) was the second highest among the assessed methods. The sensitivity (0.645) of the Backward + ECDF technique was relatively poor in comparison to alternative methods assessed. The relative rankings of the two summary measures, AUC and F score, for the two methods were equivalent and as such the three contributing measures, sensitivity, specificity, and F score, were assessed. Therefore, the Backward + ECDF method was chosen due to the higher relative rankings among the three measures.

A comparison of the performance of the 3 selected methods provides support for the use of a 5 second window length and the selection of features based on those used by the Backward + ECDF methodology. Figure 3.8 gives a list of all features, and ECDF representations, included in the 5 feature subsets, generated with this methodology, during cross validation. 16 features and 1 ECDF representation, that of the Z axis, were selected in greater than 50% of cross-validation folds. A mean of 10 (± 5 SD) ECDF coefficients formed the ECDF representations resulting in an LDA model consisting of 26 individual features to use with the validation data. The model achieves an AUC of 0.429. At an optimised decision threshold of 0.776 the model achieves a sensitivity of 0.097, a specificity of 0.972, a precision of 0.750, and an F score of 0.171.

Table 3.10: The performance statistics of the identification of the directional grouped transition classes using LDA. Each permutation of feature selection and window size is presented with the highest performers highlighted

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
2 Seconds	Manual	0.534	0.819	0.303	0.506	0.626	0.387
	Manual + ECDF	0.610	0.806	0.393	0.536	0.644	0.400
	ECDF	0.667	0.658	0.635	0.611	0.634	0.457
	Forward	0.557	0.568	0.584	0.543	0.555	0.469
	Forward + ECDF	0.662	0.639	0.652	0.615	0.627	0.466
	Backward	0.652	0.665	0.601	0.592	0.626	0.468
	Backward + ECDF	0.657	0.600	0.730	0.660	0.628	0.486
	CFS	0.464	0.884	0.140	0.472	0.616	0.450
	CFS + ECDF	0.648	0.677	0.567	0.577	0.623	0.463
3 Seconds	Manual	0.483	0.535	0.506	0.485	0.509	0.463
	Manual + ECDF	0.550	0.645	0.478	0.518	0.575	0.451
	ECDF	0.641	0.561	0.719	0.635	0.596	0.472
	Forward	0.559	0.755	0.388	0.518	0.614	0.439
	Forward + ECDF	0.675	0.632	0.691	0.641	0.636	0.476
	Backward	0.654	0.619	0.669	0.619	0.619	0.474
	Backward + ECDF	0.673	0.794	0.511	0.586	0.674	0.425
	CFS	0.494	0.310	0.758	0.527	0.390	0.471
	CFS + ECDF	0.648	0.561	0.713	0.630	0.594	0.469
5 Seconds	Manual	0.598	0.568	0.612	0.561	0.564	0.475
	Manual + ECDF	0.632	0.581	0.646	0.588	0.584	0.480
	ECDF	0.622	0.613	0.596	0.569	0.590	0.465
	Forward	0.586	0.626	0.573	0.561	0.591	0.458
	Forward + ECDF	0.669	0.645	0.640	0.610	0.627	0.455
	Backward	0.687	0.755	0.596	0.619	0.680	0.439
	Backward + ECDF	0.721	0.645	0.725	0.671	0.658	0.478
	CFS	0.531	0.748	0.371	0.509	0.606	0.464
	CFS + ECDF	0.678	0.671	0.657	0.630	0.650	0.456

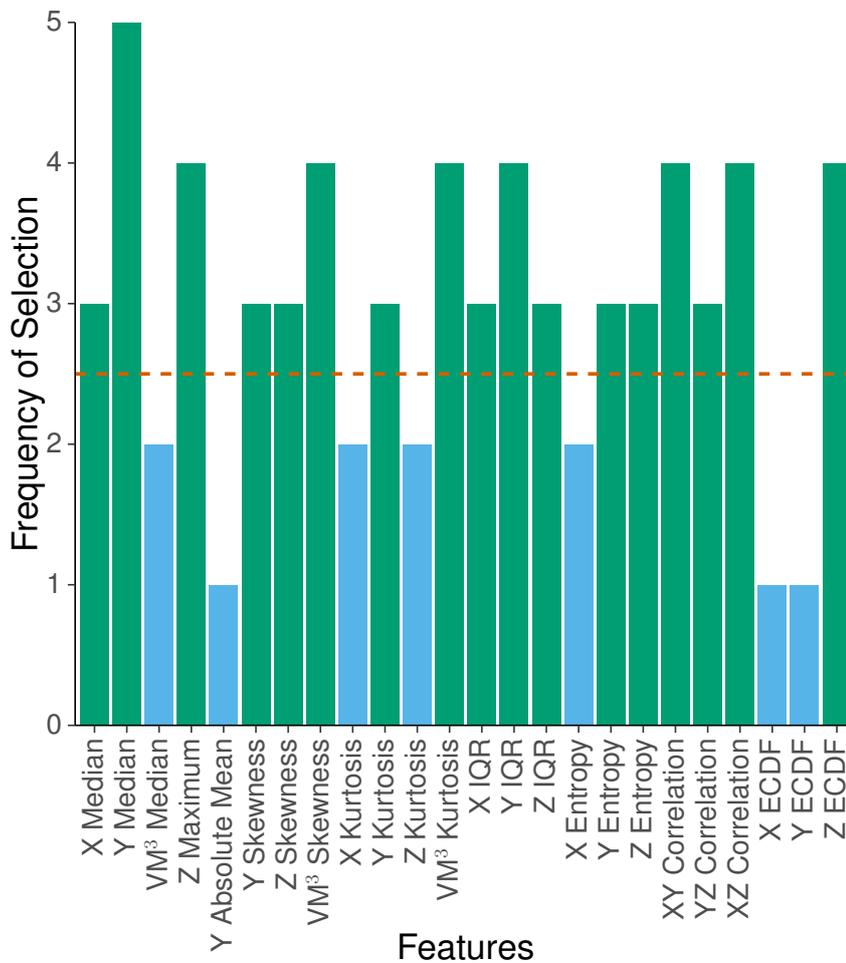


Figure 3.8: The frequency of inclusion of features within the cross-validation of the classification of directional combined transitions. The 2.5 threshold is marked.

3.5 Discussion

The detection of postural transition events is challenging when attempted on a homogenous population of healthy dogs. The potential for such event-based methods to prolong the operational lifespan of deployed devices, by strategically reducing the data captured and stored on a device, or in using transitions to postures of interest as triggers for more intensive capture or analyses, makes such a finding disappointing.

The AUC values provided by the 8 validation models presents an unexceptional picture of performance with values ranging from 0.429 to 0.660. As previously stated, the AUC is a representation of the sensitivity/specificity dichotomy and its effect on a models ability to correctly balance these aspects when discriminating between groups. Review of the F scores of the validation models gives a more variable picture of performance, ranging from a low of 0.171 to a high of 0.796. Some of this variation can be accounted for in the relative difficulty, or even the uncovered impossibility, of the classification tasks themselves. These classification tasks can be grouped into four groups, postural transition detection, identification of a single transition type, identifica-

tion of a group of inverse transitions, and detection of direction of travel. Each of these groupings serve to outline the nature of the tasks and aid in the broad formation of relevant conclusions. It should also be noted that such delineation of tasks is useful in understanding the potential practical uses of such classifiers in the formation of hierarchical classification methods (Chakravarty *et al.*, 2019; Mathie *et al.*, 2004; Zhang *et al.*, 2010). Additionally, combined measures, although useful for rapid assessment of performance, particularly when using more than one measure, are insufficient for the further assessment of models. The full examination of where model performance suffers requires discussion and interpretation of the three constituent measures of performance, sensitivity, specificity, and precision. This also ensures that high values of any one value are not masking the performance of another within the combined measures.

The first classification task, and the only one to include the randomly selected non-transitional data, aimed to identify the occurrence of a transition event regardless of the postures that are moved between or the direction of travel. The intent was to identify if there was, within the acceleration of the collar-mounted device, some common markers or features of transitional movements. The performance of the differing methods of feature selection and ECDF inclusion give promising values across the different metrics of performance but acceptable or exceptional values are often coupled with low values of opposed measures. The use of a 3 second window size for data and the Backward + ECDF method of feature selection that was selected produced a validation model that ultimately underperformed when applied to the validation data previously withheld. This suggests either a degree of over-fitting had occurred during the previous performance assessment or that the method of feature selection used, the simple voting method based on the feature subsets generated across the cross-validation, had not been sufficient. As a result it is difficult to justify the use of such a classifier for further use. The features calculated are unable to provide a distinct definition of a posture transition behaviour, which could be due to the breadth of movements this encapsulates, the variety possible within the non-transitional data, and the limited nature of the features themselves in accounting for these aspects. Class frequencies were artificially equalised, through random subsampling of the non-transitional data, and the poor performance that occurs despite the focal class being artificially amplified suggests a distinct difficulty in the data-driven definition of a "postural transition event".

These shortcomings of the methodology, and the resulting models, continue to occur when examining the second through to, and including, the fifth classification task, a group of tasks each attempting to characterise the distinctions between a focal transition and all other transitions. These models were constructed with an ideal dataset, unaffected by the shortcomings of the first classifications task, as would occur if running within a hierarchical methodology, that was composed entirely of postural transitions

and did not include any non-transitional data. These tasks each focussed on a much more limited movement that occurred in a singular direction and between set postural states. However, despite this the performance of the models produced for validation displayed little improvement. Due to the more focussed nature of the task it is difficult to justify poor performance as a result of a high level of variance in the focal movements. The similarity of the movements, particularly those that are inverted directionally from each other, may also introduce difficulties in the delineation of transitions. The shortcomings of the features are another likely explanation of this. The poor performance given by the validation model of each task, despite being trained on a dataset consisting only of postural transitions, further highlights the lack of suitability of these classifiers for further applications, where such data would not be as "clean" and the accrued error could be severe.

The Sit to Stand, Stand to Lie and Stand to Sit tasks each provide similar levels of sensitivity, with the number of identified true positive results ranging from 62.1% to 66.1% of the total positive classes, but suffer in regards to other metrics. The specificity of the Sit to Stand model, the best performing of the three when assessed by sensitivity, suggests that only 48.1% of non-"Sit to Stand" transitions are identified correctly. With greater than 50% of the majority, negative class being falsely identified as positive the precision suffers with only 32.8% of those identified as positive being true positives.

The validation model created for the inverse transition, Stand to Sit, was the highest overall performer when considering all three metrics simultaneously, despite only giving the highest value for precision. The performance of the model in terms of precision revealed a 50% split between true and false positives, again suggesting the model would be difficult to use in a real setting. The relationship between precision and prevalence, namely that with increased prevalence but consistent values of sensitivity and specificity the precision will also increase, it is likely that the perceived increase in precision is therefore partly due to the higher frequency of occurrence of the Stand to Sit transition (Table 3.2).

The Lie to Stand transition is proven the most problematic to classify, with the validation model succeeding in correctly identifying 100% of the focal transition but at the cost of mislabelling 93.3% of the negative class transitions which leads to just 14.7% of positive classified cases being true positives. This outcome illustrates the sensitivity/specificity and sensitivity/precision dichotomies under which the models operate as well as the issues in generalisability and overfitting that have occurred when applying models constructed from the feature selection methodologies to validation data.

The third classification group considered the possibility of similarity between inverse movements resulting in high confusion, and, if so, whether combining inverse transitions into a single group would result in improvements to the classification ability. Both models, the first attempting to find transitions between lie and stand posture

regardless of direction, and the second attempting to extract transitions between sit and stand postures regardless of direction, showed high levels of sensitivity, identifying 78.2% and 83.4% of true positives respectively. However, both also showed poor specificity with 64.5% and 61.2% of negative classes misclassified. A more distinct difference between the two models is presented by the values of precision given. The lie to/from stand transition model provides a very low value of precision, with 28.3% of windows identified as positive being correct, whereas the sit to/from stand model gives a higher value, with 76.2% correctly identified. This finding would make the sit/stand model a promising choice for further examination. However, the sit/stand transitional classes were the most frequently occurring within the dataset (Table 3.2) and it is likely that this result of precision is more reliant on the prevalence of the class rather than on a real improvement to model performance. This is well illustrated by the similarity of the other measures of the two models within this classification group.

The final classification task attempts to identify the potential of a directional grouping method to narrow the selection of possible transitions. It is clear from the values of both sensitivity and specificity that the validation model derived from the prior cross-validation suffered from a distinct decrease in performance. This suggests a lack of generalisability of the model to data that has not been used for training and may be compounded further by the more limited sample size of this task. This was due to the exclusion of the most frequently observed transitions, those between a sit and stand, due to the lack of upward or downward directionality.

Almost every assessed classification task presented the Backward + ECDF model as the highest performing method of feature selection. The consistency of this result lends further weight to the argument that the methods used here are not sufficient for the task. The backward stepwise selection, with feature exclusion assessed using the relative increase to AUC observed, is a greedy method of feature selection that typically results in large feature sets with a high level of redundancy (Derksen & Keselman, 1992; Mao, 2004). This suggests a difficulty in identifying features which provide a clear distinction between the classes assessed across each classification task and is shown by the frequent instability of selected feature sets, with many selected occurring in fewer than half of cross-validation folds, implying an interchangeability between features despite prior exclusion of those which are highly correlated. The difficulty of calculated models to provide a clear class distinction based on the supplied features is further supported by the consistent inclusion of the ECDF representation across 7 of the 8 assessed models. This was added as an alternative data representation which would highlight the insufficiency in the ability of the time-domain features to characterise the differences between classes (Hammerla *et al.*, 2013). The underperformance of the Backward + ECDF method despite its inclusion of a more granular representation of the windowed data suggests that this may not be a case of incorrectly selected time-

domain features, where methods including or exclusively using the ECDF would be the highest performers. It is perhaps more likely that the location of the sensor and the nature of the acceleration experienced by the sensor are obscured. This could, for the first assessed classification task, be due to the wide range of motion and behaviours that closely resemble the acceleration profiles of postural transitions, or, for the other tasks which exclude non-transitional data, the differences between transitions being difficult to define and discern when represented and summarised in the time domain. Further separation of classes could be achieved with dimensional reduction methods, such as principal component analysis (PCA), which would aim to consolidate features into those few that best characterise the variation within the dataset. However, in this case the improvement would likely be marginal and the cost of abstraction, the loss of the ability to speak in more real terms as to what features are of interest within a model, makes further exploration of such adjustments a less appealing prospect. An alternative classification method would be required due to the potential conflict of the PCA methodology with an LDA classifier (Sunderam *et al.*, 2007). Additionally, the methods used for model construction and assessment place a high level of importance on sensitivity as a performance indicator. Both F score and AUC are calculated using the sensitivity and this is apparent when viewing the results. Stepwise elimination and selection methods were also optimised using only the AUC measure. The metric was chosen due to its assessment of performance across all potential decision thresholds but its use emphasises the importance of sensitivity and specificity within model performance and excludes the consideration of alternative measures, such as precision.

The most frequently selected window size was 5 seconds. This being the maximum size assessed, and exceeding all recorded transitional durations, could suggest that the postural states occurring either side of the actual transition event are more informative, contribute more valuable information to the features, than the transition behaviour itself. However, the differences in performance between the highest performing methods of each window were often slight. This possibility will be examined in the next section in more detail. The pre-processing methodology makes an assumption about the mechanics of collar rotation that will never be truly satisfied in practice and may have led to the obfuscation of transitional acceleration. The (0,0,1) target orientation would not be obtained unless the neck of the dog is directly parallel to the ground which is highly unlikely, except when the dog is in a sternal, recumbent position. The use of a vectorial minimum in searching for the most appropriate rotation matrix relaxes this assumption as rotation around the Y axis, such as when the dog raises or lowers its head, would not effect the matrix selection as all "correct" matrices would continue to be the minimum vectorial distance from the target. In addition the use of rotation matrices preserves the total energy across all three axes, but the correction methodology used here serves to maximise this energy in the Z, or dorsoventral, axis and minimise it within the X, or

medial-lateral, axis. This results in an increased likelihood of the over and under estimation of single axis acceleration. Further work should attempt to validate the use of the correction methodology and investigate the comparative effects of its application within behaviour classification. Alternatively the inclusion of additional sensors, such as a gyroscope to directly measure orientation, may be preferable.

In future work it may be possible to refine models that focus on the event-based identification of postural transitions in domestic dogs, based on a single collar mounted sensor. The potential to correctly identify transitions and to then inspect the duration, or perceived normality of the movement could be highly informative but the limitations of an event sampling focussed approach are clear. The sample size of $n = 1345$ proved to be low for the validation requirements of the current study and would require extensive and intensive manual labelling to expand further, which would be highly expensive or impractical. The event based nature of the transitional behaviours also makes the implementation of alternative measures difficult. The current feature set attempts to capture acceleration in each axis, the relationships between the axes, the nature of the acceleration through and over time, and the energy of the movements itself across the three axes.

A further possible symptom of the limited sample sizes is the observed tendency of models to appear to over-fit to the supplied training data making the proposed uses in veterinary contexts problematic due to the likelihood of abnormality in the transitional behaviour, due to injury, illness, age, or conformation, leading to misclassification and even possible misdiagnosis where transition abnormalities indicative of illness occur only after or during certain periods (i.e. after intense activity) and only normal transitions are captured. The analysis of postural transitions themselves, particularly the sit to stand and stand to sit transitions, is a common practice in the assessment of human physical and neurological healthcare with a wealth of research on the importance and use of these assessments and in the use of accelerometer and gyroscope instrumentation in automating the approach or in extending it to include transitions occurring in free-living environments (Millor *et al.*, 2014; Pickford *et al.*, 2019; Van Lummel *et al.*, 2013). The usefulness of such a tool within veterinary practice, whether focussing on sit to stand or lie to stand, has been previously explored but focusses on controlled assessment or requires instrumentation that is impractical for longitudinal, free-living data capture (Brugarolas *et al.*, 2013; Ellis *et al.*, 2018; Thompson *et al.*, 2016). A common feature of these solutions, both for human and animal transition detection and assessment, is the inclusion of orientation focussed features, indicating a shift in device position that relates to the position of the attached body part. For the detection of human postural transitions in free-living environments this requires attachment to the thigh or waist (Janssen *et al.*, 2005; van Lummel *et al.*, 2018). In quadrupedal species the attachment location for such devices is often the trunk of the animal and

frequently involves multiple sensors (Brugarolas *et al.*, 2013; Thompson *et al.*, 2016). A single sensor equivalent to the method used in the detection of human sit to stand transitions would be the attachment of the device to a forelimb. This would allow a clear orientation based delineation between lie and stand transitions but would perhaps fail to collect information relating to sit to stand transitions or from non-focal limbs (Robert *et al.*, 2009). The use of a collar mounted sensor prevents the use of such measures due to the influence of the head orientation. However, the practicality of such an attachment method for long term data collection, the potential perception of alternative methods, such as glue, or obstructive limb or torso based attachment strategies, among dog owners and the low, or even non-existent, habituation period, due to the prevalence of collar use in domestic dog populations, presents a strong case for the assessment of a collar based solution.

3.6 Conclusion

The poor performance of the classifiers in terms of the summary measures, AUC and F score, across various window size and feature selection method combinations clearly represents the difficulty inherent in the prediction of behaviours within the constraints presented by the procedures employed here. The use of these methods to interpret the postural transition frequency and to classify the type of transition from a novel dataset would be unreliable and insufficient. This is compounded by low values for, or lack of appropriate balance between, sensitivity, specificity and precision. Ultimately the models generated here would be unfit for purpose, particularly as all classification tasks were assessed at optimal thresholds and as such the performance metrics given were optimistic.

The collection of a more expansive sample of transitional behaviours would address the tendency of the model to over-fit. However, transitional events are relatively rare due to their non-rhythmic nature and occurrence only between postures, making such an expansion difficult and time consuming. Additionally, transitional events are short, lasting a second or less in most cases, and could further compound the difficulty of the task. An alternative methodology would be to imply their existence from the identification of the preceding and subsequent postural states of the dog, and such an approach is supported by the improved performance of the lengthened data windowing shown here. This approach should be investigated further as the longitudinal postural transition monitoring of dogs could provide a valuable comparison point for the assessment of health and welfare. Particularly as similar methodologies have shown continued success within human medicine.

Chapter 4

The Automatic Categorisation of Postural States and Locomotion of the Domestic Dog

4.1 Abstract

The remote identification of postural states and behaviour has been embraced within the fields of behavioural ecology, and human medicine to provide measures of animal health, welfare and activity. The current postural state of an individual informs the behaviours which are likely to be performed and the duration and frequency of such behaviours can further indicate the ability of an individual to behave normally. The development of an automated and longitudinal method of behavioural logging for use with domestic dogs could assist in supplementing the abilities of both owners and veterinary practitioners to notice irregularities, and to diagnose and treat conditions earlier. This chapter explores the application of established posture and locomotion recognition techniques, from both human and animal focussed fields, to a sample of 20 kennel-housed Labrador Retrievers with approximately 10 hours of annotated video data and paired accelerometer data. The established technique attempts to identify between three annotated postural states (Sitting, Standing, or Lying) using features derived from a single, tri-axial, collar-worn, accelerometer. Periods where a dogs posture was identified as Standing were further analysed to investigate whether an individual was Stationary or Locomoting. Preprocessing methods, to correct for rotation of the collar and to clean potential noise from data, were also assessed for their performance impact. After feature selection Linear Discriminant Analysis classification models were applied and performance was assessed. One vs all binary classification models struggled to perform satisfactorily except in the identification of the Standing posture. Further assessment showed notable performance gains when classification tasks were simplified into one vs one binary tasks. This therefore suggests the arrangement of classifications into a hierarchy could lead to optimal performance however further testing would be required to test such a system structure beginning to end. Preprocessing methods appeared to almost universally negatively impact performance. Where such impacts were minimal further validation should be performed to investigate this trade off in performance where the issues they are attempting to correct occur. It may be

that as minimal collar rotation was observed within this sample the impact of correcting the data was greater than the impact of rotation. Equally as all 20 dogs were of a similar size and breed the further noise reduction to account for breed size discrepancies and step frequency differences did not exist and valuable information content was sacrificed. This exploration has provided satisfactory results, has informed how best to structure future classification tasks and has highlighted potential issues with preprocessing methods. The use of more sophisticated classification methods, among other refinements of the features and methodologies, should present further improved performance.

4.2 Introduction

Posture detection has been central to the remote sensing of animal behaviour for decades (Shepard *et al.*, 2008b). As stated in Chapter 3, the current posture of an individual informs the behaviours that it is possible for it to engage in (Grundy *et al.*, 2009). At a fundamental level the identification of distinct postural states, and the amount of time spent within each, provides additional context to which behaviours an individual may be engaging in at different times, and at different levels of intensity. For example, extended periods of lying, particularly during the night-time and with low intensity, would be indicative of resting behaviour (Ladha & Hoffman, 2018; Preston *et al.*, 2012; Sunderam *et al.*, 2007). Through such inferences remote, longitudinal observations of behaviour can be constructed to provide insight into an individual's daily schedule that would be impractical or time consuming to obtain using more traditional methods (Wilson *et al.*, 2008). The longitudinal monitoring of wild, often difficult to observe, animals has been a boon to the research, monitoring and assessment of the biomechanical and behavioural aspects of species, their life histories, and their interactions with conspecifics, other species, and environment (Fehlmann *et al.*, 2017; Shepard *et al.*, 2008b).

Such methods have also been adapted for use within more controlled environments with captive animals, and livestock (Martiskainen *et al.*, 2009; Robert *et al.*, 2009; Soltis *et al.*, 2012; Thompson *et al.*, 2016). Within these contexts implementations of, and research into, remote sensing technologies tend towards behavioural monitoring. Using such techniques can help to augment traditional methods of observations. Within most applications, in industrial or captive environments, there is a focus on the improvement of the efficacy and efficiency of animal management and the maintenance of welfare (Martiskainen *et al.*, 2009; Rushen *et al.*, 2012; Soltis *et al.*, 2012).

The health of an individual is fundamental to its welfare, and the freedom from disease constitutes one of the five freedoms and as such, is central to the maintenance of good animal management practices (Mellor, 2016). Therefore, measurements and

observations of behaviour taken using remote sensing technologies are often used to infer the health of individuals. One of the more common implementations of such techniques is in the gait analysis of livestock species, where lameness is a common issue and can be difficult to diagnose until severe. The approaches to such methods vary but typically revolve around the identification of behavioural states and the direct measurement of movements or intensities during behaviours of interest (Cornou *et al.*, 2011; Martiskainen *et al.*, 2009; Pastell *et al.*, 2009). Similar approaches are also used in human medical research for the assessment of both physical and cognitive disease, as well as in the general, commercial monitoring of human activity and fitness (Anan *et al.*, 2015; Pickford *et al.*, 2019; Rodríguez-Martín *et al.*, 2015; Zhang *et al.*, 2010).

As stated in Chapter 2 accelerometer data is composed of two discrete components. The first of these, the static component, relates to the pull of Earth's gravity on the device, and is therefore a distribution of the 1g pull of the planet across the 3 accelerometer axes, which in turn can be interpreted to provide device orientation (Grundy *et al.*, 2009; Shepard *et al.*, 2008b). Within both human and animal remote sensing the static component of acceleration has been combined with biomechanical knowledge and the consistent placement of the device to derive associated postures.

The dynamic component of acceleration represents all other, non-gravitational, sources of acceleration and is calculated as described in Section 2.3. The interpretation of dynamic acceleration, its intensity and rhythmicity, provides information as to the motion of the device that, when paired with a known posture, provide detail as to how an individual is moving and, potentially, what behaviours it is engaging in.

When analyses of these two components are combined a detailed image of an individual's movement can be created (Grundy *et al.*, 2009; Shepard *et al.*, 2008b). The multi-level nature of the data processing techniques allows classification problems to be arranged in hierarchical systems, well suited to automated and machine-learning classification techniques, to aid in accelerometer-based inferences (Chakravarty *et al.*, 2019; Zhang *et al.*, 2010). By identifying the differing utility of static and dynamic acceleration data the analyses can be tailored towards specific observational goals and problems (Shepard *et al.*, 2008b). Behaviours of interest can then be considered in terms static and dynamic components with the static component representing postural state and the dynamic component representing intensity, often provided simplistically as activity counts or combinations of dynamic acceleration within each, or across all, of the three axes (Wang *et al.*, 2005; Yang & Hsu, 2010). As behaviour occurs over time there also exists a periodic, or rhythmic, component (Chakravarty *et al.*, 2019; Watanabe *et al.*, 2005). Here behaviours which possess a consistent repetitive nature, such as steps when walking, are exposed through the decomposition of acceleration waveforms to the frequency components which establish it.

Companion animals, domestic dogs in particular, have seen some products de-

veloped that promise similar monitoring capabilities to devices used by their human owners (Väättäjä *et al.*, 2018; Zamansky *et al.*, 2019). Additionally, remote sensing devices have been developed with the intent of monitoring health, and welfare using methods applied within non-companion animal populations (Ladha *et al.*, 2018; Ladha *et al.*, 2017; Rhodin *et al.*, 2017). Here I have outlined a methodology which draws on a multitude of measures established across both animal and human remote sensing research, in both the time and frequency domain and using both static and dynamic representations of activity to construct a model which could provide an automated representation of the canine daily schedule. Due to the exploratory nature of the investigation the scope was limited to cover postural states, and, from there, the detection of locomotion within a standing state. These two tasks were selected as logical first steps in behavioural mapping through remote sensing as both could provide a wealth of information regarding an individual's daily activity. As in Chapter 3 LDA models are once again used to quickly and efficiently explore and interpret the practicality of such a task using a collar mounted attachment method. The effect of data pre-processing methodologies, to correct for collar rotation and to remove high frequency noise, was also assessed.

4.3 Materials and Methods

4.3.1 Data Collection and Annotation

The accelerometer data and postural transition annotations generated from the kennel-housed, healthy, Labrador sample described in Section 3.3 were used here. The postural state of the dogs was inferred from the transitional annotations previously collected. The final posture of a transition was compared to the beginning posture of the next annotation, and, if matching, the intervening time period was labelled as the coinciding posture. Where subsequent transitional annotations do not follow from each other (e.g. a Sit to Stand transition followed by a Lie to Sit), implying a non-annotated transition had occurred, no postural annotation is recorded. As such there were three possible posture annotations, Stand, Sit, or Lie. An additional series of annotations were collected, across the same sample video samples using ELAN (Max Planck Institute for Psycho-linguistics, 2019), referring to any occurrence of locomotion, all periods not labelled as locomotion were labelled as stationary. Any stationary periods that overlapped with periods not given a postural label were removed due to uncertainty inherent in the inferential method of annotation. Additionally, any stationary labels which occurred during periods where posture was not labelled as "Standing" were removed as locomotion would be unlikely to have occurred and these periods would be outliers. Both postural and locomotion annotations were divided into sliding 5 second windows

with a 2.5 second overlap, a windowing that has been successful in previous works (Ladha *et al.*, 2017). A single window length was selected for both to ensure parity between the two classification tasks and a 5 second length was observed to be sufficient for multiple stride cycles at all observed gaits. Annotations that were less than 5 seconds in length were excluded for the training sample. The resulting sample consisted of 7152 sitting periods, 8476 lying periods and 28948 standing periods. During standing periods there were 14627 windows labelled as stationary and 4362 labelled as locomotion.

4.3.2 Accelerometer Data Processing

Accelerometer data were re-sampled to 100Hz, to account for the slight drift in frequency that occurs with extended device deployments. The pre-processing of the data is extensive and as such multiple permutations of the detection models were constructed with different levels of pre-processing applied. This allows for the potential identification of accrued error that could be introduced by the pre-processing steps.

The first pre-processing stage aimed to correct for rotation of the collar during wear and was processed identically to the methods described in Section 3.3. Once corrected, data were filtered to produce both the static, the acceleration due to gravity, and dynamic, corresponding to changes in velocity due to movement, components of acceleration (Shepard *et al.*, 2008a; Shepard *et al.*, 2008b). To reduce the degree of data summarisation prior to feature calculation a low-pass filter method was used (Sato *et al.*, 2003; Shepard *et al.*, 2008a). Permutations of the classification tasks were performed where either one filter or two filters were applied to the data. The first was required in all cases to derive the static component of the data and to allow the calculation of the dynamic component. The second, optional, filter was applied to attempt to further "clean" the remaining dynamic component data.

The stop frequencies of the two filters were derived through the application of biomechanical formulae, derived by Heglund and Taylor (1988), allowing the estimation of stride frequency from the body weight of a quadruped at the preferred speeds for the trot (Formula 4.1) and the gallop (Formula 4.2) gaits. The largest and smallest dogs of a heterogeneous sample ($n = 85$) were used to create the upper and lower estimates of canine stride frequency.

$$TrottFreq(s^{-1}) = 3.35W^{-0.130} \quad (4.1)$$

$$GallopFreq(s^{-1}) = 4.44W^{-0.156} \quad (4.2)$$

To approximate the static component of acceleration, a low-pass filter with a cut-off frequency equal or lower than the stride frequency should be used to properly exclude

minor, rhythmic variances in orientation caused by the different phases of locomotion (Shepard *et al.*, 2008a). As stride frequency is negatively correlated with size the largest dog, a 64Kg Irish Setter, was used to estimate the lower bound of canine stride frequency. The stride frequency of this dog, at their preferred speed for the trot gait, was calculated as 1.95Hz, using Formula 4.1. As there are slower gaits possible, for which there are no equivalent formulae, and to account for larger or slower dogs, due to age or illness, this value was halved and rounded to the nearest whole number. Therefore, the static component was obtained through the application of a 1Hz fourth order, zero phase, low-pass Butterworth filter to the rotation-corrected data.

The dynamic component was recovered by subtracting the static component from the signal. This dynamic component was then further filtered to reduce the influence of noise or non-focal movements. To achieve this the stride frequency of the gallop of the smallest dog of the sample, a 7.6Kg Cavalier King Charles Spaniel/Poodle cross, was calculated. The calculated value of 3.26Hz, using Formula 4.2, provides an approximate upper range of the frequency of the possible rhythmic movements within canine locomotion. This value was rounded up to the nearest 0.5Hz increment to account for smaller dogs, or for dogs with conformations or aerobic ability leading to an increase in the potential preferred stride frequency. A 3.5Hz fourth order, zero phase, low-pass Butterworth filter was applied to remove noise and rapid, non-focal behaviours from the dynamic component of acceleration.

To explore the effects of preprocessing processes each classification task was attempted with all possible combinations or exclusions of the rotation correction and second dynamic filter stages applied.

4.3.3 Feature Extraction

Features for both postural and locomotion focused classification tasks were calculated for each annotated 5 second rolling window. The VM³ combination measures of both the static, postural, and dynamic, locomotion-focused, components were calculated as previously described in Formula 2.1, in Section 3.3. The VM³ of the dynamic component is equivalent to the VeDBA measure used frequently in behavioural ecology and assessed in Chapter 2 (Qasem *et al.*, 2012). Of the 13 time domain features calculated in Section 3.3 all but the measures of window skewness and kurtosis were calculated again for the static and dynamic components. The ECDF representation was also not calculated here for either classification task. These exclusions are due to the less transient nature of posture and locomotion resulting in a lesser focus on the shape and relative variance of the acceleration over time.

Locomotion is a rhythmic behaviour, as strides repeat over time following set patterns, or gaits, at a variety of frequencies. Transformation of the accelerometer signal

into the frequency domain allows the extraction of features which have been shown to allow classifiers to more effectively distinguish between the rhythmicity of animal and human gaits, and other rhythmic behaviours, and the lack of regular rhythm in behaviours occurring while stationary (Barrey *et al.*, 1994; McClune *et al.*, 2014; Pillard *et al.*, 2012; Soltis *et al.*, 2012; Watanabe *et al.*, 2005).

The frequency domain representation of the dynamic acceleration is obtained by applying a Fast Fourier Transform (FFT) to each of the three axes and the VM³ representation of the accelerometer data of each window (McClune *et al.*, 2014; Watanabe *et al.*, 2005). Data are normalised prior to the application of the FFT algorithm and, to reduce the occurrence of artefacts in the frequency spectra, a Hanning Window is applied to each 5 second window, generated using the Signal R package (Developers, 2014). The application of a windowing function is necessary as the FFT algorithm assumes that the window assessed will be repeated perfectly, infinitely and, if this is not the case it results in the occurrence of a step function where the window begins and ends at differing levels. The FFT algorithm is applied using the function available as part of R (R Core Team, 2018) and the function used to calculate the frequency spectra of windows and the features calculated from these spectra is included in Appendix E. The resulting FFT spectra consists of a series of frequency bins displaying the magnitude of signal components occurring at the relevant frequencies. The resolution of the frequency bins is dictated both by sampling rate and window size. Here each bin occurs at 0.2Hz intervals up to the Nyquist frequency of 50Hz, the highest frequency that can be observed at a 100Hz sampling rate.

5 features are calculated from the frequency spectra of the 3 axes and VM³ representation, resulting in a total of 20 frequency spectra features. Features are calculated using functions included in R 3.6.0 (R Core Team, 2018) except where specified. The magnitude of the signal relating to the 0Hz frequency is excluded from all feature calculations as this is the DC component of the signal and would have previously been calculated as it is the mean of the time domain data. The calculated features were chosen to represent the overall shape and distribution of the spectrum. The first of these is the mean magnitude which was calculated as a representation of the central tendency of the entire spectrum. The Shannon's entropy of the frequency domain was also calculated and has been used previously in human and animal literature (Bao & Intille, 2004; Benaissa *et al.*, 2017; Marais *et al.*, 2014; Wang *et al.*, 2005). The maximum power (calculated as the square of the absolute signal magnitude for a frequency) and the coinciding frequency are calculated to provide an indication of the dominant frequency of the spectrum. The use of dominant frequency and power is similar to methods used by Watanabe *et al.* (2005) to characterise actions. However, unlike Watanabe *et al.* (2005) the Power Spectral Density (PSD) used here is an estimation that has not been corrected for variance. This is typically addressed using Welch's method of spectral

density estimation involves the averaging of multiple FFT spectra, relating to the same action, but this would drastically reduce the sample size available for training models (Welch, 1967). Energy per sample is calculated as the sum of estimated PSD values divided by the total number of samples per window (500) and provides a summary of the overall energy of the signal (Bao & Intille, 2004; Benaissa *et al.*, 2017; Ravi *et al.*, 2005; Wang *et al.*, 2005).

Another variation to the pre-processing performed is the application of scaling to calculated features. Alternative feature sets were calculated that had been scaled and the absolute values taken to attempt to account for the variance in scale presented across the variety of features. This treatment was not applied to values, such as the absolute mean value of an axis, that had already been scaled or the absolute had already been taken, or features, such as those relation to correlation, where such treatment would not make sense.

4.3.4 Model Implementation and Feature Selection

Features were assessed for collinearity and selected using the same methods described in 3.3. The LDA classifiers were also constructed and cross-validated in the same manner as previously described. A validation data set was excluded from the entire process for the testing of the final model based on the feature-voting methodology used previously. All performance assessments were, as in described in Chapter 3, quantified through the use of AUC, F-score, sensitivity, specificity, and precision.

The postural classification was divided into three binary, one vs all, classification tasks, with one classifier for each of the three annotated postural states, Lie, Sit, and Stand. Alternatively the postural classification tasks were also structured as a set of three binary tasks accounting for each pairwise combination of postures. The discrimination of locomotion from stationary periods was, as previously mentioned, only performed on periods where the dogs were annotated as being in a standing posture to reduce the incidence of unnecessary classification and to reduce the level of class imbalance exhibited in the data. This arrangement of the classification tasks into a series of binary models, similar to the series of binary classifications of transitional behaviours in Section 3, allows for the construction of a hierarchical classification method to maximise the overall performance.

The manually selected feature sets were constructed to maximise the relevant information available to classifiers. The postural classification model, used for all three relevant classification tasks, prioritises device orientation, recorded in the static acceleration component, and the inter-relatedness of axes. The generated model closely resembles that of the manual model in Section 3 except in its exclusion of entropy, skewness and kurtosis. The latter two measures were not calculated for the postural

data set and entropy was excluded due to the theoretical irrelevance of the measure due to the lack of stochasticity within the gravitational component of the data. The manual model for the locomotion classification task features the same time domain features as the postural model but also includes 5 frequency domain features calculated for the 3 axes. The frequency domain features of the VM³ were not included as it is a summary representation of the three individual axes already included. Validation models were constructed based on the best performing permutation per classification task and used feature subsets derived from the prominence of features, across the 5-fold cross-validation, within those models, as described in Chapter 3.

4.4 Results

4.4.1 One VS All Postural Identification

The manual model was consistent across all one vs all posture identification tasks and consisted of the mean, minimum, maximum, and standard deviation of the Y and Z axes. The correlation of the two axes was also included as a feature.

4.4.1.1 Standing Detection

One sixth of data relating to postural states was removed for validation (N = 7430) resulting in a dataset of 37146 5 second windows, with 24123 labelled as Standing, to be used in a 5 fold cross-validation. The performance for the detection of periods where the individual is adopting a standing posture from among all recorded 5 second windows is shown in Table 4.1 with results relating to all combinations of preprocessing and feature selection methodologies.

The overall best performing methodology was a backward stepwise feature selection with data that had not been corrected for rotation or had additional filtering of dynamic data applied. This produced the highest value for the AUC (0.787) when compared across all other permutations and gave an F score (0.784) second only to the backward selected feature set for the rotation corrected, no additional filter dataset. This also suggests that despite alternative combinations of feature selection and preprocessing methods outperforming in terms of specificity (0.741), sensitivity (0.724), or precision (0.832) values the three measures are more consistent and therefore produce preferable performance summary measures.

Figure 4.1 presents a summary of the features selected for inclusion within the validation model. 28 features occurred in greater than 50% of cross validation folds and were therefore included within the validation model. When the model is used to classify standing within the previously withheld validation data an AUC of 0.79 is given and an optimal decision threshold of 0.57 is selected. Using this threshold the

validation model produces an F Score of 0.78 and values of 0.74, 0.72, and 0.83 for Sensitivity, Specificity, and Precision respectively.

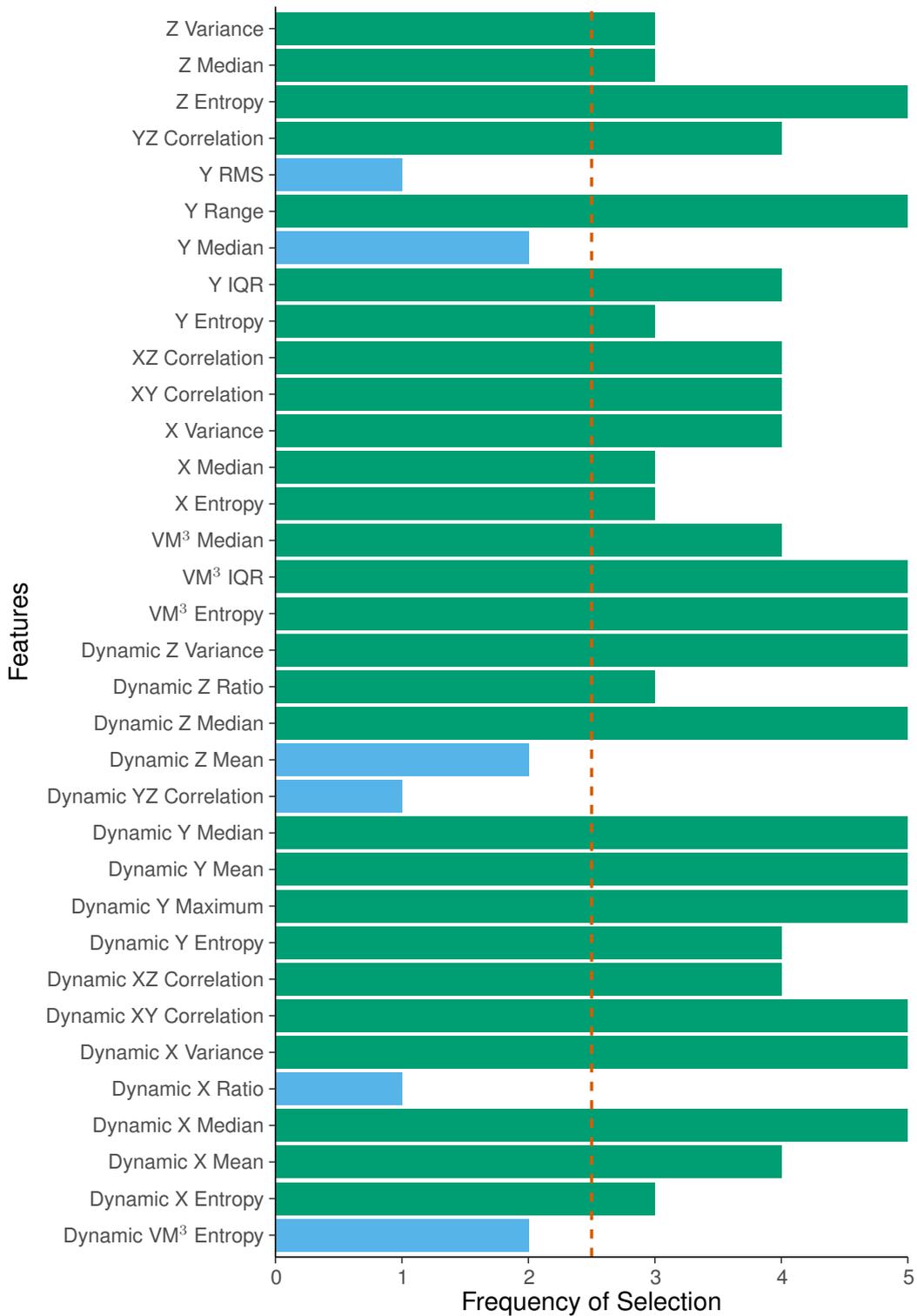


Figure 4.1: The frequency of inclusion of features within cross-validation of Postural state classification models. The 2.5 frequency threshold is marked.

Table 4.1: The performance statistics of the identification of the Standing postural state from among all other recorded postures using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
Rotation Corrected & Second Dynamic Filter	Manual	0.718	0.663	0.705	0.806	0.728	0.624
	Forward	0.763	0.739	0.684	0.812	0.774	0.587
	Backward	0.775	0.737	0.703	0.821	0.777	0.577
	CFS	0.770	0.724	0.717	0.826	0.772	0.575
Standardised Absolute Rotation Corrected & Second Dynamic Filter	Manual	0.616	0.654	0.537	0.723	0.687	0.632
	Forward	0.715	0.697	0.646	0.785	0.738	0.608
	Backward	0.743	0.736	0.653	0.797	0.766	0.606
	CFS	0.731	0.719	0.662	0.798	0.757	0.597
Non-Rotation Corrected & Second Dynamic Filter	Manual	0.765	0.732	0.695	0.816	0.772	0.580
	Forward	0.703	0.551	0.794	0.832	0.663	0.638
	Backward	0.769	0.697	0.725	0.825	0.755	0.611
	CFS	0.762	0.648	0.783	0.847	0.734	0.607
Standardised Absolute Non-Rotation Corrected	Manual	0.646	0.528	0.689	0.759	0.623	0.669
	Forward	0.731	0.641	0.733	0.817	0.718	0.609
	Backward	0.781	0.737	0.709	0.824	0.778	0.599
	CFS	0.768	0.678	0.752	0.835	0.749	0.621
Rotation Corrected No Second Dynamic Filter	Manual	0.714	0.662	0.705	0.806	0.727	0.626
	Forward	0.761	0.736	0.686	0.813	0.773	0.587
	Backward	0.773	0.768	0.672	0.812	0.790	0.570
	CFS	0.768	0.749	0.686	0.815	0.780	0.568
Standardised Absolute Rotation Corrected No Second Dynamic Filter	Manual	0.618	0.664	0.522	0.720	0.691	0.627
	Forward	0.714	0.673	0.664	0.787	0.725	0.613
	Backward	0.747	0.755	0.636	0.793	0.774	0.601
	CFS	0.738	0.735	0.643	0.792	0.763	0.602
Non-Rotation Corrected No Second Dynamic Filter	Manual	0.764	0.746	0.681	0.812	0.778	0.574
	Forward	0.685	0.528	0.787	0.821	0.643	0.644
	Backward	0.787	0.741	0.724	0.832	0.784	0.573
	CFS	0.779	0.736	0.722	0.831	0.780	0.566
Absolute Standardised Non-Rotation Corrected No Second Dynamic Filter	Manual	0.643	0.525	0.682	0.754	0.619	0.669
	Forward	0.742	0.677	0.724	0.819	0.742	0.612
	Backward	0.773	0.695	0.746	0.835	0.759	0.624
	CFS	0.762	0.671	0.754	0.835	0.744	0.631

4.4.1.2 *Sitting Detection*

As before one sixth of the data ($N = 7430$) were randomly selected and excluded from the initial model training and assessment for validation purposes. A dataset of 37146 5 second windows with 5960 of those classified as Sitting. Table 4.2 shows the performance, across a 5 fold cross-validation, for all combinations of preprocessing and feature selection methodologies.

The best performing preprocessing and feature selection methodology, in terms of the two summary measures, was the backward stepwise feature selection on data that had neither the rotation correction nor a second dynamic filter applied. Within this methodology features had also been standardised, and only the absolute value taken prior to the construction of the LDA model. Both the AUC (0.74) and the F Score (0.43) were the highest recorded across all methodologies and feature selection methods processed. This is despite the Sensitivity (0.651), Specificity (0.73), and the Precision (0.32), all being outperformed by alternative method combinations. However, such a finding suggests that where each of these were outperformed the antagonistic nature of the measures resulted in an imbalanced trade-off in another measure. This is supported through further inspection of values in Table 4.2. The optimal threshold, used for calculation of performance measures, was 0.178.

Figure 4.2 summarises the 39 features which occurred in greater than 50% of cross validation folds for the selected method and were therefore included within the validation model. When applied to the previously withheld data the summary measures of AUC and F Score are 0.73 and 0.42 respectively. Sensitivity, Specificity and Precision give values of 0.70, 0.69, and 0.30 suggesting the Precision of the model is the cause of the lower F Score value and suggests a high false positive rate and low false negative rate. An optimised decision threshold of 0.16 was selected during validation model training.

Table 4.2: The performance statistics of the identification of the Sitting postural state from among all other recorded postures using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
Rotation Corrected & Second Dynamic Filter	Manual	0.682	0.669	0.603	0.244	0.357	0.151
	Forward	0.686	0.626	0.681	0.273	0.380	0.178
	Backward	0.722	0.646	0.691	0.286	0.396	0.176
	CFS	0.697	0.631	0.664	0.264	0.373	0.172
Standardised Absolute Rotation Corrected & Second Dynamic Filter	Manual	0.640	0.387	0.829	0.302	0.339	0.223
	Forward	0.544	0.890	0.214	0.178	0.297	0.160
	Backward	0.712	0.612	0.718	0.293	0.397	0.187
	CFS	0.695	0.618	0.687	0.274	0.380	0.175
Non-Rotation Corrected & Second Dynamic Filter	Manual	0.673	0.341	0.922	0.456	0.390	0.260
	Forward	0.623	0.709	0.487	0.209	0.323	0.164
	Backward	0.714	0.555	0.751	0.299	0.388	0.191
	CFS	0.688	0.467	0.796	0.304	0.368	0.210
Standardised Absolute Non-Rotation Corrected	Manual	0.655	0.327	0.916	0.426	0.370	0.264
	Forward	0.643	0.704	0.527	0.221	0.337	0.171
	Backward	0.742	0.680	0.677	0.287	0.404	0.156
	CFS	0.712	0.577	0.731	0.291	0.387	0.181
Rotation Corrected No Second Dynamic Filter	Manual	0.680	0.627	0.640	0.250	0.357	0.157
	Forward	0.683	0.628	0.693	0.281	0.388	0.178
	Backward	0.716	0.664	0.667	0.276	0.390	0.169
	CFS	0.695	0.667	0.625	0.254	0.367	0.164
Standardised Absolute Rotation Corrected No Second Dynamic Filter	Manual	0.647	0.382	0.841	0.315	0.345	0.228
	Forward	0.577	0.625	0.505	0.194	0.296	0.162
	Backward	0.718	0.598	0.741	0.307	0.405	0.194
	CFS	0.695	0.580	0.722	0.285	0.382	0.184
Non-Rotation Corrected No Second Dynamic Filter	Manual	0.674	0.351	0.914	0.439	0.390	0.256
	Forward	0.589	0.522	0.652	0.223	0.312	0.162
	Backward	0.734	0.622	0.723	0.300	0.405	0.179
	CFS	0.712	0.688	0.619	0.256	0.373	0.154
Absolute Standardised Non-Rotation Corrected No Second Dynamic Filter	Manual	0.659	0.352	0.891	0.382	0.366	0.250
	Forward	0.623	0.580	0.620	0.226	0.325	0.170
	Backward	0.742	0.651	0.730	0.315	0.425	0.178
	CFS	0.722	0.640	0.719	0.304	0.412	0.179

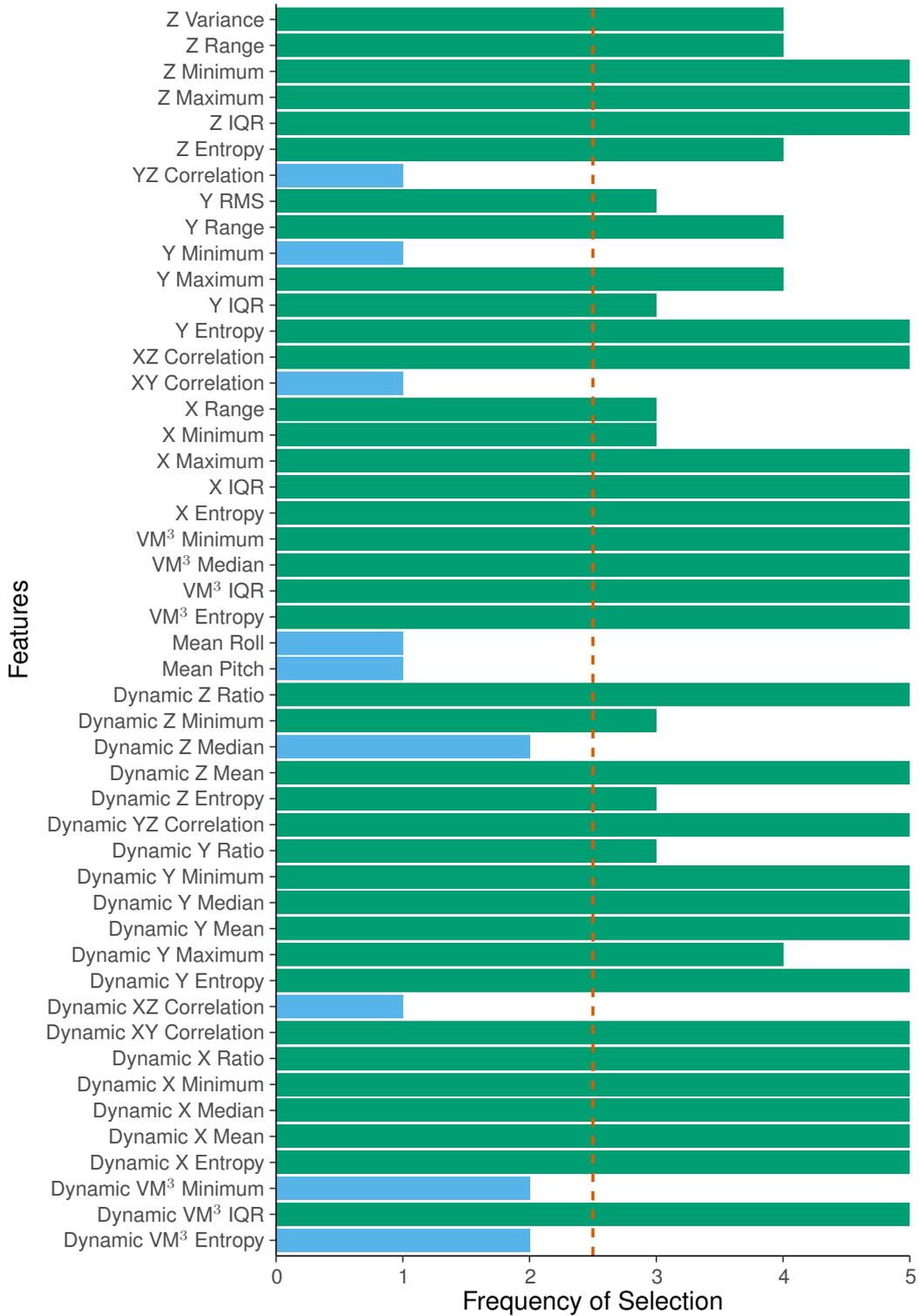


Figure 4.2: The frequency of inclusion of features within cross-validation of the One vs All Sitting postural state classification models. The 2.5 frequency threshold is marked.

4.4.1.3 Lying Detection

Approximately one sixth of the data ($N = 7429$) was excluded for use in validation of the final trained model. Within the remaining training dataset of 37417 5 second windows of acceleration data a total of 7063 were labelled as Lying. Table 4.3 displays the performance of each combination of preprocessing and feature selection methodologies which were attempted.

The performance results suggest that the best methodology for the identification of the Lying posture is the backward stepwise feature selection applied onto the absolute scaled features that have not been corrected for collar rotation but have had a second filter applied to the dynamic acceleration component. The AUC of this methodology (0.794), produced during the 5 fold cross-validation, was the highest reported for the detection of Lying down. Using an optimised decision threshold of 0.225 also produced the highest reported F Score (0.506) and value of Precision (0.392). The values of Sensitivity and Specificity reported were 0.713 and 0.741 respectively.

Figure 4.3 shows a total of 25 features were selected in greater than 50% of cross-validation folds and were therefore incorporated into the final validation model. When trained on all training data and applied to the previously excluded test data an AUC value of 0.78 was achieved. At an optimised decision threshold of 0.21 the F Score, Sensitivity, Specificity, and Precision were 0.48, 0.75, 0.68, and 0.36 respectively. Again this suggests the detection of the lying posture struggles in terms of precision and this is a trend that can also be seen across the training results presented in Table 4.3.

Table 4.3: The performance statistics of the identification of the Lying postural state from among all other recorded postures using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
Rotation Corrected & Second Dynamic Filter	Manual	0.694	0.630	0.687	0.321	0.425	0.211
	Forward	0.733	0.717	0.644	0.321	0.444	0.197
	Backward	0.756	0.701	0.687	0.344	0.462	0.198
	CFS	0.742	0.708	0.652	0.323	0.443	0.193
Standardised Absolute Rotation Corrected & Second Dynamic Filter	Manual	0.632	0.597	0.618	0.268	0.370	0.196
	Forward	0.685	0.674	0.591	0.279	0.395	0.188
	Backward	0.739	0.640	0.725	0.354	0.456	0.191
	CFS	0.714	0.582	0.733	0.338	0.428	0.201
Non-Rotation Corrected & Second Dynamic Filter	Manual	0.748	0.638	0.744	0.369	0.468	0.231
	Forward	0.765	0.802	0.604	0.323	0.460	0.217
	Backward	0.776	0.793	0.642	0.342	0.478	0.210
	CFS	0.775	0.827	0.609	0.332	0.474	0.206
Standardised Absolute Non-Rotation Corrected	Manual	0.636	0.617	0.618	0.275	0.380	0.193
	Forward	0.724	0.780	0.576	0.302	0.435	0.210
	Backward	0.794	0.713	0.741	0.392	0.506	0.225
	CFS	0.772	0.766	0.638	0.332	0.463	0.215
Rotation Corrected No Second Dynamic Filter	Manual	0.693	0.636	0.683	0.320	0.426	0.210
	Forward	0.734	0.697	0.660	0.325	0.443	0.197
	Backward	0.760	0.657	0.733	0.366	0.470	0.208
	CFS	0.751	0.649	0.731	0.362	0.465	0.207
Standardised Absolute Rotation Corrected No Second Dynamic Filter	Manual	0.630	0.621	0.589	0.262	0.368	0.190
	Forward	0.684	0.692	0.575	0.277	0.395	0.184
	Backward	0.742	0.649	0.717	0.350	0.455	0.189
	CFS	0.718	0.611	0.708	0.329	0.428	0.197
Non-Rotation Corrected No Second Dynamic Filter	Manual	0.751	0.647	0.741	0.370	0.470	0.231
	Forward	0.688	0.757	0.565	0.290	0.419	0.196
	Backward	0.775	0.781	0.671	0.358	0.491	0.218
	CFS	0.766	0.760	0.680	0.358	0.486	0.223
Absolute Standardised Non-Rotation Corrected No Second Dynamic Filter	Manual	0.636	0.638	0.596	0.271	0.380	0.189
	Forward	0.734	0.776	0.600	0.313	0.446	0.209
	Backward	0.791	0.748	0.707	0.375	0.499	0.213
	CFS	0.769	0.808	0.614	0.329	0.468	0.207

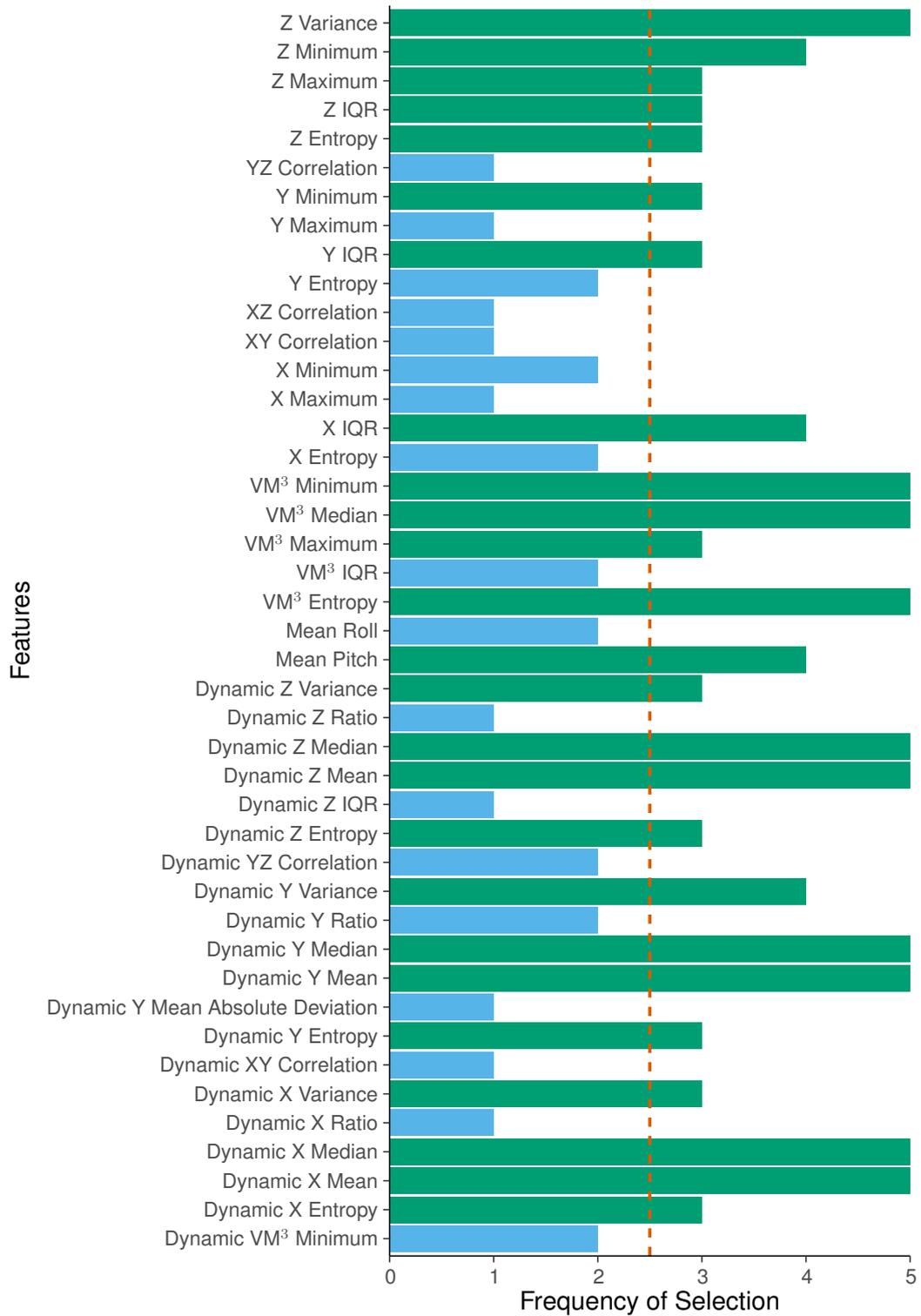


Figure 4.3: The frequency of inclusion of features within cross-validation of the One vs All Lying postural state classification models. The 2.5 frequency threshold is marked.

4.4.2 Locomotion Detection

As with the postural detection the data used for the differentiation of locomotion and stationary standing postures was split into validation (N = 3179) and training (N = 15899) datasets. 3645 windows within the training data were identified as relating to locomotion behaviour. The manually selected feature subset was consisted of the dynamic equivalents of the static features used in the postural models of Section 4.4.1. To attempt to capture the rhythmic component of locomotion the manual model here also included features derived from the frequency spectra. These were the dominant frequency, dominant frequency amplitude, entropy of the spectrogram and the energy per sample domain features for both the Z and Y axes. Table 4.4 summarises the performance of each combination of preprocessing and feature selection methodologies attempted.

The results suggest that periods of locomotion were most effectively differentiated from those where the individual was stationary when using data that had not been corrected for rotation, a second filter had not been applied to the dynamic data, and where appropriate features had been normalised and the absolutes taken. The backward stepwise feature selection, using this method of preprocessing data, gave an AUC of 0.782. The optimised decision threshold was set at 0.189. At this value the F score given was 0.544, and the sensitivity, specificity and precision were 0.669, 0.764, and 0.458.

Figure 4.4 shows the frequency of selection for each of the features included within each of the folds of the 5 fold cross validation. 66 features were used in greater than 50% of the cross validation folds and as such were used to train a final model for validation and assessment. The AUC of this final model was 0.78 and an optimised decision threshold of 0.21 was selected. The F score, Sensitivity, Specificity, and Precision of this model, on the validation data, were 0.56, 0.64, 0.80, 0.50 respectively.

Table 4.4: The performance statistics of the identification of periods of locomotion from among all standing posture periods using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
Rotation Corrected & Second Dynamic Filter	Manual	0.631	0.645	0.547	0.298	0.408	0.222
	Forward	0.630	0.611	0.609	0.318	0.418	0.211
	Backward	0.707	0.615	0.689	0.371	0.462	0.232
	CFS	0.677	0.707	0.558	0.323	0.443	0.199
Standardised Absolute Rotation Corrected & Second Dynamic Filter	Manual	0.592	0.498	0.633	0.291	0.367	0.249
	Forward	0.577	0.517	0.614	0.288	0.370	0.228
	Backward	0.691	0.687	0.596	0.339	0.454	0.205
	CFS	0.668	0.669	0.579	0.325	0.437	0.205
Non-Rotation Corrected & Second Dynamic Filter	Manual	0.696	0.755	0.537	0.329	0.458	0.178
	Forward	0.691	0.680	0.626	0.353	0.465	0.215
	Backward	0.736	0.711	0.664	0.388	0.502	0.218
	CFS	0.710	0.724	0.616	0.362	0.482	0.206
Standardised Absolute Non-Rotation Corrected	Manual	0.620	0.511	0.678	0.324	0.397	0.245
	Forward	0.655	0.669	0.560	0.315	0.428	0.200
	Backward	0.735	0.752	0.616	0.371	0.497	0.189
	CFS	0.717	0.681	0.661	0.378	0.486	0.210
Rotation Corrected No Second Dynamic Filter	Manual	0.632	0.546	0.662	0.326	0.408	0.243
	Forward	0.642	0.589	0.638	0.328	0.421	0.210
	Backward	0.704	0.727	0.575	0.339	0.462	0.196
	CFS	0.676	0.757	0.505	0.315	0.445	0.186
Standardised Absolute Rotation Corrected No Second Dynamic Filter	Manual	0.595	0.489	0.654	0.296	0.369	0.251
	Forward	0.607	0.550	0.615	0.298	0.387	0.217
	Backward	0.687	0.724	0.549	0.323	0.447	0.193
	CFS	0.655	0.677	0.552	0.310	0.425	0.201
Non-Rotation Corrected No Second Dynamic Filter	Manual	0.700	0.726	0.571	0.337	0.460	0.186
	Forward	0.667	0.608	0.649	0.342	0.438	0.225
	Backward	0.773	0.683	0.746	0.447	0.540	0.237
	CFS	0.753	0.708	0.688	0.405	0.516	0.207
Absolute Standardised Non-Rotation Corrected No Second Dynamic Filter	Manual	0.618	0.663	0.522	0.292	0.405	0.211
	Forward	0.696	0.563	0.737	0.389	0.460	0.217
	Backward	0.782	0.669	0.764	0.458	0.544	0.189
	CFS	0.741	0.628	0.761	0.439	0.517	0.180

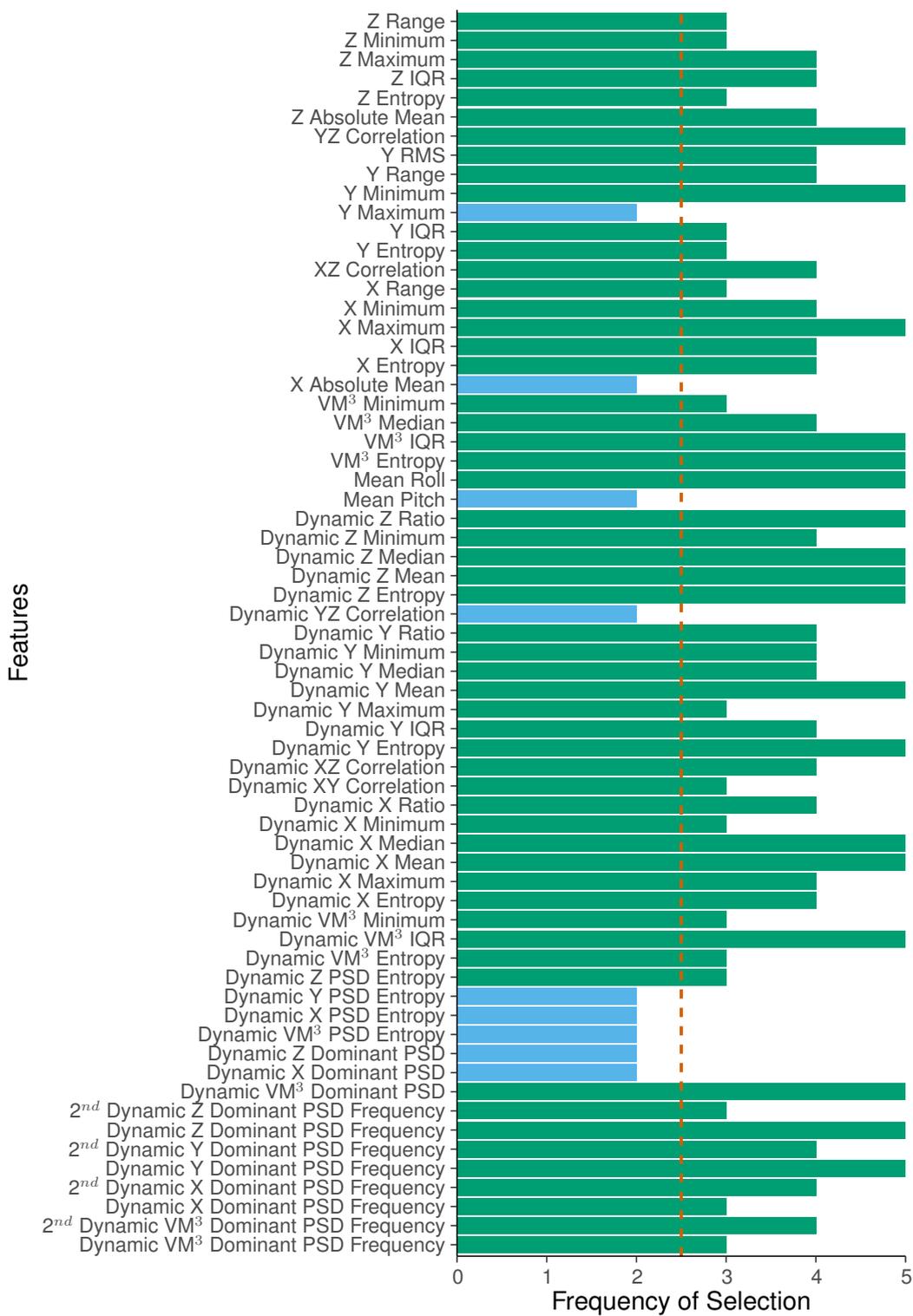


Figure 4.4: The frequency of inclusion of features within cross-validation of the Locomotion period classification models. The 2.5 frequency threshold is marked.

4.4.3 Paired Postural Binary Decisions

Re-framing the classification task into a series of binary decisions between postures, rather than a system utilizing all three one vs all classifiers, necessitates the removal of non-focal classes for model training to assess performance in ideal conditions. The manually derived feature set is the same as used in Section 4.4.1.

4.4.3.1 Standing VS Sitting

The combined dataset of Standing and Sitting (N = 36100) was, as before, split into a training dataset (N = 30083) and a validation dataset (N = 6017).

Table 4.5 gives the performance of this classification task for each combination of feature selection and preprocessing methodologies. Models using a backward feature selection on the absolute, scaled features derived from a dataset that had neither rotation correction nor a second dynamic filter applied showed the best performance with an AUC of 0.770. This is despite a less favourable F Score (0.808) at the optimised decision threshold (0.79) but, further inspection of those models which outperform in this measure shows that they do so at the expense of AUC and at least one of the two measures which contribute to it (Sensitivity or Specificity). This methodology can be observed to show a more equal distribution in performance with a Sensitivity of 0.729, a Specificity of 0.695, and a Precision of 0.906.

Figure 4.5 shows the frequency of selection for each feature included within the 5 fold cross validation of this methodology. 41 features occurred in greater than 50% of the cross-validation folds and were therefore included in the model applied to the previously excluded validation data. The AUC of this model was 0.76 and the decision threshold was set at 0.80. At this level the F Score was 0.79, the Sensitivity was 0.69, the Specificity was 0.72, and the Precision was 0.91.

Table 4.5: The performance statistics of the separation of the Standing and Sitting postural states using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
Rotation Corrected & Second Dynamic Filter	Manual	0.721	0.658	0.693	0.897	0.759	0.805
	Forward	0.729	0.668	0.721	0.907	0.769	0.774
	Backward	0.758	0.721	0.691	0.904	0.802	0.776
	CFS	0.743	0.708	0.675	0.898	0.792	0.772
Standardised Absolute Rotation Corrected & Second Dynamic Filter	Manual	0.652	0.871	0.369	0.848	0.859	0.712
	Forward	0.601	0.386	0.807	0.890	0.538	0.801
	Backward	0.743	0.787	0.588	0.886	0.833	0.746
	CFS	0.720	0.675	0.661	0.890	0.767	0.787
Non-Rotation Corrected & Second Dynamic Filter	Manual	0.723	0.712	0.618	0.883	0.788	0.775
	Forward	0.667	0.578	0.719	0.893	0.701	0.795
	Backward	0.746	0.694	0.670	0.895	0.782	0.797
	CFS	0.712	0.794	0.502	0.866	0.828	0.749
Standardised Absolute Non-Rotation Corrected	Manual	0.674	0.880	0.356	0.847	0.863	0.708
	Forward	0.700	0.635	0.685	0.891	0.742	0.780
	Backward	0.773	0.722	0.698	0.906	0.804	0.791
	CFS	0.748	0.766	0.604	0.887	0.822	0.763
Rotation Corrected No Second Dynamic Filter	Manual	0.722	0.632	0.720	0.901	0.743	0.811
	Forward	0.739	0.737	0.667	0.900	0.810	0.757
	Backward	0.758	0.742	0.681	0.904	0.815	0.772
	CFS	0.738	0.706	0.680	0.899	0.791	0.780
Standardised Absolute Rotation Corrected No Second Dynamic Filter	Manual	0.653	0.825	0.414	0.851	0.838	0.729
	Forward	0.661	0.459	0.802	0.904	0.609	0.801
	Backward	0.741	0.755	0.610	0.887	0.816	0.759
	CFS	0.726	0.746	0.604	0.884	0.809	0.764
Non-Rotation Corrected No Second Dynamic Filter	Manual	0.724	0.743	0.589	0.880	0.806	0.765
	Forward	0.589	0.450	0.683	0.852	0.589	0.802
	Backward	0.771	0.712	0.707	0.908	0.798	0.791
	CFS	0.760	0.717	0.696	0.905	0.800	0.783
Absolute Standardised Non-Rotation Corrected No Second Dynamic Filter	Manual	0.670	0.932	0.302	0.844	0.886	0.664
	Forward	0.706	0.620	0.728	0.902	0.735	0.783
	Backward	0.770	0.729	0.695	0.906	0.808	0.790
	CFS	0.755	0.733	0.672	0.900	0.808	0.782

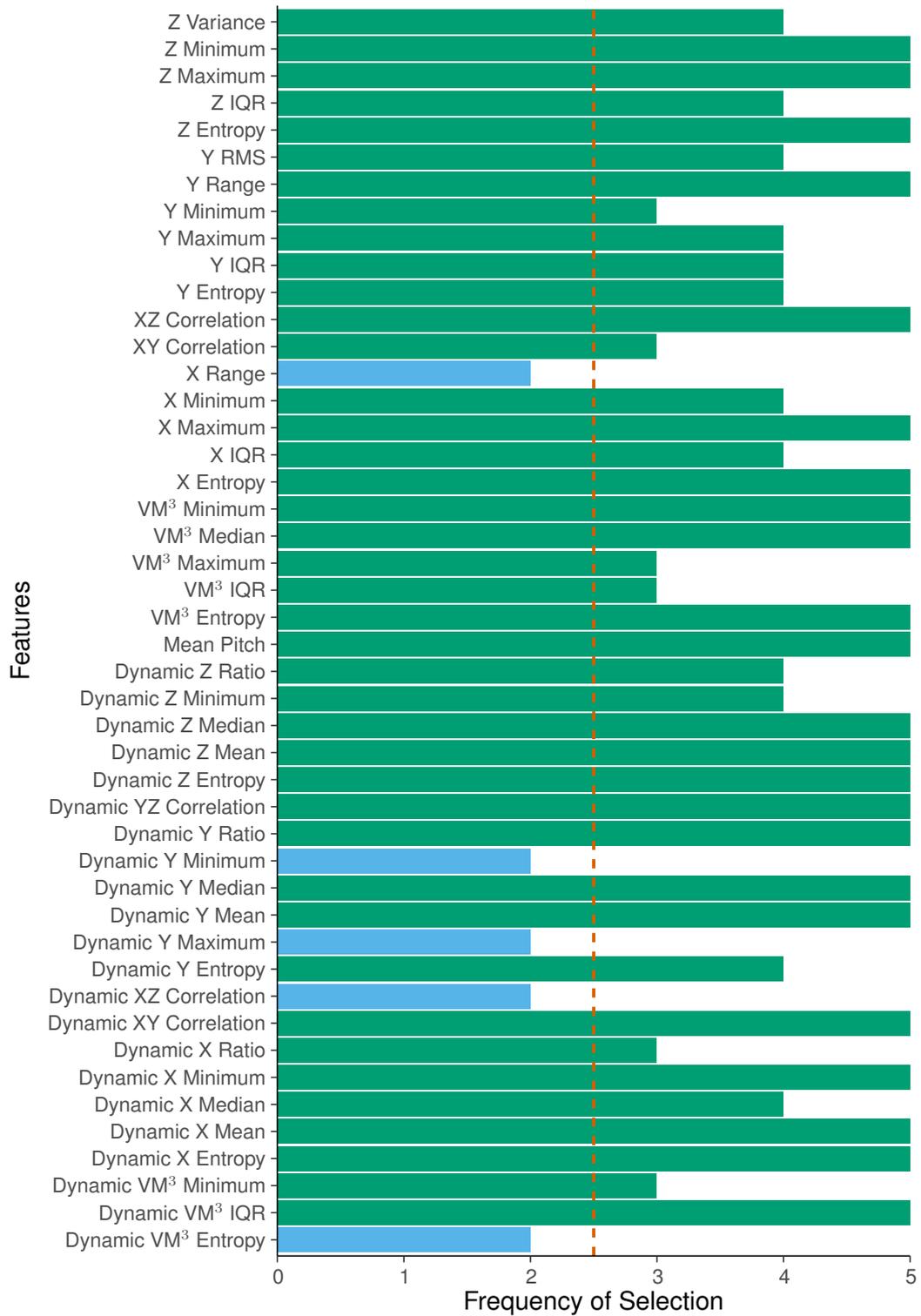


Figure 4.5: The frequency of inclusion of features within cross-validation of the combined Standing and Sitting postural state class classification models. The 2.5 frequency threshold is marked.

4.4.3.2 *Standing VS Lying*

A validation dataset (N = 6236) was excluded from a combination of both Standing and Lying periods (N = 37424) leaving a training set of N = 31118.

The performance measures in Table 4.6 suggest that a methodology consisting of backward feature selection on features derived from data that has not been corrected for rotation and has not had a second filter applied to the dynamic component. This is due to the both combined measures, the AUC (0.820), and the F Score (0.822) at the optimised threshold (0.715), are both the highest recorded among all other feature selection and pre-processing method combinations. The Sensitivity, Specificity, and Precision are 0.746, 0.763, and 0.915 respectively. All three of these contributing measures are outperformed by other method combinations but, as can be observed in Table 4.6, this is at the cost of one or both of the other contributing measures.

Figure 4.6 shows the selection frequency for each feature included by the LDA models constructed during the 5 fold cross-validation of the selected methodology. 32 features were included in greater than 50% of cross-validation models and were therefore also included within the validation model. When an LDA model was trained on the entire training dataset, using those features, and applied to the previously withheld validation data an AUC of 0.824 was achieved and a decision threshold of 0.733 was selected. At this threshold an F Score of 0.823 was obtained. The Sensitivity, Specificity, and Precision, which contribute to either the AUC and/or F Score, were 0.744, 0.780, and 0.920 respectively.

Table 4.6: The performance statistics of the separation of the Standing and Lying postural states using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
Rotation Corrected & Second Dynamic Filter	Manual	0.723	0.737	0.630	0.872	0.798	0.737
	Forward	0.779	0.748	0.697	0.894	0.815	0.740
	Backward	0.795	0.731	0.738	0.905	0.809	0.733
	CFS	0.788	0.735	0.729	0.903	0.810	0.727
Standardised Absolute Rotation Corrected & Second Dynamic Filter	Manual	0.634	0.613	0.598	0.839	0.708	0.770
	Forward	0.724	0.656	0.681	0.875	0.750	0.763
	Backward	0.766	0.719	0.695	0.890	0.795	0.765
	CFS	0.744	0.738	0.640	0.875	0.801	0.741
Non-Rotation Corrected & Second Dynamic Filter	Manual	0.786	0.725	0.737	0.904	0.804	0.731
	Forward	0.808	0.695	0.789	0.918	0.791	0.726
	Backward	0.810	0.687	0.808	0.925	0.788	0.733
	CFS	0.807	0.724	0.774	0.916	0.809	0.716
Standardised Absolute Non-Rotation Corrected	Manual	0.646	0.609	0.635	0.851	0.710	0.775
	Forward	0.752	0.614	0.788	0.908	0.733	0.748
	Backward	0.814	0.701	0.790	0.919	0.795	0.741
	CFS	0.808	0.690	0.791	0.919	0.788	0.737
Rotation Corrected No Second Dynamic Filter	Manual	0.722	0.739	0.626	0.871	0.799	0.734
	Forward	0.780	0.709	0.734	0.901	0.794	0.751
	Backward	0.800	0.757	0.714	0.900	0.822	0.737
	CFS	0.791	0.747	0.716	0.900	0.816	0.739
Standardised Absolute Rotation Corrected No Second Dynamic Filter	Manual	0.636	0.641	0.568	0.835	0.725	0.763
	Forward	0.726	0.669	0.667	0.873	0.758	0.763
	Backward	0.771	0.761	0.661	0.885	0.818	0.754
	CFS	0.752	0.732	0.657	0.879	0.799	0.755
Non-Rotation Corrected No Second Dynamic Filter	Manual	0.786	0.732	0.726	0.901	0.808	0.727
	Forward	0.809	0.692	0.842	0.937	0.796	0.736
	Backward	0.820	0.746	0.763	0.915	0.822	0.715
	CFS	0.813	0.745	0.759	0.913	0.820	0.702
Absolute Standardised Non-Rotation Corrected No Second Dynamic Filter	Manual	0.650	0.622	0.637	0.854	0.720	0.772
	Forward	0.774	0.654	0.792	0.915	0.763	0.745
	Backward	0.810	0.732	0.773	0.917	0.814	0.739
	CFS	0.802	0.685	0.818	0.928	0.788	0.746

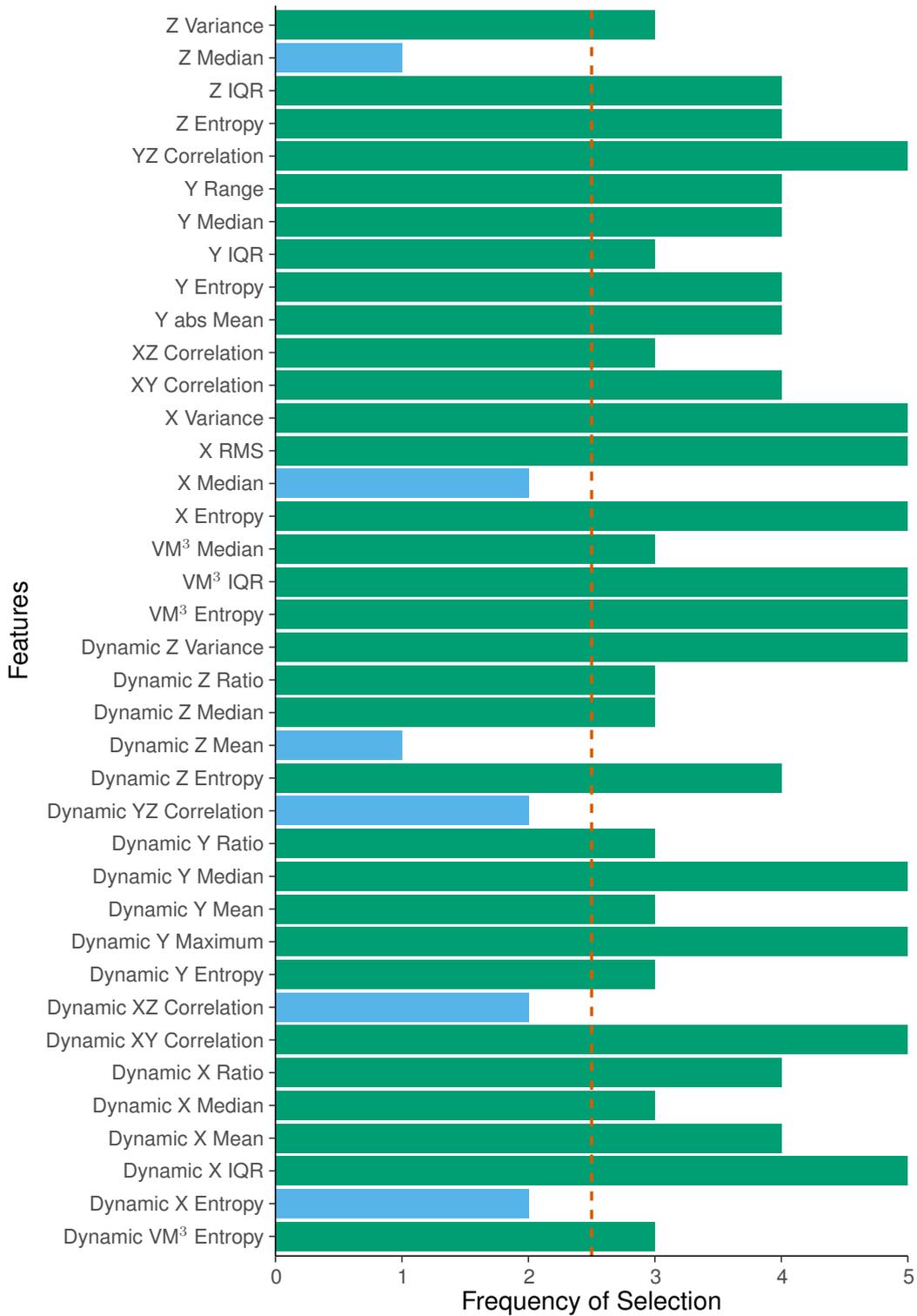


Figure 4.6: The frequency of inclusion of features within cross-validation of the combined Standing and Lying postural state class classification models. The 2.5 frequency threshold is marked.

4.4.3.3 Lying VS Sitting

The combined dataset of periods labelled as either Lying or Sitting postures (N = 15628) was split into a training dataset (N = 13024) and a validation dataset (N = 2604) prior to the initial model training and methodology performance assessment.

The performance values given in Table 4.7 suggest the standardised absolute of features calculated from data that had not been corrected for rotation, but where the dynamic acceleration component had had an additional filter applied, was the best synthesis of methods for this classification task. This is due to the AUC (0.784) being the highest reported among all alternatives but despite the reported F Score (0.743) being equivalent or worse to alternatives. The Sensitivity (0.756), Specificity (0.670), and Precision (0.731) are all outperformed by alternative combinations however further inspection of these results reveals heightened results in any one measure often coincides with reductions in one or both others.

Figure 4.7 shows the selection frequency for each feature included within the LDA models generated within the 5 fold cross validation. 50 features occurred in greater than 50% of generated models and were therefore used in the validation stage. The AUC of the validation model, trained on all training data and applied to the previously withheld validation data, was 0.757 and the optimal decision threshold was set to 0.594. At this threshold the F Score for this classification validation is 0.682, the Sensitivity is 0.613, the Specificity is 0.782, and the Precision is 0.769.

Table 4.7: The performance statistics of the separation of the Lying and Sitting postural states using LDA. Each permutation of feature selection, rotation correction, and dynamic filtering is presented with the highest performers highlighted.

		AUC	Sensitivity	Specificity	Precision	F Score	Optimised Decision Threshold
Rotation Corrected & Second Dynamic Filter	Manual	0.679	0.846	0.488	0.662	0.743	0.463
	Forward	0.681	0.777	0.562	0.678	0.724	0.474
	Backward	0.708	0.760	0.570	0.677	0.716	0.481
	CFS	0.664	0.741	0.578	0.675	0.707	0.487
Standardised Absolute Rotation Corrected & Second Dynamic Filter	Manual	0.661	0.475	0.745	0.688	0.562	0.595
	Forward	0.593	0.335	0.860	0.740	0.462	0.555
	Backward	0.742	0.755	0.632	0.708	0.731	0.498
	CFS	0.674	0.665	0.619	0.674	0.670	0.528
Non-Rotation Corrected & Second Dynamic Filter	Manual	0.664	0.777	0.513	0.654	0.710	0.491
	Forward	0.658	0.695	0.632	0.691	0.693	0.491
	Backward	0.749	0.807	0.578	0.694	0.746	0.473
	CFS	0.675	0.718	0.593	0.677	0.697	0.512
Standardised Absolute Non-Rotation Corrected	Manual	0.667	0.840	0.408	0.627	0.718	0.433
	Forward	0.636	0.672	0.572	0.650	0.661	0.531
	Backward	0.784	0.756	0.670	0.731	0.743	0.518
	CFS	0.718	0.752	0.594	0.687	0.718	0.502
Rotation Corrected No Second Dynamic Filter	Manual	0.679	0.842	0.495	0.664	0.742	0.469
	Forward	0.682	0.780	0.571	0.683	0.728	0.475
	Backward	0.707	0.807	0.520	0.666	0.730	0.465
	CFS	0.660	0.767	0.535	0.662	0.710	0.481
Standardised Absolute Rotation Corrected No Second Dynamic Filter	Manual	0.664	0.458	0.765	0.698	0.554	0.606
	Forward	0.617	0.439	0.811	0.733	0.549	0.554
	Backward	0.734	0.736	0.632	0.703	0.719	0.504
	CFS	0.679	0.648	0.647	0.685	0.666	0.528
Non-Rotation Corrected No Second Dynamic Filter	Manual	0.664	0.779	0.508	0.653	0.710	0.489
	Forward	0.648	0.709	0.590	0.672	0.690	0.491
	Backward	0.736	0.860	0.519	0.680	0.759	0.443
	CFS	0.667	0.656	0.637	0.682	0.669	0.527
Absolute Standardised Non-Rotation Corrected No Second Dynamic Filter	Manual	0.665	0.839	0.406	0.626	0.717	0.432
	Forward	0.633	0.699	0.541	0.643	0.670	0.528
	Backward	0.761	0.771	0.618	0.705	0.736	0.496
	CFS	0.710	0.723	0.609	0.687	0.705	0.512

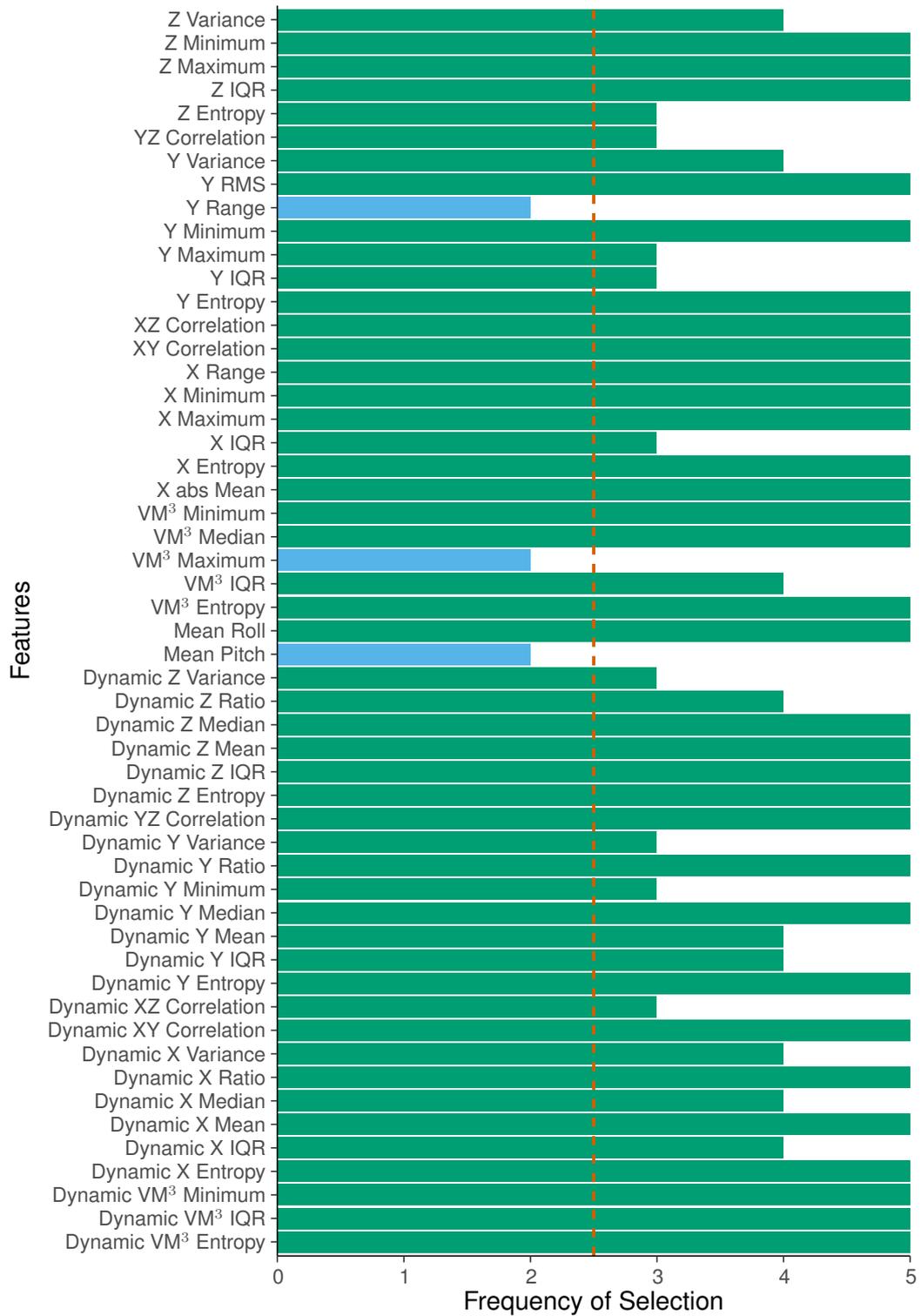


Figure 4.7: The frequency of inclusion of features within cross-validation of the classification models attempting to separate the Sitting and Lying postural states from each other. The 2.5 frequency threshold is marked.

4.5 Discussion

The classification of the postural states of dogs, using a single collar-mounted accelerometer, shows promise and would likely be deployable as currently described with minor adjustments to the structure of the classification process and the feature-sets used. However, the differentiation between stationary-standing and locomotion periods performed less well the methodology described above.

The first section aimed to identify postural states individually, using a one vs all technique the final classification could be assigned based on the highest probability output by the 3 LDA models. However, this methodology would only work if all 3 LDA models had acceptable performance across the assessed measures. The AUC values presented by the validation models for the 3 one vs all, postural classifications were between 0.73 and 0.79. Only the classification of the standing posture maintained acceptable performance when also assessed using the F score. The standing posture classifier attained a score of 0.78 which was much larger than the sitting and lying classifiers which achieved scores of 0.42 and 0.48 respectively. Further inspection reveals that both sitting and lying classifiers underperformed in terms of their precision (with values of 0.30 and 0.36 respectively), hence the lack of impact on the AUC of the model, which only summarises the measures of sensitivity and specificity. This means that both sitting and lying classifiers frequently misidentified a high proportion of positive classifications within the output. As the sensitivity typically remains high it is evident the models are including a large proportion of false positives within their results. As the standing classifier does not have poor precision it could be that the posture of standing is the most distinct from the other two, or inversely that sitting and lying are the most similar, when considering the time domain representations of acceleration included in the models. The inclusion of all three axes and the vector magnitude summaries within all 3 models suggests there are aspects of acceleration in each direction that are useful in describing the current posture. Additionally the inclusion of features relating to the dynamic component of acceleration in each axis and the vector magnitude could show that some of the ability of the classifiers to discern current postural state relates to the degree of activity which does or can occur within that posture. This is similar to past work which uses physical activity counts to assess the time spent within broad categories of activity levels, e.g. Morrison *et al.* (2013), which may lead to increased misclassification where activity level and device orientation do not both match a trained norm, e.g. stationary standing or vigorous movements within R.E.M sleep.

Section 2 focussed on the identification of locomotion, when standing, the posture wherein locomotion behaviour occurs, had been previously identified. This model and that for the identification of standing were intended to occur in series, and would ideally prevent the misclassification of movements occurring while in a non-standing posture

as locomotion. However, here this model was trained and tested with data that were labelled by traditional observation methods and was therefore operating in an artificial, ideal scenario. As such performance is an overestimate if the model were employed in a true hierarchical structure, as errors would likely be accrued. Even when tested on such data the performance of the locomotion classifier was sub-optimal. The AUC reported for the validation model, 0.78, would be promising. However, the F score of the classifier, 0.56, is once again lacking. Further inspection reveals that both the sensitivity and precision are low at 0.64 and 0.50 respectively. This arrangement of performance measures suggests a model that is reporting large proportions of false negative classifications, which in turn have led to a low proportion of the total number of periods identified as locomotion being true positives. As above summary features of both the dynamic and static acceleration components, for all 3 axes and the vector magnitude summary, were selected for inclusion in the final validation models. For this model frequency domain features were also calculated and included due to the rhythmic nature of locomotion (Watanabe *et al.*, 2005). Interestingly the frequencies at which dominant, and second dominant, PSDs occurred were included much more frequently than the PSD values themselves, suggesting the presence and rate of the rhythm is more useful to locomotion detection than the strength of that rhythm. The only dominant PSD selected frequently enough to be included within the validation model was that of the vector magnitude summary of all 3 axes of acceleration. This suggests the strength of the rhythm of the overall movement is more informative than that of the dominant rhythms of any one direction. It may be that Locomotion as a single category is too general and therefore led to a high degree of confusion with Stationary periods. Further separation of locomotion into differing gait categories may be desirable for future implementations. The inclusion of measures of activity level, or the interpretation of dynamic component measures to identify such, may also help to improve locomotion discretion by eliminating low energy periods entirely. The inclusion of additional gait categories would necessitate the validation of multiple new LDA models and further investigation of how best to implement them. Alternatively multi-class LDA or other models could be used instead. Additionally gait recognition can be challenging and often requires extensive observer training to achieve acceptable levels of reliability. The categorisation of movements into more broad, intensity-based categories, such as those used within Morrison *et al.* (2013), may therefore be more practical.

The third section involved the re-framing of the postural classification tasks into a series of binary LDA classifiers which each aimed to select between just 2 of the 3 posture categories. As previously mentioned this, when applied alongside selected classifiers from previous sections, would allow for the construction of a hierarchical structure of classification (Chakravarty *et al.*, 2019; Gao *et al.*, 2013). Arranging the classification tasks into 3 binary choices between pairs of postures, Stand vs Sit, Stand

vs Lie, and Lie vs Sit. Each of these reported higher values of both AUC and F Score measures when compared to the one vs all arrangements of the first section. AUC values for these three classifiers ranged between 0.75 and 0.82, and F Scores were between 0.68 and 0.82. The Stand vs Lie classifier was the best performing of the three with the highest values of both measures. In comparison the Lie vs Sit classifier was the worst, with both AUC and F Score values being the lowest reported. This supports previous conclusions that standing is the most easily identified and adds to the evidence that it is the sitting posture which is the most difficult to discern from the others, this is likely due to the relative lack of samples collected for that postural state. Once again the models were trained and assessed in ideal conditions, meaning they were each performed here in isolation and as a result there was no possibility of accrued errors from previous classifications.

These results help to inform potential arrangements of classifiers in future work. For example, as Standing appears the most easily identified then it can be classified earlier in the process (Figure 4.8). This arrangement would remove a large proportion of non-sitting and non-lying postures which would allow the use of the better performing Sitting vs Lying classifier of Section 3, rather than the less effective Sitting vs All or Lying vs All classifiers of Section 1. Another alternative, based on the overall lack of Sitting posture periods collected, would be to disregard the Sitting posture category entirely and instead run the best performing Standing vs Lying classifier across all data. This would invite a number false classifications as Sitting would not be a viable category. However, its rarity in comparison to the two other categories, and the higher functional importance of Standing and Lying, in terms of including behaviours of interest such as sleep or locomotion, means the performance costs may be minimal. From there consideration should be given to how the task is presented, as a Stand vs All problem, as in Section 1 where only Standing is labelled as the positive class and all non-Standing periods are assumed as Lying, or as a Stand vs Lie problem, as in Section 2 where examples of sitting and any other non-focal postures are not included in training data. As previously stated, the further subdivision of the Locomotion category may improve performance. This is also the case with the lying category and investigations into the orientation of the dog while lying (e.g. side-recumbent or sternal), and the type of behaviours which occur within each, should be performed.

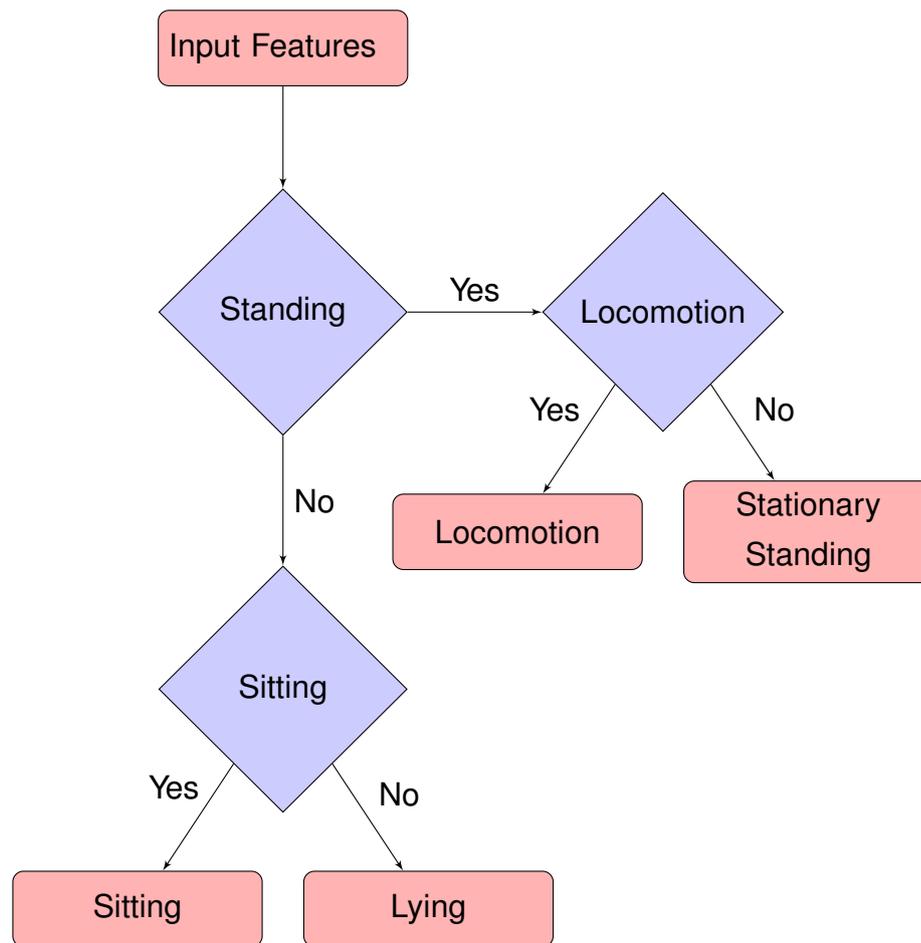


Figure 4.8: Arrangement of binary classification hierarchy suggested by results

The inclusion of different methods and combinations of pre-processing allowed some insight into the costs and benefits of employing such methods. The rotation correction methodology, as described in Chapter 3 and Ladha *et al.* (2018), works to dampen acceleration in the medial-lateral axis of the dog and to maximise acceleration in the dorsal-ventral axis. This would appear likely for the majority of classification tasks where it resulted in poorer performance during cross-validation for the models. However, video observation suggests a lack of rotation of the collars as may be expected in wider use with a more diverse range of dog breeds. As such further work should be performed to assess whether the loss in performance resulting from the implementation of the rotation correction is significant enough to outweigh any improvements that occur when regular collar rotation occurs during deployment. Similar further work should also be performed to assess the effect of the second dynamic filter. The intention of such a filter was to further clean the noise from dynamic data by only including frequencies relating to that of dog locomotion between an established range of breed heights. The filter itself may have been too strict, based as it is on biomechanical formulae and not on the actual observation of gait frequency within the sample, therefore removing information relating to the movement and acceleration of the individual which could aid in classification. Further work should explore the relaxing of the filter, as well

as its costs and benefits when applied on a more varied sample of dog breeds and heights.

As in Chapter 3 all assessed tasks performed best when using a backward stepwise feature selection methodology. This method is the greediest of those attempted, resulting in large feature-sets with high levels of feature redundancy, and suggests the tasks presented were difficult for the LDA classifiers used and there was no clear distinction (Derksen & Keselman, 1992; Mao, 2004). However, unlike Chapter 3 features were relatively rarely included within 1 or 2 cross validation folds. This suggests that unlike with postural transitions there was a lower level of feature inter-changeability within the overall feature-set.

Future work should focus on the refinement of models and the incorporation of knowledge gained during this exploration as well as within the complimentary findings of Chapters 2 and 3. Further validation of pre-processing methodologies and additional comparative investigations of more simplistic measures, such as overall activity counts and others discussed in Chapter 2, that could be used to infer postural or behavioural states should be performed. The use of additional sensors attached alongside the accelerometer would provide further details which could be useful in deriving posture. Gyroscopes are often deployed in both human and animal contexts for the direct measurement of device orientation and would be more reliable replacements for similar measures derived from static acceleration here (Wilson *et al.*, 2008). Collar-based sensor placement is suboptimal but, as previously mentioned, is more practical in longitudinal, free-ranging use-cases and should therefore be continued to be investigated despite lacking performance here. Such placements have also seen past success in other work (Hansen *et al.*, 2007; Ladha *et al.*, 2013; Ladha & Hoffman, 2018; Preston *et al.*, 2012). Alternative measures to imply locomotion, the refinement of behavioural subdivisions and hierarchy and an alternative model with a focus on the Standing vs All methodology should each be explored and assessed alongside comparisons and cost-benefit assessments of the pre-processing methodologies used here. Alternative methods of machine learning and classification, such as random forest or Support Vector Machines, may also prove more effective at the tasks than the LDA models employed here and now this initial exploration has provided some insight into the issue at hand more sophisticated models could be used, at the expense of run-speed, to better complete the task. Additionally, the stepwise feature selection methods should be replaced or developed further where possible due to the greedy nature of the method. Manually selected features and correlation based selection methods both performed poorly in comparison but investigation of features which occur frequently in the validation models here could inform future manually constructed models and produce better models with fewer features. As mentioned in Chapter 3 separation of classes could be further improved using dimensional reduction methods, such as PCA, but this would

reduce transparency and, if PCA is used, require alternative classification methods due to conflicts with LDA (Sunderam *et al.*, 2007).

4.6 Conclusion

This exploration of posture and locomotion classification suggests a selection of binary decision points produces classifiers with the highest levels of performance when using an LDA-based methodology. The differentiation of the standing posture from lying was particularly accurate and has the potential to be highly informative within an osteoarthritic context for observation of rest periods. The broader results across the many different permutations assessed here suggest that with further development acceptable levels of performance could be attained but the selection of features and the models used should be reassessed or developed further. One potential development would be the hierarchical structure outlined which may assist in the removal of false positives when assessing one vs all binary decisions. Additionally, The lists of feature selection frequency provide some insight into the influence of aspects of acceleration on this performance and could prove useful in future work. The application and assessment of methods such as PCA may allow the derivation of principal components that could be further assessed to understand the influence of groupings and commonalities between selected features. Further work with homogenous breeds should be performed prior to applying these methods on the wider range of breeds collected as part of the DogBox dataset.

Chapter 5

Circadian Rhythm and Nocturnal Activity as Canine Welfare Measures

5.1 Abstract

Patterns of activity and rest are potential indicators of health which could be automatically recorded using accelerometers. The circadian, and ultradian rhythms exhibited by the domestic dog could be vulnerable to interference by poor states of welfare such as stress, anxiety, pain or disease. Conversely many conditions have been suggested to be caused by consistent sleep disturbance. Osteoarthritis is once such condition which has been reported to be both caused and the cause of these disturbances. Modern sensor technology allows behaviour to be measured continuously over days, weeks, or even months, meaning that daily rhythms in activity and rest can be examined in free-living animals at resolutions and durations not previously possible. This allows the accurate and comprehensive assessment of the robustness of these rhythms in the face of chronic conditions, such as osteoarthritis, which would impose negative welfare costs. The aim of this study was to examine differences between patterns of activity and rest in healthy pet dogs and those with a chronic disease, osteoarthritis. We measured the periodicity of the activity of pet dogs using triaxial accelerometers mounted to their collars. Activity data were collected over 7 days in the home environment for 85 pet domestic dogs of various breeds where: 14 dogs had been diagnosed with osteoarthritis; 56 were healthy controls; and 15 had to be excluded from analyses. Summarised 24 hour periods for the 2 groups were calculated and examined for distinct group-level differences. The most active and least active 5 hour periods were calculated for the 2 groups and compared before passing the summarised data through a Fast Fourier Transformation. The number of significant harmonics occurring within the resultant frequency spectra, as selected through the calculation of Fisher's g statistic, were examined and used to inform the selection of frequencies of interest for individual-level analyses. The 24 hour period of each individual dog was calculated, the most and least active 5 hour periods and features relating to the rhythmic nature of activity were extracted from the 3 acceleration axes and the Overall Dynamic Body Acceleration summarisation. Through the application of an FFT the maximum amp-

litude observed, the average amplitude of all frequencies, the entropy of the frequency spectra, the kurtosis and skewness of the spectra, and the mean absolute deviation were all calculated for use as features. The number of significant harmonics occurring within the spectra, calculated using Fisher's g as before, was also assessed and used, as were the amplitudes of frequencies of interest that had previously been identified at the group level. Each feature was included in a mixed logistic regression with a random effect for individual. Of the 7 features found to be significantly indicative of osteoarthritic group the mean ODBA Fourier component amplitude, the number of significant harmonics within the X axis, and the amplitude of the component relating to a 12 hour rhythm were identified as the best individual predictors from among cross-correlates. The potential of the selected features to indicate the overall impact and severity of the condition upon the quality of life of the dogs was also explored using the LOAD scores as an outcome variable. Significant relationships were revealed between each of the 3 highlighted features but only the mean ODBA Fourier component amplitude was found to explain a meaningful level of LOAD score variance. Additionally, each of the 3 final features were assessed for their correlation with age. This revealed that those found to be most indicative of high LOAD scores or the presence of osteoarthritis were also the most collinear with age. These results suggest the monitoring of the ultradian and circadian rhythms of behaviour shown in dogs could be used to indicate a reduction in welfare state, specifically one produced by the presence or onset of osteoarthritis but that further adjustment to account for confounds should be investigated further to attempt to disentangle these and produce features that are both predictive and independent.

5.2 Introduction

Behaviour is often rhythmic and the most prominent behavioural rhythm is that of the mammalian sleep/wake cycle. This fluctuating pattern of activity, the circadian rhythm, is due to a mechanism present and consistent across all bodily tissues. The period is approximately 24 hours in healthy individuals and is entrained and maintained by environmental stimuli, typically light (Aschoff & Pohl, 1978; Reppert & Weaver, 2002). Within the circadian rhythm there also exist additional ultradian rhythms, those cycles occurring wholly within a single circadian period, and infradian rhythms, those which are longer than a circadian period. Changes or disruptions to any of these rhythms could be a valuable indicator of a change in welfare state and requires the inspection of behaviour at a level frequently overlooked in favour of more immediate and specific indications of behaviour such as those described in previous chapters. In unhealthy or welfare-impaired subjects, for example those suffering from neurodegeneration, ageing, inflammation, chronic stress, or chronic pain, circadian rhythms of activity and rest

are often reported to exhibit varying degrees of disruption (Cheeta *et al.*, 1997; Drewes, 1999; Fast *et al.*, 2013; Hamilton *et al.*, 2007; Volicer *et al.*, 2001; Wulff *et al.*, 2010).

In humans, disruption to the initiation and maintenance of sleep has been well established as a co-morbidity of osteoarthritis and is a frequently reported concern by sufferers (Parmelee *et al.*, 2015). Such disruptions may be due to the degradation of the maintenance of the circadian rhythm or are in spite of the influence of the metabolic processes responsible. Additionally, it has been suggested that disturbances to human sleep could be a causal factor in the development of many diseases, including osteoarthritis (Berenbaum & Meng, 2016). Berenbaum and Meng (2016) describes the bidirectional nature of the relationship between circadian rhythm and osteoarthritis in humans may potentially form a negative feedback loop and necessitates intervention as early as is practical. It is reasonable to expect that due to similarities between canine and human osteoarthritis that a similar co-morbidity could occur in dogs which could be exploited for the monitoring and assessment of the condition (Meeson *et al.*, 2019; Pond & Nuki, 1973).

Sleep and rest disturbances have been noted by owners of dogs in chronic pain and disturbances to rest behaviour have been reported in dogs with surgically induced arthritis (Knazovicky *et al.*, 2015; Little *et al.*, 2016). Previous research has recognised the potential for the monitoring of circadian rhythms in activity and rest as an indicator of health and welfare (Langford & Cockram, 2010; Owczarczak-Garstecka & Burman, 2016). Together this suggests monitoring of the rhythmic nature of activity could provide effective indicators of health and welfare in dogs with osteoarthritis. However the inherent differences in the activity patterns of dogs necessitates the development of a methodology that accounts for the idiosyncrasies of domestic dogs. For example, the sleep-wake cycle of dogs differs from that of humans as they display polyphasic, rather than biphasic, patterns and periods of sleep and wakefulness are short when compared directly (Adams & Johnson, 1993; Bódizs *et al.*, 2020; Kis *et al.*, 2017; Takahashi *et al.*, 1978). As a result the nature of domestic dog activity may be better described by ultradian rhythms, those with periods less than the 24 hour circadian rhythm, and as such the measurement of any impact to these shorter periods of behaviour is also of interest in the recognition of osteoarthritis-related disruption.

Anderson *et al.* (2018) has shown that, in humans, we can measure both rest and activity using a wrist worn accelerometer and hence obtain a convenient and non-invasive window into the underlying circadian sinusoid signal. In this chapter I aim to build on previous work on the investigation of behavioural rhythmicity in both human and non-human species to develop a methodology that investigates the potential of rhythmic disruption as an indicator of osteoarthritis in domestic dogs, and related chronic pain and welfare impacts. The use of tri-axial accelerometers attached to the collars of domestic dogs has previously been demonstrated to be an effective tool in

the measurement of canine rest and activity (Ladha *et al.*, 2013; Preston *et al.*, 2012; Zanghi *et al.*, 2013). I hypothesise that osteoarthritic dogs will exhibit more fragmented activity rhythms, including those likely related to sleep and wakefulness, and that there would be less distinction between rest and active states due to the previously observed decrease in overall activity.

5.3 Materials and Methods

5.3.1 Data Collection and Preprocessing

The previously used sample of $n = 86$ dogs of mixed breed, age and sex, as described in Section 2.3 and detailed in full in Appendix A, were each fitted with a collar-mounted tri-axial accelerometer for a 7-day period. The sample details, attachment protocol, sampling protocol, and data download and interpolation methodology were all identical to those described in Chapter 2. All collected periods were trimmed to start and end at midnight 7 days apart. 18 dogs were removed due to less than 7 days of continuous data remaining when processed in this way. This resulted in a final sample of $n = 68$ dogs. 14 of these dogs had been diagnosed with osteoarthritis and 54 dogs formed the control group.

The 100Hz acceleration data of each separate axis and of the Overall Dynamic Body Acceleration (ODBA) were summarised to a 1 minute sampling rate, 1/60Hz, as sub-minute period behaviours were not of interest. The ODBA was chosen due to the findings of Chapter 2, where it was the measure showing the most distinction between both control and osteoarthritic groups. The X, Y and Z axes were included as indications of data features excluded by the use of the ODBA.

As the data collection for each dog began at different points during the week due to the inconsistencies in owner availability the 7 day period collected was further summarised into a mean 24 hour period. This was generated for each dog by taking the mean of the 7 values for each minute of a day, resulting in a total of 1440 values per dog. The mean 24 hours of both the control and osteoarthritic groups were calculated in the same fashion, by taking the mean of each equivalent minute across the dogs in each of the 2 groups. Mean days, both per group and per dog, were further summarised to an hourly resolution and a rolling window of 5 hours applied. Values of ODBA were summed for each 5 hour period and the maximum and minimum values extracted, giving the most active and least active 5 hour periods of overall activity (M5 and L5 respectively) often used as indications of the phase of activity rhythms (Korte *et al.*, 2004; Ortiz-Tudela *et al.*, 2014). The mean ODBA per minute over 24 hours for each of the 2 groups are shown in Figure 5.1.

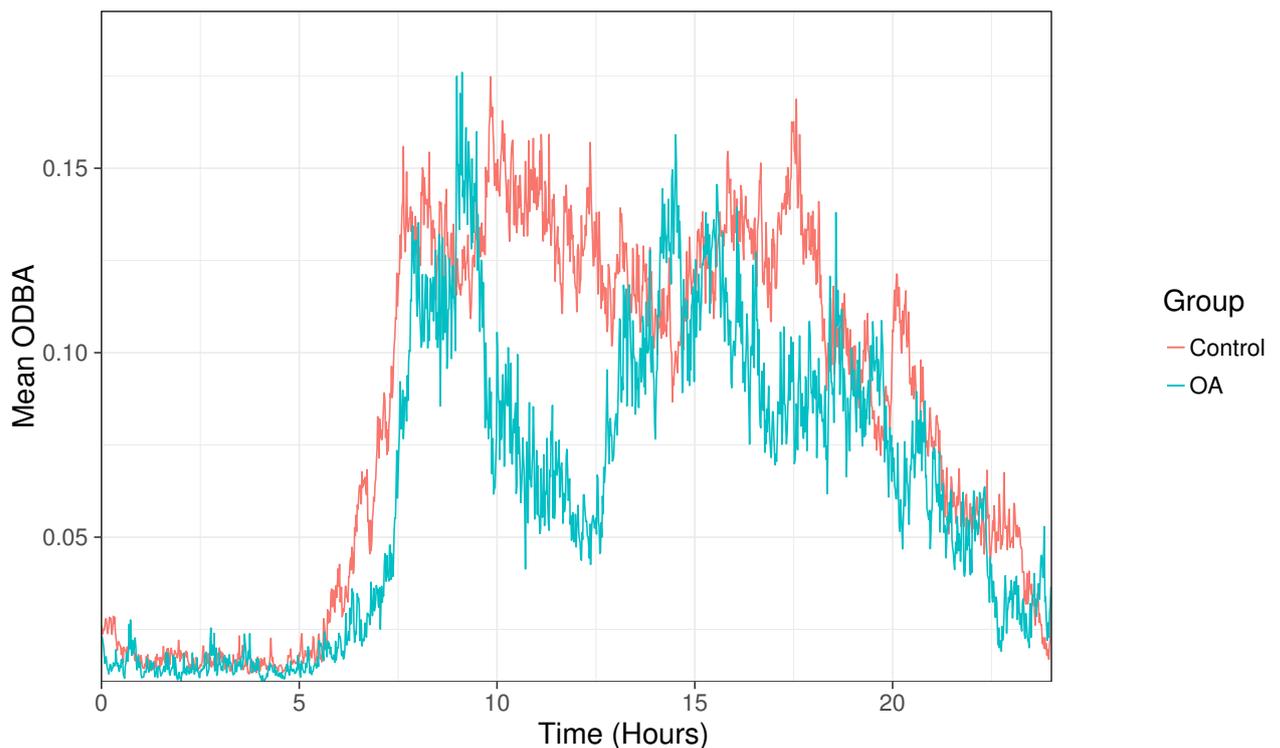


Figure 5.1: The mean 24 hour ODBA of Control and Osteoarthritic groups

A Fast Fourier Transform (FFT) algorithm was applied, following the methodology described in Chapter 2, to the X, Y, Z, and ODBA data of the mean 24 hour periods at both the individual and group levels. The resulting periodograms for the mean 24 hour ODBA of the 2 groups are shown in Figure 5.2. Once again following the methodology described in Chapter 2 the spectra of individual dogs was calculated and the dominant and second dominant magnitudes and frequencies were extracted as features for further analysis along with the mean, entropy, skewness, and kurtosis of the spectra. As can be seen from the group level spectra of 24 hour ODBA presented in Figure 5.2 there are more than 2 dominant frequencies. The strongest would be expected to be the circadian frequency however each of the others would likely represent an ultradian rhythm which may be impacted by the presence of a condition such as osteoarthritis. To identify frequencies which are considered significantly different from noise Fisher's g statistic was calculated and assessed (Fisher, 1929; Li *et al.*, 2020; Liew *et al.*, 2009; Percival & Walden, 1993). The same was performed at the individual level.

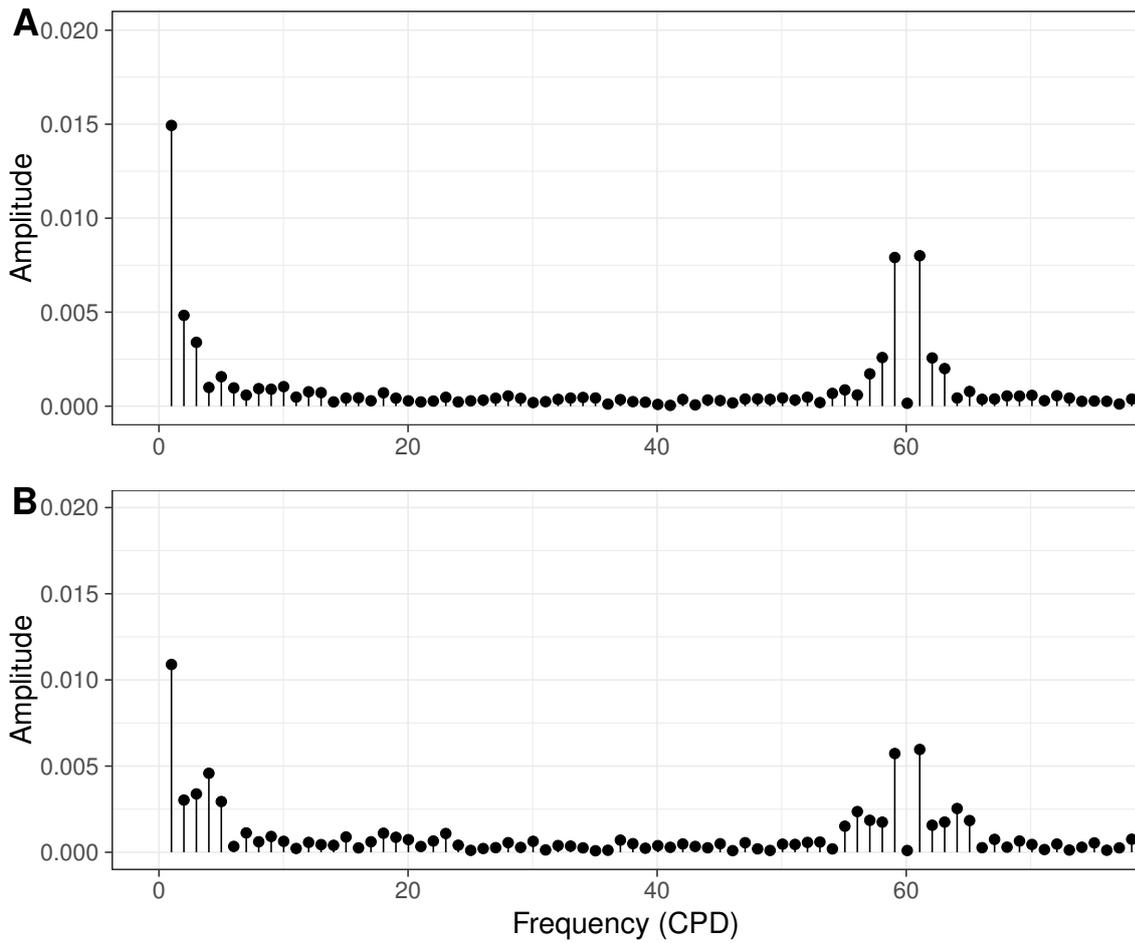


Figure 5.2: Frequency domain representations of the mean 24 hour ODBA of the A) Control and B) Osteoarthritic groups. Frequency represented as Cycles per Day (CPD).

5.3.2 Statistical Analyses

To identify significant Fourier components from the calculated spectra Fisher's g statistic was used (Fisher, 1929; Li *et al.*, 2020; Liew *et al.*, 2009; Percival & Walden, 1993). The g statistic is the ratio of the periodogram magnitude for the focal frequency to the sum of all magnitudes, up to the Nyquist frequency, and is frequently used across disciplines when interpreting spectra (Fisher, 1929; Li *et al.*, 2020; Liew *et al.*, 2009; Percival & Walden, 1993). The calculation for Fisher's g was implemented in R (R Core Team, 2018) using code available in Appendix G. The definition of Fisher's g is shown in equation 5.1 from Fisher (1929).

$$g = \frac{\max_k I(\omega_k)}{\sum_{k=1}^{N/2} I(\omega_k)} \quad (5.1)$$

$I(\omega)$ denotes the periodogram generated by the Fourier transform, N is the sample size, and $k = 0, 1, \dots, [N/2]$. The value of g is calculated for each fourier frequency where large values will result in a rejection of the null hypothesis that the frequency

magnitude is not significantly distinct from noise. The p value used to assess the significance of each g statistic was calculated using Equation 5.2.

$$P(g > \chi) = n(1 - \chi)^n - 1 - \frac{n(n-1)}{2}(1 - 2\chi)^n - 1 + \dots + (-1)^p \frac{n!}{p!(n-p)!}(1 - p\chi)^n - 1 \quad (5.2)$$

Here $n = \lceil N/2 \rceil$, χ is the observed value of the g statistic and p is equivalent to the largest value less than $1/\chi$. Once calculated for each Fourier frequency the bonferroni correction was applied to the p values to account for the high number of comparisons (Cabin & Mitchell, 2000; Li *et al.*, 2020). Frequencies were then ordered by descending g statistic and a significance threshold of $p = 0.05$ was set. Once a frequency component was found to be insignificantly different from others all remaining ordered components were discounted. This method was used to characterise the dominant frequencies of the mean 24 hour values of the ODBA for the osteoarthritic and control groups to investigate the presence of rhythmic aspects of overall activity at a group level. Fisher's g statistic was also applied to the X, Y, and Z axes, and the ODBA representation of each dog to extract the total number of significant harmonic periods as a feature for individual level comparison. This was included, along with the previously mentioned M5, L5 and FFT features, in mixed effect logistic regression models. The Fourier amplitudes of any significant rhythms found to differ between the 2 groups were also included as features.

A series of mixed effects logistic regression models were devised, using the lme4 package in R, to assess the ability of the calculated features to differentiate between the control and osteoarthritic groups. The individual was included as a random effect in each of the models. All features, except the M5, L5, and amplitudes of significant components as they were generated or selected from the ODBA, were calculated for the ODBA representation and the 3 raw axes. This was to investigate whether the information lost through summarisation was not of use to the task. Those found to be significant were assessed for correlation and a final selection of features was produced for discussion.

Following this, linear models aiming to explore the potential of identified features for their potential to indicate the severity and extent of the osteoarthritic condition were generated. This was achieved through the use of LOAD scores provided by owners. This measure has, in Chapter 2, been shown to be potentially more informative than a binary classification.

The potential effect of confounds on these features was also considered through the exploration of the collinearity of age. This continues to expand on the initial exploration of such relationships demonstrated in Chapter 2.

5.4 Results

The M5 value of the control group's mean 24 hour ODBA trace was between 09:00 and 14:00, the L5 for the same group was between 01:00 and 06:00. For the osteoarthritic group the M5 was between 14:00 and 19:00, and the L5 was also between 01:00 and 06:00.

Analysis of the Fisher's g statistic for each Fourier component of the control group ODBA waveform revealed 5 frequencies with statistically significant magnitudes. The first, and strongest component, is that relating to the 24 hour (1 cycle per day) circadian rhythm ($p = 7.913 \times 10^{-35}$). The two components to also be significantly strong within the signal are much shorter with periods of 0.393 and 0.406 hours, approximately 24 minutes ($p = 1.018 \times 10^{-15}$ and $p = 1.823 \times 10^{-15}$ respectively). An intermediate frequency component can be seen in Figure 5.2 which is not significant. As these are within seconds of each other, and due to the nature of the Fourier transform approximation and discretization of frequencies causing an increased likelihood of spectral leakage, it is probable that this pair of rapid ultradian rhythms is instead a single activity cycle (Zeng *et al.*, 2011). The mean 24 hour ODBA trace of the control dogs exhibited 2 further significant components. The first occurred approximately twice per day, a 12 hour period ($p = 2.417 \times 10^{-7}$), and the last occurred approximately three times daily, with an 8 hour period ($p = 0.00124$).

Table 5.1: The frequencies (CPD) and period length (hours) of harmonic components of the mean ODBA of Control dogs found to be significantly distinct through the use of Fisher's g statistic.

CPD	Period Length	Fisher's g	P value
1.001	23.967	0.120	7.913×10^{-35}
61.085	0.393	0.064	1.018×10^{-15}
59.082	0.406	0.064	1.823×10^{-15}
2.003	11.983	0.039	2.417×10^{-7}
3.004	7.989	0.027	0.00124

When the same analysis is applied to the mean 24 hour activity waveform of the osteoarthritic group 5 significant components are identified. The first three are identical to the first three of the control group with the first representing the circadian period ($p = 4.226 \times 10^{-19}$), and the 2 subsequent frequencies likely representing the same 24 minute rhythm ($p = 5.704 \times 10^{-8}$ and $p = 1.902 \times 10^{-7}$). The values of Fisher's g for each of these components are consistently lower than those reported for the control group, suggesting they are less distinct from the remaining signal noise (Fisher, 1929;

Percival & Walden, 1993). In lieu of the 12 hour period rhythms of the control group the osteoarthritic activity trace instead features an ultradian rhythm with a 6 hour period ($p = 6.434 \times 10^{-5}$). The 8 hour period ultradian rhythm is again present ($p = 0.02667$). As a result of these findings the amplitudes of the 12 hour and the 6 hour period rhythms were calculated as features in the mixed effects models.

Table 5.2: The frequencies (CPD) and period length (hours) of harmonic components of the mean ODBA of Osteoarthritic dogs found to be significantly distinct through the use of Fisher's g statistic.

CPD	Period Length	Fisher's g	P value
1.001	23.967	0.074	4.226×10^{-19}
61.085	0.393	0.041	5.704×10^{-8}
59.082	0.406	0.039	1.902×10^{-7}
4.006	5.992	0.031	6.434×10^{-5}
3.004	7.989	0.023	0.02667

The results for when each generated feature is included individually in mixed effects models alongside a random effect accounting for subject difference are shown in Table 5.3. Of the 33, 7 models are significant. The mean ODBA FFT amplitude of 24 hours, or the energy per sample, OR = 0.205 (CI: 0.078-0.541, $P = 0.001$), the entropy of the Fourier spectra of the X axis, OR = 0.518 (CI: 0.277-0.968, $P = 0.039$), the entropy of the Fourier spectra of the Z axis, OR = 0.472 (CI: 0.244-0.913, $P = 0.026$), the mean absolute deviation of the fourier transform magnitude, OR = 0.223 (CI: 0.087-0.575, $P = 0.002$), the number of detected significant harmonics in the x axis, OR = 2.621 (CI: 1.33-5.17, $P = 0.005$), and the Z axis, OR = 2.302 (CI: 1.18-4.49, $P = 0.014$), and the circadian amplitude of the ODBA, OR = 0.249 (CI: 0.088-0.700, $P = 0.0008$).

Table 5.3: The results of logistic regression models including each feature alongside a random effect accounting for the individual dog.

Variable	Odd's Ratio	2.5% CI	97.5% CI	P Value	Pseudo R ²
Dominant X Axis Amp.	1.318	0.751	2.312	0.336	0.023
Dominant Y Axis Amp.	1.050	0.573	1.926	0.874	0.001
Dominant Z Axis Amp.	1.473	0.651	3.335	0.353	0.038
Dominant ODBA Amp.	0.469	0.213	1.033	0.060	0.148
Mean X Axis Amp.	0.726	0.337	1.565	0.414	0.028
Mean Y Axis Amp.	0.894	0.490	1.634	0.716	0.004
Mean Z Axis Amp.	0.868	0.477	1.579	0.643	0.006
Mean ODBA Amp.	0.205	0.078	0.541	0.001	0.433
X Axis Spectrum Entropy	0.518	0.277	0.968	0.039	0.116
Y Axis Spectrum Entropy	0.826	0.425	1.608	0.574	0.010
Z Axis Spectrum Entropy	0.472	0.244	0.913	0.026	0.147
ODBA Spectrum Entropy	0.887	0.459	1.716	0.723	0.004
X Axis Spectrum Kurtosis	1.149	0.629	2.097	0.651	0.005
Y Axis Spectrum Kurtosis	1.103	0.603	2.017	0.751	0.003
Z Axis Spectrum Kurtosis	1.051	0.575	1.923	0.872	0.001
ODBA Spectrum Kurtosis	1.033	0.573	1.860	0.915	0.000
X Axis Spectrum Skewness	1.308	0.636	2.690	0.465	0.019
Y Axis Spectrum Skewness	1.180	0.614	2.269	0.620	0.008
Z Axis Spectrum Skewness	1.233	0.623	2.442	0.547	0.012
ODBA Spectrum Skewness	0.973	0.537	1.765	0.929	0.000
X Axis Spectrum Amp. MAD	0.841	0.434	1.628	0.607	0.009
Y Axis Spectrum Amp. MAD	1.013	0.553	1.857	0.967	0.000
Z Axis Spectrum Amp. MAD	0.985	0.544	1.784	0.960	0.000
ODBA Spectrum Amp. MAD	0.223	0.087	0.575	0.002	0.406
ODBA Sig. Harmonics	0.714	0.314	1.624	0.422	0.031
X Axis Sig. Harmonics	2.621	1.329	5.169	0.005	0.220
Y Axis Sig. Harmonics	1.144	0.608	2.152	0.677	0.005
Z Axis Sig. Harmonics	2.302	1.181	4.485	0.014	0.174
Mean ODBA M5	1.023	0.565	1.852	0.940	0.000
Mean ODBA L5	0.604	0.255	1.430	0.251	0.071
ODBA 24 hour Amp.	0.450	0.199	1.015	0.054	0.163
ODBA 12 hour Amp.	0.249	0.088	0.700	0.008	0.370
ODBA 6 hour Amp.	0.657	0.271	1.592	0.352	0.047

The correlation of the 7 selected features was assessed in Table 5.4. Where pairs of features exhibited a correlation coefficient of greater than 0.7 the feature whose model gave the highest pseudo R^2 value was retained. High correlation can be seen between mean absolute deviation of the ODBA Fourier components, and the mean amplitude of ODBA Fourier components. The latter was retained due to its higher R^2 value of 0.433. Similarly correlation was revealed between the entropy of the fourier spectra of the X and Z axes and the count of reported significant harmonics from within the 2 spectra. The number of detected significant harmonics of the X axis was selected with an R^2 value of 0.220.

The 3 remaining features were the mean amplitude of ODBA Fourier components, the number of detected significant harmonics of the X axis, and the amplitude of the 12 hour period Fourier component of the ODBA. Figures 5.3 A), B), and C) give an indication of the distribution of the features within and between groups. A simple indication of class separation is given in Figure 5.4 A), B), and C), where it can be observed that dogs suffering from osteoarthritis exhibit an overall reduced amplitude of Fourier components, including the 12 hour period component, along with an increased number of significant harmonics in the acceleration of the X axis.

Table 5.4: The correlation of features revealed to be significantly related to osteoarthritic group when included in logistic regression models with only a random term for individual

	Mean ODBA Amp.	X Spectrum Ent.	Z Spectrum Ent.	ODBA Spectrum Ent.	MAD	X Sig. Harmonics	Z Sig. Harmonics	ODBA 12hr Amp.
Mean ODBA Amp.	1.00	0.37	0.40	0.98	-0.34	-0.36	0.30	
X Spectrum Ent.	0.37	1.00	0.73	0.38	-0.92	-0.62	0.16	
Z Spectrum Ent.	0.40	0.73	1.00	0.39	-0.75	-0.93	0.16	
ODBA Spectrum Amp. MAD	0.98	0.38	0.39	1.00	-0.34	-0.36	0.34	
X Sig. Harmonics	-0.34	-0.92	-0.75	-0.34	1.00	0.70	-0.21	
Z Sig. Harmonics	-0.36	-0.62	-0.93	-0.36	0.70	1.00	-0.20	
ODBA 12hr Amp.	0.30	0.16	0.16	0.34	-0.21	-0.20	1.00	

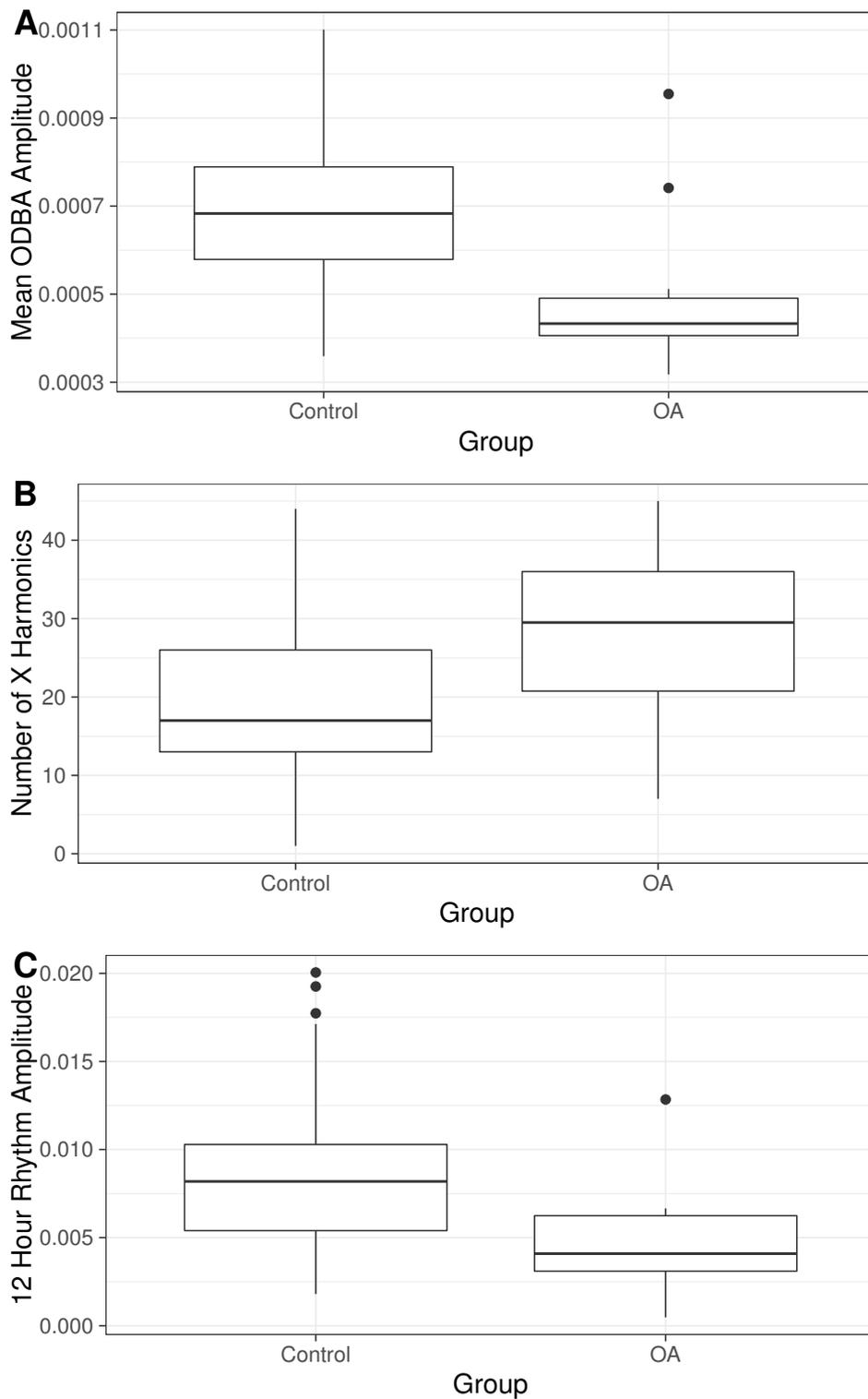


Figure 5.3: A series of plots indicating the inter and intra-group variance of significant, uncorrelated features. A) The Amplitude of the mean ODBA, B) the number of identified X axis harmonics, C) the amplitude of the 12 hour period ultradian rhythm.

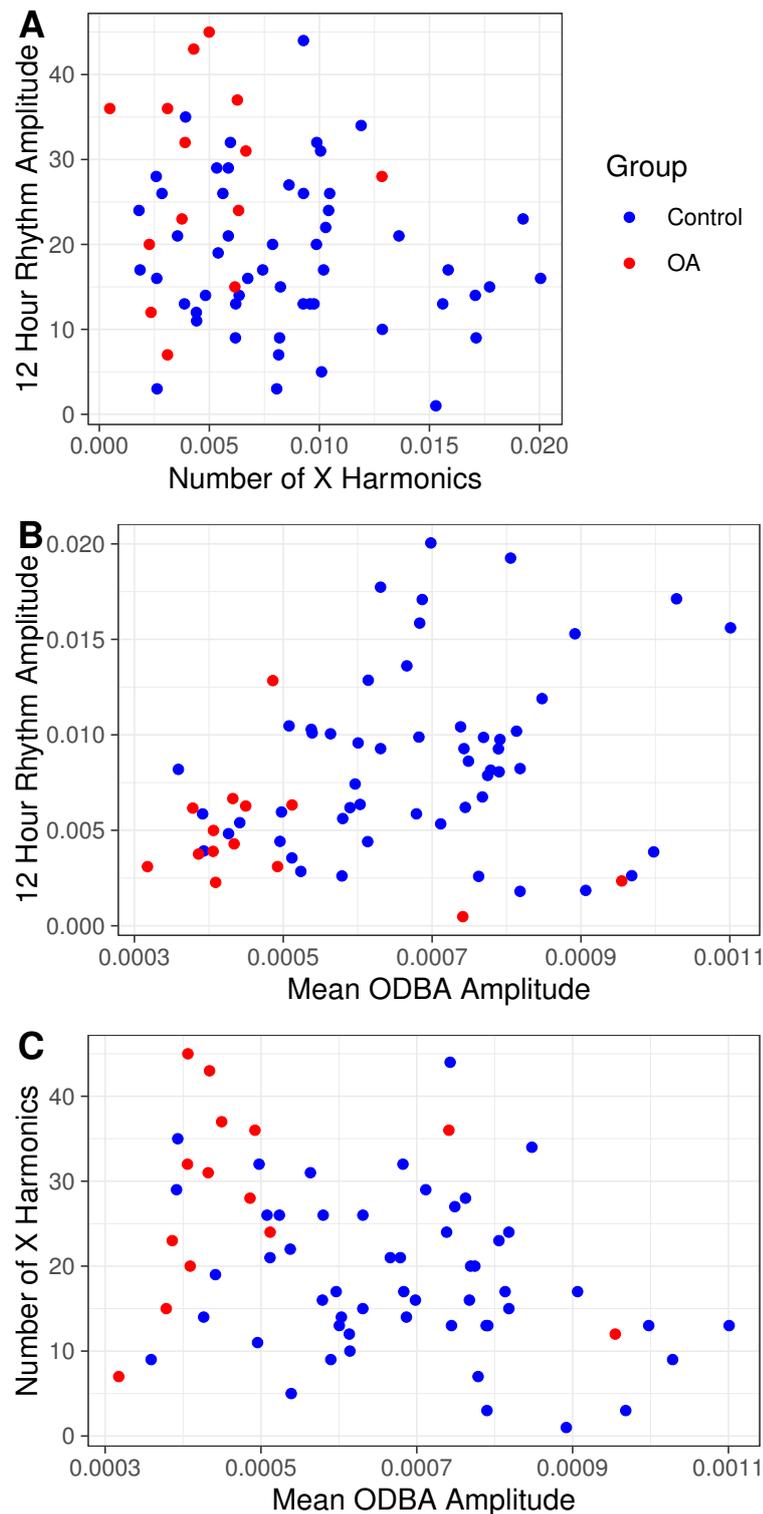


Figure 5.4: A selection of plots providing an indication of class separability when using the selected features as predictors of group. A) The amplitude of the 12 hour period ultradian rhythm against the number of identified X axis harmonics, B) the amplitude of the 12 hour period ultradian rhythm against the Amplitude of the mean ODBA, C) the number of identified X axis harmonics against the Amplitude of the mean ODBA.

5.4.1 Exploration of Variance in the Osteoarthritic Condition

The three remaining features were included in univariate linear regression models with the LOAD score as the output variable (Figure 5.5). The amplitude of the 12 hour period ultradian rhythm was significantly related to the LOAD scores of the dogs ($F(1,62) = 6.79$, $p < 0.05$), but the model struggled to explain the full extent of the variance of LOAD with an R^2 of 0.1. Similarly the number of significant X axis harmonics was indicative of LOAD score ($F(1,62) = 4.83$, $p < 0.05$), but again indicated poor model fit by presenting a low R^2 value of 0.07. The amplitude of the mean ODBA signal was a more effective indicator ($F(1,62) = 21$, $p < 0.001$), with an R^2 value of 0.25.

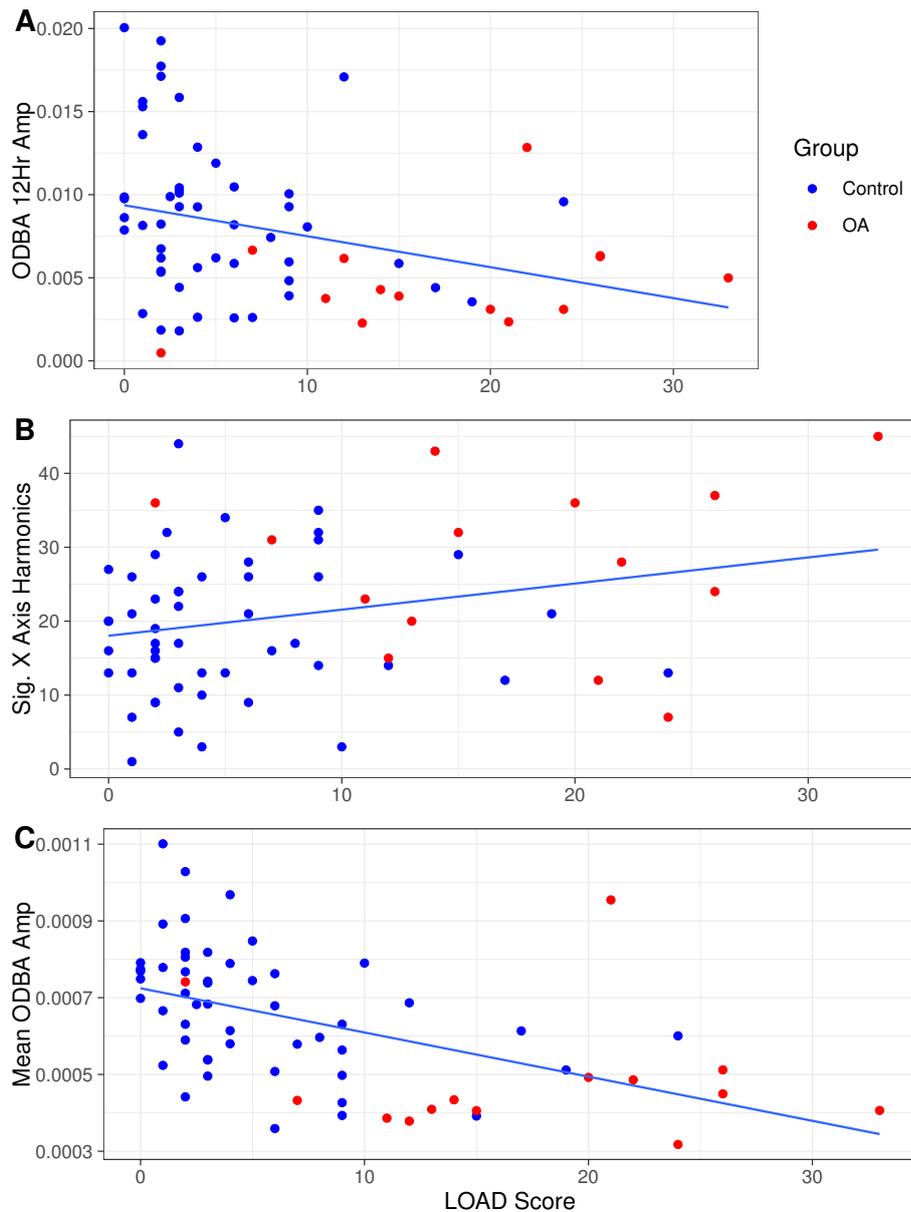


Figure 5.5: Regression of LOAD score against A) the amplitude of the 12 hour ultradian rhythm of the mean ODBA signal, B) the number of significant X axis harmonics, C) the amplitude of the mean ODBA signal

5.4.2 Age Relatedness

The correlation of age with the three selected rhythmic features was examined (Figure 5.6). The amplitude of the 12 hour ultradian cycle within the ODBA was found to have a small negative correlation with age ($r = -0.24$) and was marginally insignificant ($p = 0.057$). The number of significant harmonics of the X axis was found to correlate positively with Age ($r = 0.39$) and was significant ($p = 0.001$). Of the three the most related to age was the negative correlation of the amplitude of the mean ODBA signal ($r = -0.60$, $p < 0.001$).

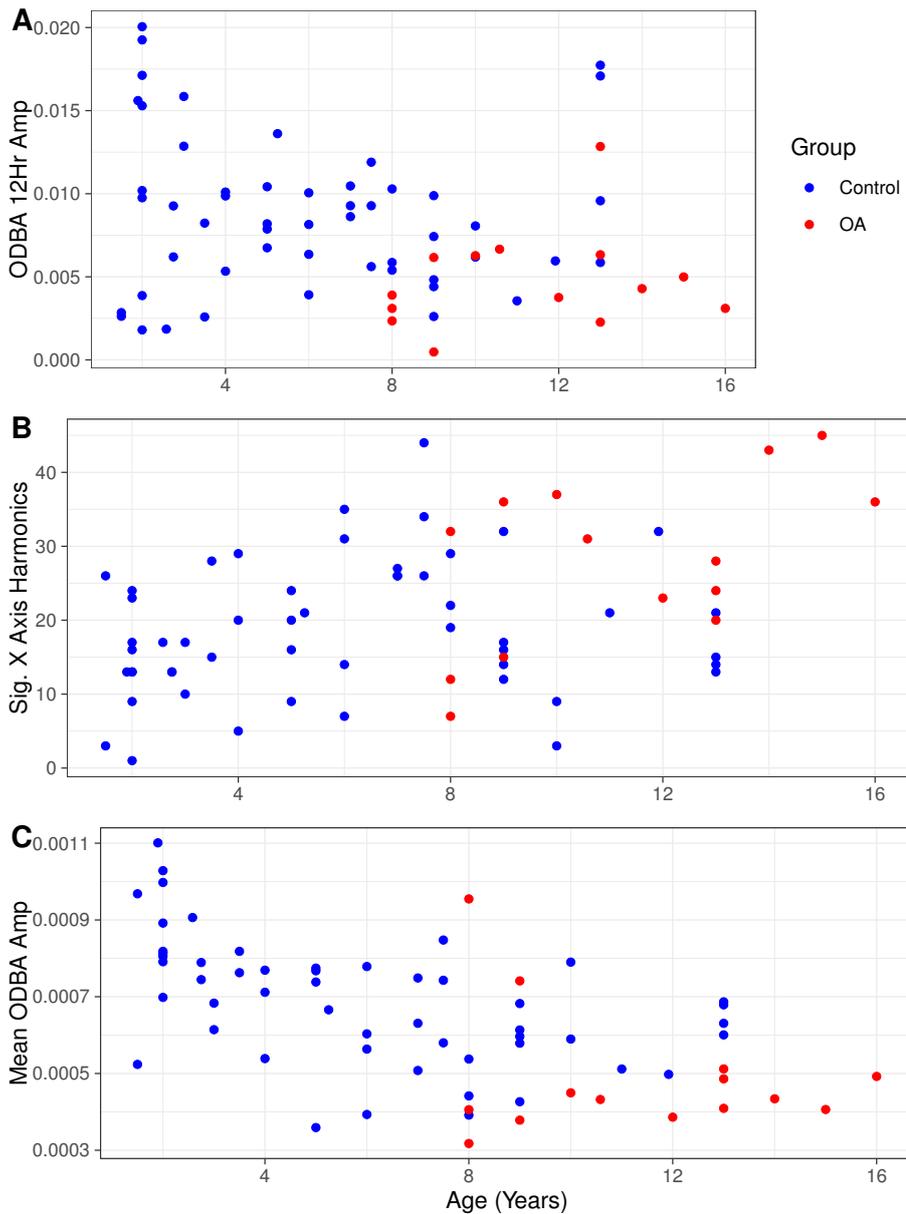


Figure 5.6: The relationships between dog age and A) the amplitude of the 12 hour ultradian rhythm of the mean ODBA signal, B) the number of significant X axis harmonics, C) the amplitude of the mean ODBA signal

5.5 Discussion

The assessment of the rhythmic aspects of acceleration has been shown to have potential for the future discretization of osteoarthritic dogs from healthy controls and the work here serves as a robust foundation to guide further development of these methods both in terms of osteoarthritis and other conditions or welfare issues.

The phase shift of the M5 period from a start time of 09:00 in the control group to 14:00 in the osteoarthritic appears present only when a mean 24 hour signal is taken per group. The mixed model featuring the mean calculated M5 period per dog was non-significant showing no relation between such changes and osteoarthritic health. The L5 reported in the group level assessment remained the same between control and osteoarthritic dogs suggesting daily sleep periods remained consistent with those expected of healthy dogs (Adams & Johnson, 1993).

Comparisons of the Fourier components identified as significant across the mean group ODBA representations show minimal differences. For both groups the circadian period is the strongest observed with 4 additional ultradian rhythms identified. 3 of these are consistent between groups but the control group exhibits a significant 12 hour period rhythm whereas the osteoarthritic group instead shows a significant 6 hour rhythm. It is not possible to ascribe cause to these ultradian rhythms without additional data but hypotheses can be made for future testing particularly as the majority of ultradian rhythms relate to physiological processes. The 2 periods reported, in both groups, that possess 24 minute periods are so short as to be difficult to place and may be an artefact generated through the processing methodologies. However, the canine sleep-wake cycle is reported as possessing a mean period of 21 minutes which may also cause such a rhythm to appear within the data (Adams & Johnson, 1993). The longer period ultradian rhythms could be representative of behaviours that occur more intermittently throughout the day or whose occurrence is restrained by the husbandry of the animal. Piccione *et al.* (2014) outlines various factors that can dramatically impact the activity patterns of domestic dogs. The home environment (Siwak *et al.*, 2002), feeding times (Zanghi *et al.*, 2013), and owner engagement (Dow *et al.*, 2009; Piccione *et al.*, 2014) have all been previously reported to significantly affect the activity levels of dogs. It is possible that one of these behaviours, or others, is more intermittently engaged in by the osteoarthritic group causing the presence of the related Fourier component at a 6 hour, rather than 12 hour, period. The collection of further data, such as video for visual behaviour verification, owner activity and diaries for control of husbandry patterns, and EEG for the assessment of sleep-wake cycles, could help to explain each of these rhythms and why the groups differ.

The majority of features calculated from the Fourier spectra were shown to be unrelated to the presence or absence of osteoarthritis. As can be seen in Figure 5.2 the

spectra of the mean group ODBA values are similar with visible differences occurring predominately in the amplitude of components. This shift in Fourier amplitude also appears to be consistent across individuals as shown in the results of the mixed models where the mean amplitude of the Fourier spectra of the ODBA, and the amplitude of the 12 hour period were both found to be significantly related to osteoarthritic diagnosis. The odds ratios, of 0.205 and 0.249 respectively, suggesting a decrease in amplitude is indicative of an increased probability of osteoarthritis. Interestingly the two are not highly correlated ($r = 0.30$) which shows the decrease in the 12 hour period component is distinct from the overall decrease of amplitudes across the spectrum. As the decrease in overall amplitude is illustrative of an overall decrease in activity levels, as found to be indicative of an osteoarthritis diagnosis in Chapter 2 and is known to be indicative of the disease in humans (Berenbaum & Meng, 2016), the decreased strength of the 12 hour rhythm could be a useful indicator even when activity is constrained by husbandry or other circumstances (i.e post-operative recovery) and should be investigated further with particular attempts to identify the behavioural cause of the rhythm itself.

The number of significant harmonics identified within the X axis was also found to be significant and does not correlate with the ODBA amplitude features. This suggests that summarisation of acceleration to the ODBA has excluded information held within the X axis that could assist in the identification of osteoarthritis. The X axis relates to the medial-lateral acceleration of the device and with an odds ratio of 2.621 an increasing number of significant harmonics can drastically increase the probability of the presence of osteoarthritis. As can be seen in Table 5.4 this feature is highly positively correlated with the number of identified significant harmonics in the Z axis ($r = 0.70$), and negatively with the entropy of both the X and Z axes ($r = -0.92$ and $r = -0.75$ respectively). This suggests that whatever movement or behaviour is related to this increased number of harmonics involves similar acceleration within the devices Z axis, which is aligned with the dogs dorsal-ventral plane. Simultaneously the entropy of the Fourier spectra of the X and Z axes are decreasing, and therefore becoming more deterministic. One potential explanation for this could be the presence of a consistent gait or movement impairment caused by the osteoarthritis or related pain or that behaviours involving movement in these directions are more fractured and activities and behaviours are engaged in for shorter periods before resting. Additionally, as covered in Chapters 3 and 4 both the X and Z axes static components can be representative of postural state, for example whether the individual is standing or recumbent on their side. The use of ODBA which removes this static component, and thus discounts posture in favour of the non-gravitational acceleration, would have caused the rhythms of postures, and the transitions between them, to be excluded from this analysis. It could be that these features are reflecting more fragmented periods spent in different pos-

tures. Further research into the specific frequencies that have been identified along with supplementary data sources, such as video, would allow for the identification of causal behaviours.

The potential use of the rhythmic features generated here in, not just the detection of changes due to osteoarthritis, but of changes due to the effect of the condition on the quality of life of the dog also shows some promise but, as shown by the relationship of mean ODBA signal amplitude with LOAD, only when monitoring the overall distinction between peaks and troughs of activity over a day. The effective fit of the model was shown to be poor using this single predictor and such a feature is likely closely related to similar measures of activity level taken from the time domain and without need of Fourier Transformation.

As in Chapter 2 an initial exploration of the effects of age upon potential models was explored. These correlation results support the further investigation of the effect of age on both the number of significant X axis harmonics and the amplitude of the mean ODBA signal. Additionally, the potential for the 12 hour harmonic amplitude to be a predictor of LOAD while also being robust to age differences should be explored.

Future work should build on the findings outlined here to further refine rhythm based methodologies within the unique context of domestic animal attached, collar-mounted accelerometers. It may be that the methodology used caused the obfuscation of rhythmic components of interest, for example by taking the mean day per dog. A summarisation such as the mean may be effected by the existence of outliers or high levels of intra-week activity variability. Similarly intra-group variation requires further assessment as the summarised 24 hour period of each group formed an informative foundation for the individual-level analyses. Humans express a number of distinct chronotypes and this also appears to be true to some extent in domestic dogs (Little *et al.*, 2016; Nauha *et al.*, 2020; Randler *et al.*, 2018; Refinetti *et al.*, 2016). The current study also assumed all dogs within the assigned groups would share consistent patterns of disruption, which may be too strong an assumption and should be considered when interpreting results and developing future methodologies. The further assessment of LOAD scores for individual dogs may help to disentangle such a phenomena as would further information as to the specific nature of the condition within each individual. It is also possible the chronotype of the owner has a high level of influence on the exhibited rhythms of the dog. This would also be true of ultradian rhythms, which from the results shown here appear to be more indicative of osteoarthritis than the circadian rhythm, and which would be impacted by intra-week schedule changes such as the weekend or owner work patterns. For example, Zanghi *et al.* (2013) found that alteration of feeding schedules would drastically effect the rest-activity rhythms of dogs.

Further investigation of the dataset with unsupervised tools such as clustering algorithms may help to reveal groupings of dogs that could account for variation in chro-

notypes or other unidentified source of intra-group variation. Another strategy that may improve the ability of the methodology to discern between groups is the application of a transformation, such as the logarithm, to the Fourier amplitude values, which often differ over orders of magnitude. This process could reveal additional rhythmic components hidden within the low level noise. Finally constraint of the assessed epoch could help disentangle ultradian rhythms of interest. As the circadian rhythm itself has been found to be unchanged in period length between the groups the reduction of the processed acceleration to only the nocturnal periods could reveal rhythmic components of sleep disturbance which would be consistent with owner reports of night time restlessness in osteoarthritic dogs (Knazovicky *et al.*, 2015). In the specific case of nocturnal rhythms such a change to the sampling protocol could also help to reduce owner impact by removing daytime periods where contact is more common. Additional factors of canine health and welfare should also be investigated once methods are further refined to ensure features are indicating osteoarthritic health and not co-morbidities or other correlated features such as age (Siwak *et al.*, 2002).

5.6 Conclusion

The detection of the presence of osteoarthritis from a single collar-mounted, tri-axial accelerometer using a frequency-based methodology focussed on the extraction of the circadian and ultradian rhythms of behaviours shows promise. From this initial exploration the mean amplitude of the ODBA Fourier components, the number of identified significant Fourier components within the X axis, and the amplitude of the 12 hour period ODBA Fourier component are each shown to be related to the dogs osteoarthritic state. This supports the hypothesis that such measures could, as in Humans, be used in dogs to monitor for welfare change, such as the development of osteoarthritis. However, initial exploration of the known collinear feature of age and its effect suggests that these features need further refinement and perhaps reformulation to account for such confounds. Specific data concerning the extent and management of the condition along with other idiosyncrasies would be informative. The LOAD score assessment of the variance of condition effect on quality of life as perceived by owners appears to show promise for future work focussing on welfare as a measure of disease impact, or probability of condition presence based on linear changes in lifestyle rather than a binary assignment of diagnosis.

The circadian period of dogs was not found to differ between groups, with the ultradian rhythms appearing to be more indicative. Increased ultradian fragmentation may be present in the X and Z axes of acceleration as there are correlating changes in the entropy of the 2 axes and both feature an increased number of detected significant Fourier components. Further development of these methods alongside the

collection of supplemental observational data would assist in the assignment of causal relationships between identified components of interest and behavioural patterns. The use of frequency related features and the Fourier transform at these circadian and ultradian scales may allow the creation of rhythmic fingerprints of conditions, such as osteoarthritis, or welfare states to supplement other identification methods and, in much the same way it is used in commercial human gait and behaviour recognition algorithms, improve overall classification results (Bao & Intille, 2004; Foerster & Fahrenberg, 2000).

Chapter 6

Discussion

This thesis investigated the potential of using a single, collar-mounted, tri-axial accelerometer to characterise canine welfare, using osteoarthritis as a model negative state. The focus of this research was on establishing the breadth of potential applications of these devices, drawing from human and animal research, to enhance the ability of both owners and vets to recognise subtle changes in behaviour or physical function that could be indicative of underlying welfare issues and could, with further development, allow earlier intervention. The initial groundwork put forward in the thesis was broadly successful in this task and as a result it appears that the monitoring of a reduction in physical function would be possible with further work to address areas where classification ability was lacking.

This discussion will consider the thesis as a whole and the interplay of the methods examined within. The limitations and intricacies common to multiple chapters will be assessed further.

6.1 Overview of Findings

Across the thesis the potential of accelerometers to understand the bio-mechanistic and behavioural differences of osteoarthritic dogs, when compared to healthy controls, has been established. However, success has been limited in performance with each revealed indicator of osteoarthritic health, or of behavioural states requiring further work to refine. A combination of these methods with such refinements however could prove effective and informative by accounting for behavioural and biomechanical changes at multiple levels and resolutions. Such methods could also be applied in different welfare compromising conditions to attempt to characterise each independently.

Where this work struggles is the distinction between negative physical and negative cognitive welfare. The current methodologies do not account for the potential differences in the symptoms of these two facets of poor welfare.

Chapter 2 showed the ability of existing measures of activity entropy and overall activity to differentiate between osteoarthritic and control dogs when summarised to a 7 day period. However, when applying this to daily summaries of activity the variabil-

ity of day to day data presented a masking effect that obscured such difference. The success of ODBA and its entropy reveals an effect of osteoarthritic grouping on the overall energy expenditure of individuals and the increased stochasticity of expenditure patterns. Due to this relationship with energy expenditure models using these two features could also be indicative of other activity impacting conditions, *e.g.* the monitoring of obesity in dogs (Morrison *et al.*, 2013). The mean ODBA is potentially more susceptible to influence from husbandry practices, such as the frequency of walks dictated by owners. It may therefore be that entropy is a more robust indicator and could be indicative of sporadic stopping for pain avoidance or other cognitive welfare related reasons. Further study would be required to investigate this further.

The findings of Chapter 3 indicated that a postural transition recognition algorithm, using a collar-mounted tri-axial accelerometer, is not reliable using the methods outlined within this thesis. Within the specific case of osteoarthritis this is disappointing where such a methodology would allow the replication of similar outcome measures assessed with human literature (Podsiadlo & Richardson, 1991). This would have been a good measure of physical welfare but may perhaps have been specific to the osteoarthritic case, or only those conditions causing significant biomechanical impacts.

The detection of postural states and locomotion assessed in Chapter 4 would allow the longitudinal monitoring of basic behavioural patterns in free-ranging domestic dogs. This limited number of states would serve as an important foundational step to further behavioural classifications informed by the behavioural repertoire available to dogs within each posture, or when locomoting. This method would provide an alternative observation method for the behavioural assessment of welfare as changes to behaviour can be indicative of changes to welfare state. Therefore, this method could be applied to other cases of compromised welfare to investigate the behaviours and patterns exhibited. The longitudinal nature of the device also potentially allows for the monitoring of behaviour during negative welfare state onset and of recovery using within dog comparisons. In contrast to the physical nature of osteoarthritis the tool could also be applied when cognitive welfare is compromised, such as during periods of chronic stress, where behaviours are also known to deviate from the norm (Beerda *et al.*, 1997). The performance of the method requires further development however the potential collection of long-term data could provide sufficient indication of 'welfare degradation to offset issues of classification accuracy.

The final exploration, within Chapter 5, examined the rhythmic nature of acceleration and its applicability to osteoarthritic detection. Differences reported were subtle between groups. As in Chapter 4 the potential for longitudinal collection of behaviour and activity reveals the exciting prospect of the collection of data difficult to obtain with traditional or alternative methods. For example, sleep behaviours are often difficult to collect in a dogs natural environment due to the requirement of set sleeping positions,

if using vision based methods, or the potential impact of human observers and time required, when using traditional methods. Changes to circadian and ultradian rhythms of behaviour were anticipated to be indicative of welfare disruption. However, differences between groups were subtle. This could be due to the same inter-day variability seen in Chapter 2 and is perhaps indicative of the daily difference in severity reported by dog owners (Belshaw, 2017; Belshaw *et al.*, 2016; Belshaw & Yeates, 2018). However, this could also be due to the influence of owner behaviour, such as the night time routines effect on circadian rhythm, or the heterogeneity of the osteoarthritic sample itself, with potential differences in pain management and severity. The difference detected in circadian amplitude is likely a reflection of the same reduced energy expenditure observed in Chapter 2. Similarly the indication that ultradian rhythms are more indicative of osteoarthritic group, and by extension potentially of welfare state, lends further support to the suggestion made regarding Chapter 4 that individual rhythms of behaviour could be more sensitive and robust measures which could be investigated further in the future. Across the chapters are common limitations and extensions which are explored in the following sections.

6.2 Attachment Method

The unique context offered by domestic dogs provided a number of obstacles to the project. The use of a single, collar-mounted device is not ideal for the monitoring of canine behaviour and movement for several reasons. Most prominent of these was that it was believed to be more agreeable to owners, in terms of their perception of their dogs comfort, and would not require additional habituation as the majority of domestic dogs are already accustomed to wearing a collar.

Collar attachment is particularly problematic for Chapters 3 and 4 which attempt to replicate automated postural recognition used in the assessment of gait impairment in humans, and livestock. For humans devices are typically positioned at the hip or on the thigh. Here changes from lying to standing or sitting to standing and vice versa involve distinct orientation transitions that are not replicated in the vertical movement exhibited by devices placed at the neck of a dog. Similarly measures of rest and lameness in livestock, particularly cattle, are taken from devices attached to the leg or ankle, again where changes would be most apparent (Martiskainen *et al.*, 2009). As a result of this compromise features of interest required reassessment and the task of classification was likely more difficult due to the lack of distinct orientation changes and separation from the area of interest of the dog. The classifier was no longer attempting to classify features relating to the direct results of impaired movement but instead was assessing the effects of the condition at a holistic level. However, this shifted focus allows the methodologies discussed throughout the thesis, particularly of Chapters 2,

3 and 4, to concentrate on the behavioural rather than the bio-mechanistic which could be more meaningful to owners and allow discussion of symptoms and intervention to be couched in more immediately recognisable terms by veterinarians.

The initial problem with a collar attachment method is the ability of the sensor to rotate around the neck of the dog, an issue that was highlighted in Chapter 2 when considering measures that were robust to such changes, and was addressed in Chapters 3 and 4 with attempts to rectify such an eventuality with a rotational correction algorithm (provided in Appendix D). However, Chapter 4 reported a reduction in the classification performance as a result of the application of the algorithm. This, coupled with a lack of perceived rotation during video recording, resulted in the correction for an eventuality that was present only in theory, and not observed.

One potential solution suggested was the addition of rotational sensors, such as a gyroscope. This would be useful, and has shown much success when used both in human and non-domestic free-ranging animal remote sensing, but would further increase the price of the device, the battery draw, and the memory requirements, which could each impact the size, deployment time or the price (Wilson *et al.*, 2008). Within research environments such changes may not be problematic, particularly as the inclusion of a gyroscope would be unlikely to significantly impact device size or weight. However, this thesis situates its findings within the domestic context. Accessibility, both in price and in usability, should be considered throughout the development of such techniques. The current deployment of devices allows a set it and forget it approach by owners that would be preferable to repeated interventions by owners which would increase the perceived burden and introduce further points of failure.

An alternative approach would be the exclusive use of rotationally robust features, in particular the use of VeDBA in lieu of the ODBA, however measures derived from the vector of acceleration have been shown in Chapters 2, 3, and 4 to have reduced classification ability due to their exclusion during feature selection methods. This is despite showing a strong correlation with the ODBA when compared directly, and suggests there is some obscuration of information, possibly the rotational component that has been corrected for, that is useful in discerning between the groups when data are summarised in such a way (Qasem *et al.*, 2012).

The placement of the device on the collar may not be optimal for the detection of the activities discussed throughout the thesis. Placement of the device at the neck introduces increased influence of head orientation upon the static acceleration component and of head movements upon the dynamic component.

Features of Chapter 2 relying on the dynamic acceleration component could be less susceptible to corruption, as brief movements of the head are limited in scope and should not produce high dynamic acceleration readings. Similarly, such an attachment position would be unlikely to significantly affect the ODBA derived measures of Chapter

5 as it once again concerns itself with overall dynamic activity.

The slower movements of the head, relating more to a shift of orientation and therefore a change in the static component, could be a potential influence on features calculated in all methodologies depending on their inclusion of the static component of acceleration. Chapter 2 lists several orientation and static component related features which could be affected by frequent adjustments of head position. Similarly, Chapters 3 and 4, are both dependant on the interpretation of the static component of acceleration to identify postural transitions or states and increased potential variation within a postural state would reduce classification ability. Alongside features relating to the rhythmicity of the ODBA summarisation Chapter 5 includes rhythmic features of the 3 axes which have not been filtered to remove the static or dynamic components.

However, despite this potential influence each chapter has produced indications that such static-inclusive features are related to osteoarthritic grouping. For example, in Chapter 5 rhythmic features of both the X and Z axes were significant. Both of these were unfiltered representations of the axes and as such included both dynamic and static components. It could be that rhythm of the head position static component is more indicative of osteoarthritic state than that of the overall posture, as detected from the neck. Head orientation would likely be highly correlated with rest, alertness, feeding from a bowl, and other such behaviours where deviation from normality could indicate issues.

The relationship between head orientation and the calculated features should be further investigated with an aim to disentangle the influence of such a distinct movement on the ability of calculated features to classify or represent behaviours of interest. Comparisons against devices situated on the back or underbelly of the dog, where head movement would be less influential, would be sufficient particularly when observing behaviours that require distinct shifts in head position. If found to be distinct then adaptation of methods should be pursued to account for such orientation changes. However, this should be alongside the investigation and addition of a head orientation algorithm. This is due to the ability of head position identification through acceleration to be indicative of behaviours such as feeding, or resting (Ladha & Hoffman, 2018; Moreau *et al.*, 2009; Watanabe *et al.*, 2005).

6.3 The DogBox Sample

The DogBox sample, consisting of the 7-day acceleration profiles of $N = 85$ dogs, contained a wide assortment of breeds, ages, living environments and husbandry practices (Appendix A). When the entire sample was used, in Chapters 2 and 5, each of these factors presents an additional source of variation that was controlled for through the inclusion of a random effect term accounting for the individual within generated models

or in the use of the LOAD score to assess the perceived effect of the condition on quality of life, and therefore accounting for aggravating or mitigating factors. Additionally, the osteoarthritic group was under-represented with only 14 dogs confirmed to suffer from the disease by veterinary diagnosis. This under-representation of the focal class could be problematic where classification is attempted due to the increased probability of misclassification due to the accuracy paradox (Valverde-Albacete & Peláez-Moreno, 2014). To combat this, alternative measures of classification ability were used that are more robust to such occurrences and that produce more conservative results. However, the proportion of healthy to osteoarthritic dogs within the DogBox sample is approximately consistent with estimates of prevalence estimated within the UK and USA (Anderson *et al.*, 2018). As such the prior probability of belonging to either of the two classes is consistent and may in fact improve the accuracy of models by reflecting the true population.

The postural state and transition recognition chapters, 3 and 4, used a reduced sample of 20 healthy control dogs, from a separate and pre-existing dataset, that were all the same breed and subject to identical husbandry conditions. This allowed more precise claims to be made in terms of the predictive ability of the models for the prediction of postural state. However, the limited sample created difficulties in the cross validation of predictive models and a large degree of class imbalance between rare and common behaviours. The sample itself also lacks generalisability and if further developed will need to be adapted to account for the wide range of breed variation within dogs.

The DogBox sample lacked detail in terms of the daily schedules of the dogs. The addition of owner diary methods or the additional instrumentation of owners would help to untangle the relationship between the features calculated through the thesis and husbandry, owner schedules and related confounds. The additional completion of further owner-perception related questionnaires would also be of use to interpret any changes in husbandry that may result from the perceived welfare change of the dogs due to their positive osteoarthritis, for example reduced activity may be originating from the owner providing fewer exercise opportunities than from the dogs unwillingness to engage (Belshaw, 2017). Such additional tasks however would be difficult to collect across such a large sample due to the increased level of engagement required by owners and increased likelihood of non-continuous compliance to the protocol.

6.4 Osteoarthritis as a model of reduced welfare

The core assumption of this thesis is that osteoarthritis is equivalent to a reduction in welfare and by extension that the automated detection of the presence of the disease would be equivalent to an automated detection of negative welfare state. It is widely re-

ported that human sufferers of osteoarthritis report chronic pain, and stiffness of movement at levels which impact daily function (Duncan *et al.*, 2007; Duncan *et al.*, 2011). As the mechanisms of osteoarthritis in both species are equivalent, and dogs are often used as models for human osteoarthritic research, it is assumed that this is also true in canine sufferers. As such any measurement of pain within dogs with osteoarthritis is assessed and measured second hand rather than by the sufferer themselves. In this study indications of osteoarthritic health were provided by owners and their personal perception of their dogs quality of life and level of pain. Such phenomena as the caregiver placebo effect, where owners are found to be poor reporters of acute pain, could have resulted in the under-diagnosis of osteoarthritis within the sample. Such an eventuality was attempted to be controlled for using quality of life questionnaires, such as the Liverpool Osteoarthritis in Dogs questionnaire, which more extensively probed the daily function of dogs but remained owner dependent (Belshaw, 2017; Belshaw *et al.*, 2016; Belshaw & Yeates, 2018).

Additionally, accelerometers measure the actual acceleration of the device, which is then used to draw conclusions regarding the biomechanics and behaviour of the animal it is attached to. As such it must be questioned whether the methodologies described are detecting osteoarthritis or welfare at all and just what the generated features coincide with. The features generated across the thesis may each be indicative of different symptoms of osteoarthritis but fall along two interrelated axes, pain avoidance behaviours or reduced welfare behaviours. Physical adjustment to gait, the reduction of the number of postural transitions, or changes to the patterns and degree of activity or postural states, may each be indicative of pain avoidance. If this is the case then similar results would be found when investigating other chronic pain conditions, particularly those affecting similar areas of the body. Simultaneously behavioural shifts in terms of rhythmicity, frequency, or amplitude of activity or posture could be indicative of the stress and impact on affective state of the condition, including effect of the inability of the animal to engage in natural behaviours due to pain. This presents 3 possibilities as to what the features calculated throughout the thesis are detecting: the occurrence and development of specific symptoms of osteoarthritis, evidence of increased pain avoidance behaviour in osteoarthritic dogs, or the changes of behaviour resulting from a reduction in welfare irrespective of physical health. To disentangle these additional work would be required on a population suffering from a differing chronic pain causing condition, and with a non-physical negative welfare causing condition, such as a cognitive impairment (e.g. epilepsy) or a known negative experience or stress causing situation (e.g. anxiety).

Across both Chapters 2 and 5 the collinear nature of age with the condition is explored with relation to highlighted features and the LOAD score. The nature of ageing within dogs is such that both activity and health are reduced as it advances. As such,

there was no clear feature that was found to be entirely disentangled from age as a confounding factor. This is a difficult problem to fully solve and should be considered with future work. Further work into the biomechanical and/or the behavioural information that could be obtained via remote sensing methods, such as the postural methods illustrated in Chapters 3 and 4, may be a more robust methodology to pursue when attempting to account for this as they focus more on alterations to how and when behaviours are being performed that may be easier to attribute to the condition itself.

6.5 The role of Technology in animal welfare

This thesis attempts to avoid sophisticated machine learning methodologies for the classification of the osteoarthritic grouping throughout. This is a conscious attempt to preserve the interpretability of results in terms of the features and their potential meanings. The implementation of more sophisticated methods that have shown a high degree of performance in other work, such as random forest methods, would create a black-box environment (Rudin, 2019). This would result in a disconnect between the classification of dogs and specific causes. At this foundational stage and in an effort to better understand the changes that occur and how they relate to the osteoarthritic state it was decided that the preservation of features and the use of simplistic statistical models was of more importance than the performance attained. Despite this sacrifice the methods proved promising and have presented several features and insights that are worth pursuing further throughout the thesis.

The black box issue is present throughout the application of machine learning in animal behaviour and work is frequently published with outstanding performance but with little interpretation possible beyond the proclamation of relationships or the ability to classify. In some situations this is desirable and further interpretation of cause and effect is superfluous. The extensive use of such methods for behaviour detection in wild free-ranging species or in commercial agricultural settings are examples of this. The presence or absence of the behaviour is of the highest importance as this can then be used to gain further insight into the focal animals behavioural ecology. Within work such as this thesis, where the interest lies not in the presence of the disease, but on its wider effects and repercussions, a more transparent methodology that can convey meaning and cause more easily is more desirable. By aiming for transparency of methods it is hoped that if applied such methods could provide insight that better allows the treatment or alleviation of the condition, pain or negative welfare state. This is in direct contrast to the ever-increasing commercial solutions which operate as true black boxes and often obscure the methodologies used for their output due to concerns regarding competition.

6.6 Future work

The most prominent avenue to pursue further is the incorporation of the methods with a naïve dataset to truly assess performance. This could be coupled with the combination of the more successful methods described in Chapters 2, 4, and 5 into a multi-faceted classifier consisting of either a hierarchical or vote based system of classification which could help to address any shortfalls exhibited when used individually. The issues relating to collar rotation and positioning should be addressed as described and such improvements would allow for the addition of head orientation analysis that, if coupled with postural identification, could further inform the ability of the sensors to infer behavioural state. There also exists the potential to combine the procedures of chapters 4 and 5 to analyse the rhythmicity of postural state and the time spent within each state.

The application of knowledge regarding the interrelationship of features and osteoarthritic health gained here could also be used to inform the future use of less transparent machine learning methodologies. Such methods may allow the achievement of levels of performance akin to those seen in similar literature however care should be taken to minimise their black-box nature. The information required in management and care of dogs the vast, varying domestic should be accessible and unobscured where possible. Data condensation through methods such as PCA or MFA could also present improvement to performance and should be assessed, however care should be taken in the derivation of real-world meaning from the presented components.

The potential effect of owner schedule and husbandry should be accounted for through simultaneous instrumentation of owners and/or the completion of diary questionnaires. Additional condition groupings would be a valuable addition to the samples to better understand the level of specificity the classification procedures are achieving (i.e. whether the classifier is identifying osteoarthritis, pain, or negative welfare).

This work focussed on the negative welfare of dogs affected by the osteoarthritic disease. As the field of animal welfare research increasingly involves the assessment of positive welfare state and behaviour it would be of interest to also attempt to devise a classifier that detects it. Combined with the broad methods described in this thesis it could be that rather than an increase in pain-related behaviour the grouping of the dogs is instead seen in a decrease of positive welfare related behaviours, such as play. This is already claimed to be detected by commercial solutions and in research. However, play is often conflated with periods of vigorous activity that do not infer valence and although this may be broadly true the diversity of play behaviour exhibited by dogs presents the distinct possibility that such simplifications may be insufficient and should be assessed.

6.7 Conclusions

This thesis highlights the potential of accelerometer based methods, particularly those relating to behavioural rhythmicity and energy expenditure, for the remote monitoring of conditions relating to a negative welfare state, specifically osteoarthritis. Methodologies that have been deployed in past works across multiple species and within a variety of fields have been adapted to the specific context of domestic dog welfare. A promising foundation for further work and refinement has been laid with initial explorations of established acceleration measures revealing the difficulties in their adaptation to this welfare context. It may be that instead higher order interpretations of acceleration measures are required to attribute behaviour and biomechanical abnormalities to individuals. The hierarchical model of Chapter 4 poses a framework for the development of meaningful remote sensing ethograms that, from a ground up perspective, could provide a modular and effective method of observation, similar to that established in wild animal behavioural ecology. I would suggest that such a system could then form the basis which provides the additional context allowing meaning to be found from the methodologies proposed throughout the rest of the thesis. The attempt to recognise transitional movements was largely unsuccessful but an automated "stand up and go" test or similar could be implemented at probable postural change windows if such a hierarchical system were further developed. Similarly, the varied measures of overall activity of Chapter 2 may be more widely informative when taken during specific periods or behaviours. Measures of circadian and ultradian rhythms may also benefit from details relating to behaviour, such as nocturnal orientation changes and disturbance recognition, or rhythmic nature of specific behaviours and events that may be difficult to recognise when interpreting raw acceleration. The overall focus of this thesis has centred around interpretation by the end user and the transparency of revealed relationships and potential confounds. This is core to the useability of such methods going forward and their practicality within real-world home and veterinary environments. A system that more closely resembles existing behavioural and health monitoring methods, by representing acceleration in behavioural and biomechanical terms, would also greatly help in further promoting this useability and ensuring that such tools are accessible within research, veterinary care and the domestic environment to ensure the maximal improvement to canine welfare.

References

- Adams, G. J. & Johnson, K. (1993). Sleep-wake cycles and other night-time behaviours of the domestic dog *canis familiaris*. *Applied Animal Behaviour Science*, *36*(2-3), 233–248.
- Aminian, K. & Najafi, B. (2004). Capturing human motion using body-fixed sensors: Outdoor measurement and clinical applications. *Computer animation and virtual worlds*, *15*(2), 79–94.
- Anan, M., Shinkoda, K., Suzuki, K., Yagi, M., Ibara, T. & Kito, N. (2015). Do patients with knee osteoarthritis perform sit-to-stand motion efficiently? *Gait & Posture*, *41*(2), 488–492.
- Anderson, K. L., O'Neill, D. G., Brodbelt, D. C., Church, D. B., Meeson, R. L., Sargan, D., Summers, J. F., Zulch, H. & Collins, L. M. (2018). Prevalence, duration and risk factors for appendicular osteoarthritis in a uk dog population under primary veterinary care. *Scientific reports*, *8*(1), 1–12.
- Anderson, K. L., Zulch, H., O'Neill, D. G., Meeson, R. L. & Collins, L. M. (2020). Risk factors for canine osteoarthritis and its predisposing arthropathies: A systematic review. *Frontiers in Veterinary Science*, *7*, 220.
- Aschoff, J. & Pohl, H. (1978). Phase relations between a circadian rhythm and its zeitgeber within the range of entrainment. *Naturwissenschaften*, *65*(2), 80–84.
- Asher, L., Buckland, E. L., Phylactopoulos, C. I., Whiting, M. C., Abeyesinghe, S. M. & Wathes, C. M. (2011). Estimation of the number and demographics of companion dogs in the uk. *BMC Veterinary Research*, *7*(1), 74.
- Balzani, A. & Hanlon, A. (2020). Factors that influence farmers' views on farm animal welfare: A semi-systematic review and thematic analysis. *Animals*, *10*(9), 1524.
- Bao, L. & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. *International conference on pervasive computing*, 1–17.
- Barker, Z., Diosdado, J. V., Codling, E., Bell, N., Hodges, H., Croft, D. & Amory, J. (2018). Use of novel sensors combining local positioning and acceleration to measure feeding behavior differences associated with lameness in dairy cattle. *Journal of Dairy Science*, *101*(7), 6310–6321.
- Barnard, S., Calderara, S., Pistocchi, S., Cucchiara, R., Podaliri-Vulpiani, M., Messori, S. & Ferri, N. (2016). Quick, accurate, smart: 3d computer vision technology helps assessing confined animals' behaviour. *PLoS ONE*, *11*, 7.

- Barrey, E., Hermelin, M., Vaudelin, J., Poirel, D. & Valette, J. (1994). Utilisation of an accelerometer device in equine gait analysis. *Equine Veterinary Journal*, 26(S17), 7–12.
- Barthelemy, 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(11), 788–796.
- Beerda, B., Schilder, M., Van Hooff, J., De Vries, H. & Mol, J. (2000). Behavioural and hormonal indicators of enduring environmental stress in dogs. *Animal Welfare-potters Bar-*, 9(1), 49–62.
- Beerda, B., Schilder, M. B., van Hooff, J. A. & de Vries, H. W. (1997). Manifestations of chronic and acute stress in dogs. *Applied Animal Behaviour Science*, 52(3-4), 307–319.
- Belshaw, Z. (2017). *Decision making and welfare assessment in canine osteoarthritis* (Doctoral dissertation). University of Nottingham.
- Belshaw, Z., Asher, L. & Dean, R. S. (2016). Systematic review of outcome measures reported in clinical canine osteoarthritis research. *Veterinary Surgery*, 45(4), 480–487.
- Belshaw, Z. & Yeates, J. (2018). Assessment of quality of life and chronic pain in dogs. *The Veterinary Journal*, 239, 59–64.
- Ben Sassi, N., Averós, X. & Estevez, I. (2016). Technology and poultry welfare. *Animals*, 6(10), 62.
- Benaissa, S., Tuytens, F. A., Plets, D., de Pessemier, T., Trogh, J., Tanghe, E., Martens, L., Vandaele, L., Van Nuffel, A., Joseph, W. et al. (2017). On the use of on-cow accelerometers for the classification of behaviours in dairy barns. *Research in veterinary science*.
- Berenbaum, F. & Meng, Q.-J. (2016). The brain–joint axis in osteoarthritis: Nerves, circadian clocks and beyond. *Nature Reviews Rheumatology*, 12(9), 508–516.
- Berrar, D. & Flach, P. (2011). Caveats and pitfalls of roc analysis in clinical microarray research (and how to avoid them). *Briefings in Bioinformatics*, 13(1), 83–97.
- Bidder, O. R., Qasem, L. A. & Wilson, R. P. (2012). On higher ground: How well can dynamic body acceleration determine speed in variable terrain? *Plos One*, 7(11), 1–7.
- Bódizs, R., Kis, A., Gácsi, M. & Topál, J. (2020). Sleep in the dog: Comparative, behavioral and translational relevance. *Current Opinion in Behavioral Sciences*, 33, 25–33.
- Boissy, A., Manteuffel, G., Jensen, M. B., Moe, R. O., Spruijt, B., Keeling, L. J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I. & Aubert, A. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiology and Behavior*, 92(3), 375–397.

- Botreau, R., Bonde, M., Butterworth, A., Perny, P., Bracke, M., Capdeville, J. & Veissier, I. (2007). Aggregation of measures to produce an overall assessment of animal welfare. part 1: A review of existing methods. *Animal*, 1(8), 1179–1187.
- Brambell, F. W. R. et al. (1965). Report of the technical committee to enquire into the welfare of animals kept under intensive livestock husbandry systems.
- Brondeel, R., Rahimipour Anaraki, J., KhataeiPour, S. & Fuller, D. (2019). *Activity-counts: Generate actilife counts* [R package version 0.1.2].
- Broom, D. M. (1986). Indicators of poor welfare. *British veterinary journal*, 142(6), 524–526.
- Broom, D. M. (2010). Animal welfare: An aspect of care, sustainability, and food quality required by the public. *Journal of veterinary medical education*, 37(1), 83–88.
- Broom, D. M. (2011). A history of animal welfare science. *Acta biotheoretica*, 59(2), 121–137.
- Broom, D. M. (2016). Sentience, animal welfare and sustainable livestock production. *Indigenous*, eds KS Reddy, RMV Prasad and KA Rao.
- Brown, D. D., Kays, R., Wikelski, M., Wilson, R. & Klimley, A. (2013). Observing the unwatchable through acceleration logging of animal behavior. *Animal Biotelemetry*, 1(1), 20.
- Brugarolas, R., Roberts, D., Sherman, B. & Bozkurt, A. (2013). Machine learning based posture estimation for a wireless canine machine interface. *2013 IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems*, 10–12.
- Burghardt, T. & Calic, J. (2006). Analysing animal behaviour in wildlife videos using face detection and tracking. *IEE Proceedings - Vision, Image, and Signal Processing*, 153(3), 305.
- Butterworth, A., Knowles, T., Whittington, P., Matthews, L., Rogers, A. & Bagshaw, C. (2007). Validation of broiler chicken gait scoring training in thailand, brazil and new zealand. *Animal Welfare*, 16(2), 177–179.
- Cabin, R. J. & Mitchell, R. J. (2000). To bonferroni or not to bonferroni: When and how are the questions. *Bulletin of the Ecological Society of America*, 81(3), 246–248.
- Caeiro, C. C., Burrows, A. M. & Waller, B. M. (2017). Development and application of catfacs: Are human cat adopters influenced by cat facial expressions? *Applied Animal Behaviour Science*, 189, 66–78.
- Caeiro, C. C., Waller, B. M., Zimmermann, E., Burrows, A. M. & Davila- Ross, M. (2013). Orangfacs: A muscle-based facial movement coding system for orangutans (pongo spp.) *International Journal of Primatology*, 34(1), 115–129.

- Cagnacci, F., Boitani, L., Powell, R. A. & Boyce, M. S. (2010). Animal ecology meets gps-based radiotelemetry: A perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550), 2157–2162.
- Casper, R. M. (2009). Guidelines for the instrumentation of wild birds and mammals. *Animal Behaviour*, 78(6), 1477–1483.
- Cattell, R. & Jaspers, J. (1967). A general plasmode for factor analytic exercises and research (no. 30-10-52). *MBR Monographs*, 3, 67.
- Chakravarty, P., Cozzi, G., Ozgul, A. & Aminian, K. (2019). A novel biomechanical approach for animal behaviour recognition using accelerometers. *Methods in Ecology and Evolution*.
- Cheeta, S., Ruyt, G., van Proosdij, J. & Willner, P. (1997). Changes in sleep architecture following chronic mild stress. *Biological psychiatry*, 41(4), 419–427.
- Chen, Y. Y. & Hong, C. Y. (2019). Taiwanese consumers willingness to pay for broiler welfare improvement. *Animals*, 9(5), 231.
- Church, J. S., Cook, N. J. & Schaefer, A. L. (2009). Recent applications of infrared thermography for animal welfare and veterinary research: Everything from chicks to elephants. *InfraMation*, (0), 53.
- Clark, C., Gruffydd-Jones, T. & Murray, J. (2012). Number of cats and dogs in uk welfare organisations. *Veterinary Record*.
- Coleman, G. (2007). Public perceptions of animal pain and animal welfare. *Proceedings of the Australian Animal Welfare Strategy Science Summit on Pain and Pain Management*, 27(1), 8.
- Conzemius, M. G. & Evans, R. B. (2012). Caregiver placebo effect for dogs with lameness from osteoarthritis. *Journal of the American Veterinary Medical Association*, 241(10), 1314–1319.
- Cooke, S. J., Hinch, S. G., Wikelski, M., Andrews, R. D., Kuchel, L. J., Wolcott, T. G. & Butler, P. J. (2004). Biotelemetry: A mechanistic approach to ecology. *Trends in ecology & evolution*, 19(6), 334–343.
- Cornou, C., Lundbye-Christensen, S. & Kristensen, A. R. (2011). Modelling and monitoring sows' activity types in farrowing house using acceleration data. *Computers and electronics in agriculture*, 76(2), 316–324.
- Crawford, C. & Koopman, P. (1973). A note on horn's test for the number of factors in factor analysis. *Multivariate Behavioral Research*, 8(1), 117–125.
- Dawkins, M. S. (1990). From an animal's point of view: Motivation, fitness, and animal welfare. *Behavioral and brain sciences*, 13(1), 1–9.
- Dawkins, M. S. (2004). Using behaviour to assess animal welfare. *Animal welfare*, 13(1), 3–7.

- Dawkins, M. S. (2015). Animal welfare and the paradox of animal consciousness. *Advances in the study of behavior* (pp. 5–38). Elsevier.
- Dawkins, M. S. (2017). Animal welfare with and without consciousness. *Journal of Zoology*, 301(1), 1–10.
- Dawkins, M. S., Cook, P. A., Whittingham, M. J., Mansell, K. A. & Harper, A. E. (2003). What makes free-range broiler chickens range? in situ measurement of habitat preference. *Animal behaviour*, 66(1), 151–160.
- Dawkins, M. S., Lee, H.-j., Waitt, C. D. & Roberts, S. J. (2009). Optical flow patterns in broiler chicken flocks as automated measures of behaviour and gait. *Applied Animal Behaviour Science*, 119(3-4), 203–209.
- Degeling, C. & Johnson, J. (2015). Citizens, consumers and animals: What role do experts assign to public values in establishing animal welfare standards? *Journal of Agricultural and Environmental Ethics*, 28(5), 961–976.
- de Passillé, A., Jensen, M., Chapinal, N. & Rushen, J. (2010). Technical note: Use of accelerometers to describe gait patterns in dairy calves. *Journal of Dairy Science*, 93(7), 3287–3293.
- Derksen, S. & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45(2), 265–282.
- Descovich, K. (2017). *Facial expression: An under-utilised tool for the assessment of welfare in mammals*. Altex.
- Developers, S. (2014). *signal: Signal processing*.
- Diosdado, J. A. V., Barker, Z. E., Hodges, H. R., Amory, J. R., Croft, D. P., Bell, N. J. & Codling, E. A. (2015). Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry*, 3(1).
- Donnell, J. R., Frisbie, D. D., King, M. R., Goodrich, L. R. & Haussler, K. K. (2015). Comparison of subjective lameness evaluation, force platforms and an inertial-sensor system to identify mild lameness in an equine osteoarthritis model. *Veterinary Journal*, 206(2), 136–142.
- 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(4), 444–448.
- Drewes, A. (1999). Pain and sleep disturbances with special reference to fibromyalgia and rheumatoid arthritis. *Rheumatology*, 38(11), 1035–1038.
- Duncan, I. J. (2006). The changing concept of animal sentience. *Applied Animal Behaviour Science*, 100(1-2), 11–19.

- Duncan, R., Peat, G., Thomas, E., Hay, E., McCall, I. & Croft, P. (2007). Symptoms and radiographic osteoarthritis: Not as discordant as they are made out to be? *Annals of the Rheumatic Diseases*, 66(1), 86–91.
- Duncan, R. P., Leddy, A. L. & Earhart, G. M. (2011). Five times sit-to-stand test performance in parkinson's disease. *Archives of physical medicine and rehabilitation*, 92(9), 1431–1436.
- Ekman, P. & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1), 56–75.
- Ellis, R. G., Rankin, J. W. & Hutchinson, J. R. (2018). Limb kinematics, kinetics and muscle dynamics during the sit-to-stand transition in greyhounds. *Frontiers in Bioengineering and Biotechnology*, 6, 162.
- Engel, B., Bruin, G., Andre, G. & Buist, W. (2003). Assessment of observer performance in a subjective scoring system: Visual classification of the gait of cows. *The Journal of Agricultural Science*, 140(3), 317.
- Fast, R., Schütt, T., Toft, N., Møller, A. & Berendt, M. (2013). An observational study with long-term follow-up of canine cognitive dysfunction: Clinical characteristics, survival, and risk factors. *Journal of Veterinary Internal Medicine*, 27(4), 822–829.
- Fehlmann, G., O'Riain, M. J., Hopkins, P. W., O'Sullivan, J., Holton, M. D., Shepard, E. L. & King, A. J. (2017). Identification of behaviours from accelerometer data in a wild social primate. *Animal Biotelemetry*, 5(1), 6.
- Feltenstein, M. W., Ford, N. G., Freeman, K. B. & Sufka, K. J. (2002). Dissociation of stress behaviors in the chick social-separation-stress procedure. *Physiology & behavior*, 75(5), 675–679.
- Fenwick, N., Duffus, S. E. & Griffin, G. (2014). Pain management for animals used in science: Views of scientists and veterinarians in canada. *Animals*, 4(3), 494–514.
- Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 125(796), 54–59.
- Foerster, F. & Fahrenberg, J. (2000). Motion pattern and posture: Correctly assessed by calibrated accelerometers. *Behavior research methods, instruments, & computers*, 32(3), 450–457.
- Fong, D. T. P. & Chan, Y. Y. (2010). The use of wearable inertial motion sensors in human lower limb biomechanics studies: A systematic review. *Sensors (Switzerland)*, 10(12), 11556–11565.

- Fourati, H., Manamanni, N., Afilal, L. & Handrich, Y. (2011). Posture and body acceleration tracking by inertial and magnetic sensing: Application in behavioral analysis of free-ranging animals [Biomedical signal processing(Extended selected papers from the 7th IFAC Symposium on Modelling and Control in Biomedical Systems(MCBMS'09))]. *Biomedical Signal Processing and Control*, 6(1), 94–104.
- Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N. J. & Pedrotti, L. (2010). Resolving issues of imprecise and habitat-biased locations in ecological analyses using gps telemetry data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550), 2187–2200.
- Ganea, R., Paraschiv-Ionescu, A. & Aminian, K. (2012). Detection and classification of postural transitions in real-world conditions. *IEEE transactions on neural systems and rehabilitation engineering*, 20(5), 688–696.
- Gao, L., Campbell, H. A., Bidder, O. R. & Hunter, J. (2013). A web-based semantic tagging and activity recognition system for species' accelerometry data. *Ecological Informatics*, 13, 47–56.
- Geiger, M., Gendron, K., Willmitzer, F. & Sánchez-Villagra, M. R. (2016). Unaltered sequence of dental, skeletal, and sexual maturity in domestic dogs compared to the wolf. *Zoological letters*, 2(1), 1–8.
- Gillette, R. L. & Angle, T. C. (2008). Recent developments in canine locomotor analysis: A review. *Veterinary Journal*, 178(2), 165–176.
- Gleerup, K. B., Andersen, P. H., Munksgaard, L. & Forkman, B. (2015a). Pain evaluation in dairy cattle. *Applied Animal Behaviour Science*, 171(1), 25–32.
- Gleerup, K. B., Forkman, B., Lindegaard, C. & Andersen, P. H. (2015b). An equine pain face. *Veterinary Anaesthesia and Analgesia*, 42(1), 103–114.
- Gleiss, A. C., Wilson, R. P. & Shepard, E. L. (2011). Making overall dynamic body acceleration work: On the theory of acceleration as a proxy for energy expenditure. *Methods in Ecology and Evolution*, 2(1), 23–33.
- González, L., Tolkamp, B., Coffey, M., Ferret, A. & Kyriazakis, I. (2008). Changes in feeding behavior as possible indicators for the automatic monitoring of health disorders in dairy cows. *Journal of Dairy Science*, 91(3), 1017–1028.
- Greenhough, B. & Roe, E. (2018). Exploring the role of animal technologists in implementing the 3rs: An ethnographic investigation of the uk university sector. *Science, Technology, & Human Values*, 43(4), 694–722.
- Grégoire, J., Bergeron, R., D'Allaire, S., Meunier-Salauín, M.-C. & Devillers, N. (2013). Assessment of lameness in sows using gait, footprints, postural behaviour and foot lesion analysis. *Animal*, 7, 07.

- Grundy, E., Jones, M. W., Laramée, R. S., Wilson, R. P. & Shepard, E. L. (2009). Visualisation of sensor data from animal movement. *Computer Graphics Forum*, 28(3), 815–822.
- Hacker, C. E., Horback, K. M. & Miller, L. J. (2015). Gps technology as a proxy tool for determining relationships in social animals: An example with african elephants. *Applied Animal Behaviour Science*, 163, 175–182.
- Hall, M. A. (1999). *Correlation-based feature selection for machine learning* (tech. rep.).
- Hamilton, N. A., Catley, D. & Karlson, C. (2007). Sleep and the affective response to stress and pain. *Health Psychology*, 26(3), 288.
- Hammerla, N. Y., Kirkham, R., Andras, P. & Ploetz, T. (2013). On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution. *Proceedings of the 2013 International Symposium on Wearable Computers*, 65–68.
- Handcock, R. N., Swain, D. L., Bishop-Hurley, G. J., Patison, K. P., Wark, T., Valencia, P., Corke, P. & O'Neill, C. J. (2009). Monitoring animal behaviour and environmental interactions using wireless sensor networks, gps collars and satellite remote sensing. *Sensors*, 9(5), 3586–3603.
- Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. [Pmid: 7063747]. *Radiology*, 143(1), 29–36.
- Hannan, M. T., Felson, D. T. & Pincus, T. (2000). Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. *The Journal of rheumatology*, 27(6), 1513–1517.
- Hansen, B. D., Lascelles, B. D., Keene, B. W., Adams, A. K. & Thomson, A. E. (2007). Evaluation of an accelerometer for at-home monitoring of spontaneous activity in dogs. *Am. J. Vet. Res.*, 68(5), 468–475.
- Harrison, R. (2013). *Animal machines*. Cabi.
- Hart, B. L. (1988). Biological basis of the behavior of sick animals. *Neuroscience & Biobehavioral Reviews*, 12(2), 123–137.
- Hausser, J. & Strimmer, K. (2014). *Entropy: Estimation of entropy, mutual information and related quantities* [R package version 1.2.1].
- Heglund, N. C. & Taylor, C. R. (1988). Speed, stride frequency and energy cost per stride: How do they change with body size and gait? *Journal of Experimental Biology*, 138(1), 301–318.
- Held, S. D. E. & Spinka, M. (2011). Animal play and animal welfare. *Animal Behaviour*, 81(5), 891–899.
- Henchoz, Y., Bastardot, F., Guessous, I., Theler, J.-M., Dudler, J., Vollenweider, P. & So, A. (2012). Physical activity and energy expenditure in rheumatoid arthritis patients and matched controls. *Rheumatology*, 51(8), 1500–1507.

- Hickey, A., Galna, B., Mathers, J. C., Rochester, L. & Godfrey, A. (2016). A multi-resolution investigation for postural transition detection and quantification using a single wearable. *Gait & Posture*, *49*, 411–417.
- 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.
- Hokkanen, A. H., Hlanninen, L., Tiusanen, J. & Pastell, M. (2011). Predicting sleep and lying time of calves with a support vector machine classifier using accelerometer data. *Applied Animal Behaviour Science*, *134*(1-2), 10–15.
- Holdgate, M. R., Meehan, C. L., Hogan, J. N., Miller, L. J., Soltis, J., Andrews, J. & Shepherdson, D. J. (2016). Walking behavior of zoo elephants: Associations between gps-measured daily walking distances and environmental factors, social factors, and welfare indicators. *PLoS ONE*, *11*, 7.
- Holman, A., Thompson, J., Routly, J. E., Cameron, J., Jones, D. N., White, G., D., S., F., R. & Dobson, H. (2011). Comparison of oestrus detection methods in dairy cattle. *Veterinary Record*, *169*(2), 47–47.
- Howe, C. A., Staudenmayer, J. W. & Freedson, P. S. (2009). Accelerometer prediction of energy expenditure: Vector magnitude versus vertical axis. *Med Sci Sports Exerc*, *41*(12), 2199–206.
- Huzzey, J., Veira, D., Weary, D. & von Keyserlingk, M. (2007). Prepartum behavior and dry matter intake identify dairy cows at risk for metritis. *Journal of Dairy Science*, *90*(7), 3220–3233.
- Iwasaki, M. & Noguchi, Y. (2016). Hiding true emotions: Micro-expressions in eyes retrospectively concealed by mouth movements. *Scientific Reports*, *6*, 22049.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). Linear regression. *An introduction to statistical learning: With applications in r* (pp. 59–126). Springer New York.
- Janssen, M., Busch, C., Rödiger, M. & Hamm, U. (2016). Motives of consumers following a vegan diet and their attitudes towards animal agriculture. *Appetite*, *105*, 643–651.
- Janssen, M., Bussmann, J. B., Horemans, H. L. & Stam, H. J. (2005). Analysis and decomposition of accelerometric signals of trunk and thigh obtained during the sit-to-stand movement. *Medical and Biological Engineering and Computing*, *43*(2), 265–272.
- Jukan, A., Masip-Bruin, X. & Amla, N. (2017). Smart computing and sensing technologies for animal welfare: A systematic review. *ACM Computing Surveys (CSUR)*, *50*(1), 1–27.

- Junior, C. F. C., Pederiva, C. N., Bose, R. C., Garcia, V. A., Lino-de-Oliveira, C. & Marino-Neto, J. (2012). ETHOWATCHER: Validation of a tool for behavioral and video-tracking analysis in laboratory animals. *Computers in Biology and Medicine*, 42(2), 257–264.
- Kanyongo, G. Y. (2005). Determining the correct number of components to extract from a principal components analysis: A monte carlo study of the accuracy of the scree plot. *Journal of Modern Applied Statistical Methods*, 4(1), 13.
- Keeling, L. (1995). Spacing behaviour and an ethological approach to assessing optimum space allocations for groups of laying hens. *Applied Animal Behaviour Science*, 44(2-4), 171–186.
- Kestin, S., Knowles, T., Tinch, A. & Gregory, N. (1992). Prevalence of leg weakness in broiler chickens and its relationship with genotype. *Veterinary Record*, 131(9), 190–194.
- Kis, A., Gergely, A., Galambos, Á., Abdai, J., Gombos, F., Bódizs, R. & Topál, J. (2017). Sleep macrostructure is modulated by positive and negative social experience in adult pet dogs. *Proceedings of the Royal Society B: Biological Sciences*, 284(1865), 20171883.
- 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.
- Knight, S. & Barnett, L. (2008). Justifying attitudes toward animal use: A qualitative study of people's views and beliefs. *Anthrozoös*, 21(1), 31–42.
- Korte, J., Hoehn, T. & Siegmund, R. (2004). Actigraphic recordings of activity-rest rhythms of neonates born by different delivery modes. *Chronobiology international*, 21(1), 95–106.
- Kozak, M., Tobalske, B., Springthorpe, D., Szkotnicki, B. & Harlander- Matauschek, A. (2016). Development of physical activity levels in laying hens in three-dimensional aviaries. *Applied Animal Behaviour Science*, 185, 66–72.
- Kramer, K. & Kinter, L. B. (2003). Evaluation and applications of radiotelemetry in small laboratory animals. *Physiological genomics*, 13(3), 197–205.
- Krause, J., Krause, S., Arlinghaus, R., Psorakis, I., Roberts, S. & Rutz, C. (2013). Reality mining of animal social systems. *Trends in Ecology and Evolution*, 28(9), 541–551.
- Kuhn., M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C. & Hunt., T. (2019). *Caret: Classification and regression training* [R package version 6.0-84].

- Kumpulainen, P., Valldeoriola, A., Somppi, S., Törnqvist, H., Väättäjä, H., Majaranta, P., Surakka, V., Vainio, O., Kujala, M. V., Gizatdinova, Y. & Vehkaoja, A. (2018). Dog activity classification with movement sensor placed on the collar. *Proceedings of the Fifth International Conference on Animal-Computer Interaction*, 4:1–4:6.
- 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), 107.
- 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*, 415–418.
- Ladha, C. & Hoffman, C. L. (2018). A combined approach to predicting rest in dogs using accelerometers. *Sensors*, 18(8), 2649.
- Ladha, C., O’Sullivan, J., Belshaw, Z. & Asher, L. (2017). Gaitkeeper: A system for measuring canine gait. *Sensors*, 17(2).
- Lane, D. M., Hill, S. A., Huntingford, J. L., Lafuente, P., Wall, R. & Jones, K. A. (2015). Effectiveness of slow motion video compared to real time video in improving the accuracy and consistency of subjective gait analysis in dogs. *Open Veterinary Journal*, 5(2), 158–165.
- Langford, F. & Cockram, M. (2010). Is sleep in animals affected by prior waking experiences? *Animal Welfare*, 19(3), 215–222.
- Lê, S., Josse, J. & Husson, F. (2008). FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, 25(1), 1–18.
- Leach, M. C., Coulter, C. A., Richardson, C. A. & Flecknell, P. A. (2011). Are we looking in the wrong place? implications for behavioural-based pain assessment in rabbits (*Oryctolagus cuniculi*) and beyond? *PloS one*, 6(3), e13347.
- Leighty, K. A., Soltis, J. & Savage, A. (2010). Gps assessment of the use of exhibit space and resources by african elephants (*Loxodonta africana*). *Zoo Biology*, 29(2), 210–220.
- Li, X., Kane, M., Zhang, Y., Sun, W., Song, Y., Dong, S., Lin, Q., Zhu, Q., Jiang, F. & Zhao, H. (2020). Circadian rhythm analysis using wearable device data: A novel penalized machine learning approach. *bioRxiv*.
- Liew, A. W.-C., Law, N.-F., Cao, X.-Q. & Yan, H. (2009). Statistical power of fisher test for the detection of short periodic gene expression profiles. *Pattern Recognition*, 42(4), 549–556.
- Little, D., Johnson, S., Hash, J., Olson, S. A., Estes, B. T., Moutos, F. T., Lascelles, B. D. X. & Guilak, F. (2016). Functional outcome measures in a surgical model of hip osteoarthritis in dogs. *Journal of experimental orthopaedics*, 3(1), 1–16.

- Lund, J. D. & Jørgensen, M. C. (1999). Behaviour patterns and time course of activity in dogs with separation problems. *Applied Animal Behaviour Science*, 63(3), 219–236.
- Mancuso, C. A., Rincon, M., Sayles, W. & Paget, S. A. (2007). Comparison of energy expenditure from lifestyle physical activities between patients with rheumatoid arthritis and healthy controls. *Arthritis & Rheumatism*, 57(4), 672–678.
- Mansfield, E. R. & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician*, 36(3a), 158–160.
- Mao, K. Z. (2004). Orthogonal forward selection and backward elimination algorithms for feature subset selection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1), 629–634.
- Marais, J., Le Roux, S. P., Wolhuter, R. & Niesler, T. (2014). Automatic classification of sheep behaviour using 3-axis accelerometer data. *Proceedings of the twenty-fifth annual symposium of the Pattern Recognition Association of South Africa (PRASA)*, 97–102.
- Martin, K. W., Olsen, A. M., Duncan, C. G. & Duerr, F. M. (2017). The method of attachment influences accelerometer-based activity data in dogs. *BMC Veterinary Research*, 13(1), 48.
- Martiskainen, P., Järvinen, M., Skön, J. P., Tiirikainen, J., Kolehmainen, M. & Mononen, J. (2009). Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied Animal Behaviour Science*, 119(1-2), 32–38.
- Mathie, M., Celler, B. G., Lovell, N. H. & Coster, A. (2004). Classification of basic daily movements using a triaxial accelerometer. *Medical and Biological Engineering and Computing*, 42(5), 679–687.
- Max Planck Institute for Psycho-linguistics. (2019). *Elan* (Version 5.7). Nijmegen, The Netherlands.
- McClune, D. W., Marks, N. J., Wilson, R. P., Houghton, J. D., Montgomery, I. W., McGowan, N. E., Gormley, E. & Scantlebury, M. (2014). Tri-axial accelerometers quantify behaviour in the eurasian badger (*meles meles*): Towards an automated interpretation of field data. *Animal Biotelemetry*, 2(1), 5.
- Mccracken, M. J., Kramer, J., Keegan, K. G., Lopes, M., Wilson, D. A., Reed, S. K., Lacarrubba, A. & Rasch, M. (2012). Comparison of an inertial sensor system of lameness quantification with subjective lameness evaluation. *Equine Veterinary Journal*, 44(6), 652–656.
- McLennan, K. M., Rebelo, C. J. B., Corke, M. J., Holmes, M. A., Leach, M. C. & Constantino-Casas, F. (2016). Development of a facial expression scale using footrot and mastitis as models of pain in sheep. *Applied Animal Behaviour Science*, 176, 19–26.

- Meeson, R. L., Todhunter, R. J., Blunn, G., Nuki, G. & Pitsillides, A. A. (2019). Spontaneous dog osteoarthritis a one medicine vision. *Nature Reviews Rheumatology*, 15(5), 273–287.
- Mellor, D. J. (2016). Updating animal welfare thinking: Moving beyond the “five freedoms” towards “a life worth living”. *Animals*, 6(3).
- Mendes, E. D. M., Carstens, G. E., Tedeschi, L. O., Pinchak, W. E. & Friend, T. H. (2011). Validation of a system for monitoring feeding behavior in beef cattle. *Journal of Animal Science*, 89(9), 2904–2910.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. (2019). *E1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), tu wien* [R package version 1.7-0.1].
- Millor, N., Lecumberri, P., Gomez, M., Martinez-Ramirez, A. & Izquierdo, M. (2014). Kinematic parameters to evaluate functional performance of sit-to-stand and stand-to-sit transitions using motion sensor devices: A systematic review. *IEEE transactions on neural systems and rehabilitation engineering*, 22(5), 926–936.
- Miwa, M., Oishi, K., Nakagawa, Y., Maeno, H., Anzai, H., Kumagai, H., Okano, K., Tobioka, H. & Hirooka, H. (2015). Application of overall dynamic body acceleration as a proxy for estimating the energy expenditure of grazing farm animals: Relationship with heart rate. *PLoS one*, 10(6), e0128042.
- Monath, T. P., Kahn, L. H. & Kaplan, B. (2010). One health perspective. *ILAR Journal*, 51(3), 193–198.
- 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(3-4), 158–170.
- Morris, A. J., Whittingham, M. J., Bradbury, R. B., Wilson, J. D., Kyrkos, A., Buckingham, D. L. & Evans, A. D. (2001). Foraging habitat selection by yellowhammers (*emberiza citrinella*) nesting in agriculturally contrasting regions in lowland england. *Biological Conservation*, 101(2), 197–210.
- 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(11), 570–574.
- Mosaferi, B., Babri, S., Mohaddes, G., Khamnei, S. & Mesgari, M. (2015). Post-weaning environmental enrichment improves BDNF response of adult male rats. *International Journal of Developmental Neuroscience*, 46(1), 108–114.
- Murray, J., Gruffydd-Jones, T., Roberts, M. & Browne, W. (2015). Assessing changes in the uk pet cat and dog populations: Numbers and household ownership. *Veterinary Record*.

- Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M. & Getz, W. M. (2012). Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: General concepts and tools illustrated for griffon vultures. *The Journal of experimental biology*, 215(Pt 6), 986–996.
- Nauha, L., Jurvelin, H., Ala-Mursula, L., Niemelä, M., Jämsä, T., Kangas, M. & Korpelainen, R. (2020). Chronotypes and objectively measured physical activity and sedentary time at midlife. *Scandinavian journal of medicine & science in sports*, 30(10), 1930–1938.
- Office International des Epizooties. (2015). Section 7. animal welfare. terrestrial animal health code.
- Open Lab. (2018). *Open movement gui application* (Version 1.0.0.42). Newcastle Upon Tyne, U.K.
- Ortiz-Tudela, E., Martinez-Nicolas, A., Díaz-Mardomingo, C., García-Herranz, S., Pereda-Pérez, I., Valencia, A., Peraita, H., Venero, C., Madrid, J. A. & Rol, M. A. (2014). The characterization of biological rhythms in mild cognitive impairment. *BioMed Research International*, 2014.
- 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(10), e0163620.
- Parmelee, P. A., Tighe, C. A. & Dautovich, N. D. (2015). Sleep disturbance in osteoarthritis: Linkages with pain, disability, and depressive symptoms. *Arthritis care & research*, 67(3), 358–365.
- 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.
- Pastell, M., Tiusanen, J., Hakojärvi, M. & Hänninen, L. (2009). A wireless accelerometer system with wavelet analysis for assessing lameness in cattle. *Biosystems engineering*, 104(4), 545–551.
- Paul, E. S., Harding, E. J. & Mendl, M. (2005). Measuring emotional processes in animals: The utility of a cognitive approach. *Neuroscience & Biobehavioral Reviews*, 29(3), 469–491.
- Pawlyk, A. C., Morrison, A. R., Ross, R. J. & Brennan, F. X. (2008). Stress-induced changes in sleep in rodents: Models and mechanisms. *Neuroscience and Biobehavioral Reviews*, 32(1), 99–117.
- Percival, D. B. & Walden, A. T. (1993). *Spectral analysis for physical applications*. Cambridge university press.

- Pfau, T., Fiske-Jackson, A. & Rhodin, M. (2016). Quantitative assessment of gait parameters in horses: Useful for aiding clinical decision making? *Equine Veterinary Education*, 28(4), 209–215.
- Piccione, G., Marafioti, S., Giannetto, C., Di Pietro, S., Quartuccio, M. & Fazio, F. (2014). Comparison of daily distribution of rest/activity in companion cats and dogs. *Biological Rhythm Research*, 45(4), 615–623.
- Pickford, C. G., Findlow, A. H., Kerr, A., Banger, M., Clarke-Cornwell, A. M., Hollands, K. L., Quinn, T. & Granat, M. H. (2019). Quantifying sit-to-stand and stand-to-sit transitions in free-living environments using the activpal thigh-worn activity monitor. *Gait & Posture*, 73, 140–146.
- Pillard, P., Gibert, S. & Viguiet, E. (2012). Development of a 3d accelerometric device for gait analysis in dogs. *Computer methods in biomechanics and biomedical engineering*, 15(sup1), 246–249.
- Platz, S., Ahrens, F., Bendel, J., Meyer, H. & Erhard, M. (2008). What happens with cow behavior when replacing concrete slatted floor by rubber coating: A case study. *Journal of Dairy Science*, 91(3), 999–1004.
- Podsiadlo, D. & Richardson, S. (1991). The timed up & go: A test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*, 39(2), 142–148.
- 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.
- Pond, M. & Nuki, G. (1973). Experimentally-induced osteoarthritis in the dog. *Annals of the rheumatic diseases*, 32(4), 387.
- Poole, T. (1997). Happy animals make good science. *Laboratory animals*, 31(2), 116–124.
- Popova, A., Tsvirkun, D., Dolgov, O., Anokhin, K., Alberts, J., Lagereva, E., Custaud, M.-A., Gauquelin-Koch, G., Vinogradova, O. & Andreev-Andrievskiy, A. (2017). Adaptation to a blood pressure telemetry system revealed by measures of activity, agility and operant learning in mice. *Journal of Pharmacological and Toxicological Methods*, 85, 29–37.
- Poursaberi, A., Bahr, C., Pluk, A., Van Nuffel, A. & Berckmans, D. (2010). Real-time automatic lameness detection based on back posture extraction in dairy cattle: Shape analysis of cow with image processing techniques. *Computers and Electronics in Agriculture*, 74(1), 110–119.

- Preisig, D. F., Kulic, L., Krulger, M., Wirth, F., McAfoose, J., Splani, C., Gantenbein, P., Derungs, R., Nitsch, R. M. & Welt, T. (2016). High-speed video gait analysis reveals early and characteristic locomotor phenotypes in mouse models of neurodegenerative movement disorders. *Behavioural Brain Research*, *311*, 340–353.
- 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*(1), 412–416.
- Purohit, R. C. & McCoy, M. D. (1980). Thermography in the diagnosis of inflammatory processes in the horse. *American Journal of Veterinary Research*, *41*(8), 1167–1174.
- Qasem, L., Cardew, A., Wilson, A., Griffiths, I., Halsey, L. G., Shepard, E. L., Gleiss, A. C. & Wilson, R. (2012). Tri-axial dynamic acceleration as a proxy for animal energy expenditure; should we be summing values or calculating the vector? *PloS one*, *7*(2), e31187.
- Quinn, M. M., Keuler, N. S., Lu, Y., Faria, M. L., Muir, P. & Markel, M. D. (2007). Evaluation of agreement between numerical rating scales, visual analogue scoring scales, and force plate gait analysis in dogs. *Veterinary Surgery*, *36*(4), 360–367.
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Raabymagle, P. & Ladewig, J. (2006). Lying behavior in horses in relation to box size. *Journal of Equine Veterinary Science*, *26*(1), 11–17.
- Randler, C., Daz-Morales, J. F. & Jankowski, K. S. (2018). Synchrony in chronotype and social jetlag between dogs and humans across europe. *Time & Society*, *27*(2), 223–238.
- Rault, J.-L., Lawrence, A. & Ralph, C. (2018). Brain-derived neurotrophic factor in serum as an animal welfare indicator of environmental enrichment in pigs. *Domestic Animal Endocrinology*, *65*, 67–70.
- Ravi, N., Dandekar, N., Mysore, P. & Littman, M. L. (2005). Activity recognition from accelerometer data. *Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence - Volume 3*, 1541–1546.
- Reaney, S. J., Zulch, H., Mills, D., Gardner, S. & Collins, L. (2017). Emotional affect and the occurrence of owner reported health problems in the domestic dog. *Applied Animal Behaviour Science*, *196*, 76–83.
- Refinetti, R., Wassmer, T., Basu, P., Cherukalady, R., Pandey, V. K., Singaravel, M., Giannetto, C. & Piccione, G. (2016). Variability of behavioral chronotypes of 16 mammalian species under controlled conditions. *Physiology & behavior*, *161*, 53–59.

- Reid, J., Wiseman-Orr, M. L., Scott, E. M. & Nolan, A. M. (2013). Development, validation and reliability of a web-based questionnaire to measure health-related quality of life in dogs. *Journal of Small Animal Practice*, 54(5), 227–233.
- Reppert, S. M. & Weaver, D. R. (2002). Coordination of circadian timing in mammals. *Nature*, 418(6901), 935–941.
- Rhodin, M., Bergh, A., Gustås, P. & Gómez Álvarez, C. (2017). Inertial sensor-based system for lameness detection in trotting dogs with induced lameness. *The Veterinary Journal*, 222, 54–59.
- Ringgenberg, N., Bergeron, R. & Devillers, N. (2010). Validation of accelerometers to automatically record sow postures and stepping behaviour. *Applied Animal Behaviour Science*, 128(1-4), 37–44.
- Ringgenberg, N., Fröhlich, E. K. F., Harlander-Matauschek, A., Toscano, M. J., Wuírbel, H. & Roth, B. A. (2015). Effects of variation in nest curtain design on pre-laying behaviour of domestic hens. *Applied Animal Behaviour Science*, 170, 34–43.
- Ripley, B. D. (2008). *Pattern recognition and neural networks* (Vol. 11).
- Robert, B., White, B., Renter, D. & Larson, R. (2009). Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. *Computers and Electronics in Agriculture*, 67(1-2), 80–84.
- Rodríguez-Martín, D., Samà, A., Pérez-López, C., Cabestany, J., Català, A. & Rodríguez-Molinero, A. (2015). Posture transition identification on pd patients through a svm-based technique and a single waist-worn accelerometer. *Neurocomputing*, 164, 144–153.
- Rodríguez-Martín, D., Samà, A., Pérez-López, C., Català, A., Cabestany, J. & Rodríguez-Molinero, A. (2013). Svm-based posture identification with a single waist-located triaxial accelerometer. *Expert Systems with Applications*, 40(18), 7203–7211.
- Romanski, P. & Kotthoff, L. (2018). *Fselector: Selecting attributes* [R package version 0.31].
- Rooney, N. J., Gaines, S. A. & Bradshaw, J. W. (2007). Behavioural and glucocorticoid responses of dogs (*canis familiaris*) to kennelling: Investigating mitigation of stress by prior habituation. *Physiology & Behavior*, 92(5), 847–854.
- Ropert-Coudert, Y., Grémillet, D., Kato, A., Ryan, P. G., Naito, Y. & Le Maho, Y. (2004). A fine-scale time budget of cape gannets provides insights into the foraging strategies of coastal seabirds. *Animal Behaviour*, 67(5), 985–992.
- Roubenoff, R., Walsmith, J., Lundgren, N., Snyderman, L., Dolnikowski, G. J. & Roberts, S. (2002). Low physical activity reduces total energy expenditure in women with rheumatoid arthritis: Implications for dietary intake recommendations. *The American Journal of Clinical Nutrition*, 76(4), 774–779.

- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215.
- Rufener, C., Berezowski, J., Sousa, F. M., Abreu, Y., Asher, L. & Toscano, M. J. (2018). Finding hens in a haystack: Consistency of movement patterns within and across individual laying hens maintained in large groups. *Scientific Reports*, 8(1).
- Rushen, J., Chapinal, N., De Passille, A. et al. (2012). Automated monitoring of behavioural-based animal welfare indicators. *Animal Welfare-The UFAW Journal*, 21(3), 339.
- Rutherford, K. (2002). Assessing pain in animals. *Animal Welfare*, 11(1), 31–53.
- Salmeri, K., Bloomberg, M., Scruggs, S. L. & Shille, V. (1991). Gonadectomy in immature dogs: Effects on skeletal, physical, and behavioral development. *Journal of the American Veterinary Medical Association*, 198(7), 1193–1203.
- Sato, K., Mitani, Y., Cameron, M. F., Siniff, D. B. & Naito, Y. (2003). Factors affecting stroking patterns and body angle in diving weddell seals under natural conditions. *Journal of experimental Biology*, 206(9), 1461–1470.
- Scaglia, E., Cannas, S., Minero, M., Frank, D., Bassi, A. & Palestrini, C. (2013). Video analysis of adult dogs when left home alone. *Journal of Veterinary Behavior*, 8(6), 412–417.
- Schoenbaum, I., Kigel, J., Ungar, E. D., Dolev, A. & Henkin, Z. (2017). Spatial and temporal activity of cattle grazing in mediterranean oak woodland. *Applied Animal Behaviour Science*, 187, 45–53.
- Shepard, E. L., Wilson, R. P., Halsey, L. G., Quintana, F., Gómez Laich, A., Gleiss, A. C., Liebsch, N., Myers, A. E. & Norman, B. (2008a). Derivation of body motion via appropriate smoothing of acceleration data. *Aquatic Biology*, 4(3), 235–241.
- Shepard, E. L., Wilson, R. P., Quintana, F., Gómez Laich, A., Liebsch, N., Albareda, D., Halsey, L., Gleiss, A. C., Morgan, D., Myers, A., Newman, C. & Macdonald, D. (2008b). Identification of animal movement patterns using tri-axial accelerometry. *Endangered Species Research*, 10, 47–60.
- Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. (2005). Rocr: Visualizing classifier performance in r. *Bioinformatics*, 21(20), 7881.
- Siwak, C. T., Murphey, H. L., Muggenburg, B. A. & Milgram, N. W. (2002). Age-dependent decline in locomotor activity in dogs is environment specific. *Physiology & behavior*, 75(1-2), 65–70.
- Soltis, J., Wilson, R. P., Douglas-Hamilton, I., Vollrath, F., King, L. E. & Savage, A. (2012). Accelerometers in collars identify behavioral states in captive african elephants *loxodonta africana*. *Endangered Species Research*, 18(3), 255–263.
- Spivey, R. J. & Bishop, C. M. (2013). Interpretation of body-mounted accelerometry in flying animals and estimation of biomechanical power. *Journal of The Royal Society Interface*, 10(87), 20130404.

- Staley, M., Conners, M. G., Hall, K. & Miller, L. J. (2018). Linking stress and immunity: Immunoglobulin a as a non-invasive physiological biomarker in animal welfare studies. *Hormones and Behavior*, *102*, 55–68.
- Stephens, A., O'Donnell, K., Marmot, M. & Wardle, J. (2008). Positive affect, psychological well-being, and good sleep. *Journal of Psychosomatic Research*, *64*(4), 409–415.
- Stewart, M., Webster, J., Verkerk, G., Schaefer, A., Colyn, J. & Stafford, K. (2008). Non-invasive measurement of stress in dairy cows using infrared thermography. *PhD Thesis*, 520–525.
- Sunderam, S., Chernyy, N., Peixoto, N., Mason, J. P., Weinstein, S. L., Schiff, S. J. & Gluckman, B. J. (2007). Improved sleepwake and behavior discrimination using mems accelerometers. *Journal of Neuroscience Methods*, *163*(2), 373–383.
- Tabachnick, B. & Fidell, L. (2001). Using multivariate statistics.(4th edn.) needham heights. *MA: Allyn & Bacon*.
- Takahashi, Y., Ebihara, S., Nakamura, Y. & Takahashi, K. (1978). Temporal distributions of delta wave sleep and rem sleep during recovery sleep after 12-h forced wakefulness in dogs; similarity to human sleep. *Neuroscience letters*, *10*(3), 329–334.
- Thompson, R., Matheson, S. M., Plötz, T., Edwards, S. A. & Kyriazakis, I. (2016). Porcine lie detectors: Automatic quantification of posture state and transitions in sows using inertial sensors. *Computers and electronics in agriculture*, *127*, 521–530.
- Titulaer, M., Blackwell, E. J., Mendl, M. & Casey, R. A. (2013). Cross sectional study comparing behavioural, cognitive and physiological indicators of welfare between short and long term kennelled domestic dogs. *Applied Animal Behaviour Science*, *147*(1-2), 149–158.
- Tomkiewicz, S. M., Fuller, M. R., Kie, J. G. & Bates, K. K. (2010). Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1550), 2163–2176.
- Torres, J. B. & Solberg, V. S. (2001). Role of self-efficacy, stress, social integration, and family support in latino college student persistence and health. *Journal of Vocational Behavior*, *59*, 53–63.
- Tuytens, F., de Graaf, S., Heerkens, J. L., Jacobs, L., Nalon, E., Ott, S., Stadig, L., Van Laer, E. & Ampe, B. (2014). Observer bias in animal behaviour research: Can we believe what we score, if we score what we believe? *Animal Behaviour*, *90*, 273–280.
- Ungar, E. D., Henkin, Z., Gutman, M., Dolev, A., Genizi, A. & Ganskopp, D. (2005). Inference of animal activity from gps collar data on free-ranging cattle. *Rangeland Ecology & Management*, *58*(3), 256–266.

- Urton, G., von Keyserlingk, M. & Weary, D. (2005). Feeding behavior identifies dairy cows at risk for metritis. *Journal of Dairy Science*, *88*(8), 2843–2849.
- Väätäjä, H., Majaranta, P., Isokoski, P., Gizatdinova, Y., Kujala, M. V., Somppi, S., Vehkaoja, A., Vainio, O., Juhlin, O., Ruohonen, M. et al. (2018). Happy dogs and happy owners: Using dog activity monitoring technology in everyday life. *Proceedings of the Fifth International Conference on Animal-Computer Interaction*, 1–12.
- 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.
- Valverde-Albacete, F. J. & Peláez-Moreno, C. (2014). 100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox (M. G. A. Paris, Ed.). *PLoS ONE*, *9*(1), e84217.
- Van Lummel, R., Ainsworth, E., Lindemann, U., Zijlstra, W., Chiari, L., Van Campen, P. & Hausdorff, J. (2013). Automated approach for quantifying the repeated sit-to-stand using one body fixed sensor in young and older adults. *Gait & Posture*, *38*(1), 153–156.
- van Lummel, R., Evers, J., Niessen, M., Beek, P. & van Dieën, J. (2018). Older adults with weaker muscle strength stand up from a sitting position with more dynamic trunk use. *Sensors*, *18*(4), 1235.
- Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth) [Isbn 0-387-95457-0]. Springer.
- Vianna, D. M. L. & Carrive, P. (2005). Changes in cutaneous and body temperature during and after conditioned fear to context in the rat. *European Journal of Neuroscience*, *21*(9), 2505–2512.
- Viazzi, S., Bahr, C., Van Hertem, T., Schlageter-Tello, A., Romanini, C. E. B., Halachmi, I., Lokhorst, C. & Berckmans, D. (2014). Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows. *Computers and Electronics in Agriculture*, *100*, 139–147.
- Vick, S. J., Waller, B. M., Parr, L. A., Pasqualini, M. C. S. & Bard, K. A. (2007). A cross-species comparison of facial morphology and movement in humans and chimpanzees using the facial action coding system (facs). *Journal of Nonverbal Behavior*, *31*(1), 1–20.
- Vij, M., Naik, V. & Gunturi, V. M. V. (2017). Use of ecdf-based features and ensemble of classifiers to accurately detect mobility activities of people using accelerometers. *2017 9th International Conference on Communication Systems and Networks (COMSNETS)*, 39–46.
- Volicer, L., Harper, D. G., Manning, B. C., Goldstein, R. & Satlin, A. (2001). Sundowning and circadian rhythms in alzheimers disease. *American Journal of Psychiatry*, *158*(5), 704–711.

- Walton, M. B., Cowderoy, E., Lascelles, D. & Innes, J. F. (2013). Evaluation of construct and criterion validity for the liverpool osteoarthritis in dogs(load) clinical metrology instrument and comparison to two other instruments. *PLoS One*, 8(3), e58125.
- Wang, S., Yang, J., Chen, N., Chen, X. & Zhang, Q. (2005). Human activity recognition with user-free accelerometers in the sensor networks. *2005 International Conference on Neural Networks and Brain*, 2, 1212–1217.
- 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(1-2), 117–131.
- Wathan, J., Burrows, A. M., Waller, B. M. & McComb, K. (2015). Equifacs: The equine facial action coding system. *PLoS ONE*, 10, 8.
- Webster, J. (2016). Animal welfare: Freedoms, dominions and a life worth living. *Animals*, 6(6), 35.
- Weiss, A., Herman, T., Plotnik, M., Brozgol, M., Giladi, N. & Hausdorff, J. (2011). An instrumented timed up and go: The added value of an accelerometer for identifying fall risk in idiopathic fallers. *Physiological measurement*, 32(12), 2003.
- Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2), 70–73.
- Westgarth, C. & Ladha, C. (2017). Evaluation of an open source method for calculating physical activity in dogs from harness and collar based sensors. *BMC Vet. Res.*, 13(1), 322.
- Whitham, J. & Miller, L. (2016). Using technology to monitor and improve zoo animal welfare. *Animal Welfare*, 25(4), 395–409.
- Whitham, J. C. & Wielebnowski, N. (2013). New directions for zoo animal welfare science. *Applied Animal Behaviour Science*, 147(3-4), 247–260.
- Wilson, R. P., Liebsch, N., Davies, I. M., Quintana, F., Weimerskirch, H., Storch, S., Lucke, K., Siebert, U., Zankl, S., Müller, G., Zimmer, I., Scolaro, A., Campagna, C., Plíotz, J., Bornemann, H., Teilmann, J. & McMahon, C. R. (2007). All at sea with animal tracks; methodological and analytical solutions for the resolution of movement. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 54(3-4), 193–210.
- Wilson, R. P., White, C. R., Quintana, F., Halsey, L. G., Liebsch, N., Martin, G. R. & Butler, P. J. (2006). Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: The case of the cor- morant. *Journal of Animal Ecology*, 75(5), 1081–1090.

- Wilson, R. P., Grundy, E., Massy, R., Soltis, J., Tysse, B., Holton, M., Cai, Y., Parrott, A., Downey, L. A., Qasem, L. et al. (2014). Wild state secrets: Ultra-sensitive measurement of micro-movement can reveal internal processes in animals. *Frontiers in Ecology and the Environment*, 12(10), 582–587.
- Wilson, R. P., Shepard, E. & Liebsch, N. (2008). Prying into the intimate details of animal lives: Use of a daily diary on animals. *Endangered species research*, 4(1-2), 123–137.
- Wulff, K., Gatti, S., Wettstein, J. G. & Foster, R. G. (2010). Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nature Reviews Neuroscience*, 11(8), 589–599.
- Xie, J. & Qiu, Z. (2007). The effect of imbalanced data sets on lda: A theoretical and empirical analysis. *Pattern Recognition*, 40(2), 557–562.
- Yam, P. S., Penpraze, V., Young, D., Todd, M. S., Cloney, A. D., Houston-Callaghan, K. A. & 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(2), 86–91.
- Yang, C. C. & Hsu, Y. L. (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, 10(8), 7772–7788.
- Yeates, J. W. & Main, D. C. (2008). Assessment of positive welfare: A review. *The Veterinary Journal*, 175(3), 293–300.
- Zamansky, A., van der Linden, D., Hadar, I. & Bleuer-Elsner, S. (2019). Log my dog: Perceived impact of dog activity tracking. *Computer*, 52(9), 35–43.
- Zanghi, B. M., Kerr, W., Gierer, J., de Rivera, C., Araujo, J. A. & Milgram, N. W. (2013). Characterizing behavioral sleep using actigraphy in adult dogs of various ages fed once or twice daily. *Journal of Veterinary Behavior*, 8(4), 195–203.
- Zeng, B., Teng, Z., Cai, Y., Guo, S. & Qing, B. (2011). Harmonic phasor analysis based on improved fft algorithm. *IEEE Transactions on Smart Grid*, 2(1), 51–59.
- Zhang, L., Gray, H., Ye, X., Collins, L. & Allinson, N. (2019). Automatic individual pig detection and tracking in pig farms. *Sensors*, 19(5), 1188.
- Zhang, S., McCullagh, P., Nugent, C. & Zheng, H. (2010). Activity monitoring using a smart phone's accelerometer with hierarchical classification. *2010 Sixth International Conference on Intelligent Environments*, 158–163.
- Zuur, A. F., Ieno, E. N. & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14.

Appendices

Appendix A

Sample Details

Table A.1: Sample details of the Osteoarthritic dog group

ID	Age	Breed	Sex	Neutered	LOAD
OA001MZB	10	Greyhound	M	Y	14
OA002FZB	11	Labrador	F	Y	52
OA003MZB	14	Springer Spaniel	M	Y	36
OA004MCL	13	Border Collie	M	Y	22
OA005FZB	14	Labrador	F	Y	25
OA006FZB	11	Labrador	F	Y	26
OA007MCL	9	German Shorthaired Pointer	M	Y	2
OA008FZB	9	Springer Spaniel	F	Y	23
OA009MZB	11	Border Collie	M	Y	21
OA010MZB	8	German Shepherd	M	Y	24
OA011MZB	8	Spinone	M	Y	15
OA012FZB	10	Rottweiler	F	Y	36
OA013FZB	11	Labrador	F	Y	16
OA014FZB	11	Boxer	F	Y	20
OA015FCL	12	Cross	F	Y	11
OA016FZB	9	Basset	F	Y	23
OA017FZB	13	Toy Poodle	F	Y	10
OA018FCL	14	Border Collie	F	Y	14
OA019MCL	15	Jack Russell	M	Y	33
OA020MZB	13	Collie cross	M	Y	13
OA021MZB	9	Collie cross	M	Y	16
OA022MZB	9	Toy poodle	M	Y	7
OA023MCL	10	Boxer	M	Y	26
OA024MDB	8	Unknown	M	Y	21
OA025MDB	10.58	Lab x Collie	M	Y	7
OA026MDB	13	Border Collie	M	Y	26
OA027MDB	8	Labrador	M	Y	15
OA028FDB	9	Collie cross	F	Y	12
OA029MDB	16	Fox Terrier	M	Y	20

Table A.2: Sample details of the Control dog group

ID	Age	Breed	Sex	Neutered	LOAD
C001FGK	5	Labrador	F	Y	0
C002FGK	6	Irish Wolfhound	F	Y	9
C003MGk	6	Lurcher	M	Y	1
C004MGk	2	Cockerpoo	M	Y	2
C005MGk	7.5	Rhodesian Ridgeback	M	Y	4
C006MGK	2	Labrador	M	N	0
C007FGK	10	Springer Spaniel	F	Y	10
C008FGK	9	Rhodesian Ridgeback	F	Y	8
C009MGK	NA	Lurcher	M	Y	2
C010FGK	1.5	Cavapoo	F	N	4
C011FGK	8	Mongrel	F	Y	2
C012FGK	9	Jack Russell	F	Y	2.5
C013MGK	4	Jack Russell	M	Y	2
C014FGK	5	Labrador	F	Y	3
C015FGK	10	Cross (Terrier/Lurcher style)	F	Y	2
C016FGK	2	Corgi	F	Y	3
C017FGK	5	Springer/Collie Cross	F	Y	2
C018FGK	1.9	Lurcher	F	Y	1
C019MGK	NA	Labrador	M	Y	3
C020FGK	3.5	Collie cross	F	Y	2
C021MDB	1.5	Springer Spaniel/Labrador	M	Y	1
C022MDB	2	Miniature Schnauzer	M	Y	2
C023MDB	4	Kelpie	M	Y	0
C024MDB	3.5	Labradoodle	M	Y	3
C025MDB	7	Lurcher	M	Y	0
C026MDB	2	Sprocker (Working cocker/Working springer)	M	Y	0
C027MDB	4	Staffordshire bull terrier	M	Y	3
C028MDB	7	Utonagan	M	N	6
C029FDB	3.5	Mongrel (small)	F	Y	6
C030FDB	8	Labrador Retriever	F	Y	3
C031FDB	2.75	Border Terrier	F	Y	4
C032FDB	2.75	Border Terrier	F	Y	5
C033FDB	4	Working Cocker Spaniel	F	Y	3
C034FDB	5	Bedlington Terrier	F	Y	6
C035MDB	6	Beagle	M	Y	9
C036FDB	2	Border collie	F	N	1
C037MDB	7.5	Ibizan Hound	M	Y	5
C038MDB	7.5	Ibizan Hound	M	Y	3
C039FDB	13	Collie Alsatian Cross	F	Y	6
C040MDB	11	West highland terrier	M	Y	19
C041FLA	7	Dachshund	F	Y	9
C042FJO	9	Jack Russell	F	Y	9
C043FJO	11.92	Jack Russell	F	Y	9
C044FLA	6	Golden Doodle	F	Y	NA
C045FLA	2	Vizla	F	Y	NA
C046MLA	2	Cocker Spaniel	M	Y	NA
C047MDB	2.58	Working collie cocker cross	M	Y	2
C048MDB	13	Jack Russell	M	Y	12
C049FDB	9	Welsh Springer Spaniel	F	Y	7
C050FDB	3	Greyhound	F	Y	4
C051FDB	13	Border collie	F	Y	24
C052FDB	3	Sprocker spaniel	F	Y	3
C053FDB	5.25	Alsatian	F	Y	1
C054FDB	13	Jack Russell	F	Y	2
C055FDB	8	Border collie	F	Y	15
C056FDB	9	Border collie	F	Y	17

Appendix B

Read.ELAN function code

```
1 library(data.table)
2
3 read.elan<-function(d,vstart ,aoffset , tz){
4     if (missing(d)==T){
5         stop("No_ELAN_export_specified._ELAN_exported_
6             files_should_be_in_CSV_format_and_include_
7             Start,_Stop,_Duration_and_Behaviour_Class.")
8     }
9     if (missing(vstart)==T){
10        vstart<-strptime("0", format = "%S")
11        warning(paste('Time_of_video_not_specified_
12                    default_to_midnight_', strptime("0", format
13                    = "%S"), sep=""))
14    }
15    if (missing(aoffset)==T){
16        aoffset<-"0"
17        warning('Offset_of_annotations_to_video_not_
18                specified._Default_to_0.')
19    }
20    if (missing(tz)==T){
21        tz<-Sys.timezone()
22        warning(paste('Timezone_not_specified._Default_
23                    to_system_timezone:_', Sys.timezone(), sep="
24                    "))
25    }
26    # vstart = video start time calculated, convert to unix
27    # using tz
```

```
22     vstart <- as.numeric(as.POSIXct(vstart,format="%d/%m/%Y
      _%H:%M:%OS",tz=tz))
23     aoffset<-as.numeric(substring(aoffset,7))
24
25     # Identify second counts within annotation file
26     # finds numeric columns (those input as second counts)
      & includes class label
27     if (typeof(d)== "character"){
28         d <- fread(d)
29     } else {d<-as.data.table(d)}
30
31     d<-cbind(d[,which(unlist(lapply(d, function(x)all(is.
      numeric(x)==TRUE)))]),with=F], d[[ncol(d)]]
32     # adds vstart and annotation offset to convert to UNIX
      time and align
33     # currently aoffset only takes number of seconds -
      needs to be adjusted to take timestamp min:sec.msec
      style input too
34     d[,1:2] <- d[,1:2]+vstart+aoffset
35     names(d)<-c("start","end","duration","class")
36     return(d)
37 }
```

Appendix C

Unknown annotation generation function code

```
1 library(data.table)
2
3 unknown.gen<-function(data,wlen,fs=100,ann){
4
5     # adjust annotations to window length by finding
        difference from actual window length to calculated
        window length
6     wlen.edit<-(wlen-(ann$end-ann$start))/2
7     ann$start<-ann$start-wlen.edit
8
9     for ( i in 1:nrow(ann)){
10         start.ind <- which.min(abs(data[[1]]-ann$start[
11             i]))
12         ann$start[i] <- data[[1]][start.ind]
13         ann$end[i] <- data[[1]][start.ind+(fs*wlen-1)]
14     }
15
16     # check if unknown data already exists - if it does
        then this function just ensures everything matches
        the wlen
17     if (any(ann$class=="unknown")==FALSE){
18
19         data<-data[data$Time>=min(ann$start) & data$
20             Time<=(max(ann$end)),]
21
22         # Generate unknown data from within min/max
            collection period to train against non-
            postural transition
```

```
23         unknown.ann<-data.table(start=double(nrow(ann))
24             ,end=double(nrow(ann)),duration=double(nrow(
25                 ann)),class=character(nrow(ann)))
26
27     for (i in 1:nrow(ann)){
28         unknown.ann$start[i] <- sample(data$
29             Time,1)
30         start.ind <- which.min(abs(data[[1]] -
31             unknown.ann$start[i]))
32         unknown.ann$start[i] <- data[[1]][start
33             .ind]
34         end.ind<-start.ind+(fs*wlen-1)
35         unknown.ann$end[i] <- data[[1]][end.ind
36             ]
37         unknown.ann$duration[i] <- wlen
38         unknown.ann$class[i] <- "unknown"
39     }
```

```
ann<-rbind(ann,unknown.ann)}
```

```
return(ann)
```

```
}
```

Appendix D

Rotation correction function code

```
1 library(data.table)
2 library(doSNOW)
3 library(foreach)
4 library(pracma)
5 library(signal)
6 library(utils)
7 library(doParallel)
8
9 r.fix <- function(d,fs ,gDev,wlen ,writeCSV=FALSE, path){
10
11     # Check for arguments and add defaults if missing.
12     if (missing(fs)==T){
13         fs<-100
14         warning('Frequency_(fs)_not_specified._Default_
15                 fs_=_100')
16     }
17     if (missing(gDev)==T){
18         gDev<-0.2
19         warning('Gravity_deviation_threshold_not_
20                 specified._Default_gDev_=_0.2')
21     }
22     if (missing(wlen)==T){
23         wlen<-5
24         warning('Window_length_not_specified._Default_
25                 wlen_=_5')
26     }
27     if (writeCSV==TRUE){
```

```
28         if (missing(path)==T){
29             path<-getwd()
30             warning('path_not_specified._Default_is
                 _your_working_directory ')
31         }
32     }
33
34     cores <- detectCores()-1
35
36     # Variables held constant/not defined within function
        call
37     targ.cart <- c(0,0,1)
38
39     # create temporary data variable
40     temp.d<-d[,2:4] # removes character string datetime
41
42     # Build filter
43     butter.fil <- butter(4,0.1/(fs/2), type="low")
44
45     # Apply to temp data
46     temp.d[,1] <- as.numeric(filtfilt(butter.fil , as.matrix
        (temp.d[,1])) )# apply filter to x
47     temp.d[,2] <- as.numeric(filtfilt(butter.fil , as.
        matrix(temp.d[,2])) )# apply filter to y
48     temp.d[,3] <- as.numeric(filtfilt(butter.fil , as.
        matrix(temp.d[,3])) )# apply filter to z
49
50     temp.d$Time <- seq(1/fs ,length(temp.d[[1]])/fs ,by=1/fs)
51
52     window.size = fs*wlen
53     step.size = window.size/2
54     iter.seq<-seq(0, length(temp.d[[1]])-1, step.size)
55
56
57     # progress bar 1
58     sink(tempfile()) # prevents the empty 0% bar being
        printed
59     pb1 <- txtProgressBar(min=0, max=length(iter.seq),
        style=3)
```

```

60     progress <- function(n) setTxtProgressBar(pb1, n)
61     opts1 <- list(progress=progress)
62     sink()
63     cat('\nWindowing_data:\n')
64
65     cl <- makeSOCKcluster(cores)
66     registerDoSNOW(cl)
67
68     win.sum<-as.data.frame(foreach(i=iter.seq, combine = '
        rbind', .packages="data.table", .options.snow=opts1
        ) %dopar% {
69         x=mean(as.numeric(unlist(temp.d[i:(i+window.
            size),1]))) #each data frame has to be
            unlisted and parsed
70         y=mean(as.numeric(unlist(temp.d[i:(i+window.
            size),2])))
71         z=mean(as.numeric(unlist(temp.d[i:(i+window.
            size),3])))
72         t=temp.d[(i+1),4] # if i=0 it will throw an
            error
73         cbind(x,y,z,t)
74     })
75     close(pb1)
76     stopCluster(cl)
77
78     # convert cartesian to spherical coords for the average
        of the window
79     sph.coord <- cart2sph(as.matrix(win.sum[,1:3]))
80
81     # Select candidate windows for correction by
        establishing all that differ from the origin by 1g
82     cand.index <- which(sph.coord[,3]>=(1-gDev) & sph.coord
        [,3]<=(1+gDev))
83
84     #subset by cand.index
85     win.sum<-win.sum[cand.index,]
86
87     # build theta list to iterate through
88     theta.iter<-seq(0, 359.5, by=1)*(pi/180)

```

```
89
90     # create dataframe for storing converted coords prior
      to selection
91     theta.coords<-data.frame(x=double(length(theta.iter)),y
      =double(length(theta.iter)),z=double(length(theta.
      iter)))
92
93     # dataframe to store vector between target and
      corrected
94     t.vect<-data.frame(x=double(length(theta.iter)),y=
      double(length(theta.iter)),z=double(length(theta.
      iter)))
95
96     # create output copy of data
97     temp.d.rot<-as.data.frame(temp.d)
98
99     # progress bar 2
100    sink(tempfile())
101    pb2 <- txtProgressBar(min=1, max=(length(cand.index)),
      style=3)
102    progress <- function(n) setTxtProgressBar(pb2, n)
103    opts2 <- list(progress=progress)
104    sink()
105    cat('\nFinding optimal adjustment angles:\n')
106    # loop through the list of candidates finding value of
      theta resulting in minimum distance from target to
      corrected
107
108    cl <- makeSOCKcluster(cores)
109    registerDoSNOW(cl)
110
111    theta.ang <- foreach (i=1:(length(cand.index)),.combine
      = 'rbind', .packages="data.table",.options.snow=
      opts2) %dopar% {
112
113        # create list of coords for each angle and each
          candidate frame
114        theta.coords[1:length(theta.iter),1]<-(win.sum[
          i,1]*cos(theta.iter))+(win.sum[i,3]*sin(
```

```

        theta.iter))
115     theta.coords[,2]<-win.sum[i,2]
116     theta.coords[,3]<-(win.sum[i,1]*-sin(theta.iter
        ))+(win.sum[i,3]*cos(theta.iter))
117
118     # vector from corrected coords to target frame
119     t.vect[1:length(theta.iter),1]<-targ.cart[1]-
        theta.coords[,1]
120     t.vect[,2]<-targ.cart[2]-theta.coords[,2]
121     t.vect[,3]<-targ.cart[3]-theta.coords[,3]
122
123     # the index of the smallest vector distance
124     t.vect.ind <- which(sqrt(t.vect[1]^2+t.vect
        [2]^2+t.vect[3]^2)==min(sqrt(t.vect[1]^2+t.
        vect[2]^2+t.vect[3]^2)))
125
126     # the value of theta to build the rotation
        matrix from start of window to start of next
127     theta.ang <- theta.iter[t.vect.ind]
128   }
129
130   close(pb2)
131   stopCluster(cl)
132
133   # cbind theta values to win.sum
134   win.sum<-cbind(win.sum,theta.ang)
135
136   # Collapse consecutive, identical values of theta
137   win.sum<-win.sum[nrow(win.sum):1,]
138   win.sum<-win.sum[c(win.sum$theta.ang[-1] != win.sum$
        theta.ang[-nrow(win.sum)],TRUE),]
139   win.sum<-win.sum[nrow(win.sum):1,]
140
141   if (win.sum[1,4]!=temp.d.rot[1,4]){
142     win.sum<-rbind(c(0,0,0,temp.d.rot[1,4],0),win.
        sum)
143   }
144
145   sink(tempfile())

```

```

146     # progress bar 3
147     pb3 <- txtProgressBar(min=0, max=nrow(win.sum), style
      =3)
148     progress <- function(n) setTxtProgressBar(pb3, n)
149     opts3 <- list(progress=progress)
150     sink()
151     cat("\nApplying rotation correction:\n")
152
153     cl <- makeSOCKcluster(cores)
154     registerDoSNOW(cl)
155
156     # Foreach loop to apply rotation matrices
157     temp.d.rot[1:3] <- foreach (i=1:(nrow(win.sum)), .
      combine = 'rbind', .packages="data.table", .options.
      snow=opts3) %dopar% {
158
159         if (!is.na(win.sum[i+1,1])==TRUE) {
160             # correct x axis between windows
161             temp.x<-
162                 temp.d.rot[temp.d.rot[,4]>=win.
                  sum[i,4] & temp.d.rot[,4]<
                  win.sum[i+1,4],1]*cos(win.
                  sum[i,5])+
163                 temp.d.rot[temp.d.rot[,4]>=win.
                  sum[i,4] & temp.d.rot[,4]<
                  win.sum[i+1,4],3]*sin(win.
                  sum[i,5])
164             # add y axis
165             temp.y<-
166                 temp.d.rot[temp.d.rot[,4]>=win.
                  sum[i,4] & temp.d.rot[,4]<
                  win.sum[i+1,4],2]
167             #correct z axis
168             temp.z<-
169                 temp.d.rot[temp.d.rot[,4]>=win.
                  sum[i,4] & temp.d.rot[,4]<
                  win.sum[i+1,4],1]*-sin(win.
                  sum[i,5])+
170                 temp.d.rot[temp.d.rot[,4]>=win.

```

```

sum[i,4] & temp.d.rot[,4] <
win.sum[i+1,4],3]*cos(win.
sum[i,5])
171
172     #combine
173     temp.all<-cbind(temp.x,temp.y,temp.z)
174 } else {
175     # correct x axis between windows
176     temp.x<-
177         temp.d.rot[temp.d.rot[,4]>=win.
sum[i,4] ,1]*cos(win.sum[i
,5])+
178         temp.d.rot[temp.d.rot[,4]>=win.
sum[i,4] ,3]*sin(win.sum[i
,5])
179     # add y axis
180     temp.y<-
181         temp.d.rot[temp.d.rot[,4]>=win.
sum[i,4] ,2]
182     #correct z axis
183     temp.z<-
184         temp.d.rot[temp.d.rot[,4]>=win.
sum[i,4] ,1]*-sin(win.sum[i
,5])+
185         temp.d.rot[temp.d.rot[,4]>=win.
sum[i,4] ,3]*cos(win.sum[i
,5])
186
187     #combine
188     temp.all<-cbind(temp.x,temp.y,temp.z)
189 }
190     return(temp.all)
191
192 }
193 close(pb3)
194 stopCluster(cl)
195
196 temp.d.rot<-cbind(as.numeric(as.POSIXct(d[[1]])),temp.d
.rot[,1:4])

```

```
197     temp.d.rot<-temp.d.rot[,1:4]
198     names(temp.d.rot)<-c("Time","x","y","z")
199
200     # checking if the path argument has been passed, if not
201     set path to working directory
202     if(!exists("path")){
203         path <- getwd()
204     }
205
206     # checking if the user has passed a CSV file , if so and
207     write CSV is true , write to it , else create a
208     rotation_correction CSV
209     if (writeCSV == TRUE){
210         if(writeCSV && file_ext(path)=="csv"){
211             print("writing_to_CSV")
212             write.csv(temp.d.rot, file=path, row.
213                 names = FALSE)
214         }else {
215             warning("No_file_specified_in_your_path
216                 ,_creating_\\"rotation_correction.csv
217                 \"_\")
218             write.csv(temp.d.rot, file=paste(path,"
219                 rotation_correction.csv", sep=''),
220                 row.names = FALSE)
221         }
222     }
223
224     return(temp.d.rot)
225 }
226 }
```

Appendix E

Feature calculation function code

```
1 library(data.table)
2 library(doSNOW)
3 library(foreach)
4 library(pracma)
5 library(signal)
6 library(utils)
7 library(doParallel)
8
9 feat.calc<-function(data,wlen=5,fs=100,ann, FFT=FALSE, unknown.
   gen = FALSE, angle.calc = FALSE,svm.ratio = FALSE){
10
11     # Currently only calculates features for annotations (
       for train/test) – in future needs options to
       calculate across all
12
13
14     # input for annotation = start (in unixtime), end (in
       unixtime), duration, class. Check if it exists and
       if so use duration to calc wlen
15     # if/else statement to either read data from a variable
       OR if passed a string to fread it in.
16     if (missing(ann)==FALSE){
17         if (typeof(ann)=="character"){
18             ann<-fread(ann)
19         }
20         if (missing(wlen)==TRUE){
21             wlen<-ceiling(mean(ann$duration)+(2*sd(
                ann$duration)))
22             message("Window_length_set_to_", wlen,
                "_derived_from_mean_annotation_")
```

```
length_+_2_standard_deviations")
23     # Use wlen argument (or ann file
        duration) or default to 2
24     }
25 }
26
27 # Check if data is loaded or needs loading from a
    string
28 if (typeof(data)=="character"){
29     data<-fread(data)
30 }
31
32 # convert data time to UNIX
33 if (typeof(data$Time)=="character"){
34     data$Time<-as.numeric(as.POSIXct(data$Time))
35     warning("Timestamps_for_accelerometer_data_were
        character_format.If_timezone_of_data_
        differed_from_system_timezone_conversion_
        will_be_incorrect.")
36 }
37 data$Time<-as.numeric(data$Time) # accounts for times
    being in posixct format already - redundant if
    already unix
38
39 #find min start and max end unix in annotation. Then
    trim data to this to conserve memory prior to
    looping
40 data<-data[data$Time>=min(ann$start) & data$Time<=max(
    ann$end),]
41
42 # Generate SVM
43 svm <- sqrt(data[,2]^2+data[,3]^2+data[,4]^2)
44 names(svm)<-c("svm")
45 data<-cbind(data,svm)
46
47 # progress bar 1
48 sink(tempfile()) # prevents the empty 0% bar being
    printed
49 pb1 <- txtProgressBar(min=0, max=length(ann$start),
```

```

        style=3)
50     progress <- function(n) setTxtProgressBar(pb1, n)
51     opts1 <- list(progress=progress)
52     sink()
53     cat('\nCalculating_Time-Domain_features:\n')
54
55     # begin foreach loop to generate features
56     cores <- detectCores()-1
57     cl <- makeSOCKcluster(cores)
58     registerDoSNOW(cl)
59
60     feat.out <- foreach(i=1:nrow(ann), .combine = 'rbind',
        .packages=c("data.table", "e1071", "entropy"), .
        options.snow=opts1) %dopar% {
61
62         # Mean
63         x_mean <- mean(data[[2]][data[,1]>=ann$start[i]
        &data[,1]<=ann$end[i]])
64         y_mean <- mean(data[[3]][data[,1]>=ann$start[i]
        &data[,1]<=ann$end[i]])
65         z_mean <- mean(data[[4]][data[,1]>=ann$start[i]
        &data[,1]<=ann$end[i]])
66         svm_mean <- mean(data[[5]][data[,1]>=ann$start[
        i]&data[,1]<=ann$end[i]])
67
68         # median
69         x_median <- median(data[[2]][data[,1]>=ann$
        start[i]&data[,1]<=ann$end[i]])
70         y_median <- median(data[[3]][data[,1]>=ann$
        start[i]&data[,1]<=ann$end[i]])
71         z_median <- median(data[[4]][data[,1]>=ann$
        start[i]&data[,1]<=ann$end[i]])
72         svm_median <- median(data[[5]][data[,1]>=ann$
        start[i]&data[,1]<=ann$end[i]])
73
74         # min
75         x_min <- min(data[[2]][data[,1]>=ann$start[i]&
        data[,1]<=ann$end[i]])
76         y_min <- min(data[[3]][data[,1]>=ann$start[i]&

```

```
      data[,1] <= ann$end[i]))
77  z_min <- min(data[[4]][data[,1] >= ann$start[i] &
      data[,1] <= ann$end[i]])
78  svm_min <- min(data[[5]][data[,1] >= ann$start[i]
      & data[,1] <= ann$end[i]])
79
80  # max
81  x_max <- max(data[[2]][data[,1] >= ann$start[i] &
      data[,1] <= ann$end[i]])
82  y_max <- max(data[[3]][data[,1] >= ann$start[i] &
      data[,1] <= ann$end[i]])
83  z_max <- max(data[[4]][data[,1] >= ann$start[i] &
      data[,1] <= ann$end[i]])
84  svm_max <- max(data[[5]][data[,1] >= ann$start[i]
      & data[,1] <= ann$end[i]])
85
86  # range
87  x_range <- diff(range(data[[2]][data[,1] >= ann$
      start[i] & data[,1] <= ann$end[i]]))
88  y_range <- diff(range(data[[3]][data[,1] >= ann$
      start[i] & data[,1] <= ann$end[i]]))
89  z_range <- diff(range(data[[4]][data[,1] >= ann$
      start[i] & data[,1] <= ann$end[i]]))
90  svm_range <- diff(range(data[[5]][data[,1] >= ann
      $start[i] & data[,1] <= ann$end[i]]))
91
92  # sd
93  x_sd <- sd(data[[2]][data[,1] >= ann$start[i] &
      data[,1] <= ann$end[i]])
94  y_sd <- sd(data[[3]][data[,1] >= ann$start[i] &
      data[,1] <= ann$end[i]])
95  z_sd <- sd(data[[4]][data[,1] >= ann$start[i] &
      data[,1] <= ann$end[i]])
96  svm_sd <- sd(data[[5]][data[,1] >= ann$start[i] &
      data[,1] <= ann$end[i]])
97
98  # Variance
99  x_var <- var(data[[2]][data[,1] >= ann$start[i] &
      data[,1] <= ann$end[i]])
```

```

100     y_var <- var(data[[3]][data[,1]>=ann$start[i]&
      data[,1]<=ann$end[i]])
101     z_var <- var(data[[4]][data[,1]>=ann$start[i]&
      data[,1]<=ann$end[i]])
102     svm_var <- var(data[[5]][data[,1]>=ann$start[i]
      &data[,1]<=ann$end[i]])
103
104     # rms
105     x_rms <- sqrt(mean(data[[2]][data[,1]>=ann$
      start[i]&data[,1]<=ann$end[i]]^2))
106     y_rms <- sqrt(mean(data[[3]][data[,1]>=ann$
      start[i]&data[,1]<=ann$end[i]]^2))
107     z_rms <- sqrt(mean(data[[4]][data[,1]>=ann$
      start[i]&data[,1]<=ann$end[i]]^2))
108     svm_rms <- sqrt(mean(data[[5]][data[,1]>=ann$
      start[i]&data[,1]<=ann$end[i]]^2))
109
110     # MAD
111     x_mad<-mad(data[[2]][data[,1]>=ann$start[i]&
      data[,1]<=ann$end[i]],center=mean(data[[2]][
      data[,1]>=ann$start[i]&data[,1]<=ann$end[i]
      ]))
112     y_mad<-mad(data[[3]][data[,1]>=ann$start[i]&
      data[,1]<=ann$end[i]],center=mean(data[[3]][
      data[,1]>=ann$start[i]&data[,1]<=ann$end[i]
      ]))
113     z_mad<-mad(data[[4]][data[,1]>=ann$start[i]&
      data[,1]<=ann$end[i]],center=mean(data[[4]][
      data[,1]>=ann$start[i]&data[,1]<=ann$end[i]
      ]))
114     svm_mad<-mad(data[[5]][data[,1]>=ann$start[i]&
      data[,1]<=ann$end[i]],center=mean(data[[5]][
      data[,1]>=ann$start[i]&data[,1]<=ann$end[i]
      ]))
115
116     # Correlation (Ravi et al 2005)
117     xy_cor <- cov(data[[2]][data[,1]>=ann$start[i]&
      data[,1]<=ann$end[i]],data[[3]][data[,1]>=
      ann$start[i]&data[,1]<=ann$end[i]])/(x_sd*y_

```

```

sd)
118 yz_cor <- cov(data[[3]][data[,1]>=ann$start[i]&
data[,1]<=ann$end[i]], data[[4]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]])/(y_sd*z_
sd)
119 xz_cor <- cov(data[[2]][data[,1]>=ann$start[i]&
data[,1]<=ann$end[i]], data[[4]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]])/(x_sd*z_
sd)
120
121
122 # Window Gradient min/max
123
124 # Min Index
125 x_min_index <- min(which(data[[2]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]]== min(
data[[2]][data[,1]>=ann$start[i]&data[,1]<=
ann$end[i]])))
126 y_min_index <- min(which(data[[3]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]]== min(
data[[3]][data[,1]>=ann$start[i]&data[,1]<=
ann$end[i]])))
127 z_min_index <- min(which(data[[4]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]]== min(
data[[4]][data[,1]>=ann$start[i]&data[,1]<=
ann$end[i]])))
128 svm_min_index <- min(which(data[[5]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]]== min(
data[[5]][data[,1]>=ann$start[i]&data[,1]<=
ann$end[i]])))
129
130 # Max Index
131 x_max_index <- max(which(data[[2]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]]== max(
data[[2]][data[,1]>=ann$start[i]&data[,1]<=
ann$end[i]])))
132 y_max_index <- max(which(data[[3]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]]== max(
data[[3]][data[,1]>=ann$start[i]&data[,1]<=

```

```

ann$end[i]))))
133 z_max_index <- max(which(data[[4]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]]== max(
data[[4]][data[,1]>=ann$start[i]&data[,1]<=
ann$end[i]])))
134 svm_max_index <- max(which(data[[5]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]]== max(
data[[5]][data[,1]>=ann$start[i]&data[,1]<=
ann$end[i]])))

135
136 # Gradient
137 x_grad <-(x_min - x_max)/(x_min_index - x_max_
index)
138 y_grad <-(y_min - y_max)/(y_min_index - y_max_
index)
139 z_grad <-(z_min - z_max)/(z_min_index - z_max_
index)
140 svm_grad <-(svm_min - svm_max)/(svm_min_index -
svm_max_index)

141
142 # Mean of absolute acceleration
143 x_abs_mean <- mean(abs(data[[2]][data[,1]>=ann$
start[i]&data[,1]<=ann$end[i]]))
144 y_abs_mean <- mean(abs(data[[3]][data[,1]>=ann$
start[i]&data[,1]<=ann$end[i]]))
145 z_abs_mean <- mean(abs(data[[4]][data[,1]>=ann$
start[i]&data[,1]<=ann$end[i]]))
146 svm_abs_mean <- mean(abs(data[[5]][data[,1]>=
ann$start[i]&data[,1]<=ann$end[i]]))

147
148 # Skewness
149 x_skew <- skewness(data[[2]][data[,1]>=ann$
start[i]&data[,1]<=ann$end[i]])
150 y_skew <- skewness(data[[3]][data[,1]>=ann$
start[i]&data[,1]<=ann$end[i]])
151 z_skew <- skewness(data[[4]][data[,1]>=ann$
start[i]&data[,1]<=ann$end[i]])
152 svm_skew <- skewness(data[[5]][data[,1]>=ann$
start[i]&data[,1]<=ann$end[i]])

```

```
153
154     # Kurtosis
155     x_kurt <- kurtosis(data[[2]][data[,1]>=ann$
156         start[i]&data[,1]<=ann$end[i]])
157     y_kurt <- kurtosis(data[[3]][data[,1]>=ann$
158         start[i]&data[,1]<=ann$end[i]])
159     z_kurt <- kurtosis(data[[4]][data[,1]>=ann$
160         start[i]&data[,1]<=ann$end[i]])
161     svm_kurt <- kurtosis(data[[5]][data[,1]>=ann$
162         start[i]&data[,1]<=ann$end[i]])
163
164     # IQR
165     x_iqr <- IQR(data[[2]][data[,1]>=ann$start[i]&
166         data[,1]<=ann$end[i]])
167     y_iqr <- IQR(data[[3]][data[,1]>=ann$start[i]&
168         data[,1]<=ann$end[i]])
169     z_iqr <- IQR(data[[4]][data[,1]>=ann$start[i]&
170         data[,1]<=ann$end[i]])
171     svm_iqr <- IQR(data[[5]][data[,1]>=ann$start[i]
172         &data[,1]<=ann$end[i]])
173
174     # Shannon Entropy
175     x_ent <- entropy(data[[2]][data[,1]>=ann$start[
176         i]&data[,1]<=ann$end[i]])
177     y_ent <- entropy(data[[3]][data[,1]>=ann$start[
178         i]&data[,1]<=ann$end[i]])
179     z_ent <- entropy(data[[4]][data[,1]>=ann$start[
180         i]&data[,1]<=ann$end[i]])
181     svm_ent <- entropy(data[[5]][data[,1]>=ann$
182         start[i]&data[,1]<=ann$end[i]])
183
184     # combine
185     all_feats <- cbind(ann[i,4], ann[i,1:3], x_mean, y_
186         mean, z_mean, svm_mean, x_median, y_median, z_
187         median, svm_median, x_min, y_min, z_min, svm_min,
188         x_max, y_max, z_max, svm_max, x_range, y_range, z_
189         range, svm_range, x_sd, y_sd, z_sd, svm_sd, x_var,
190         y_var, z_var, svm_var,
191         x_rms, y_rms, z_rms, svm_rms, x_mad, y_mad, z_mad, svm
```

```

    _mad , x_abs_mean , y_abs_mean , z_abs_mean ,
176 svm_abs_mean , x_skew , y_skew , z_skew , svm_skew , x_
    kurt , y_kurt , z_kurt , svm_kurt , x_iqr , y_iqr , z_
    iqr , svm_iqr ,
177 x_ent , y_ent , z_ent , svm_ent , xy_cor , yz_cor , xz_cor )
178
179 if( angle.calc == TRUE){
180
181     x_seg <- data[[2]][data[,1]>=ann$start[
        i]&data[,1]<=ann$end[i]]
182     y_seg <- data[[3]][data[,1]>=ann$start[
        i]&data[,1]<=ann$end[i]]
183     z_seg <- data[[4]][data[,1]>=ann$start[
        i]&data[,1]<=ann$end[i]]
184
185     # caculate average pitch
186     mean_pitch <- mean(atan(-x_seg/sqrt(y_
        seg^2+z_seg^2)))
187
188     # calculate average roll
189     mean_roll <- mean(atan(y_seg/sqrt(x_seg
        ^2+z_seg^2)))
190
191     # combine
192     all.feats<-cbind(all.feats , mean_pitch ,
        mean_roll )
193
194 }
195 if (svm.ratio==TRUE){
196     x_ratio <- x_mean/svm_mean
197     y_ratio <- y_mean/svm_mean
198     z_ratio <- z_mean/svm_mean
199
200     # combine
201     all.feats<-cbind(all.feats , x_ratio , y_
        ratio , z_ratio )
202
203 }
204

```

```
205
206     }
207     stopCluster(cl)
208     close(pb1)
209
210     # if/else for FFT loop – if you need this let me know
211     and will send over the code we have
212     if (FFT==TRUE){
213         #Nyquist
214         fn <- fs/2
215         # Build Hanning window
216         fft_win<-hanning(wlen*fs)
217
218         # progress bar fft
219         sink(tempfile()) # prevents the empty 0% bar
220         being printed
221         pb2 <- txtProgressBar(min=0, max=length(ann$
222             start), style=3)
223         progress <- function(n) setTxtProgressBar(pb2,
224             n)
225         optsfft <- list(progress=progress)
226         sink()
227         cat("\nCalculating_Frequency-Domain_Features:\n
228             ")
229
230         cl <- makeSOCKcluster(cores)
231         registerDoSNOW(cl)
232
233         fft.out <- foreach(i=1:nrow(ann), .combine = '
234             rbind', .packages=c("data.table", "e1071", "
235             entropy", "signal"), .options.snow=optsfft) %
236             dopar% {
237
238             #Normalise data
239             norm_x <- data.matrix(data[[2]][data
240                 [,1]>=ann$start[i]&data[,1]<=ann$end
241                 [i]])
242             norm_x <- norm_x-mean(norm_x)
```

```

234 norm_y <- data.matrix(data[[3]][data
      [,1]>=ann$start[i]&data[,1]<=ann$end
      [i]])
235 norm_y <- norm_y-mean(norm_y)
236 norm_z <- data.matrix(data[[4]][data
      [,1]>=ann$start[i]&data[,1]<=ann$end
      [i]])
237 norm_z <- norm_z-mean(norm_z)
238 norm_svm <- data.matrix(data[[5]][data
      [,1]>=ann$start[i]&data[,1]<=ann$end
      [i]])
239 norm_svm <- norm_svm-mean(norm_svm)
240
241 #Apply window
242 norm_x <- fft_win*norm_x
243 norm_y <- fft_win*norm_y
244 norm_z <- fft_win*norm_z
245 norm_svm <- fft_win*norm_svm
246
247 # FFT amplitude/magnitude/power
      calculation
248 fft_x<-fft(norm_x)/length(norm_x)
249 fft_y<-fft(norm_y)/length(norm_y)
250 fft_z<-fft(norm_z)/length(norm_z)
251 fft_svm<-fft(norm_svm)/length(norm_svm)
252
253 #Frequency axis
254
255 freq_axis <- seq(0,1, length.out =
      round(length(fft_x)/2+1))*fn
256
257 # Features – Check validity of some (
      entropy etc)
258
259 # Estimated PSD
260 psd_x <- (abs(fft_x)^2)[2:(length(fft_x
      )/2)]
261 psd_y <- (abs(fft_y)^2)[2:(length(fft_x
      )/2)]

```

```
262 psd_z <- (abs(fft_z)^2)[2:(length(fft_x
      )/2)]
263 psd_svm <- (abs(fft_svm)^2)[2:(length(
      fft_x)/2)]
264
265 # Dominant Estimated PSD
266 dpsd_x <- max(psd_x)
267 dpsd_y <- max(psd_y)
268 dpsd_z <- max(psd_z)
269 dpsd_svm <- max(psd_svm)
270
271 # Dominant Estimated PSD Frequency
272 dpsd_freq_x <- freq_axis[2:(length(fft_
      x)/2)][which.max(psd_x)]
273 dpsd_freq_y <- freq_axis[2:(length(fft_
      x)/2)][which.max(psd_y)]
274 dpsd_freq_z <- freq_axis[2:(length(fft_
      x)/2)][which.max(psd_z)]
275 dpsd_freq_svm <- freq_axis[2:(length(
      fft_x)/2)][which.max(psd_svm)]
276
277 # Second Dominant EPSD
278 psd_x2 <- psd_x[-which.max(psd_x)]
279 psd_y2 <- psd_y[-which.max(psd_y)]
280 psd_z2 <- psd_z[-which.max(psd_z)]
281 psd_svm2 <- psd_svm[-which.max(psd_svm)
      ]
282
283 dpsd_x2 <- max(psd_x2)
284 dpsd_y2 <- max(psd_y2)
285 dpsd_z2 <- max(psd_z2)
286 dpsd_svm2 <- max(psd_svm2)
287
288 # Second Dominant EPSD Frequency
289 freq_axis2_x <- freq_axis[2:(length(fft
      _x)/2)][-which.max(psd_x)]
290 freq_axis2_y <- freq_axis[2:(length(fft
      _y)/2)][-which.max(psd_y)]
291 freq_axis2_z <- freq_axis[2:(length(fft
```

```

    _z)/2)][-which.max(psd_z)]
292 freq_axis2_svm <- freq_axis[2:(length(
    fft_svm)/2)][-which.max(psd_svm)]
293
294 dpsd_freq_x2 <- freq_axis2_x[which.max(
    psd_x2)]
295 dpsd_freq_y2 <- freq_axis2_y[which.max(
    psd_y2)]
296 dpsd_freq_z2 <- freq_axis2_z[which.max(
    psd_z2)]
297 dpsd_freq_svm2 <- freq_axis2_svm[which.
    max(psd_svm2)]
298
299 # Mean Estimated PSD
300 mean_psd_x <- mean(psd_x)
301 mean_psd_y <- mean(psd_y)
302 mean_psd_z <- mean(psd_z)
303 mean_psd_svm <- mean(psd_svm)
304
305 # Entropy of PSD
306 ent_psd_x <- entropy(psd_x)
307 ent_psd_y <- entropy(psd_y)
308 ent_psd_z <- entropy(psd_z)
309 ent_psd_svm <- entropy(psd_svm)
310
311 # Energy per sample from Estimated PSD
312 eps_psd_x <- sum(psd_x)/(wlen*fs)
313 eps_psd_y <- sum(psd_y)/(wlen*fs)
314 eps_psd_z <- sum(psd_z)/(wlen*fs)
315 eps_psd_svm <- sum(psd_svm)/(wlen*fs)
316
317 # build output dataframe
318 fft.feats <- cbind(dpsd_x,dpsd_y,dpsd_z,
    dpsd_svm,dpsd_freq_x,dpsd_freq_y,
    dpsd_freq_z,dpsd_freq_svm,
319 dpsd_x2,dpsd_y2,dpsd_z2,dpsd_svm2,dpsd_
    freq_x2,dpsd_freq_y2,dpsd_freq_z2,
    dpsd_freq_svm2,
320 mean_psd_x,mean_psd_y,mean_psd_z,mean_

```

```
321         psd_svm,  
          ent_psd_x, ent_psd_y, ent_psd_z, ent_psd_  
          svm, eps_psd_x, eps_psd_y, eps_psd_z,  
          eps_psd_svm)  
322  
323     }  
324     stopCluster(cl)  
325     close(pb2)  
326  
327     feat.out<-cbind(feat.out, fft.out)  
328 }  
329 return(feat.out)  
330 }
```

Appendix F

ECDF calculation function code

```
1 library(data.table)
2 library(doParallel)
3 library(doSNOW)
4 library(e1071)
5 library(entropy)
6
7 ecdf.calc<-function(data,wlen=2,fs=100,ann, write.csv=FALSE){
8
9     # ECDF code
10    inv_ecdf <- function(f){
11        x <- environment(f)$x
12        y <- environment(f)$y
13        approxfun(y, x)
14    }
15
16    # Currently only calculates features for annotations (
17    for train/test)
18    # Column names are specifically referenced throughout
19    code
20
21    # input for annotation = start (in unixtime), end (in
22    unixtime), duration, class. Check if it exists and
23    if so use duration to calc wlen
24    # if/else statement to either read data from a variable
25    OR if passed a string to fread it in.
26    if (missing(ann)==FALSE){
27        if (typeof(ann)=="character"){
28            ann<-fread(ann)
29        }
30        if (missing(wlen)==TRUE){
```

```
26         wlen<-ceiling(mean(ann$duration)+(2*sd(
           ann$duration)))
27         message("Window_length_set_to_", wlen,
           "_derived_from_mean_annotation_
           length_+_2_standard_deviations")
28     }
29
30 }
31
32 # FFT defaults to false – set to TRUE to calculate FFT
   features
33 # Check if data is loaded or needs loading from a
   string
34 if (typeof(data)=="character"){
35     data<-fread(data)
36 }
37
38 # convert data time to UNIX
39 if (typeof(data$Time)=="character"){
40     data$Time<-as.numeric(as.POSIXct(data$Time))
41     warning("Timestamps_for_accelerometer_data_were
           _character_format._If_timezone_of_data_
           differed_from_system_timezone_conversion_
           will_be_incorrect.")
42 }
43 data$Time<-as.numeric(data$Time) # accounts for times
   being in posixct format already – redundant if
   already unix
44
45 #find min start and max end unix in annotation. Then
   trim data to this to conserve memory prior to
   looping
46 data<-data[data$Time>=min(ann$start) & data$Time<=(max(
   ann$end)),]
47
48 # Generate SVM
49 svm <- sqrt(data[,2]^2+data[,3]^2+data[,4]^2)
50 names(svm)<-c("svm")
51 data<-cbind(data,svm)
```

```

52
53   # progress bar 1
54   sink(tempfile()) # prevents the empty 0% bar being
      printed
55   pb1 <- txtProgressBar(min=1, max=length(ann$start),
      style=3)
56   progress <- function(n) setTxtProgressBar(pb1, n)
57   opts1 <- list(progress=progress)
58   sink()
59   cat('\nProcessing_ECDF\n')
60
61   # begin foreach loop to generate features
62   cores <- detectCores()-1
63   cl <- makeSOCKcluster(cores)
64   registerDoSNOW(cl)
65
66   ecdf.out <- foreach(i=1:nrow(ann), .combine = '
      rbind', .packages=c("data.table", "e1071", "
      entropy"), .options.snow=opts1) %do% {
67
68   # ECDF
69   x_ecdf <- inv_ecdf(ecdf(data[[2]][data[,1]>=ann
      $start[i]&data[,1]<=ann$end[i]]))
70   y_ecdf <- inv_ecdf(ecdf(data[[3]][data[,1]>=ann
      $start[i]&data[,1]<=ann$end[i]]))
71   z_ecdf <- inv_ecdf(ecdf(data[[4]][data[,1]>=ann
      $start[i]&data[,1]<=ann$end[i]]))
72   svm_ecdf <- inv_ecdf(ecdf(data[[5]][data[,1]>=
      ann$start[i]&data[,1]<=ann$end[i]]))
73
74   x_ecdf_full <- numeric(length=0)
75   y_ecdf_full <- numeric(length=0)
76   z_ecdf_full <- numeric(length=0)
77   svm_ecdf_full <- numeric(length=0)
78
79   for ( j in 1:100){
80       x_ecdf_full <- cbind(x_ecdf_full , x_
          ecdf(j/100))
81       y_ecdf_full <- cbind(y_ecdf_full , y_

```

```

      ecdf(j/100))
82      z_ecdf_full <- cbind(z_ecdf_full , z_
      ecdf(j/100))
83      svm_ecdf_full <- cbind(svm_ecdf_full ,
      svm_ecdf(j/100))
84      }
85
86      # combine
87      x_ecdf.out<-cbind(ann[i,4],ann[i,1:2],axis='x',
      x_ecdf_full)
88      y_ecdf.out<-cbind(ann[i,4],ann[i,1:2],axis='y',
      y_ecdf_full)
89      z_ecdf.out<-cbind(ann[i,4],ann[i,1:2],axis='z',
      z_ecdf_full)
90      svm_ecdf.out<-cbind(ann[i,4],ann[i,1:2],axis='
      svm',svm_ecdf_full)
91
92      ecdf.combined <- rbind(x_ecdf.out,y_ecdf.out,z_
      ecdf.out,svm_ecdf.out)
93
94      }
95      stopCluster(cl)
96      close(pb1)
97
98      return(ecdf.out)
99 }
```

Appendix G

Calculation of Fisher's g Statistic

Code given for example list of mean ODBA values (mean.odba). Processes FFT of values, then calculates and statistically assesses Fisher's g statistic.

```
1 library(data.table)
2 library(signal)
3 library(seewave)
4
5 temp.dat <- mean.odba
6
7 # normalise
8 norm_odba <- temp.dat - mean(temp.dat)
9
10 #Apply window
11 norm_odba <- fft_win*norm_odba
12
13 fft_odba<-fft(norm_odba)/length(norm_odba)
14
15 #Frequency axis
16 freq_axis_mean <- (seq(0,1, length.out = round(length(fft_odba)
    /2))*fn)
17
18 d.raw_odba <- abs(fft_odba[2:(length(fft_odba)/2)-2])
19 # remove nyquist frequency and 0 Hz (DC component)
20 a <- length(abs(fft_odba[2:(length(fft_odba)/2)-2]))
21 pval.out <- data.frame()
22
23 for ( j in 2:((length(fft_odba)/2)-2)){
24
25     fisherg_odba <- d.raw_odba[j]/sum(d.raw_odba)
26     b <- 1:floor(1/fisherg_odba)
```

```
27     pval.out <- rbind(pval.out,cbind(j, fisherg_odba ,as.
      numeric(sum(((−1)^(b−1))*exp(lchoose(a,b))*((1−(b*
      fisherg_odba))^(a−1)), na.rm = T))))
28 }
29
30 pval.out <- pval.out[!is.nan(pval.out[[3]])&!is.infinite(pval.
      out[[3]]) ,]
31
32 pval.out[,3]<-p.adjust(pval.out[,3], method = "bonferroni")
33
34 pval.out <- pval.out[order(−pval.out$fisherg_odba) ,]
35
36 firstinsig <- which(pval.out[,3]>pthresh)[1]
37
38 sigharm.mean <- pval.out[1:(firstinsig −1) ,]
39
40 sigharm.mean[1] <- freq_axis_mean[sigharm.mean[[1]]]*86400
41
42 sigharm.mean <- cbind(sigharm.mean[1],24/sigharm.mean[1] ,
      sigharm.mean[2:3])
43
44 names(sigharm.mean) <- c("CPD" ,"Period_Length" ,"Figher 's_g" ,"P_
      value")
45
46 sig.harmonic <- length(sigharm.mean[[1]])
```