Classifying Frog Calls Using Gaussian Mixture Model and Locality Sensitive Hashing

Kathryn Hollowood\textsuperscript{1}, Olatide Omojaro\textsuperscript{2} Dalwindeerjeet Kular\textsuperscript{3} and Eraldo Ribeiro\textsuperscript{4}

\textsuperscript{1}Department of Computer Science, Mathematics, and Physics, Roberts Wesleyan College, Rochester, New York; Hollowood.Kathryn@Roberts.edu
\textsuperscript{2}Engineering, Mathematics, & Computer Science, Georgia Perimeter College, Clarkston, Georgia; omojaroo@student.gpc.edu
\textsuperscript{3}Department of Computer Science, Florida Institute of Technology, Melbourne, Florida; dkular2009@my.fit.edu
\textsuperscript{4}Department of Computer Science, Florida Institute of Technology, Melbourne, Florida; eribeiro@fit.edu

ABSTRACT

Biodiversity conservation is a major concern of modern society. Pressures such as littering, pollution, and climate change deteriorate the natural habitats and reduce biodiversity. To help repair this issue, we need some way to track these habitat damages. We need a bio-indicator. Frogs make excellent bio-indicators due to their permeable skin. This skin allows them to absorb water and Oxygen. However, when they live in polluted areas they also absorb toxins. As pollution rise, some frog populations suffer decline. Citizen scientists have been monitoring these frog populations by recording their calls for years. However, this process requires expensive equipment, training, and a sizable time commitment. This paper focuses on the automatic classification of frog calls using computer programs. Features were extracted from the audio data and classified using two classification methods: Locality Sensitive Hashing and Gaussian Mixture Modeling. Tests performed on a dataset of frog calls of 15 different species produced promising classification results for the Gaussian mixture model approach.

Keywords: Anuran Population Control, Classification, Clustering, Frog, Frog Call Classification, Gaussian Mixture Model, Locality Sensitive Hashing, Machine Learning, Mixture Modeling, Supervised Learning

1. INTRODUCTION

The conservation of Earth’s natural habitats is increasingly becoming a major concern of modern society. Worldwide interest in preserving natural habitats is driven not only by the desire to having a greener planet but also by the understanding that a region’s biodiversity wealth can lead to economic wealth.\textsuperscript{10} Regardless the underlying motivation, countries worldwide spend billions of dollars yearly on reducing industrial pollution and on promoting the conservation of marine and terrestrial ecosystems. Among ecosystems, wetlands are considered to have the highest degree of biodiversity. Wetlands are areas that are inundated or saturated with water periodically or for long periods of time. In the U.S., preservation of wetlands is particularly relevant to the economy of southern and coastal states, supporting activities such as tourism, fishing, agriculture, and recreation.\textsuperscript{2}

Efforts aimed at preserving wetlands include the biodiversity monitoring and pollution control. Ecologists monitor wetland health in a number of ways including by taking water samples, analyzing satellite images, and measuring vegetation cover. Another proxy to habitat health is the decline in populations of certain animals that are more sensitive to environmental changes. These animals are called bioindicators. In wetlands, amphibians such as frog and toads (i.e., Anurans) are suitable bio-indicators.\textsuperscript{12,14} Anurans retain toxins by feeding on animals that carry pollutants.\textsuperscript{17} In addition to retaining prey toxins, the permeable nature of the ventral skin of anurans allows absorption of water and moisture into their respiratory and digestive system. If the habitat is polluted, they absorb the pollutants from the water around them, which cause decline in certain anuran species.\textsuperscript{4} While certain anuran population decline can be partially credited to infectious diseases,\textsuperscript{16} the influence of climate change, pollution, the introduction of invasive species, and habitat loss cannot be ignored.\textsuperscript{15}

To obtain estimates of anuran populations, scientists listen to calls of these animals in the wild while identifying the various species. These type of data-collection procedure is time consuming and require expert training...
to improve classification ability. There are also a number of frog-monitoring programs in the U.S. that count with the participation of volunteer citizen scientists. These volunteers are known as citizen scientists because they contribute to the scientific community without any formal scientific education.

The Wisconsin Frog and Toad Survey (WFTS) began in 1981. This program relies on volunteers and researchers identifying frog species by their breeding calls. The WFTS has many routes equipped with listening stations. At each listening station, volunteers listen for about five minutes, and a route takes a couple of hours to complete. Volunteers complete each route three times a year. The reported data is checked and entered into a data bank. The WFTS takes special care to make sure the data that is added into the data bank is correct and the data is being taken at optimum times.

The North American Amphibian Monitoring Program (NAAMP) established the Southwest Florida Amphibian Monitoring Network, also known as Frog Watch, in 2000. Frog Watch involves a group of volunteers from different areas of Southwest Florida going to different stops each month during the rainy season and recording the temperature, humidity, wind speed, degree of cloud cover, precipitation, and changes to adjacent habitat. Here, volunteers listen to frog calls for three minutes and record species and the intensity of the calls. While these frog-monitoring programs are a great way to monitor the local frog populations, they require a sizable time commitment from volunteers to be trained and to perform the frog classifications. This is why there have been efforts to try to classify frogs automatically using computer programs. One of these efforts was from a group in Australia. Researchers were concerned with the introduction of the Cane Toad, *Bufo marinus*, to northern Australia because it was a predator that went after many different types of prey. The goal was to find some way to take a census of the native fauna. They used a classification method based on decision trees. The training data consisted of twenty-two species of frog and a few cricket species that sounded similar to frogs. The species were classified using the peaks of energies on the spectrograms that they produce.

Recent advances in machine learning have motivated the development of a number of new methods for frog-call recognition. Classification techniques such as Support Vector Machines and Hidden Markov Models are able to recognize frog calls with good accuracy. However, most automatic methods for frog-call classification rely on the extraction of syllabus from the acoustic signal, which is a major limitation in applying these methods to monitoring frog communities in the field. While automatic syllable-segmentation methods do exist, they can be unreliable when applied to complex of outdoor signals.

In another work by Belyaeva et al., a number of classification algorithms were compared for precision in effectively recognizing anuran species vocalization that made use of features defined in similar techniques as Chen et al. and Xie et al. Belyaeva et al. extended the feature defined for a frog call to include human-input for improved classification precision of anuran species vocalization.

In the work reported in this paper, we describe two methods for classifying frog calls using machine-learning techniques. In the first method, we build model signatures of calls for each species using locality sensitive hashing (LSH). This method uses principles of bucketing and hashing to reduce the dimensionality of the data. Signatures are created using pairs of features detected on the spectrograms of the calls. Matching is performed by comparing pairs of features detected from unknown calls with pairs from the pre-learned signatures. Our second classification method consists of modeling each frog species as a mixture of Gaussian densities. Here, instead of using signatures of pairs of spectral landmarks, we create feature vectors from filter response maps obtained by convolving the spectrogram of the audio signal with a bank of multi-scale multi-orientation filters.

We tested our methods on a dataset containing calls of 15 different species of frog. Classification rates for the landmark-based LSH method were quite low. While we are still studying the exact reason, we believe that variations in calls of a same species as well as noise may be the causes of the low accuracy of the LSH method. In contrast, the method based on the Gaussian Mixture model produced promising results, reaching classification rates of 80%, which approximates human performance.

2. METHODS

The basic approach for either method begins by taking the audio data and creating a spectrogram. The spectrogram of an audio signal maps the frequency as a function of time. There are different colors on a spectrogram
that represent the different levels of energy for the different frequencies. The highest levels of energy are in red and these are what you would hear when listening to a frog call. These high peaks of energy are our landmarks and from these landmarks we extract features. These features are then used to classify the frog call using either Gaussian Mixture Modeling or Locality Sensitive Hashing. This process is shown in Figure 1.

Figure 1: Basic methodology of our classification methods. Feature extraction uses peaks detected from the spectrogram of the audio signal. For the method based on locality sensitive hashing, landmarks are created for each pair of spectral peaks. A collection of these landmarks form the signature for the frog species. For the method based on Gaussian mixture model, MFCC coefficients located at detected high-energy peaks are converted into feature vectors. Gaussian mixture models are learned for each frog species. Classification is based on a maximum-likelihood approach.

3. GAUSSIAN MIXTURE MODELING

3.1 Definition

GMM uses clustering. Clustering is the concept of grouping together data based on similar characteristics. There are two different kinds of clustering, there is hard clustering and there is soft clustering. Hard clustering is the principle that when data is grouped together the different data points can only completely belong in one cluster or the other. Soft clustering is the principle that the data points can partially belong in both groups. Gaussian Mixture Modelings (GMMs) basically state that data belongs in one group or the other based on the probability that it belongs in one model or the other. GMMs cluster data based on a normal distribution. The normal distribution is a distribution of data points where the mean is equal to the median and the median is equal to the mode. A one-dimensional normal distribution is commonly seen in statistics as the bell curve. The center of the distribution is the mean/median/mode and the data is distributed outwardly based on the variance. The GMM can be one-dimensional, two-dimensional, and more. The major difference is that when the GMM is one-dimensional, the variance determines how spread out the data is. When the model is two-dimensional or more, the co-variant matrix determines how spread out the data is. To produce these models, a random mean and variance is selected for the amount of models desired. Then the probability of each data point belonging to one of the data points or the other is calculated. Once all of these probabilities are calculated, a new mean and variance for each model is calculated. This process continues until either a certain amount of iterations are hit or the amount of change between the new and old means or variances is within a threshold. A GMM can contain multiple components, or Gaussian distributions. The mixture model is a combination of all of these components. In our case, each species had its own model.
3.2 Feature Extraction

The audio data and sampling rate of each frog call were extracted into the program and key features were selected to be used for training. The sampling rate and audio data were used to find the fingerprint of the data, the mapping of the frequencies and their energies. This fingerprint was a two-dimensional array. This array was used to create a spectrogram. A spectrogram is a visual representation of the audio sample with the y-axis containing the frequencies, the x-axis containing the time, and different colors represented the amounts of energies each frequency had. The spectrogram of the call of a Barking Tree Frog is shown in Figure 2.

![Figure 2: Spectrogram of a Barking Tree Frog Call](image)

We convolved the spectrograms with a subset of the bank of filters used for texture classification by Varma and Zisserman. The bank contains 48 filters with multiple scales and orientations. We used 12 of these filters for feature extraction. These were the twelfth through the seventeenth filters and the thirtieth through the thirty-fifth of these filters. These filters were selected because they included the most orientations. Figure 3 shows the full filter bank. The 12 filters we used for the feature extraction are enclosed by the red rectangles.

![Figure 3: The complete filter bank as in Varma and Zisserman](image)

Each of these filters were convolved with the spectrogram. After the convolution, a threshold was applied to the filtered spectrograms, this threshold removed the lower-energy peaks, leaving only the highest peaks of energy that have been filtered through. At each of these peaks, the frequency and the shape of the peak were
used as features. Figure 4 shows a spectrogram being filtered by one of these filters and then a threshold being applied to the same spectrogram. The remaining red peaks were used as features.

![Figure 4: Filtered spectrogram. From left to right: filter bank is convolved with the spectrogram of the audio signal. The result of the convolution is a map of filter responses (i.e., convolution coefficients). We threshold the filter response to keep only the locations with the highest energy.](image)

In addition to using the filter response values as features, we also used Mel-Frequency Cepstral Coefficients extracted at locations of prominent peaks in the spectrogram. Mel-Frequency Cepstral Coefficients, or MFCCs, are commonly used in speech recognition. MFCCs are designed to extract data that reflects what the human ear actually hear. This helps extract the most relevant features of the signal that define the sound. Using MFCCs extracts another important features of the sound. The sound is split up into time frames and at each time frame a certain amount of coefficients are extracted. For the purpose of this experiment, the only MFCCs that were used were ones extracted from time frames that contained a high peak of energy to assure that the MFCCs corresponded to key features of the frog call. Figure 5 shows the MFCCs for a Barking Tree Frog Call. The horizontal x-axis represents the time, the vertical y-axis represents the energy for each MFCC, and the z-axis represents the MFCCs.

![Figure 5: MFCCs for a Barking Tree Frog call.](image)

Once these MFCCs were found and the features from the filtered spectrograms were found, they were combined into one vector of features. This feature vector contained all of the data points used for training and classification. The diagram in Figure 6 shows the process of feature extraction for Gaussian Mixture Modeling.

### 3.3 Training and Classification

Once these features were extracted, the training process began. Each species would be fit to its own GMM. Figure 7 shows the training process for a Gaussian Mixture Model.

Then these models would be tested with a frog call that was not used in training. In order to do this the leave one out method was used. The leave one out method involves using every data point except one data point
to train the data with. In this specific experiment, one frog call out of every frog call recorded was taken out to be used as a test call. Then every other call was put in categories with the corresponding species and features were extracted. These features were fit to GMMs. Once this process was completed, features were extracted from the test call a test was run to determine to which model the test call was most likely to belong.

Each frog species was represented by a GMM. Data points are determined to belong to certain GMMs based on probability. The probability density function for a given species is:

$$p(x|c) = \sum_{k=1}^{K} r_k \mathcal{N}(x, \mu_k, \Sigma_k).$$

(1)

The $\mathcal{N}(x, \mu_k, \Sigma_k)$ represents the Gaussian density over the data $x$. The mean is represented by $\mu_k$ and the co-variance matrix is represented by $\Sigma_k$. $r_k$ represents the mixture weights, these are the proportions of the density for the different components in the mixture model. The posterior probability of a frog species is then
calculated using:

\[ p(c|x) = \frac{p(x|c)p(c)}{p(x)} \]
\[ = \frac{p(x|c)p(c)}{\sum_{c=1}^{C} p(x|c)p(c)}. \]  

Here, \( C \) represents the number of frog species, or classes. The species with the maximum posterior probability is assumed to be the correct species for that frog call, i.e.:

\[ \hat{c} = \arg \max_{c \in C} \{ p(c|x)p(c) \}. \]

For simplicity, we assume that each species is equally likely and set \( p(c) = 1 \). A diagram of the classification process is shown in Figure 8.

4. LOCALITY SENSITIVE HASHING

4.1 Definition

LSH as a process is an effective technique in grouping data as similar based on some relative similarities in high-dimensional spaces. Locality Sensitive Hashing (LSH) takes advantage of common principles associated with data structures and its theories, particularly, collision and bucketing. LSH has been an effective application to resolving problems plagued by the curse of dimensionality with its capabilities of exponential feature space reduction, where it is often implemented as a nearest neighbor search algorithm. One other common implementation of LSH is as a schema of feature manipulation and classification in a machine learning process. The exponential feature space reduction abilities of LSH translates a data that normally exists in high multidimensional spaces into solitary dimensional spaces, upon which simple and popular measures of relatability can be used in searching for similitude among data observations for classification.

A practical application of LSH is in musicology with music pattern recognition. Shazam Entertainment Limited is a pioneering company in the musicology industry that utilizes the process of LSH in querying a database of songs and videos with a sample audio or visual data to find a precise match. Briefly summarizing the process utilized, as described in, all audio files - both model and query files - are subject to similar feature extraction and labeling processes, as fingerprints. The generated fingerprints are used to measure correspondence between a large set of models and unknown queries to find a match and precisely label queries, following subsequent accuracy analysis. Additionally, highlights the importance of the fingerprints to be temporally localized, translation-invariant, robust, and highly entropic within satisfiable boundaries. The method of audio fingerprinting has also been implemented by a number of other musicology entities, such as SoundHound and Echoprint, to adapt their individual audio recognition software.

Keeping in mind the problem statement currently defined for this project, the implementation of LSH will be similar in principles to the schema adhered to by. The implementation of a classification technique, using LSH will fingerprint audio signal data for models and queries from frog calls and use similarities between signatures generated for distinct species to identify an unknown frog call query audio data signal as sourcing from a frog of a particular known species. The technique of fingerprinting and classification is broken down into three steps: Feature extraction, Feature definition, and Classification. Figure 9 summaries the classification method.
4.2 Feature Extraction

The extraction of feature from frog calls is the first stage in the automated identification process of frog calls. This stage involves the processing of an inputted frog call audio data into a signal from which a definitive feature vector, (or signature) is extracted and used in defining the observed data. The captured frog call audio data is read in as a wave file, represented as a relationship between amplitude \( A \) and time \( t \). The wave representation of the frog call audio data signal is put through a Short-Time Fourier Transform (STFT) to get spectral representation of the wave signal as a relationship between frequency \( f \), intensity \( I \), and time \( t \).

\[ R^2 = t, A \]

\[ R^3 = t, f, I \]

In this project, a spectrogram is used in visualizing the transformed spectral representation of a frog call audio data signal from its wave form in \( R^2 = t, A \) to its spectral form in \( R^3 = t, f, I \), as illustrated in Figure 10.
Mathematically, the representation is as shown in Eq. (4).

\[ \text{FFT}(w) : \{t, A\} \mapsto \{t, f, I\}. \] (4)

The spectral information from the spectrogram provide defining details about a particular sourced audio signal as features that are sufficiently discriminating. The defining details are depicted as regions of high intensities on the spectrogram and are key data points of interests, referred to as landmarks. Landmarks on a spectrogram are either trivial or non-trivial, as they are deduced by their reference to a locally defined window. As stepped through in the landmark detection algorithm Algorithm 1, within a defined landmark window \((\omega_\gamma)\), local maxima are deduced and threshed within satisfiable criterion. The \(\omega_\gamma\) shifts to the next \(\Delta t\) time frame for the detection process to be iterated, over the entire spectrogram representation of the frog call audio data signal. The set of detected landmarks are represented as \(\Gamma = \{\gamma_{1..N}|\gamma = (t, f)\}\).

\[ \text{Algorithm 1 Landmark Detection algorithm} \]

**Input:** Spectral data \(R^3 = t, f, I, \text{maximaThreshold} \)

**Output:** \(\Gamma = \{\gamma_{1..N}|\gamma = (t, f)\}\)

1. Initialize \(\{\Delta t, \Delta f\}\)
2. For each defined window, \(\omega_\gamma = (\Delta t, \Delta f)\) in timeFrame Do
3. \(\gamma_n = \text{LocMax}\{\{(t, f)\}_j\}\) that satisfy maximaThreshold
4. End For

Subsequent to detecting the collection of landmarks, the definition of these landmarks as a feature vector that describes an observation utilizes the hashing principles associated with LSH. In a similar fashion to, the hashing of the landmarks generates a signature set, \(\psi_n \ni \Psi | n = 1\ldots N\), that represents a feature vector to describe an observation data. The generation of the hash signatures for a data observation is defined by a form of relativity between a reference landmark and local pairing landmarks. This process begins by defining a reference window \((\omega_r)\) that contains a subset of landmarks. A pairing window \((\omega_p)\) is also defined to contain a subset of landmarks, such that \(\gamma_i \in \omega_r, \gamma_i \notin \omega_p, \gamma_j \in \omega_p, \) and \(\gamma_j \in \omega_r\). Each landmark contained in \(\omega_r\) will be a reference point \((\gamma_r)\) to pair with the set of pairing landmarks \((\gamma)\) in the \(\omega_p\). The relationship between the paired \(\omega\) will define the hashing for the pair as \(\psi_p = [f_r : \Delta f_p : \Delta t_p],\) as illustrated in Figure 11a.

![Figure 11: Hash generation for landmark pairs](image)

(a) Landmark pairing  
(b) Example of landmark pairing

As an example of a defined hashing as in Figure 11b, assume \((t, f)_r = (1.3, 1.6), (t, f)_p = (2.8, 3.5).\) The defined hash signature for the pairing as \(\psi_p = [f_r : \Delta f_p : \Delta t_p]\) will deduce \(\psi_p = [1.6 : 1.9 : 1.5]\).
4.3 Training and Classification

Thereafter, the hash signatures set that define the landmark pairs represents the feature vector as a description of an observed frog call audio data signal. This is to say that $\Psi$ describes an observation of frog call data. The defined feature vector is utilized in characterizing the correct label category (or specie) of a frog call, as stated in. In order to use $\Psi$ in the classification procedure of LSH, the processes as described above for landmark detection and definition is done for frog call audio data signal with known labels and unknown labels, similarly. The classification, as illustrated in Figure 13, models a labeled frog data data. The model is checked against an unlabeled frog call data for similarity in $\Psi$.

![Figure 12: LSH feature extraction schema](image)

![Figure 13: Classification with LSH - hash matching](image)
The classification process of LSH takes advantage of the hashing principles of collision and bucketing. Since hash values are to be unique to the data it describes, an occurrence of similar (or same) hash values for two different data causes leads to a collision. In an event of a collision, LSH uses bucketing to group together data that have similitude in their hash values, and labels all data in a bucket to be of the same source. With the $\Psi$ that describes an observed frog call data, bucketed $\Psi$ are labeled as observations that sources from the same frog species.

In addition to the feature defined from the landmarks extracted from the frog call audio data, an additional feature component is introduced into the feature vector. The added feature component represents the shape of landmarks, poised to give provide more discriminatory representation of observation data. The extension of the feature vector was proposed to improve the effectiveness of LSH in vocalization classification of frog species. In our approach, filters are used to describe the shape of landmarks. The filter bank used in our approach is the LM filter bank that is made up of 48 filters (6 orientations and 3 scales of first and second derivatives of Gaussian and 8 Laplacian of Gaussian filters). However, based on the expected shape and orientation of the spectral representation of a frog call sample, 12 filters highlighted in Figure 3 (under Section 3.2) were chosen to define the shape of detected landmark. The addition of landmark shapes as a feature component, caused a redefinition of the hashing of the landmark pairs as shown in Eq. (5):

$$\psi_p = [f_p : \Delta f_p : \Delta t_p : \eta_r]$$

5. EXPERIMENTS

We tested the Gaussian Mixture Modeling (GMM) and Locality Sensitive Hashing (LSH) methods on both synthetic and actual datasets. The synthetic data is modeled to simulate the actual data-set and is used for experimenting with the separate classification process to test the validity of the process. We adopted the leave-one-out training approach. The Leave-one-out approach defined the training data set with all observation data with known labels, but one that was left out. The left-out observation was the test sample. The actual dataset contains 226 audio samples of frog calls supplied by the Florida Tech’s Paleo Ecology Laboratory. The dataset contains calls from 15 frog species from wetland regions of Florida. A list of the frog species are shown in Table 1.

<table>
<thead>
<tr>
<th>Species Name</th>
<th>Common Name</th>
<th>Acronym</th>
<th>No. of Calls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lithobates catesbeianus</td>
<td>Bull Frog</td>
<td>BF</td>
<td>17</td>
</tr>
<tr>
<td>Hyla gratiosa</td>
<td>Barking Tree Frog</td>
<td>BTF</td>
<td>10</td>
</tr>
<tr>
<td>Pseudacris nigrita</td>
<td>Southern Chorus Frog</td>
<td>CF</td>
<td>11</td>
</tr>
<tr>
<td>Gastrophyne carolinensis</td>
<td>Eastern Narrow Mouth Toad</td>
<td>ENMT</td>
<td>9</td>
</tr>
<tr>
<td>Lithobates clamitans</td>
<td>Green Frog</td>
<td>GF</td>
<td>19</td>
</tr>
<tr>
<td>Hyla cinera</td>
<td>Green Tree Frog</td>
<td>GTF</td>
<td>23</td>
</tr>
<tr>
<td>Pseudacris ocularis</td>
<td>Little Grass Frog</td>
<td>LGF</td>
<td>12</td>
</tr>
<tr>
<td>Anaxyrus quercicus</td>
<td>Oak Toad</td>
<td>OT</td>
<td>7</td>
</tr>
<tr>
<td>Rana grylio</td>
<td>Pig Frog</td>
<td>PF</td>
<td>18</td>
</tr>
<tr>
<td>Hyla femoralis</td>
<td>Pine Woods Tree Frog</td>
<td>PWTF</td>
<td>9</td>
</tr>
<tr>
<td>Acris Gryllus</td>
<td>Southern Cricket Frog</td>
<td>SCF</td>
<td>21</td>
</tr>
<tr>
<td>Hyla squirella</td>
<td>Squirrel Tree Frog</td>
<td>SF</td>
<td>9</td>
</tr>
<tr>
<td>Lithobates phenocephalus</td>
<td>Southern Leopard Frog</td>
<td>SLF</td>
<td>16</td>
</tr>
<tr>
<td>Pseudacris crucifer</td>
<td>Spring Peeper</td>
<td>SP</td>
<td>23</td>
</tr>
<tr>
<td>Anaxyrus terrestris</td>
<td>Southern Toad</td>
<td>ST</td>
<td>22</td>
</tr>
</tbody>
</table>

5.1 GMM Experiments

Before we applied the Gaussian mixture model to a data set of real frog calls, synthetic data was used. Synthetic data was used first because it is easier to control the outcome of the results. This synthetic data was generated
using a random Gaussian data generator function in Python. We grouped this data together and labeled these groups as different species. A Gaussian mixture model was fit for each species. The scatter plots of these species are shown below.

![Figure 14: Synthetic Species Data](image)

(a) Species 1  
(b) Species 2  
(c) Species 3  
(d) Species 4

In order to test this method, we used two different sets of test data. The first set of test data was the same exact data used to create 2 species. The second set of test data was the same data used to create species 4, but more test data points were introduced to replicate background noise.

![Figure 15: Synthetic Test Data](image)

(a) Test 1  
(b) Test 2

For both tests, the outcomes were as expected. The first test data tested to be from species 2 and the second test data tested to be from species 4. After the synthetic data tested well, the Gaussian Mixture Model was applied to the data set of 226 calls across 15 species. The algorithm that was originally used to test all of these calls is shown below. In order to speed up the experimentation process for Gaussian mixture models, a model was generated for each species before any testing began. Then a loop was created to go through all of the frog calls and pull out one at a time to act as the testing call. The species that the testing call belonged to was re-trained to fit to a new Gaussian mixture model without including the testing call. This was done to insure that the test call was completely left out of the training process. After this is complete, features are extracted from the test call. The posterior probability of the test call belonging to any of the species is calculated and the species corresponding to the greatest posterior probability is said to be the correct species.
5.2 LSH Experiments

Similarly, the approach generates a mode for each observed frog call audio data signal. The One specie-Many models will give a count of models equivalent to the count of observed frog call data with known labels. This implies the classification of new data will require the number of match comparisons equal to the count of models available. The unlabeled observation will be bucketed with the model with an acceptable feature similitude. In non-ideal circumstances, a lack of found match will signify the non-existence of an effectively descriptive model that represents the likely label group of the unmatched frog call data.

The second of the two implementation approaches for Locality Sensitive Hashing for classifying frog species from frog call audio data is the One specie-One model approach. This approach, unlike the former generates a single model that is descriptive for an entire label group. Hereafter, this approach follows the same structure as the One specie-many model approach.

5.3 Experimental Results

The confusion matrix shows the probability (P(C)) of a test call belonging to a label group. In an ideal scenario, the highest probability will be along the diagonal in the darkest shade of boxes. Below are the confusion matrices for both GMM and LSH classification techniques.

Figure 16: Confusion Matrix of Classification Results with LSH
GMM was tested using 13, 64, and 128 MFCCs. The table below shows the results for each. The classification results with the 128 MFCCs ended up showing the highest accuracy percentage.

Table 2: GMM Results

<table>
<thead>
<tr>
<th>Number of MFCCs</th>
<th>Accuracy Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>79.66 %</td>
</tr>
<tr>
<td>64</td>
<td>69.38 %</td>
</tr>
<tr>
<td>13</td>
<td>66.22 %</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

The use of Gaussian Mixture Model and Locality Sensitive Hashing as classification techniques to identify frog calls is novel. GMM has been used in the past for speaker recognition. LSH has been used in the past as a classification method for song recognition. Both of these methods have performed adequately effective in these domains. This is why we decided to try them out for frog call classification. We found, according to our results, that GMM seems to be the more accurate method of classification for frog calls. However, we would like to look at other ways to increase the accuracy of both methods in the future. One of these ways would be modifying the filter component of the feature extraction. Instead of using the Leung-Malik filter bank, we would like to look into creating our own filters to improve the quality of the discrimination of the filter responses on the spectrograms. We would also like to look into increasing the amount of Mel-Frequency Cepstral Coefficients (MFCCs) per time frame. The only dilemma is we would like to make sure the run time of the program does not become too long. We would also like to try to experiment with these methods on a more uniform data set. Some of the species had more frog calls than others. Also, many of the frog calls differed in duration. Some were almost two minutes long, while others were only one second long. If we are able to organize a data set that includes frog calls of similar length with species containing the same amount of frog calls each, we believe this would increase the accuracy percentage for both methods.
This has brought us a step closer to the overall goal of innovating a more efficient and effective method of data gathering of frog populations. The use of readily available devices without the need of extensive training would allow almost anyone to contribute to the process of data gathering and environmental conservation. The use of frog populations as a bio-indicator has been extensively researched and stressed upon by renowned biological and environmental scientists. This has lead to many different programs to train citizen scientists to recognize frog species based on their calls which would lead to some subjective classification. This also requires the citizen scientists to be trained, have the correct equipment, and be able to commit a sizable amount of time. The development of a mobile application to automatically identify frog species based on their call is going to open many doors. This will allow so many other people to get involved. This would require no extra training, no extra equipment, and a much smaller time commitment. This would also spread the word about environmental monitoring very quickly because the application will be connected to social media. This application will also impact education. Teachers may use this application to teach their students about environmental conservation and respect. Since many of the users of smart phones and social media are in the younger generations, this would inspire citizen scientists for years to come.

ACKNOWLEDGMENTS

The authors acknowledge support from National Science Foundation (NSF) grant No. 1263011. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

The authors would also like to thank Dr. Eraldo Ribeiro who served as a faculty mentor and Dalwinderjeet Kular who served as a graduate mentor. Further acknowledgments would be expressed to the Advances of Machine Learning in Theory and Application (AMALTHEA) Research Experiences for Undergraduates (REU) program coordinators, Dr. Georgios Anagnostopoulos and Dr. Adrian Peter.

The authors would also like to extend their humblest appreciation to the Florida Institute of Technology as a host institution and the College of Engineering for the use of campus and laboratory facilities.

REFERENCES


