Ethogram-based automatic wild animal monitoring through inertial sensors and GPS data

Jessica Leoni[a,*], Mara Tanelli[a], Silvia Carla Strada[a], Tanya Berger-Wolf[b]

[a] *Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano, Italy*
[b] *Department of Computer Science, University of Illinois at Chicago, 851 S. Morgan St, Chicago, Illinois*

Abstract

Direct monitoring of wild animals' behavior is challenging and data tampering. Instrument the animals with collars that embeds sensors, such as tri-axial accelerometer and GPS, allows obtaining sufficient information for remotely classifying the performed activities. In this work is presented an accurate and human intelligible framework, designed leveraging the authors' skills in machine-learning and data analysis. The system covers all the steps required to accurately map the raw signals to the activities carried out, grouped in pre-processing, features extraction and selection, and classification phases.

A case of study consists of a dataset collected by the Crofoot Lab at the Mpala Centre, in Kenya, instrumenting free-ranging Olive Baboons. This dataset provides both sensors time-series paired with respective activity labels and unlabeled ones. Labeled data was used to tune the parameters of the framework phases and to train and test the employed boosted-trees classifiers, while unlabeled ones were used for further system validations. The average accuracy obtained on a single activity is 94.5%. At best of the authors knowledge, this is the first work that aims to solve the problem of direct human monitoring with such high accuracy, outperforming the state of the art by a lift about 10%.

Another novelty is given by the attention paid to the ethologist's needs. Together with the predictions, the framework also returns a ranking for the features considered, based on their importance in the decision-making process of the classifier. Therefore, the extracted features are consistent with the logical human path that the ethologist follows in performing direct monitoring. The produced framework has also been designed consistently with the ethogram structure to be easily interpretable and to allow activities classification at different aggregation levels.

Keywords: Animal Behavior, Ethogram, Pre-processing, Machine-learning, Ensemble tree

---

*Corresponding author

Email addresses: jessica.leoni@polimi.it (Jessica Leoni), mara.tanelli@polimi.it (Mara Tanelli), silvia.strada@polimi.it (Silvia Carla Strada), tanyabw@uic.edu (Tanya Berger-Wolf)

1. Introduction

Animal monitoring is a fundamental phase in the process of understanding individual behavior and social dynamics that characterize each species. Observing the species of interest ethologists are able to understand their needs and health status but also to study phenomena such as habitat utilization [1] or influence of external conditions on their normal behavior [2], [3].

To date, animal monitoring is directly performed by the ethologist. Its task in this phase is to observe the individuals of the species of interest, annotating over time the activities carried out, choosing from those proposed by the reference ethogram. An ethogram is defined as the set of natural behavior for an individual, each with its own description [4]. Direct observation causes several problems [5], as the presence of man is perceived by the animals observed. The perceived state of alert causes data tampering, as it involves a deviation from normal behavior [6]. In addition, periods of prolonged observation are necessary to collect a consistent dataset, but increase the risk of distraction, not to mention that human reflexes could introduce a delay in recording transitions from one activity to another [7] and [8]. In addition, as the number of species observed increases, the number of required ethologists will inevitably increase, which is a major logistical problem.

Recent developments in wearable tracking technologies and machine-learning algorithms pave the way for animal automatic remote monitoring systems. The devices used to instrument animals are usually composed of a GPS, for geolocation, and tri-axial accelerometers, suitable for discriminating the different activities [9]. In literature, the best results are obtained in the automatic recognition of activities carried out by captive animals, mainly cattle as cows [10], or sheep [11]. Less satisfactory are the performances of systems for automatic remote monitoring of wild animals. The different scenario offers major challenges, such as the inability to instrument the entire area or to repair the sensors in case of damage. Although good results have been obtained in the discrimination between stationary and non-stationary activities in free-ranging animals [12], by decreasing granularity and expanding the pool of recognized activities, the accuracy of classification models falls significantly [13]. At best of the authors' knowledge, the best performance in wild animal monitoring activities is between 69.0% [14] and 77.8% [15].

The proposed framework aims to support automatically and effectively the process of remote animal monitoring, returning to the ethologist accurate prediction, each combined with an interpretable and comprehensive explanation of the decision-making process that produced it. In detail, the innovative contributions brought by the work here present are:

- A raw-data pre-processing, conduced both in time and frequency domain. In this step data analysis techniques as noise filtering and bias removal has been leveraged to ease the following framework phases;

- Features selection properly designed to match ethologist needs. In fact, all the features used by the system are coherent with those considered by the ethologist in the direct performing of animal monitoring. Also, just the minimum number of features has been selected, taking into account the well-known *Curse of Dimensionality Problem* [16];

- An inconsistent observation detector, to augment prediction reliability and robustness A classifier has been properly trained to identify outliers and anomalies in the data, for which mislabeling is highly probable;

- Interpretability of the classification model. In fact, it has been structured so as to be consistent with the ethogram. This also allows to classify activities at different aggregation level;

- A classifier whose decision-making process is transparent to the ethologist. In fact, each prediction produced is associated with a ranking of features, based on the importance they had in the outcome computation.

The performance the proposed framework outperforms the ones reported in the literature by a lift about 10%, revealing that it constitutes valid support for the ethologist, accelerating his work in the identification phase, allowing him to focus more on the phases of inference and analysis, aware that he can rely on accurate monitoring.

The rest of the document is organized as follows: Section 2 illustrates framework implementation and explain the choices that have allowed us to overcome the issues related to wild animal monitoring. Section 3 describes the case study used to validate the produced framework. Evaluation results are reported in Section 4 and then discussed in Section 5. Finally, the conclusions are presented in Section 6.

2. Methods

This section describes the three main phases of the proposed framework: pre-processing, features extraction and selection, and classification. Particular attention is paid to justify the implementation choices, explaining how they aim to solve the problems related to the addressed scenario.

2.1. Pre-Processing

The dataset provided consists of time-series measured by the sensors used to instrument the animals, usually tri-axial accelerometer and GPS. As these sensors use to work at different sampling rates, the first pre-processing step consists of the equalization of the frequency resolution for the time-series provided. Usually, the frequency resolution of the labeling process is chosen as the reference. Therefore, time-series sampled at a higher frequency are averaged, while those sampled at lower frequency are interpolated.

Dealing with real data, it is also necessary to handle incomplete, inconsistent and noisy data. Incomplete data may be due to sensor failures, or even to problems in downloading data. Among the different solutions proposed in the literature, we decide as a rule for our framework to drop all the instances that contain one or more missing values, rather than interpolating the values. This choice was made to avoid introducing further artifacts and noise.

The superimposed noise component present in the measured time-series represents one of the main issues to face dealing with free-ranging animal monitoring. This noise is due to pure measurement error and is in good approximation composed of high-frequency oscillations. Although it is difficult to eliminate it, low pass filters are able to reduce it. In particular, the proposed framework uses two identical second-order Chebyshev low-pass filters, arranged in a cascade. The choice of the filter family is due to its zero transition band and reduced ripple amplitude. The use of two filters in cascade is instead necessary to compensate for the non-linearity of the filter phase, which if applied individually would alter the morphology of the signal [17]. In addition to the high-frequency components, the zero frequency harmonic must also be removed. This process is called debiasing because this frequency represents the bias, i.e. the systematic error that indicates a persistent distortion that the sensor produces on the measured values.

At this point, inconsistencies are managed. To distinguish anomalous instances and normal we leverage an event detection algorithm based on the consistency between features directly interpretable and associated label. In particular, in the proposed framework we have chosen to consider the speed time-series provided by GPS. This choice was made considering that a common split among the ethograms of wild animals species consists in dividing the activities between stationary and non-stationary. The choice to use an event detection algorithm is due to the excellent results reported in the literature regarding their use in behavior identification [18]. In our framework, however, this algorithm was not used to perform the classification, but as a pre-processing technique to identify inconsistent instances. The algorithm is based on two thresholds, the value of which is set by the ethologist, based on his background knowledge. One corresponds to the speed threshold beyond which the animal cannot longer be considered in a stationary state, while the second corresponds to the value below which the animal cannot be considered non-stationary. Instances whose label is related to a stationary state but whose speed value is lower than the minimum threshold and instances whose label is related to a non-stationary state, but whose speed is greater than the maximum threshold are considered to be inconsistent. In this scenario, the most common inconsistencies cause is GPS failures due to connection problems with the satellite. The output of the algorithm is used to create a new label, based on which a specific classifier is properly trained to recognize the inconsistencies. In fact, the speed values measured during failures will be inconsistent with other features provided. Therefore it is possible to train a classifier to recognize these patterns and become robust even to inconsistencies.

2.2. Features Extraction and Selection

Once processed, time-series are used to extract new features. Perform this procedure starting from the raw time-series would be deleterious because they could amplify the measurement errors contained. In this phase, we have chosen to extract the minimum number of features necessary to allow the classifier to extract effective patterns to perform its task. In fact, increasing the number of features does not necessarily improve the performance of the classifier, but rather increases the risk of overfitting. This phenomenon is well-known in literature as *The Curse of Dimensionality* [16]. Bedised, as a selection criterion for a feature, it is required to be significant for the ethologist, i.e. consistent with direct animal behavior monitoring and reasonably related to the activities to identify.

2.3. Classification

In many machine-learning problems, the output of a single classifier is not enough for the results obtained to be reliable [19]. This has led to the development of ensemble algorithms. Our framework leverages an algorithm based on boosted tree, which involves combining several weak learners sequentially, where each one is constructed in such a way as to minimize the classification error committed by the previous one. The function used to define the first tree is

$$F_o(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma) \tag{1}$$

where gamma is an introduced term to allow the model to generalize, preventing overfitting. The following trees are computed according considering each m sample as

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{2}$$

The function that defines the single tree is defined considering the first derivative of the loss function

$$r_{im} = -\alpha[\frac{\partial(L(y_i, F(x_i)))}{\partial F(x_i)}] \tag{3}$$

Where y is the target variable and α is the learning rate. The result, shown in Figure 1 is a classifier composed of many interpretable and shallow trees. This is important in order to be able to estimate the importance of each feature in the decision-making process. Each tree is also defined by several hyperparameters, the most important of which are learning rates and the maximum depth, which are calibrated during the fine-tuning phase.

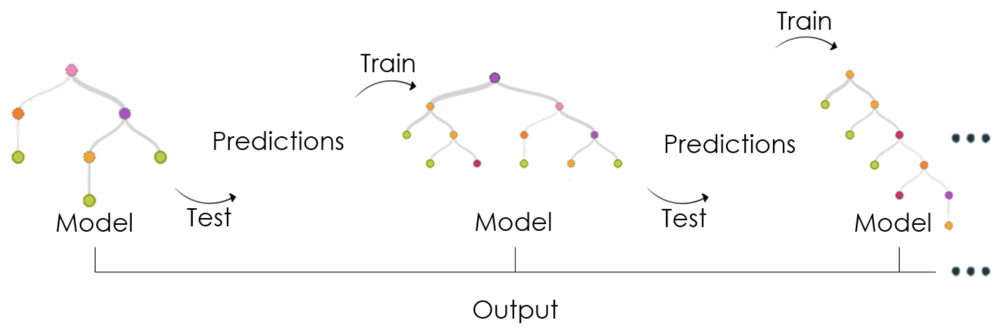

Figure 1: Boosting Tree Algorithm Pipeline. Boosted trees algorithm leverages many shallow tree, sequentially combined so that each one minimizes the classification errors produced by the previous one.

In particular, as a boosted tree algorithm XGBoost was chosen, state of the art among the boosting algorithms in the literature, thanks to the parallelization of the computation and the techniques of regularization properly designed for preventing overfitting [20]. As a metric to estimate the importance that each feature has assumed in the decision-making process was chosen instead F-score, expressed as the ratio between the number of times that a feature is used as a criterion of split and the total of the splits in each boosted trees-based classifier.

In this work, more boosted tree-based classifiers have been combined recreating the structure of the reference ethogram for the species of interest. This results in a hierarchical model, where each split node corresponds to a classifier. This choice has been made to make the model as much as possible interpretable by the ethologist, following his logical path in performing direct animal monitoring. Also, it allows classifying activities at different levels of aggregation.

3. Experimental Framework

In this section are described the dataset employed as a case study and the implementation of the developed framework to perform the proposed classification.

3.1. Dataset Structure and Related Issues

The analyzed dataset was collected between the 1st of August and the 4th of September 2012 at The Mpala Research Centre, in Laikipia County, Africa. The Crofoot Lab was responsible for both the data collection and labeling process. The dataset is retrievable at [21]. It contains the time-series measured by instrumenting 26 Olive baboons with collars equipped with tri-axial accelerometer and GPS, for a total recording period of 220 hours. The first sensor provides a time-series for each of its axes, while GPS measures speed, heading, latitude, and longitude. During the entire duration of the experiment, the animals are left free in their natural state. The combination of the hostile environment and the poor battery life of the collars is a limiting factor for data collection. The number of active collars decreases over time, starting from 26 on the first day and reaching 1 on the last. Animal monitoring always occurs during the day, namely from 6am to 6 p.m., while the night hours have been used to download data, accessible remotely via Bluetooth connection.

The labeling process was conducted by majority voting by a group of biologists, observing video recordings of the animals, acquired synchronously to the sensors measurements. The labeled data constitute 4 hours of recording and are related to just 10 of the 26 baboons in 10 of the 35 days of total observation. Also, they are all related to a time span between 3 p.m. and 5 p.m. . This is an issue, as only this small percentage of the available data can be used for classifier training. The labels were assigned according to two ethograms, supplied together with the dataset and shown in Figure 2. These ethograms are related to the individual positional state and the individual activity assumed by the animal. The first one distinguishes between stationary and non-stationary status. The stationary state is in turn distinguished in sitting and standing at rest, while the non-stationary state in walking and running. As for the activity state, instead, the individual can be in a state of feeding or not feeding.

A balance analysis of the dataset reveals that the individual activity states are equally represented by 51.0% of observations related to feeding and 49.0% related to non-feeding. However, individual positional states distribution appears strongly unbalanced. The non stationary states are only 13.9%, and the running class constitutes 0.9% of the total of the

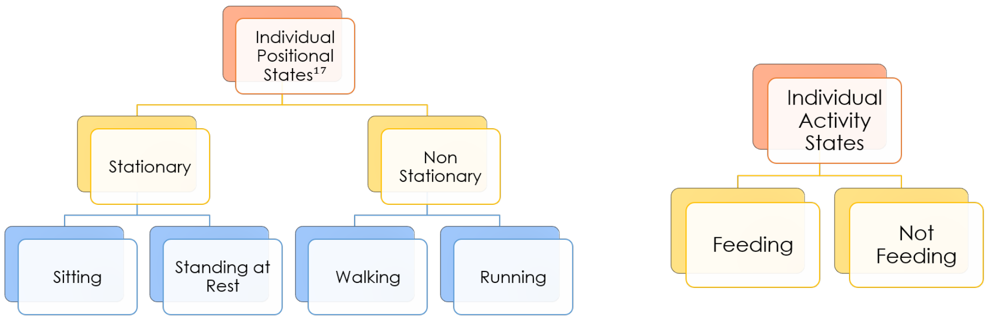

Figure 2: Ethograms for Individual Positional States and Individual Activity States for Olive Baboons. This figure shows the ethogram for the individual positional state, on the left, and for the individual activity state, on the right. They are referred to Olive baboons and have been produced and used by the biologists of Crofoot Lab to provide the ground truth for the video recordings.

observations. This is an extremely important issue, as some distinguishing factors of the less represented classes can be hidden by the preponderance of the most frequent classes, making it difficult for the classifier to identify them. The distribution is strongly unbalance for both individual positional and activity states if we consider the observations recorded per baboon. Even more, some baboons have been monitored just in some of the states. These results are shown in Table 1 and in Table 2. Moreover, the dataset is also unbalanced considering the amount of observations recorder per day of observation and per individuals monitored. This issue is fundamental to overcome to build a classifier capable of generalizing the results obtained on individuals never proposed in the training phase.

3.2. Framework Implementation

First of all, we removed from the dataset the features not related to classification purposes. Therefore heading, latitude and longitude were discarded. Subsequently, according to our framework design, we standardized the temporal resolution of the time-series. In fact, the sampling frequency of the accelerometer used is 12Hz, while that of GPS is 1Hz. The labeling process was performed assigning a label for every second of recording, so its frequency is also 1Hz. The time-series measured by the accelerometer were then averaged, reducing the frequency resolution from 12 to 1Hz. At this point, the time-series were low-pass filtered and debiased.

Subsequently, the thresholds of the event detection algorithm for inconsistencies detection were tuned. The minimum threshold below which an activity can no longer be considered non-stationary has been set to $0.2\frac{m}{s}$, while the maximum threshold above which an activity

Table 1: Baboons Individual Positional State Label. This table shows the strongly unbalanced distribution of the individual positional state label for each monitored baboon.

Collar ID	**Individual Positional State**			
	Sitting	Standing at Rest	Walking	Running
2426	51.55%	29.15%	17.33%	1.97%
2427	77.09%	7.11%	14.46%	1.34%
2428	58.82%	✗	41.18%	✗
2436	95.75%	3.30%	0.95%	✗
2443	80.40%	6.46%	13.14%	✗
2447	100.00%	✗	✗	✗
2449	76.34%	2.23%	20.09%	1.34%
2451	60.07%	18.90%	20.41%	0.62%
2454	54.29%	10.00%	31.90%	3.81%
2457	86.64%	3.31%	9.11%	0.94%
Average	68.5%	13.0%	17.6%	0.9%

Table 2: Baboons Individual Activity State Label. This table shows the strongly unbalanced distribution of the individual activity state label for each monitored baboon.

Collar ID	**Individual Activity State**	
	Feeding	Non-Feeding
2426	83.24%	16.76%
2427	66.47%	33.53%
2428	✗	100%
2436	2.83%	97.17
2443	71.05%	28.95
2447	✗	100%
2449	37.50%	62.50%
2451	30.82%	69.18%
2454	98.10%	1.90%
2457	45.38%	54.62%
Average	51.00%	49.00%

can no longer be considered stationary is $1.75\frac{m}{s}$. The thresholds have been chosen in such a way as to guarantee a wide margin, going to identify only the instances with extremely anomalous characteristics. A new label was then produced, with a value of 1 in case the event detection algorithm recognized the instance as inconsistent, 0 otherwise. Basing on it, a classifier is properly trained in order to distinguish inconsistent observations. This step is also fundamental to solve the problems related to the unbalancing of the dataset, without performing undersampling, which is the technique most employed in literature [22]. This is extremely important, as undersampling would have drastically reduced the size of the available data, which would have lead to classifiers not be sufficiently reliable.

During the features extraction phase, four new features were calculated

- Acceleration norm

$$|a| = \sqrt{|a_x|^2 + |a_y|^2 + |a_z|^2}. \tag{4}$$

- Acceleration ratios

$$\frac{a_x}{ay}; \frac{a_y}{az}; \frac{a_x}{az}. \tag{5}$$

.

Then, correlation between features and labels has been calculated. Since all appeared highly correlated, none was discarded.

As explained in Section 2, the classification model is arranged according to the ethogram structure. Since two ethograms have been provided, two classification models have been designed. These two models were trained on all the labeled observations provided to recognize the individual positional state label and the individual activity state label, respectively. The first, shown in the Figure 3, is composed of the hierarchical combination of four classifiers. The first one separates the observations between consistent and inconsistent, and it is trained basing on the labels provided as the output of the event detection algorithm. The second further divides the consistent observations between those related to the stationary and non-stationary state. A third classifier distinguishes the observations classified as stationary between sitting and standing, while the last one separates those related to the non-stationary state between walking and running. Each classifier works using XGBoost. To identify the best parameters, fine tuning using grid search was performed. This lead us to set the learning rate equals to 0.03,the maximum depth to 40, a number of estimators equals to 13, the subsample ratio of columns when constructing each tree equals to 0.8 and α, the L1 regularization weight, to 10.

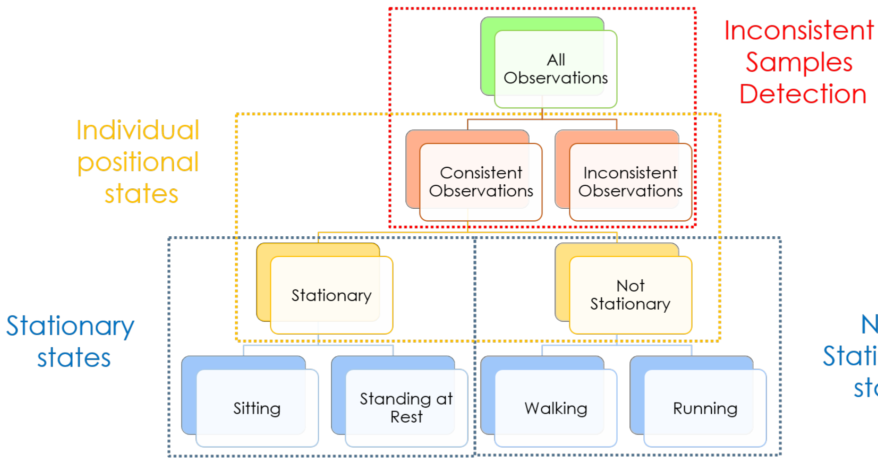

Figure 3: Individual Positional State Classifier. This figure shows the design of the classifier aimed at identifying individual positional states.

4. Results

In this section are reported the validation techniques leveraged and the obtained results related to the case study under analysis.

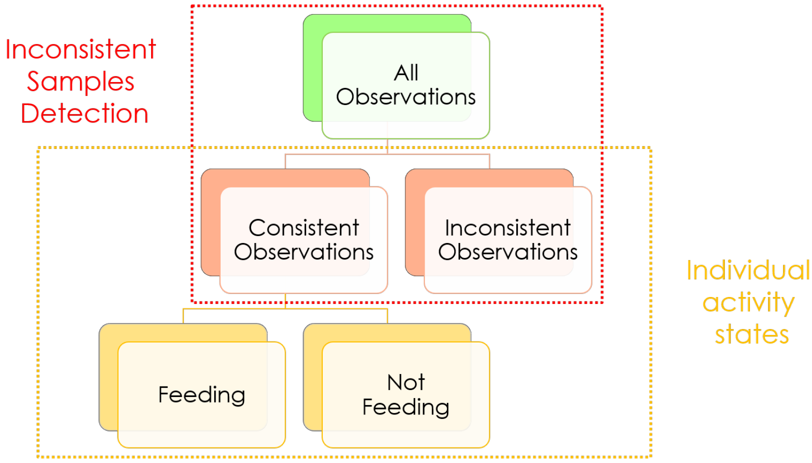

Figure 4: Individual Activity State Classifier. This figure shows the design of the classifier aimed at identifying individual activity states.

The holdout k-fold cross-validation whit k equals to 10 was used as a validation technique to assess the classification models' performance. We arbitrarily chose to train each classifier over 70% of the data and to test it over the 30% left. The split was performed leveraging stratification to lock the distribution of classes in train and test sets to face classes unbalance. The procedure was iterated 10 times to ensure the reliability of the results, changing the composition of training and test sets from time to time. The metrics measured were

- Accuracy, i.e. the percentage of positive classifications that are correct

$$acc = \frac{tp + tn}{tp + fp + fn + tn} \tag{6}$$

- Recall, i.e. percentage of positive elements that have been classified as positive

$$rec = \frac{tp}{fn + tp} \tag{7}$$

- Specificity, i.e. percentage of negative elements that have been classified as negative.

$$spe = \frac{tn}{tn + fp} \tag{8}$$

Although accuracy is the most widely used metric to evaluate the performance of a classifier, in the case of an unbalanced dataset, recall and precision are fundamental, as they allow to assess if even the less represented classes are correctly recognized. The results obtained are reported in the Table 3. Considering individual positional states, it turns out that the average accuracy at first aggregation level is 99.9%, while on the single activity is 94.5%. For what concerns individual activity states, the average accuracy is 87.9%. The classifier tasked to recognize inconsistent samples achieves an average accuracy of 89.0%. This result is satisfactory considering that the inconsistent observations constitute just 17.5% of the labeled data.

Table 3: Performance Evaluation of the Produced Classifiers. This Table reports the performance, in terms of accuracy, recall and specificity of the produced classifiers respect to holdout validation.

Behavior	**Metrics**		
	Accuracy (%)	Recall(%)	Specificity(%)
Stationary	100 ± 0.02	100 ± 0.02	100 ± 0.04
Non-Stationary		100 ± 0.03	100 ± 0.03
Sitting	90.0 ± 0.7	90.5 ± 0.5	98.4 ± 0.1
Standing at Rest		84.8 ± 0.3	45.8 ± 1.5
Walking	99.0 ± 1.0	98.9 ± 1.7	90 ± 1.8
Running		99.0 ± 1.7	83.3 ± 1.2
Feeding	87.9 ± 0.4	87.6 ± 0.6	88.5 ± 0.8
Non-Feeding		88.6 ± 0.7	87.2 ± 1.0
Consistent	89.0 ± 0.9	90.7 ± 0.6	96.7 ± 0.7
Inconsistent		76.6 ± 0.6	52.5 ± 0.1

The Table 4 reports a comparison between the results achieved by our framework and the state of the art solutions proposed in the literature. It turns out that the lift in accuracy respect to the identification of stationary and non-stationary activities is 14.25%, while for the recognition of the single positional state it is 11.6%. For the recognition of the positional activity state, the accuracy lift is 1.6%.

Table 4: Gold Standard Comparison. In this figure, the results, in terms of accuracy, achieved leveraging the proposed approach are compared to the gold standards' ones, respect to holdout validation.

Recognized behavior	**Accuracy**		
	Our Approach	Wang et al. [14]	G.Muscioni [15]
Stationary vs Non-Stationary	100%	✗	85.75%
Sitting, Standing at Rest, Walking, Feeding	94.50%	69.00%	82.90%%
Feeding vs Non-Feeding	87.90%	63.70%	86.30%%

The Table 5 shows the results obtained in terms of F-score for the different classifiers.

Table 5: F-Scores. This figure shows the F-Score that characterized each features for all the designed classifiers.

Considered Classifier	**Features**							
	$Speed$	$\|a\|$	a_x	a_y	a_z	$\frac{a_x}{a_y}$	$\frac{a_x}{a_z}$	$\frac{a_y}{a_z}$
Consistent vs Non-Consistent	97.0%	17.4%	17.0%	17.3%	17.4%	17.3%	17.5%	17.5%
Stationary vs Non-Stationary	98.4%	1.4%	1.2%	1.4%	1.4%	1.4%	1.2%	1.1%
Sitting vs Standing at Rest	33.3%	97.6%	52.0%	59.2%	59.5%	62.2%	92.5%	70.5%
Walking vs Running	98.1%	41.9%	36.2%	✗	19.4%	12.8%	✗	25.9%
Feeding vs Non-Feeding	98.2%	33.4%	29.5%	27.4%	45.1%	90.0%	31.0%	39.0%

5. Discussion

In this section, we will discuss the results exposed in Section 4. In addition, a practical application of the proposed framework for automatic remote animal monitoring will be shown considering the unlabeled data.

The results obtained in terms of accuracy demonstrate that the proposed methodology is effective to perform automatic remote animal monitoring. Both models produced allow to obtain better results than those reported in the literature. To allow for a fair comparison, we select works aimed at developing frameworks for automatic remote wild animal monitoring from the time-series measured by a tri-axial accelerometer and a GPS. Moreover, the activities that they intend to recognize are the same identified in the case of studies we analyze in Section 3. As the number and type of classes recognized vary, the comparability of the results is no longer reliable. The performances in accuracy obtained by the inconsistencies detection classifier have not been compared with other works, as it is a novelty introduced by our approach. At the best of the authors' knowledge, no work in the literature includes a similar classifier in their frameworks. Its accuracy is certainly high, but the results obtained in terms of recall and sensitivity are affected by the class unbalancing and could be improved by training the classifier on a consistent and balanced dataset. Performing undersampling on available data would not be considered as a solution, as the final amount of information left will not be sufficient to train a reliable classifier. The performance assessed by the classification model for the individual positional states reveals that it is able to accurately recognize all the activities contained in the reference ethogram. Even running, which appeared to be poorly represented within the dataset, is identified with a 100% recall and a 83.3% specificity. The most challenging split related to sitting and standing at rest. The cause of this difficulty is probably related to the placement of the sensors as, in both states, it is fair to assume that the patterns in time-series measured at neck level are very similar. Finally, the classification model for the recognition of individual activity states has slightly lower performances than the others, although higher than those reported in the literature. This could be because instrumenting the neck alone is not sufficient to better recognize this activity. Probably, having an accelerometer also on one of the front limbs would give additional information to the system and would allow obtaining better performances.

The results returned in terms of F-Score, fully satisfy the objective of producing a model whose decision-making process was clear and intelligible for the ethologist. The key feature in the decision-making process of the inconsistencies detection classifier is speed. As for the classification model aimed at recognizing individual stationary states, it seems reasonable that speed is the key feature considered in distinguishing stationary activities, while it is minimally considered to distinguish sitting from standing. It is also interesting how for the split between walking and running, even if the speed is a key feature, also the accelerations become determinant. This makes sense considering that some speed ranges may be common to running and walking, but the difference between the two behaviors is given by the different locomotor mechanism, which results in different for the acceleration trends measured by the sensors. Therefore, an accurate distinction between the two activities must consider the combination of these features. In all the classifiers that compose this model, the acceleration norm appears to be a relevant feature in the decision-making process. Analyzing this feature trend, it emerges that it has higher values for stationary activities than for the non stationary ones. We can deduce that, once in motion, the animal tends to keep its neck stable. In the decision-making process for the recognition of individual activity states, both speed and acceleration are relevant. The importance of speed suggests that as the speed of the animal increases, it becomes less likely that the resulting motion is compatible with the feeding activity. The importance of acceleration trends is related to the repetitive neck motion performed by the animal while eating. This pattern can be learned by the classifier during its training phase and used in the decision-making process. These considerations show that results returned in terms of features importance are easy to interpret and offer important information about the species under analysis.

5.1. Automatic Remote Animal Monitoring

Once we assessed the effectiveness of the produced framework, we used it for automatic remote animal monitoring considering the unlabeled data available, which we could not use in the training and testing phase. The predictions were then analyzed. First of all, it emerged that the distribution of the classes is consistent both with the distribution of the 4 hours labeled data and with the real behavior of the Olive baboons, that behave to minimize the energy spent [23]. In fact, it turns out a clear preponderance of stationary activities. Moreover, sitting is preferred over standing at rest, while walking is preferred over running.

Besides, we compared the characteristic speed ranges of walking and running for the different baboons observed. The Figure 5 shows the results obtained for running. It is clear that speed values for running range from 2 $\frac{m}{s}$ to 4 $\frac{m}{s}$, with an average value of 2.5 $\frac{m}{s}$. The fastest baboon, with collar ID 2434, has an average speed of 25% faster than his teammates. The slowest baboon, on the other hand, with collar ID 2430 runs at 2.1 $\frac{m}{s}$ on average, which is 17.5% slower than the others. For walking, the speed values range from 0.15 $\frac{m}{s}$ to 0.6 $\frac{m}{s}$, and the average speed is 0.36 $\frac{m}{s}$.

Finally, we represented the positional state assumed by the baboons hour per hour on the different days, shown in the Figure 6. Each column represents the frequency for the four individual positional assumed on average by the animals during the considered hour. It turns out that animals tend to behave according to repetitive patterns day after day. This seems reasonable, considering that during the observation period no particular external trigger was recorded that should have caused deviation from their normal behavior. Moreover, it is evident that the most represented activity is sitting, in light blue, considering stationary states, and walking, in orange, considering non-stationary states. Standing at rest, in green, and running, in red, are instead infrequent.

6. Conclusion

In this work, an accurate and intelligible framework for automatic remote monitoring of wild animals has been proposed.

The entire system is designed to be consistent with the logical human path followed by the ethologist in the monitoring process. At best of author knowledge is the first time that

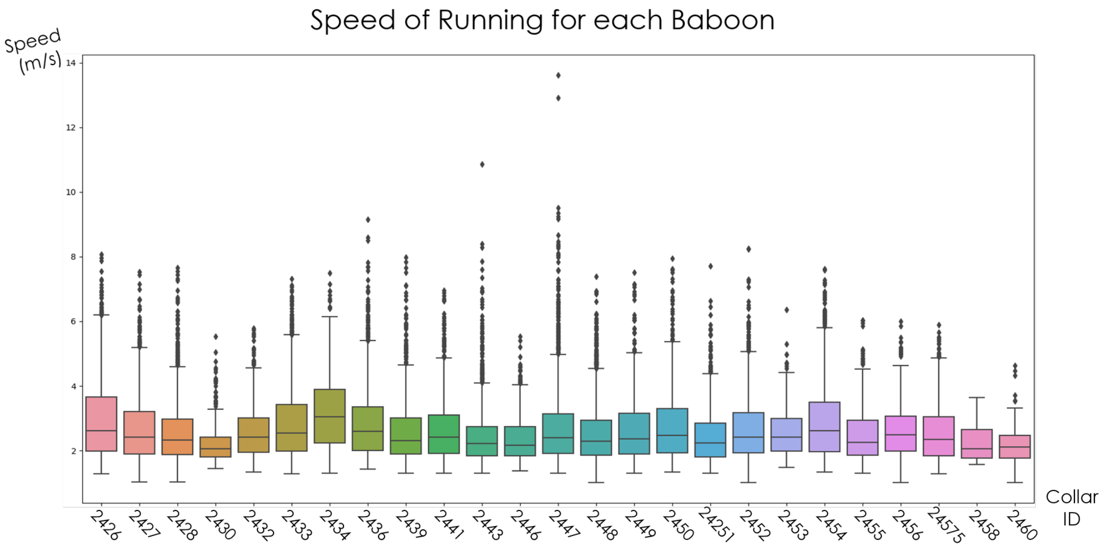

Figure 5: Running Speeds of Observed Baboons. This figure shows the speeds grouped for each baboon of those observations whose individual positional state's prediction is running.

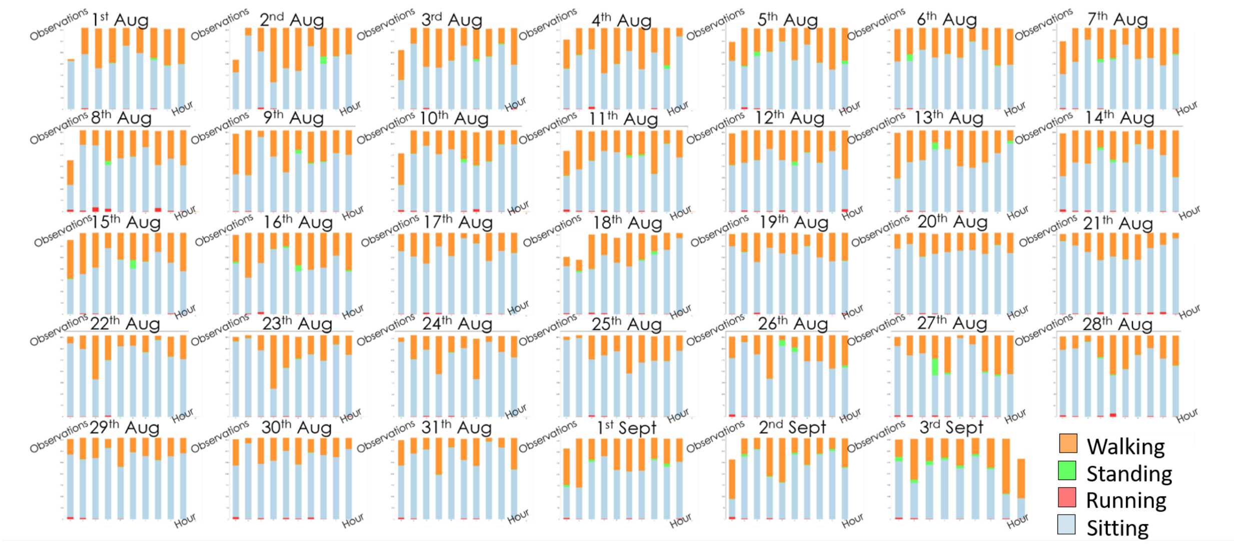

Figure 6: Daily Individual Positional State's Pattern for a Single Baboon. This figure shows the individual positional state assumed by the considered baboons hour by hour, during the different observation days. Each plot represents an observation day, while each column represents an hour of observation.Y-axis ranges from 0% to 100% while X-axis ranges from 7 a.m. to 6 p.m. and each column refers to one hour. Notice that in columns that do not reach the 100% the missing percentage is referred to missing or corrupted data.

interpretability is considered as a parameter to be optimized in the creation of a framework for animal monitoring. Moreover, our system is more accurate than the ones presented in the literature.

The developed framework is therefore valid for automating the process of animal monitoring, allowing the ethologist to devote himself to the phase of analysis and inference.

References

[1] J. Mench, Why it is important to understand animal behavior, ILAR journal 39 (1) (1998) 20–26.

[2] R. Williams, D. Lusseau, P. S. Hammond, Estimating relative energetic costs of human disturbance to killer whales (orcinus orca), Biological Conservation 133 (3) (2006) 301–311.

[3] J. Krause, D. Lusseau, R. James, Animal social networks: an introduction, Behavioral Ecology and Sociobiology 63 (7) (2009) 967–973.

[4] I. L. Boyd, A. Kato, Y. Ropert-Coudert, et al., Bio-logging science: sensing beyond the boundaries, National Institute of Polar Research.

[5] T. T. Hammond, D. Springthorpe, R. E. Walsh, T. Berg-Kirkpatrick, Using accelerometers to remotely and automatically characterize behavior in small animals, Journal of Experimental Biology 219 (11) (2016) 1618–1624.

[6] K. M. Gaynor, C. E. Hojnowski, N. H. Carter, J. S. Brashares, The influence of human disturbance on wildlife nocturnality, Science 360 (6394) (2018) 1232–1235.

[7] D. D. Brown, R. Kays, M. Wikelski, R. Wilson, A. P. Klimley, Observing the unwatchable through acceleration logging of animal behavior, Animal Biotelemetry 1 (1) (2013) 20.

[8] R. P. Wilson, E. Shepard, N. Liebsch, Prying into the intimate details of animal lives: use of a daily diary on animals, Endangered Species Research 4 (1-2) (2008) 123–137.

[9] A. Papailiou, E. Sullivan, J. L. Cameron, Behaviors in rhesus monkeys (macaca mulatta) associated with activity counts measured by accelerometer, American Journal of Primatology: Official Journal of the American Society of Primatologists 70 (2) (2008) 185–190.

[10] P. Martiskainen, M. Järvinen, J.-P. Skön, J. Tiirikainen, M. Kolehmainen, J. Mononen, Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines, Applied animal behaviour science 119 (1-2) (2009) 32–38.

[11] J. Barwick, D. W. Lamb, R. Dobos, M. Welch, M. Trotter, Categorising sheep activity using a tri-axial accelerometer, Computers and Electronics in Agriculture 145 (2018) 289–297.

[12] P. Löttker, A. Rummel, M. Traube, A. Stache, P. Šustr, J. Müller, M. Heurich, New possibilities of observing animal behaviour from a distance using activity sensors in gps-collars: an attempt to calibrate remotely collected activity data with direct behavioural observations in red deer cervus elaphus, Wildlife Biology 15 (4) (2009) 425–434.

[13] J. Li, K. Asif, H. Wang, B. D. Ziebart, T. Y. Berger-Wolf, Adversarial sequence tagging., in: IJCAI, 2016, pp. 1690–1696.

[14] Y. Wang, B. Nickel, M. Rutishauser, C. M. Bryce, T. M. Williams, G. Elkaim, C. C. Wilmers, Movement, resting, and attack behaviors of wild pumas are revealed by tri-axial accelerometer measurements, Movement Ecology 3 (1) (2015) 2.

[15] G. Muscioni, R. Pressiani, M. Foglio, M. C. Crofoot, M. D. Santambrogio, T. Berger-Wolf, A framework for identifying group behavior of wild animals, arXiv preprint arXiv:1907.00932.

[16] V. Spruyt, The curse of dimensionality in classification, Computer Vision for Dummies 21 (3) (2014) 35–40.

[17] J. Kormylo, V. Jain, Two-pass recursive digital filter with zero phase shift, IEEE Transactions on Acoustics, Speech, and Signal Processing 22 (5) (1974) 384–387.

[18] S. Gelmini, S. Strada, M. Tanelli, S. Savaresi, A. Guzzon, Analysis and development of an automatic ecall algorithm for wearable devices, in: 2018 IEEE Conference on Control Technology and Applications (CCTA), IEEE, 2018, pp. 1240–1245.

[19] J. H. Friedman, Greedy function approximation: a gradient boosting machine, Annals of statistics (2001) 1189–1232.

[20] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, ACM, 2016, pp. 785–794.

[21] S.-P. et al., Collective movement in wild baboons (2015).
URL https://www.datarepository.movebank.org/handle/10255/move.405

[22] A. Dal Pozzolo, O. Caelen, R. A. Johnson, G. Bontempi, Calibrating probability with undersampling for unbalanced classification, in: 2015 IEEE Symposium Series on Computational Intelligence, IEEE, 2015, pp. 159–166.

[23] S. A. Altmann, Baboons, space, time, and energy, American Zoologist 14 (1) (1974) 221–248.