Classifying horse activities with big data using machine learning

Derya Birant*, Emircan Tepe
Dept. of Computer Engineering, Dokuz Eylul University, Turkey
**Corresponding author: derya@cs.deu.edu.tr*

Abstract

Using big data-assisted machine learning methods in animal science has received increasing attention in recent years since they extract useful insights from large-scale animal datasets. Especially, animal activity recognition is the task of identifying the actions performed by animals and can provide rich insight into their health, welfare, reproduction, survival, foraging, and interaction with humans/other animals. This paper aims to propose a new solution for this purpose by building a machine learning model that classifies the actions of horses based on big sensor data. Unlike the previous studies, our study is original in that it compares the accuracies of per-subject (personalized) and cross-subject (generalized) models. It is the first study that especially compares different ensemble learning algorithms for horse activity recognition in terms of classification accuracy, including bagging trees, extremely randomized trees, random forest, extreme gradient boosting, light gradient boosting, gradient boosting, and categorical boosting. The purpose of the study is to classify five horse activities: walking, standing, grazing, galloping, and trotting. The experimental results showed that our solution achieved very good performance (94.62%) on average on a real-world dataset. Furthermore, the results also showed that our method outperformed the state-of-the-art methods on the same dataset.

Keywords: Animal activity recognition; big data; classification; machine learning; wearable sensors

1. Introduction

In the context of big data, machine learning methods enable us uncovering of complex patterns and making accurate predictions, e.g., for diagnosis in health (Al-Dousari *et al.*, 2021; Oshinubi *et al.*, 2021; Nallamuth & Palanichamy, 2015; Colak *et al.*, 2016), opinion target identification (Khan *et al.*, 2016), text classification (Jain & Kumar, 2018), image retrieval (Mehmood *et al.*, 2018), and recognition (Datsi *et al.*, 2021). In this study, we focus on the power of machine learning with big data for animal activity recognition.

Animal activity recognition (AAR) is the task of monitoring and identifying the actions performed by an animal in various environmental conditions by analyzing big sensor or video data using an automated system. It is a growing research field that can help with the tracking of wildlife, livestock, and pets. AAR has been widely studied with the use of various sensors such as accelerometers, magnetometers, and gyroscopes, which are used to measure acceleration, magnetism, and angular velocity, respectively. A typical AAR system uses machine learning methods to distinguish various activities such as walking, standing, grazing, and jumping.

There are many reasons why animal activities are important to be recognized. The behavior of animals is a rich source of information for numerous application domains, including wildlife monitoring, poaching detection, livestock management, and domestic animal tracking. Animal activity recognition can provide rich insight into their health and welfare. For example, declining in food intake, tracking the irregularities of the animal gaits, or recognizing pain-related animal activities like pawing and rolling might be a sign of health issues and help early identification of health disorders. Monitoring the movements of a pregnant animal through an AAR system provides information on the time of birth. Furthermore, the behavior of animals can be an indicator of the occurrence of environmental problems such as forest fires, poaching activities, impending earthquakes, or climate changes. Animals have physical reactions when they sense danger. For example, when an animal suddenly stops grazing and starts to run, it may indicate the presence of an external factor such as a predator or poacher. Therefore, with an AAR system, abnormal animal behaviors can be recognized by humans and can be utilized as an alarm or early warning. Moreover, the ability to automatically recognize animal activity from sensor data provides information about social interaction within a herd. Besides, AAR may improve livestock management by reporting a possible event to the farmer such as abnormal gait or posture of an animal, grazing behavior of an animal, or depressed states of animals. For example, some horse activities such as shaking, rolling, rubbing, and scratch biting are good indicators for the psychological states of horses. Manual observations of animal behaviors are time-consuming, labor-intensive, and prone to subjective judgments of individuals. Therefore, developing an automatic system by using sensors and machine learning is of significant importance for monitoring animals in an easy way. Besides, pet owners can use a real-time AAR system to keep track of the activities of their pets during the day.

This study focuses on *horse activity recognition* (HoAR) since monitoring the behavior of horses can yield significant knowledge about their mental and physical status. Manual observation of horse activities is expensive, time-consuming, and largely relies on subjective judgments of humans derived from previous experiences. To solve this problem, this study aims to propose an automated high-performance model for HoAR using wearable sensor data and machine learning techniques. Automated behavior classification for horses through sensors can provide rich insight into their health, welfare, reproduction, subjective state, survival, and interaction with humans and other animals. For example, it can be useful to recognize the movement patterns of horses to monitor how they are affected by changes in weather.

This study aims to be a comprehensive investigation of the impact of personalized/generalized information on the recognition of horse activities. For this purpose, we consider two different models:

- *Per-subject model* (personalized model): Both the training and testing set are all from the same horse. Each horse has unique characteristics of motions, such as speed and range corresponding to its physical properties (i.e., gender, age, weight, and height) and habits. For instance, the step frequency of an old horse is lower and the step length is shorter compared to a young horse. Therefore, it is natural to construct a classifier trained only on a single horse's activity data.

- *Cross-subject model* (generalized model): The training and testing data come from all the horses. In general, for a multi-class classification problem, a huge amount of training data is required, especially when the feature-vector dimension is high. Nevertheless, for the horse activity recognition task, the data collection and labeling processes are difficult. For this reason, the training data from a single horse may not sufficient to build robust machine learning models. Therefore, we may combine the training data of several horses in practice. A large amount of data from different horses would be of great use to achieve a high activity recognition accuracy. Besides, a generalized model is expected to be more reliable and robust than a simple model since it can defend against abnormal data interference. Furthermore, constructing a separate model for each horse is impractical, especially when a large number of horses has to be monitored. These are the motivations why we construct the generalized model. Hence, in this study, we also examined the cross-subject performance by checking whether a single predictive model trained on data from all horses is applicable to data from another horse.

The main contributions of this paper include the following aspects. (i) It is the first study that especially and deeply explores the use of *ensemble learning* methods to detect major activities of horses from big sensor data. (ii) Unlike the previous studies, our study investigated the accuracy for both *per-subject* (for each horse separately) and *cross-subject* (for all horses jointly) models for the first time. (iii) Our study achieved higher performance than the state-of-art studies carried out on the same dataset.

The experiments were conducted on a publicly available dataset. We proposed a new machine learning model by comparing different ensemble learning methods for horse activity recognition, including Bagging Trees, Extremely Randomized Trees (Extra Trees), Random Forest, Gradient Boosting (GBoost), Light Gradient Boosting (LGBoost), Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost). Especially, we aim to recognize five major horse activities with big data: walking, standing, grazing, galloping, and trotting. The experimental results showed that our solution achieved very good performance (94.62%).

The rest of the paper is organized as follows. Section 2 gives an overview of the literature. Section 3 explains the proposed horse activity recognition approach. Section 4 presents the description of the dataset used in this study and the details of the experiments along with their results. Section 5 gives the conclusion and some remarks for future work.

2. Related work

Horse activity is a sensitive indicator that may be potentially used to track horse health (Mao *et al.*, 2021; Eerdekens *et al.*, 2021). Monitoring the activity of horses (i.e., walking, eating) can provide useful information about horses' welfare, reproduction, survival, and interaction with humans and other animals (Nunes *et al.*, 2021; Alves *et al.*, 2021). To achieve this, animal activity recognition systems with the aid of wearable sensors and the use of machine learning methods over the gathered data have been developed in the past decade (Casella *et al.*, 2020; Eerdekens *et al.*, 2020; Braganca *et al.*, 2020). Some previous studies related to horse activity recognition (HoAR) are given in Table 1.

Automated HoAR has been widely studied with the use of different kinds of sensors such as accelerometers (Kamminga *et al.*, 2019a; Lee *et al.*, 2018), magnetometers (Gutierrez-Galan *et al.*, 2018; Lee *et al.*, 2016), and gyroscopes (Mao *et al.*, 2021; Braganca *et al.*, 2020). The sensors can be attached to horses in various locations such as the horse's neck (Mao *et al.*, 2021; Kamminga *et al.*, 2019a), leg (Eerdekens *et al.*, 2021, 2020), and wrist (Casella *et al.*, 2020) for collecting observations at a particular time point. Some studies (Lee *et al.*, 2018, 2016) show that the placement of sensors on the horse's body is possible to construct successful machine learning models. In the literature, previous work (Eerdekens *et al.*, 2021; Braganca *et al.*, 2020; Kamminga *et al.*, 2019a) showed that employing machine learning methods on wearable sensor data provided correct estimations on recognizing various horse activities such as walking (W), grazing (GR), standing (S), trotting (T), galloping (GL), cantering (C), rolling (R), pawing (P), flank watching (F), motionless (M), tölt (TL), pace (PC), paso fino (PF), and trocha (TR).

Until now, various machine learning methods have been utilized for HoAR such as Naive Bayes (NB) (Mao *et al.*, 2021; Kamminga *et al.*, 2019a; Lee *et al.*, 2016), Support Vector Machine (SVM) (Mao *et al.*, 2021; Alves *et al.*, 2021; Braganca *et al.*, 2020; Lee *et al.*, 2018, 2016), Decision Tree (DT) (Mao *et al.*, 2021; Casella *et al.*, 2020; Braganca *et al.*, 2020; Lee *et al.*, 2018), Random Forest (RF) (Eerdekens *et al.*, 2021; Alves *et al.*, 2021; Braganca *et al.*, 2020), K-Nearest Neighbors (KNN) (Casella *et al.*, 2020; Lee *et al.*, 2018), Neural Network (NN) (Alves *et al.*, 2021; Casella *et al.*, 2020; Braganca *et al.*, 2020; Gutierrez-Galan *et al.*, 2018; Lee *et al.*, 2016), Quadratic Discriminant Analysis (QDA) (Braganca *et al.*, 2020), Linear Discriminant Analysis (LDA) (Braganca *et al.*, 2020; Lee *et al.*, 2018), and Extreme Learning Machine (ELM) (Lee *et al.*, 2018). Furthermore, some deep learning methods have been investigated for HoAR such as Convolutional Neural Network (CNN) (Mao *et al.*, 2021; Eerdekens *et al.*, 2021; Alves *et al.*, 2021; Eerdekens *et al.*, 2020), Long-Short Term Memory (LSTM (Nunes *et al.*, 2021; Braganca *et al.*, 2020), Auto-Encoder (AE) (Lee *et al.*, 2018), and Recurrent Neural Network (RNN) (Nunes *et al.*, 2021).

Our work differs from the aforementioned studies in two important aspects. First, we assessed the use of ensemble learning methods to recognize the behaviors of horses from big data. Unlike previous works, we utilized different techniques such as Extremely Randomized Trees, Categorical Boosting, and Extreme Gradient Boosting. Second, we compared two approaches in terms of accuracy: the *per-subject approach* (building a model for each horse separately) and the *cross-subject approach* (building a single model for all horses jointly). The per-subject approach gives us a reliable insight into horse activities since the physical differences between horses make the wearable sensor data unique for each horse.

Table 1. Previous work related to the classification of horse activities.

Ref	Methods	Activities	Sensor Type	Sensor Location	#Subjects	#Samples
Mao *et al.* (2021)	NB, DT, SVM, CNN	GR, S, T, GL, W	Acc, Gry	Neck	6	87,621
Eerdekens *et al.* (2021)	CNN, RF	S, W, T, C, R, P, F	Acc	Leg	6	2,238
Nunes *et al.* (2021)	RNN, LSTM	Bite Chew	Video, Audio	Jaw	-	2,309
Alves *et al.* (2021)	RF, SVM, NN, CNN	Gaiting	Video, Audio	External	-	196
Casella *et al.* (2020)	NN, DT, KNN	W, T, C	Acc	Saddle, Wrist	2	140,000
Eerdekens *et al.* (2020)	CNN	S, W, T, C, R, P, F	Acc	Leg	6	959,075
Braganca *et al.* (2020)	LDA, QDA, DT, RF, SVM, NN, LSTM	W, T, C, TL, PC, PF, TR	Acc, Gry	Head, Withers, Pelvis, Limbs	120	7,576
Kamminga *et al.* (2019a)	NB	GR, S, T, GL, W	Acc	Neck	6	1,800,000
Lee *et al.* (2018)	SVM, LDA, DT, KNN, ELM, AE	W, T, C	Acc	Rider's Hip, Backbone, Elbow, Knee	1	2,400
Gutierrez-Galan *et al.* (2018)	NN	W, T, M	Acc, Gry, Magn	Collar	-	30,000
Lee *et al.* (2016)	SVM, NN, NB, Fuzzy NN	W, T, C	Acc, Gry, Magn	Rider's Hip, Breast, Neck, Elbow, Knee, Toe, etc.	1	80
Proposed approach	ET, BT, RF, GBoost, XGBoost, LGBoost, CatBoost	GR, S, T, GL, W	Acc	Neck	6	1,284,961

3. Material and methods

In an animal activity recognition system, big data is collected from sensors, stored, and then processed by machine learning methods. All three aspects of big data are at work – the big *volume*, *variety*, and *velocity*. The *volume* of the data is large since the data is collected in a continuous period of about 2 seconds for an animal performing an activity. From the perspective of *variety*, data is collected from various types of sensors such as accelerometers, gyroscopes, magnetometers, and biosensors (heart rate, blood pressure, and temperature). The *velocity* of data flow is also rapid since raw data is collected with a high sampling rate, e.g. 100 Hz, 1 s of data equals 100 rows.

The flowchart of the proposed horse activity recognition approach is shown in Figure 1. As the first step, the approach includes raw data collection via wearable sensors which are attached to horses. After that, the obtained data is stored in a data source and labeled (annotated) with activities. In the *data selection* stage, the relevant records belonging to some horses and activities are retrieved from the data source for the machine learning task. In the next stage, named *data preprocessing*, sensor data is filtered by a low-pass filter technique with the aim of avoiding noise data. In the *feature extraction* phase, data is segmented into small partitions by utilizing a windowing technique, and features are automatically extracted for each window such as max, min, mean, covariance, skewness, kurtosis, entropy, and energy. After that, data is split into training and test set. Alternative ensemble learning algorithms are applied to the training data for model construction. The predictive models are evaluated on testing data in terms of classification rate (accuracy). Finally, the best-fit model is used to recognize horse activities (i.e., walking, standing, grazing, galloping, and trotting) on new unseen data.

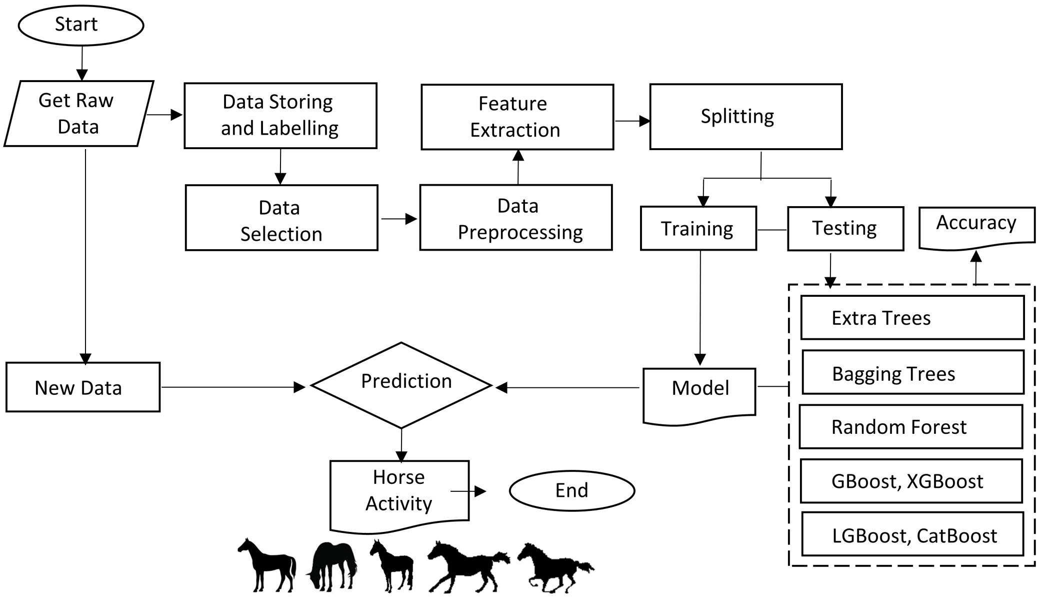


Fig. 1. Flowchart of the proposed horse activity recognition approach.

Figure 2 shows some samples from accelerometer data that were collected during different horse activities over time. As can be seen, each movement has different characteristics. For example, *walking* is often more periodic for horses than *standing*. *Trotting* requires higher power and effort than *walking* since it involves more intense muscle contractions. It can also be seen that a higher z-direction of the accelerometer for *galloping* is observed. Moreover, a horse may require different relaxation times when *grazing*.

In this study, both *personalized* and *generalized* models were built by using different ensemble learning methods to perform horse activity recognition. Each horse has a unique behavior habit during performing an activity since the behavior habit consists of strength, speed, and other characteristics of the horse's body. Each horse has its own movement pattern, which possibly affects wearable sensing HoAR data. Physical characteristics, such as weight and height, affect the electromagnetic wave propagation around the horse's body. This in turn creates unique models. Based on this insight, we built individual models for each horse separately. We can build a model for each horse that contains personalized information; however, we have to deal with sufficient data, robustness, and managing many models. Sometimes, an adequate amount of personalized data cannot be gathered, especially when a large number of horses has to be monitored. On the other hand, we can build a single model for all the horses. A generalized model can tolerate the differences among horses regarding horse activity recognition. Furthermore, managing a single model is more practical. Moreover, training data from other horses can help reduce the burden on data acquisition. Considering these facts, in this study, we examined the impacts of both personalized and generalized data on the performance of HoAR.

Animal activity recognition from horses is a hard technological task because of several issues that need to be solved: (i) the development of an efficient wearable device, called Inertial Measurement Unit (IMU), to attach to the horse, i.e., low energy-consumption, high battery life, and ergonomically designed, (ii) the implementation of wireless communication to collect data from the device, and (iii) storing and processing data using machine learning methods.

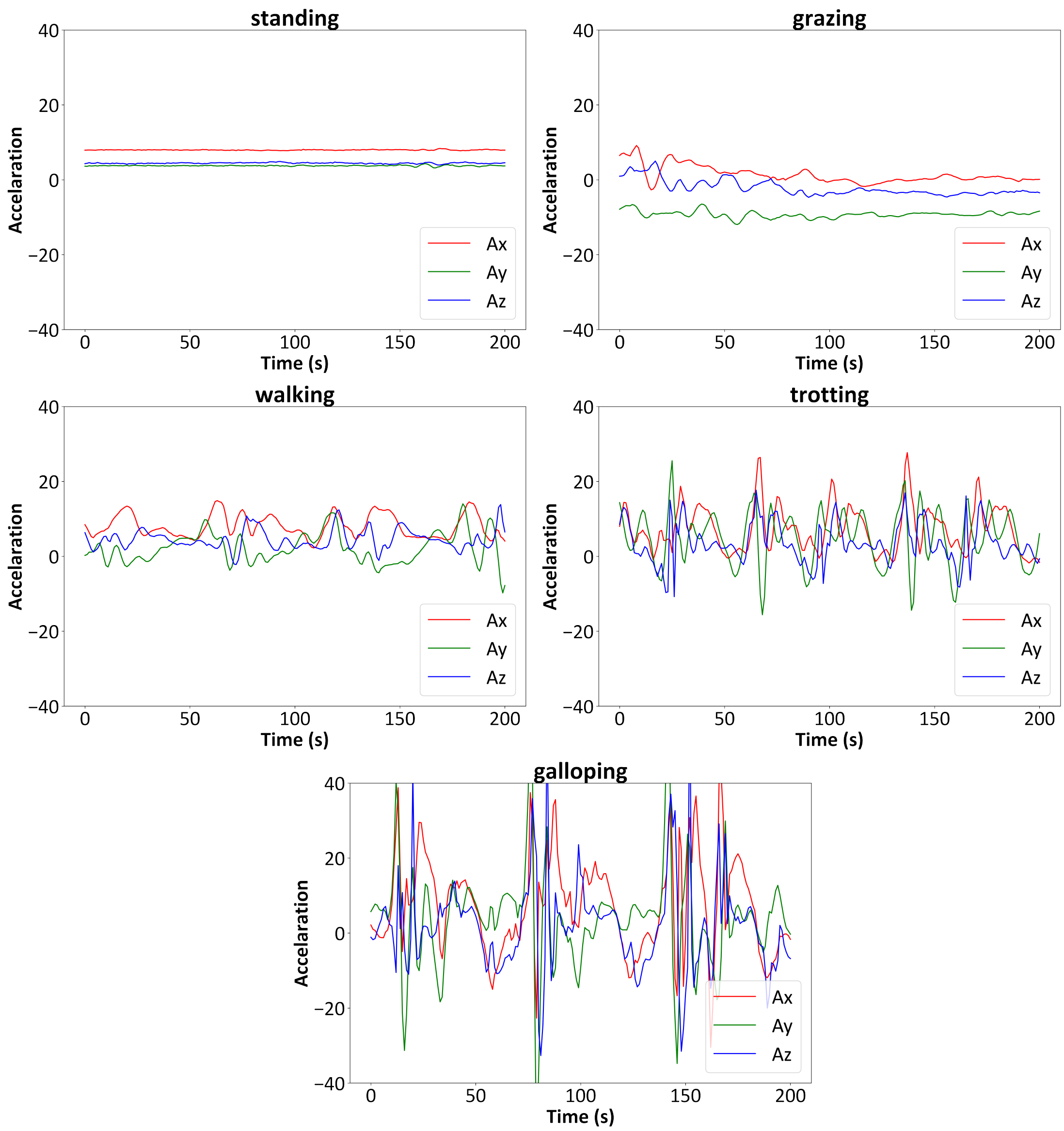


Fig. 2. A typical example of the x-, y-, and z-axis of the 3D- accelerometer sensor data from different horse activities.

HoAR provides a potential for the following purposes:

- *Health*: Automatically tracking animal behavior can be a valuable solution for detecting potential health problems at an early stage by identifying hidden signs in horse activities. In other words, monitoring activities via a HoAR system can provide significant information about the horse's health since the irregularities of the activities may be because of health problems of the horse. This can be achieved by assessing specific links between horse activities and health issues. For example, tracking pain-related horse activities such as rolling and pawing may act as a health indicator such as colic. Similarly, the irregularities of the horse gaits may be due to medical problems and monitoring canter lead allows early identification of health disorders. Since foraging and eating are also health-related behaviors for horses, they can also be detected through a sensor-based HoAR system.

- *Horse training*: Gait analysis during training such as cantering and walking provide useful knowledge in training and can be used to improve the performance of sport horses. For example, how much time a horse has spent at each activity during a training section. This information is important since balancing the time spent at each activity based on the goals of training is a primary requirement in horse training.

- *Rider coaching*: Horse riding is a complex sport involving all the movements of galloping and walking on horseback and requires coaching. The expensive and regularly long coaching requirements make the horse riding sport difficult for a normal person to approach. Therefore, a self-coaching system can be developed for horse riding by analyzing both the rider's and horse's motions. Here, the aim is to design a real-time coaching mechanism for horse riding by giving feedback to the rider about movements and postures. This can be achieved by recognizing some horse-related tasks; for example, how many jumps the horse has taken or how many movements (i.e., pirouette, passage, piaffe) the horse has done. The rider may be coached in the accurate motion by recognizing horse gaits. Furthermore, the horse activities recognized by an expert rider and an amateur rider can be compared since the actions of an expert can be an example for a beginner.

- *Reproduction*: Monitoring a pregnant mare's movements through a HoAR system provides information on the time of birth. Furthermore, a HoAR system enables analyzing highly accurate phenotypes of horse gaits for breeding and genetic research.

- *Injuries*: The fitness level of a sport horse is a major determinant of injuries, and thus monitoring horse movements via a HoAR system during exercise sessions can provide significant information in the performance evaluation and predicting possible future injuries for a sport horse.

- *Foraging*: The monitoring of grazing activity is significant for considering horse management. However, human-based monitoring is a time-consuming and expensive task. Automatically tracking foraging behavior via a HoAR system can be cheaper and can give vital information to livestock managers. For example, modeling grazing behavior can help in decision-making about pasture use. Furthermore, it provides significant information on the use of food resources.

- *Psychology*: The horse is an animal with unique psychological features that can be recognized by specific movements since the horse's activities are orchestrated by the nervous system. Some horse activities such as shaking, rolling, rubbing, and scratch biting are good indicators for the psychological states of horses. Therefore, automatically tracking these horse movements via a HoAR system can provide rich insight into their psychological well-being. Monitoring activities can be performed for determining if a horse is depressed or agitated.

4. Experimental studies

In this study, we implemented a HoAR application with the Python programming language. We used several libraries (i.e., Pandas, NumPy, Pyplot) for data manipulation and drawing processes. Furthermore, we utilized the Sklearn library to construct predictive models. The models were validated by utilizing the k-fold cross-validation method, where the data is split into k subsets, then $k-1$ subsets are utilized for training and the remaining set for testing. The whole process is repeated k times with each subset serving once as a test set and the final accuracy is calculated by averaging the accuracies obtained at each iteration. In this work, the number of folds (k) was set to 10 because of low variance and low bias in the data.

In this study, we investigated the performances of *per-subject* and *cross-subject* models. From one point of view, building a separate model for each horse is logical since each horse has a unique body movement and activity style. In other words, each horse has its own discriminative and specific movement patterns, which means that data from each horse should be used in both training and testing phases. On the other hand, building a general model for all horses is more practical and robust. There may be two reasons: first, the generalized model has been built on more training data compared to individual horse data; second, aberrant samples, which may be caused by noise interference, can be tolerated by a generalized model. The augmented training data can help eliminate the influence of abnormal samples and enhance the robustness of a HoAR system. Therefore, as wearable sensor data are collected from multiple horses, a predictive model can cover a large variety of activity executions. In summary, personalized data can be used to promote recognition accuracy; on the other hand, generalized data can help to strengthen the stability of abnormal samples in the training data. Based on the aforementioned analysis, we investigated whether the personalized or generalized information can achieve the horse activity recognition task with acceptable accuracy.

4.1 Dataset description

In this study, the experiments were carried out on a publicly-available big data (Kamminga *et al.*, 2019b). The complete dataset involves 2-second data samples collected from 18 horses (subjects) when performing 17 different activities within a period of seven days. The data were gathered during both horse riding sessions and while the horses moved freely in an outdoor pasture. The sensor module was placed on a collar, which was wrapped around the neck of a horse, especially attached to the manes of the horses by utilizing an elastic band. The sampling rate was 100 Hz for both the tri-axial gyroscope and accelerometer, and 12 Hz for the tri-axial magnetometer when recording different horse activities. Sensor orientation was not strictly fixed to provide a robust evaluation. The raw data was organized into segments. Each one has a unique ID and different length that depends on how long the horse exercised a particular activity. For the ground truth, video data was also collected by cameras. During the labeling process, sensor data were synchronized with video data using metadata. To minimize mistakes in the labeling process, both the raw sensor data and video data were visually inspected and annotated by a human.

Data selection: In this study, the commonly labeled six horses (namely Bacardi, Driekus, Galoway, Happy, Patron, and Zafir) and five activities (walking, standing, grazing, galloping, and trotting) were used to provide sufficient accelerometer data for training.

Data preprocessing: The rider and natural activities (i.e., trotting-rider and trotting-natural) were combined since they are the same type of activity. A part of the data is unlabeled (denoted by unknown and null). Since we employed supervised learning, which utilizes data labels, we removed the unlabeled part of the dataset. To remove noise from the raw data, we used the low-pass Butterworth filter method with a cut-off frequency of 30 Hz.

Feature extraction: Raw sensor data usually doesn't have sufficient information in its own format to describe a horse activity since it includes specific values obtained at a certain time instant. Therefore, we extracted features by transforming the raw sensor data into more informative and statistical values. Table 2 shows the features extracted from the raw wearable sensor data. Features were extracted with a sliding window of two seconds with 50% overlap.

4.2 Experimental results

This work determined the best EL algorithm in terms of accuracy by comparing their performances on horse activity recognition. We compared seven methods: Bagging Trees, Extremely Randomized Trees (Extra Trees), Random Forest (RF), Gradient Boosting (GBoost), Light Gradient Boosting (LGBoost), Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost). We used the default parameter settings for all algorithms.

Table 3 presents the accuracy rates obtained by different ensemble learning (EL) algorithms on each horse dataset. All the experimental results range between 87.18% and 97.67%, hence, it is possible to say that EL techniques have the ability to achieve good accuracy in HoAR. The accuracy values obtained for the horse named Bacardi were lower than other horses, because, the size of data collected from this horse is the smallest, not sufficient enough like others. It is clearly seen that the highest accuracy rate (97.67%) was reached by the GBoost and Random Forest methods for the horse named Driekus. This is probably due to the fact that GBoost has some powerful properties that can provide solving machine learning problems in a correct way, i.e., the flexibility of choosing a loss function that fits the classification task to be solved, reducing the largest errors at each level of training, and not requiring feature normalization. The Random Forest method has also many advantages such as its resistance to outliers, adaptive feature selection, and ability to deal with high dimensional data. The high accuracy results in Table 3 indicate that the proposed approach is efficient and accurate for practical application.

Table 2. Features extracted from raw sensor data.

ID	Feature
F1	Min
F2	Max
F3	Mean
F4	Median
F5	Standard Deviation
F6	Covariance
F7	Percentile (25)
F8	Percentile (75)
F9	Number of Zero Crossing
F10	Peak-to-Peak Value
F11	Root Mean Squared (RMS)
F12	Kurtosis
F13	Skewness
F14	Crest Factor
F15	Sample Entropy
F16	Spectral Entropy
F17	Energy

Table 3. Comparison of per-subject and cross-subject models built with different machine learning algorithms in terms of accuracy (%).

Dataset	Random Forest	Bagging Trees	Extra Trees	GBoost	XGBoost	LGBoost	CatBoost
Bacardi	87.22	87.18	88.01	88.75	88.50	88.81	88.14
Driekus	97.67	96.77	97.44	97.67	97.16	97.41	97.29
Galoway	93.12	92.55	93.09	92.99	93.46	93.32	92.84
Happy	96.55	95.92	96.62	96.34	96.52	96.44	96.56
Patron	96.13	95.82	96.23	95.80	96.11	96.01	96.04
Zafir	96.21	95.12	96.35	95.99	95.85	95.53	95.99
Per-Subject Model	94.48	93.89	**94.62**	94.59	94.60	94.59	94.48
Cross-Subject Model	92.46	91.62	92.02	92.42	92.57	**92.65**	92.37

As can be observed in Table 3, each horse has different classification accuracy due to its own characteristics. This result indicates that each horse has its own discriminative and specific movement patterns. Comparing the per-subject model with the cross-subject model, we can observe a slight improvement in performance in the personalized model. For example, by bagging trees, accuracy rates for the personalized model and generalized model were reached 93.89% and 91.62%, respectively. This is probably because of the fact that horses usually show different movement patterns even for the same type of activity since they differ in body size, age, gender, and other physiological properties. Recognition performance is higher in the personalized model compared to the generalized model because it is more like a horse-dependent validation. Therefore, for the horse activity recognition task, it is better to use individual horse data. In other words, training data from each horse would be of great use to achieve a high activity recognition accuracy. The highest accuracy values (94.62% and 92.65%) on average were obtained by the Extra Trees and LGBoost methods for the per-subject and cross-subject models, respectively.

For cross-subject models, LGBoost achieved the highest accuracy since it has a good generalization ability. The main difference between LGBoost and other gradient boosting algorithms is that it grows a tree vertically while others grow a tree horizontally. This means that LGBoost grows a tree leaf-wise while other algorithms grow a tree level-wise. A leaf-wise algorithm can reduce more loss than a level-wise algorithm because it chooses the leaf with max delta loss to grow. Furthermore, two important strategies are used in LGBoost: Exclusive Feature Bundling (EFB) and Gradient-based One-Side Sampling (GOSS). Thanks to the GOSS property, the algorithm allows the classifier to learn from the data points with larger gradients while randomly sampling out data points with lower gradients. In this way, a quite accurate estimation can be obtained since data points with larger gradients play a more significant role in the calculation of information gain. Furthermore, thanks to the EFB property, the algorithm selects the most relevant and mutually exclusive features from the whole feature space. In this way, it reduces the dimensionality for improving both efficiency and prediction accuracy. Besides, LGBoost is faster than other gradient-based algorithms and uses much less memory to run, and therefore, it is suitable to analyze large-scale data.

For per-subject models, Extra Trees showed its superiority over other methods in terms of accuracy since it has different behaviors and advantages over others. The main advantages of Extra Trees are the reduction in the variance of the tree and computational efficiency, therefore higher prediction ability. There are two main differences between the Extra Trees method and other methods. First, the Extra Trees method uses all the training samples rather than a bootstrap replica to minimize bias when growing the trees. Second, it completely randomly selects both attribute and cut-points to split nodes, which reduces variance more strongly than the weaker randomization and deterministic approaches used by other methods. In brief, it provides an advantage to the personalized models by decreasing variance resulting from stronger tree randomization.

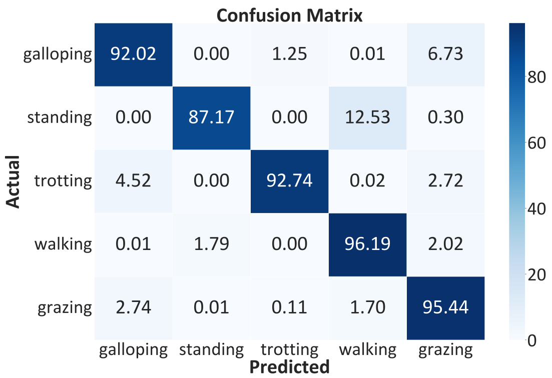
In addition to accuracy, we also compared the performances of per-subject and cross-subject models in terms of precision, recall, and F1-measure as given in Table 4. The values of these metrics are ranged between 0 and 1, where 1 is the best value. As can be seen from Table 4; the precision values of per-subject models are closer to 1 than cross-subject models. For example, by the Random Forest algorithm, precision values for the personalized model and generalized model were observed as 0.9451 and 0.9242, respectively. This indicates that it is better to use individual horse data for a better horse activity recognition task. It is clearly seen that the Extra Trees and LGBoost methods achieved better performance on average than other methods for the per-subject and cross-subject models, respectively. The same achievement was also observed for recall and F1-measure metrics on average. For example, Extra Trees (0.9466) outperformed Bagging Trees (0.9378) on average in terms of recall value. All the F1-measure values range between 0.8660 and 0.9769, hence, it can be concluded that the proposed models can be successfully used to recognize the movement patterns of horses.

Figure 3(a) presents the confusion matrix of the best per-subject model built by the Extra Trees method on average and Figure 3(b) gives the confusion matrix of the best cross-subject model built by the LGBoost method to show the performances of the algorithms on each horse activity separately. According to the matrices, it is possible to say that the models usually had no difficulty in distinguishing horse activities. For example, 94.75% of "*walking*" activities were correctly classified by the cross-subject model. Though horse activities were identified well in general, the "*standing*" activity was confused slightly by the "*walking*" activity during estimation. The largest difference between per-subject and cross-subject models appeared for the "*galloping*" activity (92.02% vs 84.13%), followed by the "*standing*" activity (87.17% vs 81.29%). This means that the physical conditions and habits of horses mostly affect these activities. On the other hand, the performances of the models in distinguishing other activities (trotting, walking, and grazing) are close to each other. It is clearly seen that the best accuracy (96.19%) was achieved on the "*walking*" activity by the per-subject model.

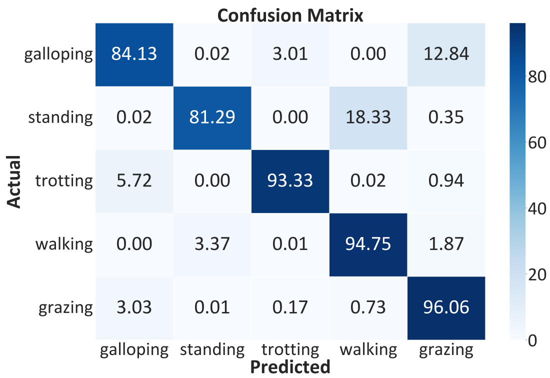
1

Table 4. Comparison of per-subject and cross-subject models in terms of precision, recall, and F1-measure.

Dataset	**Metric**	**Random Forest**	**Bagging Trees**	**Extra Trees**	**GBoost**	**XGBoost**	**LGBoost**	**CatBoost**
Bacardi	Precision	0.8723	0.8664	0.8855	0.8895	0.8857	0.8886	0.8819
	Recall	0.8722	0.8665	0.8846	0.8882	0.8851	0.8882	0.8815
	F1-measure	0.8722	0.8660	0.8847	0.8886	0.8853	0.8883	0.8816
Driekus	Precision	0.9773	0.9699	0.9748	0.9773	0.9726	0.9751	0.9738
	Recall	0.9767	0.9692	0.9739	0.9769	0.9717	0.9741	0.9729
	F1-measure	0.9768	0.9693	0.9739	0.9769	0.9718	0.9742	0.9730
Galoway	Precision	0.9309	0.9256	0.9304	0.9296	0.9343	0.9330	0.9281
	Recall	0.9312	0.9255	0.9308	0.9300	0.9346	0.9332	0.9285
	F1-measure	0.9310	0.9255	0.9306	0.9298	0.9344	0.9331	0.9282
Happy	Precision	0.9655	0.9585	0.9659	0.9636	0.9653	0.9645	0.9656
	Recall	0.9655	0.9582	0.9660	0.9635	0.9653	0.9644	0.9657
	F1-measure	0.9654	0.9583	0.9658	0.9635	0.9652	0.9644	0.9655
Patron	Precision	0.9613	0.9566	0.9616	0.9584	0.9612	0.9602	0.9606
	Recall	0.9614	0.9561	0.9618	0.9585	0.9611	0.9601	0.9605
	F1-measure	0.9612	0.9562	0.9615	0.9583	0.9610	0.9600	0.9604
Zafir	Precision	0.9630	0.9526	0.9630	0.9613	0.9593	0.9570	0.9604
	Recall	0.9621	0.9510	0.9623	0.9607	0.9586	0.9553	0.9599
	F1-measure	0.9623	0.9513	0.9624	0.9608	0.9587	0.9555	0.9600
Per-Subject Model	Precision	0.9451	0.9383	**0.9469**	0.9466	0.9464	0.9464	0.9451
	Recall	0.9449	0.9378	**0.9466**	0.9463	0.9461	0.9459	0.9448
	F1-measure	0.9448	0.9378	**0.9465**	0.9463	0.9461	0.9459	0.9448
Cross-Subject Model	Precision	0.9242	0.9184	0.9214	0.9238	0.9257	**0.9264**	0.9235
	Recall	0.9246	0.9183	0.9217	0.9243	0.9256	**0.9265**	0.9237
	F1-measure	0.9240	0.9182	0.9210	0.9238	0.9252	**0.9262**	0.9232



(a)



(b)

Fig. 3. The confusion matrix of (a) the best per-subject model on average (Extra Trees), and (b) the best cross-subject model (LGBoost).

4.3 Comparison with the state-of-the-art studies

In order to show the superiority of our proposed method, we compared it with the state-of-the-art methods in the literature, such as Cross-Modality Interaction Network (CMI-Net) (Mao *et al.*, 2021), Variational AutoEncoder (VAE) (Voorend , 2021), Deep Neural Network (DNN) (Spink *et al.*, 2022; Bocaj *et al.*, 2020), and Multivariate Long Short-Term Memory Fully Convolutional Network (MLSTM-FCN) (Huveneers, 2021). Table 5 shows the related studies along with the methods and the corresponding accuracy rates. The results were directly taken from the referenced study since the researchers used the same dataset as our study. It is clearly seen that our model achieved higher accuracy (94.62%) than the previous models built on the same dataset. It can be concluded from Table 5 that our method outperformed the other methods with a 9.72% improvement on average.

Table 5. Comparison of our result against the results of the state-of-the-art studies on the same dataset.

Reference	Year	Method	Accuracy (%)
Spink *et al.*	2022	Deep Neural Network (DNN) (without active learning)	72.30
		Deep Neural Network (DNN) (with least confident)	70.30
		Deep Neural Network (DNN) (with max disagreement)	69.70
Mao *et al.*	2021	Naive Bayes	76.60
		Decision Tree	88.83
		Support Vector Machine	89.65
		Cross-Modality Interaction Network (CMI-Net) + Softmax Cross-Entropy (CE) Loss	93.37
		Cross-Modality Interaction Network (CMI-Net) + Class-Balanced (CB) Focal Loss	90.68
Huveneers	2021	Long Short-Term Memory (LSTM)	87.70
		Neural Network (NN)	85.30
		Multivariate Long Short-Term Memory Fully Convolutional Network (MLSTM-FCN)	88.50
Voorend	2021	Variational AutoEncoder (VAE) (with feature extraction)	86.00
		Variational AutoEncoder (VAE) (no feature extraction)	81.00
Bocaj *et al.*	2020	Deep Convolutional Neural Networks (ConvNet)	91.27
Kamminga *et al.*	2019	Naive Bayes (no tuning or balancing)	88.50
		Naive Bayes (tuning)	89.50
		Naive Bayes (balanced)	89.00
		Naive Bayes (tuned and balanced)	90.00
Our study		Extra Trees	94.62

5. Conclusion and future work

Horse activity is a reliable and useful indicator, which can be potentially used to track their health, welfare, reproduction, and interaction with humans. This work is concerned with the classification of horse activities using big data collected via wearable modules attached to horses. In the experiments, five horse activities (walking, standing, grazing, galloping, and trotting) were identified using the sensor data. Seven ensemble learning methods were compared to each other in terms of accuracy to detect the best one in predicting horse activities, including Extra Trees, Bagging Trees, Random Forest, GBoost, XGBoost, LGBoost, and CatBoost.

Our study is original in that it compares the accuracies of per-subject and cross-subject models. The experimental results showed that per-subject models achieved very good performance (94.62%) on average on a real-world dataset. The high accuracy results obtained in this study indicate that the proposed approach is efficient and accurate for practical application. The findings of this study are expected to help in the design of HoAR applications.

In future studies, the proposed model can be extended by using different sensors such as a gyroscope and magnetometer. Even for the same activity, different horses have distinct movement modes since their physical conditions (i.e., age, weight, height, and gender) and habits affect their movements. This means that only a horse's own movement data can reflect his own activity completely and accurately. The fact that we got the best performance in the per-subject (personalized) model can be evidence to support this conclusion. This conclusion brings us another question: can we identify horses according to their movements? In other words, can we extract the activity fingerprint for authentication? This question (horse identification) can be explored in future work. Furthermore, in the future, different predictive models can be built for other animal species such as sheep or dogs.

References

Al-Dousari, A., Qurban, M., Hussain, I., & Al-Hajeri, M. (2021). Development of novel hybrid models for the prediction of COVID-19 in Kuwait. *Kuwait Journal of Science*, special issue, 1-6.

Alves, A. A. C., Andrietta, L. T., Lopes, R. Z., Bussiman, F. O., Silva, F. F., Carvalheiro, R., Brito, L. F., Balieiro, J. C. C., Albuquerque, L. G., & Ventura R. V. (2021). Integrating audio signal processing and deep learning algorithms for gait pattern classification in Brazilian gaited horses. *Frontiers in Animal Science*, **2**, 1-19.

Braganca, F. M. S., Broome, S., Rhodin, M., Björnsdottir, S., Gunnarsson, V., Voskamp, J. P., Persson-Sjodin, E., Back, W., Lindgren, G., Novoa-Bravo, M., Gmel, A. I., Roepstorff, C., Zwaag, B. J., Weeren, P. R., & Hernlund, E. (2020). Improving gait classification in horses by using inertial measurement unit (IMU) generated data and machine learning. *Scientific Reports*, **10**, 1-9.

Bocaj, E., Uzunidis, D., Kasnesis, P. & Patrikakis, C. Z. (2020). On the benefits of deep convolutional neural networks on animal activity recognition. In *Proceedings of the International Conference on Smart Systems and Technologies* (pp. 83-88). Osijek, Croatia: IEEE.

Casella, E., Khamesi, A.R. & Silvestri, S. (2020). A framework for the recognition of horse gaits through wearable devices.*Pervasive and Mobile Computing*, **67**, 1-19.

Colak, C., Colak, Mehmet C., Ermis, N., Erdil, N., & Ozdemir R. (2016). Prediction of cholesterol level in patients with myocardial infarction based on medical data mining methods.*Kuwait Journal of Science*, **43**(3), 86-90.

Datsi, T., Aznag, K., & Oirrak, A. E. (2021). Digit recognition using decimal coding and artificial neural network. *Kuwait Journal of Science*, **49**(1), 1-17.

Eerdekens, A., Deruyck, M., Fontaine, J., Martens, L., Poorter, E. D., & Joseph, W. (2020). Automatic equine activity detection by convolutional neural networks using accelerometer data. *Computers and Electronics in Agriculture*, **168**, 1-11.

Eerdekens, A., Deruyck, M., Fontaine, J., Martens, L., Poorter, E. D., Plets, D., & Joseph, W. (2021). A framework for energy-efficient equine activity recognition with leg accelerometers. *Computers and Electronics in Agriculture*, **183**, 1-13.

Gutierrez-Galan, D., Cerezuela-Escudero, E., Dominguez-Morales, J. P., Rios-Navarro, A., Rivas-Perez, M., Tapiador-Morales, R., Dominguez-Morales, M., Linares-Barranco, A. & Jimenez-Fernandez, A., (2018). Embedded neural network for real-time animal behavior classification. *Neurocomputing*, **272**, 17-26.

Huveneers, M. (2021). The development of a horse activity recognition algorithm. *Thesis, University of Twente* (pp. 1-96). Enschede, Netherlands.

Jain, V. K., & Kumar, S. (2018). Rough set based intelligent approach for identification of H1N1 suspect using social media. *Kuwait Journal of Science*, **45**(2), 8-14.

Kamminga, J. W., Meratnia, N. & Havinga, P. J. M. (2019). Dataset: horse movement data and analysis of its potential for activity recognition. In *Proceedings of the 2nd Workshop on Data Acquisition to Analysis* (pp. 22-25). New York, USA: ACM.

Kamminga, J. W., Janßen, L. M., Meratnia, N., & Havinga, P. J. M. (2019). Horsing around-a dataset comprising horse movement. *Data*, **4**(4), 1-13.

Khan, K., Ullah, A., & Baharudin, B. (2016). Pattern and semantic analysis to improve unsupervised techniques for opinion target identification. *Kuwait Journal of Science*, **43**(1), 129-149.

Lee, J-N., Lee, M-W., Byeon Y-H., Lee W-S., & Kwak, K-C. (2016). Classification of horse gaits using FCM-based neuro-fuzzy classifier from the transformed data information of inertial sensor. *Sensors*, **16**(5), 1-18.

Lee, J-N., Byeon, Y-H., & Kwak, K-C. (2018). Design of ensemble stacked auto-encoder for classification of horse gaits with MEMS inertial sensor technology. *Micromachines*, **9**(8), 1-17.

Mao, A., Huang, E., Gan, H., Parkes, R. S. V., Xu, W., & Liu, K. (2021). Cross-modality interaction network for equine activity recognition using imbalanced multi-modal data. *Sensors*, **21**(17), 1-17.

Mehmood, Z., Anwar, S. M., & Altaf, M. (2018). A novel image retrieval based on rectangular spatial histograms of visual words. *Kuwait Journal of Science*, **45**(1), 54-69.

Nallamuth, R., & Palanichamy, J. (2015). Cyber deception: Virtual networks to defend insider reconnaissance. *Kuwait Journal of Science*, **42**(2), 189-205.

Nunes, L., Ampatzidis, Y., Costa, L., & Wallau, M. (2021). Horse foraging behavior detection using sound recognition techniques and artificial intelligence. *Computers and Electronics in Agriculture*, **183**, 1-8.

Oshinubi, K., Al-Awadhi, F., Rachdi, M., & Demongeot, J. (2021). Data analysis and forecasting of COVID-19 pandemic in Kuwait based on daily Observation and basic reproduction number dynamics. *Kuwait Journal of Science*, special issue, 1-30.

Spink, S. J., Kamminga, J. W. & Kamilaris, A. (2022). Improving the annotation efficiency for animal activity recognition using active learning. In *Proceedings of the 12th International Conference on Methods and Techniques in Behavioral Research (Measuring Behavior 2022)* (pp. 1-6). Krakow, Poland: Noldus.

Voorend, R. (2021). Deep unsupervised representation learning for animal activity recognition. *Thesis, University of Twente* (pp. 1-70). Enschede, Netherlands.

Submitted:	18/03/2022
Revised:	21/05/2022
Accepted:	23/05/2022
DOI:	10.48129/kjs.splml.19571