TOWARDS A ROBUST ANALYSIS AND CLASSIFICATION OF DOG BARKING

Maja Schneider[1] *and Oliver Jokisch*[2]

[1]*Department of Computer Science, Universität Leipzig, Germany*
[2]*Institute of Communications Engineering, HfT Leipzig, Germany*
majaschneider@gmx.de, jokisch@hft-leipzig.de

Abstract: The analysis of animal sounds or even communication is an emerging research topic, e.g. in biodiversity research, climate studies or digital farming. Considering animal sounds in a natural environment, it becomes clear, that the underlying signal processing may be quite challenging, e.g. by a low signal-to-noise ratio due to a large microphone distance or other acoustic peculiarities, e.g. additional sound sources. Furthermore, the classification of the signals depends on the availability and interpretability of appropriate (and annotated) sound data, e.g. representative recordings of dog barking in our contribution. We investigated, whether specific dog barking can be distinguished from silence or other sounds, like animal or traffic noise, to control a window-closing mechanism in a smart home scenario. The sound recordings have been collected and improved with a wavelet de-noising technique and notch filters. The analysis included varying analysis frames between 21 and 168 ms, and up to 8,239 temporal or spectral features that are reduced to a set of 51 features by a Linear Discriminant Analysis (LDA). Additionally, we applied a Correlation-based Feature Selection (CFS) method. We then classified the samples by various methods, namely AdaBoost, Random Forest, Support Vector Machine (SVM), Multi-layer Perceptron (MLP) and decision tree C4.5. Preliminary results show the best performance for a selection of all 51 features (after LDA) without any CFS, based on analysis frames of 21 ms. The described methods are useful to detect barking of one specific dog.

1 Introduction

Most studies on animal sounds focus on the identification of species in amphibian and bird sounds during the investigation of biodiversity or ecological health of an environment. For example, Acevedo et al. [1] compared several methods of machine learning for different bird and amphibian calls. Further studies show that automated classification of animals is relevant for animal behavior research. For example Oliver et al. [2] develop a classifier to research the breeding times of certain bird species. Automated studies of this kind will in future provide more detailed information about the population size, the sex, the age structure and social relations of animals living together in social groups.

Studies on dog barking however are mainly investigating the communicative role of such sounds. Feddersen-Petersen [3] studied this question already in 2000, while Molnár et al. [4] investigated possible context-specific and individual features of dog bark by applying a machine-learning algorithm. Molnár et al. [4] conclude that dog barking is context-dependently distinguishable by humans as well as by the means of machine-learning algorithms, for example when classifying situations like fight, loneliness or the presence of strangers. Molnár et al. [5] and Pongrácz et al. [6] prove that also dogs themselves are capable of determining the context

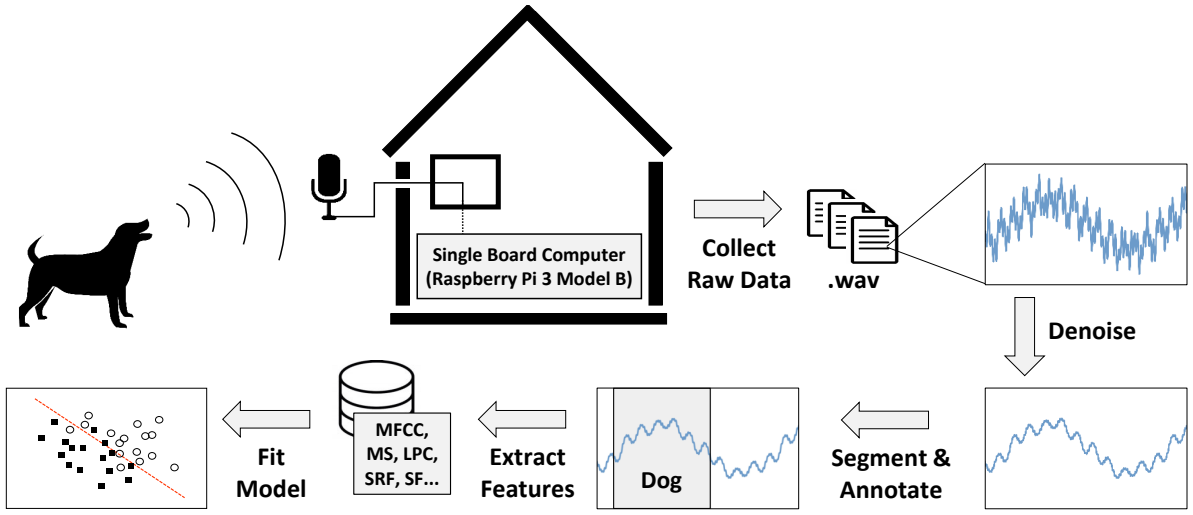


Figure 1 – Block scheme of signal enhancement, feature analysis, and classification experiments in [12]

in which such sounds appear. The question if animals are capable of exchanging information acoustically, has been researched in many studies. For example Sueur [7] and Janik [8] reveal that animals like cicada or dolphins are able to distinguish their own species from others or even to detect individuals by sound. Further studies evidence that some animals can acoustically transmit information about an approaching danger, for example suricates, monkeys or elephants [9, 10, 11].

Our contribution to animal sound classification addresses the recognition of dog barking in the context of a village environment, including low signal-to-noise ratios (SNR) due to long-distance recordings. The audio data, methodology and preliminary results are based on the experiments in [12]. In a first step of the analysis, we created a database of 260 h at one microphone position. Afterwards, we segmented and labeled the data with Audacity [13], leading to 38 min of annotated audio data, which was used to tune and compare five different machine-learning models with stratified threefold cross validation.

Section 2 describes our experimental methods including audio data collection, signal enhancement, segmentation/annotation, and feature extraction. Section 3 gives an overview about the data sampling and the evaluation process. Afterwards, we describe the selection and implementation of the machine-learning models and the tuning results. In section 4, we present and discuss our test results. Section 5 draws conclusions and gives an outlook on future research.

2 Methods and data

2.1 Recording infrastructure and data collection

The audio recordings were made in the Saxon village of Sohland/Spree with a single, omnidirectional condensor microphone (Notebook mini-micro “Hama”) connected to a Raspberry Pi 3, Model B[1], running with Raspbian. The micro, with a frequency range from 30 Hz to 16 kHz, was placed on the outside wall of a residential house on the first floor in approximately 750 m distance from a frequently barking, single dog, and wrapped with a cotton-wool ball to attenuate wind noises. Also the auto-gain control was therefore enabled.

Figure 1 depicts our data collection and processing pipeline from [12]. The audio recordings were automatically scheduled every night between midnight and 6:30 AM, from January to June 2018, resulting in more than 1200 hours. The data was sampled at 48 kHz and stored

[1] https://www.raspberrypi.org

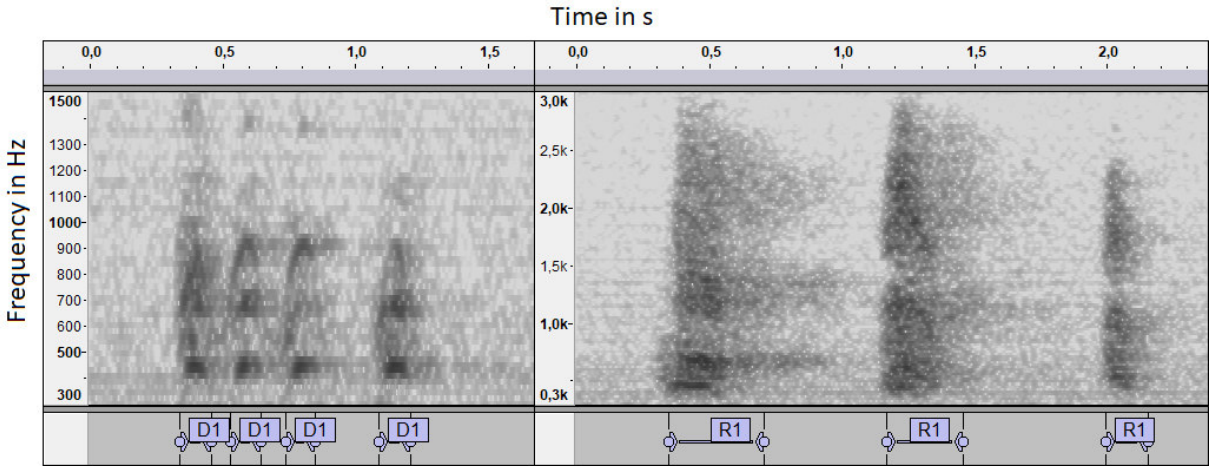


Figure 2 – Examples of the signal annotation in Audacity – left: spectrogram of consecutive dog barks (annotated with subclass D1 for very loud barks) and right: spectrogram of roebuck barks (subclass R1)

in 16 bit one-channel wav files of 30 min duration each. The recordings contain barks from one specific dog (*German Shepherd*), sounds from other animals like roebucks, crickets and a variety of singing and night birds, traffic noises from cars, trains and airplanes, other noises like door slamming, human voices, music, wind, rain or silence. The variety of sound sources leads to a representative dataset as a key factor for more robust models and to avoid data bias.

2.2 Manual data segmentation and annotation

In [12], 20 % of the recorded data (ca. 260 h of the available recordings) were randomly chosen, and searched for interesting sounds. The sounds were manually segmented and annotated into the five classes *dog barking*, *traffic*, *random noises*, *other animals*, and *silence* using the audio editing software Audacity [13], version 2.1.3[2]. All recordings were analyzed and classified based on the subjective perception. Furthermore, there is a subclass annotation with a finer granularity (e.g. animal species) to facilitate stratified sampling and to avoid the potential predominance of one specific subclass in a class. Finally, the classes were binary assigned to the target class *Dog* or anything else, called *Other*.

Figure 2 illustrates an example of the segmentation and annotation of two audio snippets, containing several consecutive dog and roebuck barks. After the segmentation process ca. 38 min of annotated audio data was available for further processing, containing (amongst other sounds) about 6,800 single dog barks.

2.3 Audio processing, filtering and SNR improvement

The challenging, long distance of 750 m between microphone and barking dog is causing a poor sound quality in the recorded signals, in which only 3 % of the possible bit depth was occupied and with a lot of interfering noise, that contributes to a low SNR of about 10 to 16 dB in the dog-barking parts. With regards to our application scenario, the microphone position was considered fixed and could not be moved closer to the sound source (to eventually reduce the influence of environmental noise).

To improve the SNR, two denoising techniques for the remaining noise were tested, reducing the mains hum (50 Hz and corresponding harmonics such as 100 Hz, 150 Hz) and other, e.g. thermal noise. To attenuate the harmonic parts, a notch filter with narrow bandwidths of 10 Hz every 50 Hz starting from 50 Hz up to 1.1 kHz was applied.

[2]https://audacityteam.org

The thermal noise in the recorded signal can be roughly characterized as colored noise ranging up to 6 kHz, which superimposes the dog-barking frequencies that typically cover the range from 400 Hz to 1 kHz. We used a wavelet-based denoising by a Daubechies wavelet of fourth order, and applied a hard-threshold approach to audio segments of 5 min duration.

In audio samples, containing dog barking with different degrees of loudness, the notch filter increases the SNR by 45 %, while the wavelet denoising contributes with improvements of 3 % only. Overall, the SNRs can be increased to a range from 16 to 22 dB in dog barks.

2.4 Feature extraction, selection and dimension reduction

We extract the features by a sliding window analysis. As the appropriate window length for a barking recognition is not surveyed yet, we compare the classification for three different window lengths (21, 84 and 168 ms). The lengths are chosen to correspond to at most the average length of a single dog bark, retrieved from the dataset, and to capture a sample count close to the next power of two. The sliding analysis is performed with a window overlap of 75 % for the lengths of 84 ms and 168 ms, and without overlap for the 21 ms window.

For each window, we calculate temporal and spectral features. The spectral features are obtained by applying a Short-time Fourier Transform (STFT) using a Hamming window with 1,024, 4,096 or 8,192 samples, depending on the analysis window. Our selected features have proven to be useful in the aforementioned studies, e.g. [4, 14, 15], on animal-sound recognition by means of Mel-frequency Cepstrum-coefficients (MFCC), Magnitude Spectrum (MS), Linear Predictive Coding (LPC) coefficients, Formants, Spectral Rolloff Frequency (SRF) or Strongest Frequency (SF).

In total 8,239 features are extracted and can be later reduced by applying Linear Discriminant Analysis (LDA) on some multidimensional features like MS or MFCC, which reduces the dimension to 51 features. Applying Correlation-based Feature Selection (CFS) [16] yields feature sets from 10 to 13 features. The classification results are compared for both strategies, with and without CF selection. The overall feature-extraction accounts for a database from 50k to 110k samples for each of the three window lengths.

3 Classification experiments

3.1 Evaluation concept and data partition

The dog-barking database is subdivided into portions of 50 % training, 25 % validation and 25 % test data, on which we performed threefold cross validation. The training and validation data sum up to 17k ... 41k samples in the *Dog* class and 5k ... 11k samples in each *Other* subclass. The folds are sequentially constructed from consecutive windows (of the same audio file) that are likely to have a high correlation, and therefore should not be distributed to the training and validation folds at the same time, to avoid any positive bias [17].

Due to the limited amount of data we did not consider algorithms from deep learning. Aiming at a proof of concept, we decided to begin just with a binary classification of the samples into *Dog* and *Other* by means of Random Forest, AdaBoost, Decision Tree C4.5, Support Vector Machine (SVM) or Multilayer Perceptron (MLP). The experiments were processed with the backend interface of the machine-learning framework "Weka" [18].

The evaluation was executed with a self-designed performance metric based upon false positive rate (FPR) and sensitivity, as FPR would be a relevant measure in a smart-home scenario, in which an overhasty window-closing event without barking can also disturb. Sensitivity, on the other hand, is a good measure about how often the system recognizes dog barks successfully. Our metric is defined by the harmonic mean of the inverse FPR and sensitivity.

3.2 Random Forest

For Random Forest classification, we used the Weka implementation of Breiman [19]. We tested different parameters in a grid search, such as the number of trees (up to 2,500, step size of 300) and the number of randomly chosen features (1 to 12). As basis learner, a Random Tree, introduced by Ho [20], chooses a random feature set at every tree node. Following the one-standard-error approach, the most reasonable model presents itself with five randomly chosen features at each node and 50 trees.

3.3 AdaBoost

As base learner in the AdaBoost M1 algorithm of Freund and Schapire [21], we specified Decision Stump [22], a simple decision tree with only two leaves. The main tuning parameter is the number of iterations, which we tested in the range from 5 to 100 by steps of 5, and from 200 to 1,000 by steps of 100. The most reasonable model is achieved at an analysis window length of 21 ms and with a full feature set without feature selection, that required 30 iterations.

3.4 Decision Tree C4.5

The C4.5 by Quinlan [23] involves a gain-ratio metric and can specify pruning strategies, such as reduced-error pruning (REP), pessimistic-error pruning (PEP) or no post-pruning at all. For PEP, the "aggressivity" of pruning is steered by a confidence factor, for which we tested values of 0.005, 0.05, 0.1, 0.25 and 0.5. The minimal amount of instances per leaf is also effecting the pre-pruning strategy.

All pruning strategies achieve the best results at an analysis window of 21 ms length and by applying feature selection (CFS). The results suggest, that the highest performance can be achieved by either constructing a complex tree and applying heavy pruning or by constructing a rather simple tree without further or just slight pruning. The most reasonable model allows at least 390 instances per leaf and involves PEP with a confidence factor of 0.001.

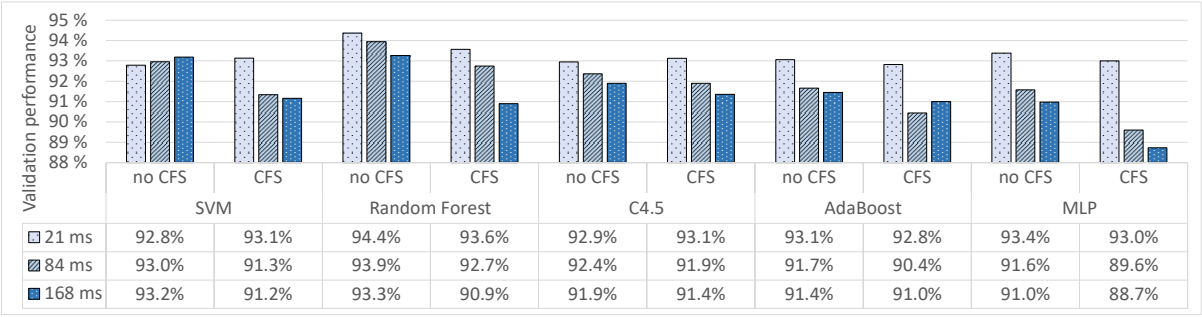
3.5 Multilayer Perceptron (MLP)

We tested MLPs with a first layer of 2 to 9 neurons, and 10 to 200 neurons with a step size of 10. An optional second layer contained 1 to 5 neurons. The learning rate was set to 0.1 or 0.2. Most of the parameters were set to the default value, suggested by Weka, which includes backpropagation and a sigmoid activation function.

The intermediate results again suggest a short analysis window of 21 ms without applying feature selection. The most reasonable model has eight neurons in the first and four neurons in the second layer.

3.6 Support Vector Machine (SVM)

For SVM parameter tuning, we used the libsvm implementation of Chang and Lin [24] with an RBF kernel and a grid search for the cost $C \in \left\{2^{-5}, 2^{-3}, \ldots, 2^{15}\right\}$ and $\gamma \in \{2^{-15}, 2^{-13}, \ldots, 2^{3}\}$. A finer search was conducted after the appropriate parameter ranges were identified. The most reasonable model was found for an analysis window of 168 ms without feature selection, a cost of 0.015625, and a γ of 0.25. The second best model included a 21 ms analysis window, which is in line with the previously observed trend.



	SVM		Random Forest		C4.5		AdaBoost		MLP	
	no CFS	CFS	no CFS	CFS	no CFS	CFS	no CFS	CFS	no CFS	CFS
21 ms	92.8%	93.1%	94.4%	93.6%	92.9%	93.1%	93.1%	92.8%	93.4%	93.0%
84 ms	93.0%	91.3%	93.9%	92.7%	92.4%	91.9%	91.7%	90.4%	91.6%	89.6%
168 ms	93.2%	91.2%	93.3%	90.9%	91.9%	91.4%	91.4%	91.0%	91.0%	88.7%

Figure 3 – Maximal validation performance (hyper-parameter tuning) per model & analysis window, with & w/o CFS. (Results refer to max. performance and do not reflect one-standard-error approach.)

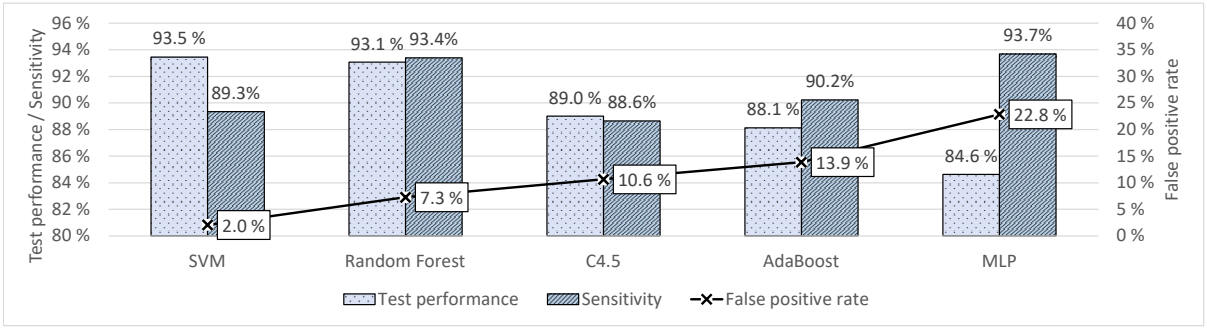


Figure 4 – Comparison of test performance, sensitivity and false positive rate for the five tested models

4 Results and discussion

Figure 3 summarizes the maximal validation performance by hyper-parameter tuning in comparison for the three window lengths, and for both cases, with and without the Correlation-based Feature Selection. All models achieve their best overall performance at 21 ms analysis length, except for SVM with slightly better results for 84 ms and 168 ms windows if not applying CFS. Except for C4.5, an additional CFS does not improve the respective best results obtained with a full set of 51 features.

For the analysis window of 21 ms, the results with and without CFS are comparably good for all models. The experimental results suggest that a CFS becomes disadvantageous for longer analysis windows.

The final performance by training all models with the full dataset from training and validation, and testing them with the separate test set, is shown in figure 4, at which also the one-standard-error approach to reduce overfitting is applied. SVM and Random Forest demonstrate the best performance, while SVM achieves a good false-positive rate of 2 % as an option for a smart-home device targeting an automatic window-closing (triggered by dog barking) and a sufficiently good sensitivity of 89.3 %. All models obtain acceptable or good sensitivities, but the FPRs vary considerably with values over 10 % (C4.5, AdaBoost) and even more than 20 % (MLP), which is not practicable for the intended target application.

The validation performance is consistent with the test performance except for MLP, which shows a significant increase of the FPR in our test. The test performances of models, created by the one-standard-error approach, is slightly lower compared to the ones without this approach.

5 Conclusions

We have investigated, whether specific dog barking can be distinguished from other sounds and silence to acoustically control a window-closing mechanism in a smart home. Therefore, we compared five machine-learning models (AdaBoost, Random Forest, SVM, MLP, C4.5), three analysis windows (duration of 21, 84 or 168 ms) and a set of 51 temporal and spectral features as well as a supplementary feature selection with the CFS algorithm.

Our results suggest, that the most promising strategy should involve SVM or Random Forest with an analysis window of 21 ms and a feature set of 51 temporal and spectral features. Against expectations, an additional feature selection does not improve the recognition results in most of the cases. Both models have also been successfully applied in the aforementioned studies to classify bird or amphibian calls.

Machine-learning algorithms are able to detect specific dog-barking in a village environment. Our future work will focus on a species-independent dog-barking recognition in flexible environments.

References

[1] ACEVEDO, M. A., C. J. CORRADA-BRAVO, H. CORRADA-BRAVO, L. J. VILLANUEVA-RIVERA, and T. M. AIDE: *Automated classification of bird and amphibian calls using machine learning: A comparison of methods*. *Ecological Informatics*, 4(4), pp. 206–214, 2009.

[2] OLIVER, R. Y., D. P. W. ELLIS, H. E. CHMURA, J. S. KRAUSE, J. H. PÉREZ, S. K. SWEET, L. GOUGH, J. C. WINGFIELD, and N. T. BOELMAN: *Eavesdropping on the arctic: Automated bioacoustics reveal dynamics in songbird breeding phenology*. *Science Advances*, 4(6), 2018.

[3] FEDDERSEN-PETERSEN, D. U.: *Vocalization of european wolves (canis lupus lupus l.) and various dog breeds (canis lupus f. fam.)*. *Archiv für Tierzucht*, 43(4), pp. 387–397, 2000.

[4] MOLNÁR, C., F. KAPLAN, P. ROY, F. PACHET, P. PONGRÁCZ, A. DÓKA, and A. MIKLÓSI: *Classification of dog barks: a machine learning approach*. *Animal Cognition*, 11(3), pp. 389–400, 2008.

[5] MOLNÁR, C., P. PONGRÁCZ, T. FARAGÓ, A. DÓKA, and A. MIKLÓSI: *Dogs discriminate between barks: the effect of context and identity of the caller*. *Behavioural Processes*, 82(2), pp. 198–201, 2009.

[6] PONGRÁCZ, P., E. SZABÓ, A. KIS, A. PÉTER, and A. MIKLÓSI: *More than just noise? - field investigations of intraspecific acoustic communication in dogs (canis familiaris)*. *Applied Animal Behaviour Science*, 159, pp. 62–68, 2014.

[7] SUEUR, J.: *Cicada acoustic communication: potential sound partitioning in a multispecies community from mexico (hemiptera: Cicadomorpha: Cicadidae)*. *Biological Journal of the Linnean Society*, 75(379-394), 2002.

[8] JANIK, V. M.: *Acoustic communication in delphinids*. In NAGUIB M., ZUBERBÜHLER K., CLAYTON N. S., JANIK V. M (ed.), *Advances in the Study of Behavior*, vol. 40, pp. 123–157. Elsevier, 2009.

[9] MANSER, M. B., R. M. SEYFARTH, and D. L. CHENEY: *Suricate alarm calls signal predator class and urgency*. *Trends in Cognitive Sciences*, 6(2), pp. 55–57, 2002.

[10] SEYFARTH, R. M., D. L. CHENEY, and P. MARLER: *Monkey responses to three different alarm calls: evidence of predator classification and semantic communication*. *Science*, 210, pp. 801–803, 1980.

[11] KING, L., M. PARDO, S. WEERATHUNGA, T. V. KUMARA, N. JAYASENA, J. SOLTIS, and S. DE SILVA: *Wild sri lankan elephants retreat from the sound of disturbed asian honey bees*. *Current biology : CB*, 28(2), pp. R64–R65, 2018.

[12] SCHNEIDER M.: *Analyse und Klassifikation von Audiodaten zur Erkennung von Hundebellen*. Master's thesis, Universität Leipzig/HfT Leipzig, Germany, 2019.

[13] AUDACITY: Retrieved 28/01/2020. URL `https://audacityteam.org/`.

[14] STOEGER, A. S., M. ZEPPELZAUER, and A. BAOTIC: *Age group estimation in free-ranging african elephants based on acoustic cues of low-frequency rumbles*. *Bioacoustics*, 23(3), pp. 231–246, 2014.

[15] FAVARO, L., M. GAMBA, C. GILI, and D. PESSANI: *Acoustic correlates of body size and individual identity in banded penguins*. *PloS one*, 12(2), 2017.

[16] HALL, M. A.: *Correlation-based Feature Selection for Machine Learning*. Dissertation, The University of Waikato, Hamilton, New Zealand, 1999.

[17] SCHEIRER, E. and SLANEY M.: *Construction and evaluation of a robust multifeature speech/music discriminator*. In *ICASSP, IEEE Intern. Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2, pp. 1331–1334. Munich, Germany, 1997.

[18] FRANK, E., M. A. HALL, I. H. WITTEN, and C. J. PAL: *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"*. Morgan Kaufmann, 4 edn., 2016.

[19] BREIMAN, L.: *Random forests*. *Machine Learning*, 45, pp. 5–32, 2001.

[20] HO, T. K.: *Random decision forests*. In *Proceedings of the third International Conference on Document Analysis and Recognition*, vol. 1, pp. 278–282. IEEE Computer Society Press, Los Alamitos, CA, 1995.

[21] FREUND, Y. and R. E. SCHAPIRE: *Experiments with a new boosting algorithm*. *Machine Learning: Proceedings of the Thirteenth International Conference*, pp. 148–156, 1996.

[22] IBA, W. and P. LANGLEY: *Induction of one-level decision trees*. In *Machine Learning: Proceedings of the Ninth International Workshop (ML92)*, pp. 233–240. Elsevier, 1992.

[23] QUINLAN, J. R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1993.

[24] CHANG, C.-C. and C.-J. LIN: *Libsvm: a library for support vector machines*. *ACM Transactions on Intelligent Systems and Technology*, 2(3), pp. 1–27, 2011.