LIVERPOOL
JOHN MOORES
UNIVERSITY

LJMU Research Online

Kleanthous, N, Hussain, A, Khan, W, Sneddon, J and Liatsis, P

Deep transfer learning in sheep activity recognition using accelerometer data

https://researchonline.ljmu.ac.uk/id/eprint/18931/

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Kleanthous, N, Hussain, A ORCID logoORCID: https://orcid.org/0000-0001-8413-0045, Khan, W ORCID logoORCID: https://orcid.org/0000-0002-7511-3873, Sneddon, J and Liatsis, P (2022) Deep transfer learning in sheep activity recognition using accelerometer data. Expert Systems with

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

http://researchonline.ljmu.ac.uk/

Deep Transfer Learning in Sheep Activity Recognition using Accelerometer Data

Natasa Kleanthous[a], *Abir Hussain*[a], *Wasiq Khan*[a], *Jennifer Sneddon*[b], *and Panos Liatsis*[c]

[a] School of Computer Science and Mathematics, Liverpool John Moores University, Liverpool, Byrom Street, L3 3AF, United Kingdom
Email: N.K.Orphanidou@2015.ljmu.ac.uk (Corresponding Author), A.Hussain@ljmu.ac.uk, W.Khan@ljmu.ac.uk

[b]School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, Byrom Street, L3 3AF, United Kingdom
Email: J.C.Sneddon@ljmu.ac.uk

[c] Department of Electrical Engineering and Computer Science, Khalifa University of Science and Technology, Abu Dhabi, UAE
Email: panos.liatsis@ku.ac.ae

ABSTRACT

Machine learning and sensor devices lined up with agriculture for the development of systems can efficiently provide real-time knowledge on animal behavior without the need of intense human observation, which is time consuming and labor demanding. In this study, we propose a system to classify three important activities of sheep namely, "grazing", "active", and "inactive" states. We acquire primary data using two types of sensors from nine Hebridean ewes resulting in source data, and target data. To address the problem of sensor heterogeneity in data and sensor orientation placement, we use convolutional neural networks in conjunction with hand-crafted features, in order to improve model generalization, i.e., robustness to sensor orientation and position. Additionally, we utilise transfer learning, which indicates substantial potential in future studies concerning animal activity recognition. More specifically, it supports the reusability of pre-trained model on unseen data without investigative training and data labelling, which is highly time-consuming. Our method obtained an overall accuracy of 98.55% on the source data, and 96.59% on the target data. This study is the first of its kind to propose convolutional neural network based transfer learning for sheep activity recognition, and demonstrate the important benefits of such an approach in the context of data gathering, data labelling, and heterogeneity of sensor devices.

Keywords

Accelerometer sensors, Animal activity recognition, Convolutional neural networks, Deep Learning, Sheep behaviour, Transfer Learning

1. Introduction

Advancements in the field of computer science and electronics engineering paved the way for renewed interest in the use of Internet of Things (IoT) systems in efficient and automated agriculture (Rutter, 2017). Machine intelligence has been deployed in this domain to provide real-time information, specifically related to animal activities. For instance, automation in real-time information processing has been used to identify the current position of animals, their activities and current state, where they mostly graze, and what their nutritional habits are during the course of a day (Anderson, Estell, Holechek, Ivey, & Smith, 2014). For this purpose, smart devices are essential in the collection, processing, and analysis of real-time behavioral data, to efficiently monitor and control animal behavior and distribution on the pasture. Having this vast amount of data, decisions about animal health, animal spatial distribution and land utilisation (Norton, Barnes, & Teague, 2013) can help prevent soil erosion and contamination, water pollution and spread of animal diseases (Rutter, 2017), and therefore, facilitate animal welfare. Furthermore, automated monitoring of animals allows early detection of illness, particularly, lameness, which is present in an estimated 80% of UK flocks (Winter, 2008). Lameness in sheep can be identified based on behavioral changes in animals (Al-Rubaye, Al-Sherbaz, McCormick, & Turner, 2016; Barwick, Lamb, Dobos, Schneider, et al., 2018; Gougoulis, Kyriazakis, & Fthenakis, 2010). Additionally, evidence showed that reduced animal activity or food intake may be an indicator of the disease (Gougoulis et al., 2010). Therefore, machine intelligence-based monitoring of animals in real-time is in demand in sheep production systems.

Grazing is of high interest in the sheep industry. This is the act of feeding on growing grasses and herbage, while standing or walking with the head down. The inactive state of the animals is the act of lying down or standing still. Animal inactivity is an indicator of their welfare as it is a sign of relaxation. Moreover, when animals are inactive (i.e., standing or resting), they take time for rumination, which is considered necessary and vital for animal welfare (Giovanetti et al., 2017). The active state in this research is defined as the behavior when the animal is walking with the head up, or scratching. The majority of existing studies focus on commercialization of smart devices, mainly in cattle, however, there is a limited number of studies in commercialization of sheep monitoring devices due to the lack of sufficient research concerned with sheep (Barwick et al., 2020). Some of these studies consider classification of sheep behavior, mostly using accelerometers, however, there are no clear recommendations on the optimal deployment of such devices to identify sheep behavior, which remains an open challenge (Radeski & Ilieski, 2017). Furthermore, sensor devices are often upgraded, however, there is no evidence that signal processing and classification algorithms designed and applied on data captured from one sensor, can be re-used on data from a new sensor. Therefore, development of a reusable technique is essential, as it can save time and effort since data gathering and labelling is the most time-consuming step of the supervised learning approach.

In this study, we propose a system for monitoring sheep activity using accelerometer data with robustness to accelerometer specifications, position, and orientation. To solve the problem of heterogeneity of accelerometer sensors, we use Transfer Learning (TF) based on Convolutional Neural Networks (CNN) in animal behavior classification. The aim is to use the pre-trained model obtained from the source dataset (sensor 1), and test it on the target dataset (sensor 2), using the learned features and weights. This will allow testing of the hypothesis of generalization in animal behavior classification, independent of the type and orientation of the accelerometer device, by leveraging the pre-trained model obtained on a larger dataset (sensor 1), to classify behavior on a smaller dataset (sensor 2).

The contributions of this research are as follows:

(1) A real time data-driven approach for animal activity recognition is proposed comprising: a) grazing, b) active, and c) inactive states using a composite of the CNN and hand-crafted features, which significantly improves classification performance.

(2) Two primary datasets are acquired through different sensors, which are made publicly available to the research community.

(3) We utilize deep Transfer Learning to data gathered from two types of sensors located on the collar of sheep in a non-fixed orientation, in order to introduce variability in the dataset, hence evaluating the generalization properties of the proposed sheep activity recognition approach. To the best of our knowledge, the proposed Transfer Learning approach is used for the first time in animal behavior recognition.

The reminder of this paper is organized as follows. Section 2 provides an overview of the state-of-the-art related to machine learning and deep learning for animal activity recognition using motion sensors. Section 3 describes the proposed activity recognition algorithm and presents the description of the datasets and the system methodology. Section 4 illustrates the experimental design. Section 5 presents the results and discussion, followed by Section 6, which summarizes the conclusion of this research and proposes avenues for future work.

2. Review of data-driven animal activity recognition methods

In the agricultural industry, various studies explored identification of animal activities based on accelerometer data (Barwick et al., 2020; L. A. A. González, Bishop-Hurley, Handcock, & Crossman, 2015; J W Kamminga, 2017; S. Le Roux et al., 2017). Wearable devices using accelerometers have been commonly used with ML techniques to recognize the behavior of cattle (Andriamandroso et al., 2017; Dutta et al., 2015; L. A. A. González et al., 2015; Gou et al., 2019; Rahman et al., 2018; Riaboff et al., 2019; Robert, White, Renter, & Larson, 2009; Smith et al., 2016; Vázquez Diosdado et al., 2015; Werner et al., 2019). Furthermore, ML was used to identify the activities of horses (Gutierrez-Galan et al., 2018), sharks (Hounslow et al., 2019), seals (Ladds et al., 2017), goats (Jacob W. Kamminga et al., 2018; Navon, Mizrach, Hetzroni, & Ungar, 2013) and other domesticated or wild animals.

While many research studies addressed classification of animal behaviour, specifically, sheep, there is still a need for further investigation on the optimization of the devices and techniques used (Barwick et al., 2020). Different studies propose diverse models, setups and devices. For example, Umstatter et al., (Umstätter, Waterhouse, & Holland, 2008) used tilt sensors to distinguish between the active and inactive states of sheep using 30s windows and achieved predictions of over 90% using linear discriminant analysis (LDA), decision trees (DT), and threshold-based decision trees (Umstätter et al., 2008). On the other hand, Nadimi et al., used accelerometer data acquired from the collar of sheep to classify behavior and achieved accuracies of 83.8%, 83.2%, 73.8%, 71.8%, and 68.5% for grazing, lying, walking, standing, and others, respectively (Nadimi, Jørgensen, Blanes-Vidal, & Christensen, 2012). Both (Umstätter et al., 2008) and (Nadimi et al., 2012) were concerned with animal behaviour on pasture, however the focus was very different, i.e., from the type of sensor device used to behaviors of interest. Furthermore, similar behaviors were investigated by Marais et al., (Marais et al., 2014), where accelerometer data was gathered from the collar of sheep at a sample rate of 100Hz. The authors extracted 10 features using a 5.12s window. The study analyzed lying, standing, walking, running, and grazing behaviors using LDA and quadradic discriminant analysis (QDA), achieving an overall accuracy of 87.1% and 89.7% for LDA and QDA, respectively. LDA was also used by Solomon et al., to identify standing, walking, grazing, running and lying down activities of sheep using a 5.3s window, achieving an overall accuracy of 82.40% (S. P. le Roux, Marias, Wolhuter, & Niesler, 2017). McLennan et al., used a 25s window to identify low activity (i.e., lying ruminating, lying), medium activity (i.e., standing, standing ruminating, grazing), and high activity (i.e., walking) (McLennan et al., 2015). The authors gathered accelerometer data from the collar of sheep and applied statistical analysis. The overall accuracy was 59.09%, 3.37% and 74.56% for high, medium and low activity behaviours, respectively. Furthermore, (McLennan et al., 2015) investigated active and inactive behavior and reported a classification accuracy of 79.98% and 74.56%, respectively, for active and inactive states.

Alvarenga et al., gathered accelerometer data from sheep, while placing the sensor at a fixed orientation and position on the halter under the jaw of the animals (Alvarenga et al., 2016). In their study, the authors examined the performance of DT using various setups. For example, they tested the algorithm using windows of 3, 5, and 10s. Additionally, the accelerometer data was gathered at three different sample rates, i.e., 5Hz, 10Hz, and 25Hz. Five mutually exclusive behaviors were examined including grazing, lying, running, standing and walking using 5 features. The study reported the best results based on a kappa score of 0.7935 and an accuracy of 85.5%.

The same sensor placement was used by Giovaneti et al., (Giovanetti et al., 2017), where they applied stepwise discriminant analysis (SDA), canonical discriminant analysis (CDA) and discriminant analysis (DA) using 60s windows to identify grazing, ruminating, and resting, and DA obtained an overall accuracy of 93%. Decandia et al., presented the evaluation of several window sizes to identify grazing, ruminating and other sheep behaviors (Decandia et al., 2018). CDA and DA were applied on 5, 10, 30, 60, 120, 180 and 300s windows at a sample rate of 62.5Hz. The authors extracted 15 parameters from accelerometer data gathered from a sensor placed under the jaw of the animals and the best accuracy of 89.7% was achieved using DA and a 30s window.

Kamminga et al., gathered accelerometer, gyroscope, and magnetometer signals from sheep and goats to classify stationary, foraging, walking, trotting, and running using one type of sensor (Jacob W. Kamminga, Bisby, Le, Meratnia, & Havinga, 2017). The authors measured the complexity in terms of memory and CPU usage between numerous ML algorithms and noted that the DNN classifier was the most promising in terms of complexity vs performance, reporting an accuracy of 94%.

All of the above studies attempt to address open challenges such as variable sampling rate, sliding window size, sensor orientation and data gathering procedures. In the context of Transfer Learning, Oquab et al., (Oquab, Bottou, Laptev, & Sivic, 2014) proposed the use of transfer learning to extract mid-level features from the ImageNet dataset ("ImageNet," n.d.) and reuse the representations on smaller datasets. Xia et al., (Xia et al., 2019) proposed ensemble concepts of multiple Transfer CNNs (TCNN) to improve model generalization, by introducing three ensemble TCNNs. The authors used several datasets and reported enhanced accuracy. They showed better generalization over CNNs and a single TCNN. Transfer learning has been successfully used in the fields of object recognition (Xiao et al., 2020) and human activity recognition (Akbari & Jafari, 2019), however, there is no study so far concerned with the application of TL in animal behavior.

The latter research works demonstrate the importance of using TF in utilizing knowledge acquired from one domain to another, which saves time and supports the application of powerful models, specifically, in the case of limited size datasets. Additionally, in the field of animal activity recognition and wearable devices in general, it provides opportunities to explore a variety of sensor devices and to easily adapt to new sensor configurations, building upon the time and effort spent on the original work. In summary, transfer learning supports the adaptation of new sensors in the problem domain, without the need for exhaustive training, in developing new predictive models. This research is the first of its kind to propose CNN transfer learning for animal activity recognition, specifically in sheep, and highlight the benefits of such an approach in the context of sensor devices, i.e., sensor heterogeneity, variations in sensor orientation and data gathering.

3. Methodology

Two types of accelerometer sensors (metamotionR, and SenseHat) are used in this research placed on the sheep collar to capture the primary data with a sample rate of 12.5Hz. Let D_S represent the data captured from metamotionR, which is considered as the source data, whereas D_T represents the data acquired through RaspberryPi (with the SenseHat board attached) and will be used to validate the reusability of transfer learning, i.e., the target data. Both datasets were labeled manually and normalized using the z-score. CNN was used to identify the activities of animals using supplementary time and frequency domain features, as illustrated in Table 1. Temporal and spectral features were extracted using a sliding window of 2s with 50% overlap, resulting to two additional datasets from the metamotionR and SenseHat sensors, which are referred to as D_{S+}, D_{T+}, respectively. It should be noted that D_{S+}, D_{T+} relate to the augmented datasets, which include the handcrafted features, whereas D_S and D_T consist only from the x, y, and z accelerometer values, and their magnitude.

Extensive simulation experiments were carried out on six CNN models, which were trained using D_S and D_{S+}. The outcome of the simulations was the selection of the top performing CNN configuration, based on the accuracy obtained on the test sets. These models were then stored, and transfer learning was used on D_T, and D_{T+}. Fig.1 illustrates the overall procedure, including the application of transfer learning.

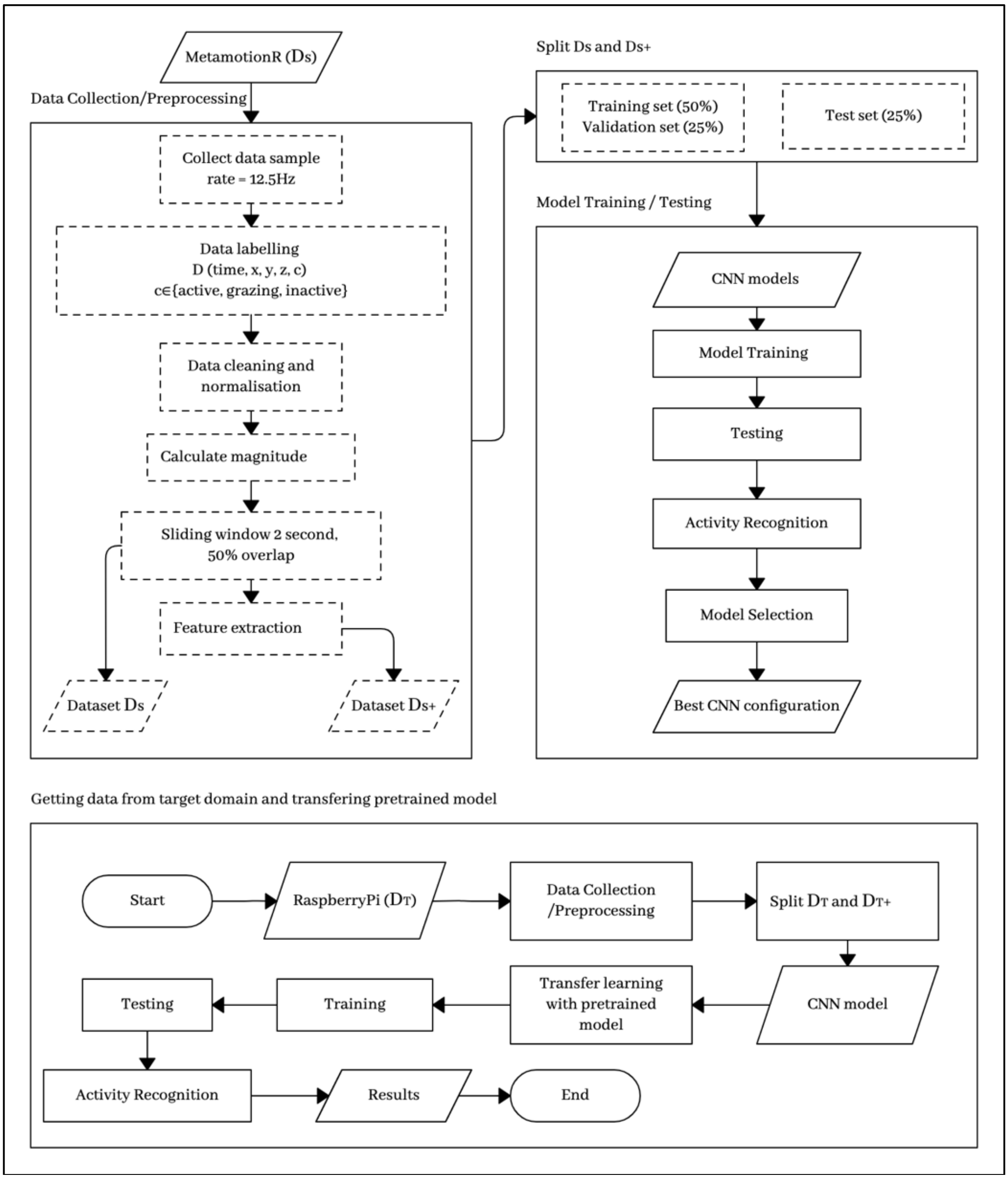

Fig. 1 System methodology

3.1. Datasets

In this work, we collected two primary datasets comprising the accelerometer measurements from a flock of 9 Hebridean ewes 35±5 kg, 9±5 years old at a farm located in Cheshire Shotwick (OS location 333781,371970). Ethical approval was obtained from Liverpool John Moores University to collect the datasets and conduct the experiments (Ref: AH_NKO/2018-13). To acquire the data, two types of devices were used, mounted on the collar of the animals. The first device was mounted on a fixed position (at 270° degrees) and orientation; however, the second device was mounted in a non-fixed manner (at 0°, 90°, or 180° degrees) to test the performance of the proposed methodology, independent of sensor orientation and position. MetamotionR (MBIENTLAB INC, 2018) was used to collect motion and environmental data, while RaspberryPi with the SenseHat board collected temperature, humidity, pressure and motion measurements. Both sensor outputs were logged and the data was stored with a sampling rate of 12.5Hz. In total, recordings of over 65 hours of activity were obtained, which resulted to a dataset of 2,925,000 samples.

The datasets collected from the ewes were loaded into the ELAN_5.7_AVFX Freeware tool ("ELAN - The Language Archive," n.d.). Behaviors such as running, fighting and shaking were very limited and, therefore, were excluded from the datasets. Walking and scratching behaviors were merged and considered as a unified behaviour, labelled as 'active'. Likewise, standing and resting were joined together and labelled as 'inactive'. Chewing with the animal head down, while walking or standing was labelled as 'grazing'. The dataset acquired with MetamotionR (D_S) contains 1,048,575 samples, whereas the dataset captured via RaspberryPi (D_T) comprises 762,860 samples. We refer to these datasets as the source and target datasets, respectively. Let D_k represent the joint dataset, where $D_k=\{t_i, x_i, y_i, z_i, c_i\}$, i=1,..,n, where n is the number of observations, and $k=\{S,T\}$. Parameter t relates to the timestamp, while x, y, z are the accelerometer measurements in the x, y, and z axes, respectively, and c is the label variable, where $c\in\{$active, grazing, inactive$\}$. Fig. 2 presents the distribution of the three activities within our datasets. The charts clearly indicate the imbalanced distribution of activities within both datasets, as expected, due to the nature of the study and the activities considered.

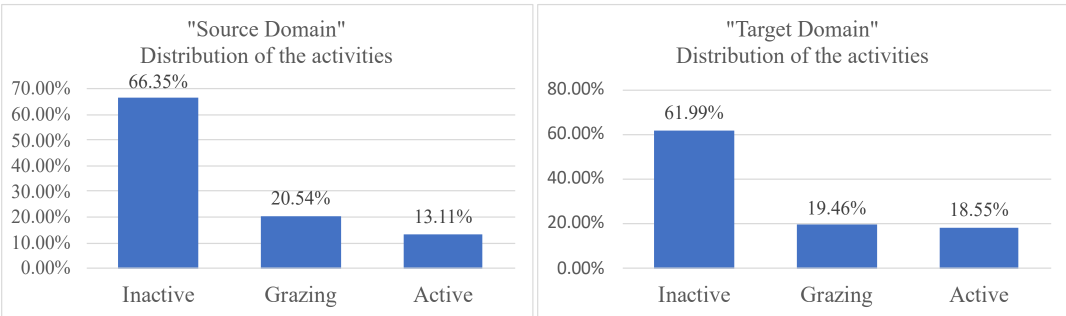

Fig. 2 Duration of the activities for source and target domains.

3.2. Data pre-processing

The magnitude of the accelerometer measurements was calculated resulting in an additional feature in $\mathbf{D_k}=\{t_i, x_i, y_i, z_i, mag_i, c_i\}$, calculated as shown in Equation (1):

$$\text{mag} = \sqrt{x^2 + y^2 + z^2} \tag{1}$$

where, x, y and z represent the 3-dimensional accelerometer data. The dataset was normalized for zero mean (μ) and a standard deviation (σ) of 1. The normalized datasets were then partitioned into training, validation, and testing, with a ratio of 50%, 25%, and 25%, respectively. Overlapping was used to enable real-time classification with a ratio of 50%, which has been shown to be effective in previous activity recognition studies (Bao & Intille, 2004). Using a larger window size in real-life classification could lead to mislabeling, since animals may exhibit more than one behavior in a short time interval. Thus, a 2s window was considered sufficient, while not compromising the device battery life. Where animals exhibited more than one behavior within the same time window, this was labelled with the most frequently occurring activity.

3.3. Feature extraction

We performed temporal and spectral analysis of the overlapping window data to extract meaningful information from the datasets, which may support behavior recognition. The features were selected based on our previous findings (Kleanthous, Hussain, Mason, & Sneddon, 2019; Kleanthous et al., 2018), as well as state-of-the-art research in terms of feature importance using gait information in human and animal activity recognition (Bouten, Westerterp, Verduin, & Janssen, 1994; Gneiting, Ševčíková, & Percival, 2012; S. González et al., 2015). A total of 13 features were calculated for each of the x, y, z accelerometer data and the magnitude of the acceleration signal for each activity, resulting in a 52-dimensional feature set as illustrated in Table 1.

We utilised the Boruta algorithm (Kursa, Jankowski, & Rudnicki, 2010) to explore and rank the level of importance for the extracted feature set and confirm that all features contribute meaningful information, as shown in Fig. 3. It can be deciphered that the mean of the x axis acceleration, skewness and kurtosis of the acceleration magnitude, fractal dimension of the z axis acceleration, and skewness of the y axis acceleration are the most significant features. On the contrary, energy, integrals, rms, peak frequency and squared integrals of the y axis acceleration are the features, which ranked as lowest. However, all features contribute to the discrimination of the three activities, which was confirmed by the Boruta algorithm. Thus, no feature elimination was performed.

Table 1 List of extracted features

#	Feature name	Equation
1	Mean	$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i$
2	Standard Deviation	$s = \frac{1}{n}\sum_{i=1}^{n}(a_i - \bar{a})^2$
3	Skewness [53]	$skewness = \frac{1}{n}\sum_{i=1}^{n}\frac{(a_i - \bar{a})^3}{s^3}$
4	Kurtosis [53]	$kurtosis = \frac{1}{n}\sum_{i=1}^{n}\frac{(a_i - \bar{a})^4}{s^4}$
5	RMS	$rms = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {a_i}^2}$
6	RMS velocity	$rmsV = \sqrt{\frac{1}{n}\sum_{i=1}^{n} diffinv(a_i)^2}$, *diffinv()* is the inverse function of the *diff()*
7	Sum of Changes	$soc = \sum_{i=1}^{n} diff(a_i)$, where *diff()* computes the consecutive differences of the vector
8	Mean of Changes	$moc = \frac{1}{n}\sum_{i=1}^{n} \lvert diff(a_i)\rvert$
9	Integrals	$Integrals = \int_{t=0}^{T} \lvert a_1\rvert\, dt + \int_{t=0}^{T} \lvert a_2\rvert\, dt + \int_{t=0}^{T} \lvert a_3\rvert\, dt$
10	Squared Integrals	$Integrals^2 = \left(\int_{t=0}^{T} \lvert a_1\rvert\, dt\right)^2 + \left(\int_{t=0}^{T} \lvert a_2\rvert\, dt\right)^2 + \left(\int_{t=0}^{T} \lvert a_3\rvert\, dt\right)^2$
11	Madogram [50]	$\gamma_p(t) = \frac{1}{2}\mathrm{E}[a_i - a_{i+u}]$, where *t=lag, E[.]=expectation*
12	Energy	$energy = \sum_{i=1}^{n} {a_i}^2$
13	Peak Frequency [52]	$pf = fmax = \arg\left(\frac{fs}{n} max_{i=0}^{n-1} P(i)\right)$ *fs = sampling frequency, P(i) = power of the spectrum*

a=accelerometer signal x, y, and z where a_1=x, a_2=y, a_3=z

n is the number of rows in the signal window

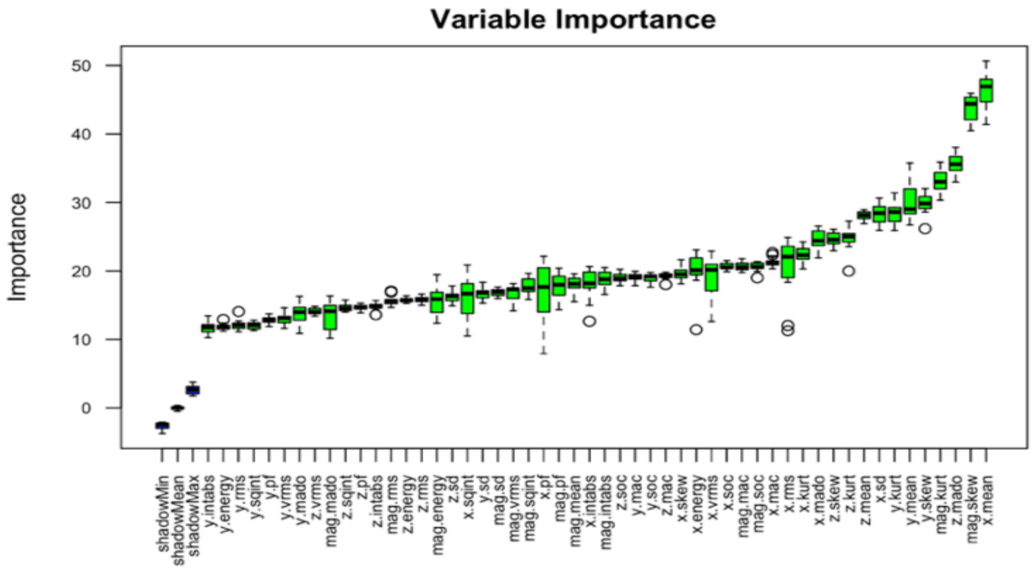

Fig. 3 Feature Importance illustration from the Boruta Algorithm for $\mathbf{D}_{S+}$. Note that features shadowMin, shadowMean, shadowMax are created by the Boruta algorithm in order to rank the original features. More information can be found in (Kursa et al., 2010).

3.4. CNN and Transfer Learning

The recorded measurements and associated features were used in the proposed Deep Learning models to classify the target activities and also during transfer learning.

CNN is a hierarchical feed-forward neural network, widely used in image classification tasks, because of its ability to perform well on very large and complex datasets (Gu et al., 2018; Wu, 2017). In recent years, CNN has also became popular in activity recognition problems (Bo Yang, Nhut Nguyen, Phyo San, Li Li, & Krishnaswamy, n.d.; San et al., 2017). Further details about CNN can be found in (Aloysius & Geetha, 2017; Wu, 2017).

The main idea behind transfer learning is to gain knowledge from a dataset (source domain D_S), and then transfer knowledge to a new dataset (target domain D_T) in order to improve learning in the target domain (Weiss, Khoshgoftaar, & Wang, 2016). Thus, we define a source domain D_S:$X_S \rightarrow Y_S$, with feature space X_S, and a label set Y_S, such that $D_S=\{(\mathbf{x_i}, y_i)\ldots,(\mathbf{x_n}, y_n)\}$, for i=1,…,n, where n is the number of observations in the dataset, and $\mathbf{x_i} \in X_S$, $y_i \in Y_S$. Additionally, we have a target domain D_T: $X_T \rightarrow Y_T$, with feature space X_T, and a label space Y_T, such that $D_T=\{(\mathbf{x_j}, y_j),\ldots,(\mathbf{x_m}, y_m)\}$, where $\mathbf{x_j} \in X_T$, and $y_j \in Y_T$, for j=1,…,m, where m is the number of observations in D_T. Furthermore, a "task" (T) consists of the label Y and the predictive function $f(\cdot)$, and is denoted as $T = \{Y, f(\cdot)\}$, which can be learned from the training data, and used to predict the labels of unseen vectors.

In the case, where $X_S \neq X_T$, i.e., source and target datasets may come from different domains, including different marginal/predictive distributions and feature/label spaces, transfer learning is described as heterogeneous. Alternatively, when $X_S = X_T$, transfer learning is defined as homogeneous. In the current study, homogeneous transfer learning is used because the feature space and domain characteristics from both source and target domains are the same. The difference between the source and target datasets in this study is the accelerometer sensors used, and the orientation and position of the sensor. Additionally, the motion measurements of the second device exhibit some noise and the size of the second dataset is smaller.

4. Experimental Design

To set the baseline for our experiments, a number of classification trials were conducted to investigate the performance of the proposed methodology and to configure the deep learning models. For the CNN, we used datasets D_S and D_{S+}, which include the original measurements and the hand-crafted features, with 50%, 25%, and 25% ratios for training, validation, and testing, respectively for both datasets.

Six CNN configurations (model A to F) were utilized, where we varied the number of convolutional layers and the dropout rate (in the range of [0.1, 0.5]). The common configuration for all models includes: a) for each layer, the ReLu activation function was applied b) smaller dropouts were applied in the convolution layers, whereas larger dropouts were applied in the fully connected layers (i.e., 0.5); c) for all six CNN models, we applied early stopping, modified to monitor the minimum validation loss with patience of 20; d) for model optimization, the Adam optimizer was applied with a learning rate lr=0.001, while the loss function was set to categorical cross-entropy (Wu, 2017); e) a SoftMax layer was used for the classification of target activities. During training, we used weight regularization with a L_2 vector norm of [0.1, 0.01, 0.001, 0.0001, 0.00001] to identify the best configuration. The vector norm was only applied to the fully connected layers of the models.

The performance of the models for each configuration is shown in Table 2. The graphical representation of the accuracy obtained from all models is illustrated in Fig. 4. Model A which has the highest accuracy was selected with L_2=0.00001 to be further used for transfer learning and the results and procedure are presented in section 5. Model A consists of two convolutional layers, one fully connected layer, and an output layer. The first convolutional layer uses 16x16 convolutional filters followed by a 10% dropout layer, and the second convolutional layer uses 32x32 convolutional filters followed by a 20% dropout. The fully connected layer has 64 filters, followed by a 50% dropout, and the output layer uses SoftMax (refer to fig. 5)

Table 2 CNN Performance on D_S and D_{S+} for each model using weight decay with the l2 norm

		Accuracy			
		D_S		**D_{S+}**	
Model	L2	Training set	Test set	Training set	Test set
A	1.00E-01	0.9600	0.9574	0.9564	0.9576
A	1.00E-02	0.9701	0.9652	0.9622	0.9624
A	1.00E-03	0.9753	0.9700	0.9716	0.9690
A	1.00E-04	0.9776	0.9700	0.9787	0.9768
A	**1.00E-05**	**0.9759**	**0.9746**	**0.9892**	**0.9855**
B	1.00E-01	0.975	0.9670	0.9799	0.9792
B	1.00E-02	0.9807	0.9715	0.9850	0.9821
B	1.00E-03	0.9813	0.9684	0.9853	0.9816
B	1.00E-04	0.9800	0.9685	0.9832	0.9809
B	1.00E-05	0.9778	0.9677	0.9847	0.9806
C	1.00E-01	0.9403	0.9365	0.8696	0.8697
C	1.00E-02	0.9606	0.9566	0.9578	0.9608
C	1.00E-03	0.9641	0.9607	0.9691	0.9700
C	1.00E-04	0.9713	0.9659	0.9689	0.9702
C	1.00E-05	0.9744	0.9658	0.9796	0.9796
D	1.00E-01	0.9535	0.9548	0.9622	0.9605
D	1.00E-02	0.9498	0.9479	0.9752	0.9733
D	1.00E-03	0.9576	0.9545	0.9804	0.9781
D	1.00E-04	0.9552	0.9501	0.9863	0.9824
D	1.00E-05	0.9664	0.9580	0.9897	0.9841
E	1.00E-01	0.7904	0.7932	0.9493	0.9485
E	1.00E-02	0.9193	0.9210	0.9683	0.9675
E	1.00E-03	0.9449	0.9453	0.9774	0.9743
E	1.00E-04	0.9183	0.9187	0.9830	0.9798
E	1.00E-05	0.9584	0.9532	0.9873	0.9821
F	1.00E-01	0.9739	0.9710	0.9826	0.9789
F	1.00E-02	0.9790	0.9716	0.9877	0.9822
F	1.00E-03	0.9795	0.9726	0.9871	0.9815
F	1.00E-04	0.9734	0.9721	0.9833	0.9810
F	1.00E-05	0.9753	0.9730	0.9851	0.9829

Fig. 4 Graphical representation of accuracy obtained from the CNN models trained on D_S^{tr}, D_{S+}^{tr}, and tested D_S^{ts}, D_{S+}^{ts}.

5. Results and Discussions

We conducted two experiments using CNN model A since it achieved the best results, while partitioning the datasets into 50%, 25%, and 25% for the training, validation, and testing purposes, respectively. The first experiment used the original source and target datasets, while the second experiment followed the same procedure as the first, but this time we used the datasets, which included the hand-crafted features. The purpose is to investigate whether the addition of the time and frequency domain features has an effect on the performance of the model and the generalization of the algorithm, since the direction and placement of the accelerometers differ between the two sensor setups. For each experiment, a number of statistical metrics, including precision, recall, F_1 score, and accuracy, were used to evaluate the performance of deep learning and transfer learning over different combinations of training and testing datasets. Figure 5 illustrates the architecture of the proposed model and transfer learning procedure.

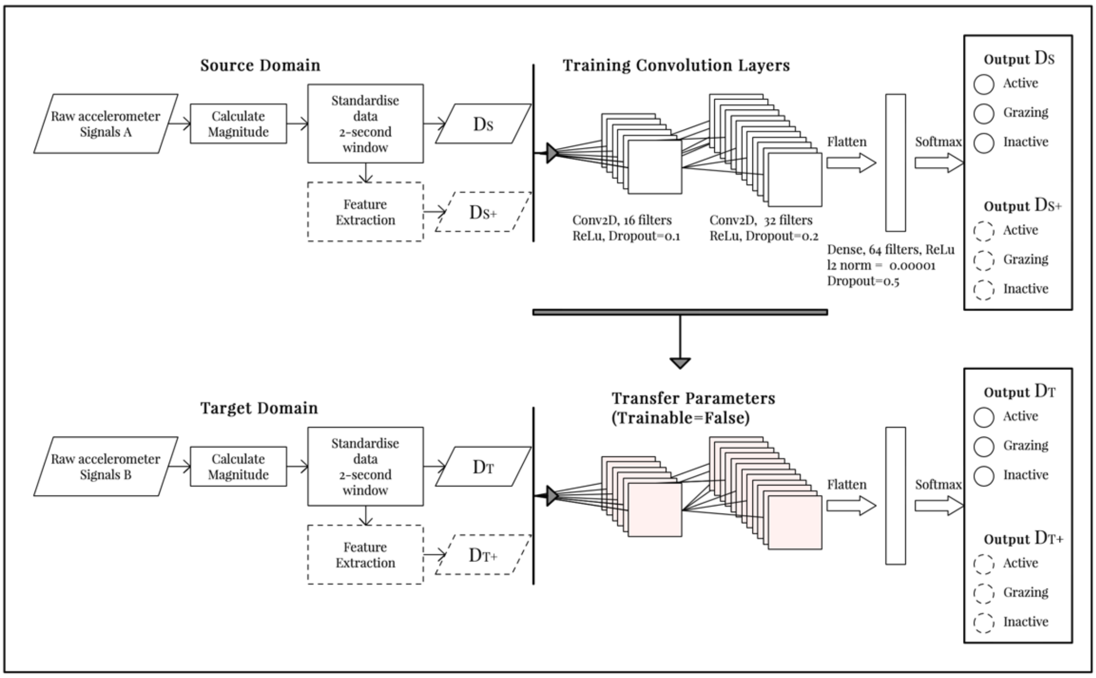

Fig. 5 CNN Architecture for the proposed model

5.1. Experiment A: Transfer learning from D_S to D_T

Model A was trained on the training set D_S^{tr}, validated on D_S^{val}, and then tested on D_S^{ts}. For transfer learning, the trained model was stored so that it can be later reused on the target domain, i.e., D_T. The only trainable layers during transfer learning on D_T^{tr} were the fully connected layers, which are responsible for the classification of the target activities. The results obtained from both experiments are shown in Table 3. Figures 6 and 7 illustrate the accuracy and loss of the model per epoch, respectively.

Table 3 CNN model A classification results on D_S^{ts}, and using Transfer Learning on D_T^{ts}

	D_S^{ts}			D_T^{ts}		
Activities	**Accuracy: 0.9746**			**Accuracy: 0.9479**		
	Precision	**Recall**	**F$_1$ score**	**Precision**	**Recall**	**F$_1$ score**
Active	0.9505	0.8942	0.9215	0.9322	0.8898	0.9105
Grazing	0.9332	0.9745	0.9534	0.8986	0.9417	0.9196
Inactive	0.9963	0.9935	0.9949	0.9982	0.9943	0.9963
Average	**0.9600**	**0.9541**	**0.9566**	**0.9430**	**0.9419**	**0.9421**

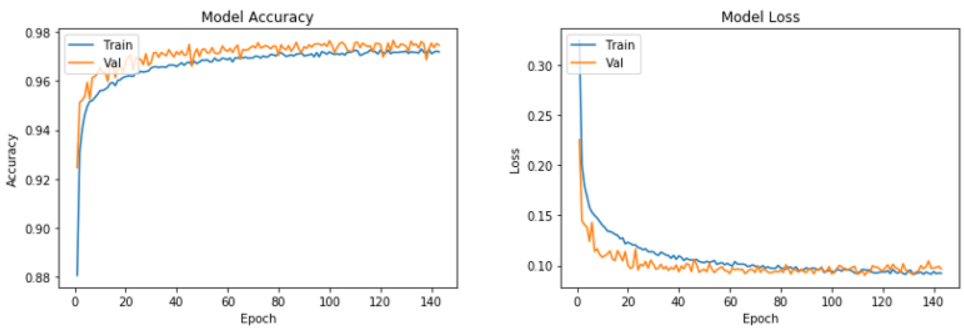

Fig. 6 Experiment A: CNN model A training on source domain D_S^{tr}.

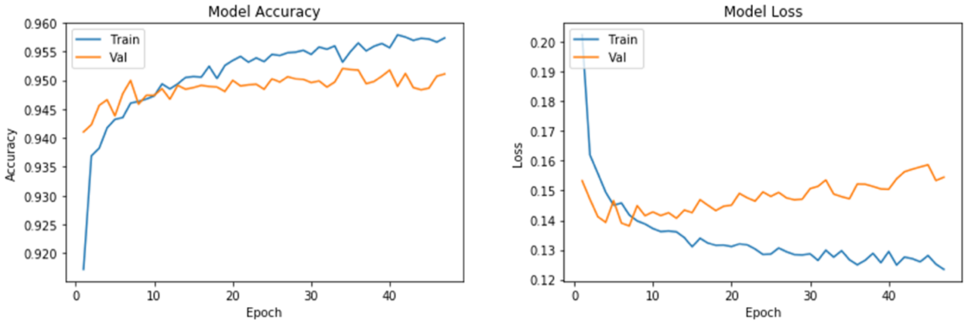

Fig. 7 Experiment A: CNN model A transfer learning on target domain D_T^{tr}.

Table 3 indicates the overall accuracy obtained is 97.46% on $D_S^{ts}$and 94.79% on D_T^{ts}. The highest precision, recall, and F_1 score on both test sets is noticed on the 'inactive' behavior, having scores above 99.35%. The lowest recall was obtained in D_T^{ts}having 88.98% on the 'active' behavior, which is similar for D_S^{ts}, indicating that recall is 89.42%. On the other hand, the precision of active behavior on both test sets is higher from that of the grazing behavior with 95.05% on $D_S^{ts}$and 93.22% on D_T^{ts}. Recall results for grazing behavior are 97.45% on D_S^{ts} and 94.17% on D_T^{ts}, respectively. The best predictive rate was achieved in 'inactive' behavior. The F_1 scores for D_S^{ts} are 92.15%, 95.34%, and 99.49% for active, grazing, and inactive behavior, respectively. For D_T^{ts}, the F_1 scores for active and grazing behavior are at 91.05, and 91.96%, respectively, which is lower than the associated scores on D_S^{ts}. However, inactive behavior on D_T^{ts} has an F_1 score of 99.63%. Overall, it can be observed that the model performed better on the source dataset in all cases and the accuracy decreased, when the model was transferred to the target dataset. A reason for this decrease may be that the orientation of the sensor on the second device was not fixed showing different patterns.

On the other hand, multiple factors may contribute towards the biased performance of the model in the case of inactive behavior (i.e., higher precision, recall and F_1 score than other classes). Firstly, grazing behavior can be easily misclassified as walking or scratching, since these behaviors exhibit similar movements in some cases. Likewise, inactive behavior can be easily classified, contrary to active and grazing, since the pattern does not exhibit changes and remains stable due to motionless behavior. Finally, the distribution of data samples for inactive behavior is comparatively larger than the other two classes. Thus, class imbalance may be one of the major causes that the model performs better in identification of inactive behavior.

5.2. Experiment B: Transfer learning from D_{S+} to D_{T+}

Experiment B is identical to experiment A, except for the datasets, which include the hand-crafted features (i.e., D_{S+}, D_{T+}) . Model A was trained on D_{S+}^{tr},validated on D_{S+}^{val}, and then transfer learning was performed on the target data, i.e., D_{T+}^{tr}. The final model was then tested on the unseen data from D_{S+}^{ts} and D_{T+}^{tr}. Similar to experiment A, only the fully connected layer was allowed to be trained. Results obtained from both tests are presented in Table 4. Figures 8 and 9 illustrate the accuracy and loss of the model per epoch.

Table 4 CNN model A classification results on D_{S+}^{ts}, and using Transfer Learning on D_{T+}^{ts}

	D_{S+}^{ts}			D_{T+}^{ts}		
Activities	**Accuracy: 0.9855**			**Accuracy: 0.9659**		
	Precision	**Recall**	**F_1 score**	**Precision**	**Recall**	**F_1 score**
Active	0.9498	0.9422	0.9460	0.9309	0.8712	0.9000
Grazing	0.9646	0.9669	0.9657	0.9248	0.9551	0.9397
Inactive	0.9987	0.9994	0.9991	0.9917	0.9949	0.9933
Average	**0.9710**	**0.9695**	**0.9703**	**0.9491**	**0.9404**	**0.9443**

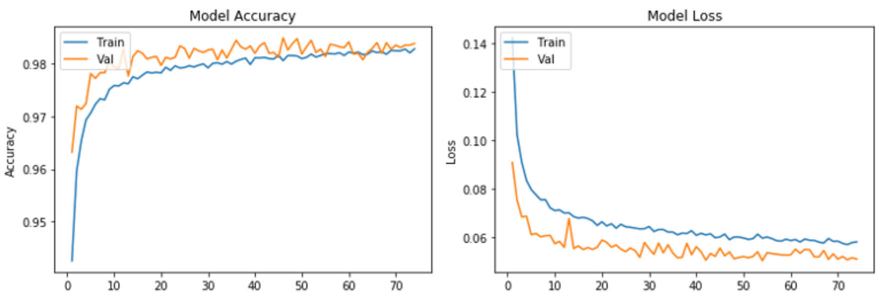

Fig. 8 Experiment B: CNN model A training on source domain D_{S+}^{tr}

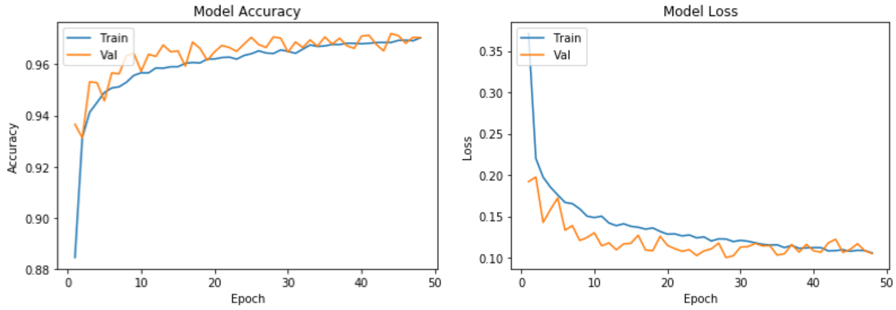

Fig. 9 Experiment B: CNN model A transfer learning on target domain D_{T+}^{tr}.

Table 4 presents the model performance over the datasets with the hand-crafted features, which indicate overall accuracies of 98.55% and 96.59% for D_{S+}^{ts} and D_{T+}^{ts}, respectively. These outcomes also align with the results from experiment A, where accuracy decreases when the model was transferred to the target domain. Likewise, the classification accuracy of inactive behavior is comparatively better with precision and recall of 99.87% and 99.94%, respectively, on D_{S+}^{ts}, and 99.17% and 99.49% on D_{T+}^{ts}, respectively. Precision and recall for the active behavior on the source domain achieved rates of 94.98% and 94.22%, respectively. There is a noticeable increase of 4.8% in the recall result on active behavior in experiment B, in comparison with experiment A, which achieved a recall of 89.42% on the source domain. The results from all behaviors for experiment B on the source domain are comparatively balanced. On the other hand, recall on active behavior in the target domain is 87.12%, which is slightly decreased (i.e., by 1.8%), compared to the recall on the target domain from experiment A. The F_1 scores obtained on the source domain are 94.60%, 96.57%, and 99.91% for active, grazing, and inactive behavior, respectively, which is also slightly higher, when compared to the F_1 scores obtained in experiment A over the source domain. On the other hand, testing the model on the target domain shows a slight decrease in the F_1 score (i.e., 1.05%) on active behavior, but an increase of 2.01% on the grazing behavior. In both experiments, the F_1 scores for inactive behavior are above 99.33% on the target domain. Regarding the grazing behavior, precision and recall on the source domain is 96.46% and 96.69%, respectively. Grazing behavior precision on the target dataset is 92.48%, while recall is 95.51%.

Table 5 Overall Results on D_S and D_{S+} for CNN

Activity	Experiment A D_S^{ts}	Experiment B D_{S+}^{ts}
Active F_1 score	92.15%	94.60%
Grazing F_1 score	95.34%	96.57%
Inactive F_1 score	99.49%	99.91%
Accuracy	97.46%	98.55%

Table 6 Results from Transfer Learning on D_T and D_{T+} for CNN

Activity	Experiment A D_T^{ts}	Experiment B D_{T+}^{ts}
Active F_1 score	91.05%	90.00%
Grazing F_1 score	91.96%	93.97%
Inactive F_1 score	99.63%	99.33%
Accuracy	94.79%	96.59%

A summary of the obtained results is presented in Tables 5 and 6. From the overall results, it is observed that the CNN achieved the best accuracy on both source and target domains, when using the datasets with the hand-crafted features. The experiments also indicate high F_1 scores for the inactive behavior, in the range of 99.33%-99.91%. The lowest F_1 score (90.00%) is achieved on the grazing behavior, when the TF is applied to D_{T+}^{ts}. However, when conducting experiment B on D_{S+}^{ts}, F_1 scores of 96.57% on grazing and 94.60% for active behaviors are obtained. These outcomes indicate the superiority of the proposed model, when compared to previously published studies. For instance, (Fogarty, Swain, Cronin, Moraes, & Trotter, 2020) indicated 69.8% and 45.2% precision performance for the lying and standing behaviors, respectively, as compared to 99.87% and 99.17% in the proposed study for the source and target domain to classify inactivity (lying, standing). Likewise, (Vázquez-Diosdado et al., 2019) and (Decandia et al., 2018) reported 85.18% and 89.7% overall accuracy, which is significantly lower than the proposed model, which achieved accuracies in the range of 96.59%-98.55%. However, it is important to note that the number of activities were different in these studies as compared to the current work.

As mentioned, active behavior is comparatively complex, comprising overlapping behaviors, such as running, shaking, scratching and walking, which exhibit more complex movements. For instance, Barwick reported poor classification (54%) for the standing activity, using collar data (Barwick, Lamb, Dobos, Welch, & Trotter, 2018). Likewise, [66] indicated limited performance with accuracies of 69.8%, 45.2%, and 25.1% for lying, standing and walking behaviors, respectively. Based on these findings and expert recommendations, the proposed model integrated the lying and standing activities into a single behavior (i.e., inactive), which significantly improved accuracy (over 99%). A similar work proposed by Umstater et al., (Umstätter et al., 2008) also indicated that there are instances, where walking and grazing show overlapping patterns, because sheep may graze while walking, which makes it more difficult for ML models to distinguish between these two behaviors.

Studies also indicated performance variations with respect to sensor attachments for data collection. For instance, (Barwick et al., 2020) reported accuracies of 67%-88% for collar-based data, as compared to 86%–95% for ear-tag sensors. Indeed, Barwick (Barwick, Lamb, Dobos, Welch, et al., 2018) obtained better results, when the sensor was attached to the ear-tag of the animal with prediction accuracies of 94%, 96% and 99% for the grazing, standing and walking behaviors, respectively.

The statistical results indicate the robustness of CNN in terms of generalization on unseen data, which support its use in real-life applications. In relation to real-life applications and real-time decision making, the CNN model is very useful, specifically, because it has the ability to automatically extract features from the raw sensor data, while producing robust results, as demonstrated in our experiments. When using CNN with the hand-crafted features in experiment B, we achieve higher results and we show that the transfer learning application is more robust, when compared to experiment A. In other words, the use of CNN with the hand-crafted features supports real-time operation in real-life scenarios of animal monitoring and warning generation.

In addition to the reliable and efficient performance, the use of the transfer learning is advantageous also because of the reusability and generalization of pre-trained models from other applications within similar domains, on unseen datasets. Furthermore, as the CNN performs better with larger datasets, transfer learning can be leveraged so as to provide a cost-effective solution (in regards to time and resources), while reusing it for limited size datasets. In this way, the new dataset can be used in transfer learning, while using the knowledge acquired from model trained in relevant larger datasets.

6. Conclusion and future directions

In this work, the problem of sheep recognition activity was considered in the context of two sensor types and configurations. The first sensor (MetamotionR) was placed at a fixed orientation and was characterized by lower noise density, while the second sensor had varying orientation and higher noise density. Our research investigated the use of CNN and transfer learning on two problem settings. Firstly, we applied CNN and transfer learning on the original datasets. Next, we extracted a large number of temporal and spectral features, which resulted to two augmented datasets. The simulation studies and associated analysis indicated that CNN and transfer learning can generate high accuracy in terms of classifying three sets of activities, including active, grazing and inactive behaviors. High quality classification results were achieved in terms of accuracy, F1 score, precision and recall quality measures, when benchmarked with other results in the literature. It was observed that the inclusion of hand-crafted features improved the performance of both the employed CNN and transfer learning solutions. Furthermore, the simulation results showed the advantage of using deep learning in terms of generalization, indicating its reusability when datasets are limited in animal behavior recognition.

Future research will consider the development of a multi-functional virtual fencing system to control the spatial distribution of the animals among the land they graze using acoustic intermittent cues for the manipulation of their location. In this case, real time monitoring and smart border fencing is proposed to automatically monitor the wellbeing and welfare of animals.

Acknowledgments

We would like to thank the Douglas Bomford Trust for the funding support of this study. We would also like to thank Dr. Jenny Sneddon for permitting access to her residential property, Willow Cottage South, located in Shotwick Village (Chester), and allowing the use of her 9 Hebridean ewes for the experiments in this study.

Availability of data

The raw datasets D_S and D_T of this article are available in the GitHub repository, https://github.com/nkleanthous2015/Sheep_activity_Data

References

Akbari, A., & Jafari, R. (2019). Transferring activity recognition models for new wearable sensors with deep generative domain adaptation. In *IPSN 2019 - Proceedings of the 2019 Information Processing in Sensor Networks* (pp. 85–96). New York, New York, USA: ACM Press. https://doi.org/10.1145/3302506.3310391

Al-Rubaye, Z., Al-Sherbaz, A., McCormick, W. D., & Turner, S. J. (2016). *The use of multivariable wireless sensor data to early detect lameness in sheep. School of Science and Technology Annual Research Conference*. Retrieved from

http://nectar.northampton.ac.uk/8311/
Aloysius, N., & Geetha, M. (2017). A review on deep convolutional neural networks. In *2017 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0588–0592). IEEE. https://doi.org/10.1109/ICCSP.2017.8286426
Alvarenga, F. A. P., Borges, I., Palkovič, L., Rodina, J., Oddy, V. H., & Dobos, R. C. (2016). Using a three-axis accelerometer to identify and classify sheep behaviour at pasture. *Applied Animal Behaviour Science*, *181*, 91–99. https://doi.org/10.1016/j.applanim.2016.05.026
Anderson, D. M., Estell, R. E., Holechek, J. L., Ivey, S., & Smith, G. B. (2014). Virtual herding for flexible livestock management - a review. *The Rangeland Journal*, *36*(3), 205–221. https://doi.org/10.1071/RJ13092
Andriamandroso, A. L. H., Lebeau, F., Beckers, Y., Froidmont, E., Dufrasne, I., Heinesch, B., … Bindelle, J. (2017). Development of an open-source algorithm based on inertial measurement units (IMU) of a smartphone to detect cattle grass intake and ruminating behaviors. *Computers and Electronics in Agriculture*, *139*, 126–137. https://doi.org/10.1016/J.COMPAG.2017.05.020
Bao, L., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. *Pervasive Computing*, *3001*, 1–17. https://doi.org/10.1007/978-3-540-24646-6_1
Barwick, J., Lamb, D., Dobos, R., Schneider, D., Welch, M., & Trotter, M. (2018). Predicting lameness in sheep activity using tri-axial acceleration signals. *Animals*, *8*(1), 1–16. https://doi.org/10.3390/ani8010012
Barwick, J., Lamb, D. W., Dobos, R., Welch, M., Schneider, D., & Trotter, M. (2020). Identifying sheep activity from tri-axial acceleration signals using a moving window classification model. *Remote Sensing*, *12*(4), 646. https://doi.org/10.3390/rs12040646
Barwick, J., Lamb, D. W., Dobos, R., Welch, M., & Trotter, M. (2018). Categorising sheep activity using a tri-axial accelerometer. *Computers and Electronics in Agriculture*, *145*, 289–297. https://doi.org/10.1016/J.COMPAG.2018.01.007
Bo Yang, J., Nhut Nguyen, M., Phyo San, P., Li Li, X., & Krishnaswamy, S. (n.d.). *Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity Recognition*. Retrieved from https://www.ijcai.org/Proceedings/15/Papers/561.pdf
Bouten, C. V, Westerterp, K. R., Verduin, M., & Janssen, J. D. (1994). Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Medicine and Science in Sports and Exercise*, *26*(12), 1516–1523. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7869887
Decandia, M., Giovanetti, V., Molle, G., Acciaro, M., Mameli, M., Cabiddu, A., … Dimauro, C. (2018). The effect of different time epoch settings on the classification of sheep behaviour using tri-axial accelerometry. *Computers and Electronics in Agriculture*, *154*, 112–119. https://doi.org/10.1016/j.compag.2018.09.002
Dutta, R., Smith, D., Rawnsley, R., Bishop-Hurley, G., Hills, J., Timms, G., & Henry, D. (2015). Dynamic cattle behavioural classification using supervised ensemble classifiers. *Computers and Electronics in Agriculture*, *111*, 18–28. https://doi.org/10.1016/j.compag.2014.12.002
ELAN - The Language Archive. (n.d.). Retrieved December 16, 2019, from https://tla.mpi.nl/tools/tla-tools/elan/
Fogarty, E. S., Swain, D. L., Cronin, G. M., Moraes, L. E., & Trotter, M. (2020). Behaviour classification of extensively grazed sheep using machine learning. *Computers and Electronics in Agriculture*, *169*, 105175. https://doi.org/10.1016/j.compag.2019.105175
Giovanetti, V., Decandia, M., Molle, G., Acciaro, M., Mameli, M., Cabiddu, A., … Dimauro, C. (2017). Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer. *Livestock Science*, *196*, 42–48. https://doi.org/10.1016/j.livsci.2016.12.011
Gneiting, T., Ševčíková, H., & Percival, D. B. (2012). Estimators of fractal dimension: Assessing the roughness of time series and spatial data. *Statistical Science*, 247–277.
González, L. A. A., Bishop-Hurley, G. J. J., Handcock, R. N. N., & Crossman, C. (2015). Behavioral classification of data from collars containing motion sensors in grazing cattle. *Computers and Electronics in Agriculture*, *110*, 91–102. https://doi.org/10.1016/j.compag.2014.10.018
González, S., Sedano, J., Villar, J. R., Corchado, E., Herrero, Á., & Baruque, B. (2015). Features and models for human activity recognition. *Neurocomputing*, *167*, 52–60. https://doi.org/10.1016/j.neucom.2015.01.082
Gou, X., Tsunekawa, A., Peng, F., Zhao, X., Li, Y., & Lian, J. (2019). MEthod for classifying behavior of livestock on fenced temperate Rangeland in Northern China. *Sensors (Switzerland)*, *19*(23), 1–15. https://doi.org/10.3390/s19235334
Gougoulis, D. A., Kyriazakis, I., & Fthenakis, G. C. (2010). Diagnostic significance of behaviour changes of sheep: A selected review. *Small Ruminant Research*, *92*(1), 52–56. https://doi.org/10.1016/j.smallrumres.2010.04.018
Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., … Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/J.PATCOG.2017.10.013
Gutierrez-Galan, D., Dominguez-Morales, J. P., Cerezuela-Escudero, E., Rios-Navarro, A., Tapiador-Morales, R., Rivas-Perez, M., … Linares-Barranco, A. (2018). Embedded neural network for real-time animal behavior classification. *Neurocomputing*, *272*, 17–26. https://doi.org/10.1016/j.neucom.2017.03.090
Hounslow, J. L. L., Brewster, L. R. R., Lear, K. O. O., Guttridge, T. L. L., Daly, R., Whitney, N. M. M., & Gleiss, A. C. C. (2019). Assessing the effects of sampling frequency on behavioural classification of accelerometer data. *Journal of Experimental Marine Biology and Ecology*, *512*(December 2018), 22–30. https://doi.org/10.1016/j.jembe.2018.12.003
ImageNet. (n.d.). Retrieved October 3, 2020, from http://www.image-net.org/
Kamminga, J W. (2017). Generic online animal activity recognition on collar tags. Data Archiving and Networked Services (DANS).
Kamminga, Jacob W., Bisby, H. C., Le, D. V., Meratnia, N., & Havinga, P. J. M. (2017). Generic Online Animal Activity Recognition on Collar Tags. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers on - UbiComp '17* (pp. 597–606). New York, NY, USA: ACM. https://doi.org/10.1145/3123024.3124407
Kamminga, Jacob W., Le, D. V., Meijers, J. P., Bisby, H., Meratnia, N., & Havinga, P. J. M. (2018). Robust Sensor-Orientation-Independent Feature Selection for Animal Activity Recognition on Collar Tags. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(1), 1–27. https://doi.org/10.1145/3191747
Kleanthous, N., Hussain, A., Mason, A., & Sneddon, J. (2019). Data Science Approaches for the Analysis of Animal Behaviours. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 11645 LNAI, pp. 411–422). https://doi.org/10.1007/978-3-030-26766-7_38
Kleanthous, N., Hussain, A., Mason, A., Sneddon, J., Shaw, A., Fergus, P., … Al-Jumeily, D. (2018). Machine learning techniques for classification of livestock behavior. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 11304 LNCS, pp. 304–315). https://doi.org/10.1007/978-3-030-04212-7_26
Kursa, M. B., Jankowski, A., & Rudnicki, W. R. (2010). Boruta – A System for Feature Selection. *Fundamenta Informaticae*, *101*(4), 271–285. https://doi.org/10.3233/FI-2010-288
Ladds, M. A., Thompson, A. P., Kadar, J.-P. P., J Slip, D., P Hocking, D., G Harcourt, R., … Harcourt, R. (2017). Super machine learning: improving accuracy and reducing variance of behaviour classification from accelerometry. *Animal Biotelemetry*, *5*(1), 8. https://doi.org/10.1186/s40317-017-0123-1

le Roux, S. P., Marias, J., Wolhuter, R., & Niesler, T. (2017). Animal-borne behaviour classification for sheep ({Dohne} {Merino}) and {Rhinoceros} ({Ceratotherium} simum and {Diceros} bicornis). *Animal Biotelemetry*, *5*, 25. https://doi.org/10.1186/s40317-017-0140-0

Le Roux, S., Wolhuter, R., Niesler, T., Roux, S., Wolhuter, R., Niesler, T., … Niesler, T. (2017). *An Overview of Automatic Behaviour Classification for Animal-Borne Sensor Applications in South Africa*. https://doi.org/10.1145/3132711.3132716

Marais, J., Wolhuter, R., Niesler, T., Le Roux, S., Wolhuter, R., Niesler, T., … Niesler, T. (2014). *Automatic classification of sheep behaviour using 3-axis accelerometer data*. *Pattern Recognition Association of South Africa*. Retrieved from https://www.researchgate.net/publication/319331093_Automatic_classification_of_sheep_behaviour_using_3-axis_accelerometer_data

MBIENTLAB INC. (2018). MetaMotionR – MbientLab. Retrieved December 15, 2019, from https://mbientlab.com/metamotionr/

McLennan, K. M., Skillings, E. A., Rebelo, C. J. B. B., Corke, M. J., Pires Moreira, M. A., Morton, A. J., & Constantino-Casas, F. (2015). Technical note: Validation of an automatic recording system to assess behavioural activity level in sheep (Ovis aries). *Small Ruminant Research*, *127*, 92–96. https://doi.org/10.1016/j.smallrumres.2015.04.002

Nadimi, E. S., Jørgensen, R. N., Blanes-Vidal, V., & Christensen, S. (2012). Monitoring and classifying animal behavior using ZigBee-based mobile ad hoc wireless sensor networks and artificial neural networks. *Computers and Electronics in Agriculture*, *82*(Supplement C), 44–54. https://doi.org/10.1016/j.compag.2011.12.008

Navon, S., Mizrach, A., Hetzroni, A., & Ungar, E. D. (2013). Automatic recognition of jaw movements in free-ranging cattle, goats and sheep, using acoustic monitoring. *Biosystems Engineering*, *114*(4), 474–483. https://doi.org/10.1016/j.biosystemseng.2012.08.005

Norton, B. E., Barnes, M., & Teague, R. (2013). Grazing Management Can Improve Livestock Distribution: Increasing accessible forage and effective grazing capacity. *Rangelands*, *35*(5), 45–51. https://doi.org/10.2111/RANGELANDS-D-13-00016.1

Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1717–1724. https://doi.org/10.1109/CVPR.2014.222

Radeski, M., & Ilieski, V. (2017). Gait and posture discrimination in sheep using a tri-axial accelerometer. *Animal*, *11*(7), 1249–1257. https://doi.org/10.1017/S175173111600255X

Rahman, A., Smith, D. V., Little, B., Ingham, A. B., Greenwood, P. L., & Bishop-Hurley, G. J. (2018). Cattle behaviour classification from collar, halter, and ear tag sensors. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2017.10.001

Riaboff, L., Aubin, S., Bédère, N., Couvreur, S., Madouasse, A., Goumand, E., … Plantier, G. (2019). Evaluation of pre-processing methods for the prediction of cattle behaviour from accelerometer data. *Computers and Electronics in Agriculture*, *165*, 104961. https://doi.org/10.1016/j.compag.2019.104961

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

Rutter, S. M. (2017). 13 - Advanced livestock management solutions. In D. M. Ferguson, C. Lee, & A. Fisher (Eds.), *Advances in Sheep Welfare* (pp. 245–261). Woodhead Publishing. Retrieved from https://www.sciencedirect.com/science/article/pii/B9780081007181000133

San, P. P., Kakar, P., Li, X. L., Krishnaswamy, S., Yang, J. B., & Nguyen, M. N. (2017). Deep Learning for Human Activity Recognition. In *Big Data Analytics for Sensor-Network Collected Intelligence*. https://doi.org/10.1016/B978-0-12-809393-1.00009-X

Smith, D., Rahman, A., Bishop-Hurley, G. J., Hills, J., Shahriar, S., Henry, D., & Rawnsley, R. (2016). Behavior classification of cows fitted with motion collars: Decomposing multi-class classification into a set of binary problems. *Computers and Electronics in Agriculture*, *131*, 40–50. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168169916303180

Umstätter, C., Waterhouse, A., & Holland, J. P. (2008). An automated sensor-based method of simple behavioural classification of sheep in extensive systems. *Computers and Electronics in Agriculture*, *64*(1), 19–26. https://doi.org/10.1016/j.compag.2008.05.004

Vázquez-Diosdado, J. A., Paul, V., Ellis, K. A., Coates, D., Loomba, R., & Kaler, J. (2019). A combined offline and online algorithm for real-time and long-term classification of sheep behaviour: Novel approach for precision livestock farming. *Sensors (Switzerland)*, *19*(14). https://doi.org/10.3390/s19143201

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

Weiss, K., Khoshgoftaar, T. M., & Wang, D. D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

Werner, J., Umstatter, C., Leso, L., Kennedy, E., Geoghegan, A., Shalloo, L., … O'Brien, B. (2019). Evaluation and application potential of an accelerometer-based collar device for measuring grazing behavior of dairy cows. *Animal*, *13*(9), 1–10. https://doi.org/10.1017/S1751731118003658

Winter, A. C. (2008). Lameness in sheep. *Small Ruminant Research*, *76*(1–2), 149–153. https://doi.org/10.1016/j.smallrumres.2007.12.008

Wu, J. (2017). *Introduction to Convolutional Neural Networks*. Retrieved from https://pdfs.semanticscholar.org/450c/a19932fcef1ca6d0442cbf52fec38fb9d1e5.pdf

Xia, S., Xia, Y., Yu, H., Liu, Q., Luo, Y., Wang, G., & Chen, Z. (2019). Transferring Ensemble Representations Using Deep Convolutional Neural Networks for Small-Scale Image Classification. *IEEE Access*, *7*, 168175–168186. https://doi.org/10.1109/ACCESS.2019.2912908

Xiao, G., Wu, Q., Chen, H., Da, D., Guo, J., & Gong, Z. (2020). A Deep Transfer Learning Solution for Food Material Recognition Using Electronic Scales. *IEEE Transactions on Industrial Informatics*, *16*(4), 2290–2300. https://doi.org/10.1109/TII.2019.2931148