Pertanika J. Sci. & Technol. 29 (4): 2383 - 2407 (2021)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

A Novel Approach of Audio Based Feature Optimisation for Bird Classification

Murugaiya Ramashini*, Pg Emeroylariffion Abas and Liyanage C De Silva

Faculty of Integrated Technologies, Universiti Brunei Darussalam, Jalan Tungku Link, BE1410, Brunei Darussalam

ABSTRACT

Bird classification using audio data can be beneficial in assisting ornithologists, bird watchers and environmentalists. However, due to the complex environment in the jungles, it is difficult to identify birds by visual inspection. Hence, identification via acoustical means may be a better option in such an environment. This study aims to classify endemic Bornean birds using their sounds. Thirty-five (35) acoustic features have been extracted from the pre-recorded soundtracks of birds. In this paper, a novel approach for selecting an optimum number of features using Linear Discriminant Analysis (LDA) has been proposed to give better classification accuracy. It is found that using a Nearest Centroid (NC) technique with LDA produces the optimum classification results of bird sounds at 96.7% accuracy with reduced computational power. The low computational complexity is an added advantage for handheld portable devices with minimal computing power, which can be used in birdwatching expeditions. Comparison results have been provided with and without LDA using NC and Artificial Neural Network (ANN) classifiers. It has been demonstrated that both classifiers with LDA outperform those without LDA. Maximum accuracies for both NC and ANN with LDA, with NC and the ANN classifiers requiring 7 and 10 LDAs to achieve the optimum accuracy, respectively, are 96.7%. However, ANN classifier with LDA is more computationally complex. Hence, this is significant as the simpler NC classifier with LDA, which does not require expensive processing power, may be used on the portable and affordable device for bird classification purposes.

Keywords: Artificial neural network (ANN), bird sounds, classification, linear discriminant analysis (LDA), nearest centroid (NC)

ARTICLE INFO

Article history:
Received: 3 February 2021
Accepted: 24 May 2021
Published: 28 October 2021

DOI: https://doi.org/10.47836/pjst.29.4.08

E-mail addresses:
ramashini@uwu.ac.lk (Murugaiya Ramashini)
emeroylariffion.abas@ubd.edu.bn (Pg Emeroylariffion Abas)
liyanage.silva@ubd.edu.bn (Liyanage C De Silva)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Birds have a considerable impact on our lives (Suthers, 2004). In the past, people had used bird sounds in their day-to-day life for stress recovery and as an attention restoration tool (Alvarsson et al., 2010). In addition, bird sounds were also used as alarms for severe weather changes and hazardous conditions (Vilches et al., 2006). They extracted this useful information from the birds' unique features and behaviours and their sounds due to their deep understanding and knowledge of birds and the environment, obtained through traditional knowledge and experiences from observing nature. However, nowadays, this knowledge is fast deteriorating and limited to only a few people due to our busy modern lifestyle and disengagement from nature.

Birdwatching, which includes identifying birds visually or by their sounds, has become a rapidly popular recreational activity. This activity positively impact the economy and the environment, especially for countries dependent on ecotourism. Tourists who visit destinations to meet their particular needs and share specific interests and motivation are called "niche tourists", and birdwatchers fall into such category. Birdwatchers may travel to specific destinations only for bird-watching purposes. The destination promotes the specific activity as one of its significant niche tourism activities (Butler, 2019). However, growing environmental changes, including potential timing mismatch for breeding, increase in ocean temperature, and the unavailability of sufficient food supply, have disrupted the bird population. As a result, it may affect the long-term sustainability of birdwatching and ecotourism. The future of ecotourism depends on wildlife tourism, such as birdwatching. It relies on the maintenance and well-being of the endemic species to draw tourists to the destinations (Kutzner, 2019). Lately, the focus has been given to surveillance and environmental monitoring related applications, with the rapid development of technologies such as artificial intelligence (Badi et al., 2019). Therefore, it is essential to find novel approaches that may strengthen the nature-based tour operators' resilience in the tourism industry and address the rapidly changing social and environmental conditions (Kutzner, 2019), It is also vital to have new technologies that can support visitors.

However, visual bird identification can be a difficult task, especially in a densely vegetated rainforest environment. Therefore, bird species identification based on sound may be a better option (Trifa et al., 2008). Consequently, audio-based bird classification has gotten the limelight in recent years.

Statistics can be computed over the audio bird-sound datasets to generate a single feature vector (Giannakopoulos & Pikrakis, 2014) that can identify the bird species. Then the relevant features extracted from sound are identified and grouped into a set of classes that it most likely fit. Depending on the application, different grouping algorithms, and feature extraction techniques may be used (Gerhard, 2003) with a wide range of supervised and unsupervised Machine Learning (ML) algorithms used for bird species identification.

For feature extraction, researchers have used time, frequency and also few cepstral domain features. Sharma et al. (2020) summarise the literature on audio signal processing for bird sound classification tasks, mainly focusing on feature extraction techniques.

Many researchers have used Artificial Neural Networks (ANNs) in their work. For example, Selouani et al. (2005) use Multi-Layer Perceptron (MLP) feedback loop to improve the architecture, using a set of selected features as input to produce different output for each species. Variations of ANN exist in the literature and have been used by many researchers to detect bird sounds (Ranjard & Ross, 2008; McIlraith & Card, 1997). Chou et al. (2008) use decision-based Neural Network (NN) to improve the accuracy of detection as well as processing time consumed by the model (Selouani et al., 2005). Probabilistic, backpropagation and Kohonen NNs have also been demonstrated by Terry and McGregor (2002). Priyadarshani et al. (2018) elaborate state of the art in bird recognition and describe the different techniques adopted over the years in their review article.

Despite the availability of many technologies, including audio signal processing and pattern recognition that have been used to study birds and their sounds, there are still plenty of research gaps in the identification of birds from their sounds due to the vast range and heritage distribution of bird species. Notably, the Borneo region is rich in biodiversity with unique and diverse animal life varieties, including many birds living in its dense and virgin tropical rainforest areas. Furthermore, although ecotourism is becoming a growing source of income for countries in the Borneo region, with bird watching as one of the main features of ecotourism in the area, there is no application using technology to assist visitors to this region.

This paper proposes a simple dimension-reduction technique, which can select the optimum feature combinations for dimension reduction purposes. When combined with two classification methods, Nearest Centroid (NC) and Artificial Neural Network (ANN), it can efficiently and effectively classify birds from their audio sounds. The proposed method has been demonstrated for the classification of 10 endemic bird species of the Borneo region. It has been shown that it can accurately identify these bird species, with a low requirement on computational power. Moreover, this is very significant, as currently, most researchers have utilised advanced and complex techniques, which require high computational power to classify bird sounds. While this is feasible for non-real-time applications with access to high-end equipment, real-time implementation of such techniques on simple portable devices has been proven very difficult.

Consequently, the proposed method may implement hardware solutions for real-time bird sound classification to assist bird watchers. The following section discusses the proposed methodology for the classification of bird sounds, composed of data collection, pre-processing, segmentation, feature extraction, dimensionality reduction and classification, followed by results and discussions and finally, the performance of the proposed method. The final section concludes the paper.

METHODOLOGY

It is necessary to pre-process the bird sound to extract essential properties as inputs to the classification model to classify a given unidentified bird sound according to its species. Pre-processing may involve passing the birds sound through a filter to remove unwanted noise and disturbance, segmenting the bird's sound into distinct parts and extracting important features from the bird sounds. Then, depending on the method adopted, selected bird sounds features may be fed directly onto the already-trained classification model to give the predicted species of the un-identified bird sound.

Figure 1 depicts a simplified process used for bird sound classification in this paper, categorised into the training of the classification model (training phase) and testing using the trained model (testing phase). Initially, the classification model is trained using a database of labelled bird sounds. Intuitively, the performance of the trained classification model in predicting unknown bird sounds shall depend on the quality of the training data, feature selections, as well as the classification model adopted. The collection of rich but reliable labelled bird sounds is a critical first step in classifying birds. Next, the labelled bird sounds are processed and used to provide selected features to train the chosen classification model. Generally, large extracted features would give better classification performance, albeit with extra computational complexity. In this regard, the dimensionality reduction technique shall be used to select combinations of the best features. Two classification models shall be adopted: Nearest Centroid (NC) and Artificial Neural Network (ANN) classifiers.

Data Collection, Pre-processing and Segmentation

The quality and quantity of bird sounds are essential to allow proper training of the classification model. However, biogenic, anthropogenic, and wind noise may influence the quality of the recordings such that bird sounds may be inaudible or only audible for a short period (Giannakopoulos & Pikrakis, 2014). Furthermore, bird sounds typically contain combinations of songs and calls, which may need to be pre-processed to obtain sections that are used for feature extraction and classification purposes. In this regard, signal pre-processing plays an important role.

Pre-filtered bird sounds are used for both training and testing phases. In the case of noisy data, researchers commonly use either low, high or bandpass filters, depending on the bird species being considered (Vilches et al., 2006), to selectively reduce noise level whilst preserving the quality of the intended bird sounds. Then, the uninterrupted bird sounds may need to be segmented into segments of homogeneous content (Giannakopoulos & Pikrakis, 2014) using quasi-periodic syllables of bird sounds. This process can be done manually (Trifa et al., 2008; Lee et al., 2008; Anderson et al., 1996) or automatically, depending on the applications. Automatic segmentation is generally preferred for real-time applications. Different methods may be adopted for segmentation, by taking advantage of the energy in the time or frequency domain (Evangelista et al., 2015) and analysing autocorrelation and roll-off of the songs (Ranjard & Ross, 2008).

Whilst manual segmentation may be used for non-real-time applications, automatic segmentation of bird sounds is adopted in this paper by implementing an energy envelop based algorithm in the time domain, and removing unwanted silent periods, to give samples of bird sounds. Many researchers have previously used this iterative time-domain algorithm (Fagerlund & Laine, 2014; Fagerlund, 2007; Härmä et al., 2004). Segmentation is performed on the training the birds' sound dataset and testing bird sound dataset.

Feature Extraction

Most nature-related data, including bird sounds, are extensive and contain much redundancy. After segmentation of bird sounds into quasi-periodic syllables, the data need to be processed further before it can be made as input onto the classification models to remove as much redundant and irrelevant information as possible whilst retaining important properties to allow efficient classification of the data. This stage, commonly referred to as feature extraction, may involve the extraction of physical or perceptual features based on measurable and reported characteristics perceived by humans (Gerhard, 2003). The same features must be extracted from every dataset to allow like-for-like comparison between different datasets to facilitate classification. Generally, features may be extracted from the time-domain representation of the data or its corresponding frequency domain representation. Obviously, for frequency-domain features, syllables of the original time-based bird sounds need to be first converted into their frequency domain representation before their features are extracted. Features may also be obtained from different syllables of the same dataset.

Each sample of the birds sound is divided into overlapping frames to perform feature extraction. M different features are derived from each frame, consisting of time, frequency and quasi-periodic features. The time-domain features include Zero Crossing Rate (ZCR), Energy (E), and Entropy of energy, whilst Spectral Centroid, Spectral Spread, Spectral Entropy, Spectral Flux, Spectral Roll-off, Mel Frequency Cepstral Coefficients (MFCCs), and Chroma Vectors form the frequency domain features from each frame (Sharma et al., 2020). Apart from these features, Harmonic Ratio and Fundamental Period are also extracted (Giannakopoulos & Pikrakis, 2014).

A total of M distinct features are derived from each frame, which is then averaged over the length of the sample. Each sample belonging to one of the L species of bird considered. For the i^{th} sample of the training data, its features are represented in an M dimensional space, in a row vector x_i, where $x_i \in R^M$; with each sample classified as one of the L species of bird under consideration i.e. $c_i \in \{w_1, w_2, \ldots\ldots w_L\}$. Consequently, for training data with N samples, feature matrix $\mathrm{X} = [x_1, x_2, \ldots, x_N]$, where $X \in R^{N\times M}$ and class column

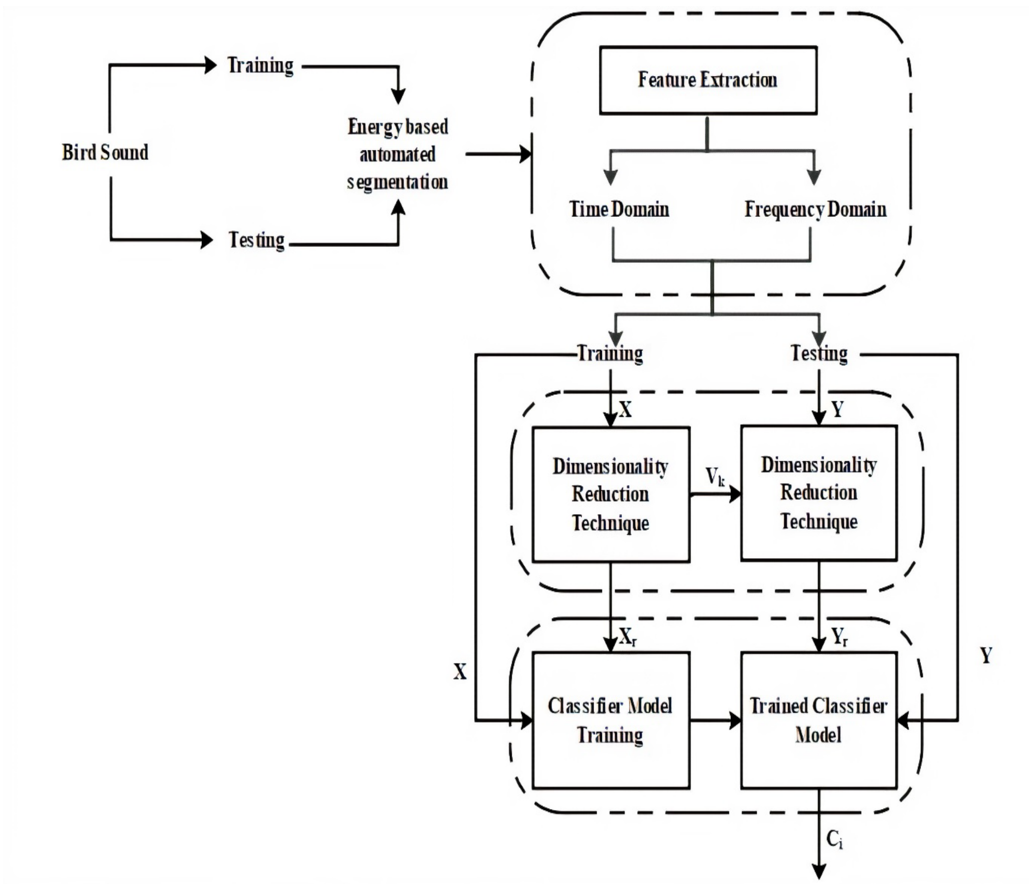


Figure 1. Proposed Methodology for Bird Sound Classification

vector $c \in R^N$ are used to populate feature information and species classification of all *N* segmented training samples, respectively.

Similar to the samples obtained from the training data, *M* features are also extracted from i^{th} test sample. This information is stored in an *M* dimensional space, in a row vector y_i, where $y_i \in R^M$. Each sample of the test data belongs to either one of the *L* species of birds under consideration. The classification model shall be performed predictions on species classification of the test sample, which has been previously pre-trained using the training data.

Dimensionality Reduction

For each training and testing sample, a feature row vector in an *M* dimensional space is obtained to describe the sample; forming feature matrix $X = [x_1, x_2, \ldots, x_N]$, where

$X \in R^{N\times M}$ for the training dataset and $y_i \in R^M$ for each of the testing sample *i*. Of course, these *M* features from a single sample may be directly fed to the classification model of choice for either training or testing. However, not all of these *M* features are applicable, or may even contribute to the classification process in differentiating between the different species of birds. Nevertheless, the *M* features, which may contain plenty of redundant information, would surely increase computational complexity significantly.

An option would be to selectively truncate the number of features from *M* features to *K* features to reduce computational complexity of the classification process. For this purpose, numerous dimensionality reduction techniques (Tharwat et al., 2017) exist in the literature. These techniques may be generally categorised into unsupervised and supervised approaches. Commonly, the training dataset is used to determine the best set of *K* features from the original *M* features to feed into the classification process, Each of the *K* features is composed of a weighted combination of the original *M* features. This weighted combination derived from the training dataset shall then be used to determine the *K* features from the testing dataset.

Species classifications of the bird i.e. class column vector $\mathrm{c} \in R^N$, are not taken into consideration in unsupervised dimensionality reduction processes, with information from the feature matrix $\mathrm{X} = [x_1, x_2, \dots, x_N]$ only used to assist in choosing the best *K* features. Common unsupervised dimensionality reduction methods are Independent Component Analysis (Mogi & Kasai, 2013), Non-negative matrix factorisation (Ranjard & Ross, 2008; Ludeña-Choez et al., 2017) and Principal Component Analysis (PCA) (Lee et al., 2008; Milani et al., 2019). PCA is one of the most popular and widely used unsupervised dimensionality reduction method (Tan et al., 2012). It aims to project the original *M* dimensional feature matrix onto alternative orthogonal *M* dimensional space, by considering linear combinations of the *M* dimensional feature matrix with an objective of finding the alternative space, which gives the largest variance. Dimensionality reduction is achieved by selecting a reduced subset of *K* dimension, which accounts for as much variability to give a reduced dimension feature matrix $X_r \in R^{N\times K}$. Commonly, a projection matrix $V_k \in R^{M\times K}$ is obtained from PCA, which projects the original feature matrix $X_r \in R^{N\times K}$ onto the reduced dimension feature matrix $X_r \in R^{N\times K}$. The projection matrix may then be used to reduce the dimension of the testing data *i*; from row vector y_i, where $y_i \in R^M$ to row vector $y_{r_i} \in R^K$.

On the other hand, supervised approach considers both features of the birds, i.e. the feature matrix $\mathrm{X} = [x_1, x_2, \dots, x_N]$, as well as the species classifications, i.e. class column vector $\mathrm{c} \in R^N$, to obtain a reduce set of *K* features from the original *M* features. Examples of supervised approaches include Neural Networks (NN), Mixture Discriminant Analysis (MDA) and Linear Discriminant Analysis (LDA) (Tharwat et al., 2017). Due to consideration of this extra information, the supervised approach can perform better in applications that require the prediction of classes, such as bird species classification. Indeed, it has been shown that LDA outperforms PCA (Martinez & Kak, 2001), particularly in cases where the number of samples per class is small. Moreover, LDA works by selecting reduced dimensions, which accounts for as much variability across different classes, as possible, instead of across all samples, as used in PCA.

LDA transforms the original feature matrix $X \in R^{N\times M}$ of the training bird sounds, which lies in a *M* dimensional space, onto a reduced matrix $X_r \in R^{N\times K}$, which lies in a *K* dimensional space, where $K \leq M$. It is done by considering the classification *c* of the segmented bird sounds of the training data. Transformation using LDA is achieved via a two-step process: finding the suitable transformation matrix to achieved maximum class separability and selecting $K \leq M$ dimensions that best discriminate between the different classes. The first step translates into an optimisation problem to find the transformation matrix *W* which maximises the ratio of the between-class variance S_B to the within-class variance S_W, of the feature matrix *X*. Mathematically, this may be represented as the Fisher's criterion, as Equations 1 and 2

$$\underset{W}{argmax}\frac{W^T S_B W}{W^T S_W W} \quad [1]$$

where

$$W = S_W^{-1}.S_B \quad [2]$$

The second step in LDA is to select the most significant dimensions of the matrix. This may be found by first finding the eigenvalues $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_M\}$ and eigenvectors $V = \{v_1, v_2, \ldots, v_M\}$ of the transformation matrix *W* , using Equation 3

$$(W - \lambda_i).v_i = 0 \;\; \text{for } i = 1,2,\ldots,M \quad [3]$$

The eigenvectors with the *K* highest eigenvalues are then chosen to construct the projection matrix $V_k \in R^{M\times K}$ to project the original feature matrix $X \in R^{N\times M}$ of the training bird sounds, onto a reduced dimension feature matrix $X_r \in R^{N\times K}$using Equation 4

$$X_r = X.V_k \quad [4]$$

The same projection matrix V_k, obtained from Equation 4, is also used to reduce the dimension of the i^{th} test bird sound $y_i \in R^M$ onto $y_{r_i} \in R^K$ using Equation 5

$$y_{r_i} = y_i.V_k \quad [5]$$

It is noted that each element of $y_{r_i} \in R^K$ is a linear combinations of the original *M* features of the i^{th} test bird sound $y_i \in R^M$. Rather than directly reducing the number of features LDA allows a reduced number of combinations of features, as inputs onto the classification model.

Due to the potentially superior performance of the supervised approach, particularly LDA, compared to the unsupervised approach, LDA shall be considered the dimensionality reduction technique of choice in this paper. Different values of $K \leq M$ shall be chosen to be fed onto the classification process for training and testing, using two different classification models. Performance of the classification models using different values of *K* shall then be compared; in terms of accuracy in predicting unknown bird sound and computational complexity. It is noted that $K = M$ represents directly feeding the classification models with all the original *M* features without performing any dimensionality reduction, which forms the basis for comparisons.

It is highlighted that the training process is normally performed non-real-time. Hence, it has the luxury of training time required and using a processor with high processing power. On the other hand, the classification of bird sounds during the testing phase requires real-time processing with limited computing power. For this reason, developers usually are more concerned with the computational complexity during the testing phase.

Using LDA for the dimensionality reduction approach requires extra computation. However, the reduced dimensions being fed onto the classification process can reduce computation process during the classification. As can be seen from the dimension reduction process in Equation 5, each element of the vector $y_{r_i} \in R^K$ is composed of *M* multiplication and $(M-1)$ addition operations. Since LDA reduces the dimensions to *K*, computational complexity of LDA during the testing phase is derived as $(M.O(n^2) + (M-1).O(n))$, where $O(n)$ is taken as the computational complexity for addition/subtraction operation and $O(n^2)$ as the computational complexities for multiplication/division operation, of an Arithmetic Logic Unit (ALU). Taking the features as type float i.e. $O(n) = O(n^2) = O(1)$, computational complexity can be simplified as $(2M-1).K.O(1)$. On the other hand, no extra computation is required for classification using all *M* features, i.e. without dimensionality reduction.

Classification

Generally, any machine learning technique aims to find the best function or mathematical model that may be used to classify bird species based on features of an unknown bird species. Training data is commonly used to derive this mathematical model. This phase is commonly referred to as the training phase. $K \leq M$ features from the training dataset, $K = M$ with representing using the classification models without any reduction to the dimensions of the original feature matrix may be used to derive the mathematical model. This mathematical model may then be used to classify unknown bird species in the inference phase, using similar $K \leq M$ features.

Two classification methods shall be considered in this paper: Nearest Centroids (NC), and Artificial Neural Network (ANN). Nearest Centroids (NC) are chosen due to their simplicity of implementation, requiring only the computation of centroids of each class from the training dataset, with classification decisions based on the nearest distance of the unknown bird species centroids of the different classes. On the other hand, Artificial Neural Network (ANN) is a basic neural network. ANN uses the training dataset to determine an appropriate mathematical model using the concept of neurons.

Nearest Centroid (NC). NC is one of the simplest supervised classification methods. Class prediction of new unknown bird sound is assigned to the class of the centroids closest to the new unknown bird sound. At the initial stage, $K \leq M$ features from the training dataset are used to determine centroids or means for the different classes of birds. Given the feature matrix $\boldsymbol{X_r} \in \mathcal{R}^{N \times K}$ where $K \leq M$ with species classification column vector $\mathbf{c} = \{c_1, c_2, \dots, c_L\}$ used for training of the classifier model, centroid for species $w_j, j \in \{1,2, \dots, L\}$, is represented by column vector $\boldsymbol{\mu_j} \in \mathcal{R}^K$ in the *K* dimensional space and is given by Equation 6:

$$\boldsymbol{\mu_j} = \frac{1}{|w_j|} \sum_{i \in w_j} \boldsymbol{x_i} \quad \forall j \in \{1,2, \dots, L\} \qquad [6]$$

where $|\boldsymbol{w_j}|$ is the number of training bird sounds belonging to species class W_j. Centroids of the different species classification are then used for predicting unknown bird sounds. Feature vector $\boldsymbol{y_{r_i}} \in \boldsymbol{\mathcal{R}^K}$ of the i^{th} test bird sound is used to predict its species classification $\boldsymbol{\hat{c}_i}$, using Euclidian's distance calculation (Ramashini et al., 2019) as follows (Equation 7):

$$\hat{c}_{\boldsymbol{i}} = \underset{j \in \{1,2,..,L\}}{\arg\min} \left\| \boldsymbol{\mu_j} - \boldsymbol{y_{r_i}} \right\| \qquad \hat{c}_i \in \{c_1, c_2, \dots, c_L\} \qquad [7]$$

Although the training dataset is commonly very bulky, which is processed simultaneously during the training phase, it has the luxury of using a high-computing facility for processing due to its non-real-time nature. For this reason, the testing phase is more of a concern. As can be seen from Equation 7, NC involves the *K* subtraction operations to determine the distance of the i^{th} test bird sound to every centroids. Consequently, computational complexity of NC during the testing phase can be derived as $(L.K + 1)O(n) = (L.K + 1)O(1)$, for features of type float. It can be seen that complexity is proportional to the number of features *K*.

Artificial Neural Network (ANN). Alternative classification considered in this paper is ANN, which is one of the basic and prevalent supervised machine learning techniques that may be used for bird species classification. Figure 2 depicts a given layer of an ANN considered in this paper. $z^k(i)$ is the output of neuron *i* before the activation function in layer *k*, $a^k(i)$ is the output of neuron *i* after the activation function, n^k is the set of neurons in layer *k*, and $b^k(i)$ is the bias of neuron *i* in layer *k*. $\theta^k(i,j)$ is the weight between the output of neuron *j* in layer *k-1* and neuron *i* in layer *k*. $g^k\left(z^k(i)\right), \forall i \in n^k$ is the activation function of every neuron in layer *k* of the ANN, and is used to provide non-linearity to the network (Equations 8 and 9),

$$z^k(i) = \sum_{j \in n^{k-1}} \left(\theta^k(i,j).a^{k-1}(j)\right) + b^k(i) \quad [8]$$

$$a^k(i) = g^k\left(z^k(i)\right) \quad [9]$$

These can alternatively represented in matrix forms (Equations 10 and 11),

$$z^k = \theta^k.A^{k-1} + b^k \quad [10]$$

$$A^k = g\left(z^k\right) \quad [11]$$

Where (Equations 12-18),

$$z^k = \left[z^k(1), z^k(2), \dots., z^k\left(\left|n^k\right|\right)\right]^T \quad [12]$$

$$\theta^k = \begin{bmatrix} \theta^k(1,1) & \cdots & \theta^k\left(1,\left|n^{k-1}\right|\right) \\ \vdots & \ddots & \vdots \\ \theta^k\left(\left|n^k\right|,1\right) & \cdots & \theta^k\left(\left|n^k\right|,\left|n^{k-1}\right|\right) \end{bmatrix} \quad [13]$$

$$\theta^k = [\theta^k(i,j)]\forall\left(i \in n^k, j \in n^{k-1}\right) \quad [14]$$

$$A^k = \begin{bmatrix} a^k(1,1) & \cdots & a^k\left(1,\left|n^k\right|\right) \\ \vdots & \ddots & \vdots \\ \theta a^k\left(\left|n^{k-1}\right|,1\right) & \cdots & a^k\left(\left|n^{k-1}\right|,\left|n^k\right|\right) \end{bmatrix} \quad [15]$$

$$A^k = [a^k(i,j)]\forall\left(i \in n^{k-1}, j \in n^k\right) \quad [16]$$

$$b^k = \left[b^k(1), b^k(2), \dots., b^k\left(\left|n^k\right|\right)\right]^T \quad [17]$$

$$g = [g^1(.), g^2(.), \dots., g^k(.)]^T \quad [18]$$

In this paper, ANN with $K \leq M$ features in the input layer, 1 hidden layer and 1 output layer is considered, as depicted in Figure 2. The number of neurons in the output layer corresponds to the number of bird species considered. Elliot Sigmoid function is used as the input layer's activation function (Elliott, 1993), which closely approximates the Hyperbolic Tangent or Sigmoid functions for small values. Whilst Softmax function is used as the output activation function in order to represent the probability distributions of a list of potential outcomes.

During the training phase of the ANN classifier, the objective is to find a set of weights θ^k so that the ANN can classify bird species accurately, using the training dataset as its basis. The ANN, together with this set of weights θ^k, then form the mathematical model used to classify unknown bird sound species during the testing phase.

Similar to NC, the testing phase of ANN represents a concern in terms of computational complexity, as it needs to be processed in real-time. The overall complexity is composed of the input-hidden and hidden-input layers, obtained using Equations 8 and 9, respectively, depending on the number of neurons. As there are K neurons at the input of the ANN, complexity of the input-hidden layer may be approximated as $KO(n^2) + KO(n))n^h + n^h O(g_h)$ where n^h is the number of neurons in the hidden layer and $O(g_h)$ is the complexity of the activation function in the hidden layer (depending on the function used). At the hidden-output layer, L represents the number of neurons at the output of the ANN. As such, complexity may be approximated as $\left(n^h O(n^2) + n^h O(n)\right) L + L.O(g_o)$, where $O(g_o)$ is the complexity of the activation function in the output layer. Overall, complexity during the testing phase of the ANN is given by $(KO(n^2) + KO(n))n^h + n^h O(g_h) + \left(n^h O(n^2) + n^h O(n)\right) L + L.O(g_o)$. Again, taking the features as type float, complexity can be simplified as $(2K + 2L)n^h O(1) + n^h O(g_h) + L.O(g_o)$. It can be seen that the complexity is also proportional to the number of features K that are being fed to the classification model as well as the number of neurons n^h in the hidden layer. Thus, there is no fixed relationship. Generally, the higher the number of features, i.e. the number of neurons in the input layer, the higher the number of neurons required in the hidden layer to achieve reasonable classification accuracy.

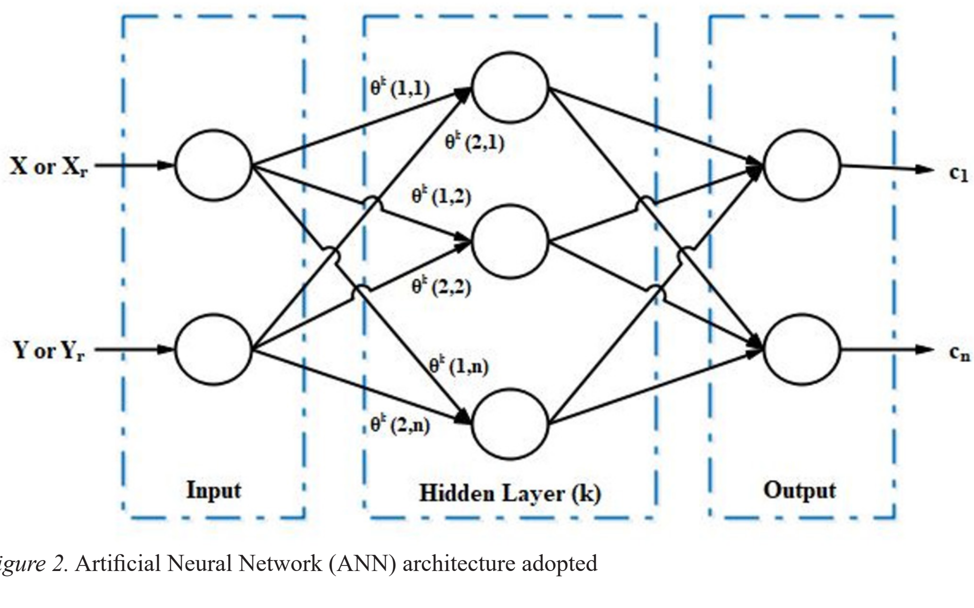
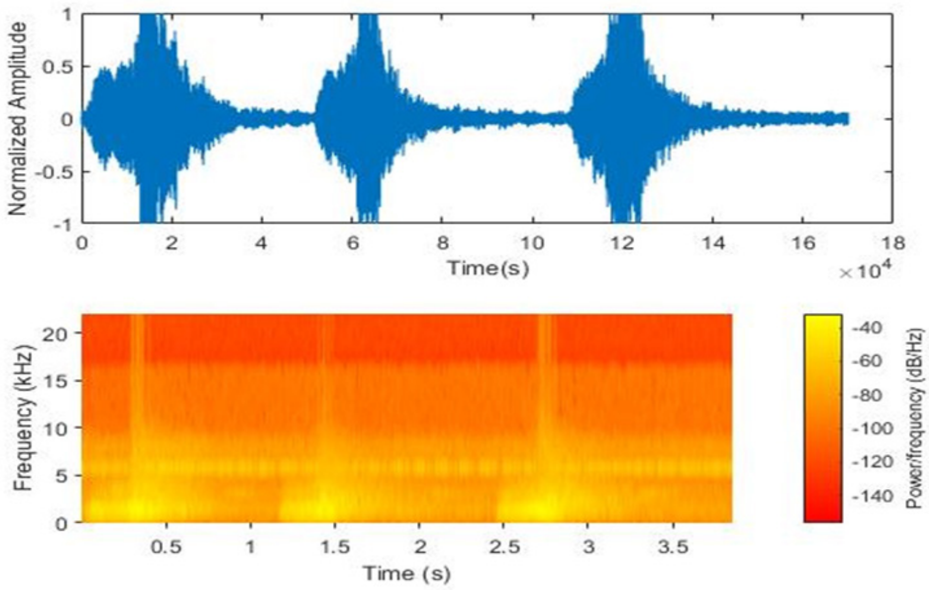


Figure 2. Artificial Neural Network (ANN) architecture adopted

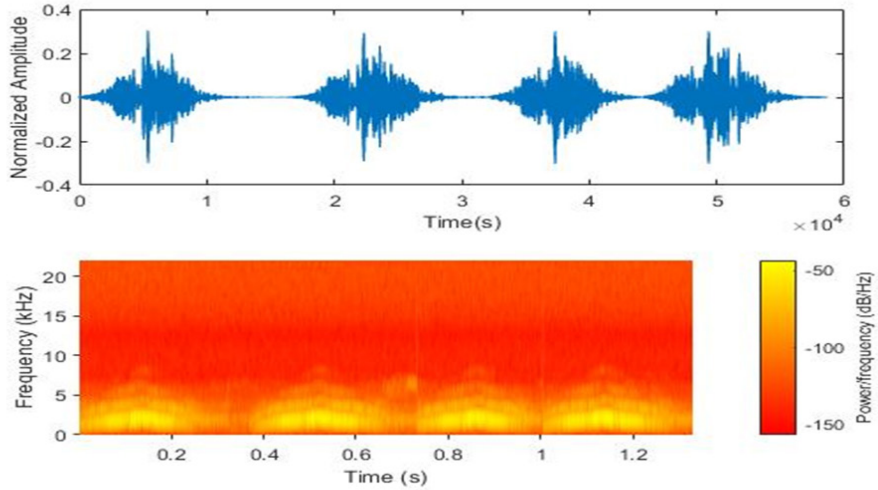
RESULTS AND DISCUSSION

Ten $L = 10$ endemic bird species of the Borneo region have been selected, i.e. Rhinoceros Hornbill (RH), Hooded Pitta (HP), Savanna Nightjar (SN), Collared Owlet (CO), Collared Kingfisher (CK), Crested Serpent Eagle (CSE), Bornean Tree Pie (BTP), Bornean Spider Hunter (BSH), Malaysian Pied Fantail (MPF), and Malaysian Banded Pitta (MBP), with audio recordings collected from the xeno-canto (https://www.xeno-canto.org/) online database, which is one of the most frequently used online databases in bird sound classification related research (Ramashini et al., 2019; Sprengel et al., 2016; Lasseck, 2015; Stowell & Plumbley, 2014). These bird species represent some of the most commonly found birds in the region. Sponsored by the Xeno-Canto Foundation, the Xeno-Canto online database contains sound recordings of wild birds from all over the world verified by experts. These recordings are shared under various Creative Commons licenses, freely available online. Thus, they can be used for education and research purposes. The time duration of each bird sound sample varies, with standard recordings lasting for a few seconds. Metadata provided with the data has indicated that most samples are recorded with a 44 kHz sampling rate. Figure 3 shows the time and frequency domain representation of selected bird sounds.

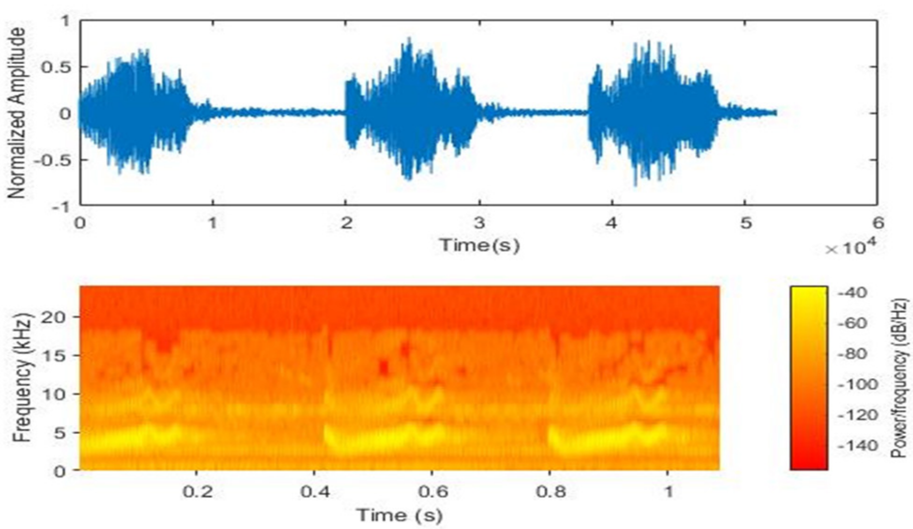


(a)

Figure 3. Time and frequency representation of bird samples; (a) Rhinoceros Hornbill (RH)

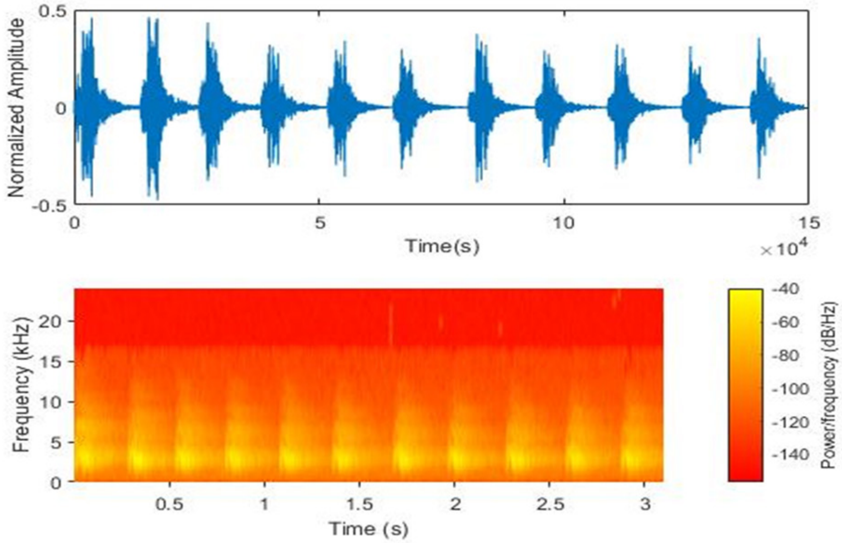


(b)



(c)

Figure 3. Time and frequency representation of bird samples; (b) Hooded Pitta (HP), (c) Savanna Nightjar (SN)



(d)

Figure 3. Time and frequency representation of bird samples; (d) Collared Kingfisher (CK)

The sounds are segmented using an energy-based automatic segmentation algorithm and divided into training and testing data sets. In this work, both 70:30 and 80:20 ratios of training to testing data are performed. In addition, 20% of the training dataset is chosen randomly for cross-validation purposes. The model is iteratively trained and validated on these different datasets. Furthermore, the training and testing data set are shuffled randomly multiple times to replicate the training and testing cycles with different combinations. These are done to avoid over-fitting as well as to ascertain the consistency of the result.

In total, 150 bird sounds have been used for training and testing, with 15 bird sounds for each class. $M = 35$ features have been extracted from each segment automatically, consisting of the frequency domain, time domain and other types of features, to form the training and testing feature matrices, giving the original feature matrix $X \in R^{N\times35}$ of training bird sounds and the i^{th} test bird sound $y_i \in R^{35}$.

The original feature matrix $X \in R^{N\times35}$ of the training bird sounds and the i^{th} test bird sound $y_i \in R^{35}$ can be fed directly to the classification algorithms. Alternatively, the dimensions may be first reduced using LDA to give a reduced matrix X_r. Similarly, in the testing phase, either the i^{th} test bird sound y_i can be fed directly, or it can be first reduced using the derived projection matrix V_k to give y_{r_i}; incurring additional complexity of $69.K.O(1)$ in the testing phase. Obviously, reducing *K* reduces complexity during the classification stage but requires extra computation during its dimension reduction process. In the classification stage using NC classifier, computational complexity is given by $(10.K + 1)O(1)$, and is dependent on the number of features *K* that are fed onto the

NC classifer. On the other hand, computational complexity of ANN classifier is given by $(2K+20)n^hO(1)+n^hO(g_h)+10.O(g_o)$; dependent on both the number of features K that are fed onto the ANN classifer, as well as the number of neurons n^k in the hidden layer of the ANN classifier. Increasing K results in an increase in the computational complexity for both NC and ANN classifiers, and the larger the number of neurons n^k, the more complex the ANN model becomes. Elliot Symmetric Sigmoid and Softmax functions are used for the hidden and output layers of the ANN, respectively.

In the case of the NC classifier, feeding all 35 features to the classification model directly without reduction requires a complexity of $351.O(1)$ in the testing phase. However, introducing LDA to reduce the number of features prior to the classification stage, requires extra computation for the reduction process. As such, collectively, the computational complexity for both the dimension reduction and NC classification sharply increases from $351.O(1)$ to $2687.O(1)$ by reducing the number of features from 35 to 34 using LDA. However, complexity decreases as the number of features K fed onto the classification model are reduced further. For instance, reducing the number of features K fed onto the classification model to a certain limit would result in lower complexity than without using LDA. Table 1 shows the relationship between the number of features K fed onto the NC classifier and the resulting computational complexity, with and without LDA.

Testing accuracy of the NC classifier, with and without LDA, for different features K being fed, is given in Figure 4. NC classifier accuracy for without any reduction is 13.3% and 10% for 80:20 and 70:30 ratios of training to testing data, respectively. It can be seen that even with 1 LDA, the accuracy of the NC is much higher than without using LDA. Thus, this is because LDA projects the original matrix X and the test sound y_i onto reduced dimensions matrix and vector that can best discriminate between different classes of bird sounds. Output matrix and vector are also ranked. The lower elements represent the most significant elements that may be used to discriminate between classes. As shown from Figure 4, improvement in accuracy is initially considerable with a small number of LDA features considered, up to a maximum accuracy upon which increasing the number of LDA features considered even further would result in a gradual reduction in accuracy using the NC classifier. Maximum accuracies of 96.7% with 7 LDA features, and 78% with 5-7 LDA features considered, are achievable for 80:20 and 70:30 ratios of training to testing data, respectively. With 7 LDA features considered, the computational complexity of the LDA/NC is $555.O(1)$.

Table 1 also gives the relationship between the number of features K fed onto the ANN classifier and the resulting computational complexity, with and without LDA. Whilst it is evident that complexity for NC reduces with the reduction in the number of LDA features K, it is not very straightforward with ANN. The complexity of ANN has a direct relationship with the number of LDA features K. However, it is also dependent on the number of neurons

n^h in the hidden layer with complexity increasing with the number of neurons n^h in the hidden layer. Figure 5 shows accuracies obtained using the ANN classifier with different values of n^h for $K = M = 35$, i.e. without LDA reduction. It can be seen that increasing the number of neurons n^h in the hidden layer does not necessarily improve the accuracy of the ANN model. The highest accuracy of 93.3% is achieved with $n^h = 100$. In fact, for a given number of features K, the number of neurons n^h in the hidden layer may need to be optimised to reach an optimal accuracy during the training phase.

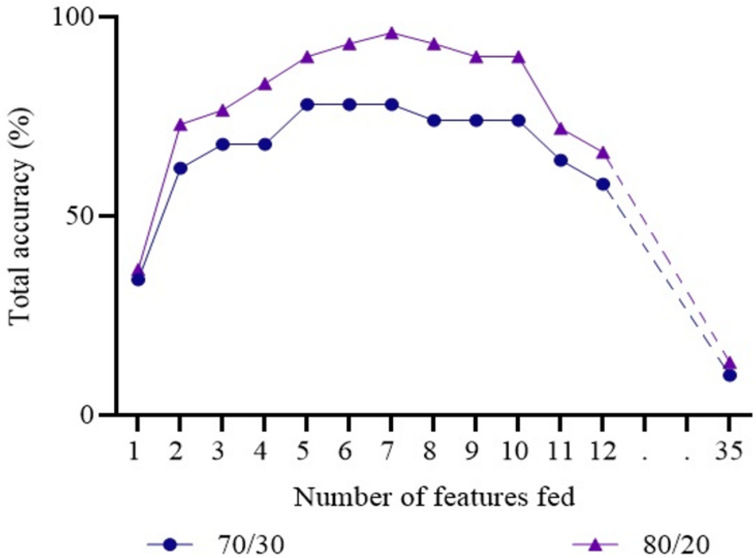


Figure 4. Testing accuracy obtained by Nearest Centroid (NC) for different number of Linear Discriminant Analysis (LDA) features K considered.

Table 1

The computational complexity of NC and ANN, with and without LDA reduction, for different number of inputs to the classification models for the testing phase

Inputs		Nearest Centroid (NC)	Artificial Neural Network (ANN)
w/o LDA	35	$351.O(1)$	$90.n^h O(1) + n^h O(g_h) + +10.O(g_o)$
W LDA	30	$2371.O(1)$	$(2070 + 80n^h)O(1) + n^h O(g_h) + 10.O(g_o)$
	25	$1976.O(1)$	$(1725 + 70n^h)O(1) + n^h O(g_h) + 10.O(g_o)$
	20	$1581.O(1)$	$(1380 + 60n^h)O(1) + n^h O(g_h) + 10.O(g_o)$
	15	$1186.O(1)$	$(1035 + 50n^h)O(1) + n^h O(g_h) + 10.O(g_o)$
	10	$791.O(1)$	$(690 + 40n^h)O(1) + n^h O(g_h) + 10.O(g_o)$
	5	$396.O(1)$	$(345 + 30n^h)O(1) + n^h O(g_h) + 10.O(g_o)$

Table 2 shows the optimum number of neurons n^h in the hidden layer and accuracies for a different number of LDA features. Generally, it can be seen that increasing the number of LDA features *K* also increases the number of neurons n^h in the hidden layer, which consequently results in an overall increase in complexity. In terms of accuracy, the initial increase in the number of LDA features *K*, increases accuracy until a maximum accuracy is reached, increasing *K* even further would result in a reduction in accuracy. Accuracies for different numbers of LDA features *K* for 70:30 and 80:20 ratios of training to testing data are given in Figure 6. Characteristics for both 70:30 and 80:20 ratios of training to testing data are similar with an initial increase in accuracy until an optimum is reached, beyond which accuracy starts to reduce.

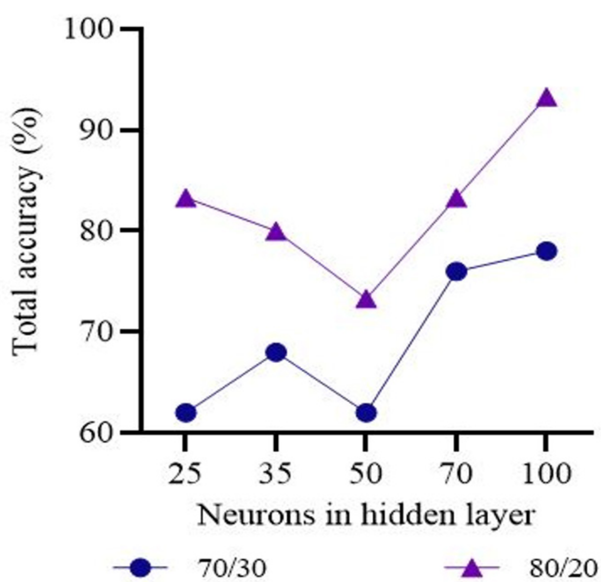


Figure 5. Testing accuracy obtained by Artificial Neural Network (ANN) for different number of neurons n^h in the hidden layer

Maximum accuracy obtained is 96.7% for 10 input LDA features with complexity of $1250.O(1) + 14.O(g_h) + 10.O(g_o)$. This is the same maximum accuracy obtained by NC classifier for 7 input LDA features with complexity of $555.O(1)$.

Comparing classification accuracies of NC and ANN classifiers with and without LDA, it can be seen that the selection of features using LDA improves performance significantly. Hence, this is especially true in the case of the NC classifier, where classification accuracy increased from 13.3% to 96.7% by using LDA. Thus, NC classification with 7 LDA and

ANN classification with 10 LDA, for 80:20 ratio of training and testing, produce the optimum testing accuracies.

Table 3 shows class-wise classification results of NC classifier with 7 LDA as input and ANN classifier with 10 LDA as input. It can be seen that the NC classifier wrongly predicted one sample of the bird BSH whilst the ANN classifier wrongly predicted one sample of the bird CO. Precision or *P*, i.e. the proportion of correct classification from the total predicted classification of a particular class. Recall or *R*, i.e. the proportion of correct classification from the actual total classification of a particular class, is another valuable measure of performance of a classification model. F_{SCORE} value of a particular class can be obtained from *P* and *R* as Equation 19,

$$F_{score} = \frac{2PR}{P + R} \quad [19]$$

Table 2

Computational complexity of artificial neural network (ANN) with linear discriminant analysis (LDA) reduction for the optimum number of neurons in the hidden layer n^h*, for the testing phase*

Input LDA, K	Neurons in Hidden Layer, n^h	Complexity	Accuracy (%)
1	3	$135.O(1) + 3.O(g_h) + 10.O(g_o)$	46.7
2	3	$210.O(1) + 3.O(g_h) + 10.O(g_o)$	73.3
3	4	$311.O(1) + 4.O(g_h) + 10.O(g_o)$	66.7
4	5	$416.O(1) + 5O(g_h) + 10.O(g_o)$	83.3
5	8	$585.O(1) + 8.O(g_h) + 10.O(g_o)$	76.7
6	9	$702.O(1) + 9.O(g_h) + 10.O(g_o)$	86.7
7	8	$755.O(1) + 8.O(g_h) + 10.O(g_o)$	90
8	8	$840.O(1) + 8.O(g_h) + 10.O(g_o)$	90
9	11	$1039.O(1) + 11.O(g_h) + 10.O(g_o)$	90
10	14	$1250.O(1) + 14.O(g_h) + 10.O(g_o)$	96.7
11	15	$1389.O(1) + 15.O(g_h) + 10.O(g_o)$	66.7
12	16	$1532.O(1) + 16.O(g_h) + 10.O(g_o)$	63.3

Figures 7 and 8 show the confusion matrices for NC classifier with 7 LDA inputs and ANN with 10 LDA inputs, respectively. As shown in Figure 7 for NC classification with 7 LDA, one sample from class 8, i.e. BSH bird, has been wrongly predicted as class 7, i.e. BTP bird. Subsequently, the precision value and F_{SCORE} for class 7 are 0.75 and 0.86, respectively, whilst for class 8, recall value and F_{SCORE} are 0.67 and 0.80, respectively. On the other hand, as shown in Figure 8, ANN classification with 10 LDA wrongly predicted one sample of class 4, i.e. CO bird, as class 8, i.e. BSH bird. As a result, the precision value and F_{SCORE} for class 7 are 0.75 and 0.86, respectively. For class 4, the recall value and F_{SCORE} are 0.67 and 0.80, respectively.

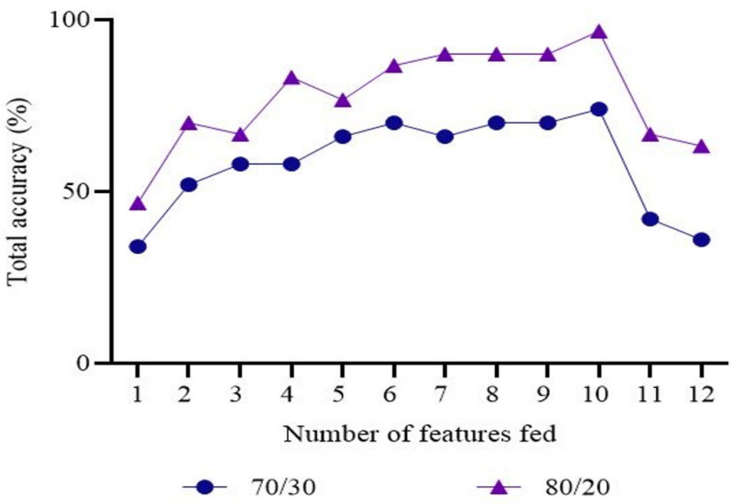


Figure 6. Testing accuracy obtained by the Artificial Neural Network (ANN) for different number of Linear Discriminant Analysis (LDA) features

Table 3

Class wise testing accuracy of Nearest Centroid (NC) classification with 7 Linear Discriminant Analysis (LDA) and Artificial Neural Network (ANN) classification with 10 Linear Discriminant Analysis (LDA)

Class Number	Birds Name	Testing Accuracy (%)	
		NC classification with 7 LDA features	ANN classification with 10 LDA features
1	RH	10	10
2	HP	10	10
3	SN	10	10
4	CO	10	6.7
5	CK	10	10
6	CSE	10	10
7	BTP	10	10

Table 3 *(Continued)*

Class Number	Birds Name	Testing Accuracy (%)	
		NC classification with 7 LDA features	ANN classification with 10 LDA features
8	BSH	6.7	10
9	MPF	10	10
10	MBP	10	10
Total Accuracy (%)		96.7	96.7

In the testing phase, NC and ANN classifiers with LDA give prediction accuracies of 96.7%. Each has one wrong prediction of the sample in different classes. Without LDA, the ANN classifier gives 93.3% accuracy with two wrong predictions. In contrast, the performance of an NC classifier without LDA is abysmal. Furthermore, in terms of computational complexity, both classifiers without LDA are comparatively less complex during the testing stage, than LDA. However, the complexity of the ANN classifier is always higher than the NC classifier, irrespective of using LDA as feature reduction, since its complexity depends on the number of neurons in both hidden and output layer and the activation function used.

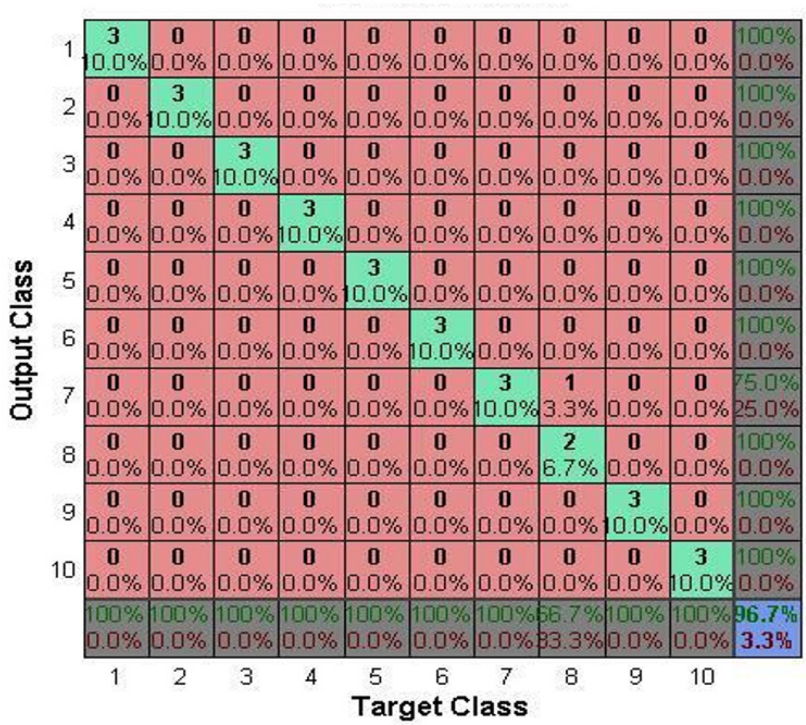


Figure 7. Testing confusion matrix: Nearest Centroid (NC) classification with 7 Linear Discriminant Analysis (LDA)

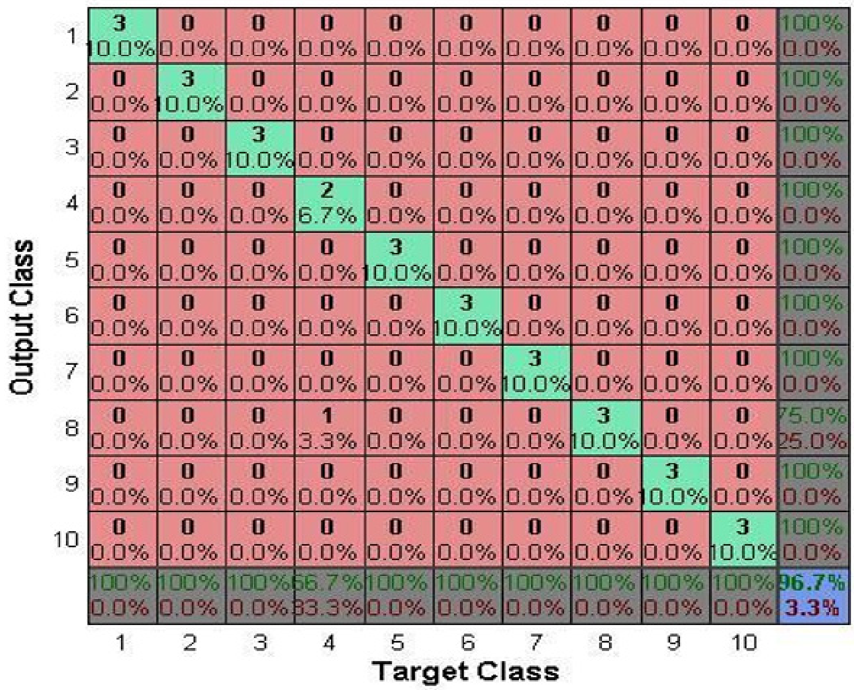


Figure 8. Testing confusion matrix: Artificial Neural Network (ANN) classification with 10 Linear Discriminant Analysis (LDA)

CONCLUSION

Classification of birds using their sound is preferable as compared to visual identification, especially in dense forests. The general processing steps for bird sound classification are pre-processing, segmentation, feature extraction and classification. This paper aims to classify ten endemic Bornean birds by their sounds, collected from an online database and pre-processed to remove unwanted noise. Then, using an energy-based automated segmentation algorithm, the recordings are segmented for further processing. Thirty-five (35) acoustic features have been extracted from the segmented samples. The LDA has been used to reduce the dimensionality and select only the significant features before feeding the transformed features onto the classifier. The NC and ANN classifiers have been used for classification. It has been shown that both NC and ANN classifiers with LDA give 96.7% accuracy, which is comparatively higher than the performance of both classifiers without LDA in terms of testing accuracy.

Nevertheless, when computational complexity is considered, the simple NC classifier produces the same accuracy with the computational complexity of only $555.O(1)$, compared to the more complex ANN classifier. Thus, the NC classifier requires 7 LDAs to produce the optimum result. On the other hand, ANN's computational complexity is $1250.O(1)+14.O(g_h)+10.O(g_o)$, requiring 10 LDAs to give the optimum classification accuracy. The result is significant, as it indicates that the simple NC classifier with LDA can give optimum classification accuracy of 96.7% with relatively low computational power.

In future work, other classification approaches such as naive Bayes and decision trees may be combined with the proposed method to reinforce the benefit of using the proposed dimensionality optimisation method in improving accuracy whilst reducing complexity. Furthermore, the variety of bird species and more samples may also be considered to reflect the rich biodiversity in the Borneo region whilst implementing real-time bird sound classification.

ACKNOWLEDGEMENT

The corresponding author would like to thank Universiti Brunei Darussalam (UBD) for awarding the University Graduate Scholarship (UGS) to pursue a PhD at the university. Further, the author like to thank Uva Wellassa University, Sri Lanka, for their support in granting study leave for her PhD work at UBD.

REFERENCES

Alvarsson, J. J., Wiens, S., & Nilsson, M. E. (2010). Stress recovery during exposure to nature sound and environmental noise. *International Journal of Environmental Research and Public Health*, *7*(3), 1036-1046. https://doi.org/10.3390/ijerph7031036

Anderson, S. E., Dave, A. S., & Margoliash, D. (1996). Template-based automatic recognition of birdsong syllables from continuous recordings. *The Journal of the Acoustical Society of America*, *100*(2), 1209-1219. https://doi.org/10.1121/1.415968

Badi, A., Ko, K., & Ko, H. (2019). Bird sounds classification by combining PNCC and robust Mel-log filter bank features. *Journal of the Acoustical Society of Korea*, *38*(1), 39-46. https://doi.org/10.7776/ASK.2019.38.1.039

Butler, R. W. (2019). Niche tourism (birdwatching) and its impacts on the well-being of a remote island and its residents. *International Journal of Tourism Anthropology*, *7*(1), 5-20. https://doi.org/10.1504/ijta.2019.10019435

Chou, C. H., Liu, P. H., & Cai, B. (2008). On the studies of syllable segmentation and improving MFCCs for automatic birdsong recognition. In *Proceedings of the 3rd IEEE Asia-Pacific Services Computing Conference, APSCC 2008* (pp. 745-750). IEEE Publishing. https://doi.org/10.1109/APSCC.2008.6

Elliott, D. L. (1993). A better activation function for artificial neural networks. *ISR Technical Report TR 93-8*. Neuro Dyne, Inc.

Evangelista, T. L., Priolli, T. M., Silla, C. N., Angelico, B. A., & Kaestner, C. A. (2015). Automatic segmentation of audio signals for bird species identification. In *2014 IEEE International Symposium on Multimedia* (pp. 223-228). IEEE Publishing. https://doi.org/10.1109/ISM.2014.46

Fagerlund, S. (2007). Bird species recognition using support vector machines. *EURASIP Journal on Advances in Signal Processing, 2007*, 1-8. https://doi.org/10.1155/2007/38637

Fagerlund, S., & Laine, U. K. (2014). New parametric representations of bird sounds for automatic classification. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings (ICASSP)* (pp. 8247-8251). IEEE Publishing. https://doi.org/10.1109/ICASSP.2014.6855209

Gerhard, D. (2003). Audio signal classification: History and current techniques. *Technical Report TR-CS 2003-07*. University of Regina.

Giannakopoulos, T., & Pikrakis, A. (2014). *Introduction to audio analysis: A MATLAB® approach*. Academic Press.

Härmä, A., Somervuo, P., Harma, A., & Somervuo, P. (2004). Classification of the harmonic structure in bird vocalization. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 5, pp. V-701). IEEE Publishing. https://doi.org/10.1109/ICASSP.2004.1327207

Kutzner, D. (2019). Environmental change, resilience, and adaptation in nature-based tourism: Conceptualizing the social-ecological resilience of birdwatching tour operations. *Journal of Sustainable Tourism*, *27*(8), 1142-1166. https://doi.org/10.1080/09669582.2019.1601730

Lasseck, M. (2015). Improved automatic bird identification through decision tree based feature selection and bagging. *LifeCLEF*. Museum für Naturkunde Berlin.

Lee, C. H., Han, C. C., & Chuang, C. C. (2008). Automatic classification of bird species from their sounds using two-dimensional cepstral coefficients. *IEEE Transactions on Audio, Speech and Language Processing*, *16*(8), 1541-1550. https://doi.org/10.1109/TASL.2008.2005345

Ludeña-Choez, J., Quispe-Soncco, R., & Gallardo-Antolín, A. (2017). Bird sound spectrogram decomposition through non-negative matrix factorization for the acoustic classification of bird species. *PLoS ONE*, *12*(6), 1-20. https://doi.org/10.1371/journal.pone.0179403

Martinez, A. M., & Kak, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(2), 228-233. https://doi.org/10.1109/34.908974

McIlraith, A. L., & Card, H. C. (1997). Birdsong recognition using backpropagation and multivariate statistics. *IEEE Transactions on Signal Processing*, *45*(11), 2740-2748. https://doi.org/10.1109/78.650100

Milani, M. G. M., Abas, P. E., & De Silva, L. C. (2019). Identification of normal and abnormal heart sounds by prominent peak analysis. In *Proceedings of the 2019 International Symposium on Signal Processing Systems* (pp. 31-35). Association for Computing Machinery. https://doi.org/10.1145/3364908.3364924

Mogi, R., & Kasai, H. (2013). *Noise-Robust environmental sound classification method based on combination of ICA and MP features*. *Journal of Artificial Intelligence Research, 2*(1), 107-121. https://doi.org/10.5430/air.v2n1p107

Priyadarshani, N., Marsland, S., & Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: A review. *Journal of Avian Biology*, *49*(5), 1-27. https://doi.org/10.1111/jav.01447

Ramashini, M., Abas, P. E., Grafe, U., & De Silva, L. C. (2019). Bird sounds classification using linear discriminant analysis. In *2019 4th International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)* (pp. 1-6). IEEE Publishing. https://doi.org/10.1109/ICRAIE47735.2019.9037645

Ranjard, L., & Ross, H. A. (2008). Unsupervised bird song syllable classification using evolving neural networks. *The Journal of the Acoustical Society of America*, *123*(6), 4358-4368. https://doi.org/10.1121/1.2903861

Selouani, S. A. S. A., Kardouchi, M., Hervet, É., Roy, D., Hervet, E., & Roy, D. (2005). Automatic birdsong recognition based on autoregressive time-delay neural networks. In *2005 ICSC Congress on Computational Intelligence Methods and Applications* (pp. 1-6). IEEE Publishing. https://doi.org/10.1109/CIMA.2005.1662316

Sharma, G., Umapathy, K., & Krishnan, S. (2020). Trends in audio signal feature extraction methods. *Applied Acoustics*, *158*, Article 107020. https://doi.org/10.1016/j.apacoust.2019.107020

Sprengel, E., Jaggi, M., Kilcher, Y., & Hofmann, T. (2016). Audio Based Bird Species Identification using Deep Learning Techniques. In *Working Notes of CLEF 2016* (pp. 547-559). Cross Language Evaluation Forum.

Stowell, D., & Plumbley, M. D. (2014). Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ*, *2*, Article e488. https://doi.org/10.7717/peerj.488

Suthers, R. A. (2004). How birds sing and why it matters. In *Nature's Music: The Science of Birdsong* (pp. 272-295). Elsevier Academic Press. https://doi.org/10.1016/B978-012473070-0/50012-8

Tan, L. N., Kaewtip, K., Cody, M. L., Taylor, C. E., & Alwan, A. (2012). Evaluation of a Sparse Representation-Based Classifier For Bird Phrase Classification Under Limited Data Conditions. In *Thirteenth Annual Conference of the International Speech Communication Association* (pp. 2522-2525). International Speech Communication Association (ISCA).

Terry, A. M. R., & McGregor, P. K. (2002). Census and monitoring based on individually identifiable vocalizations: The role of neural networks. *Animal Conservation*, *5*(2), 103-111. https://doi.org/10.1017/S1367943002002147

Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI Communications*, *30*(2), 169-190. https://doi.org/10.3233/AIC-170729

Trifa, V. M., Kirschel, A. N. G., Taylor, C. E., & Vallejo, E. E. (2008). Automated species recognition of antbirds in a Mexican rainforest using hidden Markov models. *The Journal of the Acoustical Society of America*, *123*(4), 2424-2431. https://doi.org/10.1121/1.2839017

Vilches, E., Escobar, I. A., Vallejo, E. E., & Taylor, C. E. (2006). Data mining applied to acoustic bird species recognition. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 3, pp. 400-403). IEEE Publishing. https://doi.org/10.1109/ICPR.2006.426