International journal of application or innovation in engineering and management

ijaiem.com/july 2021/ Volume 10/Issue 2/Article No-1/01-09
ISSN: 2319-4847

Combining Temporal Contextual Data to Provide Identification of Bird Species from Data

Yogita Khairnar, Nasik, Maharashtra. Yogita Khairnar @gmail.com

Article Info

Received:* *27-04-2021 ***Revised:* *11 -05-2021*** ***Accepted:* *07-06-2021*** ***Published:14/07/2021***

Abstract— We report on the development of an automated acoustic bird recognizer with improved noise robustness, which is part of a long-term project, aiming at the establishment of an automated biodiversity monitoring system at the Hymettus Mountain near Athens, Greece. In particular, a typical audio processing strategy, which has been proved quite successful in various audio recognition applications, was amended with a simple and effective mechanism for integration of temporal contextual information in the decision- making process. In the present implementation, we consider integration of temporal contextual information by joint post-processing of the recognition results for a number of preceding and subsequent audio frames. In order to evaluate the usefulness of the proposed scheme on the task of acoustic bird recognition, we experimented with six widely used classifiers and a set of real-field audio recordings for two bird species which are present at the Hymettus Mountain. The highest achieved recognition accuracy obtained on the real-field data was approximately 93%, while experiments with additive noise showed significant robustness in low signal-to-noise ratio setups. In all cases, the integration of temporal contextual information was found to improve the overall accuracy of the recognizer.

Keywords— Bioacoustics, Biodiversity Informatics, Acoustic Bird Species Recognition, Automatic Recognition

I. Introduction

Over the last years one of the most crucial issues that governments and international organizations have to deal with is the conservation of biodiversity. The protection of the endangered species is of prior importance for the conservation of biodiversity and is based primarily on the accurate monitoring of the biodiversity and secondarily on the application oftargeted conservation actions, which are based on the quantitative measures of the monitored biodiversity status. Major importance for the conservation of biodiversity has the observation and the monitoring of birds [1].

Significant amount of information about the activity of the birds has been collected by expert ornithologists. In this effort the ornithologists recognize the bird species from their vocalizations, study the interaction among them and locate their habitats. Such surveys require the repeated physical presence of expert ornithologists in the field and thus become time consuming and tedious. Moreover, the manual observations heavily rely on the visual and acoustic abilities of the surveyor as well as on the degree of his/her knowledge on the family of bird species which are under investigation. Finally, the difficulty of the task restricts most of the biodiversity monitoring surveys to take place in infrequent time intervals, especially for the hard to access areas, thus not allowing the long-term biodiversity monitoring of inhospitable habitats.

The above mentioned disadvantages of manual observations of the bird activity have led to the development and study of several approaches for automatic recognition of bird species from their vocalizations over the last decade. Automatic recognition of acoustic bird species falls in the pattern recognition task, which involves preprocessing and feature extraction of the audio signal and classification over the parameterized audio.

Several approaches in automatic bird species recognition from their vocalizations have been proposed, most of which share techniques widely used in speech and audio processing. Such techniques are the template-matching (dynamic time warping) [2, 3] and the hidden Markov models [4], which have extensively been used in the similar task of speech recognition. Hidden Markov models have been used in more recent studies [5-7], due to their well known structure. Neural networks have also been used for the recognition of bird vocalizations using spectral and temporal parameters of the audio signal [8, 9]. Other approaches use Gaussian mixture model based structures [6, 10, 11], support vector machines [12] and decision trees [13] for the recognition of bird songs. Other proposed classification schemes are based on sinusoidal modeling of bird syllables [14] and bird syllable pair histograms [15]. Different parametric representations for the bird vocalizations audio signals have been used, among which Mel frequency cepstral coefficients [5, 6, 16, 17] are the most widely used. Other audio features which have been proposed in the literature are the linear predictive coding [16], linear predictive cepstral coefficients [16], spectral and temporal audio descriptors [12], and tonal-based features [17].

In most of the previous studies on the task of bird species recognition from their vocalizations in-lab conditions of recordings were used, without the presence of real environmental noise [2, 3, 4, 6, 12, 13]. In exception of most of the published work, in [17] waterfall noise was added to bird recordings and it was shown that the recognition of bird sounds in noisy conditions reduces significantly the recognition performance. In this article, we evaluate several different machine learning algorithms on the task of bird species classification in real-field conditions, under the concept of AMIBIO project (LIFE08-NAT-GR- 000539: *Automatic Acoustic Monitoring and Inventorying of Biodiversity*, Project web-site: http://www.amibio-project.eu/).

The rest of this article is organized as follows. In Section 2, the bird species recognition task in real-field is presented. Section 3 offers description of the audio data used and the experimental setup that was followed in the present evaluation. In Section 4 the experimental results are presented. Section 5 concludes this work.

II. Acoustic Bird Species Recognition in Real-Field with Temporal Context Information

In automatic bird species recognition from audio data 24/7 monitoring of specific habitats is achieved, while the information needed for biodiversity monitoring, animal species population estimation and species behavior understanding is extracted. The recognition of bird species is an audio pattern recognition task, and in brief is structured in (i) the audio acquisition stage, (ii) the audio parameterization stage and (iii) the classification stage. When recognition of the bird species is performed in the birds habitats the captured audio signal includes interferences that are additive to the vocalizations of the bird species. Typical interferences that are met in such habitats are the rain, the wind, the sum of the leaves, vocaliza tions from other animal species of the habitat, sounds produces by human activities, etc. In general, for the recognition of species in such a noisy environment detection of audio intervals with bird vocalizations precedes the species classification stage. The concept is illustrated in Figure 1.

Briefly, the audio signal is captured by a microphone, next amplified and then sampled at 32 kHz, so that the wide frequency range of bird vocalizations from various species is covered. A precision of 16-bits per sample is used to guarantee sufficient resolution of details for the subsequent processing of the signal. After the audio acquisition stage the signal is decomposed to overlapping feature vectors of constant length, using spectral and temporal audio parameterization algorithms. The sequence of feature vectors is used as input to the bird activity detection block where the audio signal is binary segmented to intervals with or without bird vocalizations. Finally, the bird vocalization intervals are processed by the bird species classification block in order species-specific recognition to be performed.

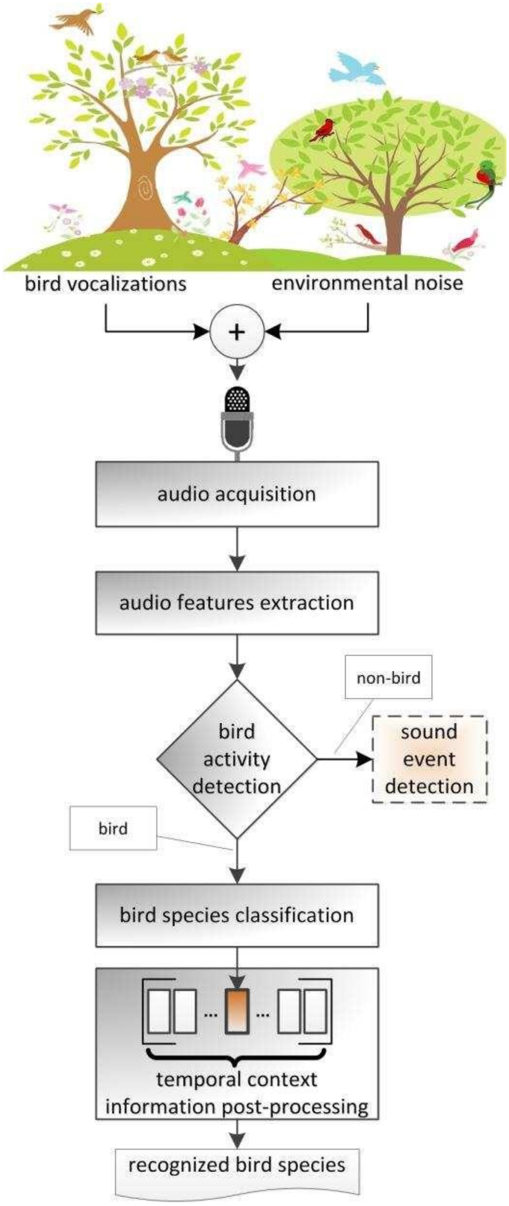

Fig. 1: Block diagram of the bird species recognition scheme in real- field conditions with temporal context information post-processing

Since classification is performed on frame level and each bird vocalization will appear in a number of consecutive audio frames the use of temporal context information for each frame as a post processing step would detect single frame misclassifications and improve the overall performance. In detail, this post- processing step aims at eliminating sporadic erroneous labeling of the current audio frame, e.g. due to momentary burst of interference, and thus contributes for improving the overall classification accuracy. The exploitation of the temporal contextual information, i.e. the labeled decision of the closest neighbor frames, is an effective way to detect and correct such sporadic erroneous frame labels. In particular, when the N preceding and the N successive audio frames, i.e. the temporal context of the current frame, are classified to one bird species vocalization then the current frame is also (re)labeled as of this bird species. The length w of the temporal context window is subject to investigation and in the general case it is equal to $w=2N+1$, where $N\geq0$. The case $N=0$, i.e. for temporal context window length w=1, corresponds to the elimination of the post- processing step of the classified labels.

In real-field the presence of non-stationary noises originating from the environment makes the species classification task more difficult and challenging. The degree of interference of the environmental noises and the actual signal-to-noise ratio are crucial for the recognition of bird species. In this work we also focus on the effect of the distance of the bird from the monitoring station (field microphone) as expressed by different signal-to-noise ratios, to the species classification performance.

III. Experimental Setup

A description of the audio data used in the present evaluation, the audio parameterization algorithms used, the machine learning classification algorithms that were tested and the experimental protocol that was followed are provided in this section.

The dataset used in the present article consists of recordings of two bird species which are known to be present at the Hymettus Mountain, a Natura 2000 site in Attica, Greece, namely the Eurasian Chaffinch (*Fringilla coelebs*) and the Common Kingfisher (*Alcedo atthis*). The recordings of the vocalizations of these bird species have been collected and manually labeled by expert ornithologists of the Zoologisches Forschungsmuseum Alexander Koenig (*ZFMK*). In order to test the bird species classification performance in different signal-to-noise ratios randomly selected recordings from the Hymettus area (from four different locations) were interfered to the bird vocalizations as additional noise. The amount of audio data used in the evaluation was approximately 14 minutes of recordings for all bird species.

The parameterization of the audio signals was performed using a diverse set of audio parameters. In particular, the audio signals were blocked to frames of
20 milliseconds length with 10 milliseconds time shifting step. Two temporal and sixteen spectral audio descriptors were used. The two temporal audio descriptors which were used are the frame intensity (*Int*) and the zero crossing rate (*ZCR*). The sixteen spectral audio descriptors which were used are the 12 first Mel frequency cepstral coefficients (*MFCCs*) as defined in the *HTK* setup [18], the root mean square energy of the frame (*E*), the voicing probability (*Vp*), the harmonics- to-noise ratio (*HNR*) by autocorrelation function and the dominant frequency (*Fd*) normalized to 500 Hz. The *openSMILE* acoustic parameterization tool [19] was used for the computation of the spectral audio parameters. After the computation of the audio parameters a post-processing with dynamic range normalization was applied to all audio features in order the range of their numerical values to be equalized.

A number of different machine learning algorithms were examined in the evaluation of the bird species classification step: (i) the k-nearest neighbors classifier with linear search of the nearest neighbor without weighting of the distance – here referred as instance based classifier (*IBk*) [20], (ii) a 3-layer Multilayer perceptron (*MLP*) neural network with architecture 18– 10–1 neurons (all sigmoid) trained with 50000 iterations [21], (iii) the support vector machines utilizing the sequential minimal optimization algorithm (*SMO*) with a radial basis function kernel [22], (iv) the pruned C4.5 decision tree (*J48*), with 3 folds for pruning and 7 for growing the tree [23], (v) the Bayes network learning (*BayesNet*) using a simple data-based estimator for finding the conditional probability table of the network and hill climbing for searching network structures [24],
(vi) the Adaboost M1 method (*Adaboost(J48)*) using the pruned C4.5 decision tree as base classifier [25], and
(vii) the bagging algorithm (*Bagging(J48)*) for reduce of the variance of the pruned C4.5 decision tree base classifier [26].

For the implementations of these algorithms the *Weka* software toolkit [24] was used. For all the above mentioned evaluated algorithms the values of the undefined parameters have been set equal to the default ones.

IV. Experimental Results

A common experimental protocol was followed in all experiments as described in Section 3. Ten-fold cross validation experiments were performed on the audio data which were described in the previous section, thus resulting to non-overlapping training and test data subsets. In Figure 2, the performance of the classification algorithms, in frame level, for various signal-to-noise ratios is shown.

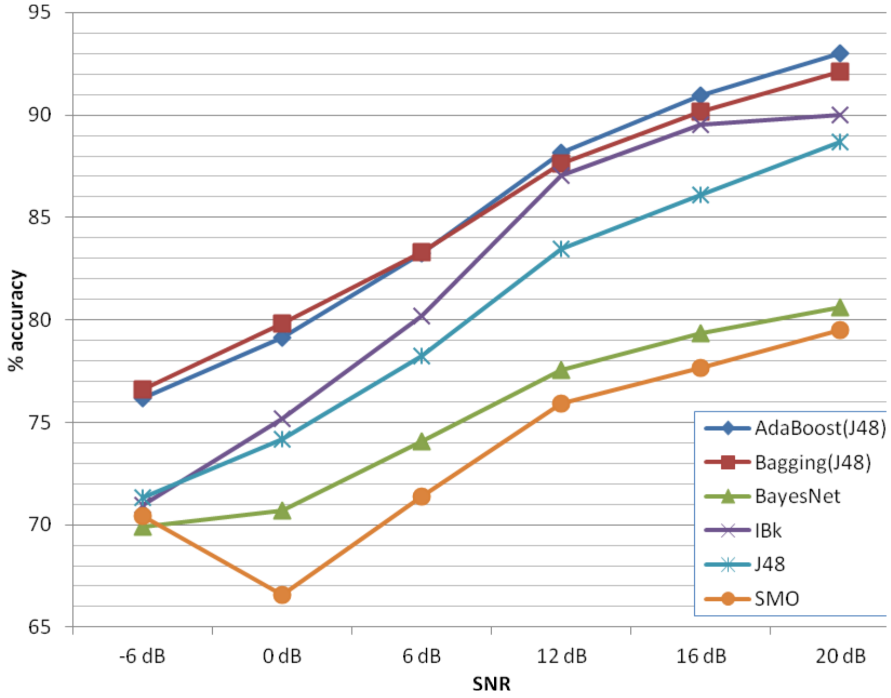

Fig. 2: Accuracy rate (in percentages) of bird species recognition for different classification algorithms and various signal-to-noise ratios

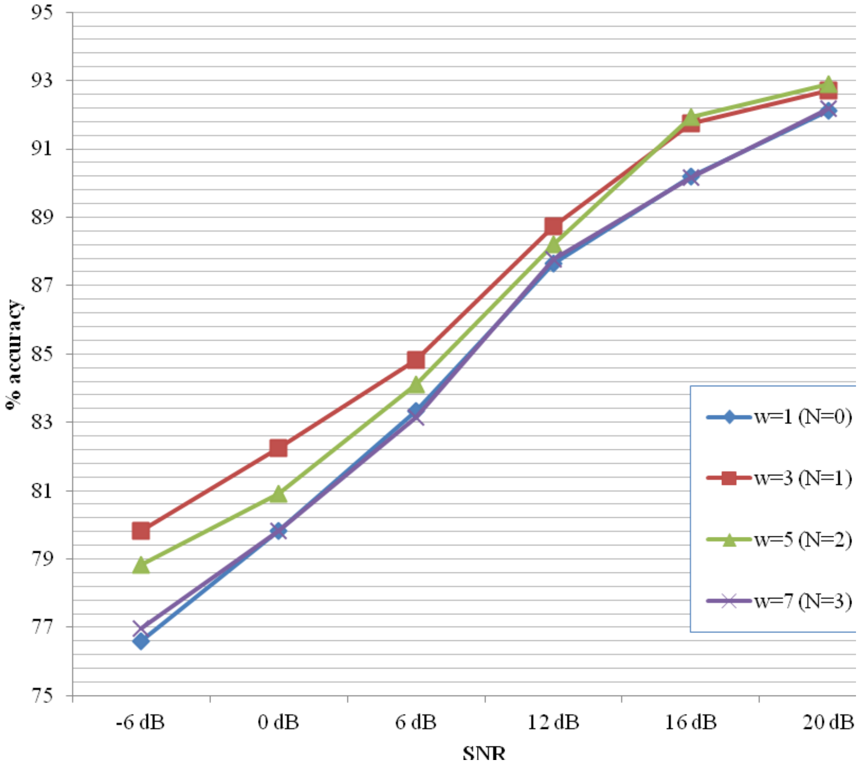

Fig. 3: Bird species classification accuracy (in percentages) for the Bagging (J48) algorithm at various signal-to-noise ratios and for different length of the temporal context information window

performance than the boosting algorithm (76.6% accuracy for -6 dB SNR). The meta-classifiers are followed in average performance by the rest of the evaluated classifiers, i.e. the k-nearest neighbor algorithm (*IBk*), the C4.5 pruned decision tree (*J48*), the Bayesian network (*BayesNet*) and the support vector machines (*SMO*).

A drop in the classification accuracy was found with the decrease of the signal-to-noise ratio for all the classifiers that were evaluated here, which is in agreement with the experimental results found in [17] for waterfall noise. It is worth mentioning that the k- nearest neighbor algorithm (*IBk*) and the decision tree algorithm (*J48*) achieved relatively high performance at noise-free conditions, i.e. for signal-to-noise ratio equal to 20 dB, approximately 3% less than the best performing meta-classifier. However, in noisy environments, such as SNR equal to 0 dB and -6 dB, they followed the best performing meta-classifier by approximately 5%. This is an indication of the advantage that the bagging and boosting algorithms can offer in real-field environments, where the presence of non-stationary interfering noises is frequent. Besides, low signal-to-noise ratio conditions are met in audio acquisition of bird vocalizations when the vocalizing bird is not close to the monitoring station installed in the field, but still captured by the microphone.

After evaluating the baseline bird species recognition performance for several classification algorithms, the overall recognition performance after post-processing the labeled audio frames using the temporal context information was examined. Specifically, the effect of the temporal context information block was applied for the best performing in noisy conditions *Bagging(J48)* algorithm and at various signal-to-noise ratios as shown in Figure 3. The best performing configuration setup of the temporal context information block for each evaluated signal-to-noise ratios (SNRs) is indicated in bold.

As can be seen in Figure 3, the effect of the temporal context information post-processing is significant for all signal-to-noise ratios and even more in the case of noisy environment, i.e. for low signal-to-noise ratios. In detail, the temporal context window length equal to three offers the best or close to the best performance across all the evaluated signal-to-noise ratios. The application of this window length (w=3) achieved approximately 3.5% absolute improvement of the bird species classification accuracy at -6 dB of signal-to-noise ratio. The improvement of the classification accuracy for signal-to-noise ratio equal to 20 dB is approximately 1% in terms of absolute recognition accuracy. The evaluated performance indicates the importance of the contextual information post-processing in the noisy real-field environment, since isolated erroneous labeling of audio frames coming from momentary bursts of interferences can be eliminated by the contextual audio frame labeling.

V. Conclusion

The results presented in Section 4, show that the integration of temporal contextual information into the decision-making process of the acoustic bird recognizer contributes to the improvement of the overall recognition accuracy, as well as supports our assumption about the importance of temporal contextual information. As the experimental results show, the integration of temporal contextual information as a post-processor in the acoustic bird recognition contributes mainly to the improvement of the recognition accuracy in low SNR conditions. This observation can be explained with the temporal smoothing effect, which the proposed post-processing scheme implements, and the resultant elimination of sporadic mislabeling of audio frames due to momentary bursts of noise which is typical in real-field conditions.

The effectiveness of the proposed post-processing scheme and its low computational and memory demands render it appropriate for mobile acoustic bird recognition applications and in other acoustic bird recognition applications with restricted energy resources.

Acknowledgments

The authors would like to thank the anonymous reviewers for their careful reading of this paper and for their helpful comments. This work was supported by the AmiBio project (LIFE08 NAT/GR/000539), which is implemented with the contribution of the LIFE+ financial instrument of the European Union. Project web-site: www.amibio-project.eu.

References

The authors of the 2009 article "Bird population density estimated from acoustic signals" (D.K. Dawson and M.G. Efford) published in the Journal of Applied Ecology, volume 46, pages 1201–1209. "Template-based automatic recognition of birdsong syllables from continuous recordings" was published in the Journal of the Acoustical Society of America in 1996 and was co-authored by S.E. Anderson,A.S.Dave,andD.Margoliash.
The paper "Application of dynamic programming matching to classification of budgerigar contact calls" was published in the Journal of the Acoustical Society of America in December 1996 and was writtenby.Ito,K.Mori,andS.Iwasaki
"A comparative study on automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models" was published in the April 1998 issue of the Journal of the Acoustical Society of America and was co-authored by J.A. Kogan and D. Margoliash.

"Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models" was published in April 2008 in the Journal of the Acoustical Society of America by V.M. Trifa, A.N.G. Kirschel, andC.E.Taylor.

The paper "Parametric representations of bird sounds for automatic species recognition" was published in 2006 by Somervuo, Harma, and Fagerlund in the IEEE Transactions on Audio, Speech, and Article published in March 2009 by Wildlife Acoustics, Inc. in Concord, Massachusetts titled "AutomaticallyIdentifyingAnimalSpeciesfromtheirVocalizations"(I.Agranat).

In 2005, at the Congress on Computational Intelligence Methods and Applications in Istanbul, Turkey, S.A. Selouani, M. Kardouchi, E. Hervet, and D. Roy presented an article titled "Automatic birdsong recognition basedonautoregressivetimedelayeuralnetworks" (pp. 1-6).

The article "Birdsong recognition using backpropagation and multivariate statistics" was published in 1997 in the IEEE Transactions on Signal Processing and was written by A. L. Mcllraith and H. C. Card.

"An automated acoustic system to monitor and classify birds" was published in 2006 in the EURASIP Journal on Applied Signal Processing by C. Kwan, K.C. Ho, G. Mei, and others. The article has 19 pages.

"Automatic identification of bird calls using spectral ensemble average voiceprints" was published in the proceedings of the 13th European Signal Processing Conference in September 2006 by Tyagi, Hegde,Murthy,andPrabhakar.

Volume 2007, Article ID 38637, 8 pages, doi:10.1155/2007/38637, S. Fagerlund, "Bird Species Recognition Using Support Vector Machines." EURASIP Journal on Advances in Signal Processing.

Data mining applied to acoustic bird species detection was presented by E. Vilches, I.A. Escobar, E.E. Vallejo, and C.E. Taylor in August 2006 in Hong Kong at the 18th International Conference on Pattern detection(Volume3,pages400–403).

In the Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 5, pages 545-548, held in Hong Kong in April 2003, A. Harma presented an algorithm for the automatic identification of bird species using sinusoidal modeling of syllables.

"Bird song recognition based on syllable pair histograms" was published in May 2004 in Montreal, Canada, by P. Somervuo and A. Harma in the Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 5, pages 825–828.

The authors of the article "Automatic Recognition of Bird Songs Using Cepstral Coefficients" (publication date: May 2006) are C.H. Lee, Y.K. Lee, and R.Z. Huang.

In the 2011 issue of the EURASIP Journal on Advances in Signal Processing, Jancovic and Kokuer published an article titled "Automatic Detection and Recognition of Tonal Bird Sounds in Noisy Environments." The article consists of 10 pages and has the DOI: 10.1155/2011/982936.

A book titled "The HTK book" was published by the Engineering Department at Cambridge University and authored by S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, Machine Learning, by T.M. Mitchell, published by McGraw-Hill International Editions in 1997.

Hiring S.S. Keerthi"Improvements to Platt's SMO algorithm for SVM classifier design," published in Neural Computation, volume 13, issue 3, pages 637-649 in 2001, by S.S., S.K. Shevade, C. Bhattacharyya,andK.R.K.Murthy.

C4.5: Machine Learning Programs by R. Quinlan, San Mateo, California: Morgan Kaufmann Publishers(1993).

Myers, H.I., and Frank, E. Data mining comprises of practical methods and tools for machine learning.

Yoav Freund and Robert E. Schapire's work on experiments using a novel boosting method is published by Morgan Kaufmann. Paper presented at the 1996 Thirteenth International Conference on Machine Learning, San Francisco, pages148–156.

Breiman, Leo (1996). Baggage predictors... Journal of Machine Learning, 24(2), 123–140.