

A survey of machine learning approaches in animal behaviour

Natasa Kleanthous^{a*}, Abir Jaafar Hussain^a, Wasiq Khan^a, Jennifer Sneddon^b, Ahmed Al-Shamma'a^c

and Panos Liatsis^d

^a Liverpool John Moores University, Computer Science Department, Liverpool, L33AF, UK

^b Liverpool John Moores University, Natural Sciences and Psychology, Liverpool, L33AF, UK

^c University of Sharjah, College of Engineering, Sharjah, UAE

^d Khalifa University of Science and Technology, Department of Electrical Engineering and Computer Science, Abu Dhabi, UAE

ARTICLE INFO

Keywords

Deep learning
Machine learning,
Sheep activity recognition,
Sheep activity survey,
Feature selection,
Multi-sensor activity

ABSTRACT

Animal activity recognition is an important topic that facilitates understanding of animal behaviour that is useful for analyzing and classifying their wellbeing. Research studies have been reporting the use of animal activity as an effective indicator of their health state. This survey focuses on recent advancements in machine intelligence utilizing wearable devices for sheep activity recognition. We summarise existing works focusing on various types of sensors used in agricultural sheep activity recognition. Furthermore, data segmentation methods used in each study, followed by the potential recommendations on window size and sample rate selection are addressed in detail. Finally, we present the features being identified as significant along with an overview of machine learning algorithms used in the domain of sheep activity recognition using accelerometer data.

1. Introduction

Animal activity recognition is a vital agricultural subject, which is helpful to understand animal behaviour, where the wellbeing of animals can be estimated and classified [1]. Various studies reported the use of animal activity as an indicator of animal health [2–4]. It is noted that the daily monitoring of animal activities and locomotion utilising sensor technology can provide information regarding stress and diseases such as lameness [5–10], and daily nutritional consumption [11]. Furthermore, a decrease in animal activities or hyper activities can provide evidence of animal disease and distress [7]. Information gathering through human observation is time-consuming and labour intensive. Therefore, devices to measure daily animal behaviour have been proposed and used over the past two decades [12,13]. It was also noted that the monitoring of animals with a human observer could influence the natural activity of animals on the pasture [14,15], which may not be the case when using technological devices.

Animal food consumption, the presence of diseases, and general level of activity or inactivity can be estimated and identified using devices with embedded machine learning (ML) algorithms [16]. These measures provide the ability to monitor and diagnose animal welfare [17]. Additionally, the position of animals and activity information can be used to nominate pasture utilisation patterns and animal distribution for pursued animal behaviour [18]. Thus, smart technologies can play a valuable part in animal health management [19] and provide vital insights to individuals and concerned bodies (e.g., farm managers). The development of information technologies has therefore had a huge impact in agriculture using data analysis. Reviews of big data analysis methods, and precision livestock farming using ML in agriculture are provided in [20–22]

A number of studies illustrated the use of computer vision [23,24] and image analysis [24–26] for monitoring animal behaviour. For example, Ren et al. proposed a system able to automatically detect sheep behaviour (standing/lying) and position using an ultra-wideband (UWB) system, infrared radiation cameras, and three-dimensional computer vision technology [23]. The authors successfully detected standing and lying with accuracies of 98.16% and 100%, respectively which is comparable with other studies, however, they noted a limitation of their system for commercial applications due to the battery life of the UWB nodes. Another limitation is the use of cameras, which may result in limited area coverage. Moreover, several studies demonstrated the collection of tracking data from livestock using GPS collars [27–30]. Likewise, recent research focus on measuring food intake through sound analysis [31,32]. All of these studies illustrate the importance of machine based intelligent monitoring of animal behaviour and providing support for the research community and their industrial counterparts. The accelerometers are

* Corresponding author. Tel.: +44-7583406628; e-mail: n.k.orphanidou@2015.ljmu.ac.uk

the most commonly used sensors due to their ability to provide information related to animal gait patterns. Due to their small size, lightweight, and low power requirements, these sensors are widely used in various applications for animal behaviour monitoring [33–35].

This paper surveys the research studies addressing sheep activity recognition based on ML, Deep Learning (DL) and accelerometer measurements. While a number of existing studies propose solutions to this problem, over the last decade, there is a significant research gap between understanding the relationships between the focus of a study and the specific solution parameters, i.e., window size, feature set and significance level, and choice of ML/ DL techniques [34]. For instance, in a study to identify multiple behaviours in dairy calves, it was found that a collar-mounted sensor, providing both accelerometer and magnetometer measurements produced superior results [16]. However, when there was a change in the targeted activities, the sensor was placed on the leg of the calves, and only accelerometer data was used [36]. Furthermore, current research on sheep activity recognition (SAR) using ML is limited and provides an opportunity to systematically analyse data processing and analysis protocols [37], as addressed in this contribution.

Reviews and survey papers regarding farm production systems dedicated to applications of machine learning for, crop management (i.e. plant diseases [38]), agricultural disease image recognition [39], livestock management, water management, and soil management are available in the literature [40,41]. However, to the best of authors' knowledge, the proposed work is first of its kind presenting detailed survey about recent advancements in SAR based on ML and DL using accelerometer data. More specifically, we provide a thorough overview of various window segmentation, feature extraction and selection, and classification algorithms which have been used over the past decade to identify the sheep activities.

The remainder of this paper is organized as follows. Section 2 presents an overview of a typical sheep activity recognition problem. In Section 3 summarizes the existing work and challenges in each of the technical aspects of SAR. Finally, conclusions of this work and potential future directions and recommendations are presented in Section 4.

2. Sheep Activity Recognition using Motion Sensors

This section describes the requirements and typical characteristics of a SAR problem, as illustrated in Figure 1. We present existing works and corresponding techniques proposed within the several steps of SAR task. Overall, a typical activity recognition problem includes: (1) data acquisition through sensor mounting on the animals' body: sensors are placed on the collar, ear, leg, or under the jaw; (2) Data labelling: once the raw data is collected, data labelling is manually performed over sensor or video data. In case of multiple data acquisition devices, the recorded data is time-synchronized to serve as ground truth; (3) pre-processing of the acquired data such as data cleansing and standardization; (4) selection of window size within the time series data and feature extraction from the windowed frames; (5) feature selection to identify the distinguishing feature set; (6) model development and training using ML/DL algorithms; (7) performance evaluation of the models and selection of the most appropriate method.

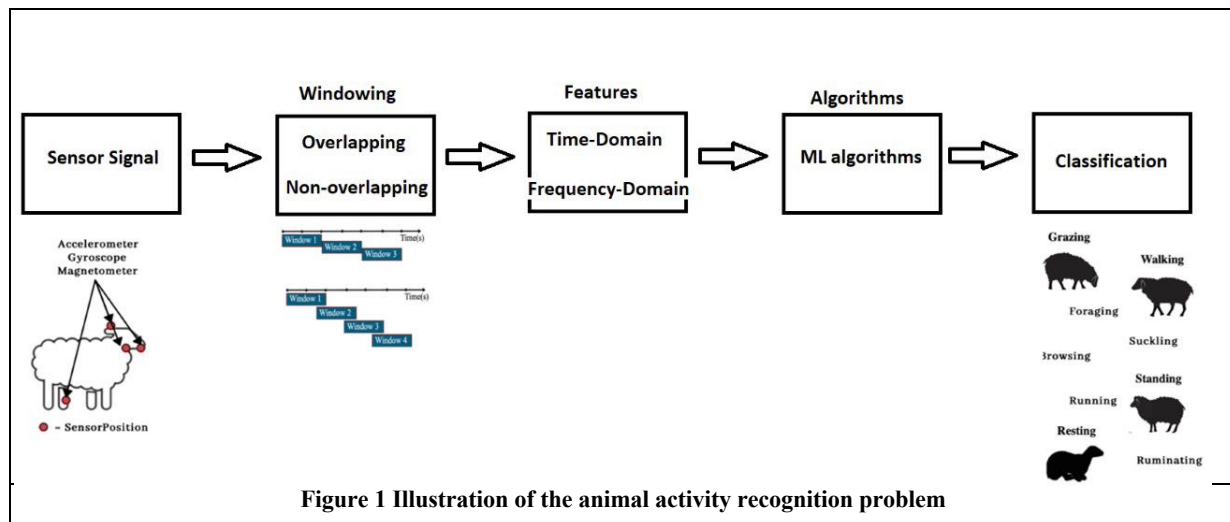


Figure 1 Illustration of the animal activity recognition problem

Figure 1 presents an overview of the core stages followed within the related research studies to investigate various animal behaviours such as, identifying when an animal is active or inactive, classifying whether an animal suffers from lameness etc. Based on the nature of the problem and type of data, multiple options for window size, feature extraction methods, and ML/DL models can be used. The following subsections provide detailed information regarding the various choices within the aforementioned stages.

2.1. Accelerometers

Accelerometers have mainly been used within the existing works in relation to SAR and therefore, this survey mainly focus on the related works, which use accelerometers for data acquisition. Accelerometers measure the acceleration of motion and are a ubiquitous type of sensor in activity recognition problems because they are light weight, small in size, inexpensive, and offer low power consumption [18,34,42–46]. Studies have reported that activities such as walking, grazing, scratching, lying, and standing can be easily recognized by using only accelerometers, yielding overall accuracies in excess of 98% [47,48]. Likewise, running activity was detected with an accuracy of 96.62% using information from only one acceleration axis of accelerometer data [49]. On the other hand, several research studies also used gyroscopes and magnetometers combined with accelerometers to obtain a deeper understanding of domestic and wild animal behaviour [1], [2–4]. Combining accelerometer and gyroscope features produced an accuracy of 98% in identifying lying activity [50]. Similarly, walking activity was predicted with an accuracy of 99% using accelerometer data [51]. Other works also reported the use of accelerometer data to discriminate between active and inactive states with 98.10% accuracy [34].

2.2. Sensor placement

To collect gait patterns from the animals, accelerometer sensors are attached to the animal body, usually in the collar and ear [5,6,34,37,50–53]. On the other hand, there are some studies, where accelerometers are mounted on the leg and under the jaw [54–58].

Table 1 presents the four most common sensor placements reported in research studies over the last decade. It can be observed that the mounting position is determined based on the activity problem under investigation. For example, a recent study [5] showed that an accelerometer attached to the animal's ear could discriminate lame walking from grazing, standing and normal walking. However, the same technique used on data collected from the animal's collar and leg failed to detect normal walking and lame walking. Placing the sensor under the jaw was used for the analysis of feeding behaviour, e.g., chewing and biting, which produced sensitivity and specificity of 97.4% and 97.7%, respectively [54]. Additionally, other types of activities were discriminated from biting and chewing with 100% sensitivity, when the sensors were attached under the jaw [54].

Table 1 Sensor placement vs animal activity

Sensor Placement	References	Animal activity domain
Collar	[2,5, 33,39,40,42–44,51,52, 53–58,]	Collar-borne devices were used to classify “active” vs “inactive” behaviour, or “grazing” vs “non-grazing”. Additionally, collar-borne devices were used in multiclass classification to discriminate behaviours such as “grazing”, “browsing”, “foraging”, “standing”, “walking”, “running”, “resting”, “lame walking”.
Leg	[5,37,51,58]	Leg-borne devices were used to identify behaviours such as “walking”, “lame walking”, “trotting”, galloping”, “running”, “resting”, “grazing”.
Ear	[5,6,34,37,50–53]	Ear-borne sensors were used to identify behaviours such as “lame” vs “not-lame”, posture (upright vs prostrate), “grazing”, “lying”, “standing”, “walking”.
Under the jaw	[54–57]	Jaw-based sensors were used to identify “biting”, “chewing”, “grazing”, “lying”, and “standing” behaviours.

2.3. Sheep Activities

Various studies related to SAR task using ML attempted to identify different types of activities. The activities of animals are discriminated based on domain experts’ knowledge. Table 2 summarizes the activities addressed within the related literature. These activities differ in their complexity in terms of being classified. The most common activities found in research studies are grazing, walking, standing, resting, and lying. However, other types of activities such as biting, chewing, ruminating, and foraging are also studied as shown in Table 2.

Table 2 Description of sheep activity and behaviour

Behaviour	Description	Reference
Grazing	Eating sward at ground level with the head down.	[2,5, 34,37,47,48,51,52,55,56,57,59–62,64–66]
Infracting	Eating from branches above a certain height.	[65]
Browsing	Eating the leaves of shrubs or trees with the head up off the ground.	[66]
Chewing	Rotation of the lower jaw after a bite activity in any head position (up or down).	[54]
Biting	Gathering forage (browse or grass) with incisor teeth.	[54]
Ruminating	Usually performed with the body lying in the sternal position. Fermentation of digesta in the reticulo-rumen complex frequently accompanied by cud-chewing.	[2,52,56,57]
Foraging	A general term for the acquisition of nutrients with the ingestive apparatus: teeth, lips, and tongue.	[63]
Walking	Four-time slow quadrupedal locomotion in sheep; speed 1.1–1.3 m.s ⁻¹ .	[2,5, 6, 34,37,47,50,51,53, 58–64,66]
Moving	This is an intended movement from one place to another. Naturally, the sheep is not looking for nutrition.	[65]
Running/ Trotting	Two-time quadrupedal locomotion in sheep; speed 1.41–2.41 m.s ⁻¹ [67].	[55,58,60,61,63–65]
Galloping	Four-time rapid quadrupedal locomotion in sheep; speed 2.28–3.56 m.s ⁻¹ [67].	[58,63,65]
Scratching	Rubbing body surface against a solid object.	[47,59,66]
Standing	Standing with all four feet on the ground.	[2,5,6,34,37,50,51,53,55,59–66]
Resting	Lying in the absence of rumination. Usually performed with the body lying in sternal position or infrequently with the body lying horizontally in a lateral position.	[2,5,6,34,37,50,51,53,55,56,59–64,66]
Active	There is body movement, e.g., locomotion, foraging, scratching.	[2,34,62]
Inactive	There is no movement; the sheep are lying down to ruminate or are asleep in sternal or lateral recumbence.	[2,47,62]
Upright	The body standing in the vertical position.	[34]
Prostrate	The body lying in the horizontal position.	[34]
Lame	Asynchronous gait commonly due to lameness in one or more limbs; usually, the hoof.	[5,6]
Not Lame	Normal gait, related to normal walking, trotting, and galloping.	[6]

Classification of animal activity to active or inactive states has a low degree of complexity to be classified and therefore, can be easily distinguished by utilizing conventional ML methods. It is noted in the literature that decreased animal activity or hyperactivity could be an indicator of disease and distress [7]. This kind of information is valuable for farm managers and related individuals for further investigation and appropriate treatment when required. On the other hand, detection of speed [60] and direction of running is necessary in hazardous cases such as identification of a thief or a predator pursuing an animal, specifically in remote locations; therefore, classifying trotting or running is essential. However, multiple studies indicated that trotting is generally the most challenging gait to determine [36,58,68].

Identification of real-time foraging activity is essential for sheep farmers working in extensive agricultural hill systems [69]. These types of grazing systems characterize the bulk of the sheep farming industry in the UK and other parts of the world. Also, changes in the eating behaviour of sheep could indicate health or management problems, e.g.,

quality of pasture [70]. Similarly, continuous monitoring of food intake in real-time could provide better estimations of carcass value at market and grazing impact on the sward. It could also be a valuable land management tool, preventing the occurrences of dangerous ecological tipping points, leading to overgrazing, soil erosion, and water contamination, particularly in sensitive upland ecosystems [71,72].

Additionally, studies are conducted to identify lameness, one of the most common and persistent health problems in sheep flocks around the world [5–10]. Unusual amounts of lying time were shown to be an indication of lameness in cattle and sheep [73–75], and therefore lying is a critical activity to be detected in autonomous monitoring systems.

2.4. Data collection and labelling

Data collection and labelling (i.e., annotation) of the raw dataset are essential steps of identification of sheep activities and behaviour. Usually, the process involves capturing the animals' activity in their natural environment, having the sensors logging motion signals from the collar, leg, ear, or mounted under the jaw (as described in Table 1). Video recordings of animals are usually observed by the domain experts to label the animal activities. There are tools for labelling the data, such as the ELAN_5.7_AVFX Freeware tool [76], however most authors perform the labelling manually having the data measurements and video recordings timestamped to serve as a ground truth during the labelling step. The camera is set to also record the time in HH(hour):MM(minute):SS(second) format, so it can be easily synchronized with the timestamped data measurements. The camera is usually placed in the pasture having a clear view of the selected animal or all the animals. During the video recordings, an observer is present and is responsible to move the camera if the animals are out of view. The animal is either recognized by the colour of the tracking device or is numbered with spray on its body, so the observer can recognize the corresponding animal. The various behaviours are labelled based on expert knowledge. Another important factor to be considered is the number of animals involved in data collection. This is because animals exhibit different characteristics such as age, height, and health status and therefore, larger sample would be required to capture such diversities. However, metadata such as age, height, and health status are not integrated in the datasets, but it can be considered during the selection of the animals for the data collection. For example, an animal suffering from lameness will differ in gait patterns from a healthy animal [7]. Additionally, younger sheep might be more active than older ones and behaviours such as walking and running may vary. Therefore, the animal selection plays a crucial role in data collection since having multiple behaviours from a variety of animals can ensure more representative training data, resulting in improved predictive model characteristics, e.g., adapting better when new animals are added to the flock.

2.5. Windowing and sample rate

Accelerometer measurements are collected in time intervals (milliseconds, seconds, minutes etc.), forming a time series dataset that needs to be analysed in overlapping intervals. A technique which is commonly used to slice the time series signals into the overlapped frames is known as 'windowing' [77]. This step is critical since accelerometer signals provide valuable information about motion patterns in slots of data and not as a single variable measure. The windows are of the same size and are either disjoint [2,5,34,37,47,54,57,59,66] or overlapping, typically at 50% [77]. Overlapping windows are suggested because of their ability to capture the transitions of activities more precisely [77]. On the other hand, very small disjoint windows can avoid transitions, but may lose important information from the signal [77]. Therefore, research studies analysed the effects of varying window size to identify the appropriate size, containing sufficient features to discriminate the gait patterns, while simultaneously avoiding misclassification. This can result from comparatively longer windows due to activity transitions, i.e., walking to grazing, standing to lying. Additionally, windowing process affects the computational complexity of the feature extraction process, which must also be taken into account [78]. In relation to SAR, various window sizes were tested and evaluated in the literature, as shown in Table 3. It can be observed that 5s and 10s windows have commonly been used.

Similar to window size, the sampling rate is an important factor presenting the quantity/quantity/number of samples in each window. The choice of sampling rate influences accelerator signal information and subsequently, feature extraction, while playing an important role in terms of the time complexity and power consumption of the device. Various choices of sampling rate were reported in the literature and are presented in Table 3. An illustration of the choice of sampling rate vs window size selection in SAR is shown in Figure 2. The plot shows that the commonly used window size varies between 3s to 10s with sampling rates between 10Hz to 20Hz. The choice of window size and the sampling rate is further analyzed in Section 3.2.

Table 3 Variation of window sizes in the literature

Window size (seconds)	Reference
-----------------------	-----------

0.50	50Hz[65]
1	25Hz[54], 200Hz[63]
3	5Hz [55], 8Hz [50], 10Hz[55], 12Hz[37], 16Hz[50], 25Hz[55][54], 32Hz[50], 33Hz[58]
5	5Hz[55], 8Hz[50], 10Hz[55], 12Hz[37], 12.5Hz[47][34], 16Hz[50], 20Hz[48], 25Hz[55][54], 32Hz[50], 62.5Hz[57]
5.12	100Hz[60]
5.30	100Hz[79]
6.4	100Hz[64]
7	8Hz[50], 16Hz[6,50,52], 32Hz[50]
10	5Hz[55], 10Hz[55,66], 12Hz[5,34,37,51], 16Hz[53], 20Hz[48], 25Hz[55], 62.5Hz[57]
15	20Hz[48]
25	32Hz [2]
30	12.5Hz[34], 200Hz[59], 62.5Hz[57]
60	62.5Hz[56], 62.5Hz[57]
120	62.5Hz[57]
180	62.5Hz[57]
300	62.5Hz[57]

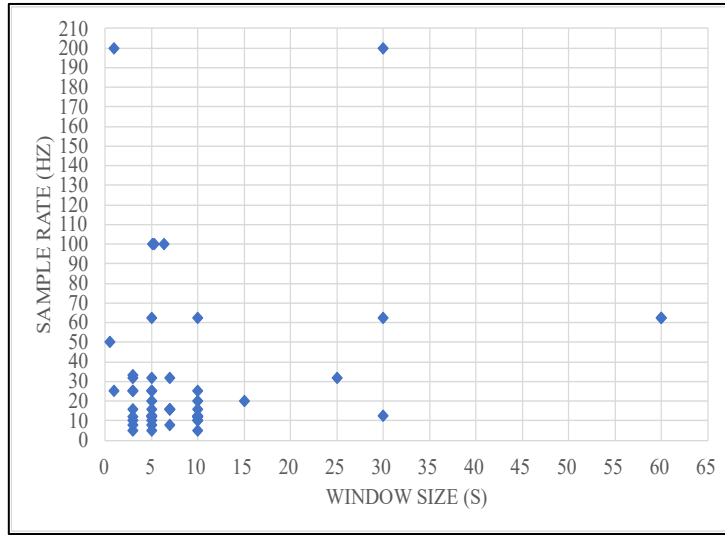


Figure 2 Window size and sample rate used in sheep activity recognition studies

2.6. Time and frequency domain features

Feature extraction is an important step in classification problems [80,81] as well as in SAR task. In case of time series data, several continuous accelerometer measurements are required in order to be able to capture the useful activity patterns. However, this is not the case with several alternative measurement modalities which can provide information from a single value. A variety of techniques have been suggested to transform the data captured through raw accelerometer into useful form (i.e., features) that are further used by the ML algorithms to classify the gait activities [82–84]. Previous research in the field considered use of an extensive number of time and frequency domain features (refer to Tables 4-5). Examples of time domain features mainly include statistical parameters such as the mean, variance, correlation [60,64], higher-order moments, and sensor-based measurements such as pitch, yaw, roll, and inclination angles [54,65]. The major advantage of time domain features is that they are easy to extract and therefore, in most cases, they are computationally efficient [45,85]. However, they are affected by measurement and calibration errors [86]. Frequency-domain features on the other hand, (e.g., signal area, spectral entropy, peak frequency, etc.,) often require additional processing that include windowing, filtering, and other form of frequency domain analysis such as Fourier transform, wavelet transform, cosine transform etc. However, they are able to robustly represent the information in the signal specifically in case of time series data. Thus, they are more computationally expensive than time-domain features [87–89]. In Tables 4 and 5, we present an extensive list of time and frequency domain features, respectively, used for the SAR task. It can be observed that most commonly used time domain features include the mean, standard deviation, minimum, maximum, movement intensity, skewness, kurtosis, energy, and entropy. Likewise, spectral entropy, and peak frequency, are identified as the most commonly used frequency domain features. Note that reference to three signals in the formulae of Table 4 is due to the x, y, and z axes of the associated sensors. A detailed discussion of feature extraction methods in the context of animal activity recognition is given in Section 3.3.

Table 4 Time-domain features in sheep activity recognition

Time-Domain Features		
Name	Description and Formula	Reference
Mean/Average	$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{j,i}, j = 1,2,3$	[5,6,21,33,39,40,42–45,46,48,49,51–55,57,58]
Average all-axis	$\bar{x} = \frac{1}{m} \sum_{j=1}^3 x_j$	[34]
Standard deviation	$s_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{j,i} - \bar{x}_j)^2}$	[6,21,39,40,42,44–46,52,53,54,55,57,58]
Variance	s_j^2	[56,57,60,61,64,65]
Inverse coefficient of variation	$\frac{\bar{x}_j}{s_j} \times 100\%$	[56,57]
Average Standard deviation	$\bar{s} = \frac{1}{3} \sum_{j=1}^3 s_j$	[34]
Median ¹	$med(x_j) = \begin{cases} x_{j,\frac{n}{2}}, & \text{if } n \text{ is even} \\ \frac{x_{j,\frac{n-1}{2}} + x_{j,\frac{n+1}{2}}}{2}, & \text{if } n \text{ is odd} \end{cases}$	[63]
Minimum ¹	$\min(x_j) = x_{j,1}$	[5,6,21,42–46,51,54,55,57,58]
Maximum ¹	$\max(x_j) = x_{j,n}$	[5,6,21,42–44,46,51,54,55,57,58]
Pairwise correlation between axes j and k	$corr(x_j, x_k) = \frac{cov(x_j, x_k)}{s_j s_k}, \text{ where } cov(j, k) \text{ is the covariance}$	[60,61,64]
Mean distance between axes j and k	$dist(x_j, x_k) = \frac{1}{n} \sum_{i=1}^n (x_{j,i} - x_{k,i})^2$	[64]
25 th /75% percentile ¹	$m_p = \left\lceil \frac{P}{100} \times n \right\rceil, \text{ where } P = 25, 75$	[63]
Movement Intensity / Average Signal Magnitude Vector	$\frac{1}{n} \sum_{i=1}^n \sqrt{x_{1,i}^2 + x_{2,i}^2 + x_{3,i}^2}$	[5,34,37,51,53,60,61,64]
Movement Variation	$\frac{1}{n} \left(\sum_{i=1}^{n-1} x_{1,i+1} - x_{1,i} + \sum_{i=1}^{n-1} x_{2,i+1} - x_{2,i} + \sum_{i=1}^{n-1} x_{3,i+1} - x_{3,i} \right)$	[5,34,37,51,54,55,65]
Signal Magnitude Area (SMA)	$\frac{1}{n} \left(\sum_{i=1}^n x_{1,i} + \sum_{i=1}^n x_{2,i} + \sum_{i=1}^n x_{3,i} \right)$	[5,34,37,48,51,54,55]

¹ Observations have been ranked in ascending order.

Signal Vector Magnitude (SVM)	$\sqrt{x_{1,i}^2 + x_{2,i}^2 + x_{3,i}^2}$	[6,48,54,55,65]
skewness	$\frac{1}{n} \sum_{i=1}^n \frac{(x_{j,i} - \bar{x}_j)^3}{s_j^3}$	[6,47,48,60,61,63,64]
kurtosis	$\frac{1}{n} \sum_{i=1}^n \frac{(x_{j,i} - \bar{x}_j)^4}{s_j^4}$	[6,47,48,50,52,53,60,61,63,64]
Interquartile Range¹	$m_{75} - m_{25}$	[6,50,52,53]
rms signal value	$\sqrt{\frac{1}{n} \sum_{i=1}^n x_{j,i}^2}$	[47,59,66]
rms velocity value	$\sqrt{\frac{1}{n} \sum_{i=1}^{n-1} \frac{1}{(x_{j,i+1} - x_{j,i})^2}}$	[47,59,66]
Zero crossing rate (per window)	$count((x_{j,i} - \bar{x}_j) == 0)$	[6,47,50,52,63,65]
Crest factor	$\frac{\max(\sqrt{x_{1,i}^2 + x_{2,i}^2 + x_{3,i}^2})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_{1,i}^2 + x_{2,i}^2 + x_{3,i}^2)}}$	[47]
Peak to Peak	$\max(x_{j,i}) - \min(x_{j,i})$	[47]
Pitch (degrees)	$\tan^{-1} \left(\frac{-x_{1,i}}{\sqrt{x_{2,i}^2 + x_{3,i}^2}} \right)$	[54,55,65]
Roll (degrees)	$atan2(x_{2,i}, x_{3,i}) \times \frac{180}{\pi}$	[54,55,65]
Yaw (degrees)	$atan2(x_{1,i}, x_{2,i}) \times \frac{180}{\pi}$	[65]
Inclination	$\tan^{-1} \left(\frac{\sqrt{x_{1,i}^2 + x_{2,i}^2}}{x_{3,i}} \right)$	[54,55]
Sum of changes	$\sum_{i=1}^{n-1} x_{j,i+1} - x_{j,i}$	[47,59,66]
Mean of absolute changes	$\frac{1}{n} \sum_{i=1}^{n-1} x_{j,i+1} - x_{j,i} $	[47,59,66]

Integrals [90]	$\int_{t=0}^T x_1(t) dt + \int_{t=0}^T x_2(t) dt + \int_{t=0}^T x_3(t) dt$	[47,59,66]
Squared integrals [90]	$\left(\int_{t=0}^T x_1(t) dt\right)^2 + \left(\int_{t=0}^T x_2(t) dt\right)^2 + \left(\int_{t=0}^T x_3(t) dt\right)^2$	[47,59,66]
Madogram [91]	$\frac{1}{2}E[x_{j,i} - x_{j,t+u}]$ where $t=lag$, $E[.]$ =expectation	[47,59,66]
Energy	$(x_{1,i}^2 + x_{2,i}^2 + x_{3,i}^2)$	[5,21,40,43,46,47,52–55,57,58]
Entropy	$(1 + (X_{1,i} + X_{2,i} + X_{3,i}))^2 \times \log(1 + (X_{1,i} + X_{2,i} + X_{3,i}))^2$	[5,34,37,47,48,51,54,55]

Table 5 Frequency-domain features in sheep activity recognition

Name	Description and Formula	Reference
Energy in 1Hz bins	$\frac{1}{N_b} \sum_{n=BN_b+1}^{BN_b+N_b} X_{j,n} ^2$ where X is the Fourier transform of x, Nb is the number of samples in each bin, and B={0,...,9}.	[64]
Spectral entropy [92]	$\sum_{n=1}^N P(X_{j,n}) \times \log \frac{1}{P(X_{j,n})}$ where P(X) is the normalized power spectrum of X.	[6,50,52,60,61,63,64]
Signal Area	$SA = \sum mag \cdot \frac{1}{f_s}$ where mag is the magnitude and f_s is the sampling frequency.	[50]
Absolute signal area	$ASA = \sum mag \cdot \frac{1}{f_s}$	[50]
Peak frequency	$\arg \left(\frac{f_s}{n} \max(P(X_{j,n})) \right)$	[6,47,50,52,59,63,66]
Frequency Magnitudes	Magnitudes of the first six components of the Fourier-transformed signal	[63]
Spectral Area	$2 \sum_{n=1}^N S(f_n) \times \Delta f$ where $S(f_n)$ is the power spectral density at frequency n.	[6]
Harmonic frequency (2nd and 3rd)	Frequencies where the Fourier-transformed signal has its second and third highest power values	[6]

Harmonic ratio	$\frac{\sum_{n=1}^{n/2} f_{2n}}{\sum_{n=1}^{\frac{n}{2}-1} f_{2n+1}}$	[6]
----------------	---	-----

2.7. Feature selection and dimensionality reduction

Feature selection and dimensionality reduction are of major importance in activity recognition. Extracted features may contain irrelevant, duplicate, or misleading information which could affect the predictive or classification tasks [93,94]. While exhaustive search algorithms may be useful in identifying distinctive features, the deployment of exhaustive search is impractical in most of the cases, specifically in high dimensionality datasets. To handle this, a variety of feature selection and dimensionality reduction algorithms have been used to identify the optimal set of features to be further used in the classification/predictive model in various fields [95]. The most commonly used algorithms can be classified into filters, wrappers, and hybrid approaches, as described in the following subsections.

2.7.1. Filter Methods

Filter methods use proxy measures for dimensionality reduction in high dimensional spaces, which mostly include the amount of information, statistical features such as variance, similarity score, consistency etc., [93,96,97]. Likewise, there exist a variety of filter methods to be used depending on the nature of data and hence the task in hand, such as prediction or classification. Various studies have used different information-based filters, e.g., information gain [98][99], gain ratio [100], fast correlation-based filter [101] and symmetrical uncertainty [101]. On the other hand, the Chi-square test [99], Fisher score [102] and feature weighting k-Means [103] are examples of works that use statistical filters. Similarly, Relief and ReliefF [101,104–106] are examples of similarity-based filter methods used in classification and regression problems. A recent work uses filter methods to identify the candidate feature set for human activity recognition with an inertial sensing unit [107]. The algorithm identifies 48 candidate features out of a set of 585 temporal and spectral features, concluding the effectiveness of the selected features and classification algorithm. Additionally, several works related to sheep activity recognition used the Relief method to reduce the number of features for the specific problem [6,48,52,63]. Overall, filter-based methods are comparatively better than wrapper and hybrid methods [108], specifically, in high dimensional feature spaces, due to lower execution times and generalization abilities, as they are independent of the employed supervised algorithm. On the other hand, filter methods are unable to eliminate inter-related features due to univariate analysis [109,110].

2.7.2. Wrapper Methods

Wrapper methods use a subset of feature space recursively to train a predictive or classification model and evaluate the performance over unseen data for the candidate feature set. Selection of subset in each round can be performed through various algorithms such as hold out (forward, backward), selection [111], and heuristic search methods [112]. Finally, the best performing feature set is identified using the test data for the corresponding trained model. One of the major issues with wrapper methods is time complexity specifically, in high dimensional feature spaces [113].

An example of wrapper methods is the Boruta [114,115], which deploys the Random Forest algorithm for recursive selection of candidate features. Work related to sheep activity recognition applied the Boruta for feature selection, prior to fitting the data to the predictive model [59]. Likewise, [116] presented a detailed comparison of RF-based feature selection and standard chemometric methods, when classifying spectral data. Suto et al., [117] presented an interesting study involving wrapper and filter methods in human activity recognition. Specifically, they presented a naïve Bayesian wrapper method, which outperformed filter methods, including Chi-Square, Fisher score and T-test for the task in hand.

Several techniques have been introduced to overcome time complexity issues in wrapper methods, specifically, for the subset selection task. For instance, Bayesian network [118], sequential search using aggregation [119], expectation maximization [120], and beam search [121] are some of the example works towards the optimization of feature search in wrapper methods. There exist several studies related to SAR which utilize wrapper methods in feature selection. For instance, Boruta algorithm is used in [59], Sequential Forward Selection (SFS) is employed in [58, 114], and Recursive Feature Elimination is used in [66]. While these methods yield high classification accuracy, they work better only for the specific models adopted for feature selection. In other words, these methods are computationally expensive as well as lacking in terms of generalization [122].

2.7.3. Embedded and Hybrid Methods

Embedded feature selection methods partly use supervised learning and hence, they are relatively faster compared to wrapper methods. These techniques utilize automated pruning, regularization, or a built-in strategy to select the candidate feature set. For instance, the SVM model can be used to recursively prune a feature with associated variance less than a set threshold. Likewise, decision tree-based approaches such as CART [123], C4.5 [124] and XGBoost [125] are other commonly used embedded methods for feature selection. A study presented in [126] introduced an embedded method for recursive feature elimination using SVM. Feature significance is measured through the associated weights in the trained model, and then used to iteratively eliminate the least important features [127]. A

widely used technique for sheep activity recognition in the literature is the Random Forest feature selection [5,37,51,55].

Alternatively, a variety of hybrid methods were introduced by integrating the properties of wrappers and filter methods [97]. For instance, a filter method (e.g., based on variance in PCA or alternative statistical features) is applied over the entire feature space to identify significant features, which are then forwarded to wrapper methods with the reduced feature set. In this way, the overall complexity and execution times can be reduced to support the use of recursive procedures within wrapper methods. Thus, a hybrid approach tends to be faster and more general than wrapper methods, but slower and less general than filter methods.

2.8. Machine Learning Algorithms

Various supervised ML techniques have been used to classify the activities of animals [128]. In [129], a combination of unsupervised and supervised learning is used. Specifically, the authors propose an automatic behaviour recognition system, which uses spatiotemporal features, incorporating temporal dynamics and invariance with respect to the animal's position and orientation. Unsupervised learning is specifically used in the context of dimensionality reduction, while a supervised classifier processes the corresponding spatiotemporal information. In [130], machine learning approaches for animal activity recognition are being considered, including unsupervised classification of movement data into behavioural modes using hidden Markov models. In this subsection, we present and describe some commonly used classification algorithms used in SAR.

The k-nearest neighbours (KNN) classifier is one of the most commonly used classification algorithms in supervised ML [131,132]. KNN is a non-parametric model, where the classification process is based on the similarity between the training and testing samples. Because of the effectiveness and simplicity of the KNN algorithm, it is widely popular in various disciplines, e.g., data science [133–136]. The algorithm determines the nearest k neighbours for an unseen sample, and then provides its category based on the maximum frequency label in the k nearest neighbours [137]. A description of the KNN is provided in Algorithm 1 in Appendix A.

The SVM is the alternative robust classifier which utilizes the kernel trick in conjunction with supervised learning [138]. The algorithm was first introduced by Boser et al., [139] and further detailed by Cortes [140]. The decision hyperplane generated through SVM depends on the so-called support-vectors. A description of the SVM classifier is provided in Algorithm 2 in Appendix A. It is effective even with high dimensional data making it a powerful ML algorithm for animal behaviour recognitions.

Logistic regression (LR) is a supervised ML algorithm that is a generalization of linear regression. LR is used for prediction as well as binary and multi-class classification. Logistic regression involves the calculation of the prediction function, building the loss function, and determining the regression parameters that are capable of minimizing the loss function. Optimal parameters are determined using iterative optimization techniques [141]. Algorithm 3 in Appendix A illustrates the LR steps.

Decision trees are well-known supervised ML algorithms used in classification and regression. There are various algorithms utilising decision trees (DT) for the classification of data, including ID3 by Quinlan et al., [142], C4.5 by Quinlan [124] and the classification and regression tree (CART) by Breiman et al., [123]. DT algorithms recursively partition the data into subsets, then assigning decision rules to their nodes, as presented in Algorithm 4 in Appendix A, which shows an overview of the DT learning process. In this case, several subsets of data are selected from the training samples to create an ensemble of various models providing more robust decision than a single model. As a result, improved sheep animal behaviour recognition is achieved.

Ensemble models further extend the functionality and hence the ability of conventional DT, while utilizing the

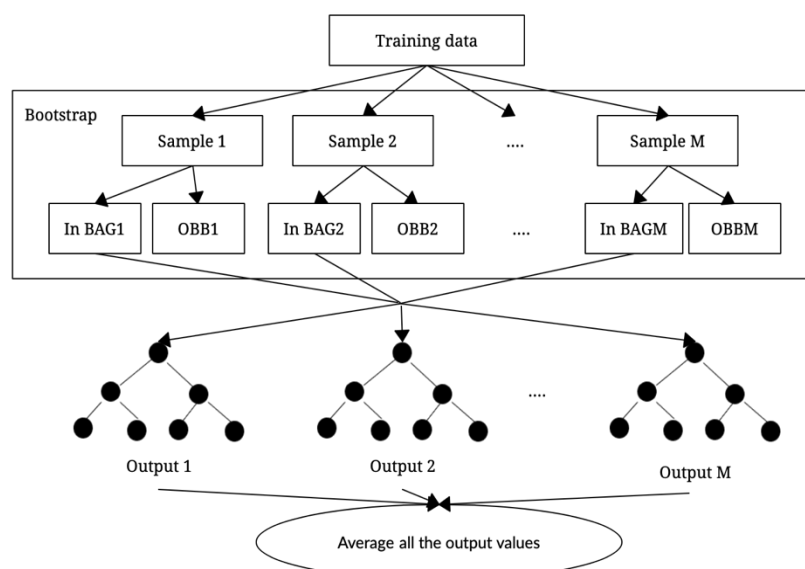


Figure 3 The RF Algorithm [182]

bagging concept (Beriman, 2001). Random Forest is a commonly used ensemble ML model that was initially proposed by Breiman [143]. The algorithm uses the Random Subspaces method [144] and bagging [145] to combine several weak classifiers leading to a robust classification. The algorithm is successfully applied to both prediction and classification tasks. Using the RF, training data are randomly received using subsets to form trees based on a random algorithm [146]. Using the RF algorithm, bootstrap samples (new training set) are selected by substituting the original data set for the tree, allowing a number of training data to be excluded, which can then be reused i.e., out of bag samples. Figure 3 demonstrates the steps used in the RF algorithm utilizing the bagging approach for decision making.

The Naïve Bayes (NB) classifier has been widely employed in a variety of data mining and classification tasks [147]. The algorithm assumes that the probability of a new data sample can belong to a particular class, when the attributes are independent of each other [148]. The algorithm works as follows. Let D_{train} represent a set of training samples of t classification objects. In this case, let the probability $P(y|x)$ for a new data sample $X = \langle x_1, x_2, \dots, x_n \rangle$ to belong to the class $y \in \{1, \dots, c\}$, where x_i represents the value of the attribute. Algorithm 5 in Appendix A shows the basic steps in NB classification.

Multilayer perceptron (MLP) or artificial neural networks (ANNs) are well-known supervised ML algorithms which has been used in a variety of domains including classification and regression [149–156] and successfully solved many computation problems related to signal processing [157]. Additionally ANNs have been applied to face recognition [158], texture classification [159], shape recognition [160] and image segmentation [161]. ANNs have also been used to improve generalization [162]. The external inputs are presented to the network through the input neurons, while the outputs are shown in the output layer. All other layers are called hidden layers. Each layer has its own weights, biases, and transfer functions. The use of more than one layers of nonlinear units makes the network more powerful than a single layer network [163]. As an example, multilayer networks can predict many functions using two layers with sigmoid and linear functions in the first and the second layers, respectively. Multilayered neural networks can be used for pattern classification and function approximation, as well as modelling and prediction [164]. For instance, (Nadimi et al., 2012) used MLP with sheep accelerometer measurements to predict grazing, lying down, walking, standing, and other activities, obtaining an accuracy of 76.2%. Figure 4 shows the structure of a MLP network, where f denotes the transfer function. Algorithm 6 in Appendix A summarizes the forward propagation process in MLP.

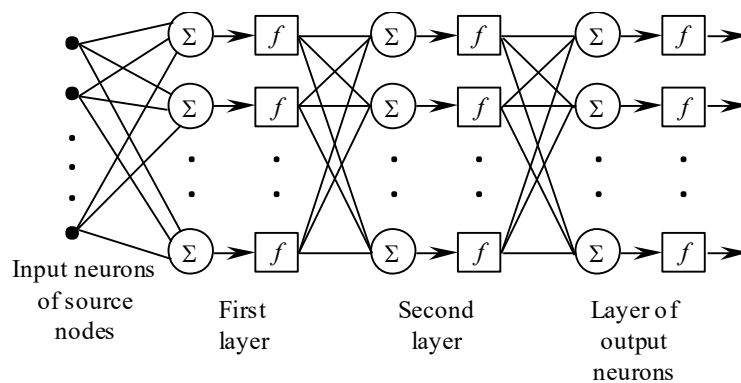


Figure 4 MLP network structure

Deep neural networks (DNN) is considered as a standard neural network with more hidden layers (i.e., a deeper structure). In this case, the depth of the network is defined by the number of hidden layers. It should be noted that there is no preset number of hidden layers for a neural network to be defined as deep, however, Schmidhuber [165] considered that if their credit assignment paths exceed 10, then the corresponding ANN is very deep. The aim of the DNN is to be trained to model complex nonlinearities in the input data by mining unique features. Each of the layers of the DNN aims to extract certain features. Examples of DNN include Convolutional Neural Networks (CNN) that are widely used in image processing [166] and deep belief networks.

The sparsity and the high dimensional properties of the data will result in increased complexity. Linear discriminant analysis [167] has been used for dimensionality and sparsity reduction and represents one of the most favourable tools for the projection into a low dimensional space. There are various applications utilizing linear discriminant analysis (LDA), including image retrieval, speech recognition, and microarray data analysis [168]. Traditional LDA assumes that the dispersion matrices are identical for all classes (i.e., common covariance matrix). When this is not possible, quadratic discriminant analysis (QDA) is used. Unlike LDA, QDA assumes that each of the classes has its own covariance matrix, thus allowing the discriminant function to contain second-order terms, effectively providing for more accurate non-linear decision boundaries. The individual covariance matrices

correspond to a higher number of parameters for QDA, compared to LDA, which means that to avoid overfitting a higher number of sample points is needed [169]. Algorithm 7 in Appendix A presents the basic steps of the LDA algorithm.

Table 6 Classification algorithms in sheep activity recognition

Model	Reference
LR ¹	[64]
Linear SVM ¹ [170]	[34,52,63,64]
Radial SVM ¹	[6]
LDA ¹	[34,48,56–58,60,61,63,64]
QDA ¹	[5,34,37,51,60,64]
CART ¹ [171][142]	[34,54,55,63–65]
RF ¹ [143]	[6,47,50,52,59,64,66]
ANN ¹ [172]	[6,59,62,63]
XGB ¹ [125]	[59]
AdaBoost ¹ [173,174]	[6,52]
DNN ¹	[63]
KNN ¹ [175,176]	[6,52,53,59,63,64]
NB ¹	[63,64]
K-means ²	[53]

1 is for supervised,
2 is for unsupervised

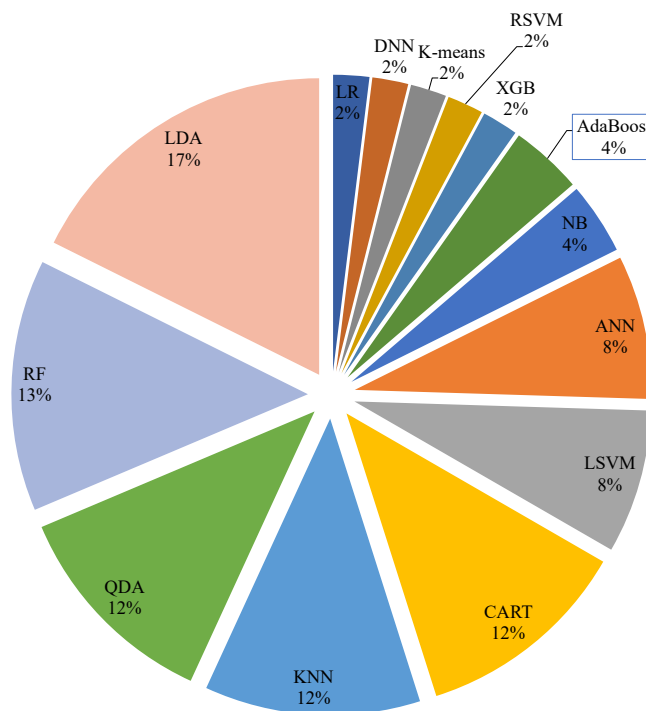


Table 6 presents the ML algorithms used in the SAR problem and illustrates the use of each algorithm in percentage manner. From the illustration, it is clear that the most used algorithm is the LDA (17%), followed by RF (13%), QDA (12%), KNN (12%), and CART (12%). Discussion regarding the performance of the algorithms in SAR is presented in section 3.3.

3. Discussion and summary

Tabular summaries and illustrations of the research works surveyed in relation to sensor placement are provided in Tables 7-10, Figure 6-7, respectively. The tables are sorted by descending accuracy and provide information about the type of activity, sample rate, window size, feature selection method, learning model, and performance (i.e., accuracy). Figure 6 illustrates an overview of the use of ML, sensors, sample rate, window size, feature selection algorithm, and resulted accuracy of the reviewed studies when the sensor is placed on the collar of the animal. Similarly, Figure 7 shows the same parameters as Figure 6, when the sensors are placed on the leg, ear, and under the jaw of the animal. In the following sub-sections, detailed observations are presented based on the information provided in Table 7-10 and Figures 6-7.

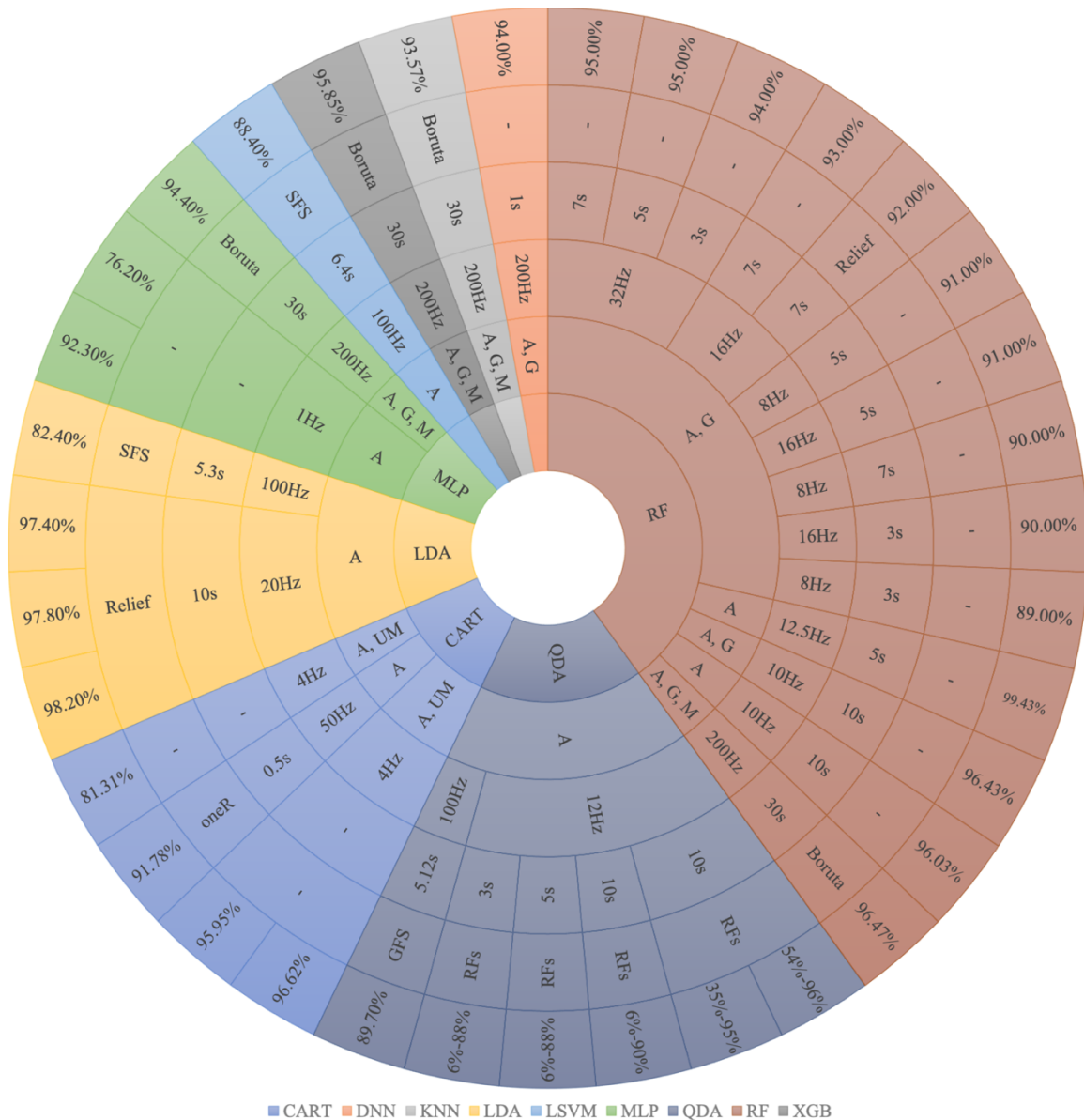


Figure 5 Summary chart for SAR studies using collar-born sensors (refer to Table 7)

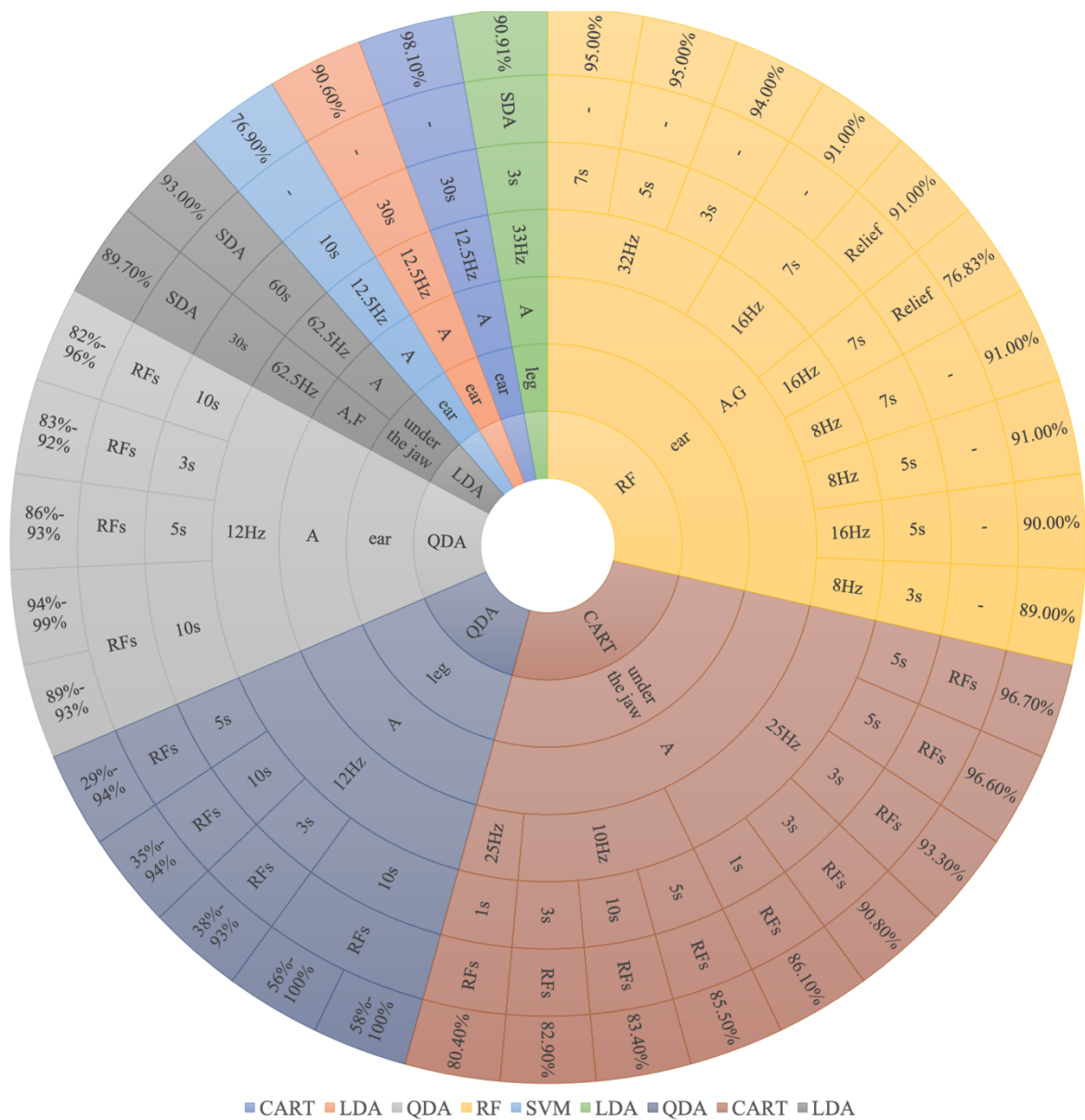


Figure 6 Summary chart for SAR studies using sensors placed on ear, leg, and under the jaw (refer to Table 8-10)

Table 7 Collar-borne sensors in sheep activity recognition

ACC=accelerometer, GYR=gyroscope, MAGN=magnetometer, RFs = random forest feature selection, SFS= Sequential forward selection, GFS=greedy feature selection

Model	Sensor	Sensor Placement	Activities	Sample Rate	Window (s)	Feature Selection	Accuracy	Reference
RF	ACC	collar	grazing, walking, scratching, inactive	12.5Hz	5	-	99.43%	[47]
LDA	ACC	collar	grazing, not-grazing (tall pasture)	20Hz	10	Relief	98.20%	[48]
LDA	ACC	collar	grazing, not-grazing (short pasture)	20Hz	10	Relief	97.80%	[48]
LDA	ACC	collar	grazing, not-grazing (medium pasture)	20Hz	10	Relief	97.40%	[48]
CART	ACC, ultrasound module	collar	Running, not-running	4Hz	-	-	96.62%	[49]
RF	ACC, GYR,	collar	grazing, lying, standing, walking,	200Hz	30	Boruta	96.47%	[59]

	MAGN		scratching					
RF	ACC, GYR	collar	grazing, lying, standing, walking, browsing, scratching	10Hz	10	-	96.43%	[66]
RF	ACC	collar	grazing, lying, standing, walking, browsing, scratching	10Hz	10	-	96.03%	[66]
CART	ACC, ultrasound module	collar	Posture (Infracting and Not Infracting)	4Hz	-	-	95.95%	[49]
XGB	ACC, GYR, MAGN	collar	grazing, lying, standing, walking, scratching	200Hz	30	Boruta	95.85%	[59]
RF	ACC, GYR	collar	Walking, standing, lying	32Hz	5	-	95.00%	[50]
RF	ACC, GYR	collar	Walking, standing, lying	32Hz	7	-	95.00%	[50]
MLP	ACC, GYR, MAGN	collar	grazing, lying, standing, walking, scratching	200Hz	30	Boruta	94.40%	[59]
DNN	ACC, GYR	collar	stationary, foraging, walking, trotting, running	200Hz	1	-	94.00%	[63]
RF	ACC, GYR	collar	Walking, standing, lying	32Hz	3	-	94.00%	[50]
KNN	ACC, GYR, MAGN	collar	grazing, lying, standing, walking, scratching	200Hz	30	Boruta	93.57%	[59]
RF	ACC, GYR	collar	Walking, standing, lying	16Hz	7	-	93.00%	[50]
MLP	ACC	collar	active, inactive	1Hz	-	-	92.30%	[62]
RF	ACC, GYR	collar	grazing, ruminating, non-eating (walking, standing, lying)	16Hz	7	Relief	92.00%	[52]
CART	ACC	collar	infracting, grazing, standing, moving, running,	50Hz	0.5	oneR	91.78%	[65]
RF	ACC, GYR	collar	Walking, standing, lying	16Hz	5	-	91.00%	[50]
RF	ACC, GYR	collar	Walking, standing, lying	8Hz	5	-	91.00%	[50]
RF	ACC, GYR	collar	Walking, standing, lying	16Hz	3	-	90.00%	[50]
RF	ACC, GYR	collar	Walking, standing, lying	8Hz	7	-	90.00%	[50]
QDA	ACC	collar	grazing, lying, standing, walking, running	100Hz	5.12	GFS	89.70%	[60]
RF	ACC, GYR	collar	Walking, standing, lying	8Hz	3	-	89.00%	[50]
Linear SVM	ACC	collar	grazing, lying, standing, walking, running	100Hz	6.4	SFS	88.40%	[64]
LDA	ACC	collar	grazing, lying, standing, walking, running	100Hz	5.3	SFS	82.40%	[61]
CART	ACC, ultrasound module	collar	Resting vs Not resting	4Hz	-	-	81.31%	[49]
MLP	ACC	collar	grazing, lying, standing, walking and others	1Hz	-	-	76.20%	[62]
QDA	ACC	collar	grazing, standing, walking, resting	12Hz	10	RFs	54%-96%	[51]
QDA	ACC	collar	grazing, standing, walking, resting, lame walking	12Hz	10	RFs	35%-95%	[5]
QDA	ACC	collar	grazing, lying, standing, walking	12Hz	3	RFs	6%-88%	[37]
QDA	ACC	collar	grazing, lying, standing, walking	12Hz	5	RFs	6%-88%	[37]

QDA	ACC	collar	grazing, lying, standing, walking	12Hz	10	RFs	6%-90%	[37]
-----	-----	--------	-----------------------------------	------	----	-----	--------	------

Table 8 Ear-borne sensors in sheep activity recognition

ACC=accelerometer, GYR=gyroscope, RFs = random forest feature selection

Model	Sensor	Sensor Placement	Activities	Sample Rate	Window (s)	Feature Selection	Accuracy	Reference
CART	ACC	ear	active, inactive	12.5Hz	30	-	98.10%	[34]
RF	ACC, GYR	ear	standing, walking, lying	32Hz	5	-	95.00%	[50]
RF	ACC, GYR	ear	standing, walking, lying	32Hz	7	-	95.00%	[50]
RF	ACC, GYR	ear	standing, walking, lying	32Hz	3	-	94.00%	[50]
QDA	ACC	ear	grazing, standing, walking	12Hz	10	RFs	94%-99%	[51]
RF	ACC, GYR	ear	standing, walking, lying	8Hz	5	-	91.00%	[50]
RF	ACC, GYR	ear	grazing, ruminating, non-eating	16Hz	7	Relief	91.00%	[52]
RF	ACC, GYR	ear	standing, walking, lying	16Hz	7	-	91.00%	[50]
RF	ACC, GYR	ear	standing, walking, lying	8Hz	7	-	91.00%	[50]
LDA	ACC	ear	posture (upright and prostrate)	12.5Hz	30	-	90.60%	[34]
RF	ACC, GYR	ear	standing, walking, lying	16Hz	5	-	90.00%	[50]
RF	ACC, GYR	ear	standing, walking, lying	8Hz	3	-	89.00%	[50]
QDA	ACC	ear	grazing, standing, walking	12Hz	10	RFs	89%-93%	[37]
QDA	ACC	ear	grazing, standing, walking	12Hz	5	RFs	86%-93%	[37]
QDA	ACC	ear	grazing, standing, walking	12Hz	3	RFs	83%-92%	[37]
QDA	ACC	ear	grazing, standing, walking, lame walking	12Hz	10	RFs	82%-96%	[5]
Linear SVM	ACC	ear	grazing, lying, standing, walking	12.5Hz	10	-	76.90%	[34]
RF	ACC, GYR	ear	lame, not lame (within Walking)	16Hz	7	Relief	76.83%	[6]

Table 9 Jaw-based sensors in sheep activity recognition

ACC=accelerometer, RFs = random forest feature selection, SDA=stepwise discriminant analysis

Model	Sensor	Sensor Placement	Activities	Sample Rate	Window (s)	Feature Selection	Accuracy	Reference
CART	ACC	under the jaw	bite, chewing, other (grazing pasture plots, different sward height treatments combined)	25Hz	5	RFs	96.70%	[54]
CART	ACC	under the jaw	bite, chewing, other (while grazing micro-sward boxes)	25Hz	5	RFs	96.6%	[54]
CART	ACC	under the jaw	bite, chewing, other (grazing pasture plots, different sward height treatments combined)	25Hz	3	RFs	93.30%	[54]
LDA	ACC	under the jaw	grazing, ruminating, and resting	62.5Hz	60	SDA	93.00%	[56]
CART	ACC	under the jaw	bite, chewing, other (while grazing micro-sward boxes)	25Hz	3	RFs	90.80%	[54]
LDA	ACC, FORCE	under the jaw	grazing, ruminating, other	62.5Hz	30	SDA	89.70%	[57]

CART	ACC	under the jaw	bite, chewing, other (grazing pasture plots, different sward height treatments combined)	25Hz	1	RFs	86.10%	[54]
CART	ACC	under the jaw	grazing, lying, standing, walking, running	10Hz	5	RFs	85.50%	[55]
CART	ACC	under the jaw	grazing, lying, standing, walking, running	10Hz	10	RFs	83.40%	[55]
CART	ACC	under the jaw	grazing, lying, standing, walking, running	10Hz	3	RFs	82.90%	[55]
CART	ACC	under the jaw	bite, chewing, other (while grazing micro-sward boxes)	25Hz	1	RFs	80.40%	[54]

Table 10 Leg-based sensors in sheep activity recognition

ACC=accelerometer, RFs = random forest feature selection, SDA=stepwise discriminant analysis

Model	Sensor	Sensor Placement	Activities	Sample Rate	Window (s)	Feature Selection	Accuracy	Reference
LDA	ACC	leg	walking, trotting, galloping (only horizontal axis was used)	33Hz	3	SDA	90.91%	[58]
QDA	ACC	leg	grazing, standing, walking, resting, lame walking	12Hz	10	RFs	58%-100%	[5]
QDA	ACC	leg	grazing, standing, walking, resting	12Hz	10	RFs	56%-100%	[51]
QDA	ACC	leg	grazing, lying, standing, walking	12Hz	3	RFs	38%-93%	[37]
QDA	ACC	leg	grazing, lying, standing, walking	12Hz	10	RFs	35%-94%	[37]
QDA	ACC	leg	grazing, lying, standing, walking	12Hz	5	RFs	29%-94%	[37]

3.1. Accelerometers and sensor placement

As mentioned in Section 1, accelerometers have been widely used in animal activity recognition due to their ability to distinguish various behavioural patterns with high accuracy. It is important to note that when using accelerometers, the absolute acceleration feature must be considered as it diminishes the effect of sensor orientation, which can adversely affect the performance of predictive models [37]. Indeed, it has been previously reported that change in sensor positioning could affect the results [177]. Several works such as e.g., [50,59,66] integrated the accelerometers with gyroscopes and magnetometers however, the outcomes in above tables indicate no substantial improvement in accuracy.. Therefore, it is suggested that an accelerometer sensor suffices in accurately identifying animal behaviour.

There is a common trend in attaching the sensor on the collar and ear. For instance, a sensor attached on the collar successfully classified grazing, walking, scratching, and inactivity with accuracies above 99.13% [47]. On the other hand, a collar sensor is not recommended when the purpose of the study is concerned with lameness [5]. Lame walking was classified with an accuracy of 87% when the sensor was attached to the leg as compared to ear attachment that produced 82% accuracy [5]. On the other hand, grazing activity can be identified with an accuracy in excess of 97% when the sensor is attached to the collar [47,48]. It can be noted from Table 7-10 that activity recognition performance depends various factors that include extracted features, ML techniques, sensor placement, and window size. When reviewing the locations of sensor placement, the advantage of ear-based sensors is the ability to be integrated to existing ear-tags on animals. Based on the reported results of the reviewed studies in Tables 7-10 [34,47,48,54], we can conclude that one sensor per animal suffices in producing satisfactory predictive results in identifying sheep activities patterns.

3.2. Windowing and sample rate

The selection of the window size (to analyze the overlapping slots of time series data) and sample rate has significant impacts on sheep activity classification results. The choice of window size always depends on the activity to be detected. For example, [54] reported that the classification of biting, chewing, and other feeding activities when acquired from under the jaw of the animal, increased the accuracy when the window size was increased from 3 s to 5 s [54]. On the other hand, [55] evaluated the classification accuracy for grazing activity using varying window size (3,

5, and 10-second windows) and reported no significant difference in the performance; however, the highest accuracy for running was achieved using the 10s window[55]. This indicates the dependency of window size on the activity to be classified. However, using larger window size in real-time animal activity classification may lead to mislabeling because, animal may exhibit more than one activity in comparatively larger time interval. A study presented in [39] conducted experiments to identify the impact of varying size window over the animal activity recognition performance. The outcomes reported 5 seconds as an optimal window size for SAR task.

In addition to window size selection, previous studies also reported that a lower sampling frequency improves the memory usage as well as less power demanding [178]. One of the existing study [50] evaluated the effect of sampling frequency and window size on power consumption to identify sheep behaviour. The study outcomes suggested that a sampling frequency of 16Hz using 7s window size is useful in terms of less power consumption. The study also reported higher accuracy using a 32Hz sampling rate however, the results were close to those reported for 16Hz from [50]. Another work presented in [61] performed spectral analysis on a sheep dataset collected with a sample rate of 100Hz. The author reported that limited spectral information is available above 10Hz to distinguish the animal activities.[61] To summarize, the sampling rate and window size need to be chosen based on the specific animal activity recognition problem, in the context of memory and power consumption application constraints.

3.3. Feature extraction, feature selection, and classification

During the course of this survey study, it became evident that there is no universal method for feature selection in SAR while using the accelerometer signals. From the reviewed studies (Section 2.6, Tables 4-5), it is observed that the majority of works use time-domain features as they are computationally efficient. While the frequency-domain features are robust to noise, they are relatively computationally expensive and therefore, requiring more power [87]. A variety of feature selection methods have been deployed for the optimal features selection from various time and frequency domain features extracted from the accelerometer data.

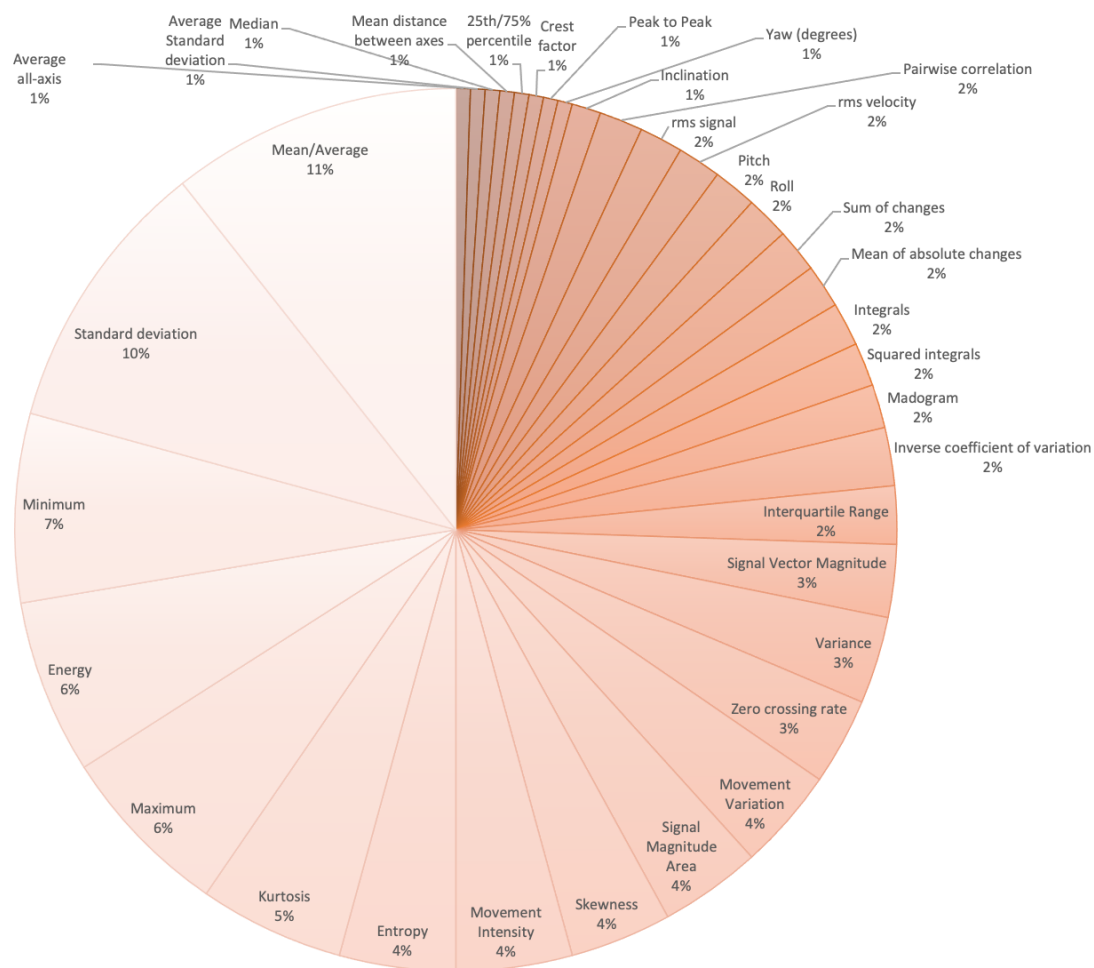


Figure 7 Use of time-domain features in SAR

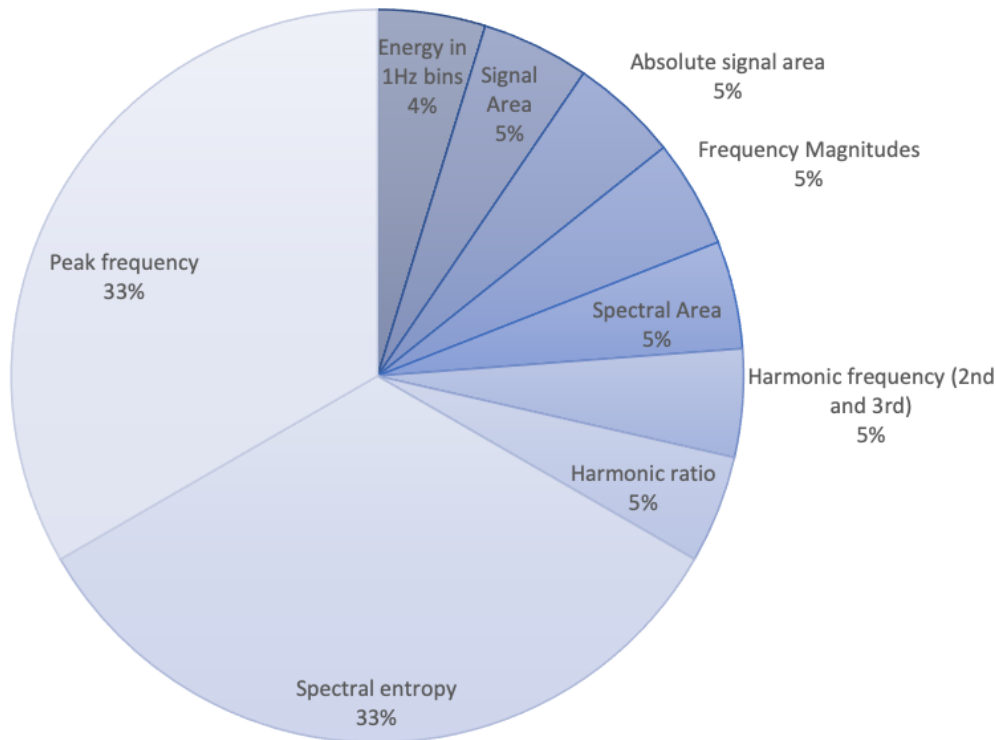


Figure 8 Use of frequency-domain features in SAR

Visualisation of time and frequency domain features used in the SAR studies are shown in Figures 8 and 9, respectively. From Figure 8, it is clear that the most commonly used time-domain features are the mean (11%), standard deviation (10%), minimum (7%), energy (6%), and maximum (6%). In Figure 9, it is shown that the most commonly used frequency-domain features are the peak frequency (33%) and spectral entropy (33%).

From the survey outcomes (Section 3, Tables 7-10), it can be observed that the RFs is identified as the most commonly used feature selection approach. In relation to optimal classification model in SAR task, it was observed that the most commonly used methods are LDA, QDA, RF, CART, and KNN (refer to Table 6, Section 2.8). However, from Tables 7-10, it can be observed that the highest accuracies are achieved using RF (99.43%), LDA (98.20%), and CART (98.10%). Given the different settings for each study, i.e., type of animal activities, and sensor position etc., a direct comparison in regard to classification or prediction performance is not straightforward. Tables 11-14 provide an overview of the applied feature selection methods and final set of features, depending on the location of the sensors. The outcomes in Tables 11-14 are sorted in terms of descending accuracy. It is envisaged that the information presented in these tables could be used to provide guidance in future research studies in relation to the sensor position, activity type, and overall system requirements.

Table 11 Top performances in collar-borne sensors: Feature selection and feature set

Model	Sensor	FS	Final Features	Activities	Results per activity	Accuracy	Ref
RF	ACC	correlation >80	mean, crest factor, root mean square velocity, skewness, kurtosis, madogram, zero crossing rate, squared integrals, and signal entropy	grazing, walking, scratching, inactive	acc 99.08% acc 99.13% acc 99.90% acc 99.85%	99.43%	[47]
LDA	ACC	Relief	entropy az, mean az, mean gy, mean gx, entropy ay	grazing, not-grazing (tall pasture)	acc 98.2%	98.20%	[48]

LDA	ACC	Relief	mean az, entropy az, mean gy, entropy ay, mean gx	grazing, not-grazing, (short pasture)	97.80%	97.80%	[48]
LDA	ACC	Relief	entropy az, mean gy, mean az, entropy ay, mean gx	grazing, not-grazing (medium pasture)	97.40%	97.40%	[48]
CART	ACC	No	the 3-axis dynamic acceleration results only considered one axis to differentiate activity	Running/not-running	96.62%	96.62%	[49]
RF	ACC, GYR, MAGN	Boruta	mean, sd, rms, rms velocity, energy, sum of changes, mean absolute change, integrals , squared integrals, madogram, peak frequency	grazing, lying, standing, walking, scratching	sens 97.66% spec 97.74% sens 93.22% spec 99.76% sens 97.32% spec 98.50% sens 96.23% spec 99.53% sens 95.70% spec 99.74%	96.47%	[59]
RF	ACC, GYR	NA	mean, sd, rms, rms velocity, energy, sum of changes, mean absolute change, integrals , squared integrals, madogram, peak frequency	grazing, lying, standing, walking, browsing, scratching	sens 94.90% spec 98.21% sens 97.29% spec 99.34% sens 95.48% spec 97.46% sens and spec 100% sens 78.91% spec 99.99% sens 90.91% spec 100%	96.43%	[66]

Table 12 Top performances in ear-borne sensors: Feature selection and feature set

Model	Sensor	FS	Final Features	Activities	Results per activity	Accuracy	Ref
CART	ACC	No	average, average all axis, minimum, maximum, sd, average sd, movement intensity, signal magnitude area, energy, entropy, movement variation	Active, inactive	sens 97.4% spec 98.5% sens 98.5% spec 97.4%	98.10%	[34]
QDA	ACC	RF	Movement Variation, Average Intensity, Average y	grazing, standing, walking	acc 94% acc 96% acc 99%	94-99%	[51]

Table 13 Top performances in leg-mounted sensors : Feature selection and feature set

Model	Sensor	FS	Final Features	Activities	Results per activity	Accuracy	Ref
QDA	ACC	RF	Average x, Signal Magnitude Area, Average Intensity	grazing, lying, standing, walking, lame walking	acc 89% acc 100% acc 58% acc 64% acc 87%	58%-100%	[5]
QDA	ACC	RF	Average x, Signal Magnitude Area, Movement Variation	grazing, lying, standing, walking	acc 91% acc 100% acc 56% acc 100%	56%-100%	[51]
QDA	ACC	RF	Average x , Signal Magnitude Area and Movement Variation	grazing, lying, standing, walking	acc 76% acc 38% acc 48% acc 93%	38%-93%	[37]
QDA	ACC	RF	Average x, Signal Magnitude Area and Movement Variation	grazing, lying, standing, walking	acc 81% acc 35% acc 47% acc 94%	35%-94%	[37]
QDA	ACC	RF	Average x, Signal Magnitude Area and Movement Variation	grazing, lying, standing, walking	acc 80% acc 29% acc 48% acc 94%	29%-94%	[37]

Table 14 Top performances in jaw-mounted sensors: Feature selection and feature set

Model	Sensor	FS	Final Features	Activities	Results per activity	Accuracy	Ref
CART	ACC	RF	mean of Movement Variation and Energy	Bite, chewing, other (grazing pasture plots, different sward height treatments combined)	sens 97.4% spec 97.7% sens 96.3% spec 96.8% sens 95.4% spec 100%	96.70%	[54]
CART	ACC	RF	mean of Movement Variation and Energy	Bite, chewing, other (while grazing micro-sward boxes)	sens 96.3% spec 98.4% sens 95.1% spec 97.4% sens 100.0% spec 99.0%	96.60%	[54]

Table 7 and Figure 6 shows that studies with the sensors attached to the animals' collar indicated high accuracy in identifying various types of activities. Best results show that accuracy range of 97.40%-99.43%, were obtained using the RF and LDA as classification models. RF yielded an overall accuracy of 99.43% for grazing, walking, scratching, and inactive [47]. On the other hand, LDA yielded an accuracy of 98.20% in binary classification, i.e., grazing, and non-grazing [48]. These two studies used different combinations of features. For example, the RF approach used features including mean, crest factor, root mean square velocity, skewness, kurtosis, madogram, zero crossing rate, squared integrals, and signal entropy to obtain an accuracy of 99.08% for grazing [47]. While [48] used entropy and mean with LDA and achieved an overall performance of 98.20% for grazing and non-grazing. On the other hand, Cardoso et al., [49] used data acquired from the collar using only one axis of the accelerometer signal to classify running activity and achieved an accuracy of 96.62%.

Another study [34] reported 98.1% accuracy towards distinguishing active and inactive sheep behaviours. They used sensor attached to the ear, CART as classification model trained over multiple set of features that include mean of each axis, mean of all axes, minimum, maximum, sd, average sd, movement intensity, sma, energy, and entropy (refer to Table 12). On the other hand, [51] acquired ear-mounted sensor data to discriminate between grazing, standing and walking, using QDA and three features (movement variation, average intensity, and average of y-axis). The study reported 94%-99% accuracies for the selected activities. The outcomes from these studies clearly indicate the association between the sensor mounting location, the activities to be distinguished, and the selected features, which indicated impact on the performance of ML models.

A limited literature reported the sensors placement on the leg and under the jaw in relation to SAR task (Refer to Figure 7). The activity classification using leg mounted sensor produced an imbalanced accuracy (refer to Table 13). For example, accuracies of 89%, 100%, 58%, 64%, and 87%, respectively, were obtained for grazing, lying, standing, walking, and lame walking respectively while using QDA and only three features (average of x-axis, signal magnitude area, and average intensity) in [5]. Considering these outcomes, it can be inferred that the selected three features are efficient to discriminate the lying activity (100%) however, demonstrate poor performance in discriminating standing and walking. Likewise, there are few studies reporting potential towards feeding activity classification when sensor are mounted under the jaw. For instance, Alvarenga et al., achieved accuracies of 96.60%, and 96.70% using CART, and identified that the most important features were the mean of movement variation and energy, to discriminate between biting, chewing, and other activities [54].

In summary, there is no universal approach for animal activity recognition. Instead, each activity recognition approach needs to be designed according to the specific aim, objectives, environment, available, and datasets, etc. Depending on the nature of problem, selection of sensor configuration, feature extraction, and ML techniques have associated advantages and disadvantages, which primarily relate to the context of the investigated activities.

4. Conclusion and future directions

Intelligent monitoring and detection of sheep activities using accelerometers and ML is an important research topic, specifically with potential to provide information which might be useful for the efficient decision-making in terms of animal welfare as well as land utilization. In this survey, the problem of SAR was considered in terms of its essential building blocks. A number of aspects were the main focus of this study, including sensor type and positioning, window size and sampling rate, feature extraction, feature selection, and classification methods in relation to SAR. Furthermore, an overview of the foundations of utilized techniques and the opportunities and challenges was presented. Based on the findings from surveyed literature, it is identified that the solution to SAR task depends upon the problem at hand and properties of the available data. For example, an essential aspect in SAR model is the choice

of activities to be identified, as this will significantly influence the proposed methodology, i.e., sensor placement, window size, feature selection, and choice of the classification algorithm. From the review of related state-of-the-art works, it was identified that lameness recognition in sheep using accelerometer signals is a challenge that has not been sufficiently studied, and therefore, offering opportunities for novel contributions in this field.

The majority of related research works focus on using collar-borne and ear-borne accelerometers. Animal activity prediction results indicated that these two sensor positions result in higher accuracy, when compared to sensors mounted on the leg and under the jaw. However, limited work exists addressing these mounting positions. Therefore, more research should be conducted to further explore the advantages of using leg and under the jaw-based sensors for recognizing varying sheep activities. Indeed, a leg-borne sensor can provide valuable information regarding movement activities, while the data collected from the jaw of the animal could provide more information regarding feeding activities, such as grazing, biting, chewing, and ruminating, which are critical behaviours for the sheep industry as well as for conservation purposes.

We further observed that limited studies exist in regard to the use of DL for the SAR task and indeed, this is an avenue that needs to be further explored. Classifying sheep activities using DL could overcome limitations that arise from conventional approaches. For example, when using traditional ML methods, there is a requirement to develop and investigate the appropriateness of feature extraction, which takes valuable time and efforts. This can be resolved by utilizing DL models, which automatically perform the feature extraction during the learning process.

Appendix A:

Consider a set X of n labelled samples, KNN performs the classification task as shown in Algorithm 1.

Algorithm 1: KNN Algorithm

Let y represent the unknown sample

Let $k \in [1, n]$

Repeat

Calculate similarity between y and x_i

If ($i \leq k$)

$x_i \in$ into k nearest neighbour of y

Else if (x_i is close to y than one of the nearest neighbours)

Eliminate the farthest neighbour in the k nearest neighbour set

$x_i \in$ into k nearest neighbour of y

End

$i++$

Until ($i > \text{No of Training data}$)

Algorithm 2. Support Vector Machine Algorithm [179]

Let S represent a set of m data points where $s = \{(x_i, y_i) | i = 1..m\}$

Where $x \in R^n$ and $y \in \{1, -1\}$ for binary classification

Let ϕ be a map function, where $Z = \phi(x)$ and ϕ maps the input space to a high-dimensional dot-product feature space.

Determine the hyperplane define by $w \cdot z + b = 0$ where $w \in R^m$ and $b \in R$.

$\exists (w, b)$

If (S is linearly separable)

$w \cdot z_i + b \geq 1, y_i = 1$

$w \cdot z_i + b < 1, y_i = -1$

Else

$w \cdot z_i + b \geq 1 - \epsilon_i$

$$\min_{w,b,\epsilon} \left(\frac{1}{2} w^T w + c \sum_i^m \epsilon_i \right)$$

$$y_i (w^T \phi(x_i) + b) \geq 1 + \epsilon_i, \forall i = 1, \dots, m$$

$$\epsilon_i \geq 0, \forall i = 1, \dots, m$$

c is a constant

$$K(u, v) = \phi(u) \phi(v)$$

K is the kernel

Algorithm 3: LR algorithm [180]

Let S be the prediction function (a sigmoid function)

Let x be the variable

$$S = \frac{1}{1 + e^{-x}}$$

Let g to be the prediction function and L the Loss function

$$g_w(x) = S(w^T x) = \frac{1}{1 + e^{-w^T x}}$$

$$L(w) = \frac{1}{m} \left[\sum_{i=1}^m (y_i \log g_w(x_i) + (1 - y_i) \log(1 - g_w(x_i))) \right]$$

Determine w using gradient descent to minimize the loss function

Algorithm 4. A general method for learning the DT [123]

Let (X, Y) , $\{x_1, x_2, \dots, x_d\}$, K_{nots} depth, $\text{depth}_{\text{max}}$ and $\text{Gain}_{\text{split}}^{\text{min}}$ as the inputs

Let N_0 the initial node

Calculate $\text{Gain}_{\text{split}}$

For $\alpha \in \{1, \dots, d\}$

 For $\beta \in k_{\text{nots}}$

 Find $\text{Gain}_{\text{split}}$

 Find $\text{Gain}_{\text{split}}^* = \max(\text{Gain}_{\text{split}}, \text{Gain}_{\text{split}}^*)$

 Record optimal split variable x_{v^*} and split the knot β^*

 End

End

Update the nodes based on x_{v^*} and β^*

depth ++

if $\text{Gain}_{\text{split}}^* > \text{Gain}_{\text{split}}^{\text{min}}$ Or $\text{depth} > \text{depth}_{\text{max}}$

 exit

end

Algorithm 4 applied to $(X, Y)_{\text{left}}$ and $(X, Y)_{\text{right}}$

Algorithm 5 shows the NB algorithm which works as follows. Let D_{train} represent a set of training samples of t classification objects. In this case, let the probability $P(y|x)$ for a new data sample $X = \langle x_1, x_2, \dots, x_d \rangle$ to belong to the class $y \in \{1, \dots, c\}$, where x_i represents the value of the attribute. Algorithm 5 shows the basic steps in NB classification.

Algorithm 5: NB algorithm

Let $P(y|x)$ be the posterior probability of the target class

Let $P(y)$ be the probability of the class

Let $P(x)$ be the probability of the predictor

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$$P(y|X) = P(x_1|y) \times P(x_2|y) \times \dots \times P(x_d|y) \times P(y)$$

Algorithm 6: The input-output equations for MLP network

Let M to be the number of inputs

Let N is the number of outputs

S is the number of hidden units

Let y represents the N -tuple outputs of the output layer,

Let x represents the M -tuple inputs to the network hidden layer, $n_j = \sum_{k=1}^M W^1_{jk} x_k$

n_j represents the net sum at the hidden neuron j

The output of this unit is:

$$v_j(n_j) = f \left(\sum_{k=1}^M W^1_{jk} x_k \right)$$

where f is a nonlinear transfer function.

The output of the hidden layer is the input to the next layer and the net input to the output unit i is:

$$n_i = \sum_{j=1}^S W^2 V_j$$

and the output unit i produces the following output value:

$$y_i = f(n_i) = f\left(\sum_{j=1}^S W^2 V_j\right)$$

Algorithm 7 presents the basic steps of the LDA algorithm.

Algorithm 7: Linear discriminant analysis (LDA) [181]

Let $x_{ij} \in R^N$ to be the training sample

Let c to be the number of unknown classes

i, j are the class numbers

Let m_i to be the mean vector of class c_i

Let S_b and S_w to be the between and within class scatter matrices, respectively, were

$$S_b = \frac{1}{n} \sum_{i=1}^c n_i (m_i - m)(m_i - m)^T$$

$$S_w = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{i:j} - m)(x_{i:j} - m)^T$$

LDA aims to find a projection that is optimal in separating data classes in a low-dimensional space

If U is a set of projection vector, then U is selected to maximize the ratio between S_b and S_w

$$U^* = \arg \max_U \frac{\text{tr}(U^T S_b U)}{\text{tr}(U^T S_w U)}$$

Acknowledgments

We would like to thank the Douglas Bomford Trust for the funding support of this study.

References

- [1] D.M. Weary, J.M. Huzzey, M.A.G. Von Keyserlingk, Board-invited Review: Using behavior to predict and identify ill health in animals, *Journal of Animal Science*. 87 (2009) 770–777. <https://doi.org/10.2527/jas.2008-1297>.
- [2] K.M. McLennan, E.A. Skillings, C.J.B.B. Rebelo, M.J. Corke, M.A. Pires Moreira, A.J. Morton, F. Constantino-Casas, Technical note: Validation of an automatic recording system to assess behavioural activity level in sheep (*Ovis aries*), *Small Ruminant Research*. 127 (2015) 92–96. <https://doi.org/10.1016/j.smallrumres.2015.04.002>.
- [3] A.. Frost, C.. Schofield, S.. Beaulah, T.. Mottram, J.. Lines, C.. Wathes, A review of livestock monitoring and the needs for integrated systems, *Comput. Electron. Agric.* 17 (1997) 139–159.
- [4] J.P. Bikker, H. van Laar, P. Rump, J. Doorenbos, K. van Meurs, G.M. Griffioen, J. Dijkstra, Technical note: Evaluation of an ear-attached movement sensor to record cow feeding behavior and activity, *Journal of Dairy Science*. 97 (2014) 2974–2979. <https://doi.org/10.3168/jds.2013-7560>.
- [5] J. Barwick, D. Lamb, R. Dobos, D. Schneider, M. Welch, M. Trotter, Predicting lameness in sheep activity using tri-axial acceleration signals, *Animals*. 8 (2018) 1–16. <https://doi.org/10.3390/ani8010012>.
- [6] J. Kaler, J. Mitsch, J.A. Vázquez-Diosdado, N. Bollard, T. Dottorini, K.A. Ellis, Automated detection of lameness in sheep using machine learning approaches: novel insights into behavioural differences among lame and non-lame sheep, *Royal Society Open Science*. 7 (2020) 190824. <https://doi.org/10.1098/rsos.190824>.
- [7] D.A. Gougoulis, I. Kyriazakis, G.C. Fthenakis, Diagnostic significance of behaviour changes of sheep: A selected review, *Small Ruminant Research*. 92 (2010) 52–56. <https://doi.org/10.1016/j.smallrumres.2010.04.018>.

- [8] Z. Al-Rubaye, A. Al-Sherbaz, W.D. McCormick, S.J. Turner, The use of multivariable wireless sensor data to early detect lameness in sheep, 2016.
- [9] A.C. Winter, Lameness in sheep, *Small Ruminant Research*. 76 (2008) 149–153. <https://doi.org/10.1016/j.smallrumres.2007.12.008>.
- [10] N. Chapinal, A.M. de Passillé, M. Pastell, L. Hänninen, L. Munksgaard, J. Rushen, Measurement of acceleration while walking as an automated method for gait assessment in dairy cattle, *Journal of Dairy Science*. 94 (2011) 2895–2901. <https://doi.org/10.3168/jds.2010-3882>.
- [11] S. Neethirajan, The role of sensors, big data and machine learning in modern animal farming, *Sensing and Bio-Sensing Research*. 29 (2020) 100367. <https://doi.org/10.1016/j.sbsr.2020.100367>.
- [12] P.D. PENNING, A technique to record automatically some aspects of grazing and ruminating behaviour in sheep, *Grass and Forage Science*. 38 (1983) 89–96. <https://doi.org/10.1111/j.1365-2494.1983.tb01626.x>.
- [13] S.M.M. Rutter, R.A.A. Champion, P.D.D. Penning, An automatic system to record foraging behaviour in free-ranging ruminants, *Applied Animal Behaviour Science*. 54 (1997) 185–195. [https://doi.org/10.1016/S0168-1591\(96\)01191-4](https://doi.org/10.1016/S0168-1591(96)01191-4).
- [14] P.P. Nielsen, Automatic registration of grazing behaviour in dairy cows using 3D activity loggers, *Applied Animal Behaviour Science*. 148 (2013) 179–184. <https://doi.org/10.1016/j.applanim.2013.09.001>.
- [15] P. Martin, P. Bateson, *Measuring Behaviour: An Introductory Guide*, in: 1986.
- [16] C. Carslake, J.A. Vázquez-Diosdado, J. Kaler, Machine learning algorithms to classify and quantify multiple behaviours in dairy calves using a sensor–moving beyond classification in precision livestock, *Sensors (Switzerland)*. 21 (2021) 1–14. <https://doi.org/10.3390/s21010088>.
- [17] A. Jukan, X. Masip-Bruin, N. Amla, Smart computing and sensing technologies for animal welfare: A systematic review, *ACM Computing Surveys*. 50 (2017) 1–15. <https://doi.org/10.1145/3041960>.
- [18] E.E.L.C.C. Shepard, R.P.R. Wilson, F. Quintana, A.G. Laich, N. Liebsch, D.A.D.D.A. Albareda, L.G.L. Halsey, A. Gleiss, D.T.D. Morgan, A.E.A. Myers, C. Newman, D.W. Macdonald, A. Gómez Laich, N. Liebsch, D.A.D.D.A. Albareda, L.G.L. Halsey, A. Gleiss, D.T.D. Morgan, A.E.A. Myers, C. Newman, D. McDonald, Identification of animal movement patterns using tri-axial accelerometry, *Endangered Species Research*. 10 (2010) 47–60. <https://doi.org/10.3354/esr00084>.
- [19] S. Neethirajan, Recent advances in wearable sensors for animal health management, *Sensing and Bio-Sensing Research*. 12 (2017) 15–29. <https://doi.org/10.1016/j.sbsr.2016.11.004>.
- [20] B.I. Evstatiev, K.G. Gabrovska-Evstatieva, A review on the methods for big data analysis in agriculture, *IOP Conference Series: Materials Science and Engineering*. 1032 (2021) 012053. <https://doi.org/10.1088/1757-899X/1032/1/012053>.
- [21] A. Kamilaris, A. Kartakoullis, F.X. Prenafeta-Boldú, A review on the practice of big data analysis in agriculture, *Computers and Electronics in Agriculture*. 143 (2017) 23–37. <https://doi.org/10.1016/j.compag.2017.09.037>.
- [22] R. García, J. Aguilar, M. Toro, A. Pinto, P. Rodríguez, A systematic literature review on the use of machine learning in precision livestock farming, *Computers and Electronics in Agriculture*. 179 (2020) 105826. <https://doi.org/10.1016/J.COMPAG.2020.105826>.
- [23] K. Ren, J. Karlsson, M. Liuska, M. Hartikainen, I. Hansen, G.H. Jørgensen, A sensor-fusion-system for tracking sheep location and behaviour, *International Journal of Distributed Sensor Networks*. 16 (2020) 155014772092177. <https://doi.org/10.1177/1550147720921776>.
- [24] Ö. Cangar, T. Leroy, M. Guarino, E. Vranken, R. Fallon, J. Lenehan, J. Mee, D. Berckmans, Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis, *Computers and Electronics in Agriculture*. 64 (2008) 53–60. <https://doi.org/10.1016/j.compag.2008.05.014>.
- [25] N. Krahnstoever, J. Rittscher, P. Tu, K. Chean, T. Tomlinson, Activity Recognition using Visual Tracking and RFID, in: *Seventh IEEE Workshops on Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1, 2005*: pp. 494–500. <https://doi.org/10.1109/ACVMOT.2005.17>.

- [26] X.F. Wang, D.S. Huang, A novel multi-layer level set method for image segmentation, 2008. <https://doi.org/10.3217/jucs-014-14-2427>.
- [27] E. Schlecht, C. Hülsebusch, F. Mahler, K. Becker, The use of differentially corrected global positioning system to monitor activities of cattle at pasture, *Applied Animal Behaviour Science*. 85 (2004) 185–202. <https://doi.org/10.1016/j.applanim.2003.11.003>.
- [28] E.D. Ungar, Z. Henkin, M. Gutman, A. Dolev, A. Genizi, D. Ganskopp, Inference of Animal Activity From GPS Collar Data on Free-Ranging Cattle, *Rangeland Ecology & Management*. 58 (2005) 256–266. [https://doi.org/10.2111/1551-5028\(2005\)58\[256:IOAAGF\]2.0.CO;2](https://doi.org/10.2111/1551-5028(2005)58[256:IOAAGF]2.0.CO;2).
- [29] M. Schwager, D.M. Anderson, Z. Butler, D. Rus, Robust classification of animal tracking data, *Computers and Electronics in Agriculture*. 56 (2007) 46–59. <https://doi.org/10.1016/j.compag.2007.01.002>.
- [30] L.W. Turner, M.C. Udal, B.T. Larson, S.A. Shearer, Monitoring cattle behavior and pasture use with GPS and GIS, *Canadian Journal of Animal Science*. 80 (2000) 405–413. <https://doi.org/10.4141/A99-093>.
- [31] H. Sheng, S. Zhang, L. Zuo, G. Duan, H. Zhang, C. Okinda, M. Shen, K. Chen, M. Lu, T. Norton, Construction of sheep forage intake estimation models based on sound analysis, *Biosystems Engineering*. 192 (2020) 144–158. <https://doi.org/10.1016/j.biosystemseng.2020.01.024>.
- [32] A. Mason, J. Sneddon, Automated monitoring of foraging behaviour in free ranging sheep grazing a biodiverse pasture, in: 2013 Seventh International Conference on Sensing Technology (ICST), 2013: pp. 46–51. <https://doi.org/10.1109/ICST.2013.6727614>.
- [33] D.N. Tran, T.N. Nguyen, P.C.P. Khanh, D.T. Trana, An IoT-based Design Using Accelerometers in Animal Behavior Recognition Systems, *IEEE Sensors Journal*. (2021) 1–1. <https://doi.org/10.1109/JSEN.2021.3051194>.
- [34] E.S. Fogarty, D.L. Swain, G.M. Cronin, L.E. Moraes, M. Trotter, Behaviour classification of extensively grazed sheep using machine learning, *Computers and Electronics in Agriculture*. 169 (2020) 105175. <https://doi.org/10.1016/j.compag.2019.105175>.
- [35] P.M. Graf, R.P. Wilson, L. Qasem, K. Hackländer, F. Rosell, The use of acceleration to code for animal behaviours; a case study in free-ranging Eurasian beavers *Castor fiber*, *PLoS ONE*. 10 (2015). <https://doi.org/10.1371/journal.pone.0136751>.
- [36] A.M. de Passillé, M.B. Jensen, N. Chapinal, J. Rushen, Technical note: Use of accelerometers to describe gait patterns in dairy calves, *Journal of Dairy Science*. 93 (2010) 3287–3293. <https://doi.org/10.3168/jds.2009-2758>.
- [37] J. Barwick, D.W. Lamb, R. Dobos, M. Welch, D. Schneider, M. Trotter, Identifying sheep activity from tri-axial acceleration signals using a moving window classification model, *Remote Sensing*. 12 (2020) 646. <https://doi.org/10.3390/rs12040646>.
- [38] A. Abade, P.A. Ferreira, F. de Barros Vidal, Plant diseases recognition on images using convolutional neural networks: A systematic review, *Computers and Electronics in Agriculture*. 185 (2021). <https://doi.org/10.1016/j.compag.2021.106125>.
- [39] Y. Yuan, L. Chen, H. Wu, L. Li, Advanced agricultural disease image recognition technologies: A review, *Information Processing in Agriculture*. (2021). <https://doi.org/10.1016/j.inpa.2021.01.003>.
- [40] K.G. Liakos, P. Busato, D. Moshou, S. Pearson, D. Bochtis, Machine learning in agriculture: A review, *Sensors (Switzerland)*. 18 (2018) 2674. <https://doi.org/10.3390/s18082674>.
- [41] L. Benos, A.C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, D. Bochtis, Machine learning in agriculture: A comprehensive updated review, *Sensors*. 21 (2021) 3758. <https://doi.org/10.3390/s21113758>.
- [42] D.D. Brown, R. Kays, M. Wikelski, R. Wilson, A. Klimley, Observing the unwatchable through acceleration logging of animal behavior, *Animal Biotelemetry*. 1 (2013) 20. <https://doi.org/10.1186/2050-3385-1-20>.
- [43] K.M. Scheibe, C. Gromann, Application testing of a new three-dimensional acceleration measuring system with wireless data transfer (WAS) for behavior analysis, *Psychonomic Society Inc.*, 2006. <https://doi.org/10.3758/BF03192796>.
- [44] D.M. Karantonis, M.R. Narayanan, M. Mathie, N.H. Lovell, B.G. Celler, Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring, *IEEE Transactions on Information Technology in Biomedicine*. 10 (2006) 156–167. <https://doi.org/10.1109/TITB.2005.856864>.

- [45] J. Morales, D. Akopian, Physical activity recognition by smartphones, a survey, *Biocybernetics and Biomedical Engineering*. 37 (2017) 388–400. <https://doi.org/10.1016/j.bbe.2017.04.004>.
- [46] M.R. Borchers, Y.M. Chang, I.C. Tsai, B.A. Wadsworth, J.M. Bewley, A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors, *Journal of Dairy Science*. 99 (2016). <https://doi.org/10.3168/jds.2015-10843>.
- [47] N. Kleanthous, A. Hussain, W. Khan, J. Sneddon, A. Mason, Feature Extraction and Random Forest to Identify Sheep Behavior from Accelerometer Data, *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 12465 LNAI (2020) 408–419. https://doi.org/10.1007/978-3-030-60796-8_35.
- [48] L. Guo, M. Welch, R. Dobos, P. Kwan, W. Wang, Comparison of grazing behaviour of sheep on pasture with different sward surface heights using an inertial measurement unit sensor, *Computers and Electronics in Agriculture*. 150 (2018) 394–401. <https://doi.org/10.1016/J.COMPAG.2018.05.004>.
- [49] A. Cardoso, J. Pereira, L. Nóbrega, P. Gonçalves, P. Pedreiras, V. Silva, SheepIT: Activity and Location Monitoring, *INForum 2018 - Simpósio de Informática*. (2018) 1–12.
- [50] E. Walton, C. Casey, J. Mitsch, J.A. Vázquez-Diosdado, J. Yan, T. Dottorini, K.A. Ellis, A. Winterlich, J. Kaler, Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour, *Royal Society Open Science*. 5 (2018). <https://doi.org/10.1098/rsos.171442>.
- [51] J. Barwick, D.W. Lamb, R. Dobos, M. Welch, M. Trotter, Categorising sheep activity using a tri-axial accelerometer, *Computers and Electronics in Agriculture*. 145 (2018) 289–297. <https://doi.org/10.1016/J.COMPAG.2018.01.007>.
- [52] N. Mansbridge, J. Mitsch, N. Bollard, K. Ellis, G.G. Miguel-Pacheco, T. Dottorini, J. Kaler, Feature selection and comparison of machine learning algorithms in classification of grazing and rumination behaviour in sheep, *Sensors (Switzerland)*. 18 (2018) 1–16. <https://doi.org/10.3390/s18103532>.
- [53] J.A. Vázquez-Diosdado, V. Paul, K.A. Ellis, D. Coates, R. Loomba, J. Kaler, 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 (2019). <https://doi.org/10.3390/s19143201>.
- [54] F.A.P. Alvarenga, I. Borges, V.H. Oddy, R.C. Dobos, Discrimination of biting and chewing behaviour in sheep using a tri-axial accelerometer, *Computers and Electronics in Agriculture*. 168 (2020) 105051. <https://doi.org/10.1016/j.compag.2019.105051>.
- [55] F.A.P. Alvarenga, I. Borges, L. Palkovič, J. Rodina, V.H. Oddy, R.C. Dobos, Using a three-axis accelerometer to identify and classify sheep behaviour at pasture, *Applied Animal Behaviour Science*. 181 (2016) 91–99. <https://doi.org/10.1016/j.applanim.2016.05.026>.
- [56] V. Giovanetti, M. Decandia, G. Molle, M. Acciaro, M. Mameli, A. Cabiddu, R. Cossu, M.G. Serra, C. Manca, S.P.G.G. Rassu, C. Dimauro, Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer, *Livestock Science*. 196 (2017) 42–48. <https://doi.org/10.1016/j.livsci.2016.12.011>.
- [57] M. Decandia, V. Giovanetti, G. Molle, M. Acciaro, M. Mameli, A. Cabiddu, R. Cossu, M.G. Serra, C. Manca, S.P.G. Rassu, C. Dimauro, The effect of different time epoch settings on the classification of sheep behaviour using tri-axial accelerometry, *Computers and Electronics in Agriculture*. 154 (2018) 112–119. <https://doi.org/10.1016/j.compag.2018.09.002>.
- [58] M. Radeski, V. Ilieski, Gait and posture discrimination in sheep using a tri-axial accelerometer, *Animal*. 11 (2017) 1249–1257. <https://doi.org/10.1017/S175173111600255X>.
- [59] N. Kleanthous, A. Hussain, A. Mason, J. Sneddon, A. Shaw, P. Fergus, C. Chalmers, D. Al-Jumeily, 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)*, 2018: pp. 304–315. https://doi.org/10.1007/978-3-030-04212-7_26.
- [60] J. Marais, R. Wolhuter, T. Niesler, S. Le Roux, R. Wolhuter, T. Niesler, S. Le Roux, S. Petrus, L. Roux, R. Wolhuter, T. Niesler, Automatic classification of sheep behaviour using 3-axis accelerometer data, 2014.
- [61] S.P. le Roux, J. Marias, R. Wolhuter, T. Niesler, Animal-borne behaviour classification for sheep (Dohne Merino) and Rhinoceros (*Ceratotherium simum* and *Diceros bicornis*), *Animal Biotelemetry*. 5 (2017) 25. <https://doi.org/10.1186/s40317-017-0140-0>.

- [62] E.S. Nadimi, R.N. Jørgensen, V. Blanes-Vidal, S. Christensen, Monitoring and classifying animal behavior using ZigBee-based mobile ad hoc wireless sensor networks and artificial neural networks, *Computers and Electronics in Agriculture*. 82 (2012) 44–54. <https://doi.org/10.1016/j.compag.2011.12.008>.
- [63] J.W. Kamminga, H.C. Bisby, D. V. Le, N. Meratnia, P.J.M. Havinga, 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*, ACM, New York, NY, USA, 2017: pp. 597–606. <https://doi.org/10.1145/3123024.3124407>.
- [64] S.P. le Roux, R. Wolhuter, T. Niesler, Energy-Aware Feature and Model Selection for Onboard Behavior Classification in Low-Power Animal Borne Sensor Applications, *IEEE Sensors Journal*. 19 (2019) 2722–2734. <https://doi.org/10.1109/JSEN.2018.2886890>.
- [65] L. Nóbrega, P. Gonçalves, M.M. Antunes, D. Corujo, L. Nóbrega, P. Gonçalves, M.M. Antunes, D. Corujo, Assessing sheep behavior through low-power microcontrollers in smart agriculture scenarios, *Computers and Electronics in Agriculture*. 173 (2020) 105444. <https://doi.org/10.1016/j.compag.2020.105444>.
- [66] N. Kleanthous, A. Hussain, A. Mason, J. Sneddon, 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)*, 2019: pp. 411–422. https://doi.org/10.1007/978-3-030-26766-7_38.
- [67] N.C. Heglund, C.R. Taylor, Speed, stride frequency and energy cost per stride: how do they change with body size and gait?, *Journal of Experimental Biology*. 138 (1988).
- [68] R.M. Alexander, Optimization and gaits in the locomotion of vertebrates, *Physiological Reviews*. 69 (1989) 1199–1227. <https://doi.org/10.1152/physrev.1989.69.4.1199>.
- [69] J. Hancock, Studies in grazing behaviour of dairy cattle: II. Bloat in relation to grazing behaviour, *The Journal of Agricultural Science*. 45 (1954) 80–95. <https://doi.org/10.1017/S0021859600045809>.
- [70] C.M. Dwyer, H.L.I. Bornett, Chronic stress in Sheep: Assessment tools and their use in different management conditions, *Animal Welfare*. 13 (2004) 293–304.
- [71] A. Bonn, T. Allott, K. Hubacek, J. Stewart, Drivers of environmental change in uplands, 2008. <https://doi.org/10.4324/9780203886724>.
- [72] R. Evans, Reducing soil erosion and the loss of soil fertility for environmentally-sustainable agricultural cropping and livestock production systems, *Annals of Applied Biology*. 146 (2005) 137–146. <https://doi.org/10.1111/j.1744-7348.2005.040055.x>.
- [73] K. Ito, M.A.G. von Keyserlingk, S.J. LeBlanc, D.M. Weary, Lying behavior as an indicator of lameness in dairy cows, *Journal of Dairy Science*. 93 (2010) 3553–3560. <https://doi.org/10.3168/jds.2009-2951>.
- [74] G.M. Borghart, L.E. O'Grady, J.R. Somers, Prediction of lameness using automatically recorded activity, behavior and production data in post-parturient Irish dairy cows, *Irish Veterinary Journal*. 74 (2021) 4. <https://doi.org/10.1186/s13620-021-00182-6>.
- [75] D. Jarchi, J. Kaler, S. Sanei, Lameness Detection in Cows Using Hierarchical Deep Learning and Synchrosqueezed Wavelet Transform, *IEEE Sensors Journal*. 21 (2021) 9349–9358. <https://doi.org/10.1109/JSEN.2021.3054718>.
- [76] ELAN - The Language Archive, (n.d.). <https://tla.mpi.nl/tools/tla-tools/elan/> (accessed December 16, 2019).
- [77] A. Dehghani, O. Sarbishei, T. Glatard, E. Shihab, A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors, *Sensors (Switzerland)*. 19 (2019) 10–12. <https://doi.org/10.3390/s19225026>.
- [78] Y. Wang, S. Cang, H. Yu, A survey on wearable sensor modality centred human activity recognition in health care, *Expert Systems with Applications*. 137 (2019) 167–190. <https://doi.org/10.1016/j.eswa.2019.04.057>.
- [79] S.P. le Roux, J. Marias, R. Wolhuter, T. Niesler, Animal-borne behaviour classification for sheep ({Dohne} {Merino}) and {Rhinoceros} ({Ceratotherium} simum and {Diceros} bicornis), *Animal Biotelemetry*. 5 (2017) 25. <https://doi.org/10.1186/s40317-017-0140-0>.
- [80] Z.L. Sun, D.S. Huang, Y.M. Cheun, Extracting nonlinear features for multispectral images by FCMC and KPCA, in: *Digital*

- Signal Processing: A Review Journal, Academic Press, 2005: pp. 331–346. <https://doi.org/10.1016/j.dsp.2004.12.004>.
- [81] Z.L. Sun, D.S. Huang, Y.M. Cheung, J. Liu, G. Bin Huang, Using FCMC, FVS, and PCA techniques for feature extraction of multispectral images, *IEEE Geoscience and Remote Sensing Letters*. 2 (2005). <https://doi.org/10.1109/LGRS.2005.844169>.
 - [82] S. Khalid, T. Khalil, S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, in: 2014 Science and Information Conference, IEEE, 2014: pp. 372–378. <https://doi.org/10.1109/SAI.2014.6918213>.
 - [83] S.J. Preece*, J.Y.Y. Goulermas, L.P.J.P.J. Kenney, D. Howard, S.J. Preece, J.Y.Y. Goulermas, L.P.J.P.J. Kenney, D. Howard, A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities From Accelerometer Data - IEEE Xplore Document, *IEEE Transactions on Biomedical Engineering*. 56 (2009) 871–879. <https://doi.org/10.1109/TBME.2008.2006190>.
 - [84] Y. Liu, L. Nie, L. Liu, D.S. Rosenblum, From action to activity: Sensor-based activity recognition, *Neurocomputing*. 181 (2016) 108–115. <https://doi.org/10.1016/j.neucom.2015.08.096>.
 - [85] P. Sarcevic, S. Pletl, Z. Kincses, Comparison of time- and frequency-domain features for movement classification using data from wrist-worn sensors, *SISY 2017 - IEEE 15th International Symposium on Intelligent Systems and Informatics, Proceedings*. (2017) 261–265. <https://doi.org/10.1109/SISY.2017.8080564>.
 - [86] W. Dargie, C. Poellabauer, *Fundamentals of Wireless Sensor Networks*, 2010. <https://doi.org/10.1002/9780470666388>.
 - [87] W. Dargie, Analysis of time and frequency domain features of accelerometer measurements, in: *Proceedings - International Conference on Computer Communications and Networks, ICCCN, 2009*. <https://doi.org/10.1109/ICCCN.2009.5235366>.
 - [88] D.-S. Huang, *Systematic theory of neural networks for pattern recognition*, Publishing House of Electronic Industry of China, Beijing. 201 (1996).
 - [89] D.S. Huang, Radial basis probabilistic neural networks: Model and application, *International Journal of Pattern Recognition and Artificial Intelligence*. 13 (1999) 1083–1101. <https://doi.org/10.1142/S0218001499000604>.
 - [90] C. V. Bouten, K.R. Westerterp, M. Verduin, J.D. Janssen, Assessment of energy expenditure for physical activity using a triaxial accelerometer, *Medicine and Science in Sports and Exercise*. 26 (1994) 1516–1523.
 - [91] T. Gneiting, H. Ševčíková, D.B. Percival, Estimators of fractal dimension: Assessing the roughness of time series and spatial data, *Statistical Science*. (2012) 247–277.
 - [92] L. Bao, S.S. Intille, Activity recognition from user-annotated acceleration data, *Pervasive Computing*. 3001 (2004) 1–17. https://doi.org/10.1007/978-3-540-24646-6_1.
 - [93] M. Dash, H. Liu, Feature selection for classification, *Intelligent Data Analysis*. 1 (1997) 131–156. <https://doi.org/10.3233/IDA-1997-1302>.
 - [94] G. Chandrashekar, F. Sahin, A survey on feature selection methods, *Computers and Electrical Engineering*. 40 (2014) 16–28. <https://doi.org/10.1016/j.compeleceng.2013.11.024>.
 - [95] C.Y. Lu, D.S. Huang, Optimized projections for sparse representation based classification, *Neurocomputing*. 113 (2013) 213–219. <https://doi.org/10.1016/j.neucom.2013.01.009>.
 - [96] V. Bolón-Canedo, N. Sánchez-Marño, A. Alonso-Betanzos, A review of feature selection methods on synthetic data, *Knowledge and Information Systems*. 34 (2013) 483–519. <https://doi.org/10.1007/s10115-012-0487-8>.
 - [97] A. Jović, K. Brkić, N. Bogunović, A review of feature selection methods with applications, in: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2015 - Proceedings, IEEE, 2015: pp. 1200–1205. <https://doi.org/10.1109/MIPRO.2015.7160458>.
 - [98] N. Hoque, D.K. Bhattacharyya, J.K. Kalita, MIFS-ND: A mutual information-based feature selection method, *Expert Systems with Applications*. 41 (2014) 6371–6385. <https://doi.org/10.1016/j.eswa.2014.04.019>.
 - [99] E.B. Hunt, P.J. Stone, J. Marin, *Experiments in induction* Earl B. Hunt, Janet Marin, Philip J. Stone, 1966.
 - [100] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2016. <https://doi.org/10.1016/c2009-0-19715-5>.

- [101] L. Yu, H. Liu, Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution, in: Proceedings, Twentieth International Conference on Machine Learning, 2003: pp. 856–863.
- [102] R. O Duda, P. E Hart, D. G. Stork, Pattern Classification, 2001.
- [103] D.S. Modha, W.S. Spangler, Feature weighting on k-means clustering, Machine Learning. 52 (2003) 217–237. <https://doi.org/10.1023/A:1024016609528>.
- [104] K. Kira, L.A. Rendell, Feature selection problem: traditional methods and a new algorithm, in: Proceedings Tenth National Conference on Artificial Intelligence, 1992: pp. 129–134.
- [105] I. Kononenko, Estimating attributes: Analysis and extensions of RELIEF, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 1994: pp. 171–182. https://doi.org/10.1007/3-540-57868-4_57.
- [106] M. Robnik-Šikonja, I. Kononenko, Theoretical and Empirical Analysis of ReliefF and RReliefF, Machine Learning. 53 (2003) 23–69. <https://doi.org/10.1023/A:1025667309714>.
- [107] S. Fan, Y. Jia, C. Jia, A feature selection and classification method for activity recognition based on an inertial sensing unit, Information (Switzerland). 10 (2019) 290. <https://doi.org/10.3390/info10100290>.
- [108] C. Freeman, D. Kulić, O. Basir, An evaluation of classifier-specific filter measure performance for feature selection, Pattern Recognition. 48 (2015) 1812–1826. <https://doi.org/10.1016/j.patcog.2014.11.010>.
- [109] B. Li, C. Wang, D.S. Huang, Supervised feature extraction based on orthogonal discriminant projection, Neurocomputing. 73 (2009) 191–196. <https://doi.org/10.1016/j.neucom.2008.09.030>.
- [110] B. Li, D.S. Huang, C. Wang, K.H. Liu, Feature extraction using constrained maximum variance mapping, Pattern Recognition. 41 (2008) 3287–3294. <https://doi.org/10.1016/j.patcog.2008.05.014>.
- [111] J. Kittler, Feature Set Search Algorithms, in: Pattern Recognition and Signal Processing, 1978: pp. 41–60. https://doi.org/10.1007/978-94-009-9941-1_3.
- [112] J.H. Holland, Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, 1975.
- [113] D. Asir, S. Appavu, E. Jebamalar, International Journal of Computer Mathematics, Technometrics. 7 (1965) 87–88. <https://doi.org/10.1080/00401706.1965.10490241>.
- [114] M.B. Kursu, A. Jankowski, W.R. Rudnicki, Boruta – A System for Feature Selection, Fundamenta Informaticae. 101 (2010) 271–285. <https://doi.org/10.3233/FI-2010-288>.
- [115] M. B. Kursu, W. Rudnicki, Feature Selection with Boruta Package, 2010.
- [116] B.H. Menze, B.M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, F.A. Hamprecht, A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data, BMC Bioinformatics. 10 (2009) 213. <https://doi.org/10.1186/1471-2105-10-213>.
- [117] J. Suto, S. Oniga, P.P. Sitar, Comparison of wrapper and filter feature selection algorithms on human activity recognition, in: 2016 6th International Conference on Computers Communications and Control, ICCCC 2016, IEEE, 2016: pp. 124–129. <https://doi.org/10.1109/ICCC.2016.7496749>.
- [118] I. Inza, P. Larrañaga, R. Etxeberria, B. Sierra, Feature Subset Selection by Bayesian network-based optimization, Artificial Intelligence. 123 (2000) 157–184. [https://doi.org/10.1016/S0004-3702\(00\)00052-7](https://doi.org/10.1016/S0004-3702(00)00052-7).
- [119] M. Grimaldi, P. Cunningham, A. Kokaram, An Evaluation Of Alternative Feature Selection Strategies And Ensemble Techniques For Classifying Music, Workshop on Multimedia Discovery and Mining. 2 (2003).
- [120] J.G. Dy, C.E. Brodley, Feature Subset Selection and Order Identification for Unsupervised Learning, ICML 00 Proceedings of the Seventeenth International Conference on Machine Learning. (2000) 247–254.
- [121] D.W. Aha, R.L. Bankert, A Comparative Evaluation of Sequential Feature Selection Algorithms, in: Springer, New York, NY,

- 1996: pp. 199–206. https://doi.org/10.1007/978-1-4612-2404-4_19.
- [122] F. Han, D.S. Huang, A new constrained learning algorithm for function approximation by encoding a priori information into feedforward neural networks, *Neural Computing and Applications*. 17 (2008) 433–439. <https://doi.org/10.1007/s00521-007-0135-5>.
 - [123] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and regression trees*, Wadsworth International Group, 2017. <https://doi.org/10.1201/9781315139470>.
 - [124] J.R. Quinlan, *{C4}.5 - Programs for Machine Learning*, Kaufmann, 1993.
 - [125] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 13-17-Aug (2016) 785–794. <https://doi.org/10.1145/2939672.2939785>.
 - [126] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, *Machine Learning*. 46 (2002) 389–422. <https://doi.org/10.1023/A:1012487302797>.
 - [127] X.-F. Wang, J.-X. Du, H. Xu, L. Heutte, Classification of plant leaf images with complicated background, *Applied Mathematics and Computation*. 205 (2008) 916–926. <https://doi.org/10.1016/J.AMC.2008.05.108>.
 - [128] J.J. Valletta, C. Torney, M. Kings, A. Thornton, J. Madden, Applications of machine learning in animal behaviour studies, *Animal Behaviour*. 124 (2017) 203–220. <https://doi.org/10.1016/j.anbehav.2016.12.005>.
 - [129] P. Ravbar, K. Branson, J.H. Simpson, An automatic behavior recognition system classifies animal behaviors using movements and their temporal context, *Journal of Neuroscience Methods*. 326 (2019) 108352. <https://doi.org/10.1016/J.JNEUMETH.2019.108352>.
 - [130] G. Wang, Machine learning for inferring animal behavior from location and movement data, *Ecological Informatics*. 49 (2019) 69–76. <https://doi.org/10.1016/j.ecoinf.2018.12.002>.
 - [131] T. Cover, P. Hart, Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*. 13 (1967) 21–27. <https://doi.org/10.1109/TIT.1967.1053964>.
 - [132] A. Mucherino, P.J. Papajorgji, P.M. Pardalos, k-Nearest Neighbor Classification, in: Springer, New York, NY, 2009: pp. 83–106. https://doi.org/10.1007/978-0-387-88615-2_4.
 - [133] G. Xiao, K. Li, K. Li, Reporting l most influential objects in uncertain databases based on probabilistic reverse top-k queries, *Information Sciences*. 405 (2017) 207–226. <https://doi.org/10.1016/j.ins.2017.04.028>.
 - [134] J. Camargo, A. Young, Feature Selection and Non-Linear Classifiers: Effects on Simultaneous Motion Recognition in Upper Limb., *IEEE Transactions on Neural Systems and Rehabilitation Engineering : A Publication of the IEEE Engineering in Medicine and Biology Society*. 27 (2019) 743–750. <https://doi.org/10.1109/TNSRE.2019.2903986>.
 - [135] G. Shakhnarovich, T. Darrell, P. Indyk, *Nearest-Neighbor Methods in Learning and Vision*, MIT Press, 2018. <https://doi.org/10.7551/mitpress/4908.001.0001>.
 - [136] G. Xiao, K. Li, X. Zhou, K. Li, Efficient monochromatic and bichromatic probabilistic reverse top-k query processing for uncertain big data, *Journal of Computer and System Sciences*. 89 (2017) 92–113. <https://doi.org/10.1016/j.jcss.2016.05.010>.
 - [137] J. Hu, H. Peng, J. Wang, W. Yu, kNN-P: A kNN classifier optimized by P systems, *Theoretical Computer Science*. 817 (2020) 55–65. <https://doi.org/10.1016/j.tcs.2020.01.001>.
 - [138] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, Y. Laurillau, EMG feature evaluation for improving myoelectric pattern recognition robustness, *Expert Systems with Applications*. 40 (2013) 4832–4840. <https://doi.org/10.1016/j.eswa.2013.02.023>.
 - [139] B.E. Boser, I.M. Guyon, V.N. Vapnik, Training algorithm for optimal margin classifiers, in: *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, ACM Press, New York, New York, USA, 1992: pp. 144–152. <https://doi.org/10.1145/130385.130401>.
 - [140] C. Cortes, V. Vapnik, Support-vector networks, *Machine Learning*. 20 (1995) 273–297. <https://doi.org/10.1007/bf00994018>.

- [141] D.S. Huang, J.X. Du, A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks, *IEEE Transactions on Neural Networks*. 19 (2008) 2099–2115. <https://doi.org/10.1109/TNN.2008.2004370>.
- [142] J.R. Quinlan, Induction of decision trees, *Machine Learning*. 1 (1986) 81–106. <https://doi.org/10.1007/bf00116251>.
- [143] L. Breiman, Random Forests, *Machine Learning*. 45 (2001) 5–32. <https://doi.org/10.1023/A:1010933404324>.
- [144] T.K. Ho, The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 20 (1998) 832–844. <https://doi.org/10.1109/34.709601>.
- [145] L. Breiman, Bagging predictors, *Machine Learning*. 24 (1996) 123–140. <https://doi.org/10.1007/bf00058655>.
- [146] P. Panov, S. Džeroski, Combining bagging and random subspaces to create better ensembles, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007: pp. 118–129. https://doi.org/10.1007/978-3-540-74825-0_11.
- [147] X. Wu, V. Kumar, Q.J. Ross, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S. Yu, Z.H. Zhou, M. Steinbach, D.J. Hand, D. Steinberg, Top 10 algorithms in data mining, *Knowledge and Information Systems*. 14 (2008) 1–37. <https://doi.org/10.1007/s10115-007-0114-2>.
- [148] S. Chen, G.I. Webb, L. Liu, X. Ma, A novel selective naïve Bayes algorithm, *Knowledge-Based Systems*. 192 (2020) 105361. <https://doi.org/10.1016/j.knosys.2019.105361>.
- [149] D.S. Huang, S. De Ma, Linear and nonlinear feedforward neural network classifiers: A comprehensive understanding, *Journal of Intelligent Systems*. 9 (1999). <https://doi.org/10.1515/JISYS.1999.9.1.1>.
- [150] D.S. Huang, H.H.S. Ip, K.C.K. Law, Z. Chi, Zeroing polynomials using modified constrained neural network approach, *IEEE Transactions on Neural Networks*. 16 (2005). <https://doi.org/10.1109/TNN.2005.844912>.
- [151] L. Shang, D.S. Huang, J.X. Du, C.H. Zheng, Palmprint recognition using FastICA algorithm and radial basis probabilistic neural network, *Neurocomputing*. 69 (2006). <https://doi.org/10.1016/j.neucom.2005.11.004>.
- [152] S. Li, Z.H. You, H. Guo, X. Luo, Z.Q. Zhao, Inverse-free extreme learning machine with optimal information updating, *IEEE Transactions on Cybernetics*. 46 (2016). <https://doi.org/10.1109/TCYB.2015.2434841>.
- [153] Z.Q. Zhao, D.S. Huang, W. Jia, Palmprint recognition with 2DPCA+PCA based on modular neural networks, *Neurocomputing*. 71 (2007). <https://doi.org/10.1016/j.neucom.2007.07.010>.
- [154] J.X. Du, D.S. Huang, G.J. Zhang, Z.F. Wang, A novel full structure optimization algorithm for radial basis probabilistic neural networks, *Neurocomputing*. 70 (2006) 592–596. <https://doi.org/10.1016/j.neucom.2006.05.003>.
- [155] D.S. Huang, W.B. Zhao, Determining the centers of radial basis probabilistic neural networks by recursive orthogonal least square algorithms, *Applied Mathematics and Computation*. 162 (2005). <https://doi.org/10.1016/j.amc.2003.12.105>.
- [156] W.B. Zhao, D.S. Huang, J.Y. Du, L.M. Wang, Genetic optimization of radial basis probabilistic neural networks, *International Journal of Pattern Recognition and Artificial Intelligence*. 18 (2004). <https://doi.org/10.1142/S0218001404003824>.
- [157] D.-S. Huang, A Constructive Approach for Finding Arbitrary Roots of Polynomials by Neural Networks, *IEEE Transactions on Neural Networks*. 15 (2004) 477–491. <https://doi.org/10.1109/TNN.2004.824424>.
- [158] B. Li, C.H. Zheng, D.S. Huang, Locally linear discriminant embedding: An efficient method for face recognition, *Pattern Recognition*. 41 (2008). <https://doi.org/10.1016/j.patcog.2008.05.027>.
- [159] Y. Zhao, D.S. Huang, W. Jia, Completed local binary count for rotation invariant texture classification, *IEEE Transactions on Image Processing*. 21 (2012). <https://doi.org/10.1109/TIP.2012.2204271>.
- [160] J.X. Du, D.S. Huang, X.F. Wang, X. Gu, Shape recognition based on neural networks trained by differential evolution algorithm, *Neurocomputing*. 70 (2007) 896–903. <https://doi.org/10.1016/j.neucom.2006.10.026>.
- [161] X.F. Wang, D.S. Huang, H. Xu, An efficient local Chan-Vese model for image segmentation, *Pattern Recognition*. 43 (2010). <https://doi.org/10.1016/j.patcog.2009.08.002>.

- [162] Z.Q. Zhao, D.S. Huang, A mended hybrid learning algorithm for radial basis function neural networks to improve generalization capability, *Applied Mathematical Modelling*. 31 (2007). <https://doi.org/10.1016/j.apm.2006.04.014>.
- [163] F. Han, D.S. Huang, Improved extreme learning machine for function approximation by encoding a priori information, *Neurocomputing*. 69 (2006) 2369–2373. <https://doi.org/10.1016/j.neucom.2006.02.013>.
- [164] B.M. Wilamowski, Neural networks and fuzzy systems, in: *Mechatronic System Control, Logic, and Data Acquisition*, Kluwer Academic, 2018: pp. 12-1-12–26.
- [165] J. Schmidhuber, Deep Learning in neural networks: An overview, *Neural Networks*. 61 (2015) 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>.
- [166] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, *Communications of the ACM*. 60 (2017) 84–90. <https://doi.org/10.1145/3065386>.
- [167] F.R.S. R. A. Fisher, Sc.D., The use of multiple measurements in taxonomic problems, *Annals of Eugenics*. 7 (1936).
- [168] C.N. Li, M.Q. Shang, Y.H. Shao, Y. Xu, L.M. Liu, Z. Wang, Sparse L1-norm two dimensional linear discriminant analysis via the generalized elastic net regularization, *Neurocomputing*. 337 (2019) 80–96. <https://doi.org/10.1016/j.neucom.2019.01.049>.
- [169] Y. Qin, A review of quadratic discriminant analysis for high-dimensional data, *WIREs Computational Statistics*. 10 (2018) e1434. <https://doi.org/10.1002/wics.1434>.
- [170] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, *IEEE Intelligent Systems and Their Applications*. 13 (1998) 18–28. <https://doi.org/10.1109/5254.708428>.
- [171] S.B. Kotsiantis, Decision trees: a recent overview, *Artificial Intelligence Review*. 39 (2013) 261–283. <https://doi.org/10.1007/s10462-011-9272-4>.
- [172] I.A. Basheer, M. Hajmeer, Artificial neural networks: Fundamentals, computing, design, and application, *Journal of Microbiological Methods*. 43 (2000) 3–31. [https://doi.org/10.1016/S0167-7012\(00\)00201-3](https://doi.org/10.1016/S0167-7012(00)00201-3).
- [173] B. Kégl, The return of ADABOOST.MH: Multi-class Hamming trees, in: *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, 2014.
- [174] Y. Freund, R.E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, *Journal of Computer and System Sciences*. 55 (1997) 119–139. <https://doi.org/10.1006/jcss.1997.1504>.
- [175] O. Kramer, K-Nearest Neighbors, in: *Dimensionality Reduction with Unsupervised Nearest Neighbors*, Springer, Berlin, Heidelberg, 2013: pp. 13–23.
- [176] N.S. Altman, An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, *The American Statistician*. 46 (1992) 175. <https://doi.org/10.2307/2685209>.
- [177] F. Foerster, M. Smeja, J. Fahrenberg, Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring, *Computers in Human Behavior*. 15 (1999) 571–583. [https://doi.org/10.1016/S0747-5632\(99\)00037-0](https://doi.org/10.1016/S0747-5632(99)00037-0).
- [178] J.L.L. Hounslow, L.R.R. Brewster, K.O.O. Lear, T.L.L. Guttridge, R. Daly, N.M.M. Whitney, A.C.C. Gleiss, Assessing the effects of sampling frequency on behavioural classification of accelerometer data, *Journal of Experimental Marine Biology and Ecology*. 512 (2019) 22–30. <https://doi.org/10.1016/j.jembe.2018.12.003>.
- [179] A. Urtubia, R. León, M. Vargas, Identification of chemical markers to detect abnormal wine fermentation using Support Vector Machines, *Computers & Chemical Engineering*. (2020) 107158. <https://doi.org/10.1016/J.COMPHEMENG.2020.107158>.
- [180] Y. Fan, J. Bai, X. Lei, Y. Zhang, B. Zhang, K.C. Li, G. Tan, Privacy preserving based logistic regression on big data, *Journal of Network and Computer Applications*. 171 (2020) 102769. <https://doi.org/10.1016/j.jnca.2020.102769>.
- [181] D. Tao, X. Li, X. Wu, S.J. Maybank, General Tensor Discriminant Analysis and Gabor Features for Gait Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 29 (2007) 1700–1715. <https://doi.org/10.1109/TPAMI.2007.1096>.
- [182] F.C. İskenderoğlu, M.K. Baltacıoğlu, M.H. Demir, A. Baldinelli, L. Barelli, G. Bidini, Comparison of support vector regression and random forest algorithms for estimating the SOFC output voltage by considering hydrogen flow rates, *International Journal of*

Hydrogen Energy. (2020). <https://doi.org/10.1016/j.ijhydene.2020.07.265>.