Investigating Generic Online Animal Activity Recognition Across Different Animal Species

Lieke Hamelers
University of Twente
PO Box 217, 7500 AE Enschede
the Netherlands

l.h.hamelers@student.utwente.nl

Abstract

Recognizing animal behaviour has proven to be very useful for detecting changes in the environment of the animal and in their wellbeing. Research has mainly focused on optimizing classifiers for one species or using different techniques for training classifiers and testing them separately on different species. In this research a different approach is taken to classifying animal behaviour. The research investigates how generic a specific animal behaviour classifier can be, applied to the group of quadruped animals. Seven different classifiers are trained with data from cows, horses, sheep and goats. The classifiers are trained with the datasets of one, two or three different species and tested on another species, a so-called non-mixed experiment. This research shows that the k-Nearest Neighbor algorithm gives the best results for non-mixed classifiers, and that with increasing of the number of species in the training set, the accuracy of the classifier increases. Further, the recognition of certain activities is investigated. The activities that are recorded are stationary, grazing, walking, trotting and running. The activities stationary and grazing are classified more accurately than the other activities. This work supports further research into non-mixed classifying to eventually develop a generic classifier for quadruped animals.

Keywords

Online Animal Activity Recognition, Generic Classifier, Animal Behaviour, Classification.

1. INTRODUCTION

The population of wildlife has decreased by 50 percent since 1970, mainly due to human influence. Poaching and forest fires are among the most important factors causing decline [22]. Worldwide preservation efforts aim to minimize poaching and forest fires. The monitoring of poaching and forest fires are a challenge as they can occur over vast areas. An innovative approach for detecting these events is to monitor animal behaviour. When a fire occurs, animals start to panic and show abnormal activity such as leaving the herd [18]. Accelerometers and global positioning systems have already been used for the conservation of rhinoceroses and to study animal behaviour [15]. Collar tags equipped with accelerometers track the individual movements of animals. This data can be analysed and classified to recognize individual animal activities. Research has mainly focused on doing offline data collection from the collar tag. It must then be classified and analysed [5,11]. To prevent poaching and to detect forest fires early-on, the data and analysis will be needed in real-time. The main challenge to online animal activity recognition is the battery life of sensors. It is proven that the battery life can be increased by using a classifier locally on the collar tag and then sending the analysed data [8,9]. With this technique, the collar tag can also be on standby when an animal is sleeping or stationary, decreasing energy use further. Therefore, an accurate, on-board classifier is needed to track animal behaviour to detect poaching and forest fires.

To develop a classifier for a certain animal, movement data must be collected, and the classifier must be trained with this data. Data for the training consists of movement data from the collar tag and classified behaviour by the researcher derived from filming. Collecting training data becomes a challenge with animals that are shy, camouflaged or live in a vast area. To train a classifier for each species will take an enormous amount of time. A generic classifier for a family of animals or different species that have some specific similarities, could overcome this challenge. It could be applied directly to different species and will save an enormous amount of time.

2. STATE OF THE ART

Frequently used monitoring techniques can be divided into two categories. The first category is the standard offline techniques. The most popular techniques in this category are live trapping, sign surveys and passive infrared camera trapping (PIR). J. Molyneux et al. have established that the PIR cameras performed best [13]. The second category includes the digital and online techniques. The use of unmanned aerial vehicles is popular to identify animals in vast areas. This is a non-invasive technique to monitor the animals and their behaviour [21]. Another technique is the use of sensors to track animals and their activities. This technique uses global positioning systems(GPS) and accelerometers and has been successfully used on animals as well as humans [2 ,3 ,4 ,5, 6, 13,10].

For cows, the technique of using sensors to monitor their behaviour is most popular [1, 3, 5, 11, 20]. Not only in research, but several companies are offering this technique to farmers to monitor their cattle [10]. For instance, Nedap offers a system where every cow is tracked and monitored. They track the duration that a cow lies, how much it eats and when the cow is in heat. This offers farmers a tool to increase efficiency and expand the farm. Most researches performed on cows, uses classifiers that mainly classify the following activities: grazing, lying, standing and walking. The most popular classifiers are: Support Vector Machine, k-Nearest Neighbor, Linear Discriminant Analysis and algorithms based on tree decisions. Most classifications performed show an accuracy of around 94% [3, 5, 11, 20]. Not much research is done on the creation of a generic classifier for animals. In the field of human activity tracking, Cvetkovic et al. have done some experiments with a generic, one

29th Twente Student Conference on IT, July 6th, 2018, Enschede, The Netherlands. Copyright 2018, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

species classifier [2]. They collected movement data from five different human beings. Several classifications were performed where the classifier was trained with the data from four persons and then tested with data from the fifth person. The accuracy was around 83%. Gao et al. have investigated several classifiers on human, badger and dog behaviour data [4]. They trained the classifiers individually and they have applied the classifiers trained with dog data on the badger data. Their results showed that the classifiers that were trained individually were more accurate than the classifier trained with dog data applied to badger data. However, the last classifier was quite accurate and acceptable when no individual classifier was available.

Relatively recently, research has started on generic online animal activity recognition with collar tags [8,9]. In this article, Kamminga et al. state that only a few studies focus on online activity tracking and there is no previous research on an online classifier for animal behaviour. They created several classifiers that have been trained and tested on data from goats and sheep and reached an accuracy of 94%. Kamminga et al. hint that this research supports the development of a generic online classifier.

3. OBJECTIVE AND RESEARCH QUESTION

The objective of this research is to investigate the performance of a previously proposed animal activity recognition classifier when it is tested with new species and how this performance can be improved. The data of sheep and goats were available from previous research. For this research, data from cows and horses were collected.

Cows are in the family of Bovinae consisting of bison's, water buffaloes and antelopes. Horses are in the family of Equidae together with antelopes, zebras and donkeys. In the Netherlands, cows and horses are present in large amounts as cows are used for dairy farming and the meat industry and horses are kept for sports and pastime. Cows and horses are interesting test subjects as they are friendly and perform similar activities to goats and sheep such as lying, standing, feeding and running. Within this research, we are mainly interested in the performance of classifiers across species, i.e. how generic is the classifier. To test this, several combinations of test and training data sets from different species were made and applied to the different classifiers. With the results, we will answer the following research question:

- How generic can a classifier be for quadruped animals?

To answer the research question, four sub-questions must be answered:

- Is the behaviour of the animals comparable to each other?
- Is overall the accuracy of the classifier acceptable?
- Are there specific observed activities that the classifier recognizes well or poorly?
- What parameters need to be changed to improve the classification performance?

4. METHODS

For this research, several phases were distinguished. These phases have all been executed and below they will be elaborated on. Approval according to clause 1, article 1B.7 of the Dutch Law of Ethics was not needed as the harm done was neglectable.

4.1 Data Acquisition

Several cows were needed to collect the data from. The research was performed at the farm "Hoeve Doorwerth" where Galloway cows are being kept. They originate from Scotland and are kept for meat. There were 14 cows present on May 23rd 2018 at "Hoeve Doorwerth", only two cows were suitable to collect data from. The others were either too wild or too timid to be captured. The two cows were lured with food into a shed where the sensors could be put on the cows. To collect the data, Promove Mini basic sensors from Inertia Technology were used [16]. They were synchronized to collect at a rate of 100 Hz and were used locally. To prolong their battery-life, a power bank was attached. The cows were filmed by the researcher during data collection. A GoPro Hero 3+ was used in combination with a power bank and tripod. The first day the sensors were put on a collar that was placed around the neck of the cows and positioned the sensors at the bottom of the neck. After evaluation, this was not the right approach. The neck collar was too big which made the sensor hang loosely. The wind could move the sensor and whenever the cow was grazing, the necklace would touch the ground; all this influenced the data and made it non-usable. In total on this day four hours of data was collected. On day two and three, the sensors were put on a special head-collar that was tightly secured on the head of the cow and placed the sensors in the neck behind the ears. The neck collars and the head collars had different colours to distinguish the animals on film. On these days a total of 12 hours of sensor data was collected.

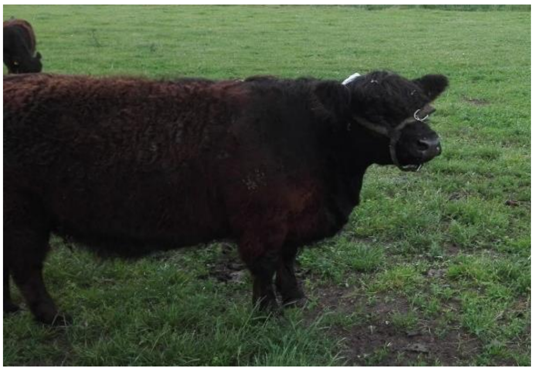

Figure 1 One of the cows with a sensor placed behind the ears.

There were some complications with the sensors. They corrupted the data once it was downloaded from the sensors. This resulted in the loss of enormous amounts of data. In total, the result was a database of 1.5 hours; approximately 6000 data points. It was concluded that is not enough data to use in the research. As a solution, horse data from another research was added to this research. Their data acquisition is still going on, however we received 4 hours of data available from in total four horses. This was in total approximately another 15.000 data points to use. The data from the horses is collected at the "Horstlinde" in Enschede. In total, data is being collected from 16 horses over several days. The sensors that are used are Human Activity Monitor from Gulf Coast Data Concepts and include an accelerometer, gyroscope and magnetometer [7]. The sensors were placed on a collar and were orientated below the neck of the horse. Data was collected at a sampling rate of 100 hz.

4.2 Data Labeling

For the process of data labeling, the raw sensor data was transformed into a MATLAB file. This file could then be used in combination with the videos in a specifically developed MATLAB GUI [8, 9, 12]. The videos and the sensor data both have timestamps to synchronize them correctly and minimize the risk of errors in the labeling process. On the screen, the video is played while a vector of the sensor data is shown. In the frame of the sensor data one can click and note when a certain activity begins. This process is saved in a csv file where the start time and the accompanying activity is noted.

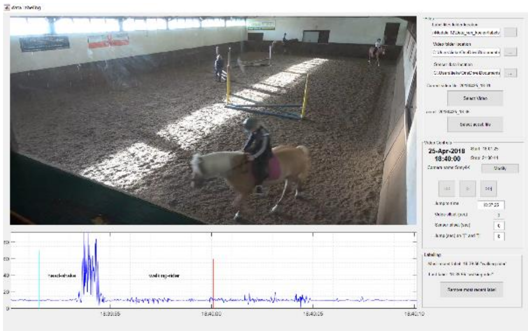

Figure 2 The MATLAB GUI that was used to label the data.

4.3 Data Processing

4.3.1 Data pre-processing

To analyse the data, the data is to be loaded into RapidMiner [16]. RapidMiner is a platform for data science that can prepare your data, apply machine learning and deploy predictive models in a Recall-friendly way. From the previous phases, two types of files are collected; a file with the labels and a file with the sensor data. These need to be combined, cleaned and the features need to be calculated to train the classifiers with. A MATLAB script that was developed by J. Kamminga was used to transform the data. For the windowing of the data, the following parameters were set. The overlap was set at 0.5, the segment length at twenty seconds and the window was two seconds. The features that were calculated are summed up in Table 1.

Table 1 The calculated features for the data points.

Feature	*Description*
Maximum	Maximum value.
Minimum	Minimum value.
Mean	Average value.
Standard deviation	The amount of variation of values.
Median	Value separating half of the values.
25th percentile	Value below which 25% of the values are found.
75th percentile	Value below which 75% of the values are found.
Mean low pass filtered signal	Average value of the DC components.
Mean rectified high pass signal	Average value of the AC components.
Skewness of the signal	The degree of asymmetry of the signal distribution.
Kurtosis	Measure of the tailedness of the signal distribution.
Zero crossing rate	The number of zero crossings per second.
Principal frequency	The frequency component that has the greatest magnitude.
Spectral energy	The sum of the squared discrete FFT component magnitudes.
Frequency entropy	Measure of the distribution of the frequency components.
Frequency magnitudes	Magnitude of the first six components of FFT analysis.

The MATLAB script created three separate files from each datafile: a training set, a test set and a cross validation set. The script does this in a 3-fold for the cow data and in a 5-fold for the horse data. This is done to ensure that every datapoint is used once in a training set, cross validation set and in a test set. This improves the data distribution for every experiment. When the data is divided with a 3-fold, the data is distributed evenly over the training, cross-validation and test sets. When the data is divided with a 5-fold, the training set is 60%, the test and cross validation sets are both 20%. The data was now added in RapidMiner. The first step is to change the activities from text to numbers that could be equalled throughout all the animals. The process that was used further within RapidMiner is depicted in figure 3 that shows the main steps: select, clean, classify and evaluate. Those steps are described in more detail below.

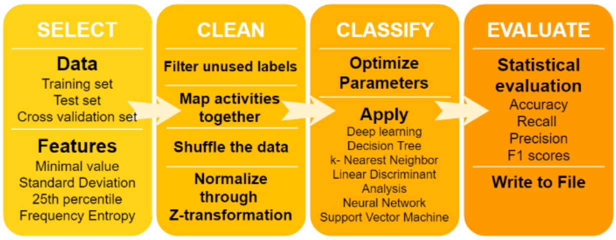

Figure 3 The process implemented in RapidMiner.

4.3.2 Select

In this section the correct data is selected. Every dataset is divided into a training, test and cross validation set. When a cross validation set is not used, it can be added to the training data. The training set is 33% or 66%, the test set is 33% and the cross-validation set is 33% or 0%. The features that are the most relevant for the classifiers were calculated. To combine all the data from the horses, goats, sheep and cows, they all need to have the same columns. A simple process from RapidMiner was created to convert the name and type of the columns of the 'old' sensor data to match the new columns. The Relief algorithm estimates the quality of the features and how well their values distinguish between instances that are nearby. The Relief algorithm is deemed highly successful due to its simplicity. It gives normalized weights to the features to extract and use the top contributing features. The features that were weighted were all derived from the 3D vector of the accelerometer. The top contributing features for the entire set of data of all species were the minimal value, standard deviation, the 25th percentile and the entropy frequency.

4.3.3 Clean

Next, the data that is to be classified needs to be cleaned. First, the labels that were not used had to be removed. After that, certain activities must be mapped together to be able to use the classifiers among different species. An example is that in the horse data there was a difference between walking with a rider on top or without a rider, these were mapped together. This was done because the data did not differ much and to have more data points. This led to Table 2. After the mapping, the data is shuffled and is normalized according to the Z-transformation. This normalization subtracts the mean of all the data and then divides them by the standard deviation. This technique is very common and preservers the distribution of data while avoids being influenced by outliers.

Table 2 The activities performed by the animals and measured with the sensors.

Horse	Cow	Sheep	Goat	Description
Stationary	Stationary	Stationary	Stationary	The animal has an upright position where he is on all four legs and does not walk or move or the animal is lying down on the ground.
Grazing		Grazing	Grazing	The animal is on four legs and lowering his head to eat the grass of the ground. During this activity he can be walking occasionally very slowly.
Walking	Walking	Walking	Walking	The animal moves in a steady but slow pace to a certain spot.
Running		Running	Running	The animal moves in a wild and very fast pace to a certain spot.
Trotting		Trotting	Trotting	The cow moves in a steady and faster pace than walking to a certain spot. This is not the pace of running yet.

4.3.4 Classify

4.3.4.1 Parameter optimization

For the classification, several classifiers were used. Before the classifier is applied, the parameters for the specific classifier are optimized. This is done by applying the classifier to the training/cross validation set and, in a loop, evaluate the performance of the classifier. The best parameters are chosen and applied to the classifier. The parameters are also saved in a text file. The value of a parameter is of influence on the performance of the classifier which makes it important to choose the correct values. The correct parameters were evolutionary selected with a tournament selection. The size of each tournament was 25% with a crossover probability of 90%. The different parameters of each classifier are summed up below with the values which are varied.

These seven classifiers were chosen based on previous research. The descriptions below are based on the information given by RapidMiner [16]. The variable parameters are mentioned in Table 7 in the appendix. The variables that are fixed are however mentioned below.

4.3.4.2 Neural Network (NN)

A Neural Network consists of a collection of connected nodes. These nodes can process signals and send these to the other neurons. The NN is a feedforward Neural Network which means that the information only goes in one direction from input to output. There are no cycles or loops. It is trained by backpropagation; the algorithm compares the output with the correct answers and adjusts the weights of each connection to minimize the errors. The function that was used is usual sigmoidal function. The number of training cycles was set to 500.

4.3.4.3 Decision Tree (DT)

A Decision Tree consists of branches and nodes. On every node a specific feature value is tested, and this decides how the values are navigated through the tree. The information gain ratio was used as a splitting criterion.

4.3.4.4 Support Vector Machine (SVM)

The Support Vector Machine is an algorithm that maps the datapoints in space and aims to maximize the gap between these datapoints. New datapoints are mapped into this space and according to the spot they are in, they are classified. A LibSVM was applied that was C-SVC and with kernel type linear.

4.3.4.5 Naïve Bayes (NB)

This classification technique assumes that the presence of features is independent of any other feature, regardless of any correlation between those features. It is based on Bayes' theorem. The parameters are not adjustable.

4.3.4.6 Linear Discrimant Analysis (LDA)

The Linear Discriminant Analysis is a generalization of Fisher's linear Discrimant. It tries to find a linear combination of features that characterises several classes. The result is used as a linear classifier. The approximate covariance inverse was activated which creates an inverse of the covariance matrix if it does not exist. There are no other parameters to be optimized.

4.3.4.7 K-Nearest Neighbors (kNN)

The k-Nearest Neighbors algorithm compares an unknown example to the k training examples that are the closest neighbors. The closeness of examples is defined by the n-dimensional space where n is the number of examples in the training set. Then several features such as the Euclidian Distance can be used. After that the example is classified by the majority vote of the weighted neighbors. The measure that was used is the Euclidian Distance.

4.3.4.8 Deep Neural Network (DNN)

The Deep Neural Network is an example of a neural network but with more hidden layers, making it deep. The Deep Neural Network was used with ten hidden layers that each had 50 neurons. The value of epsilon was 1.0×10^{-8}, and rho was 0.99.

4.3.5 Evaluate

After the classifiers have been applied in the process, the outcome is evaluated. The label together with the features, confidence and the predicted label are saved to an Excel file to evaluate specific activities separately. The performance of the classifier and its confusion matrix are saved to a text file.

4.4 Data Analysis

To analyse the data that the classifiers produce and to draw conclusions from this, several experiments must be performed. There are four categories of experiments. In the first category, *category individual*, the classifiers are tested on data from one species. In this case, every classifier is given data from one species and tested on data from the same species. This is to see how every classifier is doing within one species. The second category consists of the non-mixed experiments with one species. In *category one*, the classifiers are trained with data from one species and then tested on data from another species. In *category two*, the classifiers are trained with data from two species and tested on data from another species. In *category three*, the classifiers were trained with data from three species and tested on another species. Table 3 and 4 show every non-mixed experiment that has been performed. The horse data is data from four horses, the goat data from four goats, the cow data from one cow and the sheep data from two sheep. In total there were 994 experiments to be executed.

After the experiments have been performed, the performance data and a confusion matrix of the classification are saved. These are loaded into Excel to calculate the following features.

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

$$recall = \frac{tp}{tp + fn}$$

$$precision = \frac{tp}{tp + fp}$$

. TP stands for true positives, TN for true negatives, FP for false positives and FN for false negatives. With the use of the precision and recall, the F1-score is calculated based on the following formula. The F1 formula is the harmonic mean between the precision and recall.

$$F1 = 2 \, x \, \frac{precision * recall}{precision + recall}$$

Table 3 The experiments of category one and two.

Training with one species: 60% training and 40% cross validation data	Test: 100% test data	Training with two species: 60% training and 40% cross validation data	Test with 100% test data
Goat	Horse	Sheep, Goat	Cow
Goat	Cow	Sheep, Goat	Horse
Goat	Sheep	Sheep, Horse	Cow
Sheep	Horse	Sheep, Horse	Goat
Sheep	Cow	Sheep, Cow	Horse
Sheep	Goat	Sheep, Cow	Goat
Cow	Horse	Goat, Horse	Cow
Cow	Goat	Goat, Horse	Sheep
Cow	Sheep	Goat, Cow	Horse

Horse	Cow	Goat, Cow	Sheep
Horse	Goat	Horse, Cow	Goat
Horse	Sheep	Horse, Cow	Sheep

Table 4 The experiments of category three.

Training with three species: **60% training and 40% cross validation data**	**Test:** **100% test data**
Sheep, Goat, Horse	Cow
Goat, Horse, Cow	Sheep
Horse, Cow, Sheep	Goat
Cow, Sheep, Goat	Horse

5. RESULTS

RapidMiner was used to apply the seven classifiers to the data and conduct the experiments described in the methods section above. The number of activities was calculated per species and is shown in Table 5. These datasets were used in further experiments.

Table 5 Absolute count of activities.

Activity	**Cow**	**Goats**	**Sheep**	**Horses**
Stationary	4664	24424	8545	269
Walking	162	5860	2467	5788
Trotting	0	476	777	2882
Running	0	363	785	704
Grazing	0	8091	10092	3304

The setup for every experiment was created in RapidMiner and after cleaning the data, the processes could start. First, the accuracy of the several classifiers will be shown in figure 4 and 5. In these figures the average accuracy of every classifier is shown, applied on the experiments of *category one, two, three and individual*. The categories of experiments are shown on the top axis. For example, there are 12 experiments per classifier to be performed in *category one*. This means that the values of accuracy shown in figure 4 for experiments from *category one* are an average of these twelve experiments for each classifier.

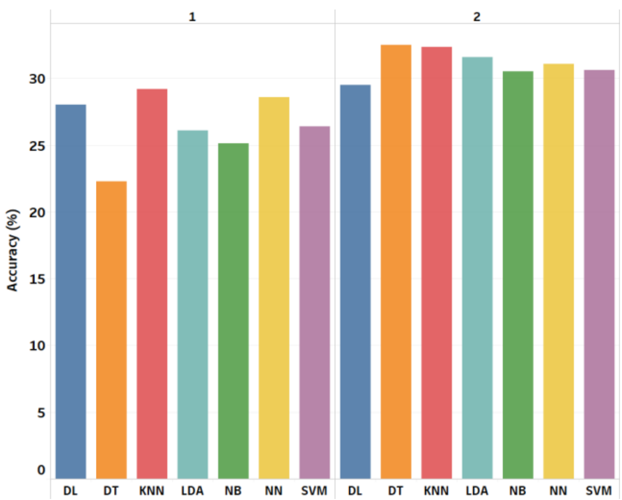

Figure 4 The average accuracies of category one and two experiments for every classifier.

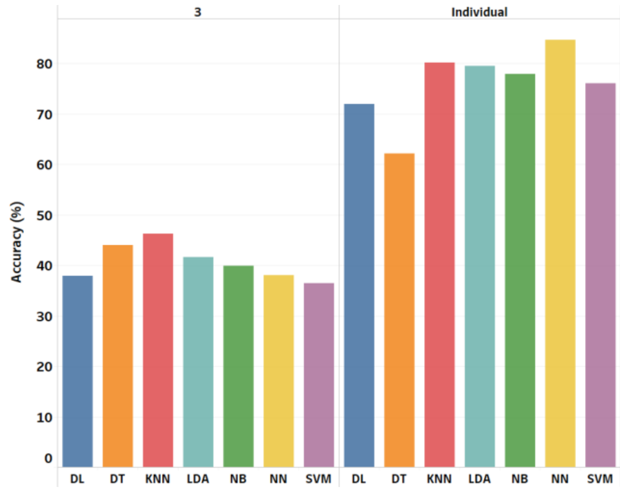

Figure 5 The average accuracies of category three and individual experiments for every classifier.

After the accuracies were analysed, the best classifier was chosen. This was the K-Nearest Neighbors for the experiments of *category one, two and three* and the Neural Network for the experiments from *category individual*. To further analyse the activities separately, the F1 scores were calculated for all the categories with the outcomes of the best performing classifiers. For every activity in every experiment of every category, the precision and recall were calculated. These were then used to calculate the F1 score. The F1 scores were averaged for every experiment within a category. This led to the results in Table 6. below. In the last column, the average was taken of all the averaged F1 scores from every category for one specific activity.

Table 6 The average F1 scores for every activity of every category experiments.

Activity	**Category one**	**Category two**	**Category three**	**Category individual**	**Average**
Grazing	37.3%	40%	47.2%	70.0%	48.6%
Running	20%	26.8%	39.9%	44.7%	32.9%
Stationary	46.6%	45.2%	62.0%	72.9%	56.7%
Trotting	32.9%	23.0%	33.6%	57.7%	36.8%
Walking	27.3%	24.7%	31.3%	67.3%	37.7%

This research focusses to show which animal species recognize each other's activities well. Figure 6 was produced to show how well each animal species classified another animal species. These are all results from the *category one* experiments with the best performing classifier applied. After the results were in, it was calculated which experiment was which combination of species. This could then be shown on the axis of the graph. On the upper x – axis is the species with which the classifier was trained. On the lower x – axis is the species with which the classifier was tested. The y-axis shows the accuracy in percentage. Figure 6 shows that the best performing animal combinations are cow-goat, goat-sheep, horse-sheep and sheep-goat. To further investigate the recognition of specific activities, the F1 scores of the activities of these four animal species combinations were calculated and plotted. Figure 7 shows the F1 scores of the specific activities from the four best performing experiments of *category one*. This gives more insight in the best recognized activities by different animal combinations. On the upper x-axis, the combination of species that the classifier was trained with and tested by is, shown. On the lower x-axis, the activity is shown from the specific combination. The y-axis shows the F1 scores.

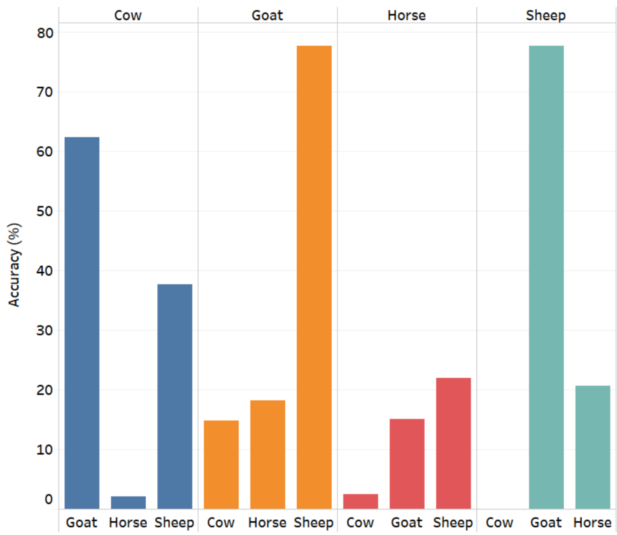

Figure 6 The accuracies of all the category one experiments with the best performing classifier.

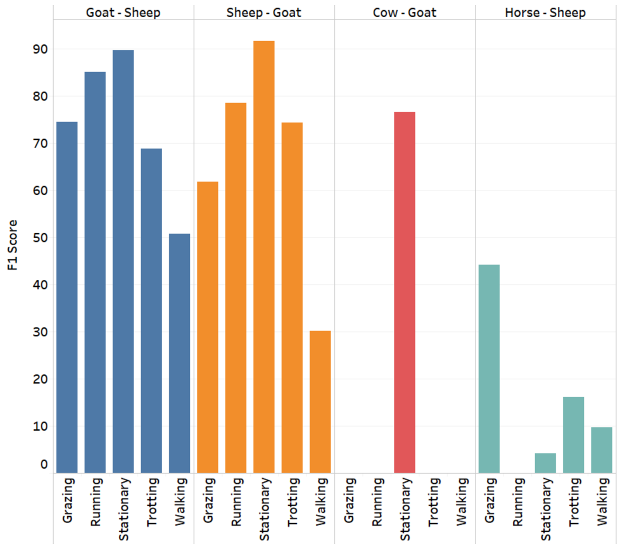

Figure 7 The F1 scores per activity of the category one experiments that showed the highest accuracy.

6. DISCUSSION

6.1 Data Distribution

The distribution of the data, meaning the size of the data set and the number of examples of activities, is of influence on the performance of the classifier. When a classifier is trained with data where some activities occur rarely, one cannot expect that this activity will be classified correctly if it occurs more often in the test data. The dataset of the cow was very small. Due to the malfunctioning sensors and the time limit of this research, there was no chance to collect more data. Moreover, the distribution of the data was not good. There was a good division between standing and lying but these were mapped together to stationary. The result is that there is a lot of data for stationary but very little on walking. For example, this led to the consequence that the classifier trained on cow data only classifies stationary activities and not walking activities. From the horses, there are roughly 13.000 datapoints. This is a more acceptable amount of datapoints, however still not comparable to the amount of goat and sheep data available.

6.2 Parameters

6.2.1 Calculation of features

The raw sensor data had to be transformed to import it in RapidMiner and to apply classifiers on it. For this a script was used developed by J. Kamminga. In this script, the data is divided into windows and different features are calculated. However, there are some parameters to be set to perform this process. The number of folds, the overlap, the window size and the segment length all must be decided. The number of folds depends on the size and distribution of the data set. The folds are made to make sure that every datapoint is once used to test, validate and train the classifier. The overlap is set at 0.5 and that is to overlap windows and to ensure that no activity is missed. The window size is set at two seconds. In this research it was necessary to keep it at this value to compare it to the previously collected data. The segment length is a parameter that can be varied. When the size is decreased, the chance is bigger that datapoints are lost but the chance of missing events is smaller. This was set at 20 seconds but should be investigated more. However, further testing of these parameters was out of the scope of this research.

6.2.2 Parameter optimization

The parameters that are used by the different classifier have been optimized by a RapidMiner implementation to increase the performance of the classifier. This was done by looping over the classifier and trying out different values of the parameters and then evaluating the performance. The settings of the optimize parameter algorithm are stated in the methods section. However, these can also be changed and will affect the process of optimizing the parameters and could possible even change the outcome. It was out of the scope of this research to research the influence of this on the results. However, this is a factor that should be noted and thought of. This could potentially be further researched upon.

6.3 Results

6.3.1 Different Activities

The activities that are mentioned in Table 2 were all observed during the data collection. All four animal species performed these activities. This answers one of the sub questions that indeed they show comparable behaviour. They all performed some species-specific activities as well such as scratch-biting, fighting, rolling etc. However, not every species exhibited such behaviour so for the sake of this research, these were let out of the comparison.

6.3.2 Count of Activities

Table 5 shows the activity counts that were collected and cleaned. As expected, the dataset from the goats and sheep is much larger than the data from the horses and the cows. It was very disappointing that the sensor from the cows corrupted the sensor data which led to this low count of activities. The dataset was quite small, and there were only two activities recorded, of which walking in a very small fraction. The horse dataset was larger, and the activities were better distributed. However, compared to the other animals there were few datapoints for the activity 'stationary'.

6.3.3 Accuracy of the Classifiers

The accuracy of each classifier is shown in Figure 4 and 5. It is clear to see from the figure that for all the non-mixed scenarios the average accuracy of the kNN classifier is the highest. The value of the kNN classifiers is not very high, however higher than the rest of the classifiers.

For the individual classifiers, the Neural Network and the K-Nearest Neighbour are the most successful. After these results, it was decided to focus further on the K-Nearest Neighbor and for the individual scenario on the Neural Network.

It seems that the overall accuracy of the classifiers on the experiments of the *individual category* is around 83%. It is hard to say what an acceptable accuracy is considering it depends on the domain of the research. However, in other research on animal classification it is considered that around 85 – 90% is acceptable [2,4]. This would make the individual classifiers just acceptable. The classifier that were applied to the experiments of *category one, two and three* are far from acceptable. They perform below 50%. This is part due to the data distribution which is explained above in this section since several combinations do perform good.

Something that shows from Figure 4 and 5 is that with the increase of species in the training set, the accuracy increases. This is an answer to sub question four on which parameters increase the accuracy. The other parameters that increase the accuracy are optimized in the process and discussed in the methods section. Apparently, this research gives rise to the suspicion that when the number of species is increased, the accuracy improves.

6.3.4 Activity Recognition

Once the best classifier was found, the focus was on the specific activities that were classified. The last column of Table 6 shows the average percentage of the F1 score which averages the recall and precision. It shows that on average the activity of stationary is recognized the best with a percentage of 56.7%. After stationary, grazing is recognized best with an average of 48.6%. Running is recognized the worst on average with 32.9%. This is an answer to the sub question of which activities are recognized the best and the worst.

This result is quite logical. Stationary is the activity of which a lot of data was available to be trained and tested with. When an animal is stationary, it does not move or very little. This is universal for every animal and the vector of the accelerometer looks almost identical for every animal.

Running is recognized the worst from all the activities. An explanation for this could be that the top speed for every species varies from 17 km/h to 64 km/h and the time in which they achieve this speed varies as well. This means that the accelerometer data for these species differs when running.

6.3.5 Species Combinations

From Figure 6, it is visible that the combination of goat-sheep as well as sheep-goat are performing best. Following, the combinations of cow-goat and cow-sheep are best. To further indicate how well specific activities were recognized, the F1 scores of the best combinations were plotted. This is Figure 7. It shows that the combinations sheep-goat and goat-sheep both classify stationary the best followed by running. In the combination cow-goat, stationary is most accurately classified. This is due to the data distribution of the cow data set that only stationary is recognised. Thus, no conclusions can be drawn from this combination. Horse-sheep was the best performing combination of the classifier trained with horse data. It is seen that grazing is the activity that is best classified.

7. CONCLUSION

The behaviour of cows, sheep, goats and horses showed enough of the same activities to compare. The following activities were shown and compared: stationary, walking, trotting, running and grazing. The best classifier for the individual species was the Neural Network. For the non-mixed species experiments, the k-Nearest Neighbor algorithm showed the most promising results. On average, the non-mixed classifier performed poorly; around 30-50%. However, the performances of the classifiers rely heavy on the combination of test and train animal species. The best combinations are sheep-goat, goat- sheep and cow-goat. These classifiers have an accuracy between 60% and 90%.

The activities which are recognized the best are stationary and grazing. These have been analysed with the recall, precision and F1 values. On average, stationary is classified correctly 56.7% of the time. Grazing is classified correctly 49% of the time. These percentages are not high but show a significant increase from the other activities.

The results of this research show that when the number of species that the classifier uses to train with is increased, the performance of the classifier is increased. Other parameters that are used to optimize the performance of the classifier, have been optimized in RapidMiner and are described in the method section.

In conclusion, this research supports the development of a generic non-mixed classifier for quadruped animals. It showed for the used dataset that when the number of species is increased in the training set, the classifier performs better. Specific activities such as grazing, and stationary are recognized better than others, due to the movement of the activities are the same for several species.

In future work, we want to collect more data from horses and cows. Due to the size of the data sets and the distributions of these two species, no hard or general conclusions could be drawn. This research, however gives some support to the possibility of developing a generic classifier and further research seems in place. It would be useful to collect more data from more different species to further research if the accuracy keeps increasing when species are added to the training set.

8. ACKNOWLEDGEMENT

First, I want to thank Johan Jurrius from “Hoeve Doorwerth” for facilitating the collection of data from cows. I want to thank Jacob Kamminga for supervising and enabling me to perform and write my research paper. Your enthusiasm about data science was truly contagious and made this a very fun process for me. Lastly, I want to thank Nirvana Meratnia for supervising and guiding me in this academic research.

9. REFERENCES

[1] Champion, R.A. et al. 1997. An automatic system to monitor lying, standing and walking behaviour of grazing animals. *Applied animal behaviour science*. 54, 4 (Nov. 1997), 291–305. DOI:https://doi.org/10.1016/S0168-1591(96)01210-5.

[2] Cvetkovic, B. et al. 2011. Semi-supervised learning for adaptation of human activity recognition classifier to the user. *Proc. of Int. Joint Conf. on*. (2011).

[3] Dutta, R. et al. 2015. Dynamic cattle behavioural classification using supervised ensemble classifiers. *Computers and Electronics in Agriculture*. 111, (Feb. 2015), 18–28. DOI:https://doi.org/10.1016/j.compag.2014.12.002.

[4] Gao, L. et al. 2013. A Web-based semantic tagging and activity recognition system for species’ accelerometry data. *Ecological informatics*. 13, (Jan. 2013), 47–56. DOI:https://doi.org/10.1016/j.ecoinf.2012.09.003.

[5] González, L.A. et al. 2015. Behavioral classification of data from collars containing motion sensors in grazing cattle. *Computers and Electronics in Agriculture*. 110, (Jan. 2015), 91–102. DOI:https://doi.org/10.1016/j.compag.2014.10.018.

[6] Grünewälder, S. et al. 2012. Movement activity based classification of animal behaviour with an application to data from cheetah (Acinonyx jubatus). *PloS one*. 7, 11 (Nov. 2012), e49120. DOI:https://doi.org/10.1371/journal.pone.0049120.

[7] Human Activity Monitor: *http://www.gcdataconcepts.com/ham.html*. Accessed: 2018-06-27.

[8] Kamminga, J.W. et al. 2017. Generic online animal activity recognition on collar tags. *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (Sep. 2017), 597–606.

[9] Kamminga, J.W. et al. 2018. Robust Sensor-Orientation-Independent Feature Selection for Animal Activity Recognition on Collar Tags. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. 2, 1 (Mar. 2018), 15. DOI:https://doi.org/10.1145/3191747.

[10] Koeien - Nedap livestock management: *http://nl.nedaplivestockmanagement.com/oplossingen/koeien/*. Accessed: 2018-05-13.

[11] Martiskainen, P. et al. 2009. Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines. *Applied animal behaviour science*. 119, 1 (Jun. 2009), 32–38. DOI:https://doi.org/10.1016/j.applanim.2009.03.005.

[12] MATLAB. 2018. version 9.4.0 (R2018a). The MathWorks Inc., Natick, Massachusetts.

[13] Molyneux, J. et al. 2018. The efficacy of monitoring techniques for detecting small mammals and reptiles in arid environments. *Wildlife research* . 44, 7 (Jan. 2018), 534–545. DOI:https://doi.org/10.1071/WR17017.

[14] O'Donoghue, P. and Rutz, C. 2016. Real-time anti-poaching tags could help prevent imminent species extinctions. *The Journal of applied ecology*. (2016).

[15] Le Roux, S.P. 2016. *A Prototype Animal Borne Behaviour Monitoring System*. University of Stellenbosch Supervisor: Dr. R. Wolhuter March.

[16] Inertia Technology. 2017. ProMove mini. online. (2017). Retrieved from: http://inertia-technology.com/

[17] RapidMiner 2018, Version 8.1.0. Rapidminer, Dortmund.

[18] Sahin, Y.G. 2007. Animals as Mobile Biological Sensors for Forest Fire Detection. *Sensors* . 7, 12 (Dec. 2007), 3084–3099. DOI:https://doi.org/10.3390/s7123084.

[19] Siirtola, P. and Röning, J. 2012. Recognizing human activities user-independently on smartphones based on accelerometer data. *IJIMAI*. 1, 5 (2012), 38–45. DOI:https://doi.org/10.9781/ijimai.2012.155.

[20] Vázquez Diosdado, J.A. et al. 2015. Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry*. 3, 1 (Jun. 2015), 15. DOI:https://doi.org/10.1186/s40317-015-0045-8.

[21] Xu, J. et al. 2015. Animal monitoring with unmanned aerial vehicle-aided wireless sensor networks. *2015 IEEE 40th Conference on Local Computer Networks (LCN)* (Oct. 2015), 125–132.

[22] 2016. *Living planet report*. WWF World Wild Fund

10. APPENDIX

Table 7 The parameters and the range of values in which they are varied per classifier.

Classifier Name	Parameter Name	Range	Parameter Name	Range	Parameter Name	Range	Parameter Name	Range	Parameter Name	Range
NN	Learning rate	4.9E-324 - 1	Momentum	0 - 1						
DT	Maximal depth	1 - 100	Minimal gain	1 - 100	Pruning confidence	1x 10^{-7} – 0.5	Minimal size for split	1 - 100	Minimal leaf size	1 - 100
SVM	Gamma	1 - 100	Cache size	1 - 100	Cost function	1 - 100				
kNN	K / Number of Neighbors	1 -100								
DNN	L1, L2	0 - 1	Learning rate Annealing, Learning rate Decay	0 - 1	Learning Rate	0 -1	Number of Epochs	0 -100	Momentum stable, Momentum ramp, Momentum start	0 -100