Dog behaviour classification with movement sensors placed on the harness and the collar

Pekka Kumpulainen [a], Anna Valldeoriola Cardó [b], Sanni Somppi [b], Heini Törnqvist [b], Heli Väätäjä [c,d], Päivi Majaranta [c], Yulia Gizatdinova [c], Christoph Hoog Antink [e], Veikko Surakka [c], Miiamaaria V. Kujala [b,f], Outi Vainio [b], Antti Vehkaoja [a,*]

[a] Faculty of Medicine and Health Technology, Tampere University, P.O. Box 692, FI-33101, Tampere, Finland
[b] Department of Equine and Small Animal Medicine, University of Helsinki, P.O. Box 57, FI-00014, University of Helsinki, Finland
[c] Research Group for Emotions, Sociality, and Computing, Faculty of Information Technology and Communication Sciences, Tampere University, P.O. Box 100, FI-33014, Tampere, Finland
[d] Lapland University of Applied Sciences, Master School, Jokiväylä 11 B, 96300, Rovaniemi, Finland
[e] Biomedical Engineering (KIS*MED), TU Darmstadt, Magdalenenstraße 2A, 64289 Darmstadt, Germany
[f] Department of Psychology, Faculty of Education and Psychology, University of Jyväskylä, PO Box 35, FI-40014, University of Jyväskylä, Jyväskylä, Finland

ARTICLE INFO

ABSTRACT

Dog owners' understanding of the daily behaviour of their dogs may be enhanced by movement measurements that can detect repeatable dog behaviour, such as levels of daily activity and rest as well as their changes. The aim of this study was to evaluate the performance of supervised machine learning methods utilising accelerometer and gyroscope data provided by wearable movement sensors in classification of seven typical dog activities in a semi-controlled test situation. Forty-five middle to large sized dogs participated in the study. Two sensor devices were attached to each dog, one on the back of the dog in a harness and one on the neck collar. Altogether 54 features were extracted from the acceleration and gyroscope signals divided in two-second segments. The performance of four classifiers were compared using features derived from both sensor modalities. and from the acceleration data only. The results were promising; the movement sensor at the back yielded up to 91 % accuracy in classifying the dog activities and the sensor placed at the collar yielded 75 % accuracy at best. Including the gyroscope features improved the classification accuracy by 0.7–2.6 %, depending on the classifier and the sensor location. The most distinct activity was sniffing, whereas the static postures (lying on chest, sitting and standing) were the most challenging behaviours to classify, especially from the data of the neck collar sensor. The data used in this article as well as the signal processing scripts are openly available in Mendeley Data, https://doi.org/10.17632/vxhx934tbn.1.

1. Introduction

Accelerometers are used in animal science in various contexts varying from the estimation of energy expenditure (Crouter et al., 2006), assessment of behaviours in wildlife (Campbell et al., 2013; Halsey et al., 2011; Moreau et al., 2009; Nathan et al., 2012; Shepard et al., 2008; Wilson et al., 2006) to applications in veterinary medicine (Helm et al., 2016; Guillot et al., 2013; Little et al., 2016).

The number of consumer-targeted activity trackers available for dogs has increased in recent years and the market is expected to grow rapidly in the forthcoming years. One of the key factors for the growth is dog owners' increased interest and awareness towards dog wellbeing. (Grand View Research, 2018) Combined with a smartphone application, solutions typically visualize the data as daily total activity, type of activity (light or heavy), and collective behaviours, such as the amount of time the dog has spent for moving or resting during the day. There is evidence that existing activity trackers are feasible for evaluating simple canine behaviours, for example, differentiating between a sedentary activity and two intensities of physical activity (Yam et al., 2011). Depending on the sensitivity of the measurement unit, spontaneous activities of a dog, such as locomotion, postural change and movement of body in each posture, can be differentiated from accelerometer data

---

(Yamada and Tokuriki, 2000). However, dog owners and dogs could benefit from even more accurate and detailed analysis of motion and body postures of the dog. Detailed detection of dog's everyday activities would improve dog owners' understanding of particular dog behaviours and reactions, such as suffering from separation anxiety while alone at home or in a kennel. Extracting more detailed behaviours from accelerometer data also has the potential to be used as an index of wellbeing and health status of the animal, for example, by detecting stress and pain-related behaviours (Morrison et al., 2014a; Brown et al., 2010). Automatic behaviour distinction would also benefit behavioural research, where behaviours are traditionally measured by manual annotation of video recordings, which is labour intensive and time consuming and it would enable more detailed behavioural research for free-roaming wild animals (Rast et al., 2020).

Some activity loggers designed for humans have been commonly used for canine activity monitoring. These include "ActiGraph GT3X / GTX3+" by Yam et al., 2011 and Morrison et al., 2014b; as well as "Actical" by Hansen et al., 2007 and Olsen et al., 2016. These devices are primarily intended for data logging (recording raw accelerometer data) and they do not classify behaviour or motions of dogs automatically. In this study, we have developed and evaluated the accuracy of behaviour classification from ActiGraph data using conventional machine learning classifier approaches. In respect to accuracy of classification algorithms developed for dogs, Ladha et al. (2013) achieved a 68.6 % of global accuracy for differentiating 16 canine behaviours in naturalistic environments. den Uijl et al. (2017) showed that walk, trot, canter/gallop, eat, drink, and headshake behaviours could be classified with 95 % accuracy using a hierarchical "one vs. the rest" classifier. Additionally, seven canine activities such as sitting, and trotting were distinguished with an accuracy of more than 80 % by using 3-axis accelerometer and 3-axis gyroscope in (Gerencsér et al., 2013). Ferdinandy et al. (2020) obtained up to 60 % or 80 % accuracy depending on the cross-validation strategy in classifying eight behaviours and Ladha and Hoffman (2018a) achieved 86 % accuracy in detecting resting of the dog.

The placement of the movement sensor (e.g. neck or back) is an important factor in the movement analysis. Regarding versatility and practicality, the best placement for an activity sensor has been concluded to be ventral attachment to the neck collar, because this placement makes it possible to detect also behaviours that do not involve movement of the whole body, such as scratching and eating (Hansen et al., 2007). On the other hand, accelerometers attached at the back may be able to differentiate behaviours that devices attached to the collar cannot detect such as elevated walking velocity (Preston et al., 2012).

The aim of this study was to evaluate four commonly used classification algorithms for distinguishing seven dog behaviours using inertial sensor data recorded in semi-controlled test situations. The behaviours were galloping, lying on chest, sitting, sniffing, standing, trotting, and walking. Furthermore, as some behavioural phenomena may be better detectable from the dog's neck or back locations, we systematically examined how the device placement, either on the harness at the back or on the collar around the neck, affected the classification accuracy. Initial results of the study with smaller number of dogs and with only one sensor location (neck) have earlier been published in ACI (Animal Computer Interaction) conference (Kumpulainen et al., 2018).

2. Materials and methods

2.1. Test setup

The experiments were conducted at the University of Helsinki, Faculty of Veterinary Medicine. The study protocol was reviewed and accepted by the Ethical Committee for the Use of Animals in Experiments at the University of Helsinki (minutes 5/2017). All dog owners signed an informed consent before participating in the study. The attendants were free to cancel their participation at any time without giving a reason.

A total of 45 healthy, middle to large -sized pet dogs from 27 breeds participated in the study. The average age of the dogs was 4.9 years (range 1–9 years) and the average weight was 24.5 kg (range 13–41 kg). Table 1 shows detailed breed, weight and age statistics of the participating dogs. The dataset is described in more detail in a separate dataset article (Vehkaoja et al., 2021a) and is freely available in Mendeley Data (Vehkaoja et al., 2021b).

2.2. Test protocol

The tests were conducted in a dog sporting hall in a testing arena of 10m × 18m covered with artificial turf. The test sequence consisted of seven tasks where the owner was instructed to guide the dog accordingly. Three of the tasks were static tasks (i.e. *sitting, standing, lying down*) and four were dynamic tasks (i.e. *trotting, walking, playing, and treat-searching*), each task lasted for three minutes. The whole procedure was repeated after a short break while changing the order of the tasks. Dogs performed tasks sequentially, alternating between static and dynamic tasks. Treat search was always performed as the final task of the sequence and it consisted of searching small pieces of dry dog food spread on the ground (area of 4m × 4m) by sniffing.

Dogs wore two ActiGraph GT9X Link (ActiGraph LLC, Florida, USA) activity sensors including 3-axis accelerometer and 3-axis gyroscope sensors (sampling rate 100 Hz). One sensor was placed inside a tight pocket made of neoprene on the back belt of the dog's harness, referred to as the back sensor in this paper. The other sensor was attached tightly with an adhesive tape on the ventral side of the neck collar and is referred to as the collar sensor in this paper. Dogs were on the leash (1.5 m) and were led by their owners or the experimenter. The leash was connected to a separate collar that was placed closer to the dogs' body than the collar to which the sensor was attached. The owners were allowed to give food rewards and command their dogs through the entire test.

2.3. Behaviour annotation

The actual behaviour of the dogs during the assigned tasks were annotated using video recordings. The test procedure was recorded with

Table 1
Characteristics of the 45 dogs that participated in the study.

Breed	Number	Weight (kg)	Age (years)
Australian Kelpie	1	18	3
Beauce Shepherd	3	30.33 (28–35)	3 (3 – 3)
Belgian Shepherd	1	29	6
Belgian Shepherd Groenendael	1	20	5
Belgian Shepherd Malinois	1	25	3
Border Collie	4	16.5 (15–20)	3.75 (3–5)
Bouvier des Flandres	1	30	7
Bouvier des Ardennes	2	22.5 (22–23)	4.5 (4–5)
Bull Terrier (Miniature)	1	17	2
Crossbreed	4	16.25 (13–20)	4.5 (3–7)
Dutch Shepherd	2	24 (23–25)	3 (3–3)
English Springer Spaniel	1	25	4
Finnish Lapphund	1	26	5
Flat-Coated Retriever	1	28	4
German Shepherd	3	32.33 (30–35)	3 (3–3)
Golden Retriever	3	30 (23–39)	4.67 (4–5)
Hovawart	2	34.5 (28–41)	5 (5–5)
Labrador Retriever	3	30 (23–37)	3 (3–3)
Lagotto Romagnolo	1	14	7
Lapponian Herder	2	21 (20–22)	3 (3–3)
Mudi	1	16	7
Nova Scotia Duck Tolling Retriever	1	20	5
Smooth Collie	1	18	3
Spanish Water Dog	2	22.5 (20–25)	7 (7–7)
Standard Poodle	1	31	7
Hungarian Vizsla	1	25	2

Panasonic HDC-SD600 and Sony HDR-CX450 video cameras positioned on the opposite lateral walls and facing towards the testing arena. The post hoc annotation of the video recordings was done using the Observer XT 10.5 software (Noldus, The Netherlands). Only segments longer than one second were included in the annotation. Dynamic behaviours (i.e. Walking, Trotting, Galloping, Sniffing; Table 2) were only encoded if unambiguous, i.e. if there was only one obvious, continuous dynamic behaviour without the dog leaning towards the handler or pulling the leash, thus affecting the gait pattern or the body position. Galloping was annotated only during the play task and sniffing during the treat search task (see Table 2 for the ethogram). Static behaviours consisted of still postures (i.e. Lying on chest, Sitting, Standing) and annotated when limbs did not move and there was no physical contact between the handler and the dog, except if a treat was given.

2.4. Feature extraction and labelling

The raw time series data produced by the movement sensors were saved with ActiLife software (ActiGraph LLC, Florida, USA) and analysed offline with MATLAB R2018b (The MathWorks, Inc., Natick, MA USA).

The time series signals were segmented into two-second time windows with 50 % overlap. A total of 27 features per sensor type were calculated for each segment and used for classification of the behaviours. The same features were used for both accelerometer (A1 – A27) and gyroscope (G1 – G27) data. The descriptions of the resulting 54 features are given in Table 3.

The interpolated inverse empirical cumulative distribution function (ecdf) has been presented by Cox and Oakes (1984) and it has earlier been used for computing movement related features by Hammerla et al. (2013). The ecdf features are based on the cumulative distribution function $P_c(x) = P(X \le x)$ in a following way. Seven values of p_i, evenly distributed between 0 and 1, in each x, y and z axis were selected. For each p_i the value x_i for which $P(X \le x_i) = p_i$ was estimated by shape-preserving piecewise cubic interpolation. Fig. 1 left panel shows an example of cumulative distribution function of a normal distribution. The right panel shows two examples derived from two-second windows of accelerometer signal in x direction during walking and trotting.

The true behaviour classes (see Table 2) of the data segments were assigned according to the video annotations and synchronized to match the timestamps in the video annotations. A behaviour class was assigned as a label to a segment if a single annotated behaviour was occurring a

Table 2
Ethogram of the behaviours included in the statistical analyses.

Behaviour	Description
Galloping	3- or 4-beat gait where the dog lifts and puts down both front and rear extremities in a coordinated manner, in 1−2-3-beat gait (canter) or in 1−2-3−4 beat gait (gallop). All four extremities are simultaneously in the air at some point in every stride. Galloping occurred only during Playing task.
Lying on chest	The dog's torso is touching the ground and hips are in the same level as shoulders. The dog can change balance point without using limbs.
Sitting	The dog has four extremities and rump on the ground. The dog can change balance point from central to hip or vice versa.
Sniffing	The dog has its head below its back line and moves its muzzle close to the ground. The dog walks, stands or performs another slow movement, but its chest and bottom do not touch the ground. Taking food from the ground and eating it can be included (eating was not coded separately).
Standing	The dog has the four extremities on the ground, without the dog's torso touching the ground.
Trotting	2-beat gait where the dog lifts and puts down extremities in diagonal pairs at a speed faster than walking.
Walking	4-beat gait where the dog moves extremities at slow speed, legs are moved one by one in the order: left hind leg, left front leg, right hind leg, and right front leg. The dog moves straight forward or at maximum in 45 degrees angle.

Table 3
Description of the features calculated for 2-second segments of time series movement sensor data. "A" refers to accelerometer and "G" to gyroscope.

Feature code	Feature description
A1, G1	Total activity: sum of standard deviation in all three axis
A2, G2	Position offset: Euclidean distance from the robust mean obtained while the dogs were standing still
A3, G3	The number of mean crossings, the sum of x, y and z axis
A4 – A6, G4 – G6	The mean value of each axis; x, y, z
A7 – A27, G7 – G27	Interpolated inverse empirical cumulative distribution function (ecdf): seven values for each axis, a total of 21 features for each sensor type

minimum of 75 % of the segment. This was done in order to increase the amount of included data, especially for the behaviours that typically occur in short durations, namely sniffing and galloping. The data of both test sequences were included for 17 dogs and only one test sequence for the remaining 28 dogs. The reason for including only one of the test sequences for the 28 was the challenges faced in reliable synchronization of the data. The 62 tests from the 45 dogs provided 54,594 instances of labelled data. Table 4 shows how the segments were distributed between the behaviours.

### 2.5. Feature selection and classification

All 54 features were Z-score normalised to zero mean and unit variance. Due to the high number of features, the feature selection was performed in two parts. First, weights of importance were calculated for each feature by the ReliefF algorithm (Robnik-Sikonja and Kononenko, 2003) based on the k-nearest neighbour approach. In order to reduce computational costs, only the most important features were used in the subsequent forward feature selection. Here, features are added one at a time in an order in which they best improve the classification accuracy. This is continued until additional features provide no improvement. The forward selection was performed for each of the four classifiers. Both, the forward selection and the final classification results were computed using Leave-One-Dog-Out cross validation. Thus, all the data of each individual dog were left out at a time as a test set and the classifiers were identified using the rest of the data. Leave one subject out cross validation has been proposed over the random splitting of the data in which case the data from the same individual easily ends up in both training and validation data sets producing overly goods classification results. This was recently shown by Ferdinandy et al. (2020) in the context of animal behaviour classification.

The whole feature selection procedure was repeated using only the accelerometer features to verify the benefit of the information provided by the gyroscopes.

The four classifiers in this study were linear and quadratic discriminant analysis classifiers (LDA and QDA, respectively), a support vector machine (SVM) classifier with gaussian kernel, and a classification tree (Duda et al., 2000). The regularisation of the classification tree was controlled by the number of cuts allowed in the tree. The number was optimised with the cross validation using the most significant features given by the relief weights and that value was used in the forward feature selection thereafter. These four classifiers were chosen based on their popularity in basic machine learning studies and their simple structure and low computational cost that would enable their integration also into a power constrained embedded measurement platform in the future. The predicted class was obtained as the highest probability class proposed by the classifier referred to as overall accuracy in (Ferdinandy et al., 2020).

2.6. Statistical analysis

The classification accuracies reported in the results were calculated
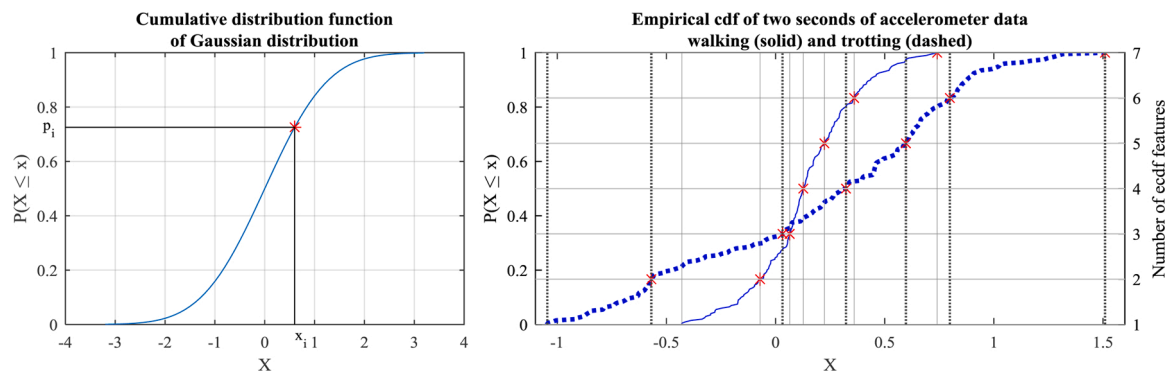
Fig. 1. Illustration of ecdf feature calculation. Left panel: one point from normal distribution at $P(X \leq x) = p_i$. Right panel: example of actual data evaluated at seven p_i values between 0 and 1.

Table 4
The number of segments assigned to each behaviour.

Behaviour	Count	% of total
Galloping	776	1.4
Lying on chest	11,062	20.3
Sitting	10,824	19.8
Sniffing	9331	17.1
Standing	9552	17.5
Trotting	6646	12.2
Walking	6403	11.7

as the averages of the percentages of correctly predicted behaviour class in all folds of the cross validation. Differences between the classifiers were tested at $p = 0.05$ level using t-test, which assumes normal distribution. The normality of the classification rate distribution of each classifier was tested by Kolmogorov-Smirnov test. The statistical analyses were conducted by MATLAB Statistics and Machine Learning Toolbox R2018b.

3. Results

The relief feature weights were calculated with six values of k: {3, 5, 9, 13, 17, 21}. For the feature set with both accelerometer and gyroscope data, the features included in the top 20 weights by any of the k values (23 features for the back and 22 for the collar sensor) were used in the forward selection phase. When evaluating the accelerometer data alone, the features were selected in the same way but choosing the features included in the top 15 weights by any of the k values (17 features for the back and 15 for the collar sensor).

The optimal numbers of cuts acquired for the classification trees were 124 for the back and 121 the collar sensor for all features, and 168 for the back and 54 for the collar sensor for the accelerometer features only. The features selected by the forward selection and the final cross-validated classification accuracies are presented in Table 5.

In all cases, more accelerometer features and fewer gyroscope features were selected. Considering also the gyroscope features in the classification provided better accuracy with all classifiers and both sensor locations. Therefore, all the detailed results presented below are presented for the cases with the selected accelerometer and gyroscope features.

Confusion matrices of the classification results are presented in Fig. 2. The most challenging behaviours to classify with the data of the collar sensor were the static postures: lying on chest, sitting, and standing. Lying on chest was most often mixed with the other static postures. The back sensor provided similar results but there the difference in the accuracies between the classes was not so clear. The activities that involved movement were generally classified very accurately, mostly higher than 90 %, sniffing being the most distinct behaviour in

Table 5
Classification accuracies and the selected features for each classifier and both feature set scenarios separately for sensors located on the back and the neck.

Classifier	Measurement	Back sensor		Collar sensor	
		Accuracy %	Features	Accuracy %	Features
LDA	A + G	89.0	9: A{3, 15, 18, 19, 25, 26} G{6, 22, 26}	71.9	10: A{5, 6, 15, 18, 19, 22, 26} G{6, 22, 27}
	A only	87.8	8: A{3, 15, 18, 19, 23, 24, 25, 26}	71.1	9: A{3, 5, 6, 15, 16, 18, 19, 23, 26}
QDA	A + G	91.1	9: A{2, 8, 17, 19, 26} G{5, 6, 22, 25}	70.6	11: A{5, 6, 15, 17, 19, 22, 23, 26} G{6, 22, 27}
	A only	90.2	5: A{8, 11, 17, 19, 26}	69.5	6: A{5, 6, 15, 19, 23, 26}
SVM	A + G	91.4	8: A{6, 8, 15, 18, 19, 24} G{22, 26}	75.6	10: A{3, 5, 6, 15, 18, 19, 22, 26} G{22, 27}
	A only	90.5	6:A{6, 8, 15, 18, 24, 26}	74.9	7: A{3, 5, 6, 15, 17, 19, 26}
Tree	A + G	91.0	10: A{5, 8, 9, 15, 18, 19} G{5, 6, 22, 25}	72.4	7: A{3, 5, 6, 17, 19, 26} G{27}
	A only	88.7	6: A{6, 8, 15, 17, 18, 25}	69.8	9: A{3, 5, 6, 15, 17, 19, 23, 24, 26}

both sensor locations and almost all classifiers. Walking was classified with slightly worse accuracy with the neck sensor and was mixed with static behaviours.

The classification results were also calculated separately for each dog to study the differences in the accuracy between individual dogs. The results are shown as boxplots in Fig. 3. The box contains the interquartile range between the 25th and 75th percentiles. The notch around the median covers 95 % confidence limits. The whiskers extend to the extreme data point up to 1.5 times the interquartile range from the box. Individual points outside the maximum length of the whiskers are marked with red crosses. Marking the confidence interval of the median value allows visual inspection of the statistically significant differences between the results obtained with different classifiers. If the notches of the results of two classifiers do not overlap, the medians are statistically different.

As seen in Fig. 3, the results vary considerably between the dogs. The highest classification accuracies for some individual dogs are above 99

Fig. 2. Confusion matrices of the four classifiers. The true classes are in the rows and the predicted classes in the columns.
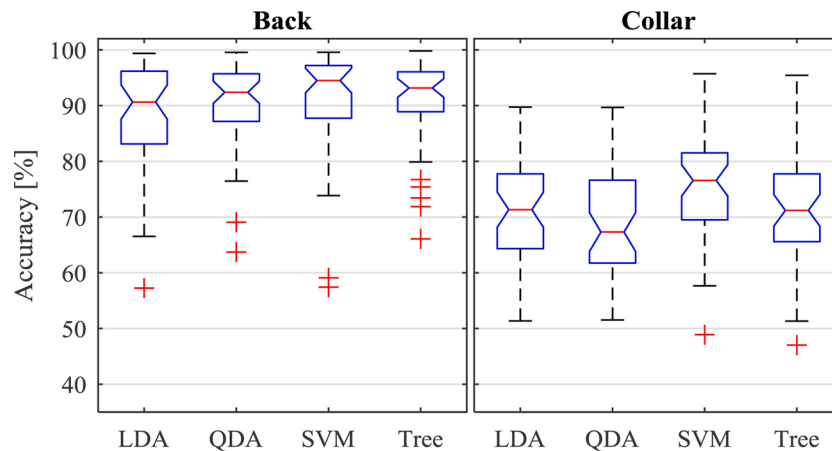

Fig. 3. Boxplots of the accuracies of the four classifiers (LDA, QDA, SVM, Tree). Each box contains the classification accuracies of the 45 dogs.

% for all classifiers with the sensor attached on the harness. The collar sensor reaches 90 % accuracy for some dogs. However, the lowest accuracies are between 47 % and 66 %.

For the back sensor, differences in the distributions of the results of individual dogs obtained with different classifiers were not statistically significant. For the collar sensor, SVM gave significantly better results than LDA and QDA. LDA and QDA were not significantly different from each other. The results obtained with the classification tree were not significantly different from any of the other classifiers. Table 6 provides all pairwise p-values of the dog-wise accuracy distributions of different classifiers.

4. Discussion

Consumer-targeted dog activity meters are widely available on the market, but the information they give for the dog owners is rather limited. The aim of this study was to evaluate the performance of

Table 6
p-values of the pairwise t- tests between the classifiers. The values for the back sensor are in the lower left and the collar sensor in the upper right part. Bold-faced values show statistical significant difference ($p < 0.05$) between the classifiers in dog-wise classification accuracy distributions.

	Collar sensor			
	LDA	QDA	SVM	Tree
LDA		0.428	**0.049**	0.786
QDA	0.289		**0.008**	0.318
SVM	0.208	0.752		0.112
Tree	0.241	0.886	0.856	
	Back sensor			

activity classification with two movement sensors located in the collar and the harness. The sensors provided both accelerometer and gyroscope data for classifying seven activities of dogs in a semi-controlled test situation. The results were promising, yielding up to 91 %

classification accuracy using the data of both sensor types from the sensor at the back. Including the gyroscope data in addition to the accelerometer data provided 0.7 %–2.6 % better accuracy with all four classifiers and both sensor locations.

Sniffing was the most distinct behaviour, resulting in 99.2 % accuracy with the collar sensor and 98.0 % with the back sensor. This is in contrast to an earlier finding by Ladha et al. (2013) who found that walking and running (of the behaviours shared with this work) had better classification performance than sniffing. However, all classes included in a classification task affect the performance of each individual class, which may explain the difference. The most challenging task for the classifiers was differentiation between the static postures with the collar sensor, namely lying down, sitting, and standing. From those time segments where the dog was lying down, only 28%–45% were classified correctly, and the rest were classified mainly as either sitting or standing. However, considering the minor orientation change in the dogs' neck and back during these tasks, the mixing of these postures is rather logical. den Uijl et al. (2017) report similar results, while they had the lowest specificity for sleep behaviour. However, they did not classify standing, sitting, and lying down separately, but a combined class as static/inactive, which makes the classification task significantly easier. In general, the results showed that activity monitors on the back yielded better results for classification than attaching an activity monitor on the neck collar. The results of the back sensor are in line with those by Gerencsér et al. (Gerencsér et al., 2013) who also used a sensor at the back.

The placement of the sensor has been shown to affect the amount of measured activity in previous studies (Hansen et al., 2007; Preston et al., 2012). In addition, tightness of the attachment may affect the accuracy. For example, Preston et al. (2012) found that accelerometers attached tightly to the back detected behaviours more accurately than devices attached loosely to the back. In our study, the attachment technique of the sensors was dependent on the placement: the collar sensor was attached to the collar with adhesive tape, but the back sensor was inserted in a neoprene pocket. This might be one reason for the difference between the results for the collar and the back sensors. Although the pocket was tight, the movements of the sensor could be affected by it, as has been earlier concluded by Martin et al. (Martin et al., 2016). In our study, a likely reason for the difference in the classification accuracy between the collar and the back sensors is that the distinct static postures result in more significant changes in the orientation of a sensor attached to the back than to the collar. Another likely reason is that, because the orientation of the neck sensor may change slightly due to turning of the collar around the neck, the orientations recorded in the static tasks may overlap. Following this, the potential rotation of the collar needs to be considered and compensated as has also been concluded in (Ladha et al., 2018b).

Walking was classified with high accuracy by the back sensor but mixed with static postures with the neck sensor. It should be noted that walking in controlled test situation was rather different from real life, where a dog rarely purely walks slowly in leash, but rather mixes walk and pace gaits.

As the classification accuracies vary considerably between individual dogs and the manually annotated video data from every dog does not necessarily contain the same amount of all behaviours. Thus, the question arises whether the proportions of the behaviours have an effect on the individual classification results. However, no significant correlation was found between the class proportions and the classification rate, except for a negative correlation for walking behaviour with QDA and SVM classifiers on the back sensor. However, walking was one of the best classified behaviours in this study, so this correlation must have happened by a pure chance.

We tested only medium to large sized dog breeds to get a homologous participant group with smaller variability. For example, Ladha et al. (2013) found better global accuracy for differentiating behaviours in small and medium sized dogs than in large ones. In the future, the

accuracy of the current behaviour classifications should be tested in a larger dog population that includes also smaller dogs as well as dogs with deviant body structure (e.g. short legs). As our test setup was also relatively controlled, future studies should also be done with natural behaviours of dogs moving freely in their familiar environment. Although the sensor was attached to a different collar than the leash, the handler's behaviour may have affected the sensor, especially in those situations when the dog moved slower than the handler did. It would be ideal if the leash could be attached to the harness and the sensor to the collar (or vice versa), but as we aimed to test both sensor placements simultaneously, this was not an option. In future studies, for reliable behaviour detection, the device should always be positioned at exactly the same orientation, or the orientation should be recalibrated after each time it is attached (as was done in the present study) as well as each time it may have been shifted.

SVM classifier provides the best results for both sensors, but the difference in performance is statistically significant only for the collar sensor. In all cases, the classification rates have high deviation between individual dogs. Thus, for practical purposes, tuning the classifier for each individual would be beneficial from the accuracy point of view but may not be necessarily feasible in practice. Including the gyroscope data in the classification lowers the misclassification rates. However, for practical embedded products, it is a compromise whether the improvement is worth the added complexity and decreased battery lifetime.

5. Conclusion

Our current results suggest that behaviour classification was more successful from the movement sensor attached to the harness at the back of the dog rather than on the neck collar. In particular, static behaviours of sitting, standing, and lying down were hard to differentiate with the sensor attached to the collar. Positioning may comprise a challenge for the usability of activity monitors for differentiating behaviours in real life. Attaching the sensor to the collar is convenient for the dog and the owner, but if it compromises differentiation of resting from other sedentary behaviours as concluded in (den Uijl et al., 2017), it can lead to misleading conclusions in cases where rest behaviour is used, for example, as an indicator for a dog's pain or stress level.

Our current results are promising in terms of development of practical methods for automatically gaining information on dog behaviour. This type of more accurate information can be useful in supporting the owner in gaining overall understanding of a dog's daily life, assessment of health or sickness, and functioning of medication, in particular for dogs suffering from chronic illness. In the future, the technique could be developed further to identify behaviour problems, their causes, their treatment as well as the effectiveness of the treatment and assess the issues and changes in overall welfare based on the data. Furthermore, our present results pave the way for developing solutions to associate the activity to the affective state of the dog, to support a more comprehensive assessment of dog welfare.

Data availability

The annotated datasets generated and used in this study as well as the framework description and scripts used in data analysis are available in a repository hosted by Mendeley Data. The DOI for the dataset is: https://doi.org/10.17632/vxhx934tbn.1. The data is described in more

detail in the related Data in Brief article: Movement Sensor Dataset for Dog Behavior Classification.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

References

Brown, D.C., Boston, R.C., Farrar, J.T., 2010. Use of an activity monitor to detect response to treatment in dogs with osteoarthritis. J. Am. Vet. Med. Assoc. 237 (1), 66–70. https://doi.org/10.2460/javma.237.1.66.

Campbell, H.A., Gao, L., Bidder, O.R., Hunter, J., Franklin, C.E., 2013. Creating a behavioural classification module for acceleration data: using a captive surrogate for difficult to observe species. J. Exp. Biol. 216 (24), 4501–4506. https://doi.org/10.1242/jeb.089805.

Cox, D.R., Oakes, D., 1984. Analysis of Survival Data. Chapman & Hall, London. https://doi.org/10.1201/9781315137438.

Crouter, S., Churilla, J., Bassett, D., 2006. Estimating energy expenditure using accelerometers. Eur. J. Appl. Physiol. 98, 601–612. https://doi.org/10.1007/s00421-006-0307-5.

den Uijl, I., Gómez Álvarez, C.B., Bartram, D., Dror, Y., Holland, R., Cook, A., 2017. External validation of a collar-mounted triaxial accelerometer for second-by-second monitoring of eight behavioural states in dogs. PLoS One 12 (11) https://doi.org/10.1371%2Fjournal.pone.0188481.

Duda, R.O., Hart, P.E., Stork, D.G., 2000. Pattern Classification. Wiley-Interscience.

Ferdinandy, B., Gerencsér, L., Corrieri, L., Perez, P., Újváry, D., Csizmadia, G., Miklósi, Á., 2020. Challenges of machine learning model validation using correlated behaviour data: evaluation of cross-validation strategies and accuracy measures. PLoS One 15 (7), e0236092. https://doi.org/10.1371/journal.pone.0236092.

Gerencsér, L., Vasarhelyi, G., Nagy, M., Vicsek, T., Miklosi, A., 2013. Identification of behaviour in freely moving dogs (Canis familiaris) using inertial sensors. PLoS One 8 (10), e77814. https://doi.org/10.1371/journal.pone.0077814.

Grand View Research, 2018. Pet Wearable Market Size, Share & Trends Analysis Report by Technology, by Application (Identification & Tracking, Behavior Monitoring, Facilitation, Safety, Medical Diagnosis), and Segment Forecasts,2018 – 2025. Market Research Report.

Guillot, M., Moreau, M., Heit, M., Martel-Pelletier, J., Pelletier, J.P., Troncy, E., 2013. Characterization of osteoarthritis in cats and meloxicam efficacy using objective chronic pain evaluation tools. Vet. J. 196, 360–367. https://doi.org/10.1016/j.tvjl.2013.01.009.

Halsey, L.G., Shepard, E.L.C., Wilson, R.P., 2011. Assessing the development and application of the accelerometry technique for estimating energy expenditure. Compara. Biochem. Physiol. Part A 158, 305–314. https://doi.org/10.1016/j.cbpa.2010.09.002.

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. https://doi.org/10.1145/2493988.2494353.

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. https://doi.org/10.2460/ajvr.68.5.468.

Helm, J., McBrearty, A., Fontaine, S., Morrison, R., Yam, P., 2016. Use of accelerometry to investigate physical activity in dogs receiving chemotherapy. J. Small Anim. Pract. 57, 600–609. https://doi.org/10.1111/jsap.12587.

Kumpulainen, P., Valldeoriola, A., Somppi, S., Törnqvist, H., Väätäjä, H., Majaranta, P., Surakka, V., Vainio, O., Kujala, M.V., Gizatdinova, Y., Vehkaoja, A., 2018.

December. Dog activity classification with movement sensor placed on the collar. Proceedings of the Fifth International Conference on Animal-Computer Interaction 1–6. https://doi.org/10.1145/3295598.3295602.

Ladha, C., Hoffman, C.L., 2018a. A combined approach to predicting rest in dogs using accelerometers. Sensors 18 (8). https://doi.org/10.3390/s18082649.

Ladha, C., Hammerla, N., Hughes, E., Olivier, P., Ploetz, T., 2013. Dog's life: wearable activity recognition for dogs. UbiComp' 13 Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing 415–418. https://doi.org/10.1145/2493432.2493519.

Ladha, C., Belshaw, Z., O'Sullivan, J., Asher, L., 2018b. A step in the right direction: an open-design pedometer algorithm for dogs. BMC Vet. Res. 14 https://doi.org/10.1186/s12917-018-1422-3, 107.

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. J. Exp. Orthop. 3, 17 https://dx.doi.org/10.1186%2Fs40634-016-0053-5.

Martin, K.W., Olsen, A.M., Duncan, C.G., Duerr, F.M., 2016. The method of attachment influences accelerometer-based activity data in dogs. BMC Vet. Res. 13 (1) https://doi.org/10.1186/s12917-017-0971-1.

Moreau, M., Siebert, S., Buerkert, A., Schlecht, E., 2009. Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behavior. Appl. Anim. Behav. Sci. 119 (3–4), 158–170. https://doi.org/10.1016/j.applanim.2009.04.008.

Morrison, R., Reilly, J.J., Penpraze, V., Pendlebury, E., Yam, P.A., 2014a. 6- month observational study of changes in objectively measured physical activity during weight loss in dogs. J. Small Anim. Pract. 55, 566–570. https://doi.org/10.1111/jsap.12273.

Morrison, R., Penpraze, V., Greening, R., Underwood, T., Reilly, J.J., Yam, P.S., 2014b. Correlates of objectively measured physical activity in dogs. Vet. J. 199 (2), 263–267. https://doi.org/10.1016/j.tvjl.2013.11.023.

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. J. Exp. Biol. 215, 986–996. https://doi.org/10.1242/jeb.058602.

Olsen, A.M., Evans, R.B., Duerr, F.M., 2016. Evaluation of accelerometer inter-device variability and collar placement in dogs. Vet Evid. 1 (2) https://doi.org/10.18849/ve.v1i2.40.

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. Res. Vet. Sci. 93 (1), 412–416. https://doi.org/10.1016/j.rvsc.2011.08.005.

Rast, W., Kimmig, S.E., Giese, L., Berger, A., 2020. Machine learning goes wild: using data from captive individuals to infer wildlife behaviours. PLoS One 15 (5), e0227317. https://doi.org/10.1371/journal.pone.0227317.

Robnik-Sikonja, M., Kononenko, I., 2003. Theoretical and empirical analysis of ReliefF and RReliefF. Mach. Learn. 53, 23–69. https://doi.org/10.1023/A:1025667309714.

Shepard, E.L., Wilson, R.P., Quintana, F., Laich, A.G., Liebsch, N., Albareda, D.A., Halsey, L.G., Gleiss, A., Morgan, D.T., Myers, A.E., Newman, C., 2008. Identification of animal movement patterns using tri-axial accelerometry. Endanger. Species Res. 10, 47–60. https://doi.org/10.3354/esr00084.

Vehkaoja, A., Somppi, S., Törnqvist, H., Valldeoriola Cardó, A., Kumpulainen, P., Väätäjä, H., Majaranta, P., Surakka, V., Kujala, M.V., Vainio, O., 2021a. Description of Movement Sensor Dataset for Dog Behavior Classification submittedData in Brief.

Vehkaoja, A., Somppi, S., Törnqvist, H., Valldeoriola Cardó, A., Kumpulainen, P., Väätäjä, H., Majaranta, P., Surakka, V., Kujala, M.V., Vainio, O., 2021b. Description of movement sensor dataset for dog behavior classification. Mendeley Data. https://doi.org/10.17632/vxhx934tbn.1 [dataset].

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 cormorant. J. Anim. Ecol. 75, 1081–1090. https://doi.org/10.1111/j.1365-2656.2006.01127.x.

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. J. Small Anim. Pract. 52 (2), 86–91. https://doi.org/10.1111/j.1748-5827.2010.01025.x.

Yamada, M., Tokuriki, M., 2000. Spontaneous activities measured continuously by an accelerometer in beagle dogs housed in a cage. J. Vet. Med. Sci. 62 (4), 443–447. https://doi.org/10.1292/jvms.62.443.