Recognizing Family, Genus, and Species of Anuran Using a Hierarchical Classification Approach

Juan G. Colonna[1(✉)], João Gama[2], and Eduardo F. Nakamura[1]

[1] Institute of Computing (Icomp), Federal University of Amazonas (UFAM),
Avenida General Rodrigo Octávio 6200, Manaus, AM 69077-000, Brazil
{juancolonna,nakamura}@icomp.ufam.edu.br
[2] Laboratory of Artificial Intelligence and Decision Support (LIAAD), INESC Tec,
Campus da FEUP, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
jgama@fep.up.pt

Abstract. In bioacoustic recognition approaches, a "flat" classifier is usually trained to recognize several species of anuran, where the number of classes is equal to the number of species. Consequently, the complexity of the classification function increases proportionally to the amount 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 level. To accomplish this, we transform the original single-label problem into a multi-dimensional problem (multi-label and multi-class) considering the Linnaeus taxonomy. Then, we develop a top-down method using a set of classifiers organized as a hierarchical tree. Thus, 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 understand the problem better and achieve additional conclusions by the inspection of the confusion matrices at the three levels of classification. In addition, we carry out our experiments using a 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 system in a dataset with sixty individual frogs, from ten different species, eight genus, and four families, achieving a final Micro- and Average-accuracy equal to 86 % and 62 % respectively.

1 Introduction

Amphibians are directly affected by environmental changes [3,4]. This observation has motivated many researchers to monitor the decline of amphibian populations through time and use it as an indicator of environmental problems [1,13]. Among all amphibian species that may be monitored anuran (frogs and toads) are preferred, because these have a semi-permeable skin which makes them sensitive to aquatic and terrestrial conditions [19]. Nowadays, the most widely used

method to monitor frog populations takes advantage of the vocalization capability to apply acoustics surveys [20,25]. However, manual application of these acoustic surveys requires many human and economic resources, as well as expert knowledge, being difficult to apply in remote tropical areas of the Amazon rainforest. Therefore, our goal is to develop an Automatic Calls Recognition (ACR) system to monitor frog populations automatically and in a less invasive manner using acoustic sensors. The general idea consists of treating the challenge of anuran monitoring as a species recognition task using their calls and Machine Learning (ML) techniques [6,7,14,28].

In bioacoustics most of the related works deal with the species recognition problem using "flat" classifiers, where each instance belongs to one class (or species name in this case), and there is no hierarchical relationship between the classes [7,12,18,22,27,28]. This work we addressed the problem of anuran species recognition through their calls using a "hierarchical" classifier that considers its family and genus taxonomy. For this purpose, the family and genus information of each species was added as new labels, transforming the original problem with a single label into a multi-dimensional approach, i.e., a problem where the outputs are multi-class and multi-label.

The hierarchical approach allows us to test three hypothesis:

(i) the decomposition of the main problem into three levels of small hierarchically related problems, in which the results may improve compared to a normal flat classifier;

(ii) this configuration allows us to understand the relationship between the misclassifications and the acoustic proximity of the species, and their taxonomy, into feature space; and

(iii) the method has better predictive capabilities for new individuals that were not present in the original training set.

Thus, the two main contributions in this work are: a customization of the existing hierarchical models, specially adapted to the anuran species taxonomy; and the advantages of this model in our bioacoustics application context.

In order to test the first hypothesis we introduce our hierarchical system in Sect. 5. To accomplish this, we give a detailed explanation about how this approach can reduce the complexity of the model from a feature space point of view and, consequently, simplify the decision function. To test the second hypothesis we compare the confusion matrix of each classification level, i.e., family, genus, and species levels in Sect. 7. To test the third hypothesis we carry all of our experiments using Cross-Validation (CV) by individuals (or specimens) as explained in Sect. 6.3. The results and conclusion are supported by the calculation of the Micro- and Average-accuracy by level (see Sects. 6.4, 7 and 8).

In addition (Sect. 4) we also discuss how hierarchical models were applied to the anuran recognition task (particularly the prediction of frogs and toads species) and, in general, to the bioacoustics problems, in which a hierarchical relationship between the labels could be modeled. Finally, we would like to emphasize that our work is the first one regarding the combination of a

hierarchical approach together with a CV procedure by individuals using the Linnaeus taxonomy.

2   Motivation for Using a Hierarchical Approach

Anura is the name of an order in the Amphibia class of animals that includes frogs and toads. According to recent reports there are more than 6600 different species of anuran in the world, classified into 56 families and several genus [9]. The anuran diversity in the tropical areas of South America is the greatest, concentrating approximately 70 % on the global biodiversity of amphibians [16]. In order to develop a flat classifier we need 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 classes, becoming an intractable problem in certain scenarios. A hierarchical approach can alleviate this problem by decomposing the classification function in several levels, similarly to a decision tree. Thus, we use the well known Linnaeus taxonomy to construct a system with three levels: family, genus, and species (see Sect. 5). With this, every time we go down through the tree to another level, the output space of possible solutions is simplified.

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. 1). Formally, the input bioacoustic signal $X = \{x_1, x_2, \cdots, x_N\}$ is a time series of length N, in which its values represent the acoustics pressure levels (or amplitude). A syllable $\mathbf{x}_k = \{x_t, x_{t+1}, \cdots, x_{t+n}\}$ is a subset of n consecutive signal values. Thus, the pre-processing step segments the signal X by identifying the beginning and the endpoints of $\mathbf{x}_k$ (Fig. 2(a)) [6].

After the syllable extraction we need to represent each $\mathbf{x}_k$ by a set of features, commonly called Low Level Descriptors (LLDs). The most frequent LLDs are the Mel-Frequency Spectral Coefficients (MFCCs). The MFCCs perform a spectral analysis based on a triangular filter-bank logarithmically spaced in the frequency
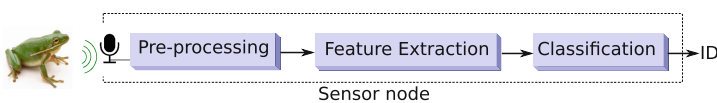
Fig. 1. An automatic call recognition system (ACR).

domain (Fig. 2(b)) [7,21]. The feature extraction using the MFCCs allows to represent any syllable by a set of coefficients (MFCC($\mathbf{x}_k$) → $\mathbf{c}_k$), i.e., X → $\{(\mathbf{c}_1, s_i), (\mathbf{c}_2, s_i), \ldots, (\mathbf{c}_k, s_i)\}$, 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 use raw data.
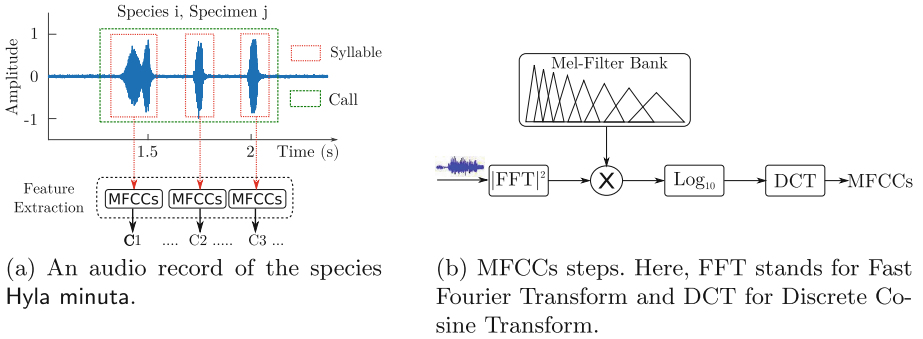

(a) An audio record of the species Hyla minuta.

(b) MFCCs steps. Here, FFT stands for Fast Fourier Transform and DCT for Discrete Cosine Transform.

Fig. 2. A framework for automatic frog's calls recognition.

Finally, the challenge is how to assign the species name to a new syllable by using the MFCC values. This is a supervised classification task and is performed by the last step of the system. For this purpose several ML algorithms could be applied to create and train 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 names.

3.2 Review of Hierarchical Classification Approaches

Hierarchical methods are widely 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. These methods help to simplify complex multi-class problems transforming these into a multi-label approach by considering the hierarchical relationship between the labels. For instance, every time we go down in a level of the hierarchy, the number of possible solutions is reduced, simplifying the decision function, as showed by Fig. 5. 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 leaf 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 [8]. In our approach we adopted the tree structure, due to the taxonomic constraints of our problem, in which every species can belong to just one genus class and one family class at the same time.
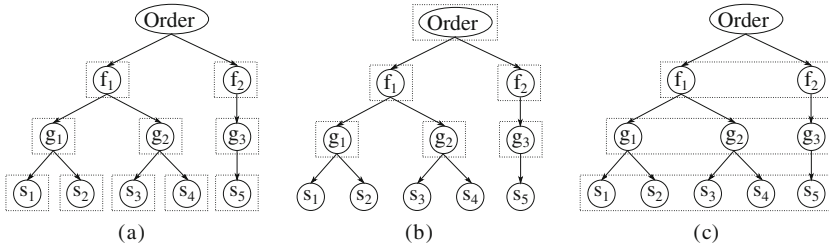
Fig. 3. Different manners to create a hierarchical classifier combining flat classifiers. From top-to-down levels: **f** stands for family, **g** for genus, and **s** for species.

Figure 3 illustrates three different approaches commonly used to construct a multi-label hierarchical tree from a set of flat classifiers. These are: (1) one classifier per node, (2) one classifier per parent node, and (3) one classifier per level [24]. These trees may be imbalanced depending on the taxonomic structure of the problem. The classifiers inside the nodes 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 between the set of predicted labels. An obvious disadvantage associated with this top-down approach is the error propagation from the higher levels of the tree. However, this approach is well suited for the context of species recognition where the number of classes is too high to train a flat classifier. Moreover, the configuration shown in Fig. 3(b) fits better the characteristics of our problem presented in Fig. 4.

4   Related Works

Several authors have already studied the problem of recognition and classification of anuran species through their calls. Among these Huang *et al.* [14] and Colonna *et al.* [7] studied the best acoustics features to recognize different species. Jaafar *et al.* [17] and Colonna *et al.* [6] focused on comparing some syllable extraction procedures as a pre-processing step. Finally, Ribas *et al.* [22] and Colonna *et al.* [5] 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 defined in 1935 by Carl Linnaeus.

Gingras et al. [11] 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.* [29]. This work studies which acoustics features should be used in each classification level, considering the taxonomy information separated in 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 including the reject option. This was developed by Phoenix *et al.* [15] for fish image recognition using underwater cameras. In this system a multi-class 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 multi-label purpose, and therefore it is not possible to evaluate the similarities between family, genus and species.

An evaluation on different hierarchical approaches applied to the bird species recognition was performed by Sillas et al. [23]. The authors compared three different approaches: a flat classification where the class hierarchy is disregarded, one classifier per parent node (see Fig. 3), 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 metrics Precision, Recall, and F-measure 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.

To the best of our knowledge, no study has yet been published integrating the family, the genus, and the species labels of anuran in one unique hierarchical approach, to be solved as a multi-dimensional problem and, at the same time, performing a CV by individuals (or specimens) to test the model generalization capabilities.

5   Proposed Approach

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

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

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. This MM problems are also called Multi-dimensional problems because the output is composed by a tuple of labels [2], which are three in our case.

This is possible because there is a 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_j 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 shown in Fig. 3(b). Our hierarchical tree is represented in Fig. 4. An example of problem simplification by the hierarchical decomposition using an example with two attributes is shown in Fig. 5. As we can note, in 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.

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 Linnaeus taxonomy and the reason why we chose the approach described in Fig. 3(b). 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.
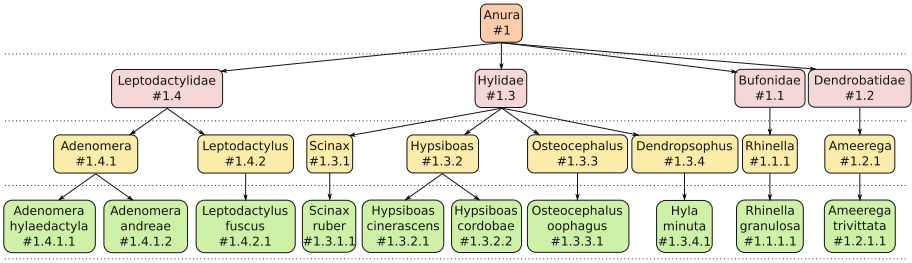
Fig. 4. Species tree. From Top-to-Down levels: Order, Family, Genus and Species. The # stands for node ID.
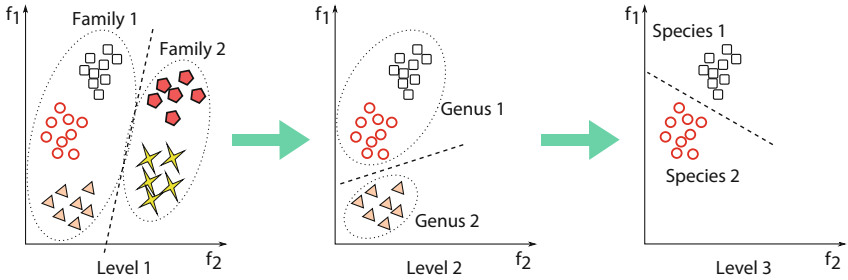
Fig. 5. Problem decomposition stages when performing the hierarchical classification from top to down described by an example of prune training data.

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. For this, we used the taxonomy information available at [9]. The dataset description, the classifier setting, the validation procedure and the metrics used are described in the following subsections.

6.1   Dataset Description

The dataset used in our experiments is summarized in Table 1. It has 10 different species, 60 specimens and 5998 syllables[1]. 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 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, 24 MFCCs were calculated by using 44 triangular filters and these coefficients were normalized between $-1 \leq c_l \leq 1$ (see Sect. 3.1). For the segmentation and syllable extraction tasks we based our approach on

---

[1] Available at https://goo.gl/61IoXc.

Table 1. Species Dataset. The **s** and the **k** stands for the number of specimens and the amount of syllables respectively.

Family	Genus	Species	s	k
Leptodactylidae	Leptodactylus	Leptodactylus fuscus[a]	4	222
	Adenomera	Adenomera andreae[a]	8	496
		Adenomera hylaedactyla[b]	11	3049
Hylidae	Dendropsophus	Hyla minuta[b]	11	229
	Scinax	Scinax ruber[b]	5	96
	Osteocephalus	Osteocephalus oophagus[a]	3	96
	Hypsiboas	Hypsiboas cinerascens[a]	4	429
		Hypsiboas cordobae[c]	4	702
Bufonidae	Rhinella	Rhinella granulosa[a]	5	135
Dendrobatidae	Ameerega	Ameerega trivittata[b]	5	544

[a]Amazonas, [b]Mata Atlântica, [c]Córdoba.

the work of Colonna *et al.* [6], however using only the energy of the signal in a batch mode setting[2]. Finally, the frame size was 0.0464 s with 66 % of overlap to obtain a good energy-time resolution.

6.2   Node Classifier Description

In our experiments we chose kNN with k = 3 as the base classifier for the parent nodes. As kNN is considered a subspace technique, then the predicted result is the similarity between the samples in the feature space and consequently, the acoustics similarity of the syllable's frequencies. 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-One (1A1) procedure [10]. After that, the result of each binary model was combined by using the majority voting rule. This decomposition technique reduces the complexity of each sub-problem compared to the multiclass approach.

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. With *k*-CV the original dataset is split into *k* disjoint folds, and for each one the conditional error (e_k) is estimated 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

---

[2] The segmentation code is available at http://goo.gl/vjVQ2c.

same individuals in the testing and training sets. This causes an overestimate on the accuracy. 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. 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 Measure (Average-Accuracy)

Diverse species of anuran have different syllable rates (amount of syllables per unit time) in their calls. This is a particular vocalization characteristic of each anuran species. Therefore, an unequal number of samples could be retrieved from each record producing an unbalanced dataset [6]. This is a secondary problem that affects the classical accuracy measure. Thus, 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 average-accuracy instead of the traditional micro-accuracy [6,26]. It means, the final accuracy value is calculated as the average accuracy of each species individually as:

$$\text{Average-Acc} = \frac{1}{m}\sum_{i=1}^{m}\text{Acc}_i = \frac{1}{m}\sum_{i=1}^{m}\frac{tp_i}{k_i}, \tag{1}$$

where Acc_i is the accuracy per row i of confusion matrix, m the total number of rows, tp_i are the true positives, and k_i the total number of syllables per row.

7   Experiments and Results

The structure of our hierarchical approach was introduced in Fig. 4. 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 owns genus labels. The last prediction takes place at the genus level being responsible for predicting their owns species names, as shown by their leaf nodes. With this configuration we can obtain a confusion matrix per level, i.e.: one matrix for the family labels (Table 2), one for the genus labels (Table 3), and one for the species labels (Table 4).

The rows of the Tables 2, 3 and 4 are the Ground Truth (GT) labels and the columns indicate the predicted labels. The main diagonal corresponds to the number of hits. From these matrix we can obtain the micro- and average-accuracy by level. The last column of each matrix (Acc) is the accuracy by class, from which we can get the average-accuracy averaging the values

Table 2. Confusion matrix of family level with kNN (k = 3). Last column (Acc) is the accuracy of each column.

	Bufonidae	Dendrobatidae	Hylidae	Leptodactylidae	Acc_i
Bufonidae	**43**	0	21	71	0.31
Dendrobatidae	27	**488**	0	29	0.89
Hylidae	3	0	**1465**	84	0.94
Leptodactylidae	16	36	322	**3393**	0.90

Table 3. Confusion matrix of genus level with kNN (k = 3). Legend: (a) Adenomera, (b) Ameerega, (c) Dendropsophus, (d) Hypsiboas, (e) Leptodactylus, (f) Osteocephalus, (g) Rhinella, and (h) Scinax. Last column (Acc) is the accuracy of each column.

	a	b	c	d	e	f	g	h	Acc_i
a	**3186**	36	18	61	58	184	0	2	0.90
b	23	**488**	0	0	6	0	27	0	0.90
c	51	0	**123**	35	0	0	0	20	0.54
d	7	0	0	**1117**	0	6	0	1	0.99
e	15	0	20	14	**134**	21	16	2	0.60
f	7	0	0	48	4	**34**	3	0	0.35
g	8	0	0	9	63	0	**43**	12	0.32
h	15	0	50	11	0	0	0	**20**	0.21

Table 4. Confusion matrix of species level with kNN (k = 3). Legend: (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. Last column (Acc) is the accuracy of each column.

	a	b	c	d	e	f	g	h	i	j	Acc_i
a	**156**	0	35	2	61	0	58	184	0	0	0.31
b	0	**3030**	1	16	0	0	0	0	0	2	0.99
c	23	0	**488**	0	0	0	6	0	27	0	0.90
d	3	48	0	**123**	3	32	0	0	0	20	0.54
e	1	6	0	0	**415**	0	0	6	0	1	0.97
f	0	0	0	0	0	**702**	0	0	0	0	1.00
g	1	14	0	20	0	14	**134**	21	16	2	0.60
h	7	0	0	0	48	0	4	**34**	3	0	0.35
i	8	0	0	0	6	3	63	0	**43**	12	0.32
j	0	15	0	50	9	2	0	0	0	**20**	0.21

Table 5. Results summary. G stands for the accuracy gains compared to the naive baseline approach for each case.

	Level					
	Family	G (%)	Genus	G (%)	Species	G (%)
Micro-Baseline	0.63		0.59		0.51	
Micro-Acc	0.90	+0.27	0.86	+0.27	0.86	+0.35
Average-Baseline	0.25		0.13		0.10	
Average-Acc	0.76	+0.51	0.60	+0.48	0.62	+0.52

of this column. A summary of the results is presented in Table 5. For the micro-accuracy case (Micro-Acc) the baseline values are given by a naive classifier which always chooses the most numerous classes (Micro-Baseline), and for the average-accuracy (Average-Acc) the baseline values are given by a naive classifier that always chooses a label randomly (Average-Baseline).

Just analyzing the confusion matrix at the family level, we can note that Bufonidae family lost about 70 % of the samples, in which almost 50 % fall in the class Leptodactylidae. That means, the Bufonidae family seems to find it hard to recognize in the presence of Leptodactylidae. However, the opposite case is not equally true. This means that the samples of the Bufonidae family are probably surrounded by samples of the Leptodactylidae in the feature space. A similar conclusion can be achieved analyzing the genus and the species levels. For instance, several samples of the Scinax were confused with Adenomera, Dendropsophus and Hypsiboas. Inside the Adenomera genus, hylaedactyla was the only confused species with the Scinax ruber.

Previously, we highlighted that we are carrying out a cross-validation by specimens, therefore we can infer that different individuals of the Leptodactylidae family share high similarities or regularities between them in the feature space. The same is valid for the Hylidae family. Therefore, we are able to recognize better the specimens that belong to these species and, consequently, they are good candidates to use them in a monitoring acoustic project.

In addition, we performed a similar test by species using only one flat classifier with the same configurations used for the nodes of the tree (that is, 1A1 and kNN with $k = 3$). In this test the Micro-accuracy was 0.85 and the Average-accuracy was 0.61 achieving results comparable to our approach. However, with our approach we can obtain several complementary information related to the taxonomy. Unfortunately, our dataset is not big enough to be able to observe the gains of the hierarchical classification at the last level.

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 are reduced.

First, we transform the original multiclass problem with a single label into a multidimensional problem adding the genus and family labels. It allows us to understand and investigate with greater depth the relationship between the samples and their taxonomy. Our hierarchical system is able to decompose and simplify this multidimensional problem into smaller subproblems avoiding the disadvantages of flat classifiers in this application context. In addition we present the confusion matrices and the Micro-accuracy and Average-accuracy at the three levels of the decomposition, useful to understand the nature of our problem and the relationship between the samples.

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 species Adenomera hylaedactyla and Hypsiboas cinerascens are clearly recognizable in the presence of other species, and therefore are good candidates for an automatic acoustic monitoring program. We would like to emphasize that these two species belong to different families and genus confirming that our hierarchical strategy is indeed advantageous for this type of application. Another interesting fact is that the species Hypsiboas cordobae, which belongs to another country, far away from the tropical area, is easy to recognize.

However, one major 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 database is imbalanced or when the hierarchy is deeper and composed from many levels, i.e., the accuracy decreases when the number of levels increases. As future work, in order to handle this problem, we propose the development of a hierarchical tree with a soft decision strategy based on the posterior probabilities of each level. With this, we intend to correct the misclassifications of the highest levels using the confidence of the lower levels.

References

1. Adams, M.J., Miller, D.A.W., Muths, E., Corn, P.S., Grant, E.H.C., Bailey, L.L., Fellers, G.M., Fisher, R.N., Sadinski, W.J., Waddle, H., Walls, S.C.: Trends in amphibian occupancy in the United States. PLoS One **8**(5), 1–5 (2013)
2. Borchani, H., Larrañaga, P., Gama, J., Bielza, C.: Mining multi-dimensional concept-drifting data streams using Bayesian network classifiers. Intell. Data Anal. **20**(2), 257–280 (2016)
3. Carey, C., Alexander, M.A.: Climate change and amphibian declines: is there a link? Divers. Distrib. **9**(2), 111–121 (2003)
4. Cole, E.M., Bustamante, M.R., Reinoso, D.A., Funk, W.C.: Spatial and temporal variation in population dynamics of andean frogs: effects of forest disturbance and evidence for declines. Glob. Ecol. Conserv. **1**, 60–70 (2014)
5. Colonna, J.G., Cristo, M.A.P., Nakamura, E.F.: A distribute approach for classifying anuran species based on their calls. In: 22nd International Conference on Pattern Recognition (2014)
6. Colonna, J.G., Cristo, M.A.P., Salvatierra, M., Nakamura, E.F.: An incremental technique for real-time bioacoustic signal segmentation. Expert Syst. Appl. **42**(21), 7367–7374 (2015)
7. Colonna, J.G., Ribas, A.D., dos Santos, E.M., Nakamura, E.F.: Feature subset selection for automatically classifying anuran calls using sensor networks. In: International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE, June 2012
8. Freitas, A.A., Carvalho, A.C.P.L.F.: A tutorial on hierarchical classification with applications in bioinformatics. In: Taniar, D. (ed.) Research and Trends in Data Mining Technologies and Applications, Chap. 7, pp. 175–208. Idea Group Publishing (2007)
9. Frost, D.R.: Amphibian species of the world: an online reference. Electronic Database accessible at http://goo.gl/3WRZhx, April 2016. American Museum of Natural History, New York, USA
10. Fürnkranz, J.: Round Robin rule learning. In: Proceedings of the Eighteenth International Conference on Machine Learning, ICML 2001, pp. 146–153 (2001)
11. Gingras, B., Fitch, W.T.: A three-parameter model for classifying anurans into four genera based on advertisement calls. J. Acoust. Soc. Am. **133**(1), 547–559 (2013)
12. Han, N.C., Muniandy, S.V., Dayou, J.: Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. Appl. Acoust. **72**(9), 639–645 (2011)
13. Houlahan, J.E., Findlay, C.S., Schmidt, B.R., Meyer, A.H., Kuzmin, S.L.: Quantitative evidence for global amphibian population declines. Nature **404**(6779), 752–755 (2000)
14. Huang, C.J., Yang, Y.J., Yang, D.X., Chen, Y.J.: Frog classification using machine learning techniques. Expert Syst. Appl. **36**(2), 3737–3743 (2009)
15. Huang, P.X.: Hierarchical classification system with reject option for live fish recognition. In: Fisher, R.B., et al. (eds.) Fish4Knowledge: Collecting and Analyzing Massive Coral Reef Fish Video Data. Intelligent Systems Reference Library, vol. 104, pp. 141–159. Springer, Switzerland (2016)
16. IUCN. Geographic patterns, April 2016. http://goo.gl/nq2qt7. The IUCN Red List of Threatened Species
17. Jaafar, H., Ramli, D.A.: Automatic syllables segmentation for frog identification system. In: 9th International Colloquium on Signal Processing and Its Applications (CSPA), pp. 224–228. IEEE (2013)

18. Jaafar, H., Ramli, D.A., Rosdi, B.A.: 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 (2014)
19. King, V.: 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. J. Physiol. **200**(2), 529–538 (1969)
20. Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti, D.J., Harris, D., Tyack, P.L.: Estimating animal population density using passive acoustics. Bio. Rev. **88**(2), 287–309 (2013)
21. Rabiner, L., Schafer, R.: Introduction to digital speech processing. Found. Trends Sign. Process. **1**, 1–194 (2007)
22. Ribas, A.D., Colonna, J.G., Figueiredo, C.M.S., Nakamura, E.F.: Similarity clustering for data fusion in wireless sensor networks using k-means. In: International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE, June 2012
23. Silla, C.N., Kaestner, C.A.A.: Hierarchical classification of bird species using their audio recorded songs. In: International Conference on Systems, Man, and Cybernetics (SMC), pp. 1895–1900. IEEE, October 2013
24. Silla Jr., C.N., Freitas, A.A.: A survey of hierarchical classification across different application domains. Data Min. Knowl. Discov. **22**(22), 31–72 (2011)
25. Da Silva, F.R.: Evaluation of survey methods for sampling anuran species richness in the neotropics. S. Am. J. Herpetology **5**(3), 212–220 (2010)
26. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. **45**(4), 427–437 (2009)
27. Vaca-Castaño, G., Rodriguez, D.: 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 (2010)
28. Xie, J., Towsey, M., Truskinger, A., Eichinski, P., Zhang, J., Roe, P.: Acoustic classification of australian anurans using syllable features. In: IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP 2015). IEEE (2015)
29. Xie, J., Zhang, J., Roe, P.: Acoustic features for hierarchical classification of australian frog calls. In: 10th International Conference on Information, Communications and Signal Processing (2015)