



ELSEVIER

Contents lists available at ScienceDirect

## Smart Agricultural Technology

journal homepage: [www.journals.elsevier.com/smart-agricultural-technology](http://www.journals.elsevier.com/smart-agricultural-technology)

# Review of movement sensor applications in livestock animal activity recognition: Communications, data collection practices, and edge-AI solutions

Bradley Patrick <sup>a,\*</sup>, Eiman Kanjo <sup>a,b</sup>, Omprakash Kaiwartya <sup>a</sup>

<sup>a</sup> Nottingham Trent University, Clifton Lane, Nottingham, NG11 8NS, Nottinghamshire, United Kingdom

<sup>b</sup> Imperial College London, Exhibition Road, London, SW7 2BX, London, United Kingdom

## ARTICLE INFO

### Keywords:

Communication technologies  
Edge-AI  
Tinyml  
Machine learning  
Data collection  
Livestock monitoring  
Animal activity recognition

## ABSTRACT

Animal Activity Recognition (AAR) is a key component of Precision Livestock Farming (PLF), enabling continuous monitoring of animal behaviour, health, and welfare. Advances in machine learning (ML) and sensor technologies have significantly improved AAR accuracy; however, most systems depend on cloud-based architectures, which are impractical in rural settings due to limited connectivity and latency constraints. Edge Artificial Intelligence (Edge-AI) offers a promising alternative by enabling local on-device inference that improves responsiveness, reliability, and autonomy. This systematic review analyses 118 peer-reviewed studies published between 2018 and 2025, examining four critical components of the AAR pipeline: communication technologies, data acquisition methodologies, ML and deep learning (DL) model development, and Edge-AI implementation. We summarise approaches to data collection across livestock species, sensor placements, sampling frequencies, labelling strategies, environments, total animals, and total samples. Furthermore, we categorise and evaluate the ML algorithms used in AAR, discussing feature engineering, windowing strategies, and validation techniques. Our findings reveal that only a limited number of studies have explored Edge-AI in real-world deployments, underscoring challenges related to model compression, resource-constrained inference, and energy efficiency. To address these gaps, we synthesise deployment strategies that include TinyML frameworks and hardware-aware model optimisation. Compared with previous surveys, this review uniquely integrates the entire AAR development cycle, from data collection and model training to real-world deployment, providing a comprehensive reference for developing scalable, on-device livestock monitoring systems.

## 1. Introduction

Precision Livestock Farming (PLF) is defined as continuous real-time monitoring of the health, welfare, production, and environmental impact of individual animals [1]. As part of PLF, there is a growing area of research known as Animal Activity Recognition (AAR). This is a useful task for farmers to gain valuable insights about their animals, including their mental and physical states [2]. Edge-AI is a relatively new research area and a natural pairing for agriculture in general, but especially AAR. Edge-AI takes Machine Learning (ML) models from the cloud to the edge, which brings on-device decision-making to the farm itself. As cloud-based methods often encounter bottlenecks on farms, as they are generally in a rural setting and lack communication infrastructure [3], edge-based solutions might effectively address these challenges [4].

The principal methodologies in AAR research encompass the deployment of movement sensors, including accelerometers, gyroscopes, and

magnetometers [5], commonly deployed around the neck of animals, which will be the focus of this article. Another common methodology for AAR is the application of computer vision techniques using cameras [6]. As mentioned above, AAR is a widely researched area, especially in the area of movement sensors. Movement sensors have been used in AAR on several livestock animals, including **cattle** [7–9], **sheep** [10–12], **horses** [2,13,14], **goats** [15–17], **chickens** [18–20], and **pigs** [21–23]. Movement sensors have also been used for AAR in domestic pets such as **dogs** [24–26] and **cats** [27–29].

The articles mentioned above and throughout this article are mostly focused on training and testing developed ML models, with deployment mainly used in the preliminary stages for collecting data from sensors deployed on the animals. Currently, there are few research articles that use Edge-AI for AAR in movement sensors, which may be due to a number of challenges in the area. Collecting and labelling data is one of the leading challenges for AAR [5,30]. There are other additional challenges

\* Corresponding author.

E-mail addresses: [brad.patrick@ntu.ac.uk](mailto:brad.patrick@ntu.ac.uk) (B. Patrick), [eiman.kanjo@ntu.ac.uk](mailto:eiman.kanjo@ntu.ac.uk) (E. Kanjo), [omprakash.kaiwartya@ntu.ac.uk](mailto:omprakash.kaiwartya@ntu.ac.uk) (O. Kaiwartya).

<https://doi.org/10.1016/j.atech.2026.101986>

Received 3 November 2025; Received in revised form 11 March 2026; Accepted 11 March 2026

Available online 19 March 2026

2772-3755/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Table 1**

Overview of existing literature surveys and reviews in comparison to this review.

	[31]	[5]	[32]	[33]	This Review
Communications	No	No	No	No	Yes
Data Collection	Yes	Some	Some	Some	Yes
ML in AAR	Yes	Yes (DL only)	Yes	Yes (DL only)	Yes
Edge-AI in AAR	No	Some	No	No	Yes

for using Edge-AI, such as model size and getting these large models compressed enough to fit small edge devices or even microcontrollers [30].

In this article, we provide a review of the four key areas that contribute to the deployment and adoption of AAR systems, along with the challenges outlined above, including the following:

- **Communications (Section 3):** We review the communication technologies used to support AAR in remote livestock farming contexts, where conventional infrastructure is often unavailable or unreliable. This includes an analysis of low-power wide-area networks (LPWANs), Bluetooth, Wi-Fi, and cellular technologies, highlighting their suitability, limitations, and deployment patterns in sensor-based systems.
- **Data Collection (Section 4):** We perform a systematic review of data collection practices for AAR using movement sensors in multiple species of livestock. This includes an in-depth evaluation of sensor configurations (e.g. placement, sampling frequency, and sensor modality), behaviour categories, annotation methods, deployment environments (barn and field conditions), total animals, and total samples. Our contribution also includes the identification and cataloguing of publicly available datasets, highlighting gaps in current data resources, and providing guidance for future data acquisition strategies.
- **ML in AAR (Section 5):** We contribute a detailed survey of the machine learning approaches applied to AAR. Our analysis covers both classical and deep learning models, pre-processing steps such as windowing and feature extraction, and evaluation methodologies. This provides a unified overview of the current algorithmic landscape and practical model design decisions.
- **Edge-AI in AAR (Section 6):** To our knowledge, we provide the first focused review of Edge-AI implementations in AAR using movement sensors. This includes an examination of model compression techniques, energy optimisation strategies, and deployment considerations on embedded hardware. We contextualise these developments within the unique constraints of rural agricultural settings and discuss the potential of Edge-AI to enable scalable real-time livestock monitoring.

Table 1 highlights that while previous reviews have addressed data collection and machine learning in AAR to varying degrees, none provide a comprehensive discussion of real-world deployment considerations such as communications or Edge-AI. In particular, no prior review systematically examines Edge-AI applications, including model compression and energy efficiency measures. This underscores the unique contribution of this review, which integrates communication technologies, Edge-AI, data collection, and pre-processing to provide guidance for practical deployment in livestock farming.

## 2. Paper collection methodology

For this systematic review, the PRISMA 2020 [34] methodology was followed to obtain a wide range of research articles spanning from 2018 to 2025. We selected articles after 2018 to keep the scope of this review current.

We started searching for articles using five popular research databases, which were Springer, IEEE, ScienceDirect, MDPI and Wiley.

Some of these databases limited the number of booleans allowed for a search query, so each database was queried using "Animal Activity Recognition" and similar synonyms using the following query, filtered with 2018 to 2025 as the year of publication: ("*animal activity recognition*" OR "*animal behavior recognition*" OR "*animal behaviour recognition*" OR "*animal activity classification*"). The total number of results in each database can be seen in the identification stages of the PRISMA flow diagram in Fig. 1.

Given the search query used on these databases does not limit Computer Vision articles (which are out of scope for this review), Google Scholar was also searched with a more optimised query that focuses on sensors specifically, coupled with ML and Edge-AI approaches, along with the specifics of livestock animals which we require for this review. To this end, Google Scholar was searched using the following query, filtered using 2018–2025: ("*animal activity recognition*" OR "*animal behavior recognition*" OR "*animal behaviour recognition*" OR "*animal activity classification*") AND ("*edge-ai*" OR "*edge ai*" OR "*tinyml*" OR "*tiny machine learning*" OR "*aiot*" OR "*machine learning*" OR "*ai*" OR "*ML*") AND ("*sensor*" OR "*movement sensor*" OR "*accelerometer*" OR "*gyroscope*" OR "*magnetometer*" OR "*IMU*") AND ("*cow*" OR "*cattle*" OR "*sheep*" OR "*ewe*" OR "*goat*" OR "*horse*" OR "*bull*" OR "*chicken*" OR "*hen*" OR "*pig*" OR "*sow*"), returning 616 results.

Before screening, any duplicates that were found in the original databases that also appeared in the Google Scholar search were dropped. We also only wanted to focus on research articles that were published, so results from Arxiv, Biorxiv, ResearchGate or PrePrints were automatically excluded using an automated script. The combination of these removals totalled up to 105 articles being removed, leaving 746 records to be screened (as seen in Fig. 1). During the screening process, titles and abstracts were reviewed to evaluate if they fall into the following criteria:

- Articles must be published and peer reviewed (e.g., not a preprint on ArXiv).
- Articles must be about livestock animals, which include: cattle, sheep, horse, goat, chickens and pigs. Some results returned dogs and cats, for example, which are out of scope for this review.
- Articles must utilise movement sensors (accelerometers, gyroscopes, and magnetometers); articles using other types of sensors, such as computer vision, for example, are out of scope for this review.
- Articles must be behaviour / activity classification, For example, some articles tried to evaluate if an animal had taken ill, rather than classifying behaviours.

This screening process resulted in 588 articles being excluded, with a total of 161 articles assessed for eligibility. All of these 161 articles were successfully retrieved. These articles were further evaluated and assessed, considering the above criteria. In the PRISMA flow diagram (Fig. 1), the reasons for exclusion and the total for each reason is provided. "Not Movement Sensors on Animals" was the largest exclusion, where the articles were either mainly Computer Vision based, or that the sensors were not placed on the animals but humans riding the animal (such as horses). Following this, 26 articles were excluded due to them not involving the five livestock animals selected for this review. The rest of the exclusions consisted of the articles not classifying behaviour / activities, for example, some articles were binary classifiers to determine if an animal was ill or how fast an animal was moving. After the screening assessments of the 161 articles, a total of 85 articles were further excluded for not fitting the criterion of this review, leaving 76 articles remaining.

In addition to database searching, we also performed citation searching in the previously mentioned literature review articles in Fig. 1, and identified 108 possible candidates that could also be considered from the titles of the articles. However, 43 of these were duplicates and were not sought for retrieval. The remaining 65 articles were all successfully retrieved and considered for eligibility using the same exclusion parameters from the original search. In total, 42 of these articles were accepted

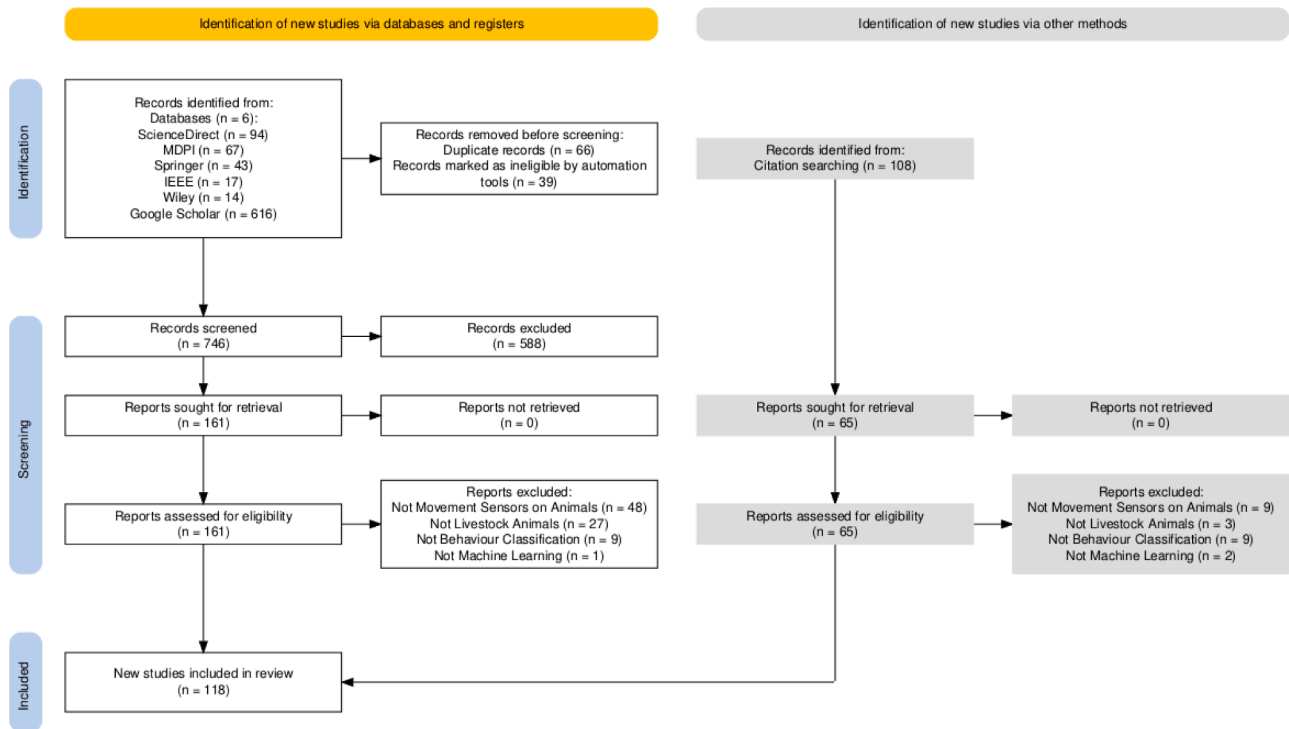


Fig. 1. PRISMA 2020 flow diagram created using [35].

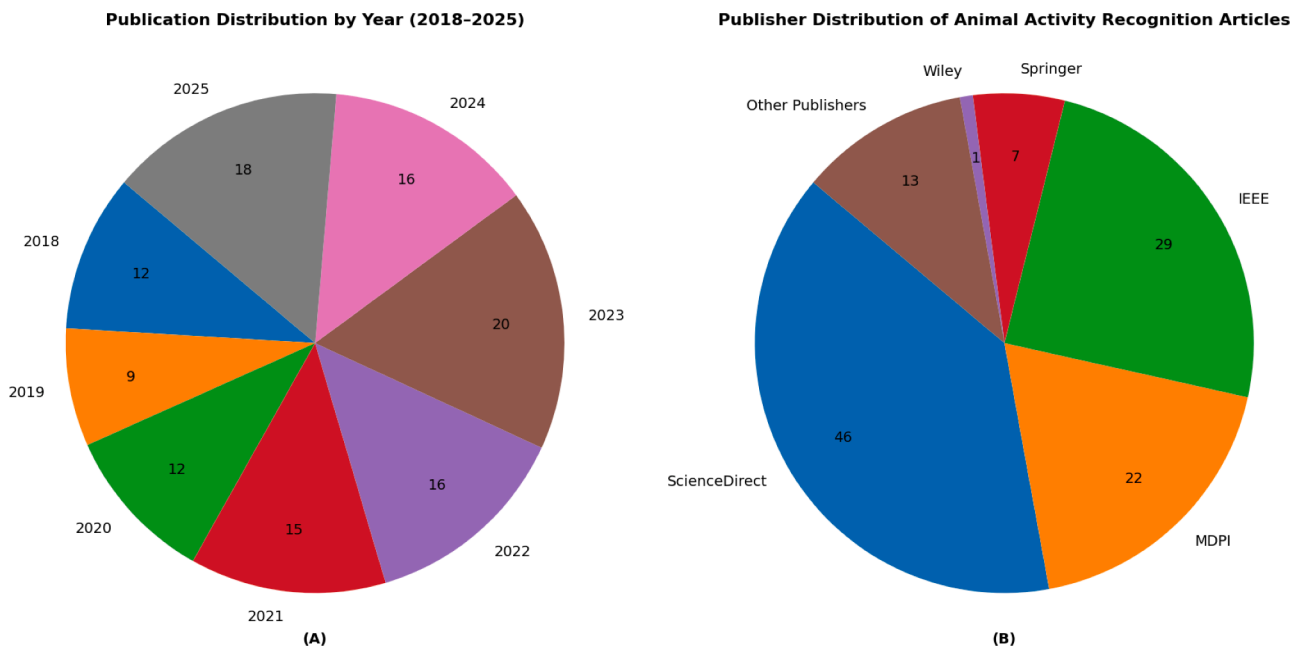


Fig. 2. (A): Distribution of articles by year (2018–2025). (B): Distribution of articles by publisher.

for consideration in this review in addition to the 76 articles from the original search, bring the total number of articles for this review to 118.

In Fig. 2A, a breakdown of the selected 118 articles has been distributed by year, showing advances in AAR using movement sensors and machine learning are still advancing, with 70 of the 118 articles being published since 2022. We have also provided a breakdown of the publishers for the 118 selected articles in Fig. 2B, which shows that ScienceDirect / Elsevier has published the most articles related to movement sensors in AAR, followed by IEEE and MDPI. Within these articles, there is a mix of standard ML papers, with some of these papers providing a contribution towards Edge-AI. In Section 4, we will review the data

collection practices and all 118 articles will be considered. In Section 5, we will review the Machine Learning side of AAR in all 118 articles. Finally, in Section 6, there are 12 articles that will be considered as are the articles that deploy models on microcontrollers, therefore contributing to Edge-AI and TinyML in AAR.

### 3. Communications technologies and networks

Establishing communication within rural areas is often a challenge due to limited connectivity. The coverage of telecommunication companies is limited in rural areas as they prioritise urban areas [3]. There

are several alternative methods for still establishing communications within rural areas with different communication technologies and types of networks that are used for livestock monitoring in general, not necessarily just AAR. This section will review the different communication technologies and the different types of network that have been used.

### 3.1. Communication technologies

There are several communication technologies that have been widely used in the literature, a brief overview of the common ones has been provided below, along with some examples from the literature of their utilisation, following some of the ongoing challenges with the technologies in the area of livestock management. From the reviewed literature identified in Section 2, we will also mention if a certain communication technology is used for data transmission in the data collection aspects of the papers. There are limited instances of communications technologies used in data collection, as many of the articles that collect data usually use local storage such as SD cards.

#### 3.1.1. Wi-Fi and cellular

Wi-Fi and cellular networks are two primary wireless communication technologies that enable Internet access and data transmission. Wi-Fi is commonly deployed in localised environments such as homes, offices, and public hotspots. The latest version of Wi-Fi is Wi-Fi 6, which offers higher throughput, more robustness outdoors, lower latency and power [36]. It offers high-speed data transmission, but is limited in range. In contrast, cellular networks are based on infrastructure managed by telecommunications providers. Cellular technology allows wide-area coverage and higher reliability.

In livestock monitoring, both of these technologies have been used to transmit data from animals to cloud servers. This includes sensor data [37–41] and GPS data [42], Wi-Fi is used where possible, with cellular used as a back-up in the cases where Wi-Fi animals are out of Wi-Fi range. Cellular technology such as SMS is also used to send emergency alerts to farmers when their cattle are entering the calving stages, an example of this is Moomall<sup>1</sup>. Both of these communication technologies have limitations; while they generally offer faster data transfer, Wi-Fi allows the creation of private networks but has limited range. Cellular is dependent on the coverage of telecommunication providers, who usually prioritise urban areas and leave poorer services in rural areas [3] due to fears of high operating costs with little return in rural areas [43]. For AAR data collection, Wi-Fi has been used to transmit data [44–46].

#### 3.1.2. Bluetooth and BLE

Bluetooth and BLE (Bluetooth Low Energy) are wireless communication technologies based on the IEEE 802.15.1 protocol. It was designed to communicate over short distances with an absolute maximum range of 100 metres [47]. BLE is a low-energy version of Bluetooth; one of the most common implementations of BLE is in beacons, which produce a constant Bluetooth signal. The latest version of BLE is BLE 5, which supports longer range, higher data-rate, eight times more advertisement capacity, making it a suitable candidate for heterogeneous IoT applications [48].

In terms of livestock monitoring, it is usually used for the purpose of tracking an animal's location and trying to locate animals within indoor environments using RSSI readings between a base station [49–51]. BLE has also been used to count the number of animals within a certain area [52]. BLE ear tags have been explored to monitor animals near each other [53]. Bluetooth and BLE do have limitations in terms of range and the data that can be transferred. The range of Bluetooth is limited, BLE while it can span a longer range is limited in data transfer, hence why it is usually used in beacons to identify a signal is present. Out of all the communications technologies discussed, Bluetooth was used in the most articles for sensor data transmission [18,54–61].

#### 3.1.3. LoRa

LoRa is a wireless communication technology managed by the LoRa Alliance<sup>2</sup> that allows private networks and integration with networked platforms around the world such as The Things Network<sup>3</sup>. LoRa technology allows for long-range communication between devices up to 30km away from each other under optimal conditions, and it does this using low power [62].

In livestock monitoring, LoRa has been used to transmit environmental sensor data information across long distances [63–65] and has also been used to transmit GPS location data [66]. LoRa has been able to transmit data from sensors and GPS readings to monitor animal health and remote grazing systems [67,68]. LoRa is one of the most popular communication technologies in livestock monitoring due to its long-range capabilities, low power, and ability to create private networks for unlicensed use. However, data speeds are much slower in comparison to other communication technologies, therefore LoRa may not be the most suitable for constant data transmission. However, for data collection in AAR, LoRa has been used to transmit sensor data [69–72].

#### 3.1.4. Sigfox

Sigfox is a wireless communication technology designed for machine-to-machine applications [47]. Sigfox requires coverage for functionality (like that of cellular) and is currently deployed in several countries at the time of writing.

In livestock monitoring, Sigfox has been used in combination with GPS to get the location of livestock in real time [73,74]. Sigfox has also been coupled with BLE to track the location of livestock using a mix of Sigfox-enabled collars and BLE ear tags to track a entire herd [53]. Sigfox has also been used to transmit sensor data [75]. Sigfox is ultra-low powered, with device battery power lasting over a year in some instances. The problem with this technology is that it is only available in limited locations, making it like that of cellular where there is a reliance on third parties to provide coverage.

#### 3.1.5. NB-IoT

Narrowband Internet of Things (NB-IoT) is a low-power wide-area network (LPWAN) technology. It operates within licensed cellular spectrum and is designed specifically for low-data-rate, delay-tolerant IoT applications [76]. NB-IoT offers wide-area coverage, strong penetration in rural and indoor environments, and significantly reduced power consumption compared to traditional cellular technologies. However, similar to cellular networks, NB-IoT depends on infrastructure provided by telecommunications operators, which may be sparse or unavailable in remote grazing regions. Additionally, uplink latency and limited data throughput restrict its suitability for applications requiring frequent or high-volume data transmission.

In livestock monitoring, NB-IoT has been used to transmit sensor data and location information from animals to cloud-based platforms [77–79]. Its ability to support long battery lifetimes while maintaining reliable connectivity makes it attractive for large-scale agricultural deployments. NB-IoT has also been explored for health monitoring applications where periodic data transmission is sufficient, such as behaviour summaries or anomaly detection alerts.

#### 3.1.6. Zigbee

Zigbee is a wireless communication technology with a range of up to 150 metres [80]. Zigbee is also low powered and when compared with the technologies discussed, it uses the least power [81].

In livestock monitoring, Zigbee has mainly been used to transfer data from a range of different sensors to on-sight gateway devices [7,82–85]. Zigbee has also been used in combination of a range of other wireless technologies, particularly for data transfer [86]. Although Zigbee uses

<sup>1</sup> <https://www.moomall.com>

<sup>2</sup> <https://lora-alliance.org>

<sup>3</sup> <https://www.thethingsnetwork.org>

little power, the range it can provide is limited and has practical uses in indoor environments [85]. For data collection in AAR, Zigbee has been used for data transmission in some articles [21,87].

### 3.1.7. Satellite

Satellite communication provides global or near-global connectivity by transmitting data via satellites, enabling communication in remote areas where terrestrial networks are unavailable [88]. Satellite-based IoT systems are particularly valuable for extensive livestock farming operations in isolated regions, such as mountainous terrain or large pastoral lands, where cellular or LPWAN coverage is unreliable or non-existent. Challenges include higher latency, limited downlink capacity, and greater terminal and infrastructure complexity, particularly for direct-to-satellite systems.

In livestock monitoring, satellite communication has been primarily used for GPS-based animal tracking and location reporting [89,90]. While satellite communication offers unmatched coverage, it is typically associated with higher hardware costs, subscription fees, increased energy consumption, and higher latency compared to terrestrial alternatives. As a result, satellite links are often used selectively, for example as a fallback communication method or for low-frequency transmissions, rather than continuous data streaming.

## 3.2. Types of networks

With an outline of the communication technologies provided, this section will explain three types of networks used within livestock monitoring and the communication technologies used within them.

### 3.2.1. Low powered wide area network

A Low Powered Wide Area Network (LPWAN) is a wireless communication technology designed for long-range, low-power, and low-bandwidth applications. It enables battery-powered devices to transmit small amounts of data over long distances while consuming minimal energy. LPWAN technologies, such as LoRa and Sigfox, are widely used in IoT applications for smart agriculture [91]. LoRa is frequently used in LPWANs to improve coverage in extensive agricultural areas [92]. This is attributed to its ability to transmit signals over long distances, reaching up to 30 km between devices under ideal conditions, all while consuming a minimal amount of power [62]. In livestock monitoring, LPWANs allow end nodes, such as animal collars that collect data, to transmit the data to a gateway device, which will then transfer the data to a cloud server [93].

### 3.2.2. Wireless sensor network

A Wireless Sensor Network (WSN) is a network of distributed sensor nodes that communicate wirelessly to collect and transmit data [94]. WSNs are designed to be energy efficient and the nodes operate at low power to extend the battery life. They typically use communication protocols such as Zigbee or Bluetooth Low Energy (BLE) due to their energy efficiency, but can use Wi-Fi for higher data rate applications [95]. In livestock monitoring, WSNs are widely used to monitor the health and behaviour of animals [96]. WSNs are typically a local network of interconnected devices that will communicate to a gateway device, which will then transfer the data to the cloud, as demonstrated in Fig. 3 from [97].

### 3.2.3. Mesh networks

A mesh network is a type of wireless network topology in which each node can communicate with multiple neighbouring nodes and can act as both a sensing device and a relay for forwarding data. This enables multi-hop communication, allowing data to be transmitted across the network without requiring all nodes to be in direct range of a gateway. Mesh networks are commonly implemented using low-power communication technologies such as LoRa [98]. In livestock monitoring, mesh networks are typically used in localised environments such as barns or

confined grazing areas, where animal-mounted or environmental sensors can form a cooperative network to extend coverage and improve reliability, however there outdoor usage using LoRa for data transmission in livestock monitoring has been explored [99]. The ability of mesh networks to reroute data around failed or mobile nodes improves robustness in dynamic environments. However, as network size increases, mesh-based communication can introduce additional latency and energy overhead due to packet forwarding, which can limit scalability for large herds or wide-area deployments.

### 3.2.4. Discussion of network types

The difference between LPWANs, WSNs, and mesh networks lies primarily in the communication technologies used and the resulting network topology. LPWANs utilise long-range communication technologies such as LoRa or Sigfox, enabling end devices to transmit data directly to a gateway over large distances. In contrast, WSNs typically employ shorter-range technologies such as Zigbee or BLE and consist of interconnected sensor nodes that communicate locally before forwarding data to a gateway, often forming mesh-based topologies. Mesh networks therefore represent a common architectural approach within WSNs, where nodes cooperate to relay data through multi-hop communication, but as shown can use the same communication technologies as LPWANs, such as LoRa.

LPWANs are well suited to wide-area livestock monitoring, where animals are distributed over large grazing regions and direct long-range communication to a gateway is required, and mesh networks are starting to be explored in this area too. WSNs are more commonly applied in localised environments, such as barns or confined grazing areas, where dense node deployment allows reliable short-range communication and network redundancy. While mesh networking can extend coverage and improve robustness by re-routing data around failed or mobile nodes, it also introduces additional latency and energy consumption due to packet forwarding, which can limit scalability as network size increases. Moving data from the field to the cloud remains a significant challenge in rural environments due to limited connectivity, and both LPWANs, WSNs and mesh networks help bridge this gap. However, both approaches have limitations. An Internet connection is still required at the gateway to upload collected data to cloud services, which may not be available on all farms. For LPWANs, fair usage policies and duty cycle limitations can restrict data transmission frequency [100]. For example, in the EU 868MHz band, LoRa devices are typically restricted to a 1% duty cycle, corresponding to approximately 36 seconds of transmission time per hour<sup>4</sup>. This severely limits continuous raw sensor streaming and instead favours AAR designs that rely on low-frequency feature transmission, batch-based reporting, or on-device inference where only behaviour labels or summary statistics are transmitted. WSNs and mesh networks also face constraints related to limited processing power and memory on sensor nodes [101], as well as battery life limitations due to frequent communication and data relaying [102]. Mesh networking approaches may also be attractive for extensive grazing systems, as multi-hop communication can theoretically extend coverage beyond the range of a single gateway. However, in large-scale outdoor deployments with highly mobile nodes, maintaining stable mesh topologies is challenging and can significantly increase energy consumption and latency due to repeated packet forwarding. As a result, while mesh networks are well suited to localised or semi-confined environments, LPWAN-based star topologies remain the dominant choice for wide-area livestock monitoring. These constraints highlight that communication technology selection directly influences AAR system design, including sampling frequency, windowing strategy, feature extraction, and the degree to which intelligence must be pushed to the edge.

<sup>4</sup> <https://www.thethingsnetwork.org/docs/lorawan/duty-cycle/>

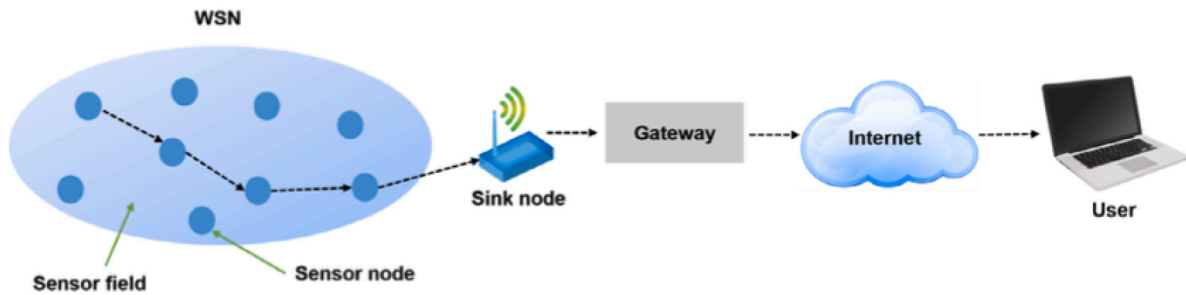


Fig. 3. Typical architecture of a WSN from [97].

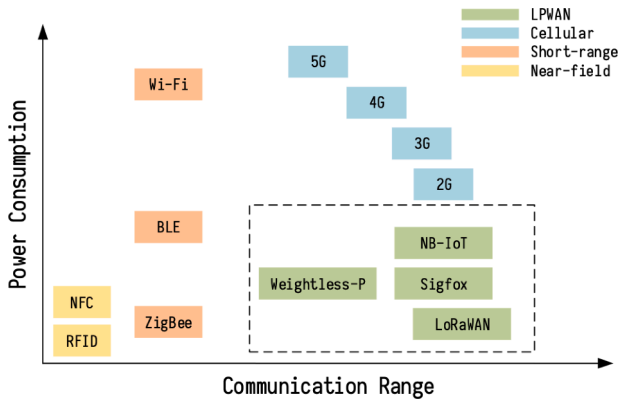


Fig. 4. Power consumption versus communication range for various communication technologies from [103].

### 3.3. Energy consumption

This review paper looks at the progress towards Edge-AI for AAR; and in Edge-AI energy efficiency is one of the most important factors as some aspects of Edge-AI target microcontrollers (more on this will be discussed in Section 6). Therefore, consideration of energy efficient communication technologies should also be taken into account. In Fig. 4, [103] provides a good overview of the distance compared to energy consumption, highlighting that common technologies such as Wi-Fi and cellular consume the most power, while LPWAN technologies such as Sigfox and LoRa consume the least power for the distance travelled, the latter being able to communicate the largest distance and use the least power. While technologies such as Wi-Fi have evolved and Wi-Fi 6 is more energy efficient, it still consumes more energy compared to its lower energy alternatives.

### 3.4. Communications discussion and recommendations

Communication requirements in AAR systems vary substantially depending on the intended application and deployment context. In many studies, communication is not required for real-time monitoring but instead serves to periodically retrieve stored data, verify device operation, or report summary information. Indeed, a significant proportion of the reviewed literature identified in Section 2 collects sensor data locally for later analysis, with 60 articles storing data on-device using SD cards, eliminating the need for continuous wireless communication during the collection deployment. Where communication is required, its purpose will vary. Sometimes it is required for real-time applications that require real-time transmission of behaviour information. It may be required for periodic reporting of summarised data or behaviour labels generated through on-device processing. Or sometimes as simple as device health monitoring to confirm correct

operation and battery status, or firmware updates. Each of these use cases imposes different constraints on latency, bandwidth, and energy consumption.

For real-time monitoring, long range communication technologies such as LoRa or Sigfox (where available) are well suited to outdoor livestock systems, particularly when paired with Edge-AI approaches that reduce transmitted data to low-rate behaviour labels or summaries. In contrast, indoor or confined environments can rely on short range technologies such as Zigbee or BLE, which provide sufficient coverage with lower energy consumption. High power-consuming technologies such as Wi-Fi and cellular networks should be avoided where possible, particularly for battery-powered devices, as lower-power alternatives exist for most livestock monitoring scenarios, albeit with constraints on data rate and packet size. Overall, communication technology selection should be guided by the specific functional requirements of the AAR system rather than assuming continuous data transmission. As Edge-AI and TinyML approaches mature, communication increasingly shifts from raw data streaming toward sparse, event-driven reporting, enabling energy-efficient operation in remote and infrastructure limited agricultural environments.

## 4. Data collection: types, methods, and datasets

This section will discuss the data collection aspects, including the types of movement sensors used, the placement of those movement sensors, the sampling rates of them, the behaviours and activities for each species of livestock, the annotation and labelling methodologies, the environment in which the data was collected, the total number of animals, and the total samples collected. This forms all the data from all 118 of the articles identified in Section 2. To begin, we breakdown the 118 articles and split them by animal species, this is visually represented in Fig. 5A. This results 52 articles for cattle (including various breeds, Calves and Bulls), 21 articles for sheep (including Ewes and Lambs), 15 for horses, 8 for goats, 10 for chickens, and 6 for pigs. The data show that most of the reviewed literature focusses on cattle, followed by sheep, then followed by horses, and then chickens, goats and finally, pigs. There were also six articles that combined a mixture of different livestock animals, three of which considered a mixture of horse and goats [104–106]. The other mixtures included cattle and pigs [107]; cattle, horses and sheep [108]; and finally, cattle, sheep, horse and goat [109].

Fig. 5B represents the number of articles that opted to collect their own data versus those articles that used pre-existing datasets. From this figure, we can see that for all animal breeds, researchers will tend to collect their own data. In the case of cattle, sheep, horses and goats, most articles will collect their own data but some articles will opt to use a dataset; while chicken and pigs, all articles collect their own data and no articles use a pre-existing dataset. The only exception to this is the articles that use multiple animals (labelled as "mixture"), as almost all of these articles use pre-existing datasets, with only one article collecting their own data [107].

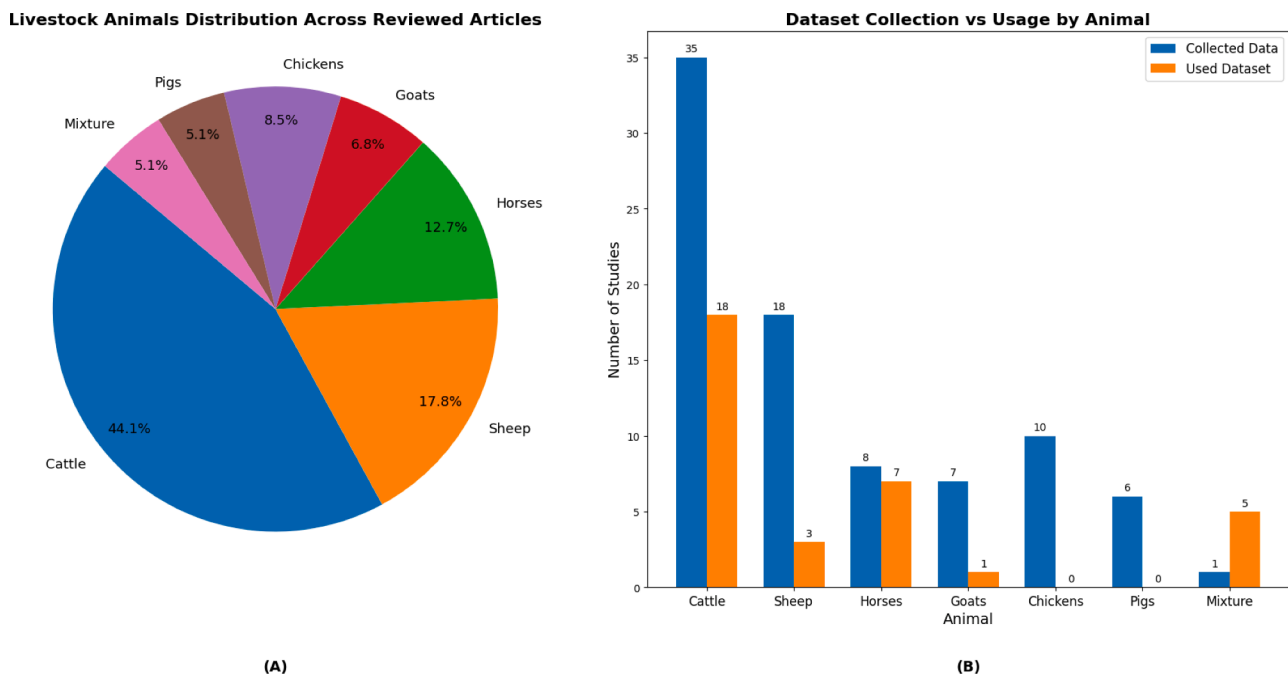


Fig. 5. (A): Livestock animal distribution across 118 reviewed articles. (B): Dataset collection vs dataset usage across 118 reviewed articles.

#### 4.1. Wearable movement sensor types

This review focusses on recognising animal behaviour through movement sensors, which include Accelerometers, Gyroscopes, and Magnetometer sensors. All of these sensors typically consist of three axes:  $X, Y, Z$ , each consisting of a direction of travel (forward and back, up and down, left and right). **Accelerometers:** This sensor is typically used to measure acceleration and is measured in metres per second squared ( $m/s^2$ ). **Gyroscopes:** This sensor is used to measure orientation and angular velocity, measured in degrees per second ( $^\circ/s$ ). **Magnetometers:** This sensor is used to measure changes in the magnetic field at a specific location, which is good for measuring rotation values, such as pitch, roll, and yaw. The magnetometer is measured in Tesla ( $T$ ) [5].

The combination of all three sensors (accelerometer, gyroscope, and magnetometer) forms an Inertial Measurement Unit (IMU) sensor, which has been widely used in the literature. Overall, the accelerometer is by far the most popular sensor used for recognising animal behaviour, with the gyroscope and magnetometer used to help improve or further classify certain behaviours in most cases. There are no review articles using the gyroscope or magnetometer on their own for AAR, meaning that the accelerometer has become the preferred movement sensing device. When presenting the sensor usage data throughout the paper, we have classified the article based on the types of data collected, not the sensor physically attached to the animal. As an example, IMUs consist of accelerometers, gyroscopes, and magnetometers, however, several articles will mention that an IMU has been placed on the animal, but only collect data from the accelerometer sensor [110] or some will only collect data from the accelerometer and gyroscope, but not the magnetometer [17,21,111]. There are a number of reasons for this; manufacturers may combine all three sensors into one unit rather than making separate units. The cost of an IMU has become cost-effective in recent years, or it could simply be that not all sensor data is needed to get good classification results, as several articles report an accuracy of 99% or more with the accelerometer alone [2,11,13,14,112,113]. Earlier studies also explored the use of gyroscope and magnetometer data, but the quality and reliability of these signals were often inconsistent, or did not offer meaningful improvements when coupled with the accelerometer [31].

This variability was an additional factor that contributed to accelerometers becoming the primary sensor used in AAR.

#### 4.2. Animal summary: cattle

Cattle are the most extensively studied species in the AAR literature, representing nearly half of the reviewed articles (44.1%, or 52 articles). This predominance can be explained by both practical and economic considerations. From a research standpoint, the large body size of cattle enables them to carry heavier or bulkier sensor devices comfortably, making them an ideal model for testing new technologies. From a farming perspective, cattle generally have higher individual economic value through multiple income streams, such as meat, milk, and hides, which increases the likelihood of adopting sensor-based systems. These factors collectively position cattle as a natural focal point for both researchers and end-users. Consequently, the strong emphasis on cattle within AAR research is logical and beneficial. For instance, disease detection in cattle can be achieved by monitoring changes in activity patterns, as illnesses often lead to reduced movement and decreased appetite [114,115]. Fig. 7 summarises the literature on cattle, focusing on 35 of the 52 reviewed articles that collected original data. The remaining 18 articles, which relied on existing datasets, will be discussed later in this section. A full breakdown of all the reviewed articles that collected their own data can be found in Appendix A, in Table A.7.

##### 4.2.1. Behaviours / activities

Throughout this review, we use the term 'activity' to describe animal behaviours, in line with the terminology adopted in most AAR studies. While some authors differentiate between 'behaviour' (e.g., grazing, ruminating, resting) and 'activity' (quantitative movement intensity), the literature often uses these terms interchangeably. Here, we follow the convention used in the source studies to accurately reflect their reporting, noting that in some cases 'activity' values are derived from accelerometer signals, whereas 'behaviours' are typically inferred via ML models.

From Fig. 7A, the most common behaviours tracked in the literature for cattle are lying (also classified as resting), eating, standing, and

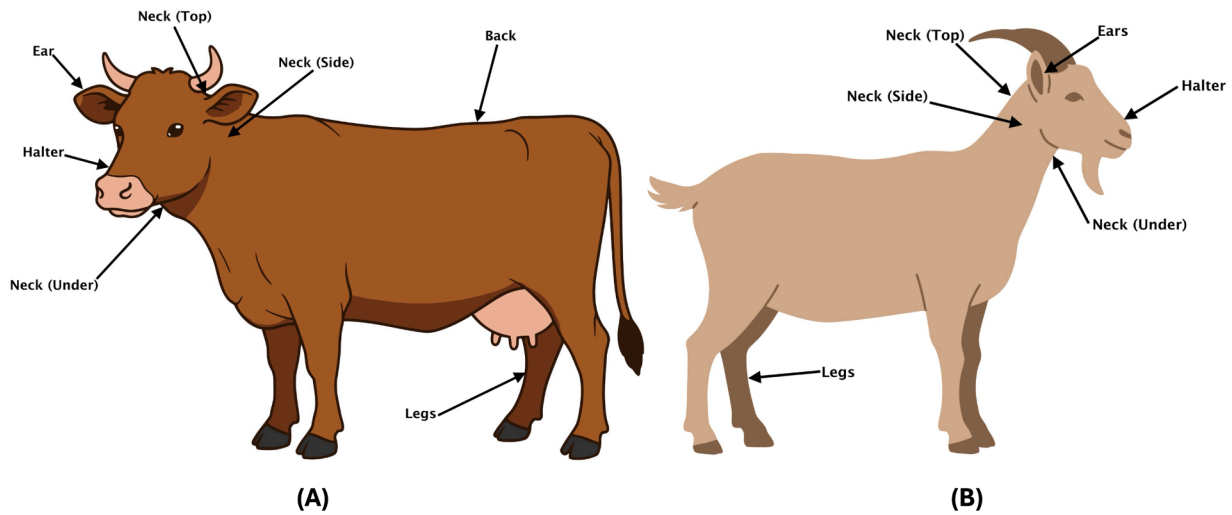


Fig. 6. Sensor placement definitions visual. (A): Cattle sensor placement. (B): Sheep and goat sensor placement.

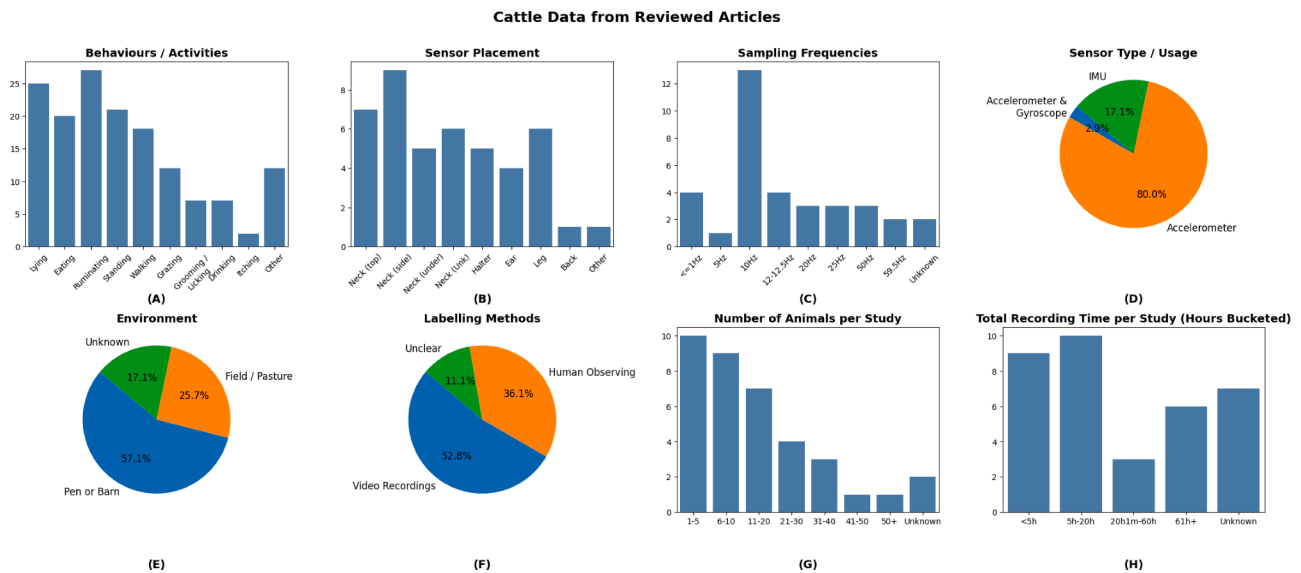


Fig. 7. Data distribution for Cattle in 35 reviewed articles that collected their own data. (A): Types of behaviours and activities classified. (B): Sensor placement. (C): Sampling frequencies used. (D): Sensor type / usage. (E): Data annotation / labelling methods. (F): Data captured environment. (G): Number of animals collected from. (H): Total time of behaviours captured.

ruminating. These behaviours are monitored for two main purposes. First, they can be directly quantified to measure the duration or frequency of each behaviour, providing information about daily behaviour patterns, feed intake, and general health. Second, changes in these behaviours can serve as indicators of events or conditions, such as stress, disease, or nutritional deficiencies [116]. For example, variations in lying or rumination patterns can indicate health issues, while decreases in eating behaviour may signal inadequate nutrition. Less commonly observed behaviours, such as grooming, licking, or itching, are more challenging to monitor but can provide valuable information; for instance, [117] successfully classified itching in cattle to detect skin disease. Behavioural monitoring is also applicable beyond dairy cattle, as demonstrated in [118], where sensors were used to track behaviours in young bulls to ensure adequate nutrition.

#### 4.2.2. Sensor positioning

Within cattle, the most common place for sensors was on the neck, attached using collars. Fig. 7B shows a breakdown of sensor placement in the reviewed articles and collar usage in the neck accounted for the majority of sensors used in the reviewed articles. A visualisation of sen-

sor placement is presented in Fig. 6A. It can be observed that most of the literature opts to place the sensors on the side or the top of the animal's neck, some with some articles opting for the lower neck. However, there were a number of articles that did not specify where on the neck the sensors were placed. The problem with using the neck could be the rotation of the collar, which can occur as the cows move; however, using a counterweight at the bottom of the collar can help keep the sensors in a mostly constant position on the animal [119], which may reduce noise.

#### 4.2.3. Sensor usage and sampling frequencies

From Fig. 7D, Accelerometers are the most widely used sensor, used in 28 of the 35 reviewed articles that used them. It should also be noted that some articles used multiple sensors that are not within the scope of this review to assist them, for example, [120–122] used accelerometers with location-based sensors to help further classify behaviours. Accelerometer coupling with a Gyroscope is not as common within cattle, with only one article opting for it [123]. IMUs were used more frequently than accelerometer and gyroscope pairing, with six articles opting for it.

Fig. 7C presents the sampling frequencies used in the reviewed articles and shows that 10Hz is the most widely used, with 13 articles. The highest sampling rate used in cattle was 59.5Hz [124,125], with the rest of the literature opting for a sampling rate below this, the lowest being 1Hz or lower [118,126–129]. In total, 29 of the 35 articles opted for a sampling frequency of 25Hz or less, successfully classifying behaviours and activities within cattle with these lower sampling frequencies across different sensor positions and different sensors used.

#### 4.2.4. Data labelling and environment

Figs. 7F and 7E break down how the cattle data collected were labelled and the environment in which the data were collected, respectively. The majority of the articles collected their own data and opted to video record the animals' behaviours and label the sensor data using them, which has the benefit of capturing animal behaviour without human interference, but this is a much longer and tedious process. Some articles opted to use live human observations, which is less tedious but more labour intensive as this would need to be done for hours at a time.

Cattle will have limited movement when in a barn or pen compared to a field or pasture, which means each environment has challenges in capturing and labelling data. Most of the articles use a pen or barn environment, as this establishes some control and limits the possible behaviours of the cattle. It also makes it easier to capture data, as the cattle will always be in view for observation. We found that sensor data is captured and labelled in one environment for cattle, only [130] captured data in both barn and field environments. For articles within a pen or barn environment (20 articles), most (13/20) use video recordings to label the captured data. Unlike the field and pasture environment articles (6 articles), half opted to label the data using human observations. Video recordings within a pen or barn environment means that the cattle are limited to where they can move, meaning that the cameras can be set up where the animal will always be in view of them. This is much more challenging in a large field or pasture due to their size, whereas a human observer can move where the animal moves to continue labelling the data.

#### 4.2.5. Total animals and data collected

Fig. 7G summarises the total number of animals used for data collection across the reviewed articles. Most studies relied on relatively small sample sizes, with 19 articles using 10 animals or fewer, suggesting that such numbers are generally considered sufficient for activity recognition research. The largest dataset was reported by [55], who collected data from 51 animals, while the smallest total number of animals was two, and was observed in four different studies [123,129,131,132]. Although small sample sizes are common, larger and more diverse datasets typically provide more robust and generalisable results. Some studies addressed this by collecting data from multiple animals across different farms, as seen in [130], who also collected data in both indoor and outdoor environments.

Reporting of data volume across the literature, however, is highly inconsistent. Some articles specify the total number of data rows, while others report duration in time units. In certain cases, data is presented per animal, whereas others provide aggregated totals. Additionally, some studies report data before cleaning, while others report post-cleaning figures. These inconsistencies make it challenging to accurately summarize the total amount of data collected. To address this, we harmonised reported sample sizes into approximate total hours where feasible, as shown in Fig. 7H. This was achieved by converting time-based entries (hours, seconds, or windows) directly into hours and, where necessary, estimating totals by dividing the number of samples by the reported sampling frequency. While this approach provides a useful overview, it should be interpreted with caution due to the assumptions involved. Nevertheless, the figure suggests that collecting around 20 hours of data or less per study is common practice.

#### 4.2.6. Data collection discussion

The predominance of cattle within the AAR literature reflects a convergence of biological suitability, economic incentives, and practical research considerations. As demonstrated in this review, cattle account for most of the surveyed studies, reinforcing their role as a benchmark species for the development and evaluation of sensor-based AAR systems. Across the reviewed studies, the emphasis on core behaviours such as lying, eating, and ruminating underscores the close relationship between these behaviours and cattle health, welfare, and productivity. These activities are relatively frequent, temporally extended, and behaviourally distinct, making them well suited to classification using wearable sensors. Their prominence also reflects applied motivations, particularly the detection of disease, stress, and nutritional deficiencies. In contrast, rarer or more transient behaviours such as grooming, licking, or itching remain under represented due to the challenges associated with capturing sufficient labelled examples. The limited number of studies targeting such behaviours suggests that current AAR systems may be biased toward high-prevalence behaviours, potentially overlooking early indicators of specific health conditions that manifest through infrequent actions. The success reported in detecting itching-related behaviours indicates that, with sufficient data and appropriate modelling, expanding the behavioural repertoire remains feasible and clinically relevant. Sensor placement choices further reflect a balance between data quality, animal comfort, and practical deployment. The widespread adoption of neck-mounted collars demonstrates their acceptability and ease of use in commercial settings. However, the variability in exact sensor positioning on the neck and the frequent lack of detailed reporting introduce challenges for reproducibility and cross-study comparison. Collar rotation remains a recognised limitation, and while mitigation strategies such as counterweights have been proposed, their adoption is inconsistent. This highlights the need for clearer reporting standards and potentially more robust, orientation-invariant feature extraction or modelling approaches that can tolerate positional variability without compromising performance. The dominance of accelerometers across the literature confirms their effectiveness as a low-power, low-cost sensing modality capable of capturing a wide range of cattle behaviours. The relatively limited use of gyroscopes and full IMUs suggests that, for many cattle-related AAR tasks, accelerometer-only solutions provide a favourable trade-off between complexity and performance. Importantly, the successful classification of behaviours at sampling frequencies of 25Hz or below in the majority of studies demonstrates that high-frequency data acquisition is not strictly necessary for effective cattle AAR. This finding has significant implications for long-term deployments, as lower sampling rates reduce energy consumption, storage requirements, and data transmission constraints in real-world agricultural systems. Data labelling practices and environmental context emerged as critical factors influencing study design and feasibility. The reliance on video-based annotation in controlled environments such as barns and pens reflects a preference for high-quality labels and reduced behavioural ambiguity. However, this approach is time-consuming and may not scale effectively to larger herds or extended monitoring periods. Conversely, studies conducted in pasture or field environments face challenges related to visibility and animal mobility, often necessitating human observers. The fact that most studies collect and label data within a single environment raises questions about the robustness and transferability of trained models across different management systems. The scarcity of multi-environment datasets suggests that many existing models may be environment-specific, limiting their applicability in heterogeneous real-world settings. The analysis of animal numbers and total data collected reveals a further limitation within the cattle AAR literature. While many studies rely on relatively small cohorts, often ten animals or fewer, this approach may be sufficient for proof-of-concept demonstrations but risks overfitting to individual animals or farm-specific conditions. Studies that incorporate larger herds or multiple farms represent an important step toward improving generalisability, yet they remain the exception rather than the norm.

**Table 2**  
Publicly available datasets for AAR within cattle.

Source	Sampling Frequency	Sensor Type(s)	Behaviours Labelled
[135]	25Hz	Accelerometer	Resting, Ruminating, Moving, Grazing, Salt licking, Feeding, Drinking, Licking, Resting, Urinating, Attacking, Escaping, Being mounted, and Other
[119]	10Hz	Accelerometer	Rumination, Eating and Other
[112]	10Hz	Accelerometer	Allogrooming, Calf suckle, Foraging high, Foraging low, Lying, Ruminating lying, Ruminating standing, Standing, Vigilance, Walking
[56]	25Hz	Accelerometer (x2)	Feeding, Ruminating and Other
[120]	50Hz	Accelerometer	Grazing, Walking, Resting, Drinking, and Other
[127]	1Hz	Accelerometer	Feeding, Lying, Standing, Lying down, Standing up, Normal walking, Active walking
[111]	10Hz	Accelerometer & Gyroscope	Ingestion, Biting, and Chewing
[134]	25Hz	Accelerometer	Standing, Lying, Defecation, Urination, Rumination, Drinking, Rising, Lying Down, Grooming, Social, Play, Sniff, Eating, Walking, Run, Pacing, Fall, Cough
[70]	10Hz	Accelerometer	Feeding, Ruminating, and Other
[149]	10Hz	Accelerometer	Drinking, Feeding, Licking, Standing, and Walking

Additionally, inconsistencies in how data volumes are reported complicate cross-study comparisons and meta-analysis. The wide variation in reporting formats, combined with differences in pre and post processing disclosure, underscores the need for standardised reporting guidelines within the AAR community. Despite these limitations, our review suggests that effective cattle AAR systems can be developed using modest data volumes, relatively low sampling frequencies, and simple sensing configurations. This is encouraging for practical deployment, particularly in resource-constrained agricultural contexts. However, the apparent standard practice of collecting 20 hours of data or less also raises questions about model robustness over longer time scales, seasonal changes, and evolving animal behaviour. Future work would benefit from longer-term longitudinal studies that capture behavioural variability over weeks or months, as well as from increased attention to dataset diversity and transparency. In summary, cattle-focused AAR research has reached a level of methodological maturity, with clear convergence around sensor types, placements, and sampling strategies. The challenge moving forward lies not in demonstrating feasibility, but in improving scalability, robustness, and generalisability. Addressing these challenges will require more diverse datasets, consistent reporting practices, and validation across animals, environments, and management systems. While cattle remain a logical and valuable focal species, expanding these insights to other animals will be essential for the broader advancement of AAR as a field.

#### 4.2.7. Datasets

Several articles reviewed captured their own data for their own specific research. Some of the reviewed articles have also published their datasets for use by other researchers [56,70,111,112,119,120]. The calf dataset used in [133] has recently been published in [134].

However, there are 18 articles that used existing datasets, and the majority of these articles used the [135] dataset, which was used by [18,136–143]. This dataset consists of accelerometer data captured at 25Hz from the neck across six different Japanese Black Beef Cows, where they were able to roam freely between a grass field and farm pens, totalling to just under 6 hours of data. The dataset contains data for 13 different behaviours, which include resting, ruminating, moving, grazing, salt licking, feeding, drinking, licking, urinating, attacking, escaping, being mounted, and other.

Other articles like [120,144–146] used a dataset from previous experiments conducted in [147], which included an accelerometer running at 50Hz that captures data at the top of the animal's neck, and included behaviours such as: Grazing, Walking, Resting, Drinking, and Other. Then [148] used a non-public data set, which consisted of an IMU running at 50Hz attached to the top of the animal's neck, totalling just over 19 hours of data from 10 different animals.

A complete overview of all publicly available cattle datasets from what we found across the literature can be found in Table 2.

#### 4.3. Animal summary: sheep and goat

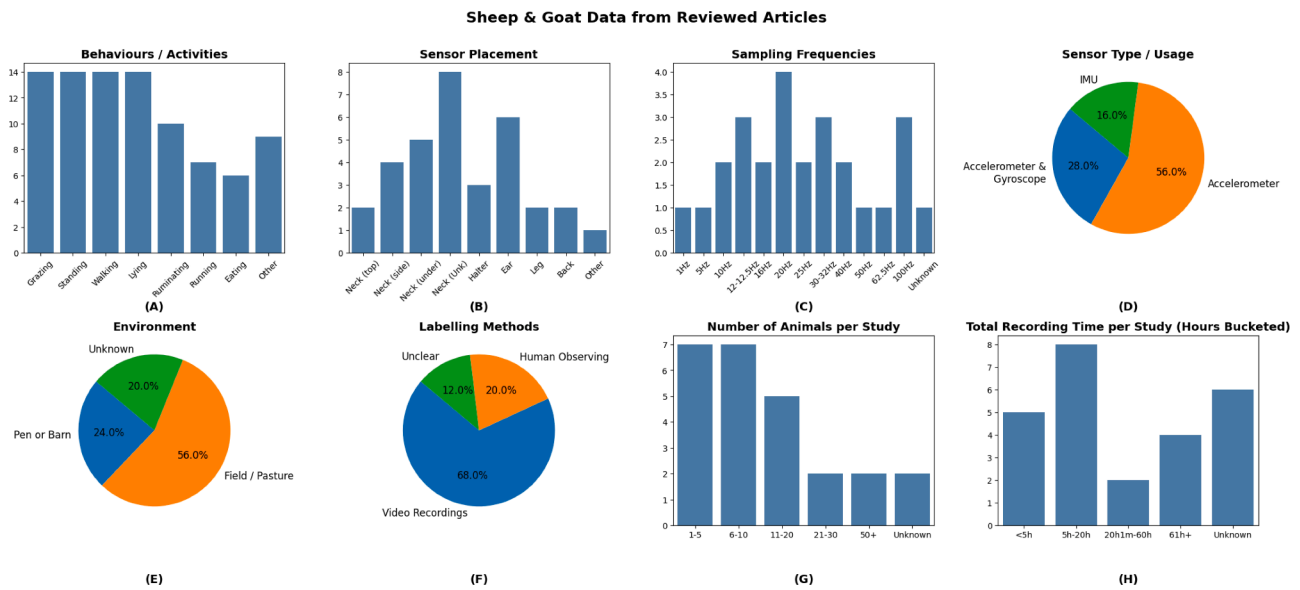
Sheep and goats are grouped together in this section due to their similar roles as ruminants in livestock monitoring. They account for approximately 17.8% and 6.8% of the reviewed literature, respectively, as shown in Fig. 5A, totalling 29 articles. While cattle are also ruminants, their higher economic value likely explains the greater research focus on them compared to sheep and goats. Nevertheless, both species provide multiple income streams for farmers, including meat and dairy, with sheep additionally contributing wool. Demand for goat milk is increasing due to its lower allergenic properties [16]. Beyond production, these animals also play an important role in land management through grazing, with goats being particularly resilient, capable of thriving on shrubs, grasses, and weeds [150]. Fig. 8 summarises the literature on sheep and goats, focusing on 25 of the 29 reviewed articles that collected original data. A full breakdown of all the reviewed articles that collected their own data can be found in Appendix A, in Table A.8. The remaining four articles, which utilised existing datasets, will be discussed later in this section.

##### 4.3.1. Behaviours / activities

The common behaviours monitored in sheep and goats, as reported in the reviewed literature, are presented in Fig. 8A. The main behaviours include grazing, walking, standing, and lying (also classified as resting). In AAR studies, these behaviours are monitored for two complementary purposes. First, they can be directly quantified to measure the duration or frequency of specific behaviours, providing insight into daily behaviour patterns and overall animal welfare. For example, tracking the amount of time sheep spend grazing or lying can inform management decisions related to feeding schedules or pasture usage. Second, deviations in these behaviours can be used for event detection or indirect inference, such as identifying health or welfare issues, or reduced forage availability. For instance, reduced grazing and increased standing or lying across multiple animals may indicate limited forage, while avoidance of certain behaviours in individual sheep may signal pain or discomfort [151,152]. Behaviours such as rumination, grazing, and lying are particularly important for managing forage intake and monitoring animal welfare [10].

##### 4.3.2. Sensor positioning

Like that of cattle, the most common position of the sensors was the neck, as seen in Fig. 8B. A visualisation of sensor placement is presented in Fig. 6B. Notably, there seems to be a more even distribution of the position on the neck, with [12,153] opting for the top of the neck, [12,45,154,155] opting for the lower part of the neck, and [15,156] using the side of the neck. More articles have used ear tags compared to cattle, some using using ear tags alone [157,158], while the other uses of ear tags have been coupled with another movement sensor placed elsewhere on the animal [151,154,159,160].



**Fig. 8.** Data distribution for sheep and goats in reviewed literature. (A): Types of behaviours and activities classified. (B): Sensor placement. (C): Sampling frequencies used. (D): Sensor usage. (E): Data labelling methods. (F): Data captured environment. (G): Number of animals collected from. (H): Total time of behaviours captured.

#### 4.3.3. Sensor usage and sampling frequencies

There is also a more uniform distribution of sensor usage compared to cattle, with 14 articles opting for accelerometer only and seven articles opting for accelerometer and gyroscope pairing, and four articles opting for the IMU (sensor distribution can be seen in Fig. 8D). Several articles used multiple sensors in different positions of the body. For example, [154] used three accelerometers placed in three different positions of the animal, including the ear, neck, and leg.

Sampling frequency usage within sheep and goats is also fairly distributed when looking at Fig. 8C. The lowest sampling frequency was 1Hz in [15] and the highest reported was 100Hz in [71,153,161]. Many articles fall in between 10Hz and 40Hz as a suitable sampling frequency, with 20Hz being the most used. This is similar to cattle, with a notable exception of the highest frequency in cattle being nearly 60Hz, whereas the highest in sheep and goat is 100Hz and was used by three articles.

#### 4.3.4. Data labelling and environment

In Figs. 8E and 8F, a breakdown of the data labelling method and the environment in which the data were captured has been provided. This shows that most articles opted to use video recordings of the animals to label the data. However, a big difference compared to cattle is that most of the articles captured the data in a field or pasture. This appears to be due to the difference in desirable behaviours to capture, as grazing is one of the most frequently occurring behaviours in sheep and goat AAR, so the majority of the research will take place in fields and pastures rather than in barns. Video recordings are still the most frequent method of capturing and labelling sheep and goat behaviour, with some articles still opting for the human observation approach. Some articles did not require data labelling as they only needed to classify two behaviours (binary classification). Engineered features extracted from the sensor data can also be used for the classification of behaviours [15,44]. For example, [15] focused on identifying grazing behaviour in goats and found that grazing produced distinctive Z-axis movements in both the accelerometer and the gyroscope, reflecting the characteristic head-down posture and chewing motion.

#### 4.3.5. Total animals and data collected

Fig. 8G summarises the total number of animals used for data collection across the reviewed articles. Like that of cattle, most articles used 10 animals or less for their data collection. The article that collected data from the most animals was [155], where the authors collected data

from 116 different animals. In contrast, one article collected data from one animal [45], and some others collecting data from three animals [58,153,162].

Like that of cattle, the reporting of data volume is highly inconsistent. We again have attempted to harmonise the reported sample sizes into approximate totals where possible, as shown in Fig. 8H. The aim of this is to provide a useful overview, but like before, it should be interpreted with caution due to the assumptions involved. Nevertheless, the figure suggests that collecting around 20 hours of data or less per study, like that of cattle, is the usual practice.

#### 4.3.6. Data collection discussion

Sheep and goats occupy a distinct yet closely related position within the AAR literature, reflecting their shared ruminant physiology, comparable management practices, and overlapping behavioural repertoires. Although they represent a smaller proportion of studies than cattle, the reviewed literature demonstrates that AAR systems for sheep and goats are increasingly motivated by productivity, welfare, and land-management applications. Their comparatively lower individual economic value, combined with larger herd sizes and more extensive grazing systems, likely contributes to the reduced research focus relative to cattle. Nonetheless, the diversity of applications identified in this review suggests that sheep and goat AAR presents unique challenges and opportunities that are not fully addressed by cattle-centric approaches. The behaviours most frequently classified are grazing, walking, standing, and lying. And these behaviours reflect the central role of foraging and movement in small ruminant management. Unlike cattle, where lying and ruminating dominate many datasets, grazing emerges as the primary behaviour of interest for sheep and goats, particularly in pasture-based systems. This emphasis aligns closely with practical decision-making, as deviations in grazing patterns can signal changes in pasture quality, health issues, or environmental stressors. The focus on these core behaviours suggests that current AAR systems for sheep and goats prioritise high-level behavioural time budgets rather than fine-grained or low-frequency behaviours. While effective for broad monitoring, this may limit the detection of more subtle welfare indicators, such as pain-related behaviour or early disease onset, which are less commonly targeted. Sensor placement strategies largely mirror those used in cattle, with neck-mounted devices remaining the most common configuration. However, a more even distribution of sensor positioning along the neck is evident, potentially reflecting differences in anatomy,

fleece, and movement dynamics between small ruminants and cattle. The greater uptake of ear-mounted sensors is a notable distinction, likely driven by practical considerations such as ease of deployment, reduced interference with grazing, and compatibility with existing identification systems. While ear tags offer clear advantages for scalability, their increased susceptibility to noise from head movements and environmental interactions may partly explain the more frequent use of multi-sensor configurations in this literature. The combination of ear-mounted and body-mounted sensors suggests an ongoing trade-off between data richness and deployment simplicity that remains unresolved. Compared to cattle, sensor usage in sheep and goat AAR is more heterogeneous, with a higher proportion of studies incorporating gyroscopes or full IMUs. This may reflect the need to better capture the complex head and body movements associated with grazing, particularly in uneven terrain or shrub-dominated environments. Despite this increased sensor complexity, the literature demonstrates that effective classification can still be achieved across a wide range of sampling frequencies. While higher frequencies like that of up to 100Hz have been explored, the concentration of studies in the 10-40Hz range indicates that moderate sampling rates are generally sufficient. This finding reinforces the potential for energy-efficient, long-term monitoring systems, even in extensive grazing systems where battery life and data transmission are critical constraints. Data labelling practices further highlight important contrasts with cattle-focused studies. The predominance of field and pasture environments introduces significant challenges for annotation, yet video recording remains the most commonly adopted method. This reflects both the necessity of capturing grazing behaviour in situ and the difficulty of maintaining consistent visibility in open environments. The continued reliance on video and human observation underscores the lack of scalable labelling solutions for small ruminants, particularly in large herds. Some studies mitigate this by narrowing the classification task to binary or highly distinct behaviours, or by leveraging engineered features that exploit characteristic movement patterns. While effective in specific contexts, these approaches may limit extensibility to more complex multi-class scenarios. As with cattle, the majority of sheep and goat studies rely on relatively small numbers of animals, raising concerns about inter-animal variability and model generalisability. Although a small number of studies demonstrate the feasibility of larger-scale data collection, these remain exceptions. The inconsistency in reporting total data volumes further complicates efforts to compare studies or establish best practices. Despite these limitations, the apparent convergence around collecting approximately 20 hours of data or less suggests that many AAR models for sheep and goats are developed under constrained data regimes. This reinforces the need for robust validation strategies and highlights the potential value of techniques such as transfer learning, semi-supervised learning, or cross-farm evaluation. Overall, the sheep and goat AAR literature reflects a field that is adapting cattle-derived methodologies to species with different behavioural priorities and environmental contexts. While many foundational design choices, such as sensor types and placements, remain similar, the stronger emphasis on grazing behaviour, pasture-based environments, and scalable sensor deployment distinguishes small ruminant AAR as a domain with its own methodological requirements. Future research would benefit from greater standardisation in reporting, increased dataset diversity, and explicit evaluation of model robustness across environments and herds. Addressing these challenges will be essential for translating AAR systems for sheep and goats from experimental settings into practical, large-scale agricultural applications.

#### 4.3.7. Datasets

In total, three research articles did publish their dataset for others to use [12,158,163]. Other researchers re-used data from previous experiments, such as [164], used a dataset that they captured in their previous research [154].

Five articles used existing datasets, and four of them used the same dataset. The articles [16,79] used the [165] dataset. This dataset con-

sisted of data from four goats and two sheep, and the data captured was from an accelerometer and gyroscope sensor running at 200Hz. The behaviours labelled in this dataset include: Stationary, Foraging, Walking, Trotting, and Running. Outside of this dataset and [12], we are not aware of any other public datasets. Again, a summary of all the datasets we found across the literature has been provided Table 3 for sheep and goats.

#### 4.4. Animal summary: horses

Horses made up nearly 13% of the reviewed articles, accounting for 15 articles in total. Horses are considered livestock, but in the reviewed literature, the benefits of monitoring horse behaviour lean toward performance within a racing context rather than livestock, so we will not delve into as much detail for this animal. By assessing daily behaviour and activities of horses, the impact of health and well-being can be monitored according to different housing and training conditions [166]. Eight articles opted to collect their own data, while the other seven opted to use existing datasets. The articles that collected their own data have been summarised in Fig. 9, and a more detailed breakdown provided in the Appendix A, inside Table A.9. As before, the other seven articles that utilised datasets will be discussed at the end of this section.

##### 4.4.1. Behaviours / activities

As can be seen in Fig. 9A, the behaviours most commonly tracked in the reviewed literature in horses are trotting, standing, and walking, which were monitored in six of the eight articles. Monitoring these behaviours during training can provide data that can improve horse performance [167]. Other behaviours mentioned within the literature include cantering, rolling, pawing, and flank watching, which are also other behaviours that can help contribute towards understanding and improving a horse's performance. Pain-related behaviours, such as rolling, can also be used for the early detection of time-sensitive health problems [168], but these were grouped in "Other" inside Fig. 9A.

##### 4.4.2. Sensor positioning

Unlike the animals discussed up to now, the most common placement for horses is on the halter and leg, as seen in Fig. 9B. The halter was used in three articles [87,169,170] and the legs used in two articles [13,14]. Two articles used the side of the neck [171] and under the neck [172], while one article used the horse saddle [57]. Compared with other ruminants, there are less articles for this animal, so it is hard to conclude a general trend or best practice.

##### 4.4.3. Sensor usage and sampling frequencies

Four of the eight articles used the accelerometer only [13,14,57,170], with two articles using the accelerometer and gyroscope pairing [171,172], and another two articles using the IMU [87,169]. The breakdown of these sensors types can be observed in Fig. 9D. However, the sampling frequencies used across the literature tend to vary, as seen in Fig. 9C. Three articles opted for a sampling frequency of 30-33Hz, with [87] using 30Hz and [57,169] using 33Hz. Each of the other articles used a different sampling frequency, with the lowest reported being 2Hz [172] and the highest being 200Hz [13]. As with the sensor positioning, due to the limited number of articles, drawing conclusions from these findings is challenging.

##### 4.4.4. Data labelling and environment

Figs. 9F and 9E breakdown the labelling method and environment, respectively. Half of the articles opt to label and annotate their data using video recordings, with two opting to use human observations and two not specifying. Interestingly, different from the other ruminants discussed so far, no articles take place within an indoor environment such as a barn or pen, with six articles specifying an outdoor environment and two not specifying an environment. We believe this is due to the nature of the animal and the types of behaviours being monitored requiring large open spaces and plenty of room for the animal to roam.

**Table 3**  
Publicly available datasets for AAR within sheep and goats.

Source	Animal	Sampling Frequency	Sensor Type(s)	Behaviours Labelled
[165]	Goats (4) & Sheep (2)	200Hz	Accelerometer & Gyroscope	Stationary, Foraging, Walking, Trotting and Running
[12]	Sheep	12.5Hz	Accelerometer (x2)	Grazing, Active, Inactive
[158]	Goat	5Hz	Accelerometer	Feeding, Movement, Social, and Other
[163]	Goat	25Hz	Accelerometer	Grazing, Displacement, Ruminating, Resting, and Other



**Fig. 9.** Data distribution for Horses in 8 reviewed articles that collected their own data. (A): Types of behaviours and activities classified. (B): Sensor placement. (C): Sampling frequencies used. (D): Sensor type / usage. (E): Data annotation / labelling methods. (F): Data captured environment. (G): Number of animals collected from. (H): Total time of behaviours captured.

#### 4.4.5. Total animals and data collected

Fig. 9G summarises the total number of animals used for data collection across the reviewed articles. All articles used 10 animals or less for their data collection. The article that collected data from the most animals was [171], where the authors collected data from 10 different animals. In contrast, two articles collected data from two horses [57,169], with three articles collecting from six different horses [13,14,172].

Like that of the other ruminants, the reporting of data volume is highly inconsistent. Again, attempting to harmonise the reported sample sizes into approximate totals where possible has been presented in Fig. 9H. Based on the assumptions made, most of the articles collected 5 hours of data or less. The highest amount of data collected was from [13], with an approximate 24 hours collected based on our estimations. Again, like that of the previous ruminants, these should be assumed as approximations and not ground truth.

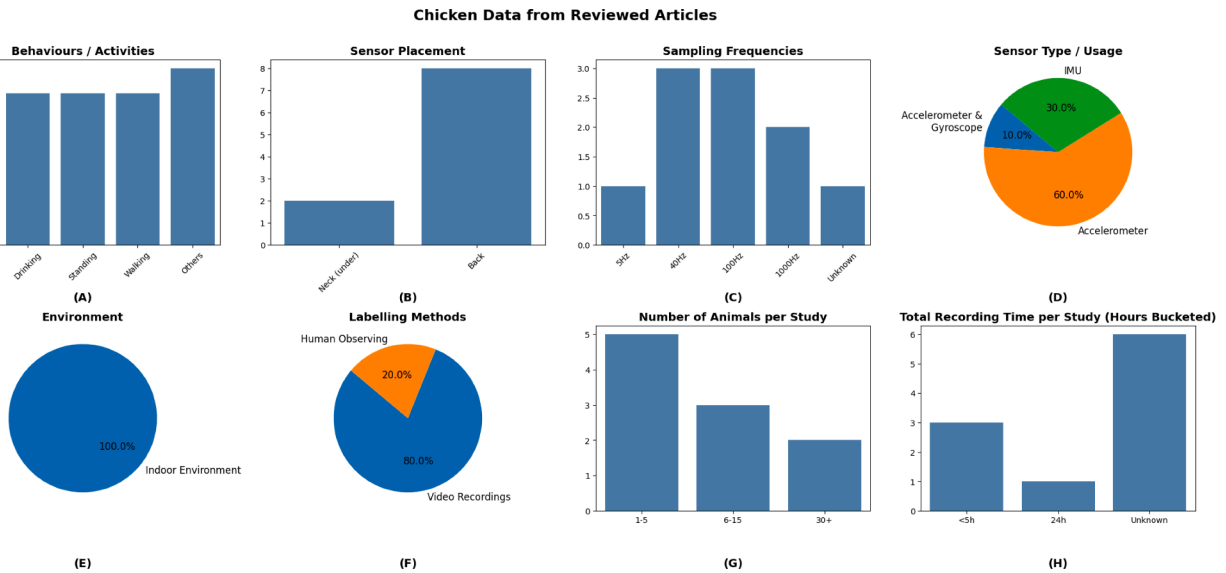
#### 4.4.6. Data collection discussion

Horse AAR research represents a smaller and more specialised body of work compared to ruminant focused studies, with primary motivations centred on performance monitoring, training optimisation, and health assessment rather than agricultural productivity. Although horses are classified as livestock, their management contexts and monitoring goals distinguish them from cattle, sheep, and goats, which is reflected in the behaviours studied, sensor placements, and data collection environments. The behaviours most commonly classified are walking, trotting, and standing. These behaviours are closely linked to gait analysis and training evaluation. These behaviours provide valuable insight into workload, movement patterns, and performance, while less frequent behaviours such as rolling or flank watching may support early detection of pain or discomfort. However, such behaviours are typically grouped into broad categories, suggesting that fine-grained health-related classification remains limited within the current literature. Sensor placement

strategies for horses differ markedly from those used for ruminants, with a greater emphasis on halter and leg-mounted sensors to capture head and limb motion. The diversity of placements observed across studies reflects both the range of monitoring objectives and the lack of established best practices. Similarly, sensor configurations vary between accelerometer only and full IMU setups, and sampling frequencies span a wide range, from very low rates to capture coarse activity patterns to high rates intended for detailed gait analysis. This variability, combined with the small number of studies, makes it difficult to draw firm conclusions regarding optimal system design. All reviewed horse AAR studies were conducted in outdoor environments, consistent with the spatial requirements of equine movement and training. Data annotation relies primarily on video recordings or human observation, which, while effective, introduces scalability challenges. Across all studies, data collection involved small numbers of animals, this being ten or fewer, and for relatively short durations, limiting the generalisability of reported results and increasing susceptibility to inter-animal variability. In summary, horse AAR research demonstrates the feasibility of wearable sensing for monitoring key equine behaviours but remains constrained by limited data scale and methodological heterogeneity. Future work would benefit from larger, more diverse datasets, clearer reporting of sensor configurations, and validation across different training and management contexts to support more robust and transferable equine AAR systems.

#### 4.4.7. Datasets

This section will discuss the other seven articles that did not collect their own data and instead opted to use an existing dataset. Five of the articles are [104,110,173–176], and all use the Horsing Around dataset by [177]. This dataset consists of accelerometer and gyroscope data collected at 100Hz, which consists of data from 18 different horses, collected from sensors attached to the necks of the horses. Behaviours within this dataset include standing, walking, trotting, galloping, and



**Fig. 10.** Data distribution for Chickens in 10 reviewed articles that collected their own data. (A): Types of behaviours and activities classified. (B): Sensor placement. (C): Sampling frequencies used. (D): Sensor type / usage. (E): Data annotation / labelling methods. (F): Data captured environment. (G): Number of animals collected from. (H): Total time of behaviours captured.

grazing. The remaining two articles used datasets from previous experiments. The authors of [2] used the dataset they collected in [13,14]. And the authors of [178] used the dataset collected in [57], which is a published dataset.

#### 4.5. Animal summary: chickens

Chicken AAR research made up 8.5% of the reviewed articles, with ten articles total. Chickens are the smallest animals contained within this review, and the only poultry animal. A summary of these ten articles are presented in Fig. 10, with a more in-depth breakdown of each article presented in Appendix A, in Table A.10. All of the articles collected their own data, so they will all be discussed in this summary. A key difference in chicken AAR is that as far as we are aware, there is currently no public dataset for this animal, so there will be no dataset section for this animal summary. We suspect that the reason for the lack of research in these animals is due to low economic value compared to the other livestock animals mentioned thus far, along with size limitations of the animal and the minimal research for this livestock animal.

##### 4.5.1. Behaviours / activities

For chicken AAR, the most common behaviours monitored are drinking, standing and walking, which was classified in six of the articles, followed by eating / feeding related behaviours, which was covered in five articles, as observed in Fig. 10A. Eggs are a key income stream for this animal, so we were surprised to see only one of the articles attempted to predict if a chicken was laying an egg [20]. In [179], the authors attempted to understand laying behaviour within chickens when they are outside of cages with maintaining productivity. While [20] had the aim of preventing abnormal behaviours within hens that lead to injury and early detection of disease, which was measured by sensing low, median and high intensity tasks. Overall, the key behaviours and behaviours monitored within this animal relate to health.

##### 4.5.2. Sensor positioning

In the case of chickens, they are naturally small animals, so the sensors would need to be small and lightweight. Given the size of this animal, there are limited places where the sensors can go, and the articles opt for two options, as seen in Fig. 10B. Among the ten articles, a small

sensor was placed on the back of the chicken in the majority of the articles (eight of the ten) [19,20,179–184]. The other two articles opted to place the sensor at the front of the chicken's neck [18,185].

##### 4.5.3. Sensor usage and sampling frequencies

The sensor usage for chickens can be observed in Fig. 10D, and this shows that like that of other animals, accelerometer is the most used sensor and used in six articles [19,181–185]. Following this, the accelerometer and gyroscope pair was only used by one article [179], and the IMU was used by three articles [18,20,180]. While the sensor usage is comparable to other animals, the selected sampling frequencies used for chickens is the most varied overall. When observing Fig. 10C, four different sampling frequencies were used, the lowest of which was 5Hz in [18], following this was 40Hz used in three articles [19,181,183]. This in itself is a large jump compared to the other animals discussed so far, as cattle, sheep and goat have sampling frequencies between these such as 10Hz and 20Hz, with 30–33Hz in sheep, goat and horses. After 40Hz, the next sampling frequency used is 100Hz in three articles [180,182,184], which again, is another large jump compared to the other animals. Finally, the largest jump and largest sampling frequency used among all of the animals included in this review is 1000Hz, used in two articles [20,179].

##### 4.5.4. Data labelling and environment

The data labelling methodology for chickens, and the environment the data was captured in is presented in Figs. 10F and 10E, respectively. From these, the approach towards the environment is consistent across the literature, with all of the articles collecting in an indoor environment, ranging from small enclosures to cages. Presumably, this is due to the fact chickens are often located within indoor environment due to external factors such as predators like foxes. The annotation methodology in chickens is mostly done through video recordings, which makes sense when comparing to other animals as they are smaller and are usually within indoor environments, so setting up a camera for the recordings is easier, as they are less likely to walk out of frame and therefore is a more convenient approach. Some articles did use human observations, however.

##### 4.5.5. Total animals and data collected

The total number of animals used for the data collection in the reviewed articles is presented in Fig. 10G. Most of the articles use five

animals or less, with two articles collecting from more than 30 animals. The authors of [184] collected data from 42 chickens, which is the highest of the reviewed articles. The smallest number of animals used for data collection was two [20,180].

We again attempt to summarise the total amount of data collected in Fig. 10H. For most of the articles, it was challenging to give a reasonable summary and consequently, six of the articles were classified as unknown. However, of the remaining four articles we were able to get some approximations (as like the rest of the animals discuss, should be taken as estimations with caution), three articles totalled to less than 5 hours of data, with one article collecting roughly 24 hours of data. Collecting five hours or less data does fall in line with the trends observed in other animals, such as horses.

#### 4.5.6. Data collection discussion

Chicken focused AAR research represents a relatively small portion of the literature and is characterised by constraints that are largely driven by the animal's size, housing conditions, and economic value. As the smallest species included in this review and the only poultry animal, chickens present unique challenges for wearable sensing, which likely contributes to the limited number of studies and the absence of publicly available datasets. Despite these challenges, the reviewed literature demonstrates growing interest in monitoring chicken behaviour for health, welfare, and productivity related applications. The behaviours most commonly classified are drinking, walking, standing, and feeding. These reflect a strong emphasis on health monitoring and early detection of abnormal activity patterns. While egg production represents a major economic driver in poultry farming, it is notable that only a one study explicitly attempt to detect egg-laying behaviour. This suggests that current chicken AAR research prioritises general health and welfare indicators over direct productivity metrics, potentially due to the difficulty of capturing subtle or infrequent behaviours using wearable sensors. Sensor placement in chickens is highly constrained, resulting in relatively consistent design choices across studies. The predominant use of back mounted sensors reflects a compromise between stability, animal comfort, and data quality, while neck mounted sensors are used less frequently. Compared to larger animals, chickens offer fewer viable attachment points, which may limit experimentation with alternative placements but also promotes greater consistency across studies. Although accelerometers remain the most widely used sensor, the sampling frequencies employed in chicken AAR are notably more varied and in some cases substantially higher than those used for other animals. The use of sampling rates up to 1000Hz is unique within this review and suggests attempts to capture fine-grained, high-frequency movements such as pecking or rapid head motion. However, such high sampling rates introduce practical challenges related to power consumption, data storage, and scalability, particularly in commercial poultry settings. The lack of convergence on an optimal sampling frequency indicates that this area remains exploratory. All chicken AAR studies were conducted in indoor environments, reflecting standard poultry housing practices and enabling relatively controlled data collection. The predominance of video-based annotation is consistent with this setting and benefits from the confined spaces and limited movement ranges of the animals. Nevertheless, the small number of animals used in most studies and the generally short data collection durations raise concerns regarding model generalisability, especially in large-scale commercial operations where flock sizes and behavioural variability are substantially higher. Overall, chicken AAR research is still in an early stage compared to that of larger livestock species. While existing studies demonstrate feasibility, particularly for health-related behaviour monitoring, limitations in data scale, sensor practicality, and methodological standardisation remain significant. Addressing these challenges, alongside the development of shared datasets, will be essential for advancing chicken AAR toward real-world deployment.

## 4.6. Animal summary: pigs

And finally, the last livestock animal of this review is pigs. Pig AAR makes up just over 5% of the reviewed articles, totalling up to 6 articles all together. Like that of chickens, we were unable to find any public datasets containing pig movement sensor data, and all of the articles reviewed collected their own data. The reasons for this may be that the animal is under represented in the research rather than less economic value, as pigs have more economic value compared to animals such as chickens, but chickens have more research articles. The summary of the six reviewed articles is presented in Fig. 11, with a more detailed breakdown is provided in Appendix A, in Table A.11.

### 4.6.1. Behaviours / activities

In pig AAR, a range of behaviours / activities were covered across the articles, there was some consistency, but there did tend to be unique behaviours across the articles since each had different use cases. For example, the focus of [21] was on gaining understanding of the behaviour of lactating sows. According to the authors, understanding their behaviour can help recover sows and piglet growth, help provide emergency measures against abnormal behaviour, and better assess their health. They monitored the behaviours within this article that were lying, eating, drinking, sleeping, nursing, and moving. Whereas [22] classified the behaviours lying, standing, walking, and exploring. That being said, Fig. 11A shows that there is some consistency in the behaviours, with eating, lying and standing being common across the articles, and half of the articles collected data on walking, drinking and moving.

### 4.6.2. Sensor positioning

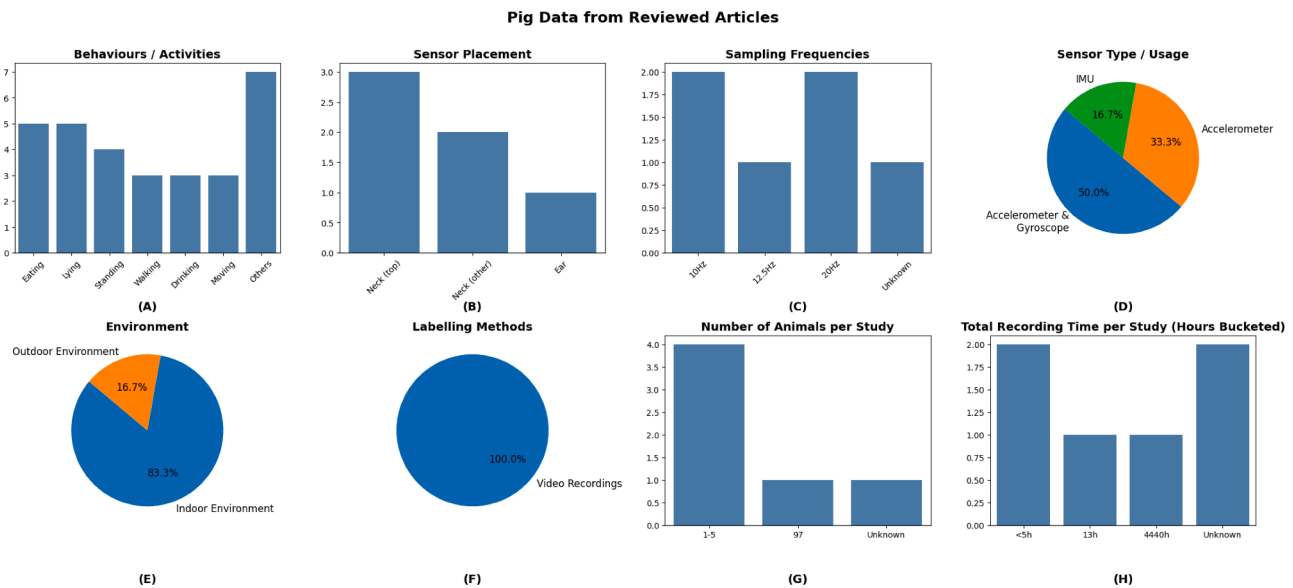
Fig. 11B shows the sensor placement across the six articles. There are a limited number of articles, and therefore a limited number of positions used compared to other livestock animals. Most of the articles placed the sensor at the back of the pigs neck (at the top), with three articles using this position [22,46,54]. Two further articles used the neck of the animal, [186] used under the neck and [21] used a collar around the neck, but the position was not specified. And finally, one article positioned their sensor on the pigs ear [187].

### 4.6.3. Sensor usage and sampling frequencies

Fig. 11D shows that sensor usage of pigs does differ from the rest of the animals discussed in this review, as half of the articles use the accelerometer and gyroscope pair [21,54,186], putting the accelerometer only setup into the minority. Two articles used the accelerometer only [22,187] and one article used an IMU [46]. We suspect this may be the case due to the types of behaviours the literature is attempting to classify and angular velocity measurements may be required to ease this. The sampling frequencies used within pig AAR lean on the lower side as seen in Fig. 11C, which the highest sampling frequency used in the articles being reported as 20Hz [21,22] and the lowest reported as 10Hz [46,54], suggesting that the higher sampling frequencies that have been used in other animals are not required in pigs.

### 4.6.4. Data labelling and environment

From Fig. 11F, the labelling methodology for pigs is consistency across all six articles, with all of the articles opting to use video recordings to label the collected data. The reason for this consistency could be due to the small sample size of articles, but we also believe that the environment in which the data is captured plays a part. Fig. 11E shows that most of the articles, five in this case, collect data within an indoor environment, with only one article collecting data in an outdoor setting. Like that of chickens, pigs are usually housed within an indoor environment, and therefore, setting up a video camera to capture behaviours is easier. Another reason to consider is that pigs are known to be social animals [188], and therefore, the human presence required for manual labelling may influence the true behaviours of the animals.



**Fig. 11.** Data distribution for Horses in 6 reviewed articles that collected their own data. (A): Types of behaviours and activities classified. (B): Sensor placement. (C): Sampling frequencies used. (D): Sensor type / usage. (E): Data annotation / labelling methods. (F): Data captured environment. (G): Number of animals collected from. (H): Total time of behaviours captured.

#### 4.6.5. Total animals and data collected

The majority of the articles have collected data from 5 animals or less, according to our summary in Fig. 11G, with one article collecting from 97 animals [187], significantly more than the rest of the reviewed articles. The articles with the smallest number of animals collected from were [21,186], which collected from two animals. The total amount of data collected from pigs had a little variety, with two articles collecting 5 hours of less data, which falls in line with the trends observed in other animals, like that of chickens. Two articles we were unable to identify in time using our estimations like that of already discussed animals (and again, should be taken with caution). Fig. 11H also shows a variety of total data, with one article collecting roughly 13 hours of data, and another article collecting 4440 hours of data, the latter of which is the article that collected from the most animals [187].

#### 4.6.6. Data collection discussion

Pig AAR research represents the smallest segment of the literature, with all reviewed studies relying on bespoke data collection rather than public datasets. Despite pigs having substantial economic value within livestock production, their limited representation suggests that methodological challenges and research focus, rather than economic importance, have constrained wider adoption of AAR in this domain. Compared to other livestock species, pig AAR remains comparatively under explored. The behaviours monitored across pig studies are diverse and strongly tied to specific use cases, such as sow welfare, nursing behaviour, and general activity monitoring. While behaviours such as eating, lying, and standing appear consistently across studies, the overall behavioural sets are less standardised than those observed for ruminants. This reflects the application-driven nature of pig AAR, where behaviour definitions are often tailored to particular production stages or welfare concerns rather than a common activity taxonomy. Sensor placement in pig AAR shows limited variation, with most studies opting for neck-based configurations and fewer exploring alternative placements such as ear-mounted sensors. This consistency likely reflects practical considerations related to animal anatomy and housing conditions, but the small number of studies makes it difficult to establish best practices. In contrast to other livestock species, pig AAR more frequently employs the accelerometer and gyroscope pairing, suggesting that rotational motion plays an important role in distinguishing pig behaviours. This is supported by the relatively low sampling frequencies used across

studies, indicating that meaningful behavioural patterns can be captured without high frequency data. Data collection and annotation practices are highly uniform, with all reviewed studies relying on video recordings for labelling. The predominance of indoor environments facilitates this approach and reduces challenges related to visibility and tracking. Moreover, video annotation may help minimise behavioural disruption in pigs, which are known to be sensitive to human presence. However, reliance on indoor settings may limit the applicability of trained models to alternative housing systems. As with chickens and horses, most pig AAR studies rely on small numbers of animals and relatively short data collection periods, raising concerns about generalisability. A single large scale study demonstrates the feasibility of collecting extensive datasets across many animals, but such efforts remain rare. Overall, pig AAR research demonstrates promising potential for welfare and behaviour monitoring, yet remains constrained by limited data availability, lack of shared datasets, and methodological heterogeneity. Addressing these limitations will be essential for advancing pig AAR toward scalable and practical deployment.

#### 4.7. Data collection discussion and recommendations

This section synthesises findings from studies that collected bespoke datasets for AAR. We focus on sensor modalities, sampling frequencies, sensor placement, annotation strategies, dataset scale, and data collection environments across livestock species, highlighting common practices, limitations, and recommendations.

##### 4.7.1. Quantity of samples

Across the reviewed literature, the quantity of data collected for model training varies substantially and is reported inconsistently. Studies differ in whether they report samples before or after preprocessing (e.g., windowing), and dataset size is expressed using heterogeneous units, including raw samples, number of windows, or total recording duration. In some cases, values are reported per animal, while others aggregate data across animals, and several studies do not report dataset size at all. These inconsistencies hinder direct comparison between studies and complicate assessments of data sufficiency. To mitigate this, where possible, we estimated the total amount of data collected by converting reported sample counts into an equivalent duration using the stated sampling frequency. This provides a more interpretable and comparable

measure of dataset scale across studies and livestock species. However, this process is inherently approximate, as it relies on assumptions regarding continuous recording, uniform sampling, and complete reporting, and should therefore be interpreted with caution. We provide detailed breakdowns for each article per livestock species in the appendix. These breakdown tables report the original values as stated by the authors, including raw sample counts, recording durations, number of animals, and sampling frequencies. This approach preserves transparency while allowing readers to assess dataset scale in the context of individual study designs. Nonetheless, the lack of standardised reporting practices remains a significant limitation within the AAR literature and highlights the need for clearer and more consistent dataset documentation in future studies.

#### 4.7.2. Total number of animals

The number of animals used during data collection also varies widely across AAR studies. Many studies rely on small cohorts, frequently involving fewer than ten animals, particularly for horses, pigs, and chickens. In contrast, cattle and sheep studies occasionally involve larger herds, reflecting easier access to animals and established monitoring practices within these domains. Small sample sizes at the animal level raise concerns regarding model generalisability, as behaviour patterns can vary substantially between individuals due to age, breed, health status, and environmental conditions. While large datasets in terms of raw samples may appear sufficient, they are often derived from prolonged monitoring of a small number of animals, which can lead to models overfitting to individual specific movement signatures. Only a limited number of studies collect data from dozens or hundreds of animals, yet these demonstrate the feasibility and value of population scale data collection for robust AAR systems.

#### 4.7.3. Total number of samples per animal

Closely related to both dataset size and animal count is the number of samples collected per individual animal. Many studies collect extensive time-series data from each animal, resulting in millions of samples derived from a small cohort. While this enables fine-grained temporal modelling, it can create an imbalance between intra-animal and inter-animal variability, potentially inflating reported model performance. Across species, the number of samples per animal varies by several orders of magnitude, depending on recording duration, sampling frequency, and windowing strategy. However, this metric is rarely reported explicitly, requiring it to be inferred from total dataset size and animal count when possible. Reporting samples per animal would provide clearer insight into data diversity and improve the interpretability and reproducibility of AAR studies. Future work should therefore explicitly report both total dataset size and per-animal data contributions.

#### 4.7.4. Sensor usage

All reviewed studies utilise accelerometers as the primary sensing modality, reflecting their effectiveness for capturing animal movement. Many studies augment accelerometer data with gyroscopes and magnetometers, with a growing trend toward full IMU configurations, particularly when targeting complex or rotational behaviours. Sampling frequencies range from below 1 Hz to 1000 Hz. Several studies collect data at high frequencies before downsampling to assess the impact on classification performance, often motivated by battery life considerations. Sampling frequency selection is therefore closely tied to both the target behaviours and practical deployment constraints. Across species, there is increasing interest in reducing sampling rates without compromising performance, particularly to support long-term monitoring and Edge-AI applications. When multiple sensors are deployed, most studies use consistent sampling frequencies to preserve temporal alignment for ML models. Exceptions arise when combining heterogeneous sensor types, such as movement and location data. Sensor placement varies by species and behaviour, reflecting anatomical constraints and task spe-

cific requirements, such as jaw mounted sensors on the halter for feeding detection in cattle or lightweight back mounted devices for chickens.

#### 4.7.5. Annotation strategies

For supervised and semi-supervised AAR approaches, labelled data are required to train machine learning models, which involves assigning an behaviour label to each instance of sensor data. Across the reviewed literature, two primary annotation strategies are used: video-based annotation and human observations.

Video-based annotation is the most common approach across livestock species. In this setup, animals are recorded using video cameras while sensor data are collected, and behaviours are later annotated by aligning the video footage with sensor timestamps. This approach reduces the need for human presence during data collection, which may otherwise influence animal behaviour, and allows annotation to be performed asynchronously. However, video annotation is time-consuming and can be challenging when animals move out of frame, particularly in outdoor or large-area environments.

Human observation is the second approach, where observers annotate behaviours in real time while monitoring the animals. This method allows observers to follow animals that move out of view and can enable faster data collection, but it requires trained personnel and may introduce behavioural bias due to human presence. Several studies employ custom software tools to support real-time annotation and timestamping.

A critical consideration in both annotation strategies is time synchronisation between sensor data and annotations. Sensor devices, video cameras, and annotation tools may operate using different time formats (e.g., local time, system time, or UTC), and even small misalignments can lead to incorrect labelling and degraded model performance. Accurate synchronisation through shared clocks, timestamp offsets, or synchronisation events should therefore be considered a fundamental requirement to ensure consistency and solid ground truth.

Despite the central role of annotation in supervised AAR, few studies quantitatively report annotation effort, cost, or inter-rater reliability. The time required for manual labelling is rarely explicitly reported in the literature, and formal measures of agreement between annotators are uncommon, even when multiple observers are involved. However, the total duration of annotated data reported across studies (summarised as H in Figs. 7–11) provides an indirect indication of the minimum effort required for annotation. As annotation is typically performed in real time or slower to ensure accuracy, these values likely underestimate the true time investment. The figures also highlight variation in annotation scale between studies, ranging from only a few hours of labelled data to datasets spanning multiple days or weeks. Such differences likely influence annotation quality and consistency, as shorter annotation periods may allow more careful review, whereas large-scale datasets may increase the risk of observer fatigue, subjective interpretation of behaviours, or inconsistencies in annotation protocols.

These differences in annotation practices also limit cross-study comparability, as performance metrics may reflect differences in annotation quality as much as differences in model capability. Inconsistent labelling granularity or observer interpretation can lead to inflated or degraded accuracy scores, particularly for behaviour classes with ambiguous boundaries. Improved reporting of annotation protocols, annotator expertise, and inter-rater agreement would therefore strengthen the reproducibility and interpretability of AAR research.

Finally, a small but growing body of work in Human Activity Recognition explores automated or semi-automated labelling strategies, such as semi-supervised active learning, self-training with pseudo-labels, and clustering combined with minimal human feedback. These approaches have been shown to propagate labels from a modestly sized labelled set to unlabelled samples, substantially reducing the manual annotation burden while maintaining classification performance in activity recognition tasks [189]. Techniques such as self-training further enable models to iteratively incorporate confident predictions on unlabelled data,

improving performance with far fewer manual labels. In animal behaviour contexts, clustering and active learning frameworks have been applied to organise behavioural segments with sparse annotations, reducing human labelling effort [190]. While still under explored in livestock AAR, these automated labelling directions represent promising solutions to the substantial time and labour costs associated with manual annotation.

#### 4.7.6. Data collection bottlenecks

The limited availability of open, standardised movement sensor datasets remains a significant bottleneck for AAR research. Even for well-studied animals such as cattle, no publicly available dataset includes all common IMU modalities. For pigs and chickens, open datasets are currently absent. Data collection is further complicated by ethical approval requirements, device robustness constraints, and the need for expert involvement during annotation. The lack of publicly available datasets for pigs and chickens likely reflects a combination of practical, ethical, and commercial constraints rather than a lack of research interest. For poultry, attaching individual IMU devices is particularly challenging due to animal size, high stocking densities, and welfare considerations, while individual-level annotation is difficult in group-housed systems. Similarly, pig monitoring faces challenges related to device durability, aggressive interactions between animals, and the high cost of annotating large cohorts over extended periods. These factors substantially increase the complexity and expense of producing high-quality, shareable datasets. Data sharing is also hindered by concerns around data ownership, privacy, and security within the agricultural sector [191]. These issues collectively restrict dataset availability and limit cross-study comparison. While this review highlights successful data collection practices, it also underscores the need for improved reporting standards, larger and more diverse animal cohorts, and greater collaboration to enable scalable and reproducible AAR research. Addressing these challenges will likely require coordinated community efforts, including the development of species-appropriate sensing hardware, adoption of semi-supervised or weakly supervised learning approaches to reduce annotation effort, and the establishment of shared benchmark tasks with clearly defined protocols.

## 5. Methods of ML in AAR and preprocessing

This section provides a comprehensive review of machine learning methodologies applied to Animal Activity Recognition (AAR), encompassing both standard machine learning and deep learning approaches. In addition to model architectures, the section examines key components of the AAR pipeline, including data pre-processing (such as cleaning, feature extraction, and windowing strategies) and model validation practices. Together, these elements play a critical role in shaping reported performance and determining the robustness and generalisability of AAR systems across different animals, sensor configurations, and deployment conditions.

Fig. 12 illustrates the temporal distribution of standard machine learning and deep learning studies included in this review. Earlier work is dominated by standard ML approaches, reflecting both the historical maturity of feature-based models and the computational constraints of early wearable sensing platforms. However, a clear shift toward deep learning is evident over time. From 2021 onwards, deep learning approaches are reported more frequently than standard ML methods, with the exception of 2022, where standard ML briefly regains prominence. This transition coincides with broader advances in deep learning for time-series analysis, improved access to computational resources, and an increasing preference for models that can operate directly on raw or minimally processed sensor data. Despite this growing dominance, standard ML remains widely used, accounting for a substantial proportion of the overall literature, as shown by the near-even split between standard ML and deep learning studies in Fig. 12B. This persistence highlights the continued relevance of feature-based models in AAR, partic-

ularly in scenarios with limited data, constrained hardware, or a need for interpretability and reproducibility. Consequently, both modelling paradigms are reviewed in depth in this section, enabling a balanced assessment of their relative strengths, limitations, and suitability for different AAR research objectives.

### 5.1. Preprocessing

Before training models, we have to preprocess the collected data which entails cleaning the data, extracting features from the data and assigning the data to windows.

#### 5.1.1. Data cleaning

Data cleaning is required after a data collection process to make sure the data is consistent, reliable, and usable. For example, there could be parts of the data that has missing or erroneous labels, especially in cases with multiple sensors being used. Within the reviewed literature, 85 of the articles either acknowledge that they clean their data, but do not elaborate on how this is done, or do not mention data cleaning at all. However, a range of data cleaning methods are mentioned across the remaining articles. There are some rather straight forward ways authors have cleaned their data, for example, some clean the data by removing invalid data [155,192], such as the data from a particular sensor that is not required, but was collected along side the other sensors. Many works focused on identifying and removing corrupted or implausible data points, including statistical outliers, physically impossible sensor values, duplicated or inconsistent timestamps, constant-value segments, and entries with invalid or unknown behaviour labels. Such filtering was used to mitigate sensor dropouts and improve overall data quality [18,54,56–59,121,126,162,176,192,193]. Missing data were handled through either deletion or imputation, depending on methodological requirements. Some studies replaced missing values with neutral placeholders (e.g., zero filling) to preserve sequence continuity for deep learning models, while others removed incomplete samples to avoid introducing synthetic bias [11,56]. Duplicate records which can occur from sensor transmission artefacts or logging errors were also removed as part of quality control, helping to prevent artificially inflated performance during cross-validation [18,58]. A number of studies applied feature scaling, normalisation, or standardisation to harmonise sensor magnitudes and to support stable model convergence. These transformations ensured that variations between devices, deployments, or animals did not distort the learning process [19,54,194]. In addition, studies could also address the influence of gravity on accelerometer signals by removing or separating the gravitational component, either through high-pass filtering or sensor fusion techniques, to better isolate dynamic movement patterns relevant to behaviour classification. Signal filtering, such as low-pass or band-pass filtering, could also be applied to suppress high frequency noise and motion artefacts, improving signal quality and reducing the impact of sensor vibrations and environmental interference on downstream feature extraction and model performance. Label integrity was another central theme, particularly for windowing pipelines. Researchers frequently enforced a single label per window, removing mixed-behaviour windows, restricting the dataset to segments with verified labels, or assigning the modal behaviour within a window. This reduced ambiguity in the target variable and improved classifier consistency [14,55,112,125,128,154,158,181,184,195,196]. Temporal preprocessing rules, such as applying minimum window lengths, discarding windows shorter than species-specific behavioural durations, or relying on already windowed datasets were also commonly implemented to enhance behavioural stationarity and signal clarity [14,127,132,173,181].

#### 5.1.2. Windows

A window in time-series or sensor data analysis is a segment of consecutive data points extracted over a defined period of time (e.g., 3 seconds, 5 seconds). Windows are used to break continuous data into manageable chunks for feature extraction or machine learning. The window

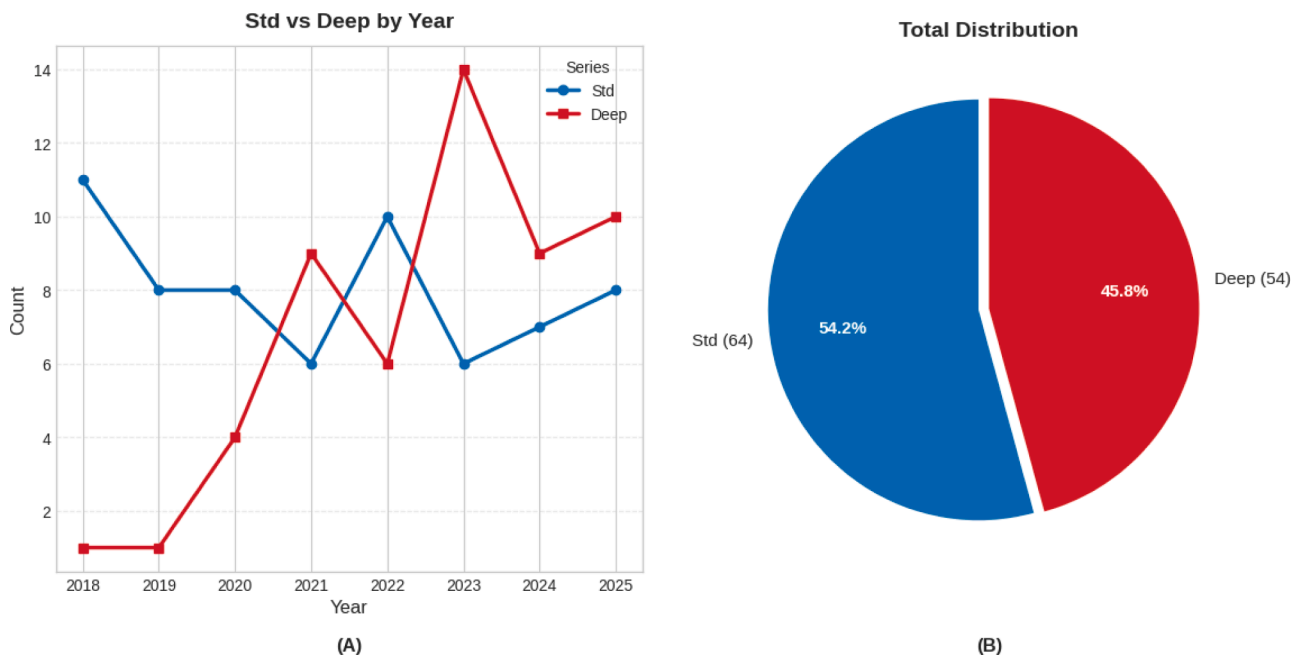


Fig. 12. Standard (Std) machine learning vs. deep learning. (A): Std and deep learning plotted by year. (B): Std and deep learning split in reviewed articles.

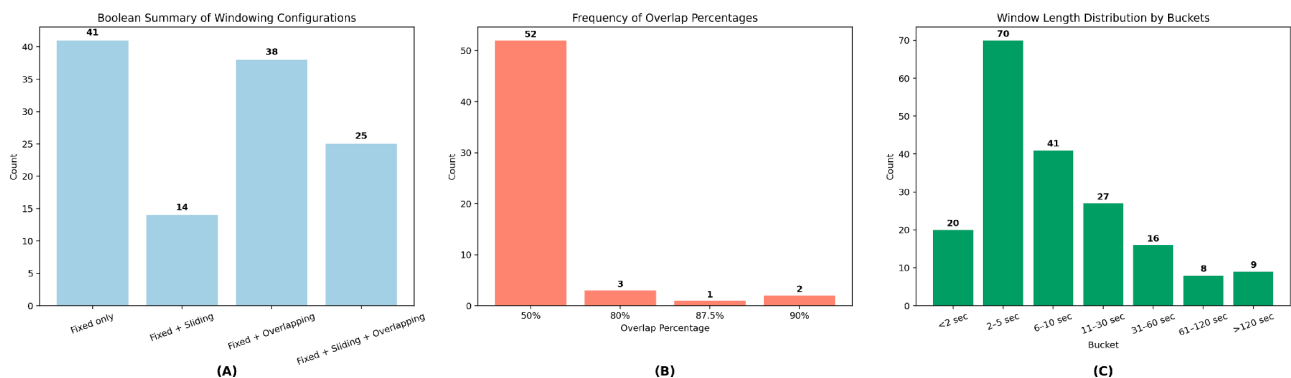


Fig. 13. Window usage across reviewed literature. (A): Windows types and combinations. (B) Overlap length usage. (C): Window length usage.

usage across all 118 of the reviewed articles have been summarised in Fig. 13, including the types of windows used, the overlapping percentage, and the lengths of windows used.

The different types of windows have been explained below considering a 5 second window length, as an example:

- **Fixed-length Windows:** each window spans for the same length of time. For example, each window could be set to a length of 5 seconds. From the reviewed literature, this is the most common windowing approach and used in all of the articles.
- **Sliding Windows:** a new window may start after a certain step size and where it may overlap with a previous one. For example, if we continue with our 5 second length example, the first window would be from 0s-5s, followed by 1s-6s, and then 2s-7s. Sliding windows are usually coupled with fixed length windows, and sometimes used with overlapping windows too, used in a total of 39 of the reviewed articles.
- **Overlapping Windows:** the windows may share some of the same data points due to the window overlapping with a certain percentage of the previous window. For example, in a 5-second window with 50% overlap, a new window would be started every 2.5 s: 0 s-5 s, followed by 2.5 s-7.5 s, then 5 s-10 s and then 7.5 s-12.5 s. Overlapping windows are often paired with fixed length windows, and occasion-

ally paired with sliding windows, and have been used in 63 of the reviewed articles.

A summary of the windowing strategies adopted across the reviewed literature is presented in Fig. 13. Overall, the results indicate a strong preference for fixed-length windowing, either used alone or combined with overlapping and sliding mechanisms. As shown in Fig. 13A, the most common approach is the use of fixed-length windows without overlap or sliding, followed closely by fixed-length windows with overlap. This suggests that most studies prioritise simplicity and reproducibility in their segmentation strategy, likely due to the ease of implementation and clear correspondence between windowed samples and behaviour labels. The widespread use of overlapping windows further reflects an effort to mitigate boundary effects, where behaviour transitions occur at the edges of windows. Fig. 13B shows that a 50% overlap is overwhelmingly dominant, appearing in the majority of studies that employ overlap. This value represents a practical compromise: it increases the effective number of training samples and improves temporal continuity without excessively inflating data redundancy or computational cost. Higher overlap ratios (80-90%) appear only sporadically in the literature, likely due to diminishing returns and increased risk of overfitting caused by highly correlated windows. Notably, several studies report the use of overlapping windows but fail to specify the overlap percentage, highlighting an ongoing lack of standardisation and

reporting clarity. Window length selection, summarised in Fig. 13C, reveals an even clearer consensus. The literature strongly favours short windows measured in seconds rather than minutes, with window lengths between two and five seconds being the most prevalent. This trend suggests that the majority of livestock activities and behaviours of interest can be captured within relatively short temporal contexts. Short windows also offer practical advantages, including faster model inference, increased training sample counts, and improved responsiveness for real-time or near-real-time systems. Longer windows, particularly those exceeding one minute, are comparatively rare and typically associated with behaviours that evolve more slowly or require broader temporal context. For example, [56,119] used window lengths of 1 minute, 1.5 minutes, and 2 minutes. Studies employing longer windows often focus on coarse grained behavioural states rather than rapid, transient actions. The fact that only a small subset of articles adopts window lengths beyond 60 seconds indicates that such long temporal contexts are the exception rather than the norm in AAR. Several studies explicitly investigate the effect of window length on model performance by experimenting with multiple window sizes. For example, in [164], they used window lengths of 10 seconds, 5 seconds, and 3 seconds. Another example could be [185], which used window lengths of 2 seconds, 5 seconds, 8 seconds, and 12 seconds. This reflects an implicit recognition that window length is a task dependent hyper-parameter, influenced by factors such as the animal species, sensor placement, sampling frequency, and target behaviours. Overall, the literature demonstrates a convergence toward short, fixed-length windows with moderate overlap, particularly 50%, as a de facto standard for livestock AAR. While this consistency supports comparability across studies, it also suggests that windowing choices are often inherited rather than optimised. Future work would benefit from more principled justifications of windowing configurations and clearer reporting, especially as AAR systems move toward deployment on resource-constrained edge devices where window size directly impacts latency, power consumption, and memory usage.

### 5.1.3. Feature extraction

Feature extraction is the process of transforming raw data into a set of meaningful, compact, and informative values (called features) that can be used as input to a machine learning model. In the context of AAR, feature extraction typically involves converting raw sensor signals into numerical values that capture key characteristics of movement over time. It is essential as features help summarise patterns in the raw data, which could be noisy or not be able to provide any useful insights directly. By having good features to summarise patterns, classification accuracy could be improved, while also reducing overfitting.

Beyond classification performance, feature extraction plays a critical role in determining computational complexity, memory usage, and energy consumption. Energy-aware human activity recognition studies highlight that feature selection and extraction incur non-negligible computational cost and energy consumption, and that careful selection can significantly reduce power use with minimal loss in accuracy [197]. The number and type of extracted features directly affect processing time, feature storage requirements, and model size, making feature selection a key design trade-off in resource-constrained deployments.

Feature extraction usually takes place after the windowing process, and there are three main types of features:

- **Time-Domain and Statistical Features:** these are features that are derived directly from raw sensor signals over a fixed time window. They summarise signal characteristics such as magnitude, variability, and shape using statistical measures such as mean, standard deviation (STD), skewness, kurtosis, and quartiles. These features are simple to compute and widely used for capturing patterns in motion, and are commonly used in the literature as demonstrated in Fig. 14A.
- **Frequency-Domain Features:** Frequency-domain features are obtained by transforming the time series signal using techniques like the Fast Fourier Transform (FFT). They capture information on the

periodicity, energy, and complexity of movement using metrics such as energy, entropy, dominant frequency, and spectral entropy. As seen in Fig. 14B, they are not as common compared to Time-Domain features.

- **Composite Features:** Composite features combine information from multiple axes or signal transformations to produce a single high-level indicator of movement. Examples include Vector Magnitude, Overall Dynamic Body Acceleration (ODBA), and Vectorial Dynamic Body Acceleration (VeDBA), which are often used as proxies for activity intensity or energy expenditure. They are not as common as the other feature categories, as shown in Fig. 14C.

Fig. 14 highlights a strong preference in the literature for time-domain and statistical features, which dominate feature usage across studies. This trend likely reflects their favourable balance between descriptive power and computational efficiency. Time-domain features are inexpensive to compute, require minimal memory overhead, and are well suited to real-time processing on low-power devices. In contrast, frequency-domain and composite features are used more selectively, often in studies where computational resources are less constrained or where periodicity and energy-related characteristics are central to the target behaviours.

In total, 35 of the reviewed articles did not specify any features they used, notably, only a minority of these papers were standard ML (5 in total), whereas the rest were deep learning. We believe that this is due to some deep learning techniques, such as CNN, which have automatic feature extraction [13], which reduces the need for manual feature engineering and extraction for deep learning approaches. However, while automatic feature learning reduces manual engineering effort, deep learning models often incur higher computational and energy costs during both training and inference, which may limit their suitability for on-device or long-term battery-powered deployments.

While frequency-domain features can provide valuable insights into periodic and repetitive movement patterns, they typically require computationally intensive transformations such as the Fast Fourier Transform (FFT), increasing computational cost [198]. These operations involve floating-point arithmetic, increased memory usage for intermediate buffers, and longer processing times, which can be prohibitive for low-power microcontrollers and embedded AAR platforms. Although approximate or low-resolution FFT implementations can reduce computational cost, this often comes at the expense of spectral precision. As a result, the inclusion of frequency-domain features represents a trade-off between representational richness and resource efficiency, and their use may not always be justified in energy-constrained or real-time applications.

### 5.2. Model validation

A range of validation strategies are employed across the reviewed literature (Table 4), reflecting differing assumptions about how model generalization should be assessed. However, while these methods are often described, they are rarely critically discussed in terms of their suitability for real-world Animal Activity Recognition (AAR) deployment.

The most frequently reported validation strategy is k-fold cross-validation (used in 38 studies), followed by train/test splits (31 studies). Both approaches are widely adopted due to their simplicity, but they can substantially overestimate performance when data from the same animal appear in both training and test sets, particularly in high-resolution time-series data where temporal autocorrelation is strong. Despite this risk, many studies do not clearly report whether splits were performed at the sample, window, or animal level.

Leave-One-Animal-Out Cross-Validation (LOAO-CV), either alone (13 studies) or combined with k-fold cross-validation (16 studies), is often presented as a more rigorous alternative intended to evaluate subject-independent generalisation.

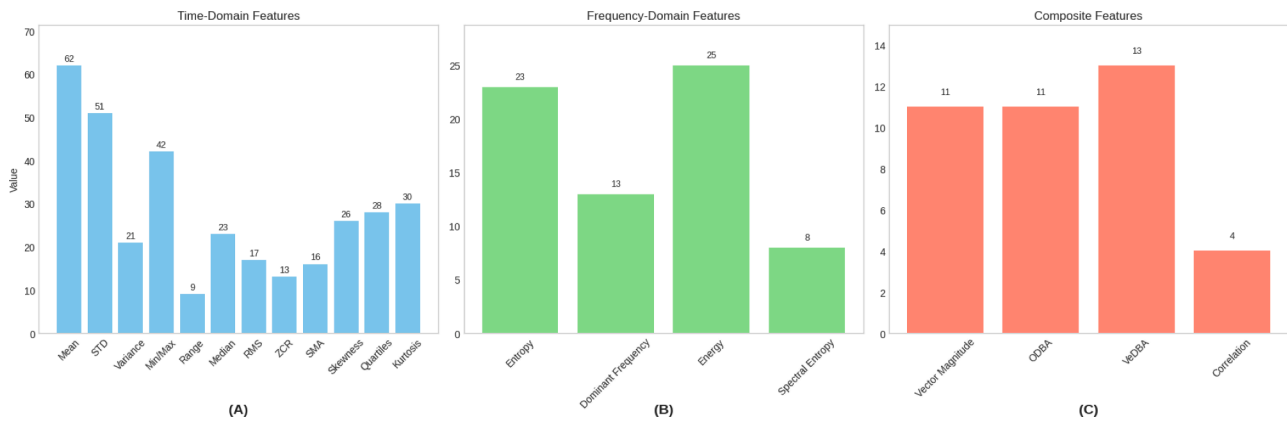


Fig. 14. Total of common features extracted from the literature.

**Table 4**  
Summary and frequency of validation strategies used in animal activity recognition studies.

Validation Type	Description	Count
<b>Leave-One-Animal-Out (LOAO)</b>	Excludes one animal from training and uses it for testing to assess generalisation across individuals.	13
<b>LOAO Combined with k-Fold CV</b>	Nested or combined approach using LOAO alongside k-fold CV (e.g., 10-fold + LOAO).	16
<b>k-Fold Cross-Validation (3, 4, 5, 10-Fold)</b>	Dataset split into <i>k</i> folds; common choices include 3-fold, 4-fold, 5-fold, and 10-fold CV.	38
<b>Train/Test Split (TTS)</b>	Simple hold-out validation; most common ratios: 80:20, 70:30, 60:40. Some include validation sets (e.g., 70:15:15).	31
<b>Leave-One-Day-Out (LODO)</b>	Excludes one day or session for testing to evaluate temporal robustness.	1
<b>External Dataset Validation / Custom Split</b>	Model validated on a separate dataset or custom subject counts (e.g., “3 animals train, 9 validate”).	3
<b>Unspecified / No Comment</b>	Validation approach not clearly reported.	14

While LOAO-CV reduces direct subject level data leakage, it does not necessarily guarantee realistic estimates of real-world generalisation. When animals are drawn from the same herd, barn, or environmental context, shared conditions such as housing, management practices, and sensor placement can lead to residual dependencies between training and test data, resulting in overly optimistic performance estimates [199]. These assumptions are rarely acknowledged explicitly in the reviewed studies.

In contrast, temporal validation strategies are notably under represented in the reviewed literature (Table 4). Only a single study employed Leave-One-Day-Out (LODO) validation, and very few studies used chronologically ordered train-test splits. The reason for this validation strategy being rarely used could be due to the low level of data collected, as highlighted in Section 4; with most articles that collected their own data containing less than 24 hours of data, so there is no day to exclude. Public datasets such as the most used cattle dataset [135] only contains approximately 6 hours of data. This is a significant limitation, as animal borne sensor data are inherently temporally structured, exhibiting strong autocorrelation and repeated behavioural routines. Validation schemes that ignore temporal ordering can therefore substantially misrepresent predictive performance on future observations. Temporal validation is critical for assessing robustness to concept drift, behavioural changes, and sensor degradation factors that are unavoidable in long-term, real-world deployments of AAR systems; whereas temporally blocked or chronological validation more accurately reflects realistic deployment scenarios [200].

Overall, Table 4 indicates a strong reliance on conventional cross-validation and random split strategies, with limited emphasis on temporal or environmental independence. This suggests that reported model performances may not fully reflect real-world operational conditions, highlighting the need for more deployment-oriented validation protocols in future AAR research.

### 5.3. Standard ML in AAR

Standard machine learning (ML) approaches remain widely used in Animal Activity Recognition (AAR), particularly in studies relying on engineered features extracted from accelerometer and inertial sensor data. Across the 64 reviewed studies employing classical ML techniques, authors typically evaluate a core set of baseline models, most commonly including Random Forest (RF), Support Vector Machines (SVM), k-Nearest Neighbours (kNN), and Decision Trees (DT). Several studies restrict their evaluations to comparisons among these baseline models only, without extending beyond the main classical ML family [57,153,181,192,195]. More commonly, however, baseline models are supplemented with additional classifiers from other methodological families. A number of studies incorporate linear or probabilistic methods such as Naïve Bayes, Linear Discriminant Analysis (LDA), or Quadratic Discriminant Analysis (QDA) to provide contrast against non-linear decision boundaries [19,54,71,128,130,131,178,183,185,201]. Others extend their comparisons by including at least one gradient-boosting approach, such as AdaBoost, XGBoost or LightGBM, alongside the main baseline models [45,125,151,176,202]. A smaller subset of studies explicitly evaluate models spanning all three of the families: baseline, boosting, and linear or probabilistic; which allows for broader comparative assessment across classical ML paradigms [18,161,186,203].

#### 5.3.1. Tree-based methods

Fig. 15A illustrates the frequency with which different classical ML models appear across the literature. Tree-based methods dominate overall usage, with Random Forest being the most frequently evaluated model, followed by SVM, kNN, and Decision Trees. This prevalence reflects their long-standing role as strong baselines for tabular, feature-engineered sensor data. Beyond frequency of use, performance trends further reinforce the dominance of tree-based approaches. As

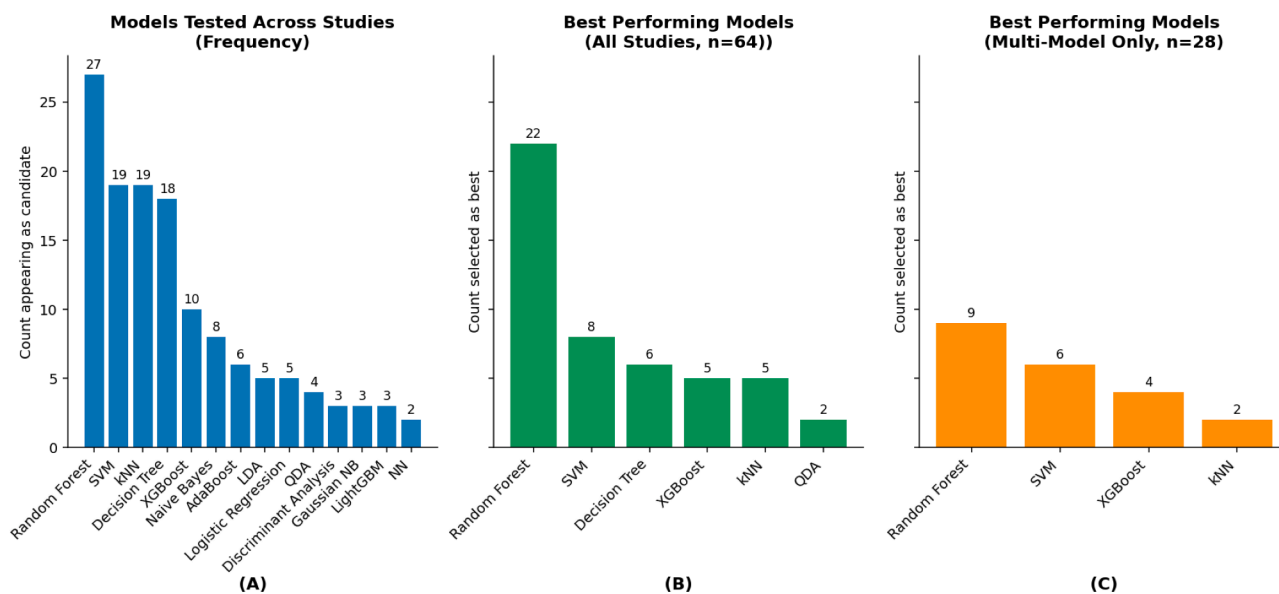


Fig. 15. Standard machine learning models used in reviewed literature. (A): Standard ML models used across studies. (B): Models that performed best in all standard ML studies. (C): Best performing models in articles that compared two or more standard ML models.

shown in Fig. 15B, Random Forest is the most consistently selected top-performing model across all studies using standard ML models, achieving the best reported performance in 22 articles. Decision Trees also were the best performing model across a number of articles, indicating that hierarchical, rule-based partitioning of feature space is well aligned with accelerometer-derived activity patterns.

The strong performance of tree-based models can be attributed to several factors. They naturally capture non-linear decision boundaries, are robust to noisy sensor measurements, and perform well with heterogeneous feature sets that may vary in scale or statistical distribution [204,205]. In addition, ensemble methods such as Random Forest are relatively insensitive to hyper parameter choices compared to more complex models, making them attractive in studies where exhaustive tuning is infeasible.

Random Forest (RF) models gave the best performance compared to other models in nine articles when the authors compared to other model types (as seen in Fig. 15C). The majority of these articles are ruminants, with three articles reporting success in cattle [130,131,206], another three articles having success in sheep [45,151,201]. There is then one article for horses [176], another one article for pigs [54] and finally, one for chickens [185].

Among the cattle articles, each of them had variation in results, behaviours and sensors used. The authors of [131] created a RF binary classification model that classified if the animal was grazing or not grazing, which achieved an accuracy score of 82.2% using an accelerometer sampling at 12Hz. The RF model created in [206] was able to classify five behaviours, with a higher accuracy score of 94% using an IMU. When comparing these articles, classifying five behaviours is more challenging and this was done with a higher accuracy score overall, however, the authors of [206] did much more feature extraction and had an additional gyroscope and magnetometer, which may have contributed towards the higher accuracy score compared to the binary classifier in [131]. The final article that produced a successful RF model had the worst accuracy score compared to the others, with a score of 87.4% [130]. However, they classified the most behaviours, classifying 12 behaviours in total, achieved using an IMU sampling at 10Hz. The larger number of behaviours this RF was able to classify is impressive, considering that it nearly matched the accuracy score of the binary classifier in [131]. We should also consider the validation techniques used in these three

articles, as [130] used 10-Fold CV, compared with the others who used train test split holdout validation only.

For the sheep articles, unlike cattle, the articles did not have as much variation. The authors of [151] created a RF model that was able to classify three behaviours with an accuracy and F1-Score of 92% using an accelerometer and gyroscope. The RF model in [45] was also able to classify three behaviours, with an accuracy and F1-Score of 92%, using an accelerometer. However, the articles vary in the number of animals they collected data from, and the types of validation used. The authors of [151] collected data from six animals and performed 10-Fold CV, while [45] collected data from only one animal and used train test split holdout validation only, meaning the RF model from [151] has been tested and validated further; and may be better at generalisation being as it was trained on more animals. The authors of [201] classified four behaviours, one more of that compared to the other sheep articles discussed. The RF model created in [201] was able to classify four behaviours with an accuracy of 94.6% and F1-Score of 88.9%. This performance is higher than that of the previous two articles, with more behaviours to classify. There are a number of reasons this may be the case, it could be that the authors collected data at a higher sampling frequency, collecting at 30Hz, compared with 16Hz in [151] and 10Hz in [45]. It could also be that the authors collected from nearly three times the animals, collecting from 17 animals; but the validation technique used by the authors was 5-Fold CV, which is a better approach compared to [45], but not as rigorous as [151].

The remaining articles had variation in the animals, number of behaviours and results. [185] created a RF model that was able to classify four behaviours within chickens, with an accuracy score of 93.9%. The model created in [54] was able to classify seven behaviours within pigs, reporting an F1-Score of 95%. And finally, the RF model in [176] was able to classify 17 behaviours in horses, with an accuracy score of 82.3% and F1-Score of 80.7%. The latter article achieved the lowest accuracy score across all of the RF articles discussed, however, they did attempt to classify the most behaviours, which is an important piece of context and is very impressive when taken into account.

Taken together, these results demonstrate that tree-based models, and Random Forest in particular, offer a strong balance between classification performance, robustness, and practical applicability in AAR. Their effectiveness is evident across species, sensor configurations, and

behaviour complexities, with many studies achieving high accuracy even when classifying a large number of behaviours. However, reported performance is strongly influenced by experimental design choices, including the number of animals, sensor modalities, feature engineering, and validation strategy. As such, while Random Forest models consistently emerge as top performers in comparative evaluations, careful consideration of dataset diversity and validation remains essential when interpreting their suitability for real-world deployment.

### 5.3.2. SVM and distance based methods

Support Vector Machines (SVMs) form a clear second tier, giving the best results in six articles (where articles compared to other models) [19,57,128,181,183,195]. While their overall win rate is lower than that of Random Forest, SVMs are often competitive on lower-dimensional feature sets with well-separated class structures. In contrast, distance based methods such as kNN and probabilistic or linear classifiers (e.g., Naïve Bayes, LDA, QDA) are frequently included as baselines but are rarely selected as top performers in most cases, with exception in [178,192], reflecting their limited capacity to model complex, non-linear activity patterns.

When investigating the results in the six articles where SVM performed best compared to other models (as seen in Fig. 15C), three of the articles were investigating chickens. In [183], the authors SVM was able to classify four behaviours with 84.8% accuracy, with an F1-Score of 86.1%. The authors later increased the behaviour count to six, and improved the SVM classification with an accuracy of 87.54% and F1-Score of 88.4% [19]. Then the authors of [181] created an SVM model that was able to classify four behaviours with an accuracy of 95.25%. When comparing these articles, they share many similarities. All three of the articles collected their own data, using the accelerometer running at a sampling frequency of 40Hz. Interestingly, the work of [181] and [183] attempted to classify the same behaviours within chickens, these being Resting, Walking, Feeding and Drinking; yet, the authors of [181] had a better performing model. This could be due to the number of differences between the articles, mainly the authors of [183] used 10-Fold CV and LOAO validation, and collected data from 12 different animals, compared to [181] who used 5-Fold CV only, and collected data from nine different animals. We could argue that the results from [183] may be more closely aligned to that of the real-world, having more validation take place and more data from different animals. In addition to this, the later work of [19] were able to create an SVM that performed better than their original work, classifying six behaviours in total: Sitting, Standing, Feeding, Drinking, Walking and Preening. Compared to their original work in [183], the authors collected data from more animals, collecting from 30 in total. They also introduced sliding windows, and performed more feature extraction, which may have helped improve classification performance with more behaviour classes.

The other articles that produced the best results using an SVM model had accuracy scores of 90% or higher when classifying three behaviours. [195] created an SVM model that scored 93% accuracy when classifying Feeding, Ruminating and other behaviours within cattle using a single accelerometer sampling at 10Hz. Compared to their other work in [128] where they again attempted to classify three behaviours in cattle, but this time the behaviours were Lying, Standing, and Feeding with two accelerometers running at 1Hz; the SVM in this work scoring 98% accuracy when combining the data from both accelerometers. Both of these works used LOAO validation, and the same windowing methodology of fixed length 60 second windows. And finally, the authors of [57] created an SVM that was able to classify three behaviours of horses with an accuracy of 99%. However, this work collected data from only two animals, which means the models generalisation may be poor, compared to the works of [195] and [128] who collected data from 10 and 16 animals, respectively.

Overall, these findings suggest that SVMs can achieve strong performance in AAR when the number of behaviours is limited, features are carefully engineered, and data collection conditions are well con-

trolled. Their success is particularly evident in species-specific studies with relatively structured behaviour patterns and moderate feature dimensionality. However, the performance of SVMs appears sensitive to validation design, dataset size, and inter-animal variability, with studies using fewer animals or less stringent validation often reporting higher accuracies. Consequently, while SVMs remain a competitive and well-established baseline for AAR, their apparent effectiveness should be interpreted in light of experimental design choices, particularly when comparing results across studies or assessing suitability for real-world deployment.

### 5.3.3. Boosting methods

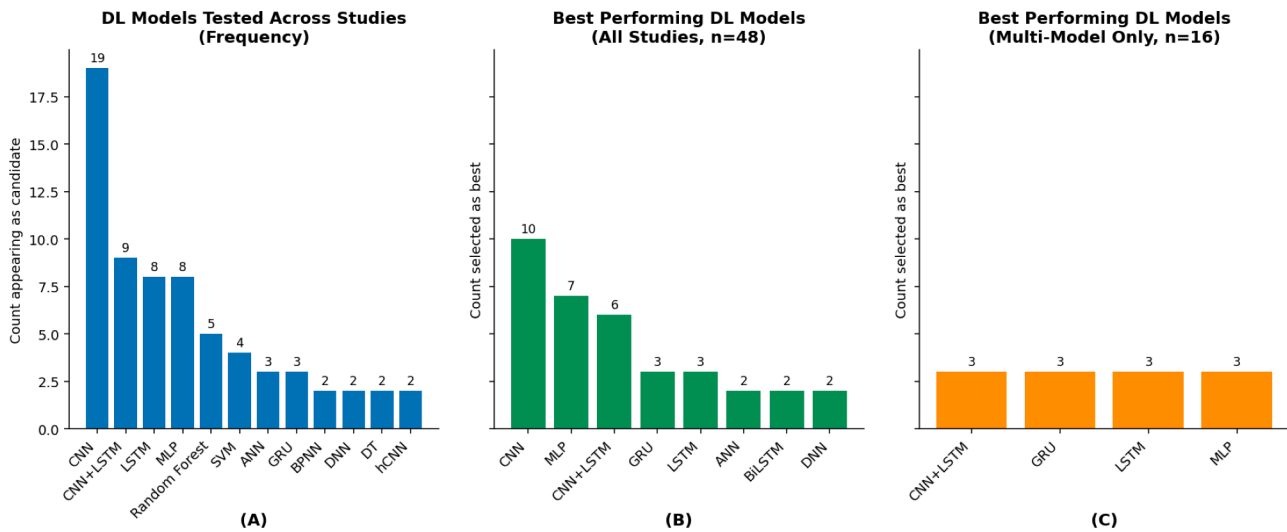
Gradient-boosting methods also feature prominently, both in terms of evaluation frequency and performance. XGBoost, in particular, is tested in a smaller number of studies but is selected as the best-performing model in approximately half of those cases [125,186,202,203]. When considered collectively, boosting approaches account for a substantial number of best-model selections, suggesting that they can outperform simpler ensembles when feature engineering is strong and class imbalance is appropriately handled. However, their higher sensitivity to hyper parameters and increased computational cost may limit their routine adoption.

The total articles that compared multiple models and XGBoost performed the best was four, and three of these articles were cattle-based [125,202,203], with one article having success with pigs [186]. All four of these articles reported XGBoost models with accuracies of over 90%, but each article was classifying a different number of behaviours, the smallest number of which was three in [203], who reported a model accuracy of 92.82% using ear tag accelerometers. The sensor position does limit the number of behaviours that can reasonably be classified compared with other sensor positions, such as collars on the neck, which was used in the other cattle articles [125] and [202], which classified six behaviours with 98% accuracy, and seven behaviours with 92.6% accuracy, respectively. The authors of [125] had the best performing XGBoost model, which was validated using 10-Fold CV. We believe that this is due to the higher sampling frequency of 59.5Hz used compared to the other articles, along with the higher number of animals the authors collected data from, collecting from 26 animals in total. Along with this, they also did the most feature extraction compared with the other articles, and was one of the rare articles that used 90% overlapping windows. The authors of [202] had the lowest performing XGBoost model with 92.6% accuracy and an F1-Score of 92%, but this model was able to classify seven behaviours, which is the most of the XGBoost models discussed. Finally, [186] was able to classify five different behaviours in pigs, creating an XGBoost model that was able to classify these behaviours with 93.4% accuracy.

Overall, gradient-boosting methods, particularly XGBoost, demonstrate strong potential in AAR when supported by rich feature engineering, high-quality data, and carefully controlled class distributions. Their success across multiple species and behaviour sets highlights their capacity to model complex, non-linear relationships beyond those captured by simpler ensembles. However, the increased sensitivity of boosting models to hyper parameter tuning, windowing design, and sampling configuration suggests that their reported performance is closely tied to experimental optimisation, which may limit reproducibility and practical deployment compared to more robust tree-ensemble baselines.

### 5.3.4. Model comparison summary

When restricting analysis to studies that explicitly compare multiple classical ML models (Fig. 15C), the dominance of Random Forest remains evident, followed by Support Vector Machines and gradient-boosting approaches such as XGBoost. This consistency across evaluation subsets reinforces the conclusion that tree-based ensemble methods represent the most reliable choice among standard ML techniques for feature-based AAR, particularly in heterogeneous datasets with varying sensor configurations, behaviour counts, and animal populations.



**Fig. 16.** Deep learning machine learning models used in reviewed literature. (A): DL models used across studies\*. (B): Models that performed best in all DL studies. (C): Best performing models in articles that compared two or more DL models. \*Studies reviewed may have compared DL models against standard ML models, hence why *Random Forest and SVM* appear in this graph.

More broadly, the reviewed literature reveals a clear performance hierarchy among classical ML models for AAR. Tree-based models, and Random Forest in particular, offer a favourable balance of accuracy, robustness to noise, and tolerance to feature heterogeneity, which likely explains their widespread adoption as both a benchmark and final model choice. Boosting methods and SVMs can achieve comparable or superior performance in carefully engineered settings, especially where feature quality is high and validation conditions are controlled; however, their effectiveness appears more sensitive to hyper-parameter tuning, dataset size, and validation design. As a result, while alternative classical models remain valuable, Random Forest continues to function as a dependable and reproducible baseline in AAR research, even as the field increasingly transitions toward deep learning approaches.

#### 5.4. Deep learning in AAR

Deep learning (DL) approaches have become increasingly prominent in Animal Activity Recognition (AAR), particularly in studies that operate directly on raw or minimally processed accelerometer time-series data. Across the reviewed DL-based literature, evaluated architectures can be broadly organised into four model families: convolutional neural networks (CNNs), recurrent neural networks (RNNs), hybrid CNN-RNN architectures, and shallow feed-forward neural networks. CNN-based models form the largest family, with studies reporting pure convolutional architectures or close variants such as hierarchical CNNs, dual-branch CNNs, and knowledge-distilled CNNs as their best-performing models [12–14,18,21,56,104,119,136,139,146,169,173]. A second major family consists of recurrent neural network (RNN) models, including LSTM, Bi-directional LSTM, hierarchical LSTM, GRU, and related variants, which are selected as top performers in a range of studies emphasising temporal sequence modelling [2,55,59,60,105,117,120,138,144,193,207]. In addition, several studies report hybrid CNN-RNN architectures, these typically combine convolutional feature extraction with LSTM or GRU-based temporal memory as their best models, reflecting an explicit integration of spatial and temporal learning [44,46,163,170,171,208,209]. Finally, a smaller subset of studies identify shallow feed-forward neural networks (FFNN), including MLPs, ANNs, DNNs, and ANN variants such as WOA-ANN, as their top-performing DL models, often in settings with lower architectural complexity or reduced data availability [16,20,22,70,87,141,145,147,148,179,210].

##### 5.4.1. Convolutional neural networks

Across the reviewed literature, convolutional neural networks (CNNs) are the most widely evaluated deep learning architecture, as illustrated in Fig. 16A. Pure CNN models appear substantially more frequently than any other DL family, followed by hybrid CNN-LSTM architectures, stand alone LSTM variants, and MLP-based networks. This prevalence reflects the relative architectural simplicity of CNNs, their suitability for raw or lightly processed accelerometer time-series data, and their ability to automatically learn discriminative features without the need for extensive manual feature engineering [13]. As a result, CNNs are commonly adopted as a first deep learning baseline when transitioning from classical ML approaches.

When performance outcomes are considered across all DL studies collectively (Fig. 16B), pure CNN architectures also emerge as the most frequently selected best-performing model. This indicates that, despite their comparatively limited temporal memory, CNNs often achieve competitive or superior performance across a wide range of datasets, activity sets, and sampling frequencies. Their strength lies in effective local pattern extraction and robust generalisation, particularly in scenarios where behaviour recognition can be driven by short-term motion signatures rather than long-range temporal dependencies, because CNNs learn convolutional filters that automatically detect discriminative patterns within local windows of time-series sensor data and build hierarchical features across layers [211,212]. However, this dominance is less pronounced in direct, head-to-head model comparisons. As shown in Fig. 16C, CNN-based models appear less frequently as the top performer when evaluated alongside alternative DL architectures within the same study. In these stricter comparative settings, hybrid CNN-LSTM models and recurrent architectures are more often selected as the best-performing approach, suggesting that the explicit modelling of longer temporal context provides a measurable advantage when datasets and evaluation protocols allow for it. This pattern indicates that while CNNs offer a strong and reliable baseline for DL-based AAR, their performance ceiling may be surpassed by hybrid architectures when temporal dynamics play a central role in behaviour discrimination.

Where there were model comparisons, a Hybrid CNN that consisted of a CNN fused with statistical features that retained knowledge of the global time-series form gave the best performance in [2]. The authors compared this hybrid CNN to a RF model and standard CNN model, with the hybrid model able to successfully classify seven behaviours within horses with 99.59% accuracy using a sampling frequency of 10Hz. The

authors also noted that the CNN and hybrid CNN models used less energy when compared with the RF model used in the same study.

In summary, CNN architectures form the backbone of deep learning approaches in AAR, serving both as the most commonly evaluated models and as strong stand alone performers across diverse experimental settings. Their ability to learn discriminative representations directly from raw sensor data makes them an effective and computationally efficient baseline, particularly when activity patterns are dominated by short-term motion characteristics. However, evidence from direct model comparisons indicates that their performance advantage diminishes when richer temporal modelling is introduced, with hybrid CNN-RNN architectures often achieving superior results. This suggests that while CNNs provide a reliable foundation for DL-based AAR, incorporating explicit temporal context can be crucial for maximising performance in more complex behaviour recognition tasks.

#### 5.4.2. Recurrent neural networks

Recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and their variants (e.g., Bidirectional-LSTM and hierarchical LSTM), are widely explored in deep-learning-based AAR due to their ability to explicitly model temporal dependencies in sequential sensor data. As shown in Fig. 16A, RNN-based architectures are evaluated less frequently than pure CNN models but appear regularly across the literature, often alongside CNNs or as part of hybrid CNN-RNN frameworks. Their continued inclusion reflects the expectation that many animal behaviours are characterised not only by short-term motion signatures, but also by longer-duration temporal structure that unfolds over time. When performance is considered across all DL studies (Fig. 16B), RNN family models are selected as the best-performing approach in a smaller but consistent subset of articles. This includes studies in which LSTM, GRU, or BiLSTM architectures outperform simpler feed-forward networks and, in some cases, pure CNNs.

Their advantage is most apparent in tasks involving behaviours with extended temporal dynamics or subtle transitions between behavioural states, where access to historical context improves class discrimination. However, RNNs are less frequently dominant in isolation and are more commonly reported as strong competitors, particularly when evaluated against hybrid CNN-RNN models that combine convolutional feature extraction with recurrent temporal modelling [213,214].

When articles compared different RNNs against other models, the LSTM and GRU were often the best performing when observing Fig. 16C (along with CNN and LSTM Hybrid models, which will be discussed in their own section). In the case of both Long Short Term Memory (LSTM) models and Gated Recurrent Unit (GRU) models, each model gave the best performance in three articles each. In terms of animal distribution, five of these six articles had focus on cattle, with one paper taking a focus on goats [163]. LSTM models (or variations of them) performed the best in [59,163,207], while GRU models performed best in [138,144,193].

The authors of [207] created an LSTM model that was capable of classifying four cattle-based behaviours, with an F1-Score of 95.13%. The LSTM model in this article was titled a VAE-LSTM, which consisted of a Variational Auto Encoder in addition to the LSTM, as this allowed more robust local features. The LSTM they created outperformed the CNN they compared it to, which was validated using 5-Fold CV. Comparatively, the authors of [59] created an LSTM model that was capable of classifying eight different behaviours within cattle, with an accuracy and F1-Score of 88.7%, which compared to the work of [207] does not perform as well, but [59] is attempting to classify double the behaviours (eight compared to four). The final LSTM model was created by the authors of [163], who created a hierarchical Bi-directional LSTM that was capable of classifying five behaviours in goats with an F1-Score of 87.9%. The idea of the hierarchical model to derive the behaviour from a given accelerometer signal step by step, as an example, they checked if the goat is grazing, or not grazing, and then check if the goat is resting, ruminating or chewing. The authors compared this

hierarchical LSTM with a standard LSTM, and their hierarchical LSTM marginally outperformed the standard LSTM. When evaluating these three works ([59,163,207]), each of them trained their models using accelerometer data using sampling frequencies of 20Hz-25Hz. The behaviours between the cattle-based articles [59,207] are also similar, as the authors of [207] used the dataset of [59]. The goat article by [163], the authors have made some unique contributions, with both the use of a hierarchical LSTM model, and collecting acceleration data from the goats horn.

In the GRU model articles, there are a number of similarities. First, all three articles focused on cattle AAR, classified three to four behaviours, and compared their GRU models to LSTM-based models, with two articles focusing on RNN models specifically [138,144]. The authors of [138] created a GRU model that was capable of classifying four behaviours with an accuracy score of 97.4%, which outperformed other RNN-based models / architectures like the LSTM, and SRL (Simple Recurrent Layer). The authors of [144] created GRU models that are capable of classifying four behaviours, one model was trained using accelerometer data collected from the neck of the animal, another model was trained using accelerometer data collected from the ear tag. Each model performed well, achieving accuracy scores of 89.5% for the neck-based model, and 80% accuracy for the ear-based model. This trend is similar to what we have discussed in the standard ML methods, models trained on ear tag sensor data do not perform as well compared to those trained on other positions. When comparing these works, they attempted to classify the same number of behaviours, but the work of [138] performs better. This could be for a number of reasons, such as the number of animals and validation technique difference between the articles, with [138] training their model with data from four animals and using 3-Fold CV on their model, and [144] training their model using ten animals and using LOAO validation. The authors of [144] also compared their model performance to LSTM models, like that of [138], but in addition they also compared performance to a state-of-the-art CNN. In [193], the authors were able to produce a GRU model that was able to classify three behaviours with an accuracy score of 97.78%, which is the best performing GRU model reported in the reviewed literature, but again, with one less class compared to [138]. The authors compared their GRU model with a range of other model types, including a RF, CNN and LSTM hybrid model, and a ANN model.

Overall, RNN-based architectures demonstrate clear strengths in modelling temporal dependencies within accelerometer time-series data and are consistently competitive in deep learning AAR. While LSTM and GRU models are less frequently dominant than pure CNNs when results are aggregated across all studies, they emerge more prominently in direct model comparisons, particularly in scenarios involving long duration behaviours, hierarchical behaviours, or subtle temporal transitions. Performance differences across studies further highlight the influence of activity complexity, sensor placement, validation strategy, and dataset size on reported outcomes. Collectively, these findings indicate that RNNs serve as a powerful complementary approach to CNNs, with their full potential most effectively realised either in carefully designed stand alone architectures or, more commonly, when integrated into hybrid CNN-RNN frameworks that jointly exploit local feature extraction and long-range temporal context.

#### 5.4.3. Hybrid CNN and LSTM models

Hybrid CNN-RNN architectures, most commonly combining convolutional neural networks (CNNs) with Long Short-Term Memory (LSTM) layers, are increasingly adopted in deep learning AAR. These models are explicitly designed to leverage the complementary strengths of their constituent components: CNNs act as automatic feature extractors that capture local temporal and spatial patterns from raw accelerometer signals, while recurrent layers model longer-range temporal dependencies across sequential windows. As illustrated in Fig. 16A, CNN and LSTM hybrids are evaluated less frequently than pure CNN architectures but appear consistently across the literature, more so than stand alone

LSTM models. In a previous literature review published in 2023, the authors reported a low number of articles using CNN and RNN hybrid approaches [5], and the increased number of these studies shows that these hybrid approaches have become more popular in recent years, supported by five research articles published in 2025 using CNN and LSTM hybrid models, reporting that they were the best performing models [46,143,170,171,209].

When performance is considered across all deep learning studies (Fig. 16B), hybrid CNN and LSTM models emerge as one of the most competitive architectural families, being selected as the best-performing approach in a substantial proportion of articles. Their relative strength becomes even more pronounced in direct, head-to-head model comparisons (Fig. 16C), where hybrid architectures are frequently favoured over both stand alone CNNs and another stand alone RNN-based models. This trend suggests that jointly modelling local motion signatures and longer term temporal structure provides a tangible advantage in many AAR scenarios, particularly when behaviours exhibit both distinctive short-term patterns and meaningful temporal evolution.

When investigating the three articles that reported CNN and LSTM models performed best in head-to-head model comparison, each article had a focus on a different animal, and a different number of behaviours they are attempting to classify. With the head-to-head comparisons, authors will generally compare their hybrid model to a stand alone CNN and stand alone LSTM. For instance, [170] created a binary classification CNN and LSTM hybrid model that was able to classify if horses were grazing or not grazing, they created a stand alone CNN, stand alone LSTM and a hybrid model, which gave the best performance with an accuracy score of 98.6%, the best performing hybrid model of the three in this discussion, but again, this was a binary classifier, and therefore classified the least number of behaviours. Comparatively, [44] created a hybrid CNN and LSTM that was able to classify seven different behaviours in sheep, with an overall accuracy score of 91.3%, which again, is a lower accuracy score compared to [170], but with five additional behaviours, totalling seven. Like that of [170], the authors compared their hybrid model to a stand alone CNN and LSTM, along with a Temporal Convolutional Network (TCN) model. In addition to this, the model in [44] was trained on IMU data, giving more features / context of movements when compared to [170] using accelerometer only data. In [46], the authors created a CNN and LSTM hybrid model that was able to classify four behaviours in pigs, with an accuracy score of 73.2% and F1-Score of 61.87%; which compared to the other articles discussed is a drop in performance, as [44] achieved better performance when classifying more behaviours, and both of these articles trained models using IMU sensor data. However, the authors of [44] trained their model using data collected at 20Hz, whereas the authors of [46] trained their model on 10Hz IMU data, which may show that higher sampling frequencies may have produced better results. Additionally, the authors compared their hybrid model to a Back Propagation Neural Network (BPNN) model, RF model and a Extreme Learning Machine (ELM) model; not to a stand alone CNN and LSTM like others.

Overall, the reviewed literature indicates that CNN-LSTM hybrid architectures represent one of the most powerful and flexible deep learning approaches for AAR. Their consistent success in direct, head-to-head comparisons highlights the advantage of integrating convolutional feature extraction with explicit temporal modelling, particularly for behaviours characterised by both distinctive short-term motion patterns and longer-term behavioural structure. While reported performance varies with factors such as the number of behaviours, sensor modality, sampling frequency, and validation strategy, hybrid models frequently outperform stand alone CNN and RNN counterparts under comparable conditions. The growing number of recent studies adopting CNN-LSTM hybrids, especially those published in 2025, further suggests that these architectures are becoming a preferred design choice for DL AAR.

#### 5.4.4. Feed-forward neural networks

Feed-forward neural networks (FFNNs), including Multi-Layer Perceptrons (MLP), Artificial Neural Networks (ANN), and deeper fully connected variants (DNN), appear regularly across DL AAR literature, most often as comparatively simple neural baselines. As shown in Fig. 16A, FFNN-based models are evaluated less frequently than convolutional or recurrent architectures but remain a common inclusion, particularly in studies transitioning from classical ML approaches to neural networks. Their continued use reflects their conceptual simplicity, low computational overhead, and ease of implementation when working with engineered or lightly processed sensor features rather than raw time-series signals.

When performance is considered across all deep learning studies (Fig. 16B), FFNN architectures are occasionally selected as the best-performing model, though less consistently than CNN-based, recurrent, or hybrid architectures. In most cases, FFNNs perform competitively on simpler classification tasks involving a limited number of classes or well-separated motion patterns, but are outperformed when temporal dynamics or complex behaviour transitions must be modelled explicitly. This performance gap reflects the inherent architectural limitation of FFNNs, which lack both convolutional mechanisms for local pattern extraction and recurrent structures for modelling temporal dependencies across time-series data.

We will discuss the three papers that did head-to-head model comparisons where FFNN provided the best performance (these all happen to be MLP models), each of these papers focused on a different livestock animal. In [147], the authors created an MLP model that was able to classify four different behaviours within cattle with an accuracy score of 93.4%. The MLP model was the only DL that was compared, with the rest of the models in the head-to-head being standard ML models, such as Logistic Regression, SVM, DT, and others. The authors of [16] were able to create an MLP model that had similar performance, with the model scoring 94.4% accuracy, being able to classify nine behaviours in goats. This is a better performance of the work in [147] and with more behaviours, however, the work of [16] used a dataset that had a sampling frequency of 200Hz across six animals using both the accelerometer and gyroscope, whereas [147] collected their own data at 50Hz across 10 animals using only the accelerometer, meaning that there may be variations in generalisation based on these factors, and additional movement context from the gyroscope could have contributed to this better performance. Both of the authors employed cross validation and LOAO validation. Unlike [147], the authors of [16] compared their MLP against other DL models, such as a CNN and a hybrid CNN that combined the CNN and MLP layers, while also comparing to standard ML models, including RF and SVM models. In [179], the authors provided results from both validation methods they used, finding that their MLP performed best when validated with LOAO, giving an accuracy score of 71%, which is low compared to the other articles, however, they were attempting to classify 12 behaviours in chickens and only collected data from four animals, meaning that the model would have only three animals worth of data and LOAO validation would naturally perform lower due to the missing context from all the animals. This is demonstrated with the article reporting a nearly 20% increase in accuracy with 10-Fold CV instead of LOAO validation, with a Light Gradient Boosting Machine (LGBM) model giving an accuracy score of 89.9%, which is the lowest score across the three papers, but they did classify the most behaviours.

Overall, FFNN-based approaches occupy a limited role in DL AAR. While MLP and ANN models can achieve strong performance in specific settings, particularly when behaviour classes are well separated, sensor features are informative, and validation protocols are favourable their success is less consistent than that of convolutional, recurrent, or hybrid architectures. The reviewed head-to-head comparisons suggest that

FFNNs can outperform more complex models when temporal dependencies are weak or implicitly captured through high sampling rates and rich feature representations. However, their lack of explicit temporal modelling constrains scalability as activity complexity increases. Consequently, FFNNs are best viewed as effective neural baselines or task-specific solutions rather than general-purpose architectures for AAR, especially in applications involving complex behavioural dynamics or limited training data.

#### 5.4.5. Model comparison summary

When restricting analysis to studies that explicitly compare multiple deep learning architectures (Fig. 16C), hybrid CNN-RNN models emerge as the most consistently selected top-performing approach, followed closely by stand alone recurrent architectures (LSTM and GRU variants). In contrast, pure CNN models while dominant in terms of overall usage and aggregate best performance counts appear less frequently as the top performer in direct, head-to-head comparisons. This shift highlights the importance of explicit temporal modelling when architectures are evaluated under comparable experimental conditions, reinforcing the advantage of integrating convolutional feature extraction with recurrent temporal memory in complex AAR tasks.

More broadly, the reviewed deep learning literature reveals a nuanced performance hierarchy that depends strongly on evaluation context. Pure CNN architectures offer a robust and computationally efficient baseline, achieving strong generalisation across diverse datasets, sampling frequencies, and behaviour sets, particularly when short-term motion patterns dominate activity discrimination. Recurrent models provide measurable benefits in scenarios involving extended temporal dynamics, hierarchical behaviours, or subtle behavioural transitions, though their performance is less consistently dominant in isolation. Hybrid CNN-RNN architectures most reliably combine these strengths, frequently outperforming both constituent models when directly compared, and are increasingly adopted in recent studies as the preferred DL design choice. In contrast, shallow feed-forward neural networks demonstrate competitive performance only in constrained settings with limited activity complexity or rich engineered features, and are rarely superior when temporal structure plays a central role. Collectively, these findings suggest that while CNNs remain a strong entry point for DL-based AAR, hybrid CNN-RNN models currently represent the most reliable and generalisable deep learning solution for behaviour recognition tasks characterised by complex temporal structure.

#### 5.5. ML in AAR discussion

Across the reviewed literature, performance in AAR must be interpreted relative to task complexity, sensor configuration, and evaluation protocol, rather than by absolute accuracy values alone. Studies addressing binary or low-complexity classification tasks (e.g., two-four behaviours) frequently report accuracies exceeding 95% using both classical and deep learning approaches, particularly when datasets are well balanced and evaluated under cross-validation. As the number of behaviours increases and behavioural boundaries become less distinct, performance naturally declines, with accuracies in the range of 85-92% commonly reported for mid-complexity tasks involving six-nine behaviours, and values between 70-85% remaining competitive for highly complex scenarios involving ten or more behaviours, limited subject numbers, or LOAO validation.

Within this context, tree-based ensemble methods, Random Forest in particular, emerge as the most reliable standard ML models, offering strong generalisation and competitive performance across a wide range of datasets with engineered features. Deep learning approaches provide additional gains when sufficient data and temporal structure are available, with pure CNNs serving as robust baselines for raw time-series modelling, recurrent architectures excelling in tasks with extended temporal dependencies, and hybrid CNN-RNN models most consistently achieving top performance in direct comparisons. Feed-forward neural

**Table 5**

Typical classification accuracy reported for different behaviour set sizes across the reviewed studies. Accuracy ranges represent the interquartile range (IQR) where sufficient data were available.

Behaviour Classes	Typical Accuracy Range (IQR)	Median Accuracy	Studies (N)
2-5	86.9% – 96.2%	~93%	75
6-7	88.4% – 99.6%	~92%	19
8-12	79.2% – 88.6%	~87%	8
>12	~82% – 95%	~86%–95%	3

networks, while computationally efficient, are generally effective only for simpler tasks or feature-based inputs. Overall, the literature indicates that strong AAR performance is not defined by a single model class, but by alignment between model capacity, behaviour complexity, data richness, and validation reinforcing the importance of contextualised evaluation when comparing results across studies, which is especially important for real-world deployment. Some studies have also explored reducing sensor sampling frequency while maintaining acceptable behaviour classification performance, typically incurring only modest reductions in accuracy [13,159]. This trade-off offers practical benefits for live-stock monitoring, including extended battery life, reduced data storage and transmission requirements, and lower computational demands, all of which are particularly relevant for long-term, field-deployed sensing systems.

In Fig. 17, we plot the accuracy of each article in line with the number of classes, while Table 5 summarises the typical performance ranges observed across studies with different behaviour set sizes. This suggests that classification performance in AAR does not decrease linearly with the number of behaviours, but there is a general trend where tasks with a larger number of classes tend to have slightly lower accuracies. For example, studies targeting two-five behaviours commonly achieved high accuracies, often exceeding 90%, indicating that simple behaviour sets are relatively straightforward to classify. In contrast, studies with more complex classification tasks, such as 7-12 behaviours, or in some cases up to 17, often exhibited more variable performance, with some models dropping below 80%. However, there are notable exceptions: some studies achieved very high accuracies (95-99%) even with seven-ten behaviours, suggesting that model choice, feature engineering, sensor type, and data quality play a critical role in mitigating the challenge of multi-class classification. Overall, while an increase in the number of behaviours can make classification more challenging, careful experimental design, sensor placement, and model selection can maintain high performance even for relatively complex behaviour sets.

Looking forward, the evolution of AAR research appears to be most strongly aligned with Edge-AI, where behaviour recognition models are deployed directly on collar-mounted or ear-tag sensor devices. By performing inference on-device, reliance on continuous connectivity is reduced, data can be kept local, and communication bottlenecks associated with remote or bandwidth-limited agricultural environments can be mitigated. This paradigm, commonly referred to as TinyML, enables compact, energy-efficient models tailored to the constraints of embedded hardware and is a natural progression for AAR given the increasing scale of IoT deployments in precision livestock farming. Edge-AI is therefore a timely and well-motivated research direction, and is discussed in detail in Section 6. A secondary, but promising, research avenue lies in semi-supervised learning. While the reviewed literature shows very limited adoption of unsupervised methods [110,215] and no clear use of semi-supervised learning to date for classification, this gap reflects practical challenges rather than lack of relevance. The scarcity of high-quality labelled datasets remains a recurring limitation in AAR, as most studies rely on labour-intensive, bespoke data collection and annotation. Semi-supervised approaches, which combine small amounts of labelled data with larger volumes of unlabelled sensor data, may therefore offer a viable pathway to improve model scalability and generalisation as AAR systems mature.

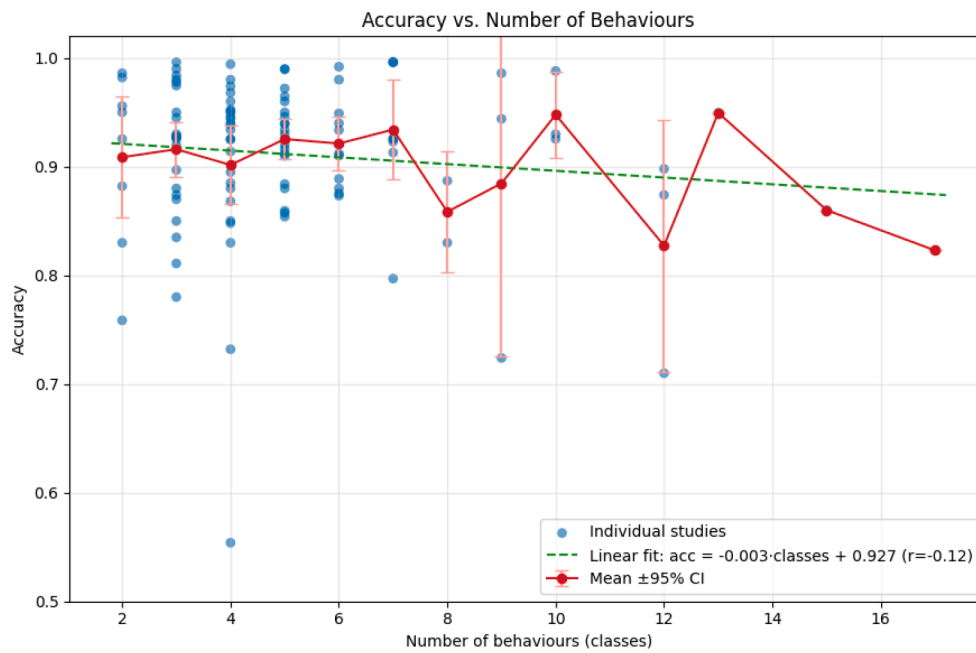


Fig. 17. Reported accuracy score plotted against number of behaviours / classes.

## 6. Edge-AI and TinyML in animal activity recognition

In this section, we will discuss the concepts of both Edge-AI and TinyML, and also discuss their usage in livestock-based AAR. Edge-AI and TinyML share the same goal of bringing AI to the edge, but each has a different definition. Edge-AI refers to AI that is implemented in smart devices [216], but it has taken on different synonyms within the literature, such as Edge Intelligence, AIoT, and TinyML [217]. However, while Edge-AI is the utilisation of AI in smart devices such as Raspberry Pi, Nvidia Jetson, or mobile phones [217]; TinyML is more specific with a focus on executing optimised ML models on ultra-low powered microcontrollers [218]. In the case of TinyML, efficiency of the ML model's and energy consumption are fundamental as microcontrollers often run on batteries, not mains power.

There are several benefits in bringing AI to the Edge using Edge-AI and TinyML [219,220], including:

- **Lower Latency:** using non-edge AI will often require communicating data between the sensor device and a cloud server, awaiting both processing and the result; whereas Edge solutions process the data on-device, removing the communications and enabling lower latency.
- **Improved Network Bandwidth:** many devices are used within WSNs, which increase the network bandwidth. Reducing the amount of data that is transmitted on the network by processing the data on the device would reduce the bandwidth of the network.
- **Scalability and Reliability:** transmitting data to the cloud constantly has limitations both in bandwidth and sometimes fair usage limits; by processing the data at the edge, this removes these limitations and also increases reliability, as this also removes a possible point of failure with network connectivity.
- **Security and Privacy:** As the data does not need to be transmitted from the device to the cloud, security increases due to fewer threats of a cyber attack. In an agricultural context, conflicts are growing between farmers and cloud providers, driven by allegations of data misuse [221]. In addition, farmers are increasingly concerned about privacy and security issues [191].

Many of these benefits described apply well to farming, especially because farms tend to be located in rural areas with limited connectiv-

ity and telecommunication infrastructure [3]. As outlined in Section 3, there are different communication methods and network types to allow data transmission between devices on farms, but these face the same issues of latency, limited bandwidth and scalability. Therefore, removing the need to communicate sensor data will significantly reduce these limitations, as instead a single behaviour label could be sent at certain times rather than a constant stream of sensor data.

### 6.1. Model size reduction, trade-offs and energy efficiency

Before moving on to how Edge-AI and TinyML have been used in the literature for AAR, we will discuss some of the main aspects required for them. One of the most fundamental steps is reducing the model size, especially in TinyML as microcontrollers are often limited on memory; therefore, the model has to be compressed small enough to fit onto the small flash memory of a microcontroller. Microcontrollers typically only have kilobytes (kB) of on-device ROM and SRAM. For example, a common microcontroller, the Arduino Nano ESP32<sup>5</sup> has a SRAM size of 512kB and ROM of 384kB. Another fundamental consideration is energy efficiency, as TinyML targets specifically battery-powered microcontrollers [222]. There are a number of established methods for reducing model size and achieving energy efficiency, which are outlined below.

- **Quantization** is the process of reducing the precision of some of the model parameters. For example, instead of using precise 32-bit floating-point numbers, 8-bit integers may be used instead. This process removes precision and, therefore, may affect the accuracy of the model. However, this will reduce the model's memory requirements, allowing the model to fit on to a microcontroller.
- **Pruning** involves removing less significant weights or neurones from the network, reducing the computational load and memory required; resulting in minimal impacts on the performance and output of the models.
- **Knowledge Distillation** is a reduction technique in which a smaller, lightweight "student" model is trained to replicate the behaviour of a larger "teacher" model, transferring its generalisation capabilities while retaining a compact architecture.

<sup>5</sup> <https://docs.arduino.cc/hardware/nano-esp32/>

- **Weight Clustering** (also known as Weight Sharing) groups similar weight values together and replaces them with shared centroids, further compressing the model and allowing efficient storage and computation.

While model compression techniques such as quantization, pruning, and knowledge distillation enable deployment on resource-constrained devices, they introduce important trade-offs that directly affect real-time feasibility, energy consumption, and classification performance in AAR. We will discuss the trade-offs for each of the methods outlined above.

In quantization, the effectiveness of it depends strongly on the selected bit-width and the underlying hardware support. Reducing weights and activations from 64-bit or 32-bit floating point numbers to 8-bit integers typically yields substantial memory savings (up to 4 times) and faster inference due to simpler arithmetic operations, which has been applied in a number of articles for AAR [119,137,138,169,209]. However, aggressive quantization (e.g., 4-bit or binary networks) can significantly degrade model accuracy, particularly for AAR tasks that rely on subtle temporal variations in accelerometer and gyroscope signals. None of the reviewed articles use aggressive quantization, with the heaviest level of quantization being 64-bit floating point numbers to 8-bit integers.

Pruning reduces the number of parameters and operations by removing redundant or low-importance weights or neurons. While unstructured pruning can substantially reduce parameter counts, it does not always translate to lower inference latency or energy consumption on microcontrollers, as irregular memory access patterns are poorly supported by embedded hardware. Structured pruning (e.g., removing entire filters or channels) is more compatible with TinyML inference engines, but may require careful architecture redesign to preserve model expressiveness. In AAR, excessive pruning risks removing features critical for distinguishing similar behaviours (e.g., grazing vs. walking), leading to class confusion even if overall accuracy appears preserved. Only two of the articles mention pruning has been used, but one does not elaborate on the scale of this [209], while the other does provide a range of details about pruning being gradually applied over five stages across 250 epochs, with each stage consisting of 35 epochs being a pruning phase, and then 15 epochs being a fine-tuning phase [119].

Knowledge distillation allows compact student models to approach the performance of larger teacher networks, making it particularly attractive for TinyML deployments. However, the student model's architecture must still be carefully chosen to match microcontroller constraints such as SRAM availability and stack usage. Additionally, while distillation improves generalisation, it does not reduce inference-time computational complexity beyond what is dictated by the student model itself. For real-time AAR, this means that distillation must be combined with architecture-level optimisation (e.g., shallow CNNs or temporal convolutions) to ensure latency constraints are met. In the reviewed articles that utilise Edge-AI, only one article utilises knowledge distillation [146]. However, the authors of [104], while not Edge-AI, did utilise knowledge distillation to obtain high accuracy using lower sampling frequencies. To achieve this, they trained the teacher model in higher sampling rates, while the student was trained on lower sampling frequencies. The authors found that their method of knowledge distillation performed better than their baseline tests in multiple sampling frequencies to classify different behaviours within horses and goats, with only minor accuracy losses.

Weight clustering further reduces memory usage by sharing parameters, but its benefits depend on hardware and runtime support. On microcontrollers, clustered weights often require lookup tables or indirect memory access, which can introduce additional computational overhead and partially offset energy savings. As a result, clustering is most effective when combined with quantization and supported by optimised TinyML frameworks. However, weight clustering was not used in any of the reviewed Edge-AI articles.

Beyond model size, real-time inference requirements impose strict limits on allowable latency per window of sensor data. In AAR systems operating at sampling frequencies of 10-25 Hz, inference must often complete within tens of milliseconds to avoid backlog and increased memory usage. This constrains not only model depth but also window length and overlap strategy. Additionally, microcontrollers are constrained by limited SRAM, meaning that intermediate activation buffers often become the dominant memory bottleneck.

Energy efficiency in Edge-AI extends beyond inference alone. While local inference reduces the need for continuous wireless transmission, compressed models must still justify their energy cost relative to communication. For low-power radios such as LoRa, transmitting raw or lightly processed sensor data can dominate energy consumption. In such cases, even moderately complex on-device inference can be energy-optimal if it enables event-driven or summary-based communication. However, if models are too computationally demanding, frequent inference may negate these savings. Thus, the feasibility of TinyML-based AAR depends on a careful balance between inference frequency, model complexity, and communication strategy. Recent microcontrollers increasingly include DSP extensions or neural accelerators that significantly improve the viability of quantized models. Nevertheless, these accelerators often support only specific operations or data types, constraining model design choices.

## 6.2. Edge-AI and TinyML uses AAR

There were few results in general for Edge-AI and TinyML articles for AAR, with some of results consisting of computer vision ML-based deployments that are out of scope for this review, with minimal articles considering the target sensors of this review, the accelerometer, gyroscope and magnetometer. This indicates that Edge-AI and TinyML is a relatively new and open research area and a natural pairing with AAR when considering the following:

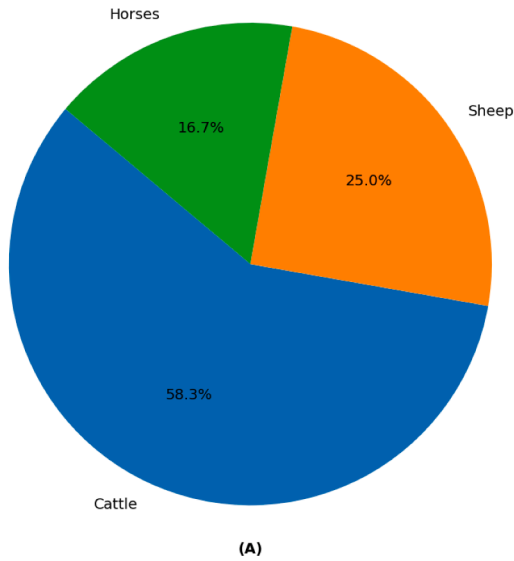
- **Communication Availability:** As outlined in Section 3, communication infrastructure is usually minimal in a farming environment due to their rural setting, meaning the cloud communications usually required for non-edge-based solutions are not possible or are limited; potentially preventing rural farms from obtaining modern AI and ML advancements, but since Edge-AI does not require data to leave the device, it therefore removes this communications barrier.
- **Labour and Maintenance:** As Edge-AI and TinyML consider battery efficiency, this would translate to less labour and maintenance for the farmer having to remove microcontroller devices from the animals from recharging. This could be further mitigated by using solar panels, reducing the need for removal almost entirely.
- **Privacy and Security:** As previously outlined, privacy and security concerns are common among farmers, and because Edge-AI and TinyML do not require the data to leave the device to go to a cloud service, it is more secure.

### 6.2.1. Overview of articles

In total, 12 papers have been reviewed that use the target sensors in this review and focus on AAR, all of which are an application of supervised learning. Most of these articles apply the model reduction and energy efficiency savings outlined above. Fig. 18A shows the distribution of livestock animals across the Edge-AI articles, and like that of standard ML, AAR strongly favours cattle. In total, cattle have seven articles, followed by sheep with three articles and two articles with horses. We suspect cattle being favoured is for the same reasons outlined in Section 4 of their economic value. However, we also suspect that more availability of public cattle datasets compared to other animals may also play a part.

When observing Fig. 18B, deep learning emerges as the dominant approach for Edge-AI AAR, with two thirds of studies employing deep learning models and one third relying on classical machine learning. A

Livestock Animals Distribution Across Edge-AI Articles



ML Types in reviewed Edge-AI Articles

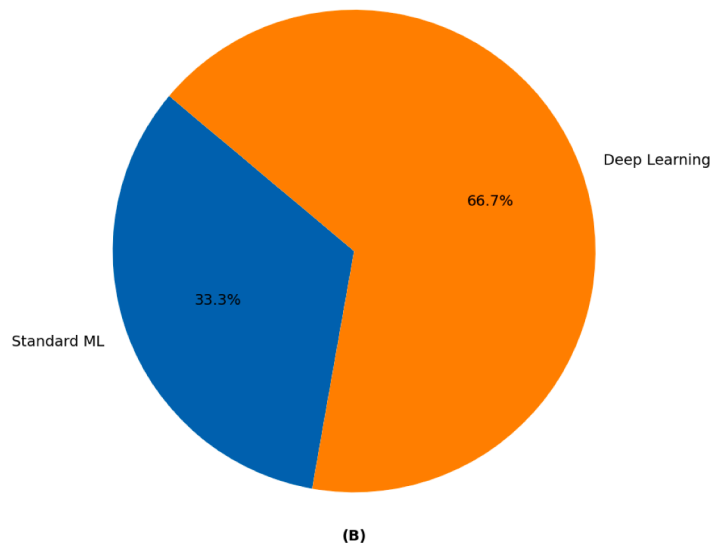


Fig. 18. (A): Livestock animal distribution across Edge-AI articles. (B): Types of ML used across Edge-AI articles.

key reason for this trend is that deep learning models perform automatic feature extraction, reducing the need for explicit preprocessing stages that can be difficult to implement on resource-constrained microcontrollers. Traditional feature-based pipelines often rely on time-frequency transformations or handcrafted statistical features, many of which require computationally intensive operations (e.g., Fourier transforms) that increase processing time, memory usage, and energy consumption. As feature extraction incurs non-negligible computational and energy costs, particularly in low-power embedded systems, avoiding complex preprocessing can significantly improve feasibility with minimal impact on classification performance [197].

There are a limited number of articles that utilise Edge-AI methodologies and deploy their ML models on to microcontrollers. However, there are also a small collection of articles within the reviewed literature that consider important Edge-AI aspects, such as energy efficiency [2,215], but ultimately do not deploy their models on a microcontroller. Some of these articles reflect that the model they created would be deployable on a microcontroller after applying model size reduction techniques, but do not do so [104,144], while some mention that they are supporting or making a contribution towards a future with on-device inference [71,147,178].

### 6.2.2. Edge-AI AAR article details

Below, we will provide a summary paragraph about each of the 12 articles that use Edge-AI and TinyML. This section also provides a summary of the models, energy, scores before and after applying compression techniques, the compression techniques applied, as well as latency and model size in Table 6, and we will conclude the findings and current state of Edge-AI and TinyML usage in AAR.

In [145], the authors did not use any model reduction techniques, as the MLP model they created was already capable of being deployed on the target microcontroller. They were able to successfully classify five different behaviours within cattle with an accuracy of 93.93%, with a small amount of latency and small model size. In the authors other work ([146]), they used knowledge distillation and quantization to produce compressed models to classify three behaviours in cattle using accelerometer data. They used a ResNet as the teacher model, and less complex student model based on GRU and MLP; finding that the accuracy of the student models improves dramatically using knowledge distillation, nearing that of the teacher model itself. Quantization was

also applied, finding that there were minimal reductions in accuracy. As an example, the model with the highest reduction after quantization was taken from 88.1% to 87.4%, which was a reduction of 0.7% to the best performing model as reported in Table 6. The lowest reduction after quantization from another model was taken from 76% before quantization to 75.8% after quantization, a reduction of only 0.2%.

Bartels et. al have made significant contributions to Edge-AI and TinyML in AAR, with three published papers [137–139]. In their first work ([137]), they start with standard ML and presented five different hardware implementations that used DTs and different features sets, which were trained on labelled accelerometer data to classify four different behaviours within cattle. During the quantization stages, there was a 10% F1-Score detriment using a conventional set of features. The best performing feature set combination was able to provide an accuracy score of approximately 87% (84.8% F1-Score) with a low power consumption of  $216\mu\text{W}$ , and energy consumption of  $557\mu\text{J}$ . They then followed up with deep learning, in [138] they compared different types of RNNs to regress four different behaviours within cattle using accelerometer data, with the aim of finding the minimal power consumption and the maximum accuracy. The best performing model provided an accuracy of 95.2% and a power consumption of  $154\text{nW}$ . Both of these works used quantization, moving from 64-bit floating point numbers down to 8-bit integers. In their other article ([139]), they created a quantised RNN capable of running on a low-power microcontroller that was able to classify four different behaviours within cattle. They were able to reduce power consumption significantly by adopting real-time processing, i.e. processing the data sample as it is received. They also set the clock frequency to complete a single RNN time step before the next sensor reading arrives, reducing power consumption as the processor is not running at the maximum clock speed. Using this approach, they were able to achieve a 95.1% accuracy score using a power of  $360\mu\text{W}$ .

The authors of [87] and [169] are the two articles that utilised Edge-AI AAR in horses. In the work of [87], they present the first instance of (as far as we are aware) an Neural Network embedded device for AAR within horses for real-time classification, presenting an embedded ANN to classify three different behaviours, achieving an average accuracy of 80.77%. Finding that reducing the communications required by using on-device classification significantly improves battery life. The authors of [169] compared two different ANNs in two different microcontroller architectures to establish a balance between performance and power

**Table 6**

Summary of Edge-AI and TinyML studies for activity recognition. Q denotes quantization, KD denotes knowledge distillation, and P denotes pruning. Latency and model size units are given in the table headers.

Citation	Model	Classes	Original Acc. (%)	Energy ( $\mu$ J)	Latency (ms)	Size (kB)	Reductions Applied	New Acc. (%)
[145]	MLP	5	93.93	Unk	85	12	None Reported	N/A
[146]	GRU-MLP	3	88.1	Unk	723	45	Q + KD	87.4 (Q), 88.1 (KD)
[137]	DT	4	85.3 (F1)	557	7469	Unk	Q (FP64 $\rightarrow$ INT8)	84.8 (F1)
[138]	GRU	4	95.7	0.216	Unk	2.043	Q (FP64 $\rightarrow$ INT8)	95.2
[139]	RNN	4	95.1	Unk	Unk	N/A	Q (INT8 only)	95.1
[79]	RF	3	78.0	Unk	Unk	1000	None (software only)	N/A
[169]	ANN	3	98.0	1,125,811.2	35,072	500	Q (FP64 $\rightarrow$ INT8)	88.15
[87]	MLP	3	81.1	Unk	Unk	Unk	None reported	80.77
[161]	LR	5	89.59	2037.4	1865	Unk	None reported	87.44
[119]	CNN	3	95.0 (F1)	16,650	438.75	23.3	Q (FP32 $\rightarrow$ FP16) + P	82.0 (F1)
[45]	RF	3	92.0	Unk	50	Unk	None (software only)	N/A
[209]	CNN-LSTM	5	98.95	Unk	495	Unk	Q (FP32 $\rightarrow$ INT8) + P	98.94

consumption. Sensor data from an IMU was collected simultaneously, with the aim of classifying three different gait behaviours within horses with an approximate accuracy of 98%. The authors then applied quantization, moving from 64-bit floating point numbers down to 8-bit integers, seeing an almost 10% reduction in accuracy, dropping to 88.15%.

The authors of [161] aimed to reduce energy consumption using energy aware features, classifying five common behaviours using nine features and a LR model for sheep. The article also performed the same tasks with rhinoceroses, an animal that is out of scope for this review, so we will focus only on the sheep related aspects. The authors originally provided a baseline LR was performed with 89.59% accuracy, which was then reduced to an accuracy of 88.4%, which is reduction of 1.19% while consuming 6.8 times less energy.

The authors of [119] created a CNN that was capable of classifying three behaviours in cattle with an F1-Score of 95%. This CNN was then deployed on a microcontroller through pruning and quantization. After 48 pruning filters were applied, and quantizing from 64-bit floating point numbers to 32-bit floating point numbers, the overall F1-Score of the model dropped to 82%. Interestingly, the authors also reported that further quantizing to 16-bit floating point number did not make any further reductions in performance.

Some articles did not apply any of the model reduction techniques discussed and instead were able to compress their model using software tools. [79] was able to move their trained RF on to a microcontroller using an open source tool called "emlearn" [223] that exports Python trained ML models on to devices with a C99 compiler. The tool is able to provide the relevant C code that houses an RF model, allowing it to run on compatible microcontrollers. The results show an accuracy of 78% on the embedded model, which the authors acknowledge as proof of concept; as the on-device decision makes up one part of the achievement in the article, with the other aspects out of scope for this review. While [45] deployed a RF model upon an ESP32 microcontroller for real-time behavioural classification from a three axis accelerometer, this model performed with an accuracy score of 92.1%. Upon classification, the result was then transmitted by Wi-Fi to be visually presented in a web dashboard. This was achieved with less than two seconds of latency from classification to display on the web dashboard. Interestingly, both of these authors that used software tools created RF models.

And finally, in the case of [209], the authors created a multi-modal system for three tasks, image classification, object detection and behaviour classification. As the image classification and the object detection are performed using methods outside of the scope of this review (computer vision), we will focus on the behaviour classification which used movement sensors. The authors created a CNN and LSTM hybrid model that was capable of classifying five behaviours in cattle. The hybrid model was quantized from 32-bit floating point numbers down to 8-bit integers, and was then pruned, with minimal differences in performance, dropping 0.01% in accuracy score.

### 6.3. Edge-AI and TinyML in AAR discussion and recommendations

The studies summarised in Table 6 demonstrate that Edge-AI and TinyML for Animal Activity Recognition (AAR) is increasingly feasible, though its success depends strongly on model choice, compression strategy, and hardware-aware optimisation. Across the reviewed literature, a clear trend emerges towards lightweight deep learning models and carefully selected classical machine learning methods that balance classification performance with stringent energy, latency, and memory constraints.

Deep learning dominates the reviewed Edge-AI AAR literature, with recurrent neural networks (RNNs), gated recurrent units (GRUs), convolutional neural networks (CNNs), and hybrid CNN-LSTM architectures. These architectures benefit from automatic feature extraction, reducing reliance on handcrafted feature pipelines that are computationally expensive on microcontrollers. Nonetheless, classical tree-based machine learning models such as Decision Trees (DTs) and Random Forests (RFs), remain relevant, like that of non-Edge-AI ML covered in Section 5. RF models appear in multiple works where software tools, such as used an open source tool called "emlearn" [223] as a method to export ML models to microcontrollers using a C99 compiler, albeit often at the cost of larger memory footprints. Google have provided their own API assistive tools for Edge-AI, called LiteRT<sup>6</sup>, which is a runtime for on-device AI and ML applications. LiteRT was formally known as TensorFlow Lite [224] and was used in some of the reviewed articles, such as [138].

Quantization is the most widely applied model reduction technique, appearing in nine of the twelve studies. Most commonly, models are quantized from 64-bit floating point to 8-bit integer representations, resulting in substantial reductions in memory usage and enabling deployment on kilobyte-scale SRAM devices. Across studies, the average reduction in classification performance due to quantization alone is modest, typically below 1%, and in several cases negligible. For example, [146] report a worst-case accuracy reduction of 0.7%, while [138] observe only a 0.5% reduction following quantization. These findings reinforce quantization as a highly effective and low-risk compression technique for AAR. Knowledge distillation is applied less frequently but demonstrates strong benefits when used. In [146], knowledge distillation allows compact GRU-MLP student models to approach the performance of a large ResNet teacher, effectively mitigating the accuracy losses typically associated with aggressive model simplification. Pruning is applied in only two studies, both involving CNN-based architectures. While pruning contributes to reduced computational complexity, its impact on accuracy is more pronounced: in [119], pruning combined with quantization results in an F1-Score reduction of 13%, highlighting the sensitivity of CNN-based AAR models to structural simplification.

Energy and latency reporting remain inconsistent across the literature, limiting direct cross-study comparison. Nevertheless, where

<sup>6</sup> <https://ai.google.dev/edge/litert>

reported, significant insights emerge. Classical ML models and small RNNs tend to achieve the lowest energy consumption, with values ranging from hundreds of microjoules per inference ([137]) down to sub-microjoule levels in optimised RNN implementations ([138]). These results demonstrate that real-time AAR is achievable within tight energy budgets when inference pipelines are hardware-aware. Several studies emphasise the importance of real-time processing strategies in reducing energy consumption. For example, [139] achieve substantial power savings by processing sensor samples as they arrive and dynamically scaling clock frequency to meet real-time constraints, rather than maximising throughput. This highlights that energy efficiency is not solely a function of model size, but also of execution scheduling and system-level design. Latency varies widely across studies, from tens of milliseconds to multiple seconds, reflecting differences in model complexity, windowing strategies, and hardware platforms. Hybrid models deployed on accelerators, such as in [209], achieve latencies as low as 16 ms on specialised hardware. These disparities underscore the importance of matching model complexity to both hardware capabilities and application-level timing requirements.

Taken together, the reviewed studies indicate that TinyML-based AAR is already feasible for a range of livestock monitoring applications, particularly when behaviour sets are limited and models are carefully optimised. Lightweight RNNs and MLPs quantized to 8-bit precision consistently achieve accuracies above 90% with low energy consumption and small memory footprints. More complex architectures, such as CNNs and hybrid models, can achieve higher accuracy but often require pruning, accelerators, or relaxed energy constraints, placing them closer to Edge-AI rather than strict TinyML deployments. Importantly, several works demonstrate that on-device inference can substantially reduce communication energy by minimising wireless data transmission. This shift from continuous data streaming to event-based or summary-based communication represents a key advantage of Edge-AI AAR systems and strengthens their practical viability in long-term, battery-powered deployments.

Overall, quantization emerges as the most reliable and widely applicable compression technique, offering significant memory and energy benefits with minimal performance degradation. Knowledge distillation provides an effective pathway for deploying compact deep learning models without sacrificing accuracy, while pruning must be applied cautiously due to its potential impact on classification performance. The reviewed literature collectively suggests that Edge-AI and TinyML for AAR are no longer aspirational, but increasingly practical, provided that model design, compression, and system-level optimisation are jointly considered.

Finally, we acknowledge a non-AAR article, as this could be an interesting research area and usage of AAR. [65] were able to use Edge-AI to create an ultra-efficient ML device used for fall detection. This coupled accelerometer and gyroscope sensors with a DT algorithm to classify the severity of falls. Using the LoRa communication protocol to notify when a switch from idle to a fall is detected. This could be further researched in AAR livestock to provide farmers with real-time alerts about any endangerment / emergencies concerning livestock. In previous work, this was conceptualised using a mesh network to transfer emergency alerts through a network of collar devices [99]. However, a challenge for this would be gathering the data to train an ML model for this, as standard window lengths used in the literature in Section 5 of three seconds or more would not be sensible here, given a fall can occur in less than a second.

## 7. Conclusions

In this review, we have examined the deployment of Edge-AI for Animal Activity Recognition (AAR), synthesising research on communication technologies, data collection, machine learning, and on-device decision making. Unlike previous reviews, we focus not only on methodological and technical aspects but also on considerations relevant for

real-world deployment. This includes communications constraints, energy efficiency, and the limited but emerging research on Edge-AI and TinyML applications in AAR. Our analysis of AAR literature revealed several patterns. Supervised learning dominates the field, while the split between classical machine learning and deep learning is roughly even.

Most studies collect their own datasets due to limited public availability, with cattle being the most commonly studied animal. In contrast, our review did not identify any publicly available datasets for pigs or poultry. This gap likely reflects several structural challenges, including the difficulty of attaching sensors to smaller animals, the logistical complexity of monitoring large group-housed populations, and economic factors that influence research priorities. Cattle often have higher individual economic value and are managed as identifiable individuals, making them more suitable for sensor-based monitoring studies, whereas pigs and poultry are typically managed at the group level. Developing publicly available benchmark datasets and standardised annotation protocols for these species would help improve cross-study comparability and accelerate progress in AAR research. Data collection varies widely in sensor type, placement, sampling frequency, labelling methods, and environmental conditions. Edge-AI studies are still minimal, but they hold promise for reducing communication overhead and energy use by enabling on-device decision making.

An important integrative question concerns whether different species require unique sampling strategies or whether universal principles can be applied. Across the reviewed studies, sampling frequencies vary by species, with cattle typically sampled at 10-25Hz, while pigs may use lower frequencies due to practical constraints, while chickens may use higher frequencies to capture their behaviours. However, our analysis suggests that minimal frequencies are sufficient to capture the fastest behaviours may serve as a baseline across species, with minor adjustments for species specific movement patterns. Similarly, the complexity of the behaviour set affects classification performance: tasks with two-five behaviours consistently achieve high accuracies of 90% or above, whereas studies with 7-12 behaviours exhibit more variable performance, sometimes dropping below 80%. Notably, high performance can still be achieved for larger behaviour sets when models, feature engineering, and sensor placement are optimised.

The choice of model family also shapes the trade-off between accuracy and deployment. Classical machine learning approaches, such as Random Forests or SVMs, often provide sufficient accuracy while remaining computationally lightweight, making them well-suited for Edge-AI deployment. Deep learning models can offer higher accuracy for complex or multi-class behaviours, but at the cost of increased memory and energy requirements, which may be comparable with the efficiency savings of automatic feature extraction compared to standard ML models. These trade-offs must be carefully considered in the context of device capabilities, power consumption, and on-device inference requirements.

Communications constraints further influence system design. In rural farm settings, connectivity limitations often preclude continuous high-bandwidth data transmission. Long-range, low-power protocols such as LoRa are advantageous for outdoor environments, while short-range, low-power technologies such as Zigbee may suffice indoors. The need to balance data fidelity, update frequency, and energy efficiency shapes feasible system architectures and highlights the importance of on-device decision making to reduce communication demands.

A key trade-off from the reviewed literature concerns the relationship between sampling frequency, communication bandwidth, and energy consumption. Higher sampling frequencies increase the temporal resolution of behaviour detection but also generate substantially larger volumes of sensor data, which in turn increases communication bandwidth requirements and transmission energy costs. In constrained agricultural environments where low-power networks such as LoRa are commonly used, continuously transmitting high-frequency raw sensor data may be impractical. Edge inference therefore provides an important mitigation strategy, allowing devices to process high-frequency data locally and transmit only summarised behavioural outputs or alerts. Designing

effective AAR systems thus requires balancing sampling frequency with communication capabilities and device energy budgets, ensuring that sufficient behavioural detail is captured while maintaining feasible data transmission and long-term battery operation.

While technological feasibility is critical for the development of AAR systems, real-world adoption depends on a broader set of economic, social, and environmental factors. Ultimately, all of the research in this area aims to improve animal wellbeing and assist farmers in making data-driven decisions. However, there are several factors that need to be considered outside of the technical space that will directly influence farmer adoption. First is the cost of the hardware. Some previous research has looked at using cost-effective "off-the-shelf" hardware to keep the cost of the devices minimal [99], as adoption of these AAR technologies may be minimal if the cost of the devices outweighs the benefits they bring to farmers.

Another consideration is the ongoing maintenance of the devices, such as battery replacement, can significantly influence farmer uptake, particularly if frequent intervention is required for recharging or device management. Equipping animals with sensor devices is a labour-intensive process, so if the battery life of the devices only lasts a few days or weeks and requires removal to charge too often, farmers are not likely to adopt such technologies. However, Edge-AI does make a significant contribution to this problem by reducing one of the more power-intensive aspects of communicating a decision instead of the raw sensor data, but nonetheless this does need to be considered in both the ML and the communication aspects. Plus, some of the articles discussed in this review have utilised solar power in an attempt to minimise removals for re-charging [145–147]. Farmer acceptance is also shaped by system reliability, ease of use, and trust in automated decisions; even highly accurate models may fail to be adopted if false positives are frequent or the mode of decision support is not aligned with farm management practices. Effective adoption, therefore, requires not only energy-efficient and computationally capable sensors but also consideration of the human-machine interface and the practical demands of daily farm operations. Together, hardware cost and battery lifespan represent key practical constraints that may ultimately determine whether AAR technologies move beyond research prototypes to sustained on-farm deployment.

Large-scale deployment of battery-powered sensors further raises environmental considerations. Energy consumption, disposal of used batteries, and the associated carbon footprint all contribute to the sustainability profile of AAR technologies. Integrating energy-saving approaches, such as Edge-AI to reduce communication demands, and exploring environmentally friendly battery or energy-harvesting solutions, can mitigate these impacts. Future research should adopt a holistic approach that balances technological performance with economic feasibility, farmer acceptance, and environmental sustainability to ensure that AAR systems are not only technically sound but also practically adoptable and ecologically responsible at scale.

Overall, while significant progress has been made in the development of AAR technologies, several bottlenecks limit their translation from laboratory validation to farm-wide deployment. These include limited public datasets, sparse research on Edge-AI across multiple livestock species, connectivity challenges, and practical constraints around energy, maintenance, and adoption. Another limitation across much of the literature is the limited use of temporal validation strategies. Many studies rely on random train-test splits or cross-validation across mixed time windows, which may overestimate model performance when applied to real-world deployments where behaviour patterns, environmental conditions, and animal populations change over time. Incorporating temporal validation protocols would therefore provide a more realistic assessment of model robustness and long-term deployment reliability.

At present, most AAR systems incorporating Edge-AI remain primarily at the experimental or early research stage rather than being widely deployed as commercially mature technologies. While prototype systems demonstrate promising accuracy and energy efficiency in controlled trials, large-scale deployment across commercial farms remains limited due to challenges including device cost, long-term energy management, dataset availability, and system reliability in diverse real-world environments. Continued progress in low-power hardware, robust datasets, and scalable communication infrastructure will be necessary before Edge-AI based AAR systems can achieve widespread commercial adoption. Addressing these challenges requires a multidisciplinary approach that integrates technical performance with economic feasibility, farmer usability, and environmental responsibility, ensuring that AAR systems are not only accurate and efficient but also adoptable and sustainable at scale.

#### CRediT authorship contribution statement

**Bradley Patrick:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation; **Eiman Kanjo:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology; **Omprakash Kaiwartya:** Writing – review & editing, Writing – original draft, Supervision.

#### Data availability

No data was used for the research described in the article.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Appendix A. Appendix

**Table A.7**  
Summary of articles collecting own data for AAR in cattle.

Citation	Sensors	Freq (Hz)	Animals	behaviours	Placement	Environment	Labelling	Activity Types	Duration / Samples
[215]	IMU	10	5	12	Back	Barn	Video	Walking, Standing, Lying, Stationary	10 hours/day over 12 days
[147]	Acc	50	4	10	Neck (top)	Field	Visual obs	Grazing, Ruminating, Resting, Other	6660 datapoints
[55]	Acc x2	50	5	51	Neck (top), Ear	Unk	Video	Grazing, Walking, Ruminating, Resting, Other	654,435 samples
[195]	Acc	10	3	10	Neck (top)	Barn	Visual obs + supporting device	Feeding, Ruminating, Other	3600 samples
[128]	Acc x2	1	3	16	Neck (side), Leg	Barn	Visual obs	Lying, Standing, Feeding	5760 samples
[56]	Acc x2	25	3	21	Neck, Leg	Barn	Visual obs	Feeding, Ruminating, Other	809 hours
[131]	Acc	12	2	45	Neck	Field	Visual obs	Grazing, Non-grazing	93,312,000 samples
[121]	Acc	10	4	3	Neck (side)	Pasture	Video	Grazing, Lying, Ruminating, Standing	5.5 hours
[126]	Acc	0.5, 1	5	33	Neck (under)	Indoors	Video	Fighting, Grazing, Resting, Ruminating, Walking	17,173 samples
[129]	Acc	1	2	6	Halter	Field	Visual obs	Grazing, Non-grazing	~30k samples
[210]	Acc	10	4	18	Neck (side)	Barn	Video	Grazing, Standing, Walking, Ruminating	136,630 samples
[133]	Acc	25	5	21	Neck	Indoors	Video	Lying, Walking, Eating, Running, Drinking milk	9.5 hours
[225]	Acc	Unk	4	5	Neck, Leg	Unk	Unk	Feeding, Lying, Standing, Walking	15,165 observations
[70]	Acc	10	3	7	Nose ring	Indoors	Video	Feeding, Ruminating, Other	1,252,522 samples
[123]	Acc + Gyro	12.5	2	3	Neck	Barn	Unk	Rumination, Non-rumination	2424 samples
[148]	IMU	50	9	10	Neck (top)	Unk	Visual obs	Grazing, Walking, Ruminating standing, Ruminating lying, Standing, Lying, Drinking, Grooming, Other	69,150 seconds
[132]	Acc	12	2	21	Halter	Paddock (indoor)	Visual obs	Suckling, Non-suckling	Unk
[119]	Acc	10	3	18	Neck (side)	Barn	Unk	Ruminating, Eating, Other	3460 hours
[59]	Acc	20	8	6	Neck (top)	Barn	Video	Feeding, Lying, Ruminating (lying), Ruminating (standing), Licking salt, Moving, Social licking, Head butt	364,544 rows
[60]	IMU	20	7	3	Neck (top)	Barn	Video	Feeding, Ruminating (lying), Ruminating (standing), Lying/Standing (72-24 h pre-calving), Lying/Standing (24 h pre-calving)	~1.4M samples
[226]	Acc x3	12	6	Unk	Neck (side), Ear, Halter	Unk	Video	Grazing, Resting, Walking, Standing, Ruminating, Other	Unk
[124]	Acc	59.5	4	10	Neck (side)	Pasture	Visual obs	Grazing, Walking, Standing, Lying	Unk
[125]	Acc	59.5	6	26	Neck (side)	Field	Visual obs	Grazing, Walking, Ruminating, Lying, Standing	57 h 22 min
[130]	IMU	10	12	35	Neck (under)	Barn	Video	Lying, Lying down, Lying bout, Standing, Walking, Grazing, Feeding, Chewing ruminating, Drinking, Other	1,526,271 samples
[192]	Acc	5	3	5	Halter	Barn	Video	Feeding, Ruminating, Other	Unk
[61]	IMU x2	10	6	Unk	Neck (side), Leg	Barn	Video	Standing, Lying, Walking, Resting, Feeding, Ruminating	Unk
[203]	Acc	25	3	4	Ear	Paddock	Video	Licking, Non-licking, Eating	23,037 (3 s windows)
[113]	Acc	20	6	10	Neck (side)	Field	Visual obs	Eating, Rumination, Lying, Standing, Walking, Drinking	Unk
[227]	Acc x2	10	4	5	Neck (under)	Unk	Visual obs	Walking, Feeding, Lying, Standing	12,195 samples
[72]	Acc	10	4	5	Neck (under), Leg	Unk	Visual + Video	Feeding, Lying, Standing, Walking	30.5k samples
[112]	Acc	10	10	38	Neck (under)	Field	Video	Allogrooming, Calf suckle, Foraging high, Foraging low, Lying, Ruminating (lying), Ruminating (standing), Standing, Vigilance, Walking	6.4M lines
[127]	Acc	1	7	5	Leg	Barn	Video	Feeding, Lying, Standing, Lying down, Standing up, Walking normally, Active walking	25,921 samples
[118]	Acc x2	0.167	3	12	Neck, Halter	Indoors	Unk	Eating, Drinking, Other	1.2M samples
[117]	IMU	10	6	12	Neck (top)	Indoors	Video	Feeding, Lying, Ruminating (lying), Rub itching leg, Social licking, Rub itching neck	300,000 samples

**Table A.8**  
Summary of articles collecting own data for AAR in sheep and goats.

Citation	Sensors	Freq (Hz)	Total Animals	behaviours	Placement	Environment	Labelling	Activity Types	Duration / Samples
[154]	Acc x3	12	4	5	Ear, Neck (under), Leg	Field	Video	Grazing, Walking, Standing, Lying	2561 samples
[163]	Acc	25	5	59	Horn	Outdoor paddock	Video	Grazing, Displacement, Ruminating, Resting, Other	144 hours
[69]	Acc	62.5	3	8	Halter	Field	Video	Grazing, Ruminating, Other	69,975 seconds
[196]	Acc	40	9	9	Neck (under)	Indoors	Video	Standing, Resting, Eating, Grooming, Shaking, Walking, Trotting, Running, Climbing	516 mins
[58]	Acc + Gyro	20	5	3	Neck	Unk	Video	Eating, Walking, Standing, Lying, Ruminating	446,043 samples
[228]	Acc + Gyro	10	4	18	Neck (side)	Indoors	Video	Movement, Lying, Feeding, Ruminating	7500 datasets
[15]	Acc + Gyro	1	2	5	Neck (side)	Unk	Unk	Grazing, Non-grazing	288,800 samples
[162]	IMU	20	2	3	Neck (under)	Field	Video	Grazing, Non-grazing	22.5 hours
[201]	Acc	30	4	17	Neck	Pasture	Visual obs	Grazing, Ruminating, Walking, Standing	1,052,475 samples
[44]	IMU	20	7	22	Halter, Neck, Legs	Outdoors	Video	Walking, Standing, Grazing, Lying, Ruminating, Running, Drinking	~40k samples
[71]	Acc + Gyro	100	6	5	Neck	Zoo (indoor)	Video	Stationary, Walking, Trotting, Running, Eating, Other	177.6 hours
[11]	Acc	12.5	4	8	Neck	Unk	Video	Grazing, Scratching, Walking, Inactive	28 hours
[12]	Acc x2	12.5	3	9	Neck (top), Neck (under)	Pasture	Unk	Grazing, Active, Inactive	2,925,000 samples
[161]	Acc	100	5	-	Neck	Unk	Visual obs	Lying, Standing, Walking, Running, Grazing	16 hours
[229]	Acc	Unk	4	Unk	Back	Indoors	Unk	Lying, Standing, Walking, Running	Unk
[230]	Acc	40	6	20	Rear / back	Field	Video	Foraging, Walking, Running, Standing, Lying, Urinating	335 mins
[151]	Acc + Gyro	16	3	6	Neck (side), Ear	Field	Video	Grazing, Ruminating, Non-eating	27,317 samples
[158]	Acc	5	4	8	Ear	Indoors	Video	Feeding, Movement, Social, Other	11 hours per goat
[155]	Acc	50	5	116	Neck (under)	Field	Visual obs	Lying, Standing, Grazing, Ruminating, Inactive, Walking, Suckling	42,935 windows
[153]	IMU	100	3	3	Neck (top)	Pasture	Video	Lying, Standing, Grazing	119,148 samples
[45]	Acc	10	3	1	Neck (under)	Unk	Visual obs	Standing, Lying, Sleeping	60,000 samples
[160]	Acc x2	30 (halter), 25 (ear)	5	30	Halter, Ear	Field	Video	Sitting, Standing, Walking, Grazing, Ruminating	28,988 samples
[157]	Acc + Gyro	16	3	17	Ear	Field	Visual obs	Walking, Standing, Lying	Unk
[159]	Acc + Gyro	8, 16, 32	3	6	Neck, Ear	Field	Video	Walking, Standing, Lying	Unk
[156]	IMU	20	5	18	Neck (side)	Indoors	Video	Feeding, Ruminating, Mating, Movement, Running	779,506 samples

**Table A.9**  
Summary of articles collecting own data for AAR in horses.

Citation	Sensors	Freq (Hz)	Total Animals	behaviours	Placement	Environment	Labelling	Activity Types	Duration / Samples
[57]	Acc	33	3	2	Saddle	Outdoors	Visual obs	Walk, Trot, Canter	90,000 samples
[172]	Acc + Gyro	2	3	6	Neck (under)	Outdoors	Visual obs	High, Mid, Low	61,649 samples
[169]	IMU	33	3	2	Halter	Unk	Unk	Motionless, Trotting, Walking	25,200 samples
[13]	Acc	25, 50, 100, 200	7	6	Legs	Outdoors	Video	Standing, Walking, Trotting, Canter, Rolling, Pawing, Flank watching	2,012,875 samples
[14]	Acc	50	7	6	Legs	Outdoors	Video	Standing, Walking, Trotting, Canter, Rolling, Pawing, Flank watching	2238 (2 s windows)
[87]	IMU	30	3	Unk	Halter	Unk	Unk	Trotting, Motionless, Walking	30,000 samples
[170]	Acc	10	2	4	Halter	Outdoors	Video	Grazing, Non-grazing	230,286 samples
[171]	Acc + Gyro	100	15	10	Neck (side)	Outdoors	Video	Standing, Walking, Trotting, Galloping, Sniffing, Eating, Chewing, Drinking, Shaking, Rolling	283,200 (over 60 hours)

**Table A.10**  
Summary of articles collecting own data for AAR in chickens.

Citation	Sensors	Freq (Hz)	Total Animals	behaviours	Placement	Environment	Labelling	Activity Types	Duration / Samples
[180]	IMU	100	3	2	Back	Indoors	Video	Low intensity, Moderate intensity, High intensity	3.5 hours
[179]	Acc + Gyro	1000	12	4	Back	Indoors	Video	Moving, Eating, Drinking, Preening, Body shaking, Head scratching, Tail flapping, Stopping, Resting, Dust bathing, Litter scratching, Other	1766 samples
[18]	IMU	5	2	5	Neck (under)	Indoors	Video	Eating, Drinking	Unk
[182]	Acc	100	3	10	Back	Indoors	Video	Pecking, Preening, Dust bathing	Unk
[184]	Acc	100	3	42	Back	Indoors	Video	Sit, Stand, Walk	Unk
[183]	Acc	40	4	12	Back	Indoors	Visual obs	Resting, Feeding, Drinking, Walking	Unk
[19]	Acc	40	6	30	Back	Indoors	Visual obs	Sitting, Standing, Feeding, Drinking, Walking, Preening	1418 mins
[20]	IMU	1000	3	2	Back	Indoors	Video	Low, Medium, High intensity	1,004,950 samples
[181]	Acc	40	4	9	Back	Indoors	Video	Walking, Resting, Feeding, Drinking	261 mins
[185]	Acc	Unk	4	5	Neck (under)	Indoors	Video	Standing, Walking, Pecking, Drinking	216k samples

**Table A.11**  
Summary of articles collecting own data for AAR in pigs.

Citation	Sensors	Freq (Hz)	Total Animals	behaviours	Placement	Environment	Labelling	Activity Types	Duration / Samples
[54]	Acc + Gyro	10	7	4	Neck (top)	Indoors	Video	Eating, Lying, Walking, Standing, Playing, Drinking, Sitting, Unknown	13 hours
[22]	Acc	20	4	3	Neck (top)	Outdoors	Video	Lying, Standing, Walking, Exploring	130k samples
[187]	Acc	12.5	2 + 8	97	Ear	Indoors	Video	High activity, Low activity + Abnormal, Agnostic, Conform, Exploration, Feeding, Locomotion, Play, Resting	185 24h periods
[21]	Acc + Gyro	20	6	2	Neck	Indoors	Video	Sleep, Nursing, Lying, Eating, Drinking, Movement	3600 samples
[186]	Acc + Gyro	Unk	5	2	Neck (under)	Indoors	Video	Lying, Standing, Walking, Drinking, Feeding	2543 samples
[46]	IMU	10	4	Unk	Neck (top)	Indoors	Video	Standing, Lying, Feeding, Body movement	54,939 entries

## References

- [1] D. Berckmans, General introduction to precision livestock farming, *Animal Front.* 7 (1) (2017) 6–11. <https://doi.org/10.2527/af.2017.0102>
- [2] A. Eerdeken, M. Deruyck, J. Fontaine, L. Martens, E. De Poorter, D. Plets, W. Joseph, A framework for energy-efficient equine activity recognition with leg accelerometers, *Comput. Electron. Agric.* 183 (2021). <https://doi.org/10.1016/j.compag.2021.106620>
- [3] D.F. Cabrera-Castellanos, A. Aragón-Zavala, G. Castañón-ávila, Closing connectivity gap: an overview of mobile coverage solutions for not-spots in rural zones, 2021. <https://doi.org/10.3390/s21238037>
- [4] L. Aldhaheri, N. Alshehhi, L.L.J. Manzil, R.A. Khalil, S. Javaid, N. Saeed, M.-S. Alouini, LoRa communication for agriculture 4.0: opportunities, challenges, and future directions (2024).
- [5] A. Mao, E. Huang, X. Wang, K. Liu, Deep learning-based animal activity recognition with wearable sensors: overview, challenges, and future directions, 2023. <https://doi.org/10.1016/j.compag.2023.108043>
- [6] A. Kapoor, A. Kumar, D.P. Singh, A.K. Karn, K.K. Gola, P. Suyal, Advancements in animal behaviour monitoring and livestock management: a review, in: *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 1206–1212. <https://doi.org/10.1109/IC3I61595.2024.10828621>
- [7] W. Shen, C. Chen, S. Zheng, S. He, M. Li, The design of system about cow activity based on SVM, *Int. J. Smart Home* 9 (3) (2015) 91–100. <https://doi.org/10.14257/ijsh.2015.9.3.09>
- [8] C. Arcidiacono, S.M.C. Porto, M. Mancino, G. Cascone, Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data, *Comput. Electron. Agric.* 134 (2017) 124–134. <https://doi.org/10.1016/j.compag.2017.01.021>
- [9] K.M. Abell, M.E. Theurer, R.L. Larson, B.J. White, D.K. Hardin, R.F. Randle, Predicting bull behavior events in a multiple-sire pasture with video analysis, accelerometers, and classification algorithms, *Comput. Electron. Agric.* 136 (2017) 221–227. <https://doi.org/10.1016/j.compag.2017.01.030>
- [10] V. Giovanetti, M. Decandia, G. Molle, M. Acciaro, M. Mameli, A. Cabiddu, R. Cossu, M.G. Serra, C. Manca, S.P.G. Rassu, C. Dimauro, Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer, *Livest. Sci.* 196 (2017) 42–48. <https://doi.org/10.1016/j.livsci.2016.12.011>
- [11] N. Kleanthous, A. Hussain, W. Khan, J. Sneddon, A. Mason, Feature extraction and random forest to identify sheep behavior from accelerometer data, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12465 LNAI, Springer Science and Business Media Deutschland GmbH, 2020, pp. 408–419. [https://doi.org/10.1007/978-3-030-60796-8\\_35](https://doi.org/10.1007/978-3-030-60796-8_35)
- [12] N. Kleanthous, A. Hussain, W. Khan, J. Sneddon, P. Liatsis, Deep transfer learning in sheep activity recognition using accelerometer data, *Expert Syst. Appl.* 207 (2022). <https://doi.org/10.1016/j.eswa.2022.117925>
- [13] A. Eerdeken, M. Deruyck, J. Fontaine, L. Martens, E. De Poorter, W. Joseph, Automatic equine activity detection by convolutional neural networks using accelerometer data, *Comput. Electron. Agric.* 168 (2020). <https://doi.org/10.1016/j.compag.2019.105139>
- [14] A. Eerdeken, M. Deruyck, J. Fontaine, L. Martens, E.D. Poorter, D. Plets, W. Joseph, Resampling and data augmentation for equines' behaviour classification based on wearable sensor accelerometer data using a convolutional neural network, in: *2020 International Conference on Omni-Layer Intelligent Systems (COINS)*, 2020, pp. 1–6. <https://doi.org/10.1109/COINS49042.2020.9191639>
- [15] A.L.C. Garces, G.M.V. Obusan, J.A.R. Ong, S.C. Perez, M.I. Alipio, G.P. Mappatao, Wi-Fi based monitoring of goat feeding behaviour using gyro- and accelerometer data, in: *International Conference on Advanced Technologies for Communications, IEEE Computer Society*, 2023, pp. 350–355. <https://doi.org/10.1109/ATC58710.2023.10318900>
- [16] A. Hollevoet, T. De Waele, D. Peralta, F. Tuytens, E. De Poorter, A. Shahid, Goats on the move: evaluating machine learning models for goat activity analysis using accelerometer data, *Animals* 14 (13) (2024). <https://doi.org/10.3390/ani14131977>
- [17] S. Antonio, P. Lama, R. Gonzalez, C. Catania, S. Paez Lama, R. Leal, P. General, S. Martin, A. Mendoza, Behavior classification of a grazing goat in the Argentine monte desert by using inertial sensors (2019). <https://doi.org/10.13140/RG.2.2.13611.18726>
- [18] L. Li, M. Di, H. Xue, Z. Zhou, Z. Wang, Feature selection model based on IWOA for behavior identification of chicken, *Sensors* 22 (16) (2022). <https://doi.org/10.3390/s22166147>
- [19] M. Sakib, S.K.S. Alam, S.H.M. Atul, F.H. Supto, M.M. Hossain, Automated detection of broiler chicken behaviors through the integration of 3D accelerometer sensor and machine learning techniques, in: *AIP Conference Proceedings*, 3245, American Institute of Physics, 2024. <https://doi.org/10.1063/5.0231888>
- [20] M. Shahbazi, K. Mohammadi, S.M. Derakhshani, P.W.G. Groot Koerkamp, Deep learning for laying hen activity recognition using wearable sensors, *Agriculture (Switzerland)* 13 (3) (2023). <https://doi.org/10.3390/agriculture13030738>
- [21] Z. Pan, H. Chen, W. Zhong, A. Wang, C. Zheng, A CNN-Based animal behavior recognition algorithm for wearable devices, *IEEE Sens. J.* 23 (5) (2023) 5156–5164. <https://doi.org/10.1109/JSEN.2023.3239015>
- [22] M. Jin, C. Wang, D.B. Jensen, Effect of de-noising by wavelet filtering and data augmentation by borderline SMOTE on the classification of imbalanced datasets of pig behavior, *Front. Anim. Sci.* 2 (2021). <https://doi.org/10.3389/fanim.2021.666855>

- [23] S. Neethirajan, AI in sustainable pig farming: IoT insights into stress and gait, *Agriculture (Switzerland)* 13 (9) (2023). <https://doi.org/10.3390/agriculture13091706>
- [24] P. Kasnesis, V. Doulerakis, D. Uzunidis, D.G. Kogias, S.I. Funcia, M.B. González, C. Giannousis, C.Z. Patrikakis, Deep learning empowered wearable-based behavior recognition for search and rescue dogs, *Sensors* 22 (3) (2022). <https://doi.org/10.3390/s22030993>
- [25] A. Hussain, S. Ali, Abdullah, H.C. Kim, Activity detection for the wellbeing of dogs using wearable sensors based on deep learning, *IEEE Access* 10 (2022) 53153–53163. <https://doi.org/10.1109/ACCESS.2022.3174813>
- [26] P. Kumpulainen, A.V. Cardó, S. Somppi, H. Törnqvist, H. Väättäjä, P. Majoranta, Y. Gizatdinova, C.H. Antink, V. Surakka, M.V. Kujala, O. Vainio, A. Vehkaoja, Dog behaviour classification with movement sensors placed on the harness and the collar (2021). <https://doi.org/10.17632/vxhx934tbn.1>
- [27] A. Hussain, S. Ali, M.I. Joo, H.C. Kim, A deep learning approach for detecting and classifying cat activity to monitor and improve cat's well-being using accelerometer, gyroscope, and magnetometer, *IEEE Sens. J.* 24 (2) (2024) 1996–2008. <https://doi.org/10.1109/JSEN.2023.3324665>
- [28] M.A. Islam Mozumder, R.I. Sumon, S.M. Imtiyaz Uddin, H.C. Kim, Domesticated cat activity detection using deep learning technique, in: *International Conference on Electrical, Computer, and Energy Technologies, ICECET 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. <https://doi.org/10.1109/ICECET61485.2024.10698381>
- [29] M.A.I. Mozumder, T.P. Theodore Armand, R.I. Sumon, S.M. Imtiyaz Uddin, H.C. Kim, Automated pipeline for robust cat activity detection based on deep learning and wearable sensor data, *Sensors* 24 (23) (2024). <https://doi.org/10.3390/s24237436>
- [30] M. El Jarroudi, L. Kouadio, P. Delfosse, C.H. Bock, A.K. Mahlein, X. Fettweis, B. Mercatoris, F. Adams, J.M. Lenné, S. Hamdioui, Leveraging edge artificial intelligence for sustainable agriculture, *Nat. Sustain.* 7 (7) (2024) 846–854. <https://doi.org/10.1038/s41893-024-01352-4>
- [31] L. Riaboff, L. Shalloo, A.F. Smeaton, S. Couvreur, A. Madouasse, M.T. Keane, Predicting livestock behaviour using accelerometers: a systematic review of processing techniques for ruminant behaviour prediction from raw accelerometer data, 2022. <https://doi.org/10.1016/j.compag.2021.106610>
- [32] S.J. Schneidewind, M.R. Al Merestani, S. Schmidt, T. Schmidt, C. Thöne-Reineke, M. Wiegand, Rumination detection in sheep: a systematic review of sensor-based approaches, 2023. <https://doi.org/10.3390/ani13243756>
- [33] A. Rohan, M.S. Razaq, M.J. Hasan, F. Asghar, A.K. Bashir, T. Dottorini, Application of deep learning for livestock behaviour recognition: a systematic literature review, 2024. <https://doi.org/10.1016/j.compag.2024.109115>
- [34] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Ak, S.E. Brennan, R. Chou, J. Glanville, J.M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, *BMJ* 372 (2021). <https://doi.org/10.1136/bmj.n71>
- [35] N.R. Haddaway, M.J. Page, C.C. Pritchard, L.A. McGuinness, PRISMA2020: an r package and shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis, *Campbell Syst. Rev.* 18 (2) (2022) e1230. <https://doi.org/https://doi.org/10.1002/cl2.1230>
- [36] E. Mozaffarihar, F. Theoleyre, M. Menth, A survey of wi-Fi 6: technologies, advances, and challenges, *Future Internet* 14 (10) (2022). <https://doi.org/10.3390/fi14100293>
- [37] Y.P. Pratama, D. Kurnia Basuki, S. Sukaridhoto, A.A. Yusuf, H. Yulianus, F. Faruq, F.B. Putra, Designing of a smart collar for dairy cow behavior monitoring with application monitoring in microservices and internet of things-based systems, in: *2019 International Electronics Symposium (IES)*, 2019, pp. 527–533. <https://doi.org/10.1109/ELECSYM.2019.8901676>
- [38] T.A. Wild, M. Wikelski, S. Tyndel, G. Alarcón-Nieto, B.C. Klump, L.M. Aplin, M. Meboldt, H.J. Williams, Internet on animals: wi-fi-enabled devices provide a solution for big data transmission in biologging, *Methods Ecol. Evol.* 14 (1) (2023) 87–102. <https://doi.org/10.1111/2041-210X.13798>
- [39] A.R. Reigones, P.D. Gaspar, Real-Time vital signs monitoring system towards livestock health furtherance, in: *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 753–758. <https://doi.org/10.1109/ICICT50816.2021.9358658>
- [40] J.K. Park, E.Y. Park, Monitoring method of movement of grazing cows using cloud-based system, *ECTI Trans. Comput. Inf. Technol.* 15 (1) (2021) 24–33. <https://doi.org/10.37936/ecti-cit.2021151.240087>
- [41] W.D. Chi-Pérez, J.A. Ríos-Martínez, F.A. Madera-Ramírez, J.J. Estrada-López, Wearable system for intelligent monitoring of assistance and rescue dogs, in: *Journal of Physics: Conference Series*, 2699, Institute of Physics, 2024. <https://doi.org/10.1088/1742-6596/2699/1/012001>
- [42] G. Ramesh, K. Sivaraman, V. Subramani, P.Y. Vignesh, S.V.V. Bhogachari, Farm animal location tracking system using arduino and GPS module, in: *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. <https://doi.org/10.1109/ICCCI50826.2021.9402610>
- [43] F. Anthony, M. Gabriel, B. Shao, Open source cellular technologies for cost effective cellular connectivity in rural areas, *Int. J. Comput. Appl.* 146 (15) (2016) 29–33. <https://doi.org/10.5120/ijca.20161910894>
- [44] Z. Jin, H. Shu, T. Hu, C. Jiang, R. Yan, J. Qi, W. Wang, L. Guo, Behavior classification and spatiotemporal analysis of grazing sheep using deep learning, *Comput. Electron. Agric.* 220 (2024). <https://doi.org/10.1016/j.compag.2024.108894>
- [45] A. Fahri Suhaimi, G.A. Mutiara, M. Fahru Rizal, Leveraging embedded accelerometers and machine learning for real-time sheep behavior classification in precision farming, *IEEE Access* 13 (2025) 165006–165024. <https://doi.org/10.1109/ACCESS.2025.3612098>
- [46] X. Zhu, L. Wei, T. Zhou, J. Zhang, C. Qian, X. Luo, A hybrid LSTM-CNN algorithm for swine posture recognition using multimodal sensor data, in: *Lecture Notes in Computer Science*, 15951 LNCS, Springer Science and Business Media Deutschland GmbH, 2026, pp. 203–211. [https://doi.org/10.1007/978-981-95-1233-1\\_19](https://doi.org/10.1007/978-981-95-1233-1_19)
- [47] B. Stiller, E. Schiller, C. Schmitt, An overview of network communication technologies for IoT, *Technical Report*, 2020.
- [48] B. Badihi, M.U. Sheikh, K. Ruttik, R. Jäntti, On performance evaluation of BLE 5 in indoor environment: an experimental study, in: *2020 Thinspace IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020, pp. 1–5. <https://doi.org/10.1109/PIMRC48278.2020.9217132>
- [49] V. Hartová, J. Hart, Livestock monitoring system using bluetooth technology, *Technical Report* 3, 2017.
- [50] J. Trogh, D. Plets, L. Martens, W. Joseph, Bluetooth low energy based location tracking for livestock monitoring, in: *8th European Conference on Precision Livestock Farming*, 2017.
- [51] M. Nikodem, Bluetooth low energy livestock positioning for smart farming applications, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12745 LNCS, Springer Science and Business Media Deutschland GmbH, 2021, pp. 55–67. [https://doi.org/10.1007/978-3-030-77970-2\\_5](https://doi.org/10.1007/978-3-030-77970-2_5)
- [52] J. Makario, C.w. Maina, A bluetooth low energy (BLE) based system for livestock tracking and localization, in: *2021 IST-Africa Conference (IST-Africa)*, 2021, pp. 1–7.
- [53] F. Maroto-Molina, J. Navarro-García, K. Príncipe-Aguirre, I. Gómez-Maqueda, J.E. Guerrero-Ginel, A. Garrido-Varo, D.C. Pérez-Marín, A low-cost IOT-based system to monitor the location of a whole herd, *Sensors (Switzerland)* 19 (10) (2019). <https://doi.org/10.3390/s19102298>
- [54] S. Alghamdi, Z. Zhao, D.S. Ha, G. Morota, S.S. Ha, Improved pig behavior analysis by optimizing window sizes for individual behaviors on acceleration and angular velocity data, *J. Anim. Sci.* 100 (11) (2022). <https://doi.org/10.1093/jas/skac293>
- [55] R. Arablouei, G.J. Bishop-Hurley, N. Bagnall, A. Ingham, Cattle behavior recognition from accelerometer data: leveraging in-situ cross-device model learning, *Comput. Electron. Agric.* 227 (2024). <https://doi.org/10.1016/j.compag.2024.109546>
- [56] V. Bloch, L. Frondelius, C. Arcidiacono, M. Mancino, M. Pastell, Development and analysis of a CNN- and transfer-Learning-Based classification model for automated dairy cow feeding behavior recognition from accelerometer data, *Sensors* 23 (5) (2023). <https://doi.org/10.3390/s23052611>
- [57] E. Casella, A.R. Khamesi, S. Silvestri, Smartwatch application for horse gaits activity recognition, in: *Proceedings - 2019 Hinspace IEEE International Conference on Smart Computing, SMARTCOMP 2019*, Institute of Electrical and Electronics Engineers Inc., 2019, pp. 409–416. <https://doi.org/10.1109/SMARTCOMP.2019.00080>
- [58] F. Pereira, L. Fonseca, D. Corujo, W. Xavier, P. Gonçalves, On the development of a wearable animal monitor (2022). <https://doi.org/10.3390/ani>
- [59] Y. Peng, N. Kondo, T. Fujiura, T. Suzuki, Wulandari, H. Yoshioka, E. Itoyama, Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units, *Comput. Electron. Agric.* 157 (2019) 247–253. <https://doi.org/10.1016/j.compag.2018.12.023>
- [60] Y. Peng, N. Kondo, T. Fujiura, T. Suzuki, S. Ouma, Wulandari, H. Yoshioka, E. Itoyama, Dam behavior patterns in Japanese black beef cattle prior to calving: automated detection using LSTM-RNN, *Comput. Electron. Agric.* 169 (2020). <https://doi.org/10.1016/j.compag.2019.105178>
- [61] Z. Shi, Z. Zhang, Y. Jia, J. Li, X. Wang, Y. Qiu, J. Miao, F. Chang, X. Han, W. Tang, Internet-of-Things behavior monitoring system based on wearable inertial sensors for classifying dairy cattle health using machine learning, in: *5th IEEE International Conference on Artificial Intelligence in Engineering and Technology, ICAIET 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 277–282. <https://doi.org/10.1109/ICAIET59451.2023.10291766>
- [62] C. Bouras, A. Gkamas, V. Kokkinos, N. Papachristos, Time difference of arrival localization study for SAR systems over LoRAWAN, in: *Procedia Computer Science*, 175, Elsevier B.V., 2020, pp. 292–299. <https://doi.org/10.1016/j.procs.2020.07.043>
- [63] X. Fu, W. Shen, Y. Yin, Y. Zhang, S. Yan, S. Kou, T. Qu, M. Jacqueline, Remote monitoring system for livestock environmental information based on lora wireless ad hoc network technology, *Int. J. Agric. Biol. Eng.* 15 (4) (2022) 79–89. <https://doi.org/10.25165/j.ijabe.20221504.6708>
- [64] B. Kumkhet, N. Pukrongta, N. Wichaipanich, P. Sangmahamad, T. Pechrkool, V. Pirajanchai, P. Sangpet, U. Issara, A. Taparuggsanagorn, Low-Cost LoRa-Based system for continuous GHGs monitoring in cattle: enhancing agricultural sustainability, *IEEE Access* (2025). <https://doi.org/10.1109/ACCESS.2025.3532471>
- [65] H. Koutsourelakis, G. Komaros, M. Coppola, Energy efficient deep-Edge computing through hardware machine learning, in: *2023 Thinspace IEEE World Forum on Internet of Things: The Blue Planet: A Marriage of Sea and Space, WF-IoT 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. <https://doi.org/10.1109/WF-IoT58464.2023.10539522>
- [66] B.R. Dos Reis, Z. Easton, R.R. White, D. Fuka, A lora sensor network for monitoring pastured livestock location and activity, *Transl. Anim. Sci.* 5 (2) (2021) 1–9. <https://doi.org/10.1093/tas/txab010>
- [67] A. Pagano, D. Croce, I. Tinnirello, G. Vitale, A survey on lora for smart agriculture: current trends and future perspectives, *IEEE Internet Things J.* 10 (4) (2023) 3664–3679. <https://doi.org/10.1109/JIOT.2022.3230505>

- [68] L. Schulthess, F. Longchamp, C. Vogt, M. Magno, A LoRa-Based and maintenance-free cattle monitoring system for alpine pastures and remote locations, in: EN-Ssys 2023 - Proceedings of the 2023 11th International Workshop on Energy Harvesting and Energy-Neutral Sensing Systems, Part of: Sensys 2023, Association for Computing Machinery, Inc, 2023, pp. 44–50. <https://doi.org/10.1145/3628353.3628549>
- [69] M. Decandia, V. Giovanetti, G. Molle, M. Acciari, 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, *Comput. Electron. Agric.* 154 (2018) 112–119. <https://doi.org/10.1016/j.compag.2018.09.002>
- [70] W. Feng, D. Fan, H. Wu, W. Yuan, Cow behavior recognition based on wearable nose rings, *Animals* 14 (8) (2024). <https://doi.org/10.3390/ani14081187>
- [71] J.W. Kamminga, D.V. Le, J.P. Meijers, H. Bisby, N. Meratnia, P.J.M. Havinga, Robust sensor-Orientation-Independent feature selection for animal activity recognition on collar tags, *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* 2 (1) (2018) 1–27. <https://doi.org/10.1145/3191747>
- [72] D.N. Tran, D.V. Manh, P. Van Thanh, A.A. Shankar, K. Joshi, D.T. Tran, Real-time behavior recognition of animal: an IoT-based system design using acceleration data, *Multimed Tools Appl.* (2024). <https://doi.org/10.1007/s11042-024-20309-5>
- [73] A. Llaría, G. Terrasson, H. Arregui, A. Hacala, Geolocation and monitoring platform for extensive farming in mountain pastures, *Proc. IEEE Int. Conf. Ind. Technol.* 2015-June (June) (2015) 2420–2425. <https://doi.org/10.1109/ICIT.2015.7125454>
- [74] D. Mancuso, G. Castagnolo, M.C.M. Parlato, F. Valenti, S.M.C. Porto, Low-power networks and GIS analyses for monitoring the site use of grazing cattle, *Comput. Electron. Agric.* 210 (2023). <https://doi.org/10.1016/j.compag.2023.107897>
- [75] M. Grane, J.D. Martinez, A. Arnaud, R. Puyol, M. Miguez, A sensor network using SigFox for temperature and humidity monitoring in the livestock industry, in: LASCAS 2024 - 15th IEEE Latin American Symposium on Circuits and Systems, Proceedings, Institute of Electrical and Electronics Engineers Inc., 2024. <https://doi.org/10.1109/LASCAS60203.2024.10506134>
- [76] R.S. Sinha, Y. Wei, S.H. Hwang, A survey on LPWA technology: LoRa and NB-IoT, 2017. <https://doi.org/10.1016/j.ict.2017.03.004>
- [77] H.S. Bhaskaran, M. Gordon, S. Neethirajan, Development of a cloud-based IoT system for livestock health monitoring using AWS and python, *Smart Agric. Technol.* 9 (2024). <https://doi.org/10.1016/j.atech.2024.100524>
- [78] A. Sutanto, A. Rakhman, I. Afriliana, R. Hernowo, W.E. Nugroho, M. Fayruz, Narrowband IoT in livestock farming: a technological innovation for productivity and sustainability, *Technical Report*, 2024.
- [79] V. Cabrera, A. Delbuggio, H. Cardoso, D. Fraga, A. Gómez, M. Pedemonte, R. Ungerfeld, J. Oreggioni, Research platform to study sheep behavior, in: 2023 Thinspace IEEE Conference on AgriFood Electronics, CAFE 2023 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 60–64. <https://doi.org/10.1109/CAFE58535.2023.10291765>
- [80] A. Kumar, S. Gupta, Study on ZIGBEE technology, *C Int. J. Eng. Sci. Res. Technol.* 2 (10) (2013) 2733–2738.
- [81] B. Evstatiev, S. Kadirova, N. Valov, Analysis of the wireless communication technologies used in livestock monitoring, in: 2022 International Conference on Communications, Information, Electronic and Energy Systems, CIEES 2022 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2022. <https://doi.org/10.1109/CIEES55704.2022.9990810>
- [82] A. Patil, C. Pawar, N. Patil, R. Tambe, Smart health monitoring system for animals, *Proc. 2015 Int. Conf. Green Comput. Internet Things, ICGCIoT 2015* (2016) 1560–1564. <https://doi.org/10.1109/ICGCIoT.2015.7380715>
- [83] A. Kumar, G.P. Hancke, A zigbee-based animal health monitoring system, *IEEE Sens. J.* 15 (1) (2015) 610–617. <https://doi.org/10.1109/JSEN.2014.2349073>
- [84] A. Yahya, Livestock: monitoring foot and mouth disease using wireless sensor networks, in: A. Yahya (Ed.), *Emerging Technologies in Agriculture, Livestock, and Climate*, Springer International Publishing, Cham, 2020, pp. 45–65. [https://doi.org/10.1007/978-3-030-33487-1\\_2](https://doi.org/10.1007/978-3-030-33487-1_2)
- [85] Z. Zeng, F. Zeng, X. Han, H. Elkhouchlaa, Q. Yu, E. Lü, Real-time monitoring of environmental parameters in a commercial gestating sow house using a zigbee-based wireless sensor network, *Applied Sciences (Switzerland)* 11 (3) (2021) 1–17. <https://doi.org/10.3390/app11030972>
- [86] N. Yamsani, K. Muthukumaran, B.S. Kumar, V. Asha, N. Singh, J.A. Dhanraj, IoT-Based Livestock monitoring and management system using machine learning algorithms, in: Proceedings of 2024 International Conference on Science, Technology, Engineering and Management, ICSTEM 2024, Institute of Electrical and Electronics Engineers Inc., 2024. <https://doi.org/10.1109/ICSTEM61137.2024.10560908>
- [87] D. Gutierrez-Galan, J.P. Dominguez-Morales, E. Cerezuola-Escudero, A. Rios-Navarro, R. Tapiador-Morales, M. Rivas-Perez, M. Dominguez-Morales, A. Jimenez-Fernandez, A. Linares-Barranco, Embedded neural network for real-time animal behavior classification, *Neurocomput.* 272 (2018) 17–26. <https://doi.org/10.1016/j.neucom.2017.03.090>
- [88] M. Centenaro, C.E. Costa, F. Granelli, C. Sacchi, L. Vangelista, A survey on technologies, standards and open challenges in satellite IoT, 2021. <https://doi.org/10.1109/COMST.2021.3078433>
- [89] Q.M. Ilyas, M. Ahmad, Smart farming: an enhanced pursuit of sustainable remote livestock tracking and geofencing using IoT and GPRS, *Wirel. Commun. Mobile Comput.* 2020 (2020). <https://doi.org/10.1155/2020/6660733>
- [90] I. Boughey, E. Hall, R. Bush, Using real-time GNSS tracking tags to monitor alpaca activity in an Australian extensive production system, *Agriculture (Switzerland)* 15 (17) (2025). <https://doi.org/10.3390/agriculture15171839>
- [91] J. Pablo Becoea, M. Grane, M. Miguez, A. Arnaud, Lora, sigfox, and NB-IoT: an empirical comparison for IoT LPWAN technologies in the agribusiness, *IEEE Embed. Syst. Lett.* 16 (3) (2024) 283–286. <https://doi.org/10.1109/LES.2024.3394446>
- [92] V.M.T. Aleluia, V.N.G.J. Soares, J.M.L.P. Caldeira, A.M. Rodrigues, Livestock monitoring: approaches, challenges and opportunities, *Int. J. Eng. Adv. Technol.* 11 (4) (2022) 67–76. <https://doi.org/10.35940/ijeat.D3458.0411422>
- [93] N. Pukrongta, P. Sangmahamad, T. Pechrcool, B. Kumkhet, V. Pirajanchai, P. Thiamsinsangwon, Cattle collars with a low-Cost LPWAN-Based system for cattle estrous monitoring, in: 2023 15th International Conference on Information Technology and Electrical Engineering, ICITEE 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 149–154. <https://doi.org/10.1109/ICITEE59582.2023.10317748>
- [94] F. Nack, *An overview on wireless sensor networks*, *Technical Report*, 2015.
- [95] J. Gray, T.M. Banhazi, A.A. Kist, Wireless data management system for environmental monitoring in livestock buildings, *Inf. Process. Agric.* 4 (1) (2017) 1–17. <https://doi.org/10.1016/j.inpa.2016.12.001>
- [96] K.M. Awan, H.H.R. Sherazi, A. Ali, R. Iqbal, Z.A. Khan, M. Mukherjee, Energy-aware cluster-based routing optimization for WSNs in the livestock industry, *Trans. Emerg. Telecommun. Technol.* 33 (3) (2022). <https://doi.org/10.1002/ett.3816>
- [97] R. Abbasi, P. Martinez, R. Ahmad, The digitization of agricultural industry - a systematic literature review on agriculture 4.0, 2022. <https://doi.org/10.1016/j.atech.2022.100042>
- [98] J.G. Panicker, M. Azman, R. Kashyap, A lora wireless mesh network for wide-area animal tracking, in: 2019 Thinspace IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2019, pp. 1–5. <https://doi.org/10.1109/ICECCT.2019.8868958>
- [99] B. Patrick, T. Johnson, E. Kanjo, Internetless low-Cost sensing system for real-time livestock monitoring, *IEEE Sens. Lett.* 8 (6) (2024) 1–4. <https://doi.org/10.1109/LSENS.2024.3399884>
- [100] R. Casas, A. Hermosa, A. Marco, T. Blanco, F.J. Zarazaga-Soria, Real-time extensive livestock monitoring using lpwan smart wearable and infrastructure, *Applied Sciences (Switzerland)* 11 (3) (2021) 1–18. <https://doi.org/10.3390/app11031240>
- [101] M.S. Farooq, O.O. Sohail, A. Abid, S. Rasheed, A survey on the role of IoT in agriculture for the implementation of smart livestock environment, *IEEE Access* 10 (2022) 9483–9505. <https://doi.org/10.1109/ACCESS.2022.3142848>
- [102] P. Thulasiraman, K.A. White, Topology control of tactical wireless sensor networks using energy efficient zone routing, *Digit. Commun. Netw.* 2 (1) (2016) 1–14. <https://doi.org/10.1016/j.dcan.2016.01.002>
- [103] C. Gu, L. Jiang, R. Tan, LoRa-Based localization: opportunities and challenges (2018).
- [104] A. Mao, M. Zhu, E. Huang, X. Yao, K. Liu, A teacher-to-student information recovery method toward energy-efficient animal activity recognition at low sampling rates, *Comput. Electron. Agric.* 213 (2023). <https://doi.org/10.1016/j.compag.2023.108242>
- [105] E. Bocaj, D. Uzunidis, P. Kasnesis, C.Z. Patrikakis, On the benefits of deep convolutional neural networks on animal activity recognition, in: 2020 International Conference on Smart Systems and Technologies (SST), 2020, pp. 83–88. <https://doi.org/10.1109/SST49455.2020.9263702>
- [106] J.W. Kamminga, D.V. Le, N. Meratnia, P.J.M. Havinga, Towards deep unsupervised representation learning from accelerometer-eter time series for animal activity recognition, *Technical Report*, 2020.
- [107] G. Ni, Z. Shi, Y. Xu, X. Han, J. Miao, W. Tang, Optimization of inertial sensor's sampling frequency and window length for classification of livestock behavior with machine learning, in: 2025 Thinspace IEEE 2nd International Conference on Electronics, Communications and Intelligent Science, ECIS 2025 - Proceeding, Institute of Electrical and Electronics Engineers Inc., 2025. <https://doi.org/10.1109/ECIS65594.2025.11087006>
- [108] A. Mao, M. Zhu, Z. Guo, Z. He, T. Norton, K. Liu, CKSP: cross-species knowledge sharing and preserving for universal animal activity recognition, *Biosyst. Eng.* 259 (2025). <https://doi.org/10.1016/j.biosystemeng.2025.104303>
- [109] A. Mao, M. Peng, G. Liu, M. Zhu, K. Wang, Advancing animal behavior recognition with self-supervised pre-training on unlabeled data, *Sci Rep.* 15 (1) (2025) 43928. <https://doi.org/10.1038/s41598-025-27736-w>
- [110] T. De Waele, A. Shahid, D. Peralta, A. Eerdeken, M. Deruyck, F.A.M. Tuytens, E. De Poorter, Time-Series-Based feature selection and clustering for equine activity recognition using accelerometers, *IEEE Sens. J.* 23 (11) (2023) 11855–11868. <https://doi.org/10.1109/JSEN.2023.3265811>
- [111] N. Tilkens, J. Bindelle, F. Lebeau, A. Siah, A.L.H. Andriamandroso, Quantification of grass-severing bites performed by grazing cattle using halter-mounted accelerometers and machine learning, 2024. <https://doi.org/10.1016/j.atech.2024.100522>
- [112] E. Versluijs, L.J. Niccolai, M. Spedener, B. Zimmermann, A. Hesse, M. Tofastrud, O. Devineau, A.L. Evans, Classification of behaviors of free-ranging cattle using accelerometer signatures collected by virtual fence collars, *Front. Anim. Sci.* 4 (2023). <https://doi.org/10.3389/fanim.2023.1083272>
- [113] T. Tamura, Y. Okubo, Y. Deguchi, S. Koshikawa, M. Takahashi, Y. Chida, K. Okada, Dairy cattle behavior classifications based on decision tree learning using 3-axis neck-mounted accelerometers, *Anim. Sci. J.* 90 (4) (2019) 589–596. <https://doi.org/10.1111/asj.13184>
- [114] K.K. Fogsgaard, C.M. Røntved, P. Sørensen, M.S. Herskin, Sickness behavior in dairy cows during escherichia coli mastitis, *J. Dairy Sci.* 95 (2) (2012) 630–638. <https://doi.org/10.3168/jds.2011-4350>
- [115] P. Sepúlveda-Varas, K.L. Proudfoot, D.M. Weary, M.A.G. von Keyserlingk, Changes in behaviour of dairy cows with clinical mastitis, *Appl. Anim. Behav. Sci.* 175 (2016) 8–13. <https://doi.org/10.1016/j.applanim.2014.09.022>

- [116] K. Schirmann, N. Chapinal, D.M. Weary, W. Heuwieser, M.A.G. von Keyserlingk, Rumination and its relationship to feeding and lying behavior in Holstein dairy cows, *J. Dairy Sci.* 95 (6) (2012) 3212–3217. <https://doi.org/10.3168/jds.2011-4741>
- [117] Y. Wu, M. Liu, Z. Peng, M. Liu, M. Wang, Y. Peng, Recognising cattle behaviour with deep residual bidirectional LSTM model using a wearable movement monitoring collar, *Agriculture (Switzerland)* 12 (8) (2022). <https://doi.org/10.3390/agriculture12081237>
- [118] R.N. Watanabe, E.P. Romanzini, P.A. Bernardes, J.L. Rodrigues, G. Alves do Val, M.M. Silva, M.H.M. d.R. Fernandes, S.L. Caetano, S.B. Ramos, R.A. Reis, D.P. Munari, Accelerometers-based position and time interval comparisons for predicting the behaviors of young bulls housed in a feedlot system, *Smart Agric. Technol.* 9 (2024). <https://doi.org/10.1016/j.atech.2024.100542>
- [119] D. Pavlovic, C. Davison, A. Hamilton, O. Marko, R. Atkinson, C. Michie, V. Crnojević, I. Andonovic, X. Bellekens, C. Tachtatzis, Classification of cattle behaviours using neck-mounted accelerometer-equipped collars and convolutional neural networks, *Sensors* 21 (12) (2021). <https://doi.org/10.3390/s21124050>
- [120] R. Arablouei, Z. Wang, G.J. Bishop-Hurley, J. Liu, Multimodal sensor data fusion for in-situ classification of animal behavior using accelerometry and GNSS data, *Smart Agric. Technol.* 4 (2023). <https://doi.org/10.1016/j.atech.2022.100163>
- [121] J. Cabezas, R. Yubero, B. Visitación, J. Navarro-García, M.J. Algar, E.L. Cano, F. Ortega, Analysis of accelerometer and GPS data for cattle behaviour identification and anomalous events detection, *Entropy* 24 (3) (2022). <https://doi.org/10.3390/e24030336>
- [122] G. Hernández, C. González-Sánchez, A. González-Arrieta, G. Sánchez-Brizuela, J.C. Fraile, Machine learning-based prediction of cattle activity using sensor-based data, *Sensors* 24 (10) (2024). <https://doi.org/10.3390/s24103157>
- [123] A.W. Hamilton, C. Davison, C. Tachtatzis, I. Andonovic, C. Michie, H.J. Ferguson, L. Somerville, N.N. Jonsson, Identification of the rumination in cattle using support vector machines with motion-sensitive bolus sensors, *Sensors (Switzerland)* 19 (5) (2019). <https://doi.org/10.3390/s19051165>
- [124] L. Riaboff, S. Aubin, N. Bédère, S. Couvreur, A. Madouasse, E. Goumand, A. Chauvin, G. Plantier, Evaluation of pre-processing methods for the prediction of cattle behaviour from accelerometer data, *Comput. Electron. Agric.* 165 (2019). <https://doi.org/10.1016/j.compag.2019.104961>
- [125] L. Riaboff, S. Couvreur, A. Madouasse, M. Roig-Pons, S. Aubin, P. Massabie, A. Chauvin, N. Bédère, G. Plantier, Use of predicted behavior from accelerometer data combined with GPS data to explore the relationship between dairy cow behavior and pasture characteristics, *Sensors (Switzerland)* 20 (17) (2020) 1–33. <https://doi.org/10.3390/s20174741>
- [126] V.A. Camargo, E.A. Pajor, S. Bayat, J.M. Pearson, Predicting main behaviors of beef bulls from accelerometer data: a machine learning framework, *Smart Agric. Technol.* 9 (2024). <https://doi.org/10.1016/j.atech.2024.100683>
- [127] J. Wang, Z. He, G. Zheng, S. Gao, K. Zhao, Development and validation of an ensemble classifier for real-time recognition of cow behavior patterns from accelerometer data and location data, *PLoS ONE* 13 (9) (2018). <https://doi.org/10.1371/journal.pone.0203546>
- [128] S. Benaisa, F.A.M. Tuytens, D. Plets, T. de Pessemier, J. Trogh, E. Tanghe, L. Martens, L. Vandaele, A. Van Huffel, W. Joseph, B. Sonck, On the use of on-cow accelerometers for the classification of behaviours in dairy barns, *Res. Vet. Sci.* 125 (2019) 425–433. <https://doi.org/10.1016/j.rvsc.2017.10.005>
- [129] L.A. Coelho Ribeiro, T. Bresolin, R.G. J.d. Magalhães, D. Rume Casagrande, D.M. de Arruda Camargo, J.R.R. Dórea, Disentangling data dependency using cross-validation strategies to evaluate prediction quality of cattle grazing activities using machine learning algorithms and wearable sensor data, *J. Anim. Sci.* 99 (9) (2021). <https://doi.org/10.1093/jas/skab206>
- [130] L. Schmeling, G. Elmamooz, P.T. Hoang, A. Kozar, D. Nicklas, M. Sünkel, S. Thurner, E. Rauch, Training and validating a machine learning model for the sensor-based monitoring of lying behavior in dairy cows on pasture and in the barn, *Animals* 11 (9) (2021). <https://doi.org/10.3390/ani11092660>
- [131] J. Brennan, P. Johnson, K. Olson, Classifying season long livestock grazing behavior with the use of a low-cost GPS and accelerometer, *Comput. Electron. Agric.* 181 (2021). <https://doi.org/10.1016/j.compag.2020.105957>
- [132] H. Kour, K.P. Patison, N.J. Corbet, D.L. Swain, Validation of accelerometer use to measure suckling behaviour in northern Australian beef calves, *Appl. Anim. Behav. Sci.* 202 (2018) 1–6. <https://doi.org/10.1016/j.applanim.2018.01.012>
- [133] O. Dissanayake, S. McPherson, J. Allyndree, E. Kennedy, P. Cunningham, L. Riaboff, Personalized weighted adaboost for animal behavior recognition from sensor data, in: 2023 31st Irish Conference on Artificial Intelligence and Cognitive Science, AICS 2023, Institute of Electrical and Electronics Engineers Inc., 2023. <https://doi.org/10.1109/AICS60730.2023.10470889>
- [134] O.I. Dissanayake, S.E. McPherson, J. Allyndree, E. Kennedy, P. Cunningham, L. Riaboff, ActBCalf: accelerometer-based multivariate time-series dataset for calf behavior classification, *Data Brief* 60 (2025). <https://doi.org/10.1016/j.dib.2025.111462>
- [135] H. Ito, K.-i. Takeda, K.K. Tokgoz, L. Minati, M. Fukawa, L. Chao, J. Bartels, I. Rachi, S. A, Japanese Black Beef Cow Behavior Classification Dataset, 2021. <https://doi.org/10.5281/zenodo.5399259>
- [136] C. Li, K. Tokgoz, M. Fukawa, J. Bartels, T. Ohashi, K.I. Takeda, H. Ito, Data augmentation for inertial sensor data in CNNs for cattle behavior classification, *IEEE Sens. Lett.* 5 (11) (2021). <https://doi.org/10.1109/LSENS.2021.3119056>
- [137] J. Bartels, K.K. Tokgoz, M. Fukawa, S. Otsubo, L. Chao, I. Rachi, K. Takeda, H. Ito, A 216  $\mu$ V, 87% accurate cow behavior classifying decision tree on FPGA with interpolated ARCTAN2, in: Proceedings - IEEE International Symposium on Circuits and Systems, 2021-May, Institute of Electrical and Electronics Engineers Inc., 2021. <https://doi.org/10.1109/ISCAS51556.2021.9401342>
- [138] J. Bartels, K.K. Tokgoz, A. Sihan, M. Fukawa, S. Otsubo, C. Li, I. Rachi, K.I. Takeda, L. Minati, H. Ito, Tinycownet: memory- and power-minimized RNNs implementable on tiny edge devices for lifelong cow behavior distribution estimation, *IEEE Access* 10 (2022) 32706–32727. <https://doi.org/10.1109/ACCESS.2022.3156278>
- [139] J. Bartels, A. Hagihara, L. Minati, K.K. Tokgoz, H. Ito, An integer-Only resource-Minimized RNN on FPGA for low-Frequency sensors in edge-AI, *IEEE Sens. J.* 23 (15) (2023) 17784–17793. <https://doi.org/10.1109/JSEN.2023.3286580>
- [140] S. Mekruksavanich, P. Jantawong, D. Tancharoen, A. Jitpattanakul, Sensor-Based cattle behavior classification using deep learning approaches, in: 2023 International Technical Conference on Circuits/Systems, Computers, and Communications, ITC-CSCC 2023, Institute of Electrical and Electronics Engineers Inc., 2023. <https://doi.org/10.1109/ITC-CSCC58803.2023.10212958>
- [141] L. Minati, C. Li, J. Bartels, P. Chakraborty, Z. Li, N. Yoshimura, M. Frasca, H. Ito, Accelerometer time series augmentation through externally driving a non-linear dynamical system, *ChaosSolitons Fractals* 168 (2023). <https://doi.org/10.1016/j.chaos.2023.113100>
- [142] K. El Moutaouakil, N. Falih, A new approach to animal behavior classification using recurrent neural networks, in: 2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET 2024, Institute of Electrical and Electronics Engineers Inc., 2024. <https://doi.org/10.1109/IRASET60544.2024.10549544>
- [143] K. El Moutaouakil, N. Falih, A hybrid CNN-LSTM approach for classifying cattle behavior using accelerometer data, in: 2025 International Conference on Circuit, Systems and Communication, ICCSC 2025, Institute of Electrical and Electronics Engineers Inc., 2025. <https://doi.org/10.1109/ICCSC66714.2025.11135242>
- [144] L. Wang, R. Arablouei, F.A.P. Alvarenga, G.J. Bishop-Hurley, Classifying animal behavior from accelerometry data via recurrent neural networks, *Comput. Electron. Agric.* 206 (2023). <https://doi.org/10.1016/j.compag.2023.107647>
- [145] R. Arablouei, L. Wang, L. Currie, J. Yates, F.A.P. Alvarenga, G.J. Bishop-Hurley, Animal behavior classification via deep learning on embedded systems, *Comput. Electron. Agric.* 207 (2023). <https://doi.org/10.1016/j.compag.2023.107707>
- [146] R. Arablouei, L. Wang, C. Phillips, L. Currie, J. Yates, G. Bishop-Hurley, In-situ animal behavior classification using knowledge distillation and fixed-point quantization, *Smart Agric. Technol.* 4 (2023). <https://doi.org/10.1016/j.atech.2022.100159>
- [147] R. Arablouei, L. Currie, B. Kusy, A. Ingham, P.L. Greenwood, G. Bishop-Hurley, In-situ classification of cattle behavior using accelerometry data, *Comput. Electron. Agric.* 183 (2021). <https://doi.org/10.1016/j.compag.2021.106045>
- [148] S. Hosseiniorbin, S. Layeghy, B. Kusy, R. Jurdak, G.J. Bishop-Hurley, P.L. Greenwood, M. Portmann, Deep learning-based cattle behaviour classification using joint time-frequency data representation, *Comput. Electron. Agric.* 187 (2021). <https://doi.org/10.1016/j.compag.2021.106241>
- [149] H. Vu, O. Prabhune, U. Raskar, D. Panditharatne, H. Chung, C.Y. Choi, Y. Kim, MmCows: a multimodal dataset for dairy cattle monitoring, in: A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, C. Zhang (Eds.), Advances in Neural Information Processing Systems, 37, Curran Associates, Inc., 2024, pp. 59451–59467. <https://doi.org/10.52202/079017-1898>
- [150] R. Primi, G. Bernabucci, C. Evangelista, P. Viola, P. Girotti, R. Spina, S. Compagnucci, B. Ronchi, Ecosystem services linked to extensive sheep and goat farming in mountain areas: a global literature analysis using text mining and topic analysis, *Animals* 15 (3) (2025). <https://doi.org/10.3390/ani15030350>
- [151] 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 (10) (2018). <https://doi.org/10.3390/s18103532>
- [152] J. Fitzpatrick, M. Scott, A. Nolan, Assessment of pain and welfare in sheep, in: *Small Ruminant Research*, 62, 2006, pp. 55–61. <https://doi.org/10.1016/j.smallrumres.2005.07.028>
- [153] K. Sakai, K. Oishi, M. Miwa, H. Kumagai, H. Hirooka, Behavior classification of goats using 9-axis multi sensors: the effect of imbalanced datasets on classification performance, *Comput. Electron. Agric.* 166 (2019). <https://doi.org/10.1016/j.compag.2019.105027>
- [154] J. Barwick, D.W. Lamb, R. Dobos, M. Welch, M. Trotter, Categorising sheep activity using a tri-axial accelerometer, *Comput. Electron. Agric.* 145 (2018) 289–297. <https://doi.org/10.1016/j.compag.2018.01.007>
- [155] E. Price, J. Langford, T.W. Fawcett, A.J. Wilson, D.P. Croft, Classifying the posture and activity of ewes and lambs using accelerometers and machine learning on a commercial flock, *Appl. Anim. Behav. Sci.* 251 (2022). <https://doi.org/10.1016/j.applanim.2022.105630>
- [156] M. Zhang, Y. Zhu, J. Wu, Q. Zhao, X. Zhang, H. Luo, Improved composite deep learning and multi-scale signal features fusion enable intelligent and precise behaviors recognition of fattening hu sheep, *Comput. Electron. Agric.* 227 (2024). <https://doi.org/10.1016/j.compag.2024.109635>
- [157] 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 (14) (2019). <https://doi.org/10.3390/s19143201>
- [158] S. Mauny, J. Kwon, N.C. Friggens, C. Duvaux-Ponter, M. Taghipoor, A pipeline with pre-processing options to detect behaviour from accelerometer data using machine learning tested on dairy goats, *Peer Community J.* 5 (2025) 000. <https://doi.org/10.24072/pci>
- [159] 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, *R. Soc. Open Sci.* 5 (2) (2018). <https://doi.org/10.1098/rsos.171442>

- [160] K.E. Turner, A. Thompson, I. Harris, M. Ferguson, F. Sohler, Deep learning based classification of sheep behaviour from accelerometer data with imbalance, *Inf. Process. Agric.* 10 (3) (2023) 377–390. <https://doi.org/10.1016/j.inpa.2022.04.001>
- [161] 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 Sens. J.* 19 (7) (2019) 2722–2734. <https://doi.org/10.1109/JSEN.2018.2886890>
- [162] 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, *Comput. Electron. Agric.* 150 (2018) 394–401. <https://doi.org/10.1016/j.compag.2018.05.004>
- [163] M. Bonneau, L. Faillot, W. Troupe, L. Riaboff, Hierarchical long short term memory recurrent neural network for goats behaviour prediction from accelerometer data, *Smart Agric. Technol.* 11 (2025) 2772–3755. <https://doi.org/10.17180/50N1-KN86>
- [164] 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 Sens. (Basel)* 12 (4) (2020). <https://doi.org/10.3390/rs12040646>
- [165] J.W. Kamminga, H.C. Bisby, D.V. Le, N. Meratnia, P.J.M. Havinga, Generic online animal activity recognition on collar tags, in: *UbiComp/ISWC 2017 - Adjunct Proceedings of the 2017 Thinspace ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 Thinspace ACM International Symposium on Wearable Computers, Association for Computing Machinery, Inc, 2017*, pp. 597–606. <https://doi.org/10.1145/3123024.3124407>
- [166] U. Auer, Z. Kelemen, V. Engl, F. Jenner, Activity time budgets-a potential tool to monitor equine welfare?, 2021. <https://doi.org/10.3390/ani11030850>
- [167] C. Giannetto, M. Bazzano, S. Marafioti, C. Bertolucci, G. Piccione, Monitoring of total locomotor activity in mares during the prepartum and postpartum period, *J. Vet. Behav.* 10 (5) (2015) 427–432. <https://doi.org/10.1016/j.jveb.2015.05.003>
- [168] J.P. A.M. van Loon, M.C. Van Dierendonck, Monitoring acute equine visceral pain with the equine utrecht university scale for composite pain assessment (EQUUS-COMPASS) and the equine utrecht university scale for facial assessment of pain (EQUUS-FAP): a scale-construction study, *Vet. J.* 206 (3) (2015) 356–364. <https://doi.org/10.1016/j.tvjl.2015.08.023>
- [169] J.P. Dominguez-Morales, L. Duran-Lopez, D. Gutierrez-Galan, A. Rios-Navarro, A. Linares-Barranco, A. Jimenez-Fernandez, Wildlife monitoring on the edge: a performance evaluation of embedded neural networks on microcontrollers for animal behavior classification, *Sensors* 21 (9) (2021). <https://doi.org/10.3390/s21092975>
- [170] U. Kamiya, K. Kakiuchi, K. Kawamura, K. Ueda, M. Kawai, A. Matsui, N. Negishi, Deep learning approach for classifying grazing behavior in yearling horses using triaxial accelerometer data: a pilot study, *J. Equine Vet. Sci.* 155 (2025). <https://doi.org/10.1016/j.jvevs.2025.105706>
- [171] K. Kirsch, S. Strutzke, L. Klitzing, F. Pilger, C. Thöne-Reineke, G. Hoffmann, Validation of a time-Distributed residual LSTM-CNN and BiLSTM for equine behavior recognition using collar-worn sensors, *Comput. Electron. Agric.* 231 (2025). <https://doi.org/10.1016/j.compag.2025.109999>
- [172] X. Chen, F. Li, J. Li, Q. Fan, P. Kwan, W. Zheng, L. Guo, Development of a device and algorithm research for akhal-Teke activity level analysis, *Applied Sciences (Switzerland)* 14 (13) (2024). <https://doi.org/10.3390/app14135424>
- [173] A. Mao, E. Huang, H. Gan, R.S.V. Parkes, W. Xu, K. Liu, Cross-modality interaction network for equine activity recognition using imbalanced multi-modal data, *Sensors* 21 (17) (2021). <https://doi.org/10.3390/s21175818>
- [174] A. Mao, E. Huang, H. Gan, K. Liu, FedAAR: a novel federated learning framework for animal activity recognition with wearable sensors, *Animals* 12 (16) (2022). <https://doi.org/10.3390/ani12162142>
- [175] T. Waele, A. Shahid, D. Peralta, F. Tuytens, E. Poorter, Towards unsupervised animal activity recognition: a deep learning based clustering algorithm for equine gait classification, 2024. <https://doi.org/10.21203/rs.3.rs-3920731/v1>
- [176] B. Cetintav, A. Yalcin, Exploring equine behavior: wearable sensors data and explainable AI for enhanced classification, *J. Equine Vet. Sci.* 149 (2025). <https://doi.org/10.1016/j.jvevs.2025.105568>
- [177] J.W. Kamminga, L.M. Janßen, N. Meratnia, P.J.M. Havinga, Horsing around-a dataset comprising horse movement (2019). <https://doi.org/10.4121/uuid:2e08745c-4178-4183-8551-f248c992cb14>
- [178] E. Casella, A.R. Khamesi, S. Silvestri, A framework for the recognition of horse gaits through wearable devices, *Pervasive Mob. Comput.* 67 (2020). <https://doi.org/10.1016/j.pmcj.2020.101213>
- [179] K. Fujinami, R. Takuno, I. Sato, T. Shimamura, Evaluating behavior recognition pipeline of laying hens using wearable inertial sensors, *Sensors* 23 (11) (2023). <https://doi.org/10.3390/s23115077>
- [180] S.M. Derakhshani, M. Overduin, T.G. C.M. van Niekerk, P.W.G. Groot Koerkamp, Implementation of inertia sensor and machine learning technologies for analyzing the behavior of individual laying hens, *Animals* 12 (5) (2022). <https://doi.org/10.3390/ani12050536>
- [181] X. Yang, Y. Zhao, G.M. Street, Y. Huang, S.D. Filip To, J.L. Purswell, Classification of broiler behaviours using triaxial accelerometer and machine learning, *Animal* 15 (7) (2021). <https://doi.org/10.1016/j.animal.2021.100269>
- [182] A.C. Murillo, A. Abdoli, R.A. Blatchford, E.J. Keogh, A.C. Gerry, Parasitic mites alter chicken behaviour and negatively impact animal welfare, *Sci Rep.* 10 (1) (2020). <https://doi.org/10.1038/s41598-020-65021-0>
- [183] M. Sakib, S.S. Alam, M.M. Hossain, Detection of broiler behaviors through a wearable sensor system and machine learning methods, in: *2023 Thinspace IEEE Engineering Informatics, EI 2023, Institute of Electrical and Electronics Engineers Inc., 2023*. <https://doi.org/10.1109/IEEECONF58110.2023.10520457>
- [184] J. Pearce, Y.M. Chang, D. Xia, S. Abeyesinghe, Classification of behaviour in conventional and slow-Growing strains of broiler chickens using tri-Axial accelerometers, *Animals* 14 (13) (2024). <https://doi.org/10.3390/ani14131957>
- [185] H. Zhang, H. Lin, C. Li, Y. Zhang, Optimization of broiler behavior recognition technology based on accelerometer, in: *2024 7th International Conference on Computer Information Science and Application Technology, CISAT 2024, Institute of Electrical and Electronics Engineers Inc., 2024*, pp. 1158–1163. <https://doi.org/10.1109/CISAT62382.2024.10695353>
- [186] H. Wu, G. Ni, Z. Shi, Y. Xu, J. Miao, W. Jiang, X. Han, M. Si, W. Tang, Intelligent classification of swine behaviors using an inertial sensor: effects of data features and machine learning models on classification performance, in: *2025 Thinspace IEEE 2nd International Conference on Electronics, Communications and Intelligent Science, ECIS 2025 - Proceeding, Institute of Electrical and Electronics Engineers Inc., 2025*. <https://doi.org/10.1109/ECIS65594.2025.11087052>
- [187] M. Knoll, C. Raudies, L. Gygas, Behavioural complexity in fattening pigs II. accelerometer-based validation of a novel welfare indicator with potential for automation, *Appl. Anim. Behav. Sci.* 286 (2025). <https://doi.org/10.1016/j.applanim.2025.106606>
- [188] A. Veit, I. Fuxjäger, M. Wondrak, L. Huber, Social influence and attention bias in free-ranging domestic pigs: effects of demonstrator rank and friendship, *Appl. Anim. Behav. Sci.* 275 (2024). <https://doi.org/10.1016/j.applanim.2024.106285>
- [189] P. Bota, J. Silva, D. Folgado, H. Gamboa, A semi-automatic annotation approach for human activity recognition, *Sensors (Switzerland)* 19 (3) (2019). <https://doi.org/10.3390/s19030501>
- [190] J. Li, M. Keselman, E. Shlizerman, Openlabcluster: active learning based clustering and classification of animal behaviors based on kinematic body keypoints, *Front. Syst. Neurosci.* 19 (2025). <https://doi.org/10.3389/fnsys.2025.1630654>
- [191] E. Jakku, B. Taylor, A. Fleming, C. Mason, S. Fielke, C. Soumness, P. Thorburn, “If they don’t tell us what they do with it, why would we trust them?” trust, transparency and benefit-sharing in smart farming, *NJAS - Wageningen J. Life Sci.* 90–91 (2019). <https://doi.org/10.1016/j.njas.2018.11.002>
- [192] W. Shen, F. Cheng, Y. Zhang, X. Wei, Q. Fu, Y. Zhang, Automatic recognition of ingestive-related behaviors of dairy cows based on triaxial acceleration, *Inf. Process. Agric.* 7 (3) (2020) 427–443. <https://doi.org/10.1016/j.inpa.2019.10.004>
- [193] D.E. Essien, S. Inyang, I. Umoren, Evaluating machine learning classifiers and explainability for monitoring cow behaviour with wearable nose rings, *Prev. Vet. Med.* 244 (2025). <https://doi.org/10.1016/j.prevetmed.2025.106630>
- [194] S. Mekruksavanich, P. Jantawong, A. Jitpattanakul, Resnet-based deep neural network using transfer learning for animal activity recognition, in: *6th International Conference on Information Technology, InCIT 2022, Institute of Electrical and Electronics Engineers Inc., 2022*, pp. 445–449. <https://doi.org/10.1109/InCIT56086.2022.10067405>
- [195] S. Benaissa, F.A.M. Tuytens, D. Plets, H. Cattrysse, L. Martens, L. Vandaele, W. Joseph, B. Sonck, Classification of ingestive-related cow behaviours using RumiWatch halter and neck-mounted accelerometers, *Appl. Anim. Behav. Sci.* 211 (2019) 9–16. <https://doi.org/10.1016/j.applanim.2018.12.003>
- [196] E.R. Dickinson, J.P. Twining, R. Wilson, P.A. Stephens, J. Westander, N. Marks, D.M. Scantlebury, Limitations of using surrogates for behaviour classification of accelerometer data: refining methods using random forest models in caprids, *Mov. Ecol.* 9 (1) (2021). <https://doi.org/10.1186/s40462-021-00265-7>
- [197] C. Contoli, V. Freschi, E. Lattanzi, Energy-aware human activity recognition for wearable devices: a comprehensive review, 2024. <https://doi.org/10.1016/j.pmcj.2024.101976>
- [198] P. de Souza, D. Silva, I. de Andrade, J. Dias, J.P. Lima, V. Teichrieb, J.P. Quintino, F.Q.B. da Silva, A.L.M. Santos, A study on the influence of sensors in frequency and time domains on context recognition, *Sensors* 23 (12) (2023). <https://doi.org/10.3390/s23125756>
- [199] S. Saeb, L. Lonini, A. Jayaraman, D.C. Mohr, K.P. Kording, The need to approximate the use-case in clinical machine learning, *GigaScience* 6 (5) (2017) 1–9. <https://doi.org/10.1093/gigascience/gix019>
- [200] D.R. Roberts, V. Bahn, S. Ciuti, M.S. Boyce, J. Elith, G. Guillera-Aroita, S. Hauenstein, J.J. Lahoz-Monfort, B. Schröder, W. Thuiller, D.I. Warton, B.A. Wintle, F. Hartig, C.F. Dormann, Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, 2017. <https://doi.org/10.1111/ecog.02881>
- [201] S. Hu, A. Ingham, S. Schmoelzl, J. McNally, B. Little, D. Smith, G. Bishop-Hurley, Y.G. Wang, Y. Li, Inclusion of features derived from a mixture of time window sizes improved classification accuracy of machine learning algorithms for sheep grazing behaviours, *Comput. Electron. Agric.* 179 (2020). <https://doi.org/10.1016/j.compag.2020.105857>
- [202] I. Marin, D.A. Ciungan, R.E. Vasiliu, D.N. Enea, Improving livestock monitoring: multi-classifier approach for cattle behavior recognition, in: *2025 Thinspace IEEE International Black Sea Conference on Communications and Networking, BlackSeaCom 2025, Institute of Electrical and Electronics Engineers Inc., 2025*. <https://doi.org/10.1109/BlackSeaCom65655.2025.11193930>
- [203] G. Simanungkalit, J. Barwick, F. Cowley, B. Dawson, R. Dobos, R. Hegarty, Use of an ear-tag accelerometer and a radio-frequency identification (RFID) system for monitoring the licking behaviour in grazing cattle, *Appl. Anim. Behav. Sci.* 244 (2021). <https://doi.org/10.1016/j.applanim.2021.105491>
- [204] A. Yaqoob, N.K. Verma, M.A. Mir, G.G. Tejani, N.H.B. Eisa, H. Mamoun Hussien Osman, M.A. Shah, SGA-Driven Feature selection and random forest classification for enhanced breast cancer diagnosis: a comparative study, *Sci Rep.* 15 (1) (2025). <https://doi.org/10.1038/s41598-025-95786-1>
- [205] O. Berezsky, O. Kovalchuk, K. Berezka, R. Ivanyskyy, Assessing smart cities’ effectiveness: machine learning approaches, *Front. Sustain. Cities* 7 (2025). <https://doi.org/10.3389/frsc.2025.1400917>

- [206] S. Saurav, J. Kumar, Exploring the effectiveness of machine learning models on livestock behavior classification using 9-axial inertial sensor, *Evol. Intell.* 18 (4) (2025). <https://doi.org/10.1007/s12065-025-01071-5>
- [207] T.H. Dang, J. Park, V.T. Tran, W.Y. Chung, VAE-LSTM data augmentation for cattle behavior classification using a wearable inertial sensor, *IEEE Sens. Lett.* 8 (5) (2024) 1–4. <https://doi.org/10.1109/LSENS.2024.3385418>
- [208] K. El Moutaouakil, N. Falih, A hybrid CNN-LSTM approach for classifying cattle behavior using accelerometer data, in: 2025 International Conference on Circuit, Systems and Communication, ICCSC 2025, Institute of Electrical and Electronics Engineers Inc., 2025. <https://doi.org/10.1109/ICCSC66714.2025.11135242>
- [209] Q. Zhang, E. Kanjo, A multicore and edge TPU-accelerated multimodal tinyML system for livestock behavior recognition, *IEEE Internet Things J.* (2025). <https://doi.org/10.1109/JIOT.2025.3624811>
- [210] T.H. Dang, L. Nkenyereye, V.T. Tran, W.Y. Chung, Self-Powered cattle behavior monitoring system using 915 MHz radio frequency energy harvesting, *IEEE Access* 12 (2024) 33779–33791. <https://doi.org/10.1109/ACCESS.2024.3360852>
- [211] A. Ignatov, Real-time human activity recognition from accelerometer data using convolutional neural networks, *Appl. Soft Comput. J.* 62 (2018) 915–922. <https://doi.org/10.1016/j.asoc.2017.09.027>
- [212] B.A. Khan, J.W. Jung, Deep learning-Based human activity recognition using dilated CNN and LSTM on video sequences of various actions dataset, *Applied Sciences (Switzerland)* 15 (22) (2025). <https://doi.org/10.3390/app152212173>
- [213] A. Murad, J.Y. Pyun, Deep recurrent neural networks for human activity recognition, *Sensors (Switzerland)* 17 (11) (2017). <https://doi.org/10.3390/s17112556>
- [214] F.J. Ordóñez, D. Roggen, Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition, *Sensors (Switzerland)* 16 (1) (2016). <https://doi.org/10.3390/s16010115>
- [215] B. Achour, M. Belkadi, R. Saddaoui, I. Filali, R. Aoudjit, M. Laghrouche, High-accuracy and energy-efficient wearable device for dairy cows' localization and activity detection using low-cost IMU/Rfid sensors, *Microsyst. Technol.* 28 (5) (2022) 1241–1251. <https://doi.org/10.1007/s00542-022-05288-7>
- [216] R. Singh, S.S. Gill, Edge AI: a survey, 2023. <https://doi.org/10.1016/j.iotcps.2023.02.004>
- [217] M. Gibbs, E. Kanjo, Realising the power of edge intelligence: addressing the challenges in AI and tinyML applications for edge computing, in: Proceedings - IEEE International Conference on Edge Computing, 2023-July, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 337–343. <https://doi.org/10.1109/EDGE60047.2023.00056>
- [218] C.R. Banbury, V.J. Reddi, M. Lam, W. Fu, A. Fazel, J. Holleman, X. Huang, R. Hurtado, D. Kanter, A. Lokhmatov, D. Patterson, D. Pau, J.-s. Seo, J. Sieracki, U. Thakker, M. Verhelst, P. Yadav, Benchmarking TinyML systems: challenges and direction (2020).
- [219] D.L. Dutta, S. Bharali, TinyML Meets IoT: a comprehensive survey, 2021. <https://doi.org/10.1016/j.iot.2021.100461>
- [220] M.M.H. Shuvo, Edge AI: leveraging the full potential of deep learning, in: Al-Emran Mostafa, K. Shaalan (Eds.), Recent Innovations in Artificial Intelligence and Smart Applications, Springer International Publishing, Cham, 2022, pp. 27–46. [https://doi.org/10.1007/978-3-031-14748-7\\_2](https://doi.org/10.1007/978-3-031-14748-7_2)
- [221] S. Neethirajan, The role of sensors, big data and machine learning in modern animal farming, 2020. <https://doi.org/10.1016/j.sbsr.2020.100367>
- [222] P.P. Ray, A review on TinyML: state-of-the-art and prospects, 2022. <https://doi.org/10.1016/j.jksuci.2021.11.019>
- [223] J. Nordby, emlearn: machine learning inference engine for microcontrollers and embedded devices, 2021. <https://doi.org/10.5281/zenodo.2589394>
- [224] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, Tensorflow: large-scale machine learning on heterogeneous distributed systems (2016).
- [225] V.M. Do, T. Quang-Huy, N. Van Son, P. Van Thanh, N.C. Minh, D.T. Tran, D.N. Tran, The effect of sensor position deflection on behavior classification performance, in: International Conference on Advanced Technologies for Communications, 2022-October, IEEE Computer Society, 2022, pp. 126–129. <https://doi.org/10.1109/ATC55345.2022.9942988>
- [226] A. Rahman, D.V. Smith, B. Little, A.B. Ingham, P.L. Greenwood, G.J. Bishop-Hurley, Cattle behaviour classification from collar, halter, and ear tag sensors, *Inf. Process. Agric.* 5 (1) (2018) 124–133. <https://doi.org/10.1016/j.inpa.2017.10.001>
- [227] D.N. Tran, T.N. Nguyen, P.C.P. Khanh, D.T. Tran, An IoT-Based design using accelerometers in animal behavior recognition systems, *IEEE Sens. J.* 22 (18) (2022) 17515–17528. <https://doi.org/10.1109/JSEN.2021.3051194>
- [228] Y. Fu, M. Zhang, D. Mao, D. Wang, H. Luo, X. Zhang, Improved feed-forward neural network and multidimensional motion data for accurate recognition of sheep behaviour at multiple fattening stages, *Biosyst. Eng.* 257 (2025). <https://doi.org/10.1016/j.biosystemseng.2025.104218>
- [229] T. Li, T. Li, R. Su, J. Xin, D. Han, Classification and recognition of goat movement behavior based on SL-WOA-XGBoost, *Electronics (Switzerland)* 12 (16) (2023). <https://doi.org/10.3390/electronics12163506>
- [230] L. Lush, R.P. Wilson, M.D. Holton, P. Hopkins, K.A. Marsden, D.R. Chadwick, A.J. King, Classification of sheep urination events using accelerometers to aid improved measurements of livestock contributions to nitrous oxide emissions, *Comput. Electron. Agric.* 150 (2018) 170–177. <https://doi.org/10.1016/j.compag.2018.04.018>