Dog's Behaviour Classification Based on Wearable Sensor Accelerometer Data

Anniek Eerdekens
IT, Ghent University
WAVES-IMEC
Ghent, Belgium
Anniek.Eerdekens@ugent.be

Arne Callaert
IT, Ghent University
WAVES-IMEC
Ghent, Belgium

Margot Deruyck
IT, Ghent University
WAVES-IMEC
Ghent, Belgium
Margot.Deruyck@ugent.be

Luc Martens
IT, Ghent University
WAVES-IMEC
Ghent, Belgium
Luc1.Martens@ugent.be

Wout Joseph
IT,Ghent University
WAVES-IMEC
Ghent, Belgium
wout.joseph@ugent.be

***Abstract*—Sensor-based behavioral detection and classification can improve dog health and welfare. Since continuous monitoring is required, an energy-efficient solution is needed. The number of logging axes, sampling rate, and selected features of accelerometer data not only have a significant impact on classification accuracy in activity recognition but also on the sensor's energy needs. Three models are designed for detecting dog's activities namely, a Random Forest classifier (RF), a Convolutional Neural Network (CNN) and a hybrid CNN, i.e. a CNN fused with statistical features that retain knowledge about the global time series form. The models are validated using an experimental dataset consisting of six different dogs performing in eight different activities i.e. lying, sitting, standing, walking, running, sprinting, eating and drinking. The results indicate that using neck and chest accelerometer data sampled at 10 Hz is sufficient for high overall classification accuracies (96.44%) for the three models. The hybrid CNN is capable of excellent performance, detecting nearly 97.87% of the behaviours at 10 Hz with a class accuracy of 80% or higher.**

***Index Terms*—Resampling, dogs, accelerometer, convolutional neural networks, random forest classifier, behaviour classification, internet-of-animals.**

I. Introduction

Tracking a dog's movements may disclose critical information about its health and welfare. Accurate data collection permits objective assessment of a dog's fitness status based on daily activities, food habits, and sleeping patterns [1]. For extended periods of time, it is difficult to supervise dogs in natural outdoor surroundings or multi-room structures. Sensors that monitor animal behavior automatically are an excellent solution for these purposes [2]. Additionally, machine learning enables novel methods for detecting potential health concerns early on by detecting indicators concealed in dogs' actions that are not obvious to the human eye. For instance, pet activity monitors have been shown to be effective in identifying and diagnosing pruritis [3], as well as in predicting early obesity [4].

Numerous sensor-based products are already available on the market that collect biometric data in a variety of ways [5]. They are often worn on the pet's collar and employ an accelerometer (sometimes in conjunction with a gyroscope, magnetometer, GPS, or other sensors) to accurately estimate the pet's activity level, step count, and distance traveled. Recent improvements in machine learning have enabled pet activity monitors to go beyond predicting general activity levels. Not only can one predict when and how long a pet will engage in basic behaviors like walking, running, lying down, or resting, but also a broader variety of actions such as drinking, feeding, scratching, and head shaking [6].

Because dog activity must be tracked over an extended length of time, the wearable device's energy consumption is an important parameter. Battery life can be increased by reducing sample rates, logging fewer accelerometer axes, or reducing the amount of features [7]. A recent study examining cows' laying, standing, and feeding behavior revealed that the use of fewer logging axes had little effect on the categorization performance [8]. However, several studies have demonstrated that reducing the sample rate has a detrimental effect on the performance of traditional machine learning tools such as Support Vector Machines, Random Forest, and Naive Bayes classifiers [9], [10]. Moreover, when sheep behaviors were monitored at a sample rate of 16 Hz versus 32 Hz, the capacity of machine learning algorithms to identify standing, walking, and lying behavior patterns increased [9]. A well-informed choice of the optimal sensor measurement settings, such as sampling rate, logging axis, and feature selection, will result in considerable reductions in the energy consumption of wearables. While there are studies in the literature that examine these features of accelerometer data for human activity classification, there are none in dog tracking that examine the effect of these variables on behavior recognition accuracy concurrently.

Three supervised machine learning (ML) classifiers – a Convolutional Neural Network (CNN), a hybrid CNN, and a Random Forest (RF) – are used in this study to generate and compare training models for predicting different activity types based on accelerometer data. Hybrid CNNs, which combine a CNN for local feature extraction with simple statistical features (mean, standard deviation, etc.) that preserve information about the global form of time series, outperform baseline approaches and achieve state-of-the-art performance on major activity classification tasks while reducing computational cost [11]. A CNN has the advantage of automatically extracting features because to its low computing complexity, whereas RFs require manual feature extraction. However, Random Forest analysis is a prominent machine learning approach that has been effectively applied to classifying behaviors in a number of animals using accelerometer data, but not yet to detecting dog behaviors [12].

To summarize, the novelty and major contributions of this paper are the following: for the first time (i) a performance assessment of three machine learning algorithms detecting eight distinct behaviors from six different dogs is carried out simultaneously with determination of the (ii) optimal sampling rate (5 Hz, 10 Hz, 12.5 Hz, 25 Hz and 50 Hz) and (iii) necessary number of accelerometers.

II. Materials and Method

A. Animals and study site

Measurements were carried out during a training session for sheep herding in Vorselaar, Belgium with five different Border Collie dogs. In addition, a second data collection campaign was performed, mainly focused on the detection of eating. Detailed information about the selected subjects is presented in Table I.

TABLE I
PARTICIPATING DOGS WITH BREED CLASS, GENDER, AGE AND WEIGHT.

No.	Breed class	Gender	Age	Weight (kg)
1	Border Collie	Male	2	17
2	Border Collie	Male	2	22
3	Border Collie	Male	11	24
4	Border Collie	Male	6	24
5	Border Collie	Male	5	20
6	Golden Retriever	Female	5	40

Fig. 1. Left: Position and orientation (X, Y, and Z axes) of the sensors. Right: top view of orientation of front and top sensor.

B. Data collection procedure

Sensor data are collected utilizing Axivity AX6 inertial movement sensors (Axivity Ltd, Newcastle, United Kingdom) with a dimension of 23 $\times$ 32.5 $\times$ 8.9 mm and a weight of 11 g. These log data with configurable sampling rates ranging from 12.5 Hz to 3200 Hz are powered by a 250 mAh lithium-polymer battery which is rechargeable via USB connection. Acceleration and angular velocity are measurable on x-, y-, z-axes with a maximum sensitivity of $\pm 16g$ $[g = m/s^2]$ and ± 2000 $[dps]$, respectively. The AX6 OMGUI Configuration and Analysis Tool, an open source application, is used to set up and configure the AX6 sensors for data recording. Data is registered on an embedded 1024 MB memory.

Each dog was fitted with two sensors. As indicated in Fig. 1, the first sensor was mounted to the top of a dog harness, while the second was attached to the front of the dog harness. The sensors were securely fastened with tape, with minimal room for vibration, slip or twist; to ensure that only the dog's motions are captured. All devices were fixed using the orientation illustrated in Fig. 1 with the three colored axes indicating the orientation of the accelerometer axes.

In this study, the device sensors were set to collect data from all the considered activities as listed in Table II with their descriptive definitions. Observations of the activities of the dogs were made with video recordings simultaneously with the collection of sensor data. After recording, the data was transferred to a computer via USB and stored in a Continuous Wave Accelerometer format. All the data is labelled by a trained observer based on the video recordings using ELAN, which is a specialized video annotation software, since it is difficult to use direct observation in conjunction with training of the dog [13].

The class discretization of each sample was determined by looking at the class labels of the individual data samples within each window. If all data samples within a window shared the same activity class, the collective label for the entire window was set to that particular activity class. Windows that contained data points with more than one activity class label were dropped. Following frames overlap by 50%, which is consistent with the state-of-the-art in sliding window activity recognition [15]. Finally, the data is normalized between 0 and 1 per batch of measurements along every axis with the normalization function of the Scikit-learn Python library [16]. In Table II the number of 2.4 s instances for all dogs and the class proportions are presented. As can be concluded from this table, the activity running is the most present with a class proportion of 53% and drinking is the least present behavior in the dataset with only 21 instances.

C. Machine learning models

To categorize dog behavior, a two-phase approach is devised since it was found that this resulted in higher classification accuracies. The first phase classified six dog movements with an optimal 2.4 s time window i.e., sitting, walking, running, sprinting, eating and drinking and one superclass 'steady' containing all standing and lying movements. Afterwards, a second classifier with an optimal 1.2 s time window is applied to determine whether the steady state of the dog is standing or lying. Both classification phases employ one of the three following models:

1) CNN: A multilayer convolutional network with two convolutional layers and a max-pooling layer is employed. The output of the last fully connected layer is sent into a softmax layer that provides a distribution over class labels.

The class proportions of the eight studied activities are not spread uniformly, therefore the class-wise weights are balanced in the training phase [17]. It penalizes loss more severely for underrepresented classes, allowing the model to better adjust to minority class characteristics.

The first convolutional layer filters the $n \times 6 \times 1$ input accelerometer data with 64 kernels of size 3 $\times$ 1 and a step size of 1. This layer uses L2 regularization with a weight decay coefficient of 0.01 [18]. To ensure the output is the same length as the input, the first convolutional layer uses zero padding. Following that, a max-pooling operation is carried out. With 16 kernels of size 5 $\times$ 1 and stride 1, the second convolutional layer filters the first convolutional layer's (pooled) output. Both layers include an activation layer composed of Rectified Linear Units (ReLUs) with a dropout of 0.55 [19]. The Adam optimizer is used for back propagation training. The training is done across 400 epochs, with a 60 epoch cutoff for halting the training [20]. The validation set is used to assess the model's performance after training.

2) Hybrid CNN: For the hybrid CNN the same convolutional network as the for CNN model is used but the max-pooling layer output is then flattened and fused with additional time domain features: mean, maximum, minimum, signal magnitude average and standard deviation.

3) RF: A random forest algorithm is a flexible ensemble classifier that is characterized by merging numerous decision trees trained on the training set. To classify a new instance from an input feature vector, it descends for each tree in the forest. Each forest forecasts a class termed a vote. The random forest chooses the class with the most votes. Random forest algorithms have the advantage of producing consistent and predictable outputs even without hyperparameter adjustment [21]. The input vector of the RF consists of a subset of selected features calculated from raw accelerometer data. The features employed in this study were shown to be important for categorizing animal activities using accelerometer data [8], [22]. Those features include thirteen time domain features namely, average (Average x, Average y, Average z), maximum (Max x, Max y, Max z), minimum (Min x, Min y, Min z), signal magnitude average (SMA) and standard deviation (Std x, Std y, Std z).

TABLE II
DESCRIPTION OF THE OBSERVED ACTIVITIES [14].

Observed activities	Description	Samples (# and %)
Lying	The movement between when the belly makes contact with the floor and when it no longer does.	372 (18%)
Sitting	The movement between when the bump makes contact with the floor and when it leaves. In comparison to lying down, the belly should not come into contact with the floor.	32 (2%)
Standing	Movement is described as the dog stands on all four legs and does not change positions.	28 (1%)
Walking	Movement is described as the dog moving forward at a pace comparable to the owner walking the dog.	75 (4%)
Running	Includes galloping and trotting actions that result in the dog moving ahead.	1095 (53%)
Sprinting	Movement is described as the dog moving forward at a high pace in galop.	296 (14%)
Eating	The series of actions that begin with the dog's tongue making contact with the meal and end with the food being swallowed. Additionally, the dog may pause between meals to breathe.	137 (7%)
Drinking	The series of actions that begins with the dog's tongue touching the liquid and continues until the dog comes to a halt. Additionally, the dog may pause between drinks to breathe. The head bobbles.	21 (1%)

Correct selection of the evaluation criteria is crucial for evaluating the merits of a model. In this work, the overall model accuracy, confusion matrices and class accuracy of validation data are considered for the model performance assessment.

III. RESULTS

The machine learning models in this work are developed and evaluated in Python language using Keras with Tensorflow as backend. The experiments are conducted on a Dell inc. computer equipped with an Intel(R) Core(TM) i7-8650U CPU (1.90GHz).

Fig. 2 shows 2.4 second data windows of the eight activities, from the top and front accelerometer mounted to the dog's harness. Lying, sitting and standing data typically appear as constant signals indicating less movement while faster gait data (walking, running and sprinting) consist of fluctuating movements.

First, the effect of sampling rate reduction on the classification accuracy is evaluated using the three machine learning models. Additionally, the models are evaluated using a reduced number of logging axes. Finally, the most important features are identified, and their impact on the classification accuracy of the random forest model is examined.

A. Resampling

The first topic which we investigate is the effect of resampling on the classification accuracy of the three machine learning models. Existing data measured at 50 Hz was resampled to lower sampling rates i.e., 5 Hz, 10 Hz, 12.5 Hz, 25 Hz. The training data for the machine learning models is obtained by automatically splitting the dataset in two parts: two thirds for training and one third for validation. Fig. 3 shows the performance of the three models using chest and neck accelerometer data with increasing sampling rate. For the RF, CNN and hybrid CNN models, the accuracy decreases sharply (with 2.3%, 4.3% and 2.4%, respectively) when the sampling rate is decreased from 10 Hz to 5 Hz. For a sampling rate of 5 Hz, the hybrid CNN performs the best, reaching an accuracy of 95.44% while the RF and CNN models only reach an average accuracy of 93.23% and 91.67%, respectively. Table III shows the class performance of the three models at two classification phases using both neck and chest accelerometer data with increasing sampling rate. It is clear that this reduction in accuracy is mainly due to the misclassification of the classes 'walking' and 'sprinting'. For all the sampling rates the Hybrid CNN classifier is best performing reaching with an average accuracy of 96.90% versus 95.44% for the RF and 95.26% for the CNN. For all sampling rates, the RF classifier misclassifies 20% or more of the sitting, walking and standing instances. All

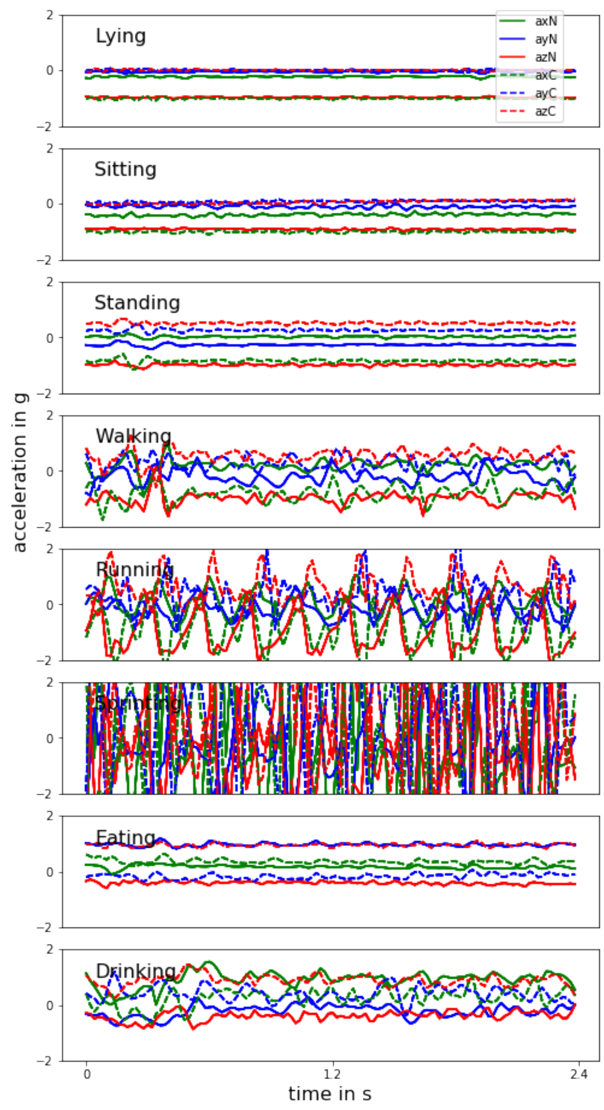


Fig. 2. Typical accelerometer patterns from top to bottom: lying, sitting, standing, walking, running, sprinting, eating and drinking in a 2.4 s window. The green, blue and red dashed lines represent X,Y,Z signals from the chest accelerometer and the red, green, blue solid lines represent X,Y,Z signals from the neck accelerometer , respectively.

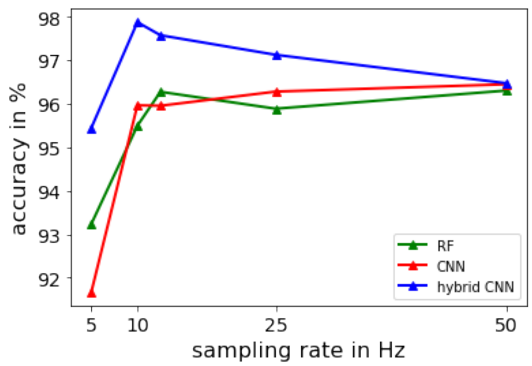


Fig. 3. Performance of the classification of the three machine learning models: a Random Forest, a CNN and a hybrid CNN with increasing sampling rate using neck and chest accelerometer data.

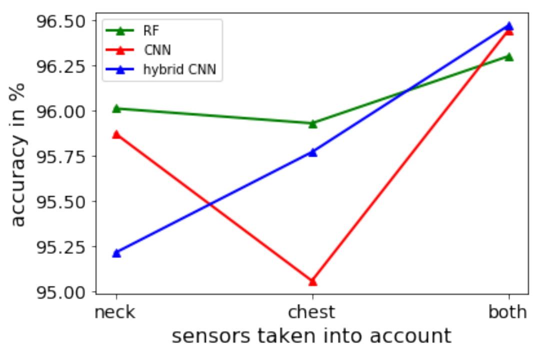


Fig. 4. Performance of the classification of the three machine learning models: a Random Forest, a CNN and a hybrid CNN with increasing sampling rate using neck, chest or both 50 Hz accelerometer data.

the models classify drinking, eating, running and lying with high accuracies (≥95%) at any sampling rate.

B. Number of sensors

The second question concerned the effect of the number of sensors on the model's performance. To account for the effect of the number of sensors on classification accuracy, the neck, chest or both tri-axial accelerometer data measured at 50 Hz were selected. Once again, training and validation data are automatically split into two thirds for training and one third for validation.

As can be concluded from Fig. 4, the models validated on two sensors reach a higher mean accuracy then when only data from the neck or chest sensor is used (96.41% vs. 95.7% or 95.6%). Table IV shows the class performance of the three models at two classification phases using accelerometer data of the neck, chest and both at a 50 Hz sampling rate. When using only chest sensor data, sitting is misclassified by all models as steady. For the CNN and hybrid CNN models more instances of standing are misclassified as lying when only data of one sensor is used.

C. Logging axes

To take the effect of the logging axes on the classification accuracy into account, existing axes from accelerometer data measured at 50 Hz were selected ranging from one to six logging axes (neck and chest tri-axial leg accelerometer data). Fig. 5 depicts the accuracy of the hybrid CNN model while adding more axes. If information from

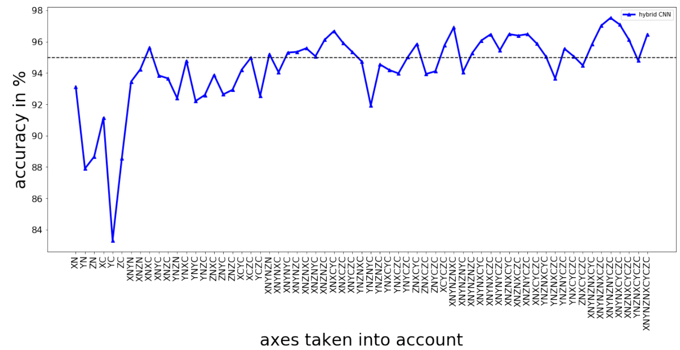


Fig. 5. Performance of the classification of the hybrid CNN model as function of the accelerometer axes for 50 Hz neck and chest accelerometer data. The striped line indicates an accuracy of 95%. (XN and XC, YN and YC and ZN and ZC are x, y and z-axis accelerometer data of the neck and chest, respectively.)

one axis is used, the model is best performing (93.13%) for information from the neck x-direction (forward direction along the dog). Good results (95.62%) are reached when x-direction information from the neck and chest (XNXC) are taken into account. The model reaches the highest accuracy of 97.52% when five axes (XNYNZNYCZC) are taken into account.

D. Feature importance

The third topic was on the impact of the amount of features on the RF's performance. Only the first phase of classification is discussed here, as the RF model does not perform well at differentiating between 'standing' and 'lying,' and hence the features for this second classification phase are irrelevant to examine. A subset of thirteen features is extracted from each 2.4 s segment of labelled sensor data to account for the effect of the number and type of features on classification accuracy. To systematically determine the usefulness and identify the most important features for classifying distinct activities, a random forest ranking of importance was undertaken. In addition, Fig. 6 depicts the accuracy of the random forest model while adding more features in order of decreasing feature relevance.

As can be concluded from the results that are presented in Fig. 6, the maximum acceleration in the x-direction of the neck (max acc x N) has the most important role (relative feature importance of 11%) in predicting the behaviours, followed by the mean of the acceleration in the x-direction of the neck (mean acc x N), the standard deviation of the acceleration in the x-direction of the neck (std acc x N), the mean of the acceleration in the x-direction of the chest (mean acc x C), etc. Fig. 1 shows the orientation of the x,y and z axes. Furthermore, Fig. 6 shows the accuracy of the random forest model while adding extra features according to descending feature importance. An accuracy of 90.14% can already be reached when the only selected feature is the 'max acc x N'. Adding 'mean acc x N' and 'std acc x N' results in the largest increase in overall accuracy to 95.42% which is an acceptable accuracy (≥ 95%). Fig. 7 shows the confusion matrices when the features max acc x N, max acc x N + mean acc x N and max acc x N + mean acc x N + std acc x N are taken into account for classification. The behaviours 'walking', 'eating' and 'drinking' get misclassified more often (34%, 24% and 33%) when only the 'max acc x N' is selected as feature. Adding 'mean acc x N' solves this misclassification problem for drinking. Adding 'std acc x N' increases the classification accuracy for all the classes except 'walking' above 90%.

TABLE III

CLASSIFICATION ACCURACY FOR EACH CLASS AS FUNCTION OF THE SAMPLING RATE FOR NECK AND CHEST ACCELEROMETER DATA FOR THREE MACHINE LEARNING MODELS: A RANDOM FOREST CLASSIFIER, A CNN AND A HYBRID CNN. THE GREEN CELLS REPRESENT CLASSIFICATION ACCURACIES $\geq 95\%$ AND THE RED CELLS REPRESENT CLASSIFICATION ACCURACIES $<80\%$.

		RF					CNN					Hybrid CNN				
		5 Hz	10 Hz	12.5 Hz	25 Hz	50 Hz	5 Hz	10 Hz	12.5 Hz	25 Hz	50 Hz	5 Hz	10 Hz	12.5 Hz	25 Hz	50 Hz
Phase 1	**Drinking**	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	Eating	100.00	98.04	100	100	100	100	100	100	100	100	100	100	100	100	100
	Running	96.87	98.17	98.17	98.43	98.43	98.17	98.43	98.43	98.17	98.43	97.65	98.17	98.96	97.65	98.69
	Sitting	76.92	76.92	69.23	61.54	69.23	0.00	92.31	100	84.62	84.62	100.00	100	100	84.62	100
	Sprinting	85.71	95.60	96.70	93.41	94.51	83.52	94.51	94.51	93.41	91.21	93.40	96.70	95.40	95.60	94.51
	Steady	97.83	99.28	99.28	99.28	99.28	97.83	97.10	96.38	94.93	97.83	94.93	97.83	98.55	97.10	93.48
	Walking	55.56	55.56	72.22	72.22	77.78	33.33	88.89	88.89	94.44	88.89	77.78	100	83.33	100	88.89
Phase 2	**Lying**	99.26	99.26	99.63	100	100	100	100	100	99.63	98.53	99.26	99.63	99.63	100	99.26
	Standing	61.90	61.90	57.14	57.14	52.38	61.90	0.00	0.00	76.19	90.48	66.67	85.71	80.95	85.71	80.95

TABLE IV

CLASSIFICATION ACCURACY FOR EACH CLASS FOR THE 50 HZ ACCELEROMETER DATA OF THE NECK, CHEST AND BOTH FOR THREE MACHINE LEARNING MODELS: A RANDOM FOREST CLASSIFIER, A CNN AND A HYBRID CNN. THE GREEN CELLS REPRESENT CLASSIFICATION ACCURACIES $\geq 95\%$ AND THE RED CELLS REPRESENT CLASSIFICATION ACCURACIES $<80\%$.

		RF			CNN			Hybrid CNN		
		Neck	Chest	Both	Neck	Chest	Both	Neck	Chest	Both
Phase 1	**Drinking**	100	100	100	100	100	100	100	100.00	100
	Eating	100	100	100	100	100	100	100	98.03	100
	Running	98.43	98.17	98.43	98.96	98.17	98.43	98.96	97.65	98.69
	Sitting	100.00	61.54	69.23	100	30.77	84.62	100	53.85	100
	Sprinting	95.60	95.60	94.51	93.41	92.31	91.21	92.31	97.80	94.51
	Steady	98.55	97.10	99.28	92.75	96.38	97.83	90.58	95.65	93.48
	Walking	61.11	83.33	77.78	83.33	94.44	88.89	88.89	88.89	88.89
Phase 2	**Lying**	98.99	100	100	98.53	99.26	98.53	97.79	99	99.26
	Standing	52.38	47.62	52.38	71.43	42.86	90.48	57.14	66.67	80.95

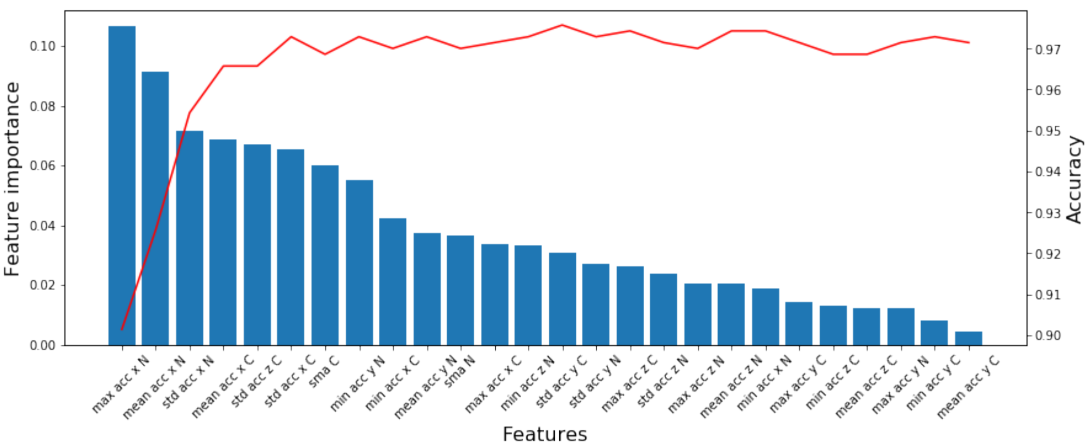


Fig. 6. Left axis: the random forest algorithm's ranking of the features. The importance is determined by which features are the most informative for the algorithm when it comes to making a choice. Right axis: the random forest algorithm's overall model accuracy when additional features are added one by one.

IV. CONCLUSION

In this study we propose a machine learning approach for dog activity recognition, which is required for monitoring dog health and welfare. Using an experimental dataset with eight behaviors from six different dogs, we compared the recognition rate of three machine learning algorithms. The findings show that the proposed hybrid CNN-based model outperforms baseline methods and achieves state-of-the-art results at reduced sample rates. Its advantage is that it can achieve 97.87 percent accuracy by using low sampling rates of up to 10 Hz. Future work will include capturing and analyzing more behaviours at additional sensor locations.

ACKNOWLEDGMENT

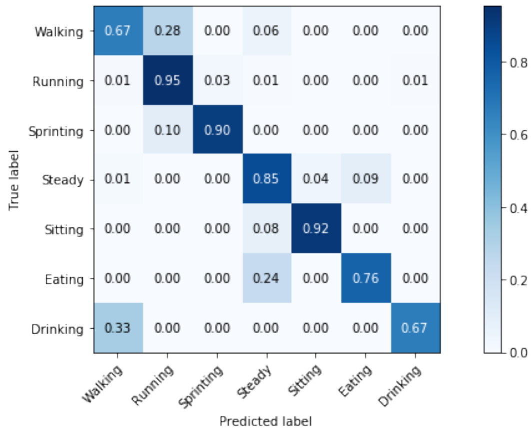
M. Deruyck is a Post-Doctoral Fellow of the FWO-V (Research Foundation – Flanders, Belgium, ref. 12Z5621N).

REFERENCES

[1] E. Courcier, R. Thomson, D. Mellor, and P. Yam, "An epidemiological study of environmental factors associated with canine obesity," *Journal of Small Animal Practice*, vol. 51, no. 7, pp. 362–367, 2010.

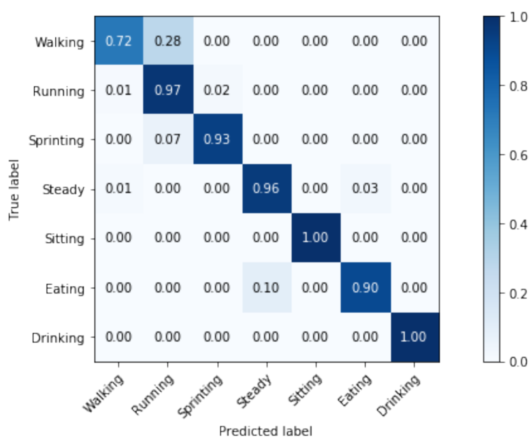
[2] J. Rushen, N. Chapinal, A. De Passille *et al.*, "Automated monitoring of behavioural-based animal welfare indicators," *Animal Welfare-The UFAW Journal*, vol. 21, no. 3, p. 339, 2012.

[3] J. D. Griffies, J. Zutty, M. Sarzen, and S. Soorholtz, "Wearable sensor shown to specifically quantify pruritic behaviors in dogs," *BMC veterinary research*, vol. 14, no. 1, pp. 1–10, 2018.



(a) max acc x N, 90.14%

(b) max acc x N + mean acc x N, 92.57%



(c) max acc x N + mean acc x N + std acc x N, 95.43%

Fig. 7. Exemplar confusion matrices of the validation data for the Random Forest model validated on 50 Hz accelerometer data with features 'max acc x N' (a), max acc x N + mean acc x N (b) and max acc x N + mean acc x N + std acc x N (c).

[4] R. Morrison, V. Penpraze, A. Beber, J. Reilly, and P. Yam, "Associations between obesity and physical activity in dogs: a preliminary investigation," *Journal of Small Animal Practice*, vol. 54, no. 11, pp. 570–574, 2013.

[5] C. Westgarth and C. Ladha, "Evaluation of an open source method for calculating physical activity in dogs from harness and collar based sensors," *BMC veterinary research*, vol. 13, no. 1, pp. 1–7, 2017.

[6] R. D. Chambers, N. C. Yoder, A. B. Carson, C. Junge, D. E. Allen, L. M. Prescott, S. Bradley, G. Wymore, K. Lloyd, and S. Lyle, "Deep learning classification of canine behavior using a single collar-mounted accelerometer: Real-world validation," *Animals*, vol. 11, no. 6, p. 1549, 2021.

[7] S. Lee and M. Annavaram, "Wireless body area networks: Where does energy go?" in *2012 IEEE International Symposium on Workload Characterization (IISWC)*. IEEE, 2012, pp. 25–35.

[8] S. Benaissa, F. A. Tuyttens, D. Plets, T. De Pessemier, J. Trogh, E. Tanghe, L. Martens, L. Vandaele, A. Van Nuffel, W. Joseph *et al.*, "On the use of on-cow accelerometers for the classification of behaviours in dairy barns," *Research in veterinary science*, vol. 125, pp. 425–433, 2019.

[9] E. Walton, C. Casey, J. Mitsch, J. A. Vázquez-Diosdado, J. Yan, T. Dottorini, K. A. Ellis, A. Winterlich, and J. Kaler, "Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour," *Royal Society open science*, vol. 5, no. 2, p. 171442, 2018.

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

[11] A. Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Applied Soft Computing*, vol. 62, pp. 915–922, 2018.

[12] A. Eerdekens, M. Deruyck, J. Fontaine, L. Martens, E. De Poorter, D. Plets, and W. Joseph, "A framework for energy-efficient equine activity recognition with leg accelerometers," *Computers and Electronics in Agriculture*, vol. 183, p. 106020, 2021.

[13] H. Brugman and A. Russel, "Annotating multimedia/multi-modal resources with elan. max planck institute for psycholinguistics, the language archive, nijmegen, the netherlands," 2004.

[14] E. M. Gilbert Jr and T. R. Brown, *K-9 Structure & Terminology*. Dogwise Publishing, 2017.

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

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[17] S. Li, W. Song, H. Qin, and A. Hao, "Deep variance network: An iterative, improved cnn framework for unbalanced training datasets," *Pattern Recognition*, vol. 81, pp. 294–308, 2018.

[18] A. A. Mallouh, Z. Qawaqneh, and B. D. Barkana, "Utilizing cnns and transfer learning of pre-trained models for age range classification from unconstrained face images," *Image and Vision Computing*, vol. 88, pp. 41–51, 2019.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[20] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert systems with applications*, vol. 59, pp. 235–244, 2016.

[21] M. T. Yazici, S. Basurra, and M. M. Gaber, "Edge machine learning: Enabling smart internet of things applications," *Big data and cognitive computing*, vol. 2, no. 3, p. 26, 2018.

[22] S. P. Le Roux, J. Marias, R. Wolhuter, and T. Niesler, "Animal-borne behaviour classification for sheep (dohne merino) and rhinoceros (ceratotherium simum and diceros bicornis)," *Animal Biotelemetry*, vol. 5, no. 1, p. 25, 2017.