Machine Learning (2018) 107:1651–1671
https://doi.org/10.1007/s10994-018-5739-8

# A comparison of hierarchical multi-output recognition approaches for anuran classification

**Juan G. Colonna[1,4] · João Gama[2] · Eduardo F. Nakamura[1,3]**

Received: 22 March 2017 / Accepted: 25 June 2018 / Published online: 5 July 2018


**Abstract**
In bioacoustic recognition approaches, a "flat" classifier is usually trained to recognize several species of anurans, where the number of classes is equal to the number of species. Consequently, the complexity of the classification function increases proportionally with the number of species. To avoid this issue, we propose a "hierarchical" approach that decomposes the problem into three taxonomic levels: the family, the genus, and the species. To accomplish this, we transform the original single-labelled problem into a multi-output problem (multi-label and multi-class) considering the biological taxonomy of the species. We then develop a top-down method using a set of classifiers organized as a hierarchical tree. We test and compare two hierarchical methods, using (1) one classifier per parent node and (2) one classifier per level, against a flat approach. Thus, we conclude that it is possible to predict the same set of species as a flat classifier, and additionally obtain new information about the samples and their taxonomic relationship. This helps us to better understand the problem and achieve additional conclusions by the inspection of the confusion matrices at the three classification levels. In addition, we propose a soft decision rule based on the joint probabilities of hierarchy pathways. With this we are able to identify and reject confusing cases. We carry out our experiments using cross-validation performed by individuals. This form of CV avoids mixing syllables that belong to the same specimens in the testing and training sets, preventing an overestimate of the accuracy and generalizing the predictive capabilities of the system. We tested our methods in a dataset with sixty individual frogs, from ten different species, eight genera, and four families, achieving a final Macro-Fscore of 80 and 70% with and without applying the rejection rule, respectively.

**Keywords** Hierarchical multi-label classification · Multi-output classification · LCPN · LCPL · Anurans taxonomy · Anuran calls recognition

Editors: Toon Calders and Michelangelo Ceci..

✉ Juan G. Colonna
juancolonna@icomp.ufam.edu.br; juan.gc@samsung.com

Extended author information available on the last page of the article

## 1 Introduction

Among all animal species, amphibians are the most sensitive to environmental changes (Carey and Alexander 2003; Cole et al. 2014). This observation has motivated many researchers to monitor the decline of amphibian populations through time and use it as an indicator of environmental problems (Adams et al. 2013; Houlahan et al. 2000). Among all amphibian species that may be monitored, anuran (frogs and toads) are preferred, because they have a semi-permeable skin which makes them sensitive to aquatic and terrestrial conditions (King 1969).

Nowadays, the most widely used method to monitor frog populations takes advantage of the vocalization capability to apply acoustics surveys (Marques et al. 2013; Silva 2010). The manual application of these surveys requires many human and economic resources, as well as expert knowledge, being difficult to apply in remote tropical areas such as the Amazon rainforest. Therefore, our goal is to develop an Automatic Calls Recognition (ACR) system to automatically monitor frog populations in a less invasive manner using acoustic sensors. The general idea consists in treating the challenge of anuran monitoring as a species recognition task using their calls combining Signal Processing and Machine Learning (ML) techniques (Colonna et al. 2015, 2012; Huang et al. 2009; Xie et al. 2015).

In bioacoustics, most of the related works deal with the species recognition problem using "flats" classifiers, where each instance belongs to one class (or "species" in this case), omitting any kind of hierarchical structure of the problem, e.g.: the taxonomy of the species (Colonna et al. 2016a, b; Han et al. 2011; Jaafar et al. 2014; Ribas et al. 2012; Vaca-Castaño and Rodriguez 2010; Xie et al. 2015).

The phylogenetic taxonomy aims to organize animal species into hierarchical categories (Fig. 1a). Then, using this pre-defined organization for anurans, we can build a hierarchical classification system combining classifiers at different levels. In this way we leave implicit the supposition that species close within the taxonomic tree may or may not produce similar sounds. Therefore, in this work, we address the problem of anuran species recognition through their calls using two hierarchical approaches considering the family and genus taxonomy as additional labels. Thus, our hypothesis is that using the phylogenetic taxonomy it is possible: (1) to better discriminate the classes by reduction of the feature space or by aggregating the decision functions of each classification level, (2) it is possible to reduce the complexity of the models involved in the classification, and (3) we can apply a decision rule based on the joint probabilities of the three levels that allows to eliminate cases with low degree of confidence.

To achieve our purpose, the family and genus information of the species were aggregated as new labels to each instance of our dataset, transforming the original multi-class dataset, with a single column of labels, into a multi-output dataset, i.e., a problem where the output space is multi-class and multi-label at the same time.[1] The addition of these two extra labels implies in an augmentation of the output space as shown in Fig. 1b.

The two hierarchical approaches compared here are (1) One Classifier per Parent Node and (2) One Classifier per Level, also known as Local Classifier per Node (LCPN) and Local Classifier per Level (LCPL), respectively (Ceci and Malerba 2007; Silla and Freitas 2011). Here, we refer to these approaches by the acronyms LCPN and LCPL to be consistent with the related literature. Then, we compare these approaches between them and against a traditional flat classifier, where we can consider the flat classifier and the LCPL model as baseline approaches.

[1] Our dataset is public available in the UCI repository through the link: https://archive.ics.uci.edu/ml/datasets/Anuran+Calls+(MFCCs).

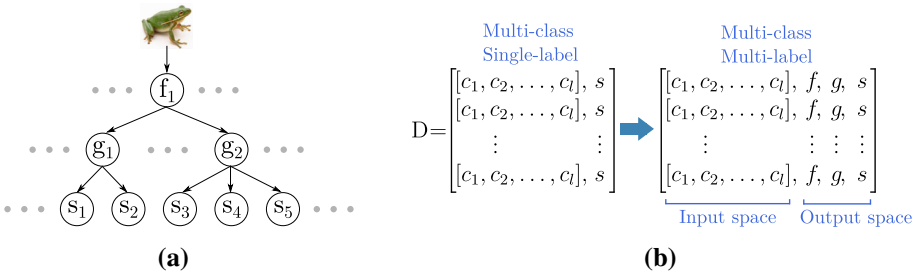


**Fig. 1** **a** Abstract example of phylogenetic taxonomy. **b** The original dataset (left) with species names as labels $s$. Output space augmentation (right) aggregating the family $f$ and genus $g$ labels obeying the phylogenetic taxonomy

Using the LCPN, we gain in simplicity being possible to decompose the feature space of the original problem into sub-problems with a smaller number of classes. In addition, we present a probabilistic decision rule that takes advantage of the constrain imposed in the LCPN model. With the second method, the LCPL, we can treat it as an ensemble of classifiers with a particular hierarchical organization taking advantage of the classifier diversity. Additionally, in both strategies, we can inspect the confusion matrix at each hierarchical level to obtain additional information about the relationship between the samples and their classes.

In summary, we present a comparative study of such hierarchical approaches for our bioacoustics application context. We would like to emphasize that our work is the first one regarding the combination and comparison of hierarchical approaches together with a CV procedure by individuals using the phylogenetic taxonomy for anuran calls recognition.

The fundamental concepts about hierarchical approaches are explained in Sect. 3.3. In addition, in Sect. 4, we discuss how hierarchical models were applied to the anuran recognition task (particularly the prediction of frog and toad species) and, in general, to the bioacoustics problems in which a hierarchical relationship between the labels could be modeled. In Sect. 5 we give an intuitive and detailed explanation about how LCPN can reduce the complexity of the model, and how the LCPL can be used. In order to test our proposal, we performed several experiments using the dataset shown in Sect. 6. Also, in all of our tests, we apply Cross-Validation (CV) by individuals (or specimens) as recommended by Colonna et al. (2016a) and explained in Sect. 6.3. The results and conclusion are supported by the calculation of Macro-metrics by level (see Sects. 6.4, 7 and 8).

## 2 Motivation for using a hierarchical approach

Anura is the name of an order in the Amphibian class of animals that includes frogs and toads. According to recent reports, there are more than 6600 different anuran species in the world, classified into 56 families and several genera (Frost 2016). The anuran diversity in the tropical areas of South America is the greatest, concentrating approximately 70% of the total global biodiversity of amphibians (IUCN 2016). Thus, in order to develop a flat classifier, we have to train it with the number of classes equal to the number of species that we intend to recognize. Therefore, the complexity of the decision function increases with the number of species, as well as the probability of misclassifications.

As mentioned early, a hierarchical approach using LCPN can alleviate the complexity of this problem by decomposing the feature space in levels. Thus, we use the well-known

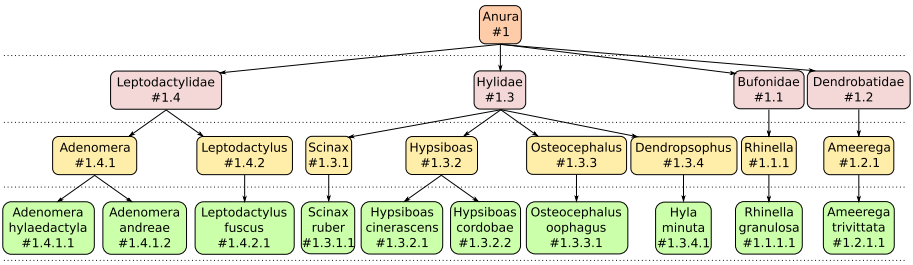


**Fig. 2** Species tree. From Top-to-Down levels: order, family, genus and species. The **#** stands for node ID

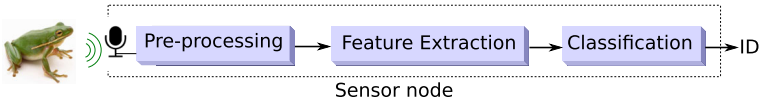


**Fig. 3** Automatic call recognition system (ACR)

taxonomy[2] to construct a tree with three levels: family, genus, and species (Fig. 2). With this, every time we go down through the levels of the tree, the input and output spaces are reduced (see Fig. 7a).

In contrast, the LCPL approach does not reduce the feature space because the last level has to be trained with the same number of classes as a flat classifier. However, the LCPL allows us to treat the problem at different levels of granularity and combine the decision in a similar way as an ensemble learning (see Fig. 7b). Another advantage compared to the LCPN lies in the cases in which the branches of the tree have only one leaf node, i.e., parent nodes that have not division in its lower levels, such as the *Bofunidae* family depicted in Fig. 2. In such cases, the LCPL gives us a posterior probability for those leaf nodes. Nonetheless, the applications of this approach have been limited in the related works and most of the time it is used as a baseline method (Silla and Freitas 2011; Silla and Kaestner 2013). We therefore decided to include it for comparison purposes.

Regardless the method (LCPN or LCPL), the paths between the highest and lowest levels of nodes are fixed and predefined by the taxonomy (see Fig. 2). These configurations allow us to investigate the error rates between families or genera to get some insights about the acoustical proximity of the samples, or to help us to develop more accurate monitoring methods in areas populated by different frog's families or genera independent of the species.

## 3 Fundamentals

In order to understand the methodology adopted in this work, two concepts are described in this section: how a bioacoustic recognition framework works, and how to create a hierarchical classification approach.

### 3.1 Bioacoustics systems

Anuran call classification systems are traditionally composed of three main steps with different purposes (see Fig. 3) (Colonna et al. 2015; Huang et al. 2009; Xie et al. 2016). Formally, the

[2] This taxonomy is maintained by the International Commission on Zoological Nomenclature (ICZN) and can be consulted in: https://amphibiaweb.org/taxonomy/index.html.

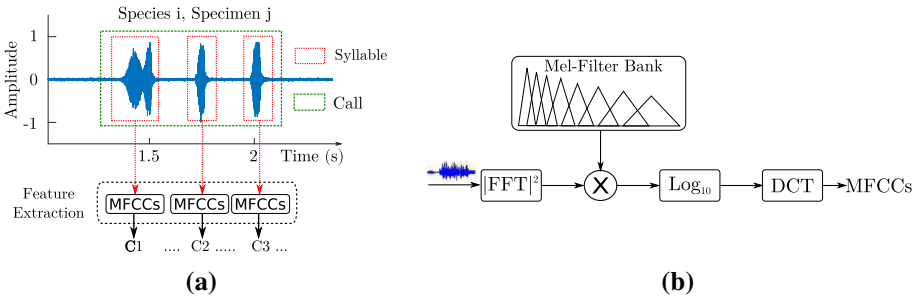


**Fig. 4** A framework for automatic frog's calls recognition. **a** An audio record with three syllables of the species *Hyla minuta*. **b** MFCCs steps. Here, FFT stands for fast Fourier transform and DCT for discrete cosine transform

input bioacoustic signal $X = \{x_1, x_2, \ldots, x_N\}$ is a time series of length $N$, in which its values represent the acoustics pressure levels (or amplitude). A syllable $\mathbf{x}_k = \{x_t, x_{t+1}, \ldots, x_{t+n}\}$ is a subset of $n$ consecutive signal values (see Fig. 4a). Thus, the pre-processing step segments the signal $X$ by identifying the beginning and the endpoints of $\mathbf{x}_k$ (Colonna et al. 2015).

After syllable extraction, we need to represent each $\mathbf{x}_k$ by a set of features, commonly called Low Level acoustic Descriptors (LLDs). The most frequent LLDs in this application context are the Mel-Frequency Spectral Coefficients (MFCCs). The MFCCs perform a spectral analysis based on a triangular filter-bank logarithmically spaced in the frequency domain (see Fig. 4b) (Colonna et al. 2012; Rabiner and Schafer 2007). Feature extraction using the MFCCs allows one to represent and reduce the dimensionality of any syllable to a finite set of coefficients ($\text{MFCC}(\mathbf{x}_k) \rightarrow \mathbf{c}_k$), i.e., $X \rightarrow \{(\mathbf{c}_1, s_i), (\mathbf{c}_2, s_i), \ldots, (\mathbf{c}_k, s_i)\}$, where each $\mathbf{c}_k = [c_1, c_2, \ldots, c_l]$ is a feature vector with $l$ coefficients, and $s_i$ is the species name (or label). The representation of $\mathbf{x}_k$ through $\mathbf{c}_k$ is more robust, more compact, and easier to recognize compared to the use of raw data.

Finally, the challenge is how to assign species names to a new syllable by using the MFCCs values. This is a supervised classification task and it is performed by the last step of the system. For this purpose, several ML algorithms could be applied to creating and training a model $f(\cdot)$ with capabilities to predict new incoming samples, i.e., given an unknown $\mathbf{c}$ estimates the most probable label by evaluating $f(\mathbf{c}) \rightarrow s_i$, where $S = \{s_1, s_2, \ldots, s_i\}$ is the set of species.

### 3.2 Syllable segmentation

For the segmentation and syllable extraction tasks we use the Spectral entropy ($\text{H}_{\text{FFT}}$) of the signal in a batch mode setting, i.e., frame-by-frame. The Spectral Entropy of a frame was defined by Sueur et al. (2008) as:

$$\text{H}_{\text{FFT}} = -\sum_{f=0}^{n/2} \frac{\text{S}(f)\ \log(\text{S}(f))}{\log\,(n/2)}, \tag{1}$$

where $n$ is the frame length and $\text{S}(f)$ is the spectrum of $\mathbf{x}_k$ calculated with the Fast Fourier Transform ($\text{S}(f) = \mathcal{F}(\mathbf{x}_k)$). In this case, only the positive part of the spectrum is used. The $\text{S}(f)$ is normalized to obtain a probability mass function:

$$\text{S}(f) = \frac{|\text{S}(f)|}{\sum_{f=0}^{n/2} |\text{S}(f)|}, \tag{2}$$

which is used to compute the normalized Spectral Entropy.

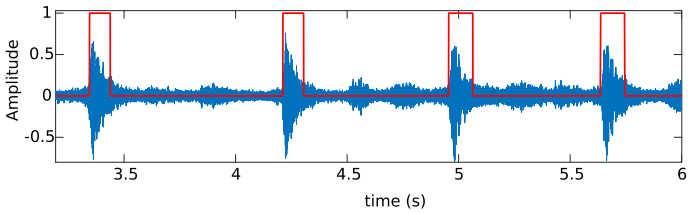


**Fig. 5** Example of syllable extraction using spectral entropy

After that, we applied the binary decision rule to detect the consecutively frames that compose a syllable:

$$\text{class}(\mathbf{x}_k) = \begin{cases} 1 & \text{if } \mathrm{H}(\mathbf{x}_k) \leq T_H \\ 0 & \text{if } \mathrm{H}(\mathbf{x}_k) > T_H \end{cases}, \tag{3}$$

in which $T_H$ is a threshold for the entropy value of each frame. With this rule, we assign the class "signal" to frames of low entropy. Since entropy can be interpreted as a measure of "impurity", the higher the value, the greater the probability that the underlying signal is a random noise.

The optimal $T_H$ value is a trade-off between the number of correctly recovered syllables and discarded noise segments. Our proposed procedure to find that threshold is summarized as follows:

1. assign $T_H^t \leftarrow \text{mean}(\{\mathrm{H}_{\mathrm{FFT}_{1:n}}\})$;
2. divide the $\{\mathrm{H}_{\mathrm{FFT}_{1:n}}\}$ set into one subset of components with entropy less than $T_H^t$ and a second subset with the remaining components;
3. obtain the mean values $\mu_1$ and $\mu_2$ of each subset;
4. update $T_H^{t+1} \leftarrow {}^{(\mu_1+\mu_2)}/_2$;
5. repeat steps 2–4 until convergence ($|T_H^{t+1} - T_H^t| < 0.01$).

This procedure, known as a binary clustering method (Sezgin and Sankur 2004), was evaluated in the context of bioacoustic signal segmentation by Colonna et al. (2018). By applying this, a different optimal $T_H$ value is found for each recording, where the input is a set of entropies $\{\mathrm{H}_{\mathrm{FFT}_1}, \mathrm{H}_{\mathrm{FFT}_2}, \ldots, \mathrm{H}_{\mathrm{FFT}_n}\}$ corresponding with the $n$ frames of the signal. Thus, this algorithm divides the entropy set into two groups of maximum separation between their means trying to maximize the inter-class distance. Different from other clustering techniques, such as $k$-means, this method attempts to balance the PDFs of the resulting separation to avoid creating thin clusters while reducing the squared error in each cluster (Magid et al. 1990).

The chosen frame size was 0.0464 s with 66% of overlap to obtain a good segmentation-time resolution. Figure 5 depicts a syllable segmentation of the *Adenomera andreae* species. Note that applying this procedure to our database of recordings each signal sample that the classifier will classify corresponds to a single species excluding the possibility of a mixture of signals.

After all the syllables were extracted, the mapping of these to the set of features is carried out according to Fig. 4b. This figure depicts the calculation of the Mel-frequency cepstral coefficients (MFCCs) applying the following equation (Rabiner and Schafer 2007):

$$\mathrm{MFCC}_m = \frac{1}{R}\sum_{r=1}^{R} \log(M_r) \cos\left(\frac{2\pi}{R}\left(r + \frac{1}{2}\right)m\right), \tag{4}$$

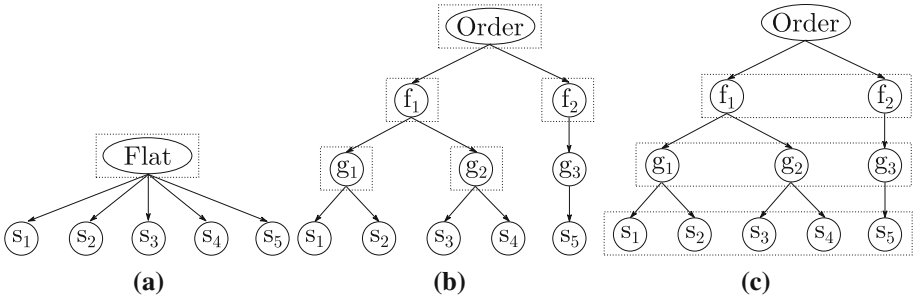


**Fig. 6** Different manners to create a hierarchical classifier combining flat classifiers. From top-to-down levels: **f** stands for family, **g** for genus, and **s** for species. **a** and **b** are examples of LCPN and LCPL approaches, respectively. **a** Flat classifier. **b** One classifier per parent node. **c** One classifier per level

where $M_r$ are the values produced by each triangular filter over the spectrum generated by the Fast Fourier Transform, $m$ is the number of coefficients and $R$ is the amount of filters. With these coefficients each syllable is represented in a unique way through a vector. Subsequently, the dataset is completed by assigning to each vector its respective label, as explained in the previous section.

### 3.3 Review of hierarchical classification approaches

Hierarchical methods are often used to solve multi-label problems in which the classes have an inherent taxonomy structure, i.e., an instance that belongs to a subclass, naturally belongs to its higher level classes (Ceci and Malerba 2007; Silla and Freitas 2011). Such methods help to simplify complex multi-class problems by transforming these into a multi-label approach by considering the hierarchical relationship between the labels. For instance, with the LCPN configuration, every time we go down to a level of the hierarchy, the number of possible solutions is reduced, simplifying the decision function, as shown in Fig. 7a. There are two common models to describe the hierarchical relationships between the classes: (a) trees, and (b) Direct Acyclic Graphs (DAG). A tree structure connects a set of leaves nodes to a single parent node forming several subtrees not interconnected on the same level. A DAG is a more flexible structure allowing the leafs to have more than one parent node (Freitas et al. 2007). In our approach, we adopted a tree structure, due to the taxonomic constraints of our problem, in which every species can belong to just one genus and one family classes at the same time.

Figure 6 illustrates the differences between a flat classifier and two commonly used hierarchical approaches. In Fig. 6b, c a set of flat classifiers is employed to construct two multi-label hierarchical trees (Silla and Freitas 2011). These trees may be imbalanced depending on the taxonomic structure of the problem. The classifiers inside each node should be trained separately and assembled after that. During the prediction phase, the strategy adopted to determine the class of a new sample is top-down. This strategy starts from the top nodes performing the corresponding predictions and goes down until it reaches a leaf node in the last hierarchical level. Thus, the decision results in a unique relationship among the set of predicted labels. An obvious disadvantage associated with this top-down approach is the error propagation coming from the higher levels of the tree. However, these hierarchical approaches are well suited for the context of species recognition where the class labels have an inherent taxonomy.

## 4 Related works

Hierarchical approaches have been shown to be useful in several application contexts when compared to flat classifiers, for example in protein classification (Zimek et al. 2010), on text categorization (Ceci and Malerba 2007) or recognizing user emotions on social networks (Angiani et al. 2016), and especially in problems with large-scale taxonomic structures (Babbar et al. 2013).

Into the bioacoustic context, several authors have already studied the problem of recognition and classification of anuran species through their calls. Among these, Huang et al. (2009) and Colonna et al. (2012) studied the best acoustic features to recognize different species. Jaafar and Ramli (2013) and Colonna et al. (2015) focused on comparing some syllable extraction procedures as a pre-processing step. Finally, Ribas et al. (2012) and Colonna et al. (2014) evaluated the possibility of embedding a classifier into the nodes of a wireless sensor network (WSN). However, little effort was made to link the hierarchical taxonomy of the species with an automatic classification system. The hierarchical taxonomic organization of the species is a standard approach in ecology since it was proposed in 1735 by Carl Linnaeus.

Gingras and Fitch (2013) formulated the hypothesis that anuran species, which are phylogenetically or taxonomically close, have more similar calls. To test this hypothesis the authors developed a three-parameter model using the mean values of dominant frequency, the variation coefficient of root-mean square energy, and spectral flux of the signals. Calls from 142 species belonging to four genera were analyzed and classified applying a logistic regression model, a Support Vector Machine (SVM), a k-Nearest Neighbors (kNN), and a Gaussian mixture model (GMM) achieving an accuracy of approximately 70%. During the test, different specimens (or individuals) were used for training and testing in order to prove the generalization capabilities of the model.

An acoustic feature extraction and a comparative analysis of these features, for developing a hierarchical classification technique of Australian frog calls, was proposed by Xie et al. (2015). This work studies which acoustics features should be used in each classification level, considering the taxonomy information separated into three levels: family, genus, and species. The contribution was a correlation method, able to select the better features for each level, but the final classification was addressed as three separate problems using SVM. The levels were not integrated into one single approach leaving two open questions: (1) how to integrate these classifiers in one single method capable of reducing the complexity by taking advantage of the hierarchical taxonomy, and (2) how to handle the disagreement between the levels.

The technique called Balance-Guaranteed Optimized Tree with Reject option (BGOTR) is a hierarchical classification system which includes the reject option. This was developed by Huang (2016) for fish image recognition using underwater cameras. In this system, a multiclass classifier and a feature selection are built together into a hierarchical tree, and this is optimized to maximize the classification accuracy by grouping the classes based on their inter-class similarities. The rejection option is performed after the hierarchical classification by applying a Gaussian mixture model (GMM) to fit the distribution of the features in the images. Despite the interesting results, the authors highlight that this approach does not consider the taxonomy of the problem. Indeed, this method was not developed for a multilabel purpose, and therefore it is not possible to evaluate the similarities between family, genus, and species.

An evaluation of different hierarchical approaches applied to the bird species recognition was performed by Silla and Kaestner (2013). The authors compared three different approaches: a flat classification where the class hierarchy is disregarded, one classifier per parent node (see Fig. 6), and one global approach where a single algorithm is used to predict classes at any level of the hierarchy based on Global-Model Hierarchical Classification Naive Bayes (GMNB). Moreover, an extension of the Precision, Recall, and F-measure metrics was introduced, tailored to the hierarchical classification scenario. The results show that the hierarchical approaches outperform flat classifiers when the number of species is large and that the labels can be organized in an adequate hierarchy.

Finally, Colonna et al. (2016b) proposed and evaluated an LCPN approach to recognizing anuran species. The results were promising, however, a baseline comparison against a flat classifier or another hierarchical approach was not presented. They applied Cross Validation on individuals (or specimens) to test the model generalization capabilities as recommended in the related works (Colonna et al. 2016a). To the best of our knowledge, the work presented here shows such baseline comparisons, integrating the family, the genus, and the species labels of anurans to be solved as a multi-output problem.

## 5 Proposed approach

The phylogenetic taxonomy aims to organize animals into hierarchical categories. Using this pre-defined organization for anurans, we can build our hierarchical classification system adding two extra labels to the original dataset ($g$ and $f$):

$$\text{Dataset} = \begin{bmatrix} \mathbf{c}_1 = [c_1, c_2, \ldots, c_l], & s, & g, & f \\ \mathbf{c}_2 = [c_1, c_2, \ldots, c_l], & s, & g, & f \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{c}_k = [c_1, c_2, \ldots, c_l], & s_i, & g_j, & f_m \end{bmatrix}$$

with these new labels, we have turned our multiclass problem with a single label into a multi-label and multi-class problem (MM). This MM is a generalization of the common multi-label problems, where the classes are binary in each column. These MM problems are also called Multi-output because the output is composed of a tuple of labels (Borchani et al. 2016), which are three in our case $[s, g, f]$.

This is possible because there is an unequivocal relationship between the species names and its genus and family names. That is, a subset $S^0 = \{s_1, \ldots s_p\}$ of species belongs to a singular genus ($S \subseteq g_m$), while a subset of genus $G^0 = \{g_1, \ldots, g_m, \ldots, g_p\}$ also belongs to a particular family ($G \subseteq f_m$) such that $f_m \subseteq F^0$. Therefore, any $s_i$ is from $G^0$ and $F^0$ without ambiguity. Thus, if a flat classifier correctly predicts a particular species, the system is effectively predicting not only the species at the last level, but also the genus and the family classes at the first two levels together.

With this concept, we can apply reverse engineering and develop a hierarchical top-down approach as depicted in Fig. 6b, c. Our hierarchical tree is represented in Fig. 2 of Sect. 2. An example of problem simplification using LCPN decomposition with two attributes is shown in Fig. 7a. As we can note, at the beginning, all the samples belong to two families (or classes). After the family classification, the problem is reduced and consequently simplified by the simple decomposition of the feature space, in which only the samples of the first family remain. This process is repeated until the last classification level is reached (the species label). Thus, the class of a leaf node is used to estimate the label of new samples.

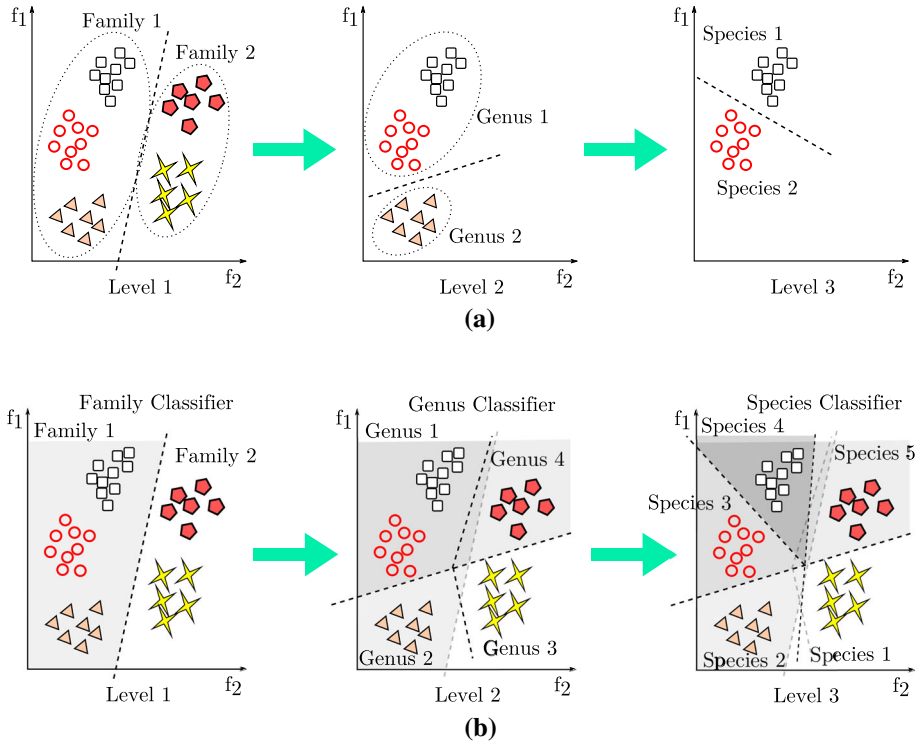


**Fig. 7** Comparison between LCPN and LCPL approaches from the feature space perspective. **a** Problem decomposition stages when performing the hierarchical LCPN from top to down described by an example of prune training data. **b** Overlap of decision functions by levels of LCPL. The final decision corresponds to the superposition of three individual decision

An example of decision function ensemble using LCPL is depicted in Fig. 7b. The final decision then corresponds to the overlapped decision of the three levels. Notice that the boundaries of such functions are not necessarily equal at all three levels. As each level is trained with a different number of classes but with the same number of samples, we can consider this method as views of the same problem with distinct granularities.

A remarkable advantage of this approach is that we do not have to perform every classification for some branches in all levels. This is the main advantage of the customization based on the taxonomy and the reason why we chose the approaches described in Fig. 6b, c. For instance, if the first classifier assigns the *Bufonidae* label to a new sample at the top level, it is not necessary to continue classifying the remaining levels, because there are no more splits for this branch. Therefore, the genus label *Rhinella* and the species label *Rhinella granulosa* are assigned automatically. The remaining settings of our approach are detailed in Sect. 6.

## 6 Methodology description

In order to develop our hierarchical method, the first step is to obtain the family and genus labels for each sample of our dataset. Given this context, we used the taxonomy information available in Frost (2016). The dataset description, the classifier setting, the validation procedure and the metrics used are described in the following subsections.

**Table 1** Species dataset

| Family | Genus | Species | *s* | *k* |
|---|---|---|---|---|
| *Leptodactylidae* | *Leptodactylus* | *Leptodactylus fuscus** | 4 | 270 |
| | *Adenomera* | *Adenomera andreae** | 8 | 672 |
| | | *Adenomera hylaedactyla*** | 11 | 3478 |
| *Hylidae* | *Dendropsophus* | *Hyla minuta*** | 11 | 310 |
| | *Scinax* | *Scinax ruber*** | 5 | 148 |
| | *Osteocephalus* | *Osteocephalus oophagus** | 3 | 114 |
| | *Hypsiboas* | *Hypsiboas cinerascens** | 4 | 472 |
| | | *Hypsiboas cordobae**** | 4 | 1121 |
| *Bufonidae* | *Rhinella* | *Rhinella granulosa** | 5 | 68 |
| *Dendrobatidae* | *Ameerega* | *Ameerega trivittata*** | 5 | 542 |

The *s* and the *k* stands for the number of specimens and the number of syllables respectively
*Amazonas, **Mata Atlântica, ***Córdoba

### 6.1 Dataset description

The dataset used in our experiments is summarized in Table 1. It has 10 different species, 60 specimens and 7195 syllables. These records were collected *in situ* under real noise conditions. Some species are from the Federal University of Amazonas, others from Mata Atlântica, Brazil and the last one from Córdoba, Argentina. These recordings were stored in *wav* format with 44.1 kHz of sampling frequency and 32 bit, which allows us to analyze signals up to 22.05 kHz. From each extracted syllable, 22 MFCCs were calculated by using 44 triangular filters and these coefficients were normalized between $-1 \leq c_l \leq 1$ (see Sect. 3.1).

### 6.2 Node classifier description

In our experiments, we chose four classifiers for comparison: a kNN with 3 neighbors, a Support Vector Machine with an RBF kernel (RBF-SVM), a polynomial kernel (Poly-SVM) of degree three, and a Decision Tree. The parameters of RBF-SVM kernel were automatically setup using the default implementation of the Matlab Software which uses cross-validation. The same is valid for the Decision Tree. Thus, these classifiers were used to fill the nodes of the hierarchical approach, as depicted in Fig. 6.

Besides that, in all parent nodes we decomposed the multiclass model $f(\cdot)$ into a combination of smaller binary models $f'(\cdot)$ applying the One-against-All procedure (Fürnkranz 2001). After that, the result of each binary model was combined by using the majority voting rule through the Error Correction Output Code procedure (ECOC). This decomposition technique reduces the complexity of each sub-problem compared to the multiclass approach and does not increase the computational load as much as the One-against-One method (Colonna et al. 2016a).

During the training phase of the internal nodes of our hierarchy the dataset is splitted according to the structure shown in Fig. 2. For instance, the classifier corresponding to node # 1.4.1.1 in the hierarchy is trained only with samples of the *Adenomera hylaedactyla* species, whereas its top level node # 1.4.1 is trained with samples of both *Adenomera hylaedactyla* and *Adenomera andreae* species since these two species belong to the same genera. This split procedure is repeated in most hierarchy pathways except for those branches having only one path.

### 6.3 Special type of cross-validation

Because we are dealing with a supervised problem, and we want to consider the generalization capabilities of the system, we need to apply a cross-validation (CV) procedure to estimate the expected error in a real situation (Friedman et al. 2001). With $k$-CV the original dataset is split into $k$ disjoint folds, and for each one, the conditional error ($e_k$) is estimated by training the model $f(\cdot)$ with $k$-1 folds. Thus, this procedure is repeated $k$ times and the expected generalized error can be obtained by averaging $e_k$. When the information of the individuals (or specimens) is omitted, we may fall into a situation in which the split could leave syllables of the same individuals in the testing and training sets causing an overestimate on the accuracy (Colonna et al. 2016a). To overcome this problem, we consider the specimen information during the $k$-CV fold splitting, i.e., we leave all the syllables that belong to the same specimen together, avoiding mixing them in the testing and training sets. In other words, we perform a Leave-one-Out Cross Validation by individuals. To accomplish this, we introduce an extra label with the record ID that will only be considered during the $k$-CV split. Thus, we assume that the generalization error will be more realistic because we are training with one specimen to predict a different one.

### 6.4 Performance measures

Diverse anuran species have different syllable rates in their calls (number of syllables per unit time). This is a particular vocalization characteristic of each anuran species. Hence, an unequal number of samples (or syllables) are retrieved from each record producing an unbalanced dataset (see Colonna et al. 2015). This problem affects the traditional accuracy measures knowing as the Micro-metrics. For instance, a classification model that always predicts the species with the higher number of samples might have a high accuracy, even in the extreme case of losing all syllables from the other classes. To overcome this matter, we suggest to use the Macro-Precision (M-Prec or Average-Precision), the Macro-Recall (M-Rec or Average-Recall) and Macro-Fscore (M-F1) (Colonna et al. 2015; Sokolova and Lapalme 2009).

## 7 Experiments and results

In this section, we present the species recognition results comparing the three approaches: a Flat classifier, an LCPN, and an LCPL. Also, we show a comparison between these approaches using four different classifiers for the nodes, described in Sect. 6.2.

### 7.1 Baseline flat classifier

With the purpose of having a baseline comparison, we test a flat Classifier. The results, divided by species, are shown in Table 2. The last rows of this table present the Macro-metrics (M-Prec, M-Rec, and M-F1). The best results are highlighted in boldface. The statistical $t$-test with 95% of confidence was applied in order to detect a tie between the distribution of results. Thus, the Precision columns are compared among themselves, and the same is valid for the Recall columns. The same considerations apply to Tables 7 and 8 in Section 7.2.

Among all the classifiers we note that the best precision was achieved by RBF-SVM, but compared to kNN and Poly-SVM the differences were not statistically significant. With

**Table 2** Flat classifier

| Species | kNN | | RBF-SVM | | Poly-SVM | | Tree | |
|---|---|---|---|---|---|---|---|---|
| | Prec. | Rec. | Prec. | Rec. | Prec. | Rec. | Prec. | Rec. |
| Adenomera a. | 0.74 | 0.37 | 0.82 | 0.24 | 0.61 | 0.35 | 0.40 | 0.36 |
| Adenomera h. | 0.99 | 0.98 | 0.99 | 0.96 | 0.99 | 0.99 | 0.97 | 0.94 |
| Ameerega t. | 0.85 | 0.61 | 0.95 | 0.34 | 0.74 | 0.57 | 0.62 | 0.40 |
| Hyla m. | 0.66 | 0.73 | 0.87 | 0.43 | 0.70 | 0.78 | 0.25 | 0.47 |
| Hypsiboas cin. | 0.86 | 0.85 | 0.67 | 0.59 | 0.72 | 0.89 | 0.81 | 0.72 |
| Hypsiboas cor. | 0.90 | 0.96 | 0.38 | 0.89 | 0.94 | 0.96 | 0.80 | 0.89 |
| Leptodactylus f. | 0.53 | 0.78 | 0.95 | 0.40 | 0.57 | 0.80 | 0.43 | 0.39 |
| Osteocephalus o. | 0.37 | 0.46 | 0.97 | 0.28 | 0.70 | 0.60 | 0.02 | 0.03 |
| Rhinella g. | 0.16 | 0.83 | 1.00 | 0.77 | 0.25 | 0.83 | 0.32 | 0.60 |
| Scinax r. | 0.84 | 0.61 | 0.91 | 0.52 | 0.91 | 0.84 | 0.32 | 0.19 |
| Average- | **0.69** | **0.72** | **0.85** | 0.54 | **0.71** | **0.76** | 0.49 | 0.50 |
| Macro-F1 | 0.70 | | 0.66 | | 0.74 | | 0.50 | |

**Table 3** Confusion matrix of species with a flat kNN classifier

| | a | b | c | d | e | f | g | h | i | j | $\text{Rec}_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 249 | 0 | 35 | 26 | 12 | 2 | 139 | 59 | 150 | 0 | 0.37 |
| b | 0 | 3436 | 0 | 37 | 0 | 4 | 0 | 0 | 0 | 1 | 0.98 |
| c | 39 | 0 | 333 | 14 | 0 | 29 | 0 | 0 | 126 | 1 | 0.61 |
| d | 34 | 14 | 19 | 229 | 0 | 0 | 0 | 0 | 1 | 13 | 0.73 |
| e | 7 | 0 | 0 | 0 | 405 | 29 | 20 | 11 | 0 | 0 | 0.85 |
| f | 1 | 6 | 0 | 2 | 13 | 1077 | 11 | 10 | 0 | 1 | 0.96 |
| g | 3 | 4 | 1 | 8 | 8 | 23 | 211 | 2 | 10 | 0 | 0.78 |
| h | 1 | 0 | 0 | 0 | 29 | 20 | 9 | 53 | 2 | 0 | 0.46 |
| i | 1 | 1 | 0 | 0 | 1 | 1 | 6 | 0 | 57 | 1 | 0.83 |
| j | 0 | 1 | 1 | 26 | 1 | 11 | 0 | 8 | 9 | 91 | 0.61 |
| $\text{Prec}_i$ | 0.74 | 0.99 | 0.85 | 0.66 | 0.86 | 0.90 | 0.53 | 0.37 | 0.16 | 0.84 | |

(a) *Adenomera andreae*, (b) *Adenomera hylaedactyla*, (c) *Ameerega trivittata*, (d) *Hyla minuta*, (e) *Hypsiboas cinerascens*, (f) *Hypsiboas cordobae*, (g) *Leptodactylus fuscus*, (h) *Osteocephalus oophagus*, (i) *Rhinella granulosa*, and (j) *Scinax ruber*. Prec and Rec stand for precision and recall respectively

RBF-SVM the species *Adenomera hylaedactyla* and *Rhinella granulosa* got the best results. Thus, if the goal is to monitor these two species, RBF-SVM is a good choice. Nonetheless, the best trade-off between precision and recall was obtained by Poly-SVM, with a Fscore equal to 0.74, showing a better generalization capability. Lastly, the decision tree approach performed the worst.

An example of a kNN classifier confusion matrix is present in Table 3. In this matrix, the rows represent the Ground Truth (GT) labels and the columns indicate the predicted labels. The main diagonal corresponds to the number of hits. The last column and row show the recall and precision by class, from which we can get the Macro-metrics by averaging these values. From this matrix, we noticed that the greatest number of confusions occur between the species *Adenomera andreae* and the species *Leptodactylus fuscus*, *Hypsiboas cordobae*

**Table 4** Confusion matrix of family level with kNN and LCPL

| | *Bufonidae* | *Dendrobatidae* | *Hylidae* | *Leptodactylidae* | $\text{Rec}_i$ |
|---|---|---|---|---|---|
| *Bufonidae* | 57 | 0 | 3 | 8 | 0.83 |
| *Dendrobatidae* | 126 | 333 | 44 | 39 | 0.61 |
| *Hylidae* | 12 | 21 | 2030 | 102 | 0.93 |
| *Leptodactylidae* | 158 | 38 | 182 | 4042 | 0.91 |
| $\text{Prec}_i$ | 0.16 | 0.84 | 0.89 | 0.96 | |

Prec and Rec stand for precision and recall respectively

**Table 5** Confusion matrix of genus level with kNN and LCPL

| | a | b | c | d | e | f | g | h | $\text{Rec}_i$ |
|---|---|---|---|---|---|---|---|---|---|
| a | 3685 | 37 | 64 | 19 | 138 | 57 | 148 | 2 | 0.88 |
| b | 39 | 333 | 13 | 29 | 0 | 1 | 126 | 1 | 0.61 |
| c | 48 | 20 | 228 | 0 | 0 | 0 | 1 | 13 | 0.73 |
| d | 17 | 0 | 2 | 1537 | 26 | 10 | 0 | 1 | 0.96 |
| e | 7 | 1 | 8 | 30 | 212 | 2 | 10 | 0 | 0.78 |
| f | 1 | 0 | 0 | 50 | 9 | 52 | 2 | 0 | 0.45 |
| g | 2 | 0 | 0 | 2 | 6 | 0 | 57 | 1 | 0.83 |
| h | 1 | 1 | 26 | 12 | 0 | 8 | 9 | 91 | 0.61 |
| $\text{Prec}_i$ | 0.96 | 0.84 | 0.66 | 0.91 | 0.54 | 0.40 | 0.16 | 0.83 | |

(a) *Adenomera*, (b) *Ameerega*, (c) *Dendropsophus*, (d) *Hypsiboas*, (e) *Leptodactylus*, (f) *Osteocephalus*, (g) *Rhinella*, and (h) *Scinax*. Prec and Rec stand for precision and recall respectively

and *Rhinella granulosa*. Since we use a flat approach, it is not possible discern at what level of taxonomy the confusions take place.

### 7.2 Hierarchical approach

The structure of our hierarchical approach was introduced in Fig. 2. The first parent node corresponds to the order (Anura) and is responsible for the classification of the samples into four family classes (column 1 in Table 1). In the second level (the family level) the parent nodes are trained with the genus labels that correspond to each particular family. Thus, the family branches are able to predict their own genus labels. The last prediction takes place at the genus level being responsible for predicting their own species names, as shown by their leaf nodes. With this configuration, we obtain a confusion matrix per level, i.e.: one matrix for the family labels (Table 4), one for the genus labels (Table 5), and one for the species labels (Table 6). These confusion matrices (Tables 4, 5 and 6) correspond to the hierarchical LCPL using kNN with three neighbors.

From these confusion matrices, it is possible to obtain the Macro-Precision, -Recall and -Fscore by level to conduct further analysis, being impossible with a flat classifier. For instance, we can note that the low precision of the *Bufonidae* family is due to the fact that the species *A. andreae* and *A. trivittata* were mostly confused with *R. granulosa*, but not the *A. hylaedactyla* species which belong to the same family as *A. andreae*. It suggests two important things, first, we can monitor with good accuracy a real scenario with the *R.*

**Table 6** Confusion matrix of species level with kNN and LCPL

| | a | b | c | d | e | f | g | h | i | j | $\text{Rec}_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| a | 249 | 0 | 37 | 27 | 13 | 2 | 138 | 57 | 148 | 1 | 0.37 |
| b | 0 | 3436 | 0 | 37 | 0 | 4 | 0 | 0 | 0 | 1 | 0.98 |
| c | 39 | 0 | 333 | 13 | 0 | 29 | 0 | 1 | 126 | 1 | 0.61 |
| d | 34 | 14 | 20 | 228 | 0 | 0 | 0 | 0 | 1 | 13 | 0.73 |
| e | 11 | 0 | 0 | 0 | 412 | 31 | 16 | 2 | 0 | 0 | 0.87 |
| f | 0 | 6 | 0 | 2 | 13 | 1081 | 10 | 8 | 0 | 1 | 0.96 |
| g | 3 | 4 | 1 | 8 | 8 | 22 | 212 | 2 | 10 | 0 | 0.78 |
| h | 1 | 0 | 0 | 0 | 30 | 20 | 9 | 52 | 2 | 0 | 0.45 |
| i | 1 | 1 | 0 | 0 | 1 | 1 | 6 | 0 | 57 | 1 | 0.83 |
| j | 0 | 1 | 1 | 26 | 1 | 11 | 0 | 8 | 9 | 91 | 0.61 |
| $\text{Prec}_i$ | 0.73 | 0.99 | 0.84 | 0.66 | 0.86 | 0.90 | 0.54 | 0.40 | 0.16 | 0.83 | |

(a) *Adenomera andreae*, (b) *Adenomera hylaedactyla*, (c) *Ameerega trivittata*, (d) *Hyla minuta*, (e) *Hypsiboas cinerascens*, (f) *Hypsiboas cordobae*, (g) *Leptodactylus fuscus*, (h) *Osteocephalus oophagus*, (i) *Rhinella granulosa*, and (j) *Scinax ruber*. Prec and Rec stand for precision and recall respectively

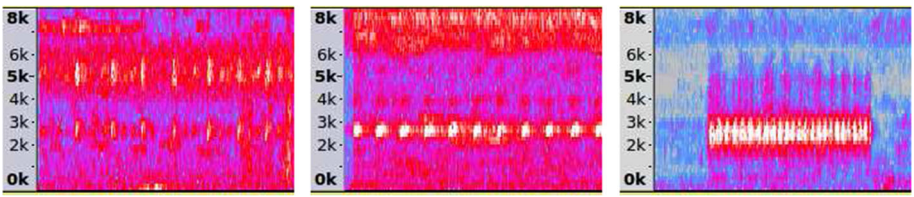


**Fig. 8** Spectrograms of three species. The white color symbolizes greater concentration of energy. Each figure depicts the main call frequencies, where we note that the frequency band centered around 2.5 kHz is present in the three species. The *A. andreae* species also presents a main frequency band of 5 kHz. **a** *A. andreae*. **b** *A. trivittata*. **c** *R. granulosa*

*granulosa* and *A. hylaedactyla* species together, using only a family classifier, and secondly, there are acoustic similarities in the feature space between these families. Figure 8 depicts three spectrograms (time versus frequency) showing the acoustic similarities. Furthermore, comparing Table 3 with Table 6, we observed that the LCPL approach achieved a greater number of true positives for *Hypsiboas cinerascens*, *Hypsiboas cordobae*, and *Leptodactylus fuscus* species.

Additionally, analyzing the confusion matrix at the family level and comparing it with the genus level, we note that *Adenomera hylaedactyla* lost about 50% of its samples at the genus level against the *Hypsiboas* class. We also found it difficult to recognize *Rhinella granulosa* in the presence of *Hypsiboas cinerascens* and *Hypsiboas cordobae*. However, the opposite case is not equally true. It suggests that the samples of the *Rhinella* genus are probably surrounded and overlapped with samples of *Hypsiboas* into the feature space. Similar conclusions can be achieved for other genera and species. For instance, several samples of *Scinax* were confused with *Dendropsophus* and *Hypsiboas*. Inside the *Hypsiboas* genus, *Hypsiboas cordobae* was mostly confused species with *Scinax ruber*.

One advantage of this hierarchical method lies in the fact that, for a real scenario with the *Rhinella granulosa*, *Hyla minuta*, *Sinax ruber*, *Osteocephalus oophagus*, *Hypsiboas cin-*

**Table 7** Hierarchical LCPN

| Species | kNN | | RBF SVM | | Poly SVM | | Tree | |
|---|---|---|---|---|---|---|---|---|
| | Prec. | Rec. | Prec. | Rec. | Prec. | Rec. | Prec. | Rec. |
| *A. andreae* | 0.73 | 0.37 | 0.80 | 0.22 | 0.60 | 0.34 | 0.72 | 0.50 |
| *A. hylaedactyla* | 0.99 | 0.98 | 0.99 | 0.94 | 0.99 | 0.98 | 0.98 | 0.97 |
| *A. trivittata* | 0.84 | 0.61 | 0.97 | 0.29 | 0.70 | 0.58 | 0.71 | 0.43 |
| *H. minuta* | 0.67 | 0.73 | 0.83 | 0.43 | 0.70 | 0.84 | 0.32 | 0.64 |
| *H. cinerascens* | 0.84 | 0.87 | 0.16 | 0.72 | 0.89 | 0.84 | 0.67 | 0.60 |
| *H. cordobae* | 0.89 | 0.96 | 0.87 | 0.82 | 0.86 | 0.96 | 0.88 | 0.91 |
| *L. fuscus* | 0.54 | 0.78 | 0.98 | 0.30 | 0.71 | 0.72 | 0.35 | 0.20 |
| *O. oophagus* | 0.42 | 0.43 | 0.97 | 0.33 | 0.15 | 0.35 | 0.21 | 0.13 |
| *R. granulosa* | 0.16 | 0.83 | 0.96 | 0.82 | 0.27 | 0.79 | 0.11 | 0.72 |
| *S. ruber* | 0.84 | 0.61 | 0.88 | 0.47 | 0.94 | 0.70 | 0.72 | 0.45 |
| Average | **0.69** | **0.72** | **0.84** | **0.54** | **0.68** | **0.71** | 0.57 | **0.56** |
| Macro-F1 | 0.70 | | 0.65 | | 0.70 | | 0.56 | |

**Table 8** Hierarchical LCPL

| Species | kNN | | RBF SVM | | Poly SVM | | Tree | |
|---|---|---|---|---|---|---|---|---|
| | Prec. | Rec. | Prec. | Rec. | Prec. | Rec. | Prec. | Rec. |
| *A. andreae* | 0.73 | 0.37 | 0.87 | 0.22 | 0.60 | 0.35 | 0.54 | 0.44 |
| *A. hylaedactyla* | 0.99 | 0.98 | 0.99 | 0.95 | 0.99 | 0.99 | 0.98 | 0.93 |
| *A. trivittata* | 0.84 | 0.61 | 0.97 | 0.30 | 0.74 | 0.56 | 0.71 | 0.43 |
| *H. minuta* | 0.66 | 0.73 | 0.83 | 0.43 | 0.65 | 0.83 | 0.31 | 0.50 |
| *H. cinerascens* | 0.86 | 0.87 | 0.36 | 0.56 | 0.72 | 0.85 | 0.58 | 0.70 |
| *H. cordobae* | 0.90 | 0.96 | 0.43 | 0.93 | 0.90 | 0.96 | 0.87 | 0.84 |
| *L. fuscus* | 0.54 | 0.78 | 0.98 | 0.31 | 0.70 | 0.74 | 0.25 | 0.20 |
| *O. oophagus* | 0.40 | 0.45 | 0.97 | 0.28 | 0.48 | 0.55 | 0.07 | 0.08 |
| *R. granulosa* | 0.16 | 0.83 | 0.96 | 0.82 | 0.22 | 0.79 | 0.11 | 0.72 |
| *S. ruber* | 0.83 | 0.61 | 0.94 | 0.45 | 0.92 | 0.79 | 0.34 | 0.17 |
| Average | **0.69** | **0.72** | **0.83** | **0.52** | **0.69** | **0.74** | 0.47 | 0.50 |
| Macro-F1 | 0.70 | | 0.64 | | 0.72 | | 0.49 | |

*erascens*, and *Hypsiboas cordobae* species, the method can get a high recognition rate, even better than a flat classifier, because these species belong to different families. Another option is to train the hierarchical method until the genus level to separate, with high precision, the set of species {*Adenomera andreae* and *Adenomera hylaedactyla*} from the set {*Hypsiboas cinerascens* and *Hypsiboas cordobae*} or from {*Scinax ruber*}, decreasing the computational load. However, the major problem with this approach occurs when misclassifications take place at the higher levels and propagate to the lowest levels.

A summary of the results by species is given in Tables 7 and 8. Again, the statistical *t*-test was applied to the precision and recall columns. It may be noted that the best precision was achieved by the RBF-SVM of the LCPN method. However, the best hierarchical M-F1 was using the combination Poly-SVM and LCPL. Comparing Tables 2, 7 and 8 we noticed that the

gains of the hierarchical models were moderate, but we showed how the constraints imposed in the hierarchies impact on the final results, and this was part of our initial hypothesis.

### 7.3 Probabilistic rule with rejection

The response of the different hierarchical levels allows us to reduce the space of possible solutions (see Fig. 7), as well as to combine such solutions to increase the confidence of the final response. If we consider the probability that each classifier associates with the class of its respective level, then we can compose a rule that uses those probabilities to obtain the final probability of each sample.

For instance, let us consider the proposed LCPN model. In this model the first level gives us the probability that any sample belongs to one of the family classes, i.e. $\mathrm{P}(f)$. Given that the second level classifiers of the LCPN hierarchy are trained only with samples from one of the four possible families, each one of them can only have confidence in recognizing the genus from a single family. Therefore, the outputs of the second level classifiers can be interpreted as $\mathrm{P}(g|f)$, i.e. the probability of a sample being of the genus $g$ given that it belongs to the family $f$. The last level of classifiers was trained only with samples of the different species belonging to the genus and the families given by taxonomy of the higher levels. In other words, the output of the last classification level of the LCPN model is $\mathrm{P}(s|f, g)$.

Therefore, by applying the chain rule we can calculate the joint probability $\mathrm{P}(s \cap g \cap f)$ in each branch of the LCPN model using only the conditional probabilities as:

$$\mathrm{P}(s, g, f) = \mathrm{P}(s|g, f)\mathrm{P}(g|f)\mathrm{P}(f), \tag{5}$$

and use it as a final decision score. Note that after classifying a new sample we will have ten $\mathrm{P}(s, g, f)$ values, one for each branch of the taxonomy proposed in Fig. 2, and then we can normalize each $\mathrm{P}(s, g, f)$ by the sum of all the joint probabilities as:

$$\mathrm{P}_i(s, g, f) = \frac{\mathrm{P}_i(s, g, f)}{\sum_{i=1}^{10} \mathrm{P}_i(s, g, f)}. \tag{6}$$

where the index $i$ indicates one of our ten species.

Now, the decision to accept or reject a new sample can be taken by applying a threshold $T$ to the maximum confidence level $\mathrm{P}_i(s, g, f)$. Thus if a classified sample gets a $\mathrm{P}_i(s, g, f)$ value greater than $T$ (e.g. $\mathrm{P}_i(s, g, f) \geq T$) we accept it as correct, otherwise (e.g. $\mathrm{P}_i(s, g, f) < T$) we reject it or store it for future inspection by a specialist. This procedure allows us to evaluate different values of T, thus obtaining the performance of the proposed model and the rejection rate.

In Fig. 9a we observe different values of Macro-F1 in relation to the variation of the threshold between $0 \leq T \leq 1$ for the LCPN. As expected, when $T$ increases, the cases with lower confidence are rejected improving the performance of our model. However, for high $T$ values, the amount of rejected samples may become unacceptable. The relationship between $T$ and the rejection rate is shown in Fig. 9b. Among all the curves illustrated in Fig. 9, we observed that using RBF as the base classifier in the LCPN model we obtain a considerable improvement in the classification rate, but the number of rejected samples grows rapidly until reaching unacceptable values of almost 50% of the total samples. The best trade-off curve corresponds to kNN. This curve showed an improvement of 10% in model performance but maintaining the lowest rejection rate. This shows that, when we used kNN in the hierarchical classifier, the intermediary nodes of the hierarchy were less likely to err. Lastly, Table 9 summarizes different trade-offs for the value $T = 0.98$.

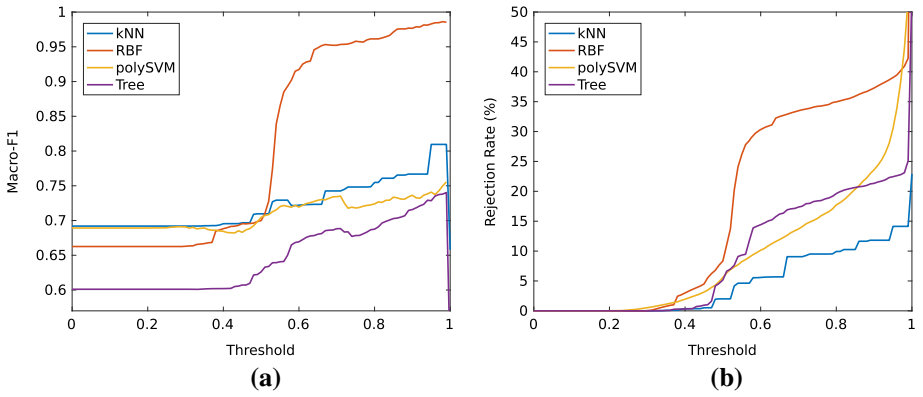


**Fig. 9** Impacts of the probabilistic decision rule on performance (**a**) and number of rejected samples (**b**), for several levels of $T$. **a** Decision versus Macro-F1. **b** Decision versus rejection rate

**Table 9** Best trade-offs between rejection and performance for a threshold equals to 0.98 in LCPN model

| | kNN | RBF SVM | Poly SVM | Tree |
|---|---|---|---|---|
| Macro-F1 | 0.80 | 0.98 | 0.74 | 0.73 |
| Rejection (%) | 14.13 | 40.92 | 43.79 | 23.10 |

Finally, we would like to point out that our rejection strategy using conditional probabilities can be applied to the LCPL model and also to discover new samples that do not correspond to the set of species used to train the hierarchical classifier. In LCPL case, the number of possible paths from the root of the hierarchy to the final nodes includes a higher number of combinations to obtain $\mathrm{P}_i(s, g, f)$, since each $\mathrm{P}_i$ should be obtained by aggregating all possible paths. Such a number of combinations grows exponentially with the amount of species and with the number of hierarchical levels. Besides that, there are some pathways on the hierarchy for which the score is meaningless. For instance, no matter how much is the score $\mathrm{P}_i(s, g, f)$ for the path: *Bufonidae* family, *Scinax* genus and *Adenomera hylaedactyla* species, such a combination is not a valid taxonomy.

## 8 Conclusion

We presented a hierarchical classification approach for frog species recognition using their calls and the biological taxonomy information. The main algorithmic contribution is how to prune the training data using a tree customized structure, i.e., after the high-level class is decided, the number of class options in the lower levels is reduced considering an LCPN. This procedure has two main advantages: (a) it helped us reduce the feature and the solution spaces for the classifier, and (b) allowed us develop a decision rule based on the conditional probabilities to access the final confidence of the hierarchical classification.

First, we transform the original multiclass problem with a single label into a multi-output problem adding the genus and family labels. It allows us to understand and investigate with greater depth, or different granularities, the relationship between the samples and their taxonomy. Our LCPN hierarchical system is able to decompose and simplify this multi-output problem into smaller subproblems avoiding the disadvantages of flat classifiers in our application context. Also, our LCPL baseline performs an aggregation of decision into

the feature space, adding more sources of information. The decision then corresponds to an overlap of three decision functions. In addition, we present an example of a confusion matrix analysis in three levels, useful for understanding the nature of our problem and the relationship between the samples. Lastly, we used Macro-metrics suitable for unbalanced datasets.

The combination of the phylogenetic taxonomy together with the cepstral frequency coefficients and the proximity obtained through the kNN classifier, enables us to notice the bioacoustics similarities between different species from a classification point of view. We can conclude that the *Adenomera hylaedactyla*, *Ameerega trivittata*, *Hypsiboas cinerascens* and *Scinax ruber* species are clearly recognizable in the presence of other species using an LCPL with kNN, and therefore are good candidates for an automatic acoustic monitoring program. We would like to emphasize that these species belong to different families and genera confirming that our hierarchical strategy is indeed advantageous for this type of application. Another interesting fact is that the *Hypsiboas cordobae* species, which belongs to another country, far away from the tropical area, is easy to recognize.

In addition, we performed a similar test using a flat classifier with the same configurations used for the nodes of the hierarchical approaches.With this, we achieved a similar performance from that obtained with the hierarchical approach without probabilistic decision rule. However, by adding the new joint probabilistic rule in the hierarchical classification we obtained better results as opposed to a flat classifier. Additionally, with our method we can obtain several complementary information related to the taxonomy and to the inner organization of the feature space. Previously, we highlighted that we are carrying out a cross-validation by specimens. Therefore, comparing the results between the flat and hierarchical approaches, we note that the error is more dependent on cross-validation than on the classification method itself.

Finally, one drawback in most of the hierarchical classification approaches is the error propagation. Unfortunately, each level of the hierarchical tree could have some misclassifications that will compound the final error when we go down through the tree. As a result, practical applications usually require corrections to eliminate the confusing cases, especially when the hierarchy is deeper and composed from many levels, i.e., the accuracy may decrease when the number of levels increases.

In order to handle this problem, we proposed a soft decision strategy based on the posterior probabilities of each level. Then, in Sect. 7.3 we showed how to get a score ($\mathrm{P}_i(s, g, f)$) for each hierarchy pathway, and how to use it to reject possibly confusing cases. With this, we intend to correct the misclassifications of the highest levels using the confidence of the lower levels. This rule may also help us identify vocalizations of new species that were absent in the original training set, assigning them a low score.

**Acknowledgements** Dr. Juan G. Colonna gratefully acknowledges to National Council of Technological and Scientific Development (CNPq, Brazil) for the Ph.D. fellowship, FAPEAM (PROTI) and CAPES for the financial support. Eduardo F. Nakamura acknowledges FAPEAM for the support granted through the Anura Project (FAPEAM/CNPq PRONEX 023/2009). We also thank professors Marcelo Gordo and the biologist Celeste Salineros for the help with the recordings. This work was supported by the research project "TEC4Growth-Pervasive Intelligence, Enhancers and Proofs of Concept with Industrial Impact/NORTE-01-0145-FEDER-000020", financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF) and by European Commission through the project MAESTRA (Grant No. ICT-2013-612944).

## References

Adams, M. J., Miller, D. A. W., Muths, E., Corn, P. S., Grant, E. H. C., Bailey, L. L., et al. (2013). Trends in amphibian occupancy in the united states. *PLoS ONE*, *8*(5), 1–5.

Angiani, G., Cagnoni, S., Chuzhikova, N., Fornacciari, P., Mordonini, M., & Tomaiuolo, M. (2016). Flat and hierarchical classifiers for detecting emotion in tweets. In G. Adorni, S. Cagnoni, M. Gori, & M. Maratea (Eds.), *AI * IA 2016 advances in artificial intelligence* (pp. 51–64). Berlin: Springer.

Babbar, R., Partalas, I., Gaussier, E., & Amini, M.-R. (2013) . On flat versus hierarchical classification in large-scale taxonomies. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 1824–1832).

Borchani, H., Larrañaga, P., Gama, J., & Bielza, C. (2016). Mining multi-dimensional concept-drifting data streams using Bayesian network classifiers. *Intelligent Data Analysis*, *20*(2), 257–280.

Carey, C., & Alexander, M. A. (2003). Climate change and Amphibian declines: Is there a link? *Diversity and Distributions*, *9*(2), 111–121.

Ceci, M., & Malerba, D. (2007). Classifying web documents in a hierarchy of categories: A comprehensive study. *Journal of Intelligent Information Systems*, *28*(1), 37–78.

Cole, E. M., Bustamante, M. R., Reinoso, D. A., & Funk, W. C. (2014). Spatial and temporal variation in population dynamics of andean frogs: Effects of forest disturbance and evidence for declines. *Global Ecology and Conservation*, *1*, 60–70.

Colonna, J. G., Cristo, M. A. P., & Nakamura, E. F. (2014) . A distribute approach for classifying anuran species based on their calls. In *22nd international conference on pattern recognition*.

Colonna, J. G., Cristo, M. A. P., Salvatierra, M., & Nakamura, E. F. (2015). An incremental technique for real-time bioacoustic signal segmentation. *Expert Systems with Applications*, *42*(21), 7367–7374.

Colonna, J. G., Gama, J., & Nakamura, E. F. (2016). How to correctly evaluate an automatic bioacoustics classification method. In *Volume 9868 of advances in artificial intelligence. Lecture Notes in Computer Science (LNCS)* (pp. 37–47). Berlin: Springer.

Colonna, J. G., Ribas, A. D., de Santos, E. M., & Nakamura, E. F. (2012). Feature subset selection for automatically classifying anuran calls using sensor networks. In *IEEE international joint conference on neural networks (IJCNN)* (pp. 1–8).

Colonna, J. G., Gama, J., & Nakamura, E. F. (2016b). *Recognizing family, genus, and species of anuran using a hierarchical classification approach, volume 9956 of Lecture Notes in Computer Science (LNCS)* (pp. 198–212). Berlin: Springer.

Colonna, J. G., Nakamura, E. F., & Rosso, O. A. (2018). Feature evaluation for unsupervised bioacoustic signal segmentation of anuran calls. *Expert Systems with Applications*, *106*, 107–120.

Da Silva, F. R. (2010). Evaluation of survey methods for sampling anuran species richness in the neotropics. *South American Journal of Herpetology*, *5*(3), 212–220.

Freitas, A. A., & Carvalho, A. (2007). A tutorial on hierarchical classification with applications in bioinformatics. In D. Taniar (Ed.), *Research and trends in data mining technologies and applications, chapter 7* (pp. 175–208). Hershey: Idea Group Pub.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning, Springer Series in Statistics* (Vol. 1). Berlin: Springer.

Frost, D. R. (2016). *Amphibian species of the world: An online reference*. Electronic Database accessible at http://goo.gl/3WRZhx. American Museum of Natural History, New York, USA.

Fürnkranz, J. (2001). Round robin rule learning. In *Proceedings of the eighteenth international conference on machine learning, ICML '01* (pp. 146–153).

Gingras, B., & Fitch, W. T. (2013). A three-parameter model for classifying anurans into four genera based on advertisement calls. *The Journal of the Acoustical Society of America*, *133*(1), 547–559.

Han, N. C., Muniandy, S. V., & Dayou, J. (2011). Acoustic classification of australian anurans based on hybrid spectral-entropy approach. *Applied Acoustics*, *72*(9), 639–645.

Houlahan, J. E., Findlay, C. S., Schmidt, B. R., Meyer, A. H., & Kuzmin, S. L. (2000). Quantitative evidence for global amphibian population declines. *Nature*, *404*(6779), 752–755.

Huang, C. J., Yang, Y. J., Yang, D. X., & Chen, Y. J. (2009). Frog classification using machine learning techniques. *Expert Systems with Applications*, *36*(2), 3737–3743.

Huang, P. X. (2016). Fish4Knowledge: Collecting and analyzing massive coral reef fish video data. In: R. B. Fisher, Y. -H. Chen-Burger, D. Giordano, L. Hardman, & F. -P. Lin (Eds.), *Hierarchical classification system with reject option for live fish recognition* (pp 141–159). Berlin: Springer.

IUCN. (2016). Geographic patterns. The IUCN Red List of Threatened Species. http://goo.gl/nq2qt7. Accessed 4 May 2016

Jaafar, H., & Ramli, D. A. (2013). Automatic syllables segmentation for frog identification system. In *IEEE 9th international colloquium on signal processing and its applications (CSPA)* (pp. 224–228).

Jaafar, H., Ramli, D. A., & Rosdi, B. A. (2014). Comparative study on different classifiers for frog identification system based on bioacoustic signal analysis. In *Proceedings of the 2014 international conference on communications, signal processing and computers*.

King, V. (1969). A study of the mechanism of water transfer across frog skin by a comparison of the permeability of the skin to deuterated and tritiated water. *The Journal of Physiology*, *200*(2), 529–538.
Magid, A., Rotman, S. R., & Weiss, A. M. (1990). Comments on picture thresholding using an iterative selection method. *IEEE Transactions on Systems, Man, and Cybernetics*, *20*(5), 1238–1239.
Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., et al. (2013). Estimating animal population density using passive acoustics. *Biological Reviews*, *88*(2), 287–309.
Rabiner, L., & Schafer, R. (2007). Introduction to digital speech processing. *Foundations and Trends in Signal Processing*, *1*, 1–194.
Ribas, A. D., Colonna, J. G., Figueiredo, C. M. S., & Nakamura, E. F. (2012). Similarity clustering for data fusion in wireless sensor networks using k-means. In *IEEE international joint conference on neural networks (IJCNN)* (pp. 1–7).
Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, *13*(1), 146–168.
Silla, C. N., & Kaestner, C. A. A. (2013). Hierarchical classification of bird species using their audio recorded songs. In *International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1895–1900). IEEE.
Silla, C. N, Jr., & Freitas, A. A. (2011). A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, *22*(22), 31–72.
Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437.
Sueur, J., Pavoine, S., Hamerlynck, O., & Duvail, S. (2008). Rapid acoustic survey for biodiversity appraisal. *PLoS ONE*, *3*(12), e4065.
Vaca-Castaño, G., & Rodriguez, D. (2010). Using syllabic mel cepstrum features and k-nearest neighbors to identify anurans and birds species. In *2010 IEEE workshop on signal processing systems (SIPS)* (pp. 466–471).
Xie, J., Towsey, M., Truskinger, A., Eichinski, P., Zhang, J., & Roe, P. (2015). Acoustic classification of Australian anurans using syllable features. In *IEEE tenth international conference on intelligent sensors, sensor networks and information processing (ISSNIP 2015)*.
Xie, J., Towsey, M., Zhang, J., & Roe, P. (2016). Frog call classification: A survey. *Artificial Intelligence Review*, *46*(161), 1–17.
Xie, J., Zhang, J., & Roe, P. (2015). Acoustic features for hierarchical classification of Australian frog calls. In *10th international conference on information, communications and signal processing*.
Zimek, A., Buchwald, F., Frank, E., & Kramer, S. (2010). A study of hierarchical and flat classification of proteins. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *7*(3), 563–571.



## Affiliations

**Juan G. Colonna[1,4] · João Gama[2] · Eduardo F. Nakamura[1,3]**

João Gama
jgama@fep.up.pt

Eduardo F. Nakamura
nakamura@icomp.ufam.edu.br; nakamura@tamu.edu

1 Institute of Computing (Icomp), Federal University of Amazonas (UFAM), Avenida General Rodrigo Octávio 6200, Manaus, AM 69077-000, Brazil

2 Laboratory of Artificial Intelligence and Decision Support (LIAAD), INESC Tec, Campus da FEUP, Rua Dr. Roberto Frias, Porto 4200-465, Portugal

3 Department of Computer Science and Engineering, Texas A&M University, H. R. Bright Building, 3112 TAMU, 710 Ross St, College Station, USA

4 Samsung R&D Institute of Amazônia (SIDIA), Av. Mário Ypiranga, 315, Manaus, AM CEP 69075-155, Brazil