International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 13 (2021) pp. 319-335

Received: 28 April, 2021; Accepted: 8 Sept, 2021; Publish: 9 October, 2021

A Comprehensive Survey on Bird Species Identification Models

Nilanjana Adhikari[1], Suman Bhattacharya[2] and Mahamuda Sultana[3]

[1] Techno International New Town, MAKAUT
Kolkata, India
nadhikari1983@gmail.com

[2] Guru Nanak Institute of Technology, JIS
Kolkata, India
suman93.2004@gmail.com

[3] Guru Nanak Institute of Technology, JIS
Kolkata, India
sg.mahamuda@gmail.com

Abstract*: Fundamental classes of ecosystem comprise of provisioning, regulating, supporting, and cultural services. Of all the living organisms, 'birds' adhere to all the earlier mentioned classes thereby motivating the current work in this paper. Humans coexist with birds and generally identify them via either their physical appearance or their vocal note. Bird species identification (BSI) has attracted global attention lately owing to the rise of technological prowess in aiding their identification. Although birds are identified by their physical appearance and vocal notes; often these two features lead to ambiguity due to image acquisition in a diverse range of postures, intensity, and varied notes of the vocal data. Recent models trained on Deep Neural Networks (DNN) have exhibited better performance in minimizing the said ambiguity. In-depth analysis of proposals in BSI mandates focused study of such prior art, majorly, to identify the inaccuracies generated through ambiguity. The present work provides the reader a detailed survey of preceding art addressing both the dimensions of BSI, namely physical appearance and vocal note based. The analysis stresses mostly on the DNN based models focused on Image Isolation, Background Reduction for physical appearance identification, and all those proposals which evaluate the vocal note datasets. The work in this paper targets those researchers who intend to pursue further study in BSI and design more efficient models and serves as a benchmark for future exploration in BSI using physical appearances and vocal notes. This work provides adequate information for amateur researchers to explore the research on bird species identification models.

***Keywords*:** Bird Species Identification, Deep Neural Network, Image Classification, Vocal Notes, Literature Survey**.**

I. Introduction

India is a vast country with various areas, like hills, plains, rivers, jungle, and sea. In this different area, different bird species are there. Birds are significant to the ecosystem because of their direct and indirect contribution to the environment. The bird lovers, creators, novelists, and the composer often find a spring of enthusiasm and inspiration from the birds' sounds from all over the world. Birds twitters everywhere, and most people can recognize birds by their sounds. But only very few species of birds can be known by watching them or by their twittering. In India, a large number of different types of birds are available. So, it is almost an impossible task to recognize every species individually without referring to anything. In this context, there are lots of books or magazines which are available in the market. So, this job is again complex to recognize the bird species by referring to these kinds of books or journals. Different types of glitches are handled by ornithologists when they identify and classify bird species. After studying the attributes of birds, ornithologists categorized them by their living atmosphere, their ecological influence so on.

A literature review on the bird species identification system using their image or call has been discussed in detail in this study. The procedure of this system approaches firstly by collecting bird images either from real-time or from the stored database. Then through an image processing system, that particular bird image has been segregated from its background. Then applying the deep learning method, the species image has been recognized. This system will reduce the overall time and effort of human beings. This system is self-adaptive, and it will reach its height of accuracy after a certain no. of iterations.

To identify the bird from an image, the model first needs to detect and differentiate the part of the bird and the irrelevant background from that given image. To feed this to the model, it has to collect an adequate number of data to train the same. The task of species identification is a procedure of classifying objects with the help of an efficient methodology trained and tested on a proper dataset and separating them into various categories. This study establishes a method using a Deep Convolutional Neural Network (DCNN) to extract information from bird images captured previously or in real-time. First, bird images were collected and then segregated part of a bird's body by excluding the irrelevant background. Then, generating a feature vector, a deep neural network model was trained with the feature vector and subsequently classified. The said trained dataset was stored on a server to identify a target object.

Moreover, capturing an image of a bird is not always possible due to the dense forest. Instead, a clear or noisy audible bird call can be here. For this constraint, sound-based identification system has also been proposed in this study. Initially, a dataset consisting of the audio file of different species of birds is collected. Afterward, the sound clips are pre-processed by framing, silence removal, and reconstruction. Then Spectrogram will be generated for each pre-processed sound clip. These spectrograms will then be trained through the Convolutional Neural Network (CNN) based on the input features of the given sound clip, which will identify the bird species. A Real-time implementation has also been proposed.

The rest of this proposed study is divided into seven segments as follows. The next section presents the Related Work followed by description of the Article Selection Process in the next section. Section IV exhibits the Comparative Analysis followed by Discussion and the Conclusion.

II. Related Work

Species recognition and classification are being performed nowadays using supervised machine learning algorithms. In this section, recent research works on bird species identification have been discussed using the image or using calls. S. Bayat et al. [*1*] have shown bird species identification using the CNN model using bird calls by comparing the raw and processed data generating acoustic Mel-spectrograms. The accuracy increased by approximately 4% using processed data. In the research paper [*2*], the Artificial Neural Network model was used to identify one bird call at a time using power spectral density by obtaining data for each bird type individually. In the proposal [*3*], the Residential Network model as a pre-trained Convolutional Neural Network has been used to train the input image to identify the individual bird. In the research work [*4*], converting input bird image into a greyscale format, autograph produced. Then examining every autograph, the score sheet generated. This score sheet was then trained and tested through Deep Convolutional Neural Network (DCNN) on the Google-Net framework bird species were identified. B. Chandu et al. [*5*] have shown that constructing the dataset with the sound recordings of different bird species and applying sound pre-processing technique spectrograms has been generated, which has been trained and tested by Convolution Neural Network (CNN) classifier and identified bird species. In the research work [*6*], from Borneo regional bird calls, features have been extracted, which have been reduced using Linear Discriminant Analysis (LDA) and then classified and identified using Nearest Centroid (NC) classifier. In the paper [*7*], using the zero-frequency filtering method and calculating the average amount of epochs per second, it has been discriminated birdcall made by a single bird or made by multiple birds. S Islam et al. [*8*] have trained and tested bird images using the VGG-16 network to identify and classify the species. According to P. Jankovic et al. [*9*], the acoustic signal of different bird species pre-processed and then trained and tested using hidden Markov models (HMMs) and unsupervised modeling to identify particular bird species. In the research work [*10*], the very low-resolution bird image was trained by deep learning methods to recognize the same. In the paper [*11*], bird calls were recorded, segmented, and features extracted by the matrix factorization method. Using the k-means clustering supervised algorithm, then those features have been tested to identify the species. According to H. Weerasena et al. [*12*], the bird's call can be experimented with bioacoustics monitoring using the SVM classifier by recording, pre-processing, and segmenting for classification. In the research work [*13*], a Gaussian mixture model (GMM) energy detector was applied to auditory events to extract spectral patterns and texture features. Then distinct features have been taken out using a Relief-based algorithm and passed to support vector machine for classification of the call. R. P. Tivarekar and H. G. Virani [*14*] generated syllables from bird calls by recording, pre-processing, and segmenting. Then these syllables have been trained using machine learning algorithms and recognized by Linear Discriminant Analysis (LDA) with a probabilistic approach. In the paper [*15*], snowy owl and toucan images have been segmented and examined nine color-related features like mean, standard deviation, and skewness of red, green, and blue (RGB) and classified using a Support Vector Machine algorithm. J. Niemi et al. [*16*] shown a bird image together with the radar image has been trained and tested by Deep Convolutional Neural Network for its identification. P. Rai et al. [*17*] from recorded bird calls, a cepstral feature matrix on the Mel-scale generated. Then this feature matrix has been trained and tested on support vector machines for classification. In this research work [*18*], the Random Decision Tree methodology was used to identify bird species or subspecies from its calls. Here in the paper [*19*], frequency domain analysis has been used to identify bird calls. In this research work [*20*], Colombian region bird calls were trained and tested through machine learning visualization techniques to identify the bird species. In this paper [*21*], both images and sounds of bird species have been experimented with for their identification. Using Scale Invariant Feature Transform (SIFT) method, visual features were taken out from unconstrained bird images, and using Mel-frequency Cepstral coefficients (MFCCs), acoustic features were taken out from bird vocalizations and then trained by a support vector machine classifier. T. Berg et al. [*22*] have used one-vs.-most classifiers for fine-grained image classification and clustering of a large dataset. In the research work [*23*], the bird calls have been recorded near the airport and segmented in pulses using high amplitude, and these pulses have been trained and tested through the SVM classifier and identified the bird. In the paper [*24*], collecting two different sets of birds calls, acoustic features have been extracted using Mel Frequency Cepstrum, and then these features have been classified using a k-NN-classifier. According to L. Jian [*25*], bird species have been identified using similarity comparison by visual feature analysis. In the research work [*26*], extracting the color feature, segmentation has been done to eliminate background elements, then splitting the segments into component planes, normalized color histograms have been computed, which has been used by the learning algorithm to distinguish bird species. The author C. N. Silla et al. [*27*] has monitored ordered classification of bird call comparing three types of approaches: classification approach, the local-model per parent node classifier approach employing a classic Naive Bayes algorithm, and the global-model hierarchical-classification approach by the Global Model Naive Bayes (GMNB) algorithm. In the research work [28], wild bird calls were recorded, pre-processed, segmented, and feature extracted, then these features have been trained and tested in FFT and to the 4-layer neural network compared the result. In this work [29], on recorded bird call wavelet transformation has been applied and mean value, the strength of the frequency and modulation spectrum was calculated then these features were trained and tested through Neural Network to classify bird call. In this paper [30], in an automatic bird species identification model K-Nearest Neighbour and Self-organizing Maps classifiers were used to classify the acoustic data. In this paper [31], generating frequency spectral from bird calls, a mean value was calculated, and then it was trained in the neural network for classification. M. T. Lopes et al. [32] have recorded bird calls, and then signal processing and machine learning techniques have been applied to identify the same. In the research work [33], a comparative analysis of probabilistic instance-based decision trees, neural networks, and support vector machines has been experimented with using bird audio signals. In this research work [34], recording bird calls from a noisy environment, segmentation has been performed on that call, then those segments were classified using Random Forest classifier. In this paper [35], different classifiers like Mel frequency Cepstrum coefficients, Principal Components Analysis, and K-Nearest Neighbors performance have been compared to identify bird species. In this research work [36], the bird image from the ImageNet dataset has been trained in five-layered Convolutional Neural networks to identify bird species. According to this paper [37], the residual network is less complicated than the deep CNN technique for species identification using the Image Net dataset. Branson, Van Horn, et al. [38] accomplished that visual categorization and identification of bird species by the object's poses for local image feature computed using Deep Convolutional Neural Network, and then graph-based clustering algorithm has been applied for categorization. All the above information has been collected from the website [39]. This survey study has been inspired a lot through the essence of the journal articles [40]. The pattern and structure of this article helped to frame this study.

III. Article Selection Process

In this study, from the last ten years of research work in bird species identification, thirty-eight research papers have been taken into consideration for the survey. Many research works are contributed related to bird species identification, in which artificial neural network or machine learning-based works are chosen for this survey. The trend and pattern can be identified or classified easily, and multidimensional and multi-variant data can also be handled efficiently by Artificial Neural Networks or Machine Learning algorithms. In association, Machine Learning algorithms expand skills and improve correctness and competence. It allows making a better decision.

In this study, thirty-eight papers have been arranged into three different clusters, namely "Bird species identification using call," "Bird species identification using an image," and "Bird species identification using the image and call both." These 38 papers have been shortlisted from a total count of 62 papers accounting for a 61% selection rate.

As discussed earlier, all the selected papers have been divided into three clusters. Table 1 shows cluster-wise references. Table 2 presents the year-wise publication count graphically further represented in Figure. 1. Figure. 2 presents the Citation count of the selected papers exhibited in Table 3.

Cluster #	Cluster Name	References
1	Bird species identification using call	[1] [2] [5], [20], [31], [32], [33] [11], [24], [30], [35] [12], [13], [17], [19], [23] [6], [7], [9], [14], [18], [27], [28] [29], [34]
2	Bird species identification using image	[3], [4], [10], [16] [8], [15], [22], [25], [26]
3	Bird species identification using image and call both.	[21]

Table 1. Clusters and related References.

Year	Publication Count
2020	5
2019	4
2018	4
2017	4
2016	3
2015	3
2014	5
2013	4
2012	2
2011	3
2010	1

Table 2. Year Wise Publication Count.

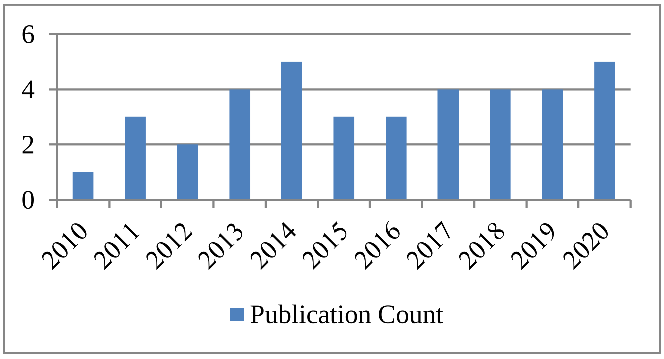


Figure 1. Graphical representation of Year Wise Publication Count

Ref	CC	Ref	CC	Ref	CC	Ref	CC
[1]	29	[10]	9	[19]	8	[28]	9
[2]	10	[11]	19	[20]	17	[29]	6
[3]	20	[12]	21	[21]	18	[30]	14
[4]	15	[13]	26	[22]	33	[31]	5
[5]	5	[14]	18	[23]	37	[32]	30
[6]	23	[15]	18	[24]	28	[33]	26
[7]	17	[16]	18	[25]	14	[34]	8
[8]	5	[17]	13	[26]	4	[35]	15
[9]	7	[18]	12	[27]	25		

Ref: Reference

CC: Citation Count

Table 3. Citation Count.

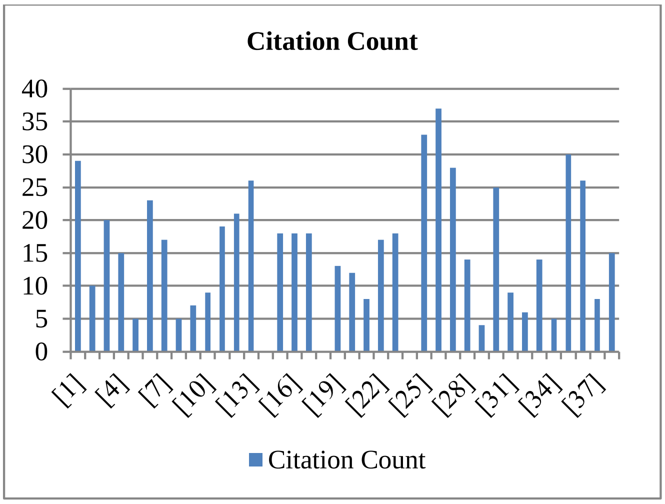


Figure 2. Graphical representation of Article wise Citation Count

A. *Algorithm for Clustering*

The tree for cluster and sub-cluster has been presented in Figure. 3. In the first segment total of twenty-five (25) papers have been studied and has been clustered into four (4) different sub-clusters depending upon the techniques applied viz

1) Using Deep CNN technique

a) CNN technique

It is a specialized case of ANN. It is used to analyze visual images. Based on the architecture of the layers, it filters input features and provides transformation responses. CNN can be constructed using a multi-layered perceptron that is fully connected networks, where neurons of a layer are connected to all the neurons of the next layer and so on. It causes the overfitting of data. The overfitting can be resolved by a different method of regularization.

b) Deep learning:

Deep learning is a part of the Machine Learning Method based on ANN techniques. It extracts visual or audio features from the given input. This learning technique can be supervised or unsupervised. The word "deep" denotes the use of multiple layers in the network. It is a network with an unrestricted number of layers of limited size. The application of this learning method is boundless. Here, it was used in the following research works for image recognition and classification system.

2) Using Mel-Frequency Cepstrum Coefficients (MFCC)KNN, and K-means clustering technique

a) Mel-Frequency Cepstrum Coefficients:

Mel-frequency cepstrum (MFC) is used to extract temporal power spectrum features of a sound frequency. Mel-frequency cepstral coefficients are factors of MFC. It is extensively used in the following papers to extract features from a segmented sound clip after applying sound preprocessing technique.

b) The MARSYAS Framework:

MARSYAS is the abbreviation of Music Analysis, Retrieval, and Synthesis for Audio Signals. It is a software tool applicable to an acoustic event to extract and analyze audio features from a sound clip.

c) KNN classifier:

K-Nearest Neighbour is a Supervised Machine Learning technique. Once a new test case is fed into the classifier, it compares the new data with the existing trained dataset for similarity. It put the new data into the cluster which is most similar to the class. It is applicable for audio or image classification.

d) Nearest Centroid (NC) classifier:

The working principle is similar to the KNN classifier.

e) K-means clustering:

It is an unsupervised machine learning technique. It solves the clustering problems. K-Means Clustering is an iterative algorithm that groups the unlabeled data into distinct clusters. "K" denotes the number of predefined categories. It also arranges the new data into a group with similar properties. The following paper used this clustering algorithm extensively.

3) Using SVM technique:

Support Vector Machine technique is a Supervised Machine Learning technique. For classification and regression-related problems, it is applicable. It creates the decision boundary to separate n-dimensional event space into distinct categories. The finest decision boundary is termed a hyperplane. It is also called a support vector.

4) Using the miscellaneous technique

a) Linear Discriminant Analysis:

Linear Discriminant Analysis is also called Normal Discriminant Analysis or Discriminant Function Analysis. It reduces the dimension of any feature vector applicable to Supervised Machine learning classification problems. It separates the dataset into different distinct classes. It projects the numerous features generated from the feature extraction method from a higher dimension to a lower dimension space.

b) Random Decision Tree:

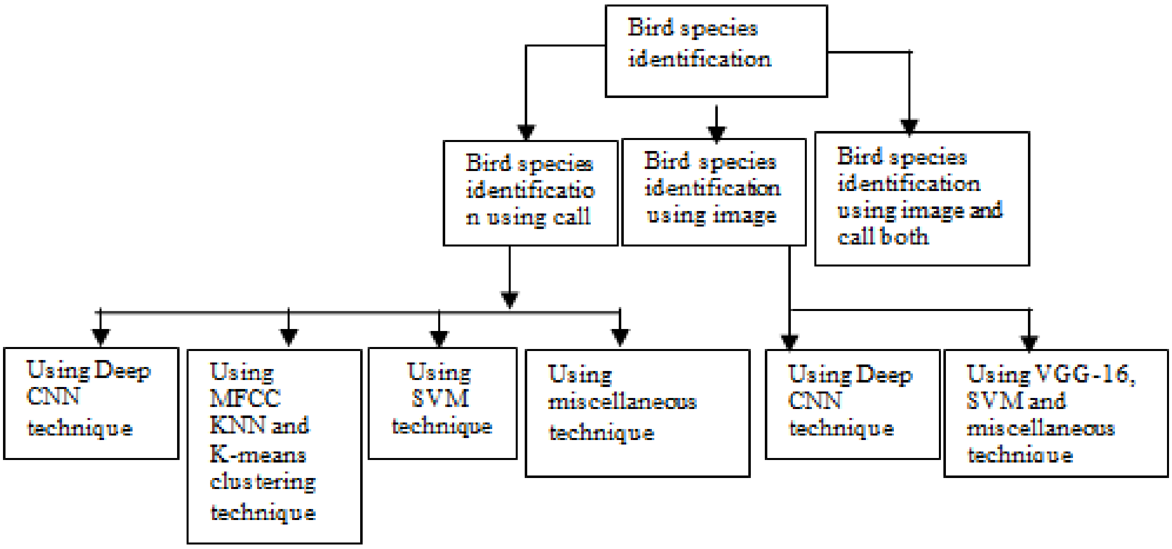
Random Forest is a supervised machine learning technique. For recognition, classification, and regression-related problems, it is extensively used. It is an ensemble learning technique. It merges several classifiers to solve a complex problem to increase the performance of the system. It has several decision trees that depend on the subsets of the input dataset. The final prediction is not dependent on a particular decision tree but the highest vote of decision.

In the next segment, a total of twelve (12) papers have been taken collectively and was clustered into two (2) different sub-clusters viz

1) *Using the Deep CNN technique*
2) *Using VGG-16, SVM, and the miscellaneous technique.*

a) Artificial Neural Network:

An artificial neural network is a sub-field of Artificial Intelligence. This network has several neurons which are connected in various layers, known as nodes. There are three



individual layers in this network.

Figure 3: Cluster and Sub-cluster

An input layer, which accepts inputs in the required format. Hidden layer, which presents in-between the input layer and the output layer. This layer finds hidden features and patterns from the given input by applying the weighted sum of input bias and the activation function. Output layer, which shows the output.

b) VGG-16:

A pre-trained CNN model is VGG-16, where16 refers the no of hidden layers.

In the last segment, a single paper is presented for this cluster, "bird species identification model using its image and tone."

IV. Brief Survey of Selected Articles

This segment contains empirically inspected approaches used for accomplishing good performance. Subsequently, a discussion about the different limitations and logical extensions of work has also been discussed in three different sections viz A. Bird species identification using call; B. Bird species identification using image; C. Bird species identification using image and call.

A. Bird species identification using call

In this section, bird species identification systems using its tone have been elaborated in detail. The different papers used different types of methodologies. In the following subsections, the research works with similar logic or algorithm have been kept collectively, and a brief discussion on each work has been presented.

1) Using Deep CNN method

In a brief discussion on the research work [*1*], the bird species were identified using the CNN method, where acoustic Mel-spectrograms have been examined to show the differences between the raw and processed data. Approximately 4% accuracy has been improved using pre-processed call data in comparison with the raw dataset.

The limitation of this approach bird species was identified for a particular area bird, and the acoustic raw input signal was pre-processed before testing. But in pre-processing, some portion of the data may loss and errors may introduce in the actual result.

In the research work [*2*], collecting and pre-processing bird calls, power spectral density was processed for each bird type individually. a Multilayer perceptron methodology of

Artificial Neural networks has been used for classification. The flowchart of this system is in Figure 4.

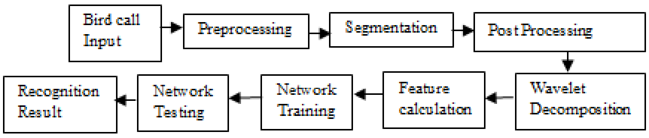


Figure.4 Flowchart of the Methodology of [*2*]

The limitation of this work is, data has been recorded for only four particular bird types. Here, the dataset is not up to the mark to train and test through Neural Network. This system may fail to recognize the bird species if multiple birds of the same species chatter with different frequencies.

A brief discussion on the paper [*5*], constructing the dataset with the sound recordings of different bird species, and applying pre-processing technique, spectrograms generated. These spectrograms have been trained and tested by Convolution Neural Network (CNN) and identified bird species. In this work, the pre-trained CNN has been used to classify millions of multiple categories of images. Image resolution is 227*227*3. This network is arranged with five CNN layers and three fully connected layers in which ReLu has also been applied to every layer. The flowchart of the methodology of this work is in Figure 5.

A limitation of this research work is, four specific bird species were considered. The dataset is weak for the convolutional neural network. An automatic identification system for different types of birds is not possible through this model. More precision has to enhance by fine-tuning performance parameters.

A brief description of the work [*20*], audio signals have been collected for bird call classification. Preprocessing the recorded audio calls and extracting the most significant features using MFCC audio vector was generated. Then, using the PCA method, the dimensionality of the vector has been reduced, and then a 2D visualization technique has been applied. A cosine distance function has also been applied to the audio vector to compare two audio calls within the audio vector. Then trained and tested using ML for recognition. Figure 6 shows the flowchart of the methodology.

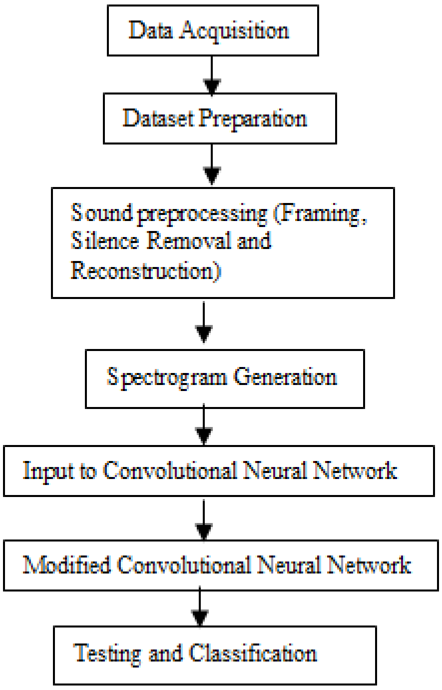


Figure. 5. Flowchart of the Methodology of paper [*5*]

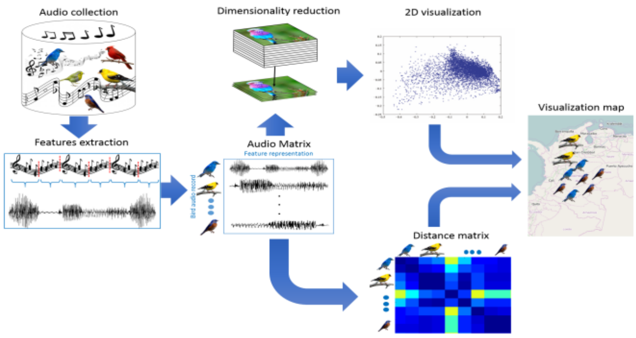


Figure. 6. Flowchart of the Methodology [*20*]

The shortfall of this work is, a group of bird species distinguishing is not possible. A short description of this paper [*31*], is wild birds call has been recorded and transformed into frequency than using wavelet transformation de-noised the signal. Signal de-noising has been done using soft threshold and hard threshold techniques. Then mean value has been calculated for that frequency, which was trained in the three-layer neural network for classification.

The limitation of this research work, the dataset is inadequate for neural networks classifier. Moreover, wavelet modulation transform performance degrades when the frequency is too high. The result after the hard threshold is somewhat good without it.

A brief discussion on paper [*32*], bird call has been recorded, and then signal processed and feature extracted. Then it was trained and tested using a machine learning technique for identification. The purpose of this work is, a large no of bird species has been taken into consideration and comparative performance analysis has been measured between pulses and raw recorded data. Two audio datasets have been experimented with; one recoded bird call and another divided into pulses with high amplitude. It shows high-performance accuracy by using the second one because of less environmental noise in the birdcall.

A limitation of this work is, the clustering of different types of birds is hard to recognize.

In this work [*33*], a comparative analysis within the different classification techniques like probabilistic, instance-based, decision trees, neural networks, and support vector machines have been experimented with using bird audio signal features. The training and classification phase are in Figure 7 and Figure 8.

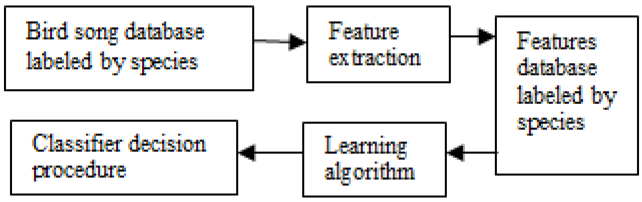


Figure. 7. Training Phase of paper [*33*]

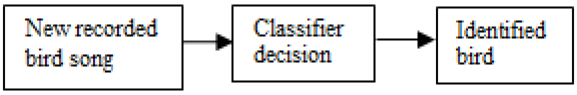


Figure. 8. Classification Phase of paper [*33*]

The limitation of this work is dataset is not up to the mark. Here, only 3 bird species have been taken into consideration with 35 bird audio calls. So, to improve the accuracy of the result larger dataset must be taken.

2) Using Mel-Frequency Cepstrum Coefficients (MFCC) KNN and K-means clustering technique

In a brief discussion on this paper [*11*], bird call has been recorded, segmented and features extracted by matrix factorization method than using k-means clustering algorithm the birds sound has been classified. Multiple bird species have been evaluated and clustered through data-driven characteristics for clustering bird sounds. The following flowing diagrams are the framework for training and testing are shown in Figure 9 and Figure 10.

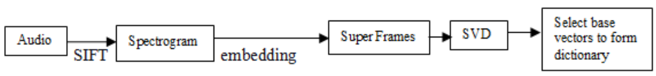


Figure. 9. Framework for training of work [*11*]

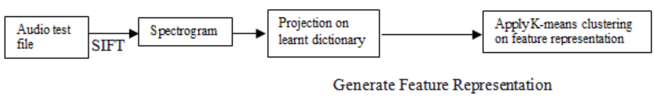


Figure. 10. Framework for testing of work [*11*]

The limitation of this work is mainly concerned with the clustering of different bird calls, but due to the lack of a multilevel dataset, the synthesized dataset has been generated, which may show an inaccurate result.

In a short description of the paper [*24*], two sets of bird calls have been recorded, and acoustic features have been extracted using the permutation features. Then, these features have been classified using a k-NN-classifier. To decrease the noise, the segmentation of these calls has been done by an iterative time-domain algorithm.

A limitation of this paper is, the PPF-matrix smoothing increases computational complexity. In association with this, more attention has to give to finding only those temporal patterns that differentiate between different species of birds. It will increase accuracy and decrease computational complexity.

A brief description of the work [*30*], three groups of bird's audio input have experimented for identification employing two classifiers-KNN (k Nearest Neighbor) and SOM (Self Organizing Maps).

The limitation of this paper is a comparative study between KNN and SOM. One main problem is the choice of features. The input audio sound has been extracted from nature. So, lots of noise was there. The dataset needs to improve for a more accurate result.

Brief discussion in this paper [*35*], bird call has been recorded, segmented using the time-domain technique, and using Mel-frequency cepstrum coefficients (MFCC) features has been extracted. These features are then evaluated by principal components analysis (PCA) and K-Nearest Neighbors (k-NN) to identify the same.

A limitation of this work, some birds have been identified correctly. But, due to a lack of data, this system may show an inaccurate result.

3) Using SVM technique

A brief description of this work [*12*], bioacoustics monitoring was applied on five bird species. The bird calls were recorded, pre-processed, and segmented. Then these segments were then trained and tested with an SVM classifier to classify the bird species. The overall design or methodology is in Figure 11.

The shortfall of this research work is that when this system is implemented in an environment, lots of sound segments will not be relevant to trains, such as other bird calls, amphibian calls, and another animal sound, wind sound. Moreover, the time and space complexity of this system will be very high.

A brief discussion about the work [*13*], through Gaussian mixture model (GMM) energy detector, recorded bird calls have been passed to detect auditory parameterization in birdcall for which spectral pattern and texture features have been taken out. Then this feature has been run in a Relief-based algorithm to find out distinctive features, which have been passed to support vector machine for classification of the call. The overall flow diagram of the method has been designed and shown in Figure 12.

The limitation of this paper is recorded bird call has been segmented into isolated events. Each event refers to either a call or a syllable. They have considered any one type of bird sound, which is the main limitation.

In a short discussion on this paper [*17*], four classes of bird calls have been recorded, and for each recording, a Cepstral feature matrix on Mel scale has been generated, which has been trained and tested on support vector machines for classification. The flow diagram of the methodology and detailed feature extraction method is in Figure 13 and Figure 14, respectively.

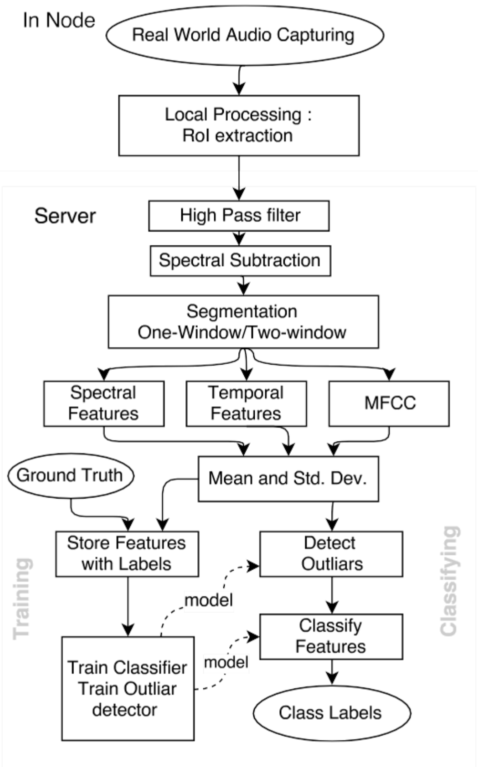


Figure. 11. Overall system design [*12*]

The limitation with this technique is, the classification was performed on a noisy dataset. Only four classes of bird species have been considered, which is not adequate for the machine learning algorithm.

A short description of this paper [*19*], a comparative study of different frequency domain analysis techniques has experimented on bird species call identification.

The limitation of this work is, the execution time will vary depending upon the size of the audio data file. Besides this, the dataset is not up to the mark.

A brief description about this paper [*23*], different birds found near the airport has been identified using its calls. Birdcall has been recorded and segmented in pulses using high amplitude. These pulses have been trained and tested through the SVM classifier and obtained the required result.

The limitation of this paper shows a manual segmentation of the bird sounds. It was assumed, the pulses contain no external noise and integrate relevant features of bird calls, which may not always be correct for real-time automatic bird species identification methods.

1) Using miscellaneous technique

A brief description of the paper [*6*], from Borneo region, birds' calls have been recorded, and features have been extracted, then dimensionality reduced using Linear Discriminant Analysis (LDA) and then classified and identified using Nearest Centroid (NC) classifier. In Figure 15 the flow diagram of the classification approach has been delivered.

A brief discussion on this paper [*14*], bird call has been recorded, pre-processed, frequency-based segmentation applied and syllable generated and using AMFCC, DELTA, and DELTA-DELTA features, then it has been trained using machine learning algorithms and recognized by Linear Discriminant Analysis (LDA) with probabilistic approach on four individual bird species. The class probabilities and recording level identification method have shown more accurate results than the syllable-based identification method.

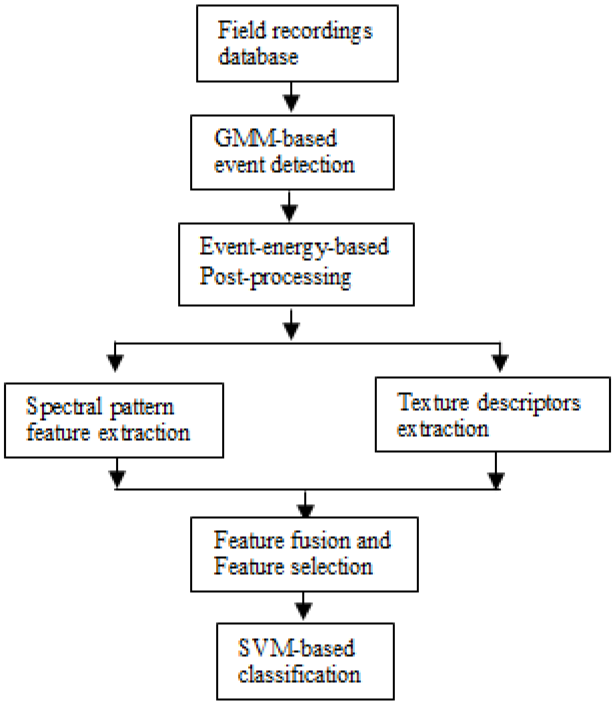


Figure 12. Overall flow diagram of the method of work [*13*]

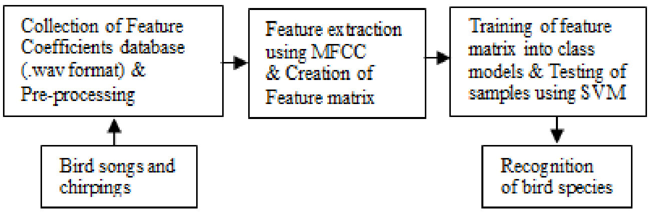


Figure. 13. Methodology flow of research work [*17*]

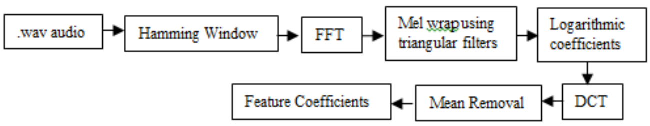


Figure. 14. Feature extraction method of research work [*17*]

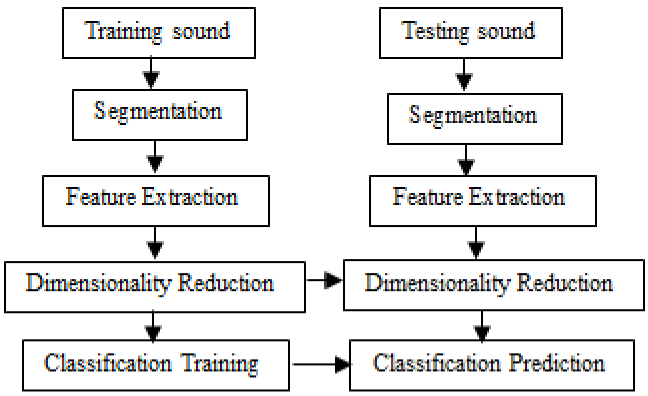


Figure. 15. Classification Approach [*6*]

The limitation in this work is, raw sound data has been recorded for four particular bird types. So, the dataset is not adequate to have better results, and if multiple birds of the same species or different species will chatter with different frequencies, then the system will not work and may not be able to identify the bird species.

A brief description of this work [*18*], the ROC performance analysis uses Random Decision Tree, which was proved better on the classification system of bird species or subspecies from its calls. In this work, audio data was processed using an image processing technique. A spectroscope image of an audio data has been examined, and relevant feature of the audio signal has been extracted for train and test using the machine learning method.

The main limitation is, the dataset is not up to the mark.

A short description of the paper [*27*] has enlightened the hierarchical bird species classification method using a 74 bird species call. It is a comparative study between three types of classification approaches the flat classification approach, the local-model per parent node classifier approach employing a classic Naive Bayes algorithm, and the global-model hierarchical-classification approach by the Global Model Naive Bayes (GMNB) algorithm.

The main limitation is, this system may fail to build a cluster of birds present in a group.

Brief discussion in this paper [*28*], wild birds calls were recorded, pre-processed, segmented, and feature extracted, then these features have been trained and tested in FFT and to the 4-layer neural network and compared the result. The flow of work is presented in Figure 16.

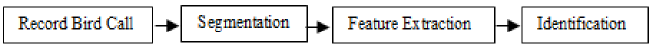


Figure. 16. flow diagram of work [*28*]

The limitation of this research work is, the dataset is inadequate for the neural network. Noise reduction methodology is missing.

A brief description of work [*29*], on thirteen wild birds call, has been recorded, pre-processed, and feature extracted. Then wavelet transform has been applied mean value, the strength of the frequency and modulation spectrum has been calculated then these features have been trained and tested through Neural Network and classified bird call.

The limitation of this research work, the dataset is not up to the mark for the neural network. One more shortfall is that, for the sake of simplicity, they have assumed an ideal environment for data recording.

A short discussion on this work [*34*], birdcall collected from a noisy environment. Then in this input audio signal, a Hamming window has been set, and a short time FFT has been applied and the frames transformed into a time-frequency domain spectrogram form then Random Forest classifier has been implemented for classification.

The limitation is, the dataset is noisy. So, after segmentation, 90% of noise examples have been produced and marked as negative examples. Moreover, below 1 kHz frequency ranges were ignored. This method must use a smaller set of features so that the runtime could decrease. It would be appropriate for a larger dataset.

B. Bird species identification using Image

In this segment of work, a brief discussion has been presented on the research work associated with bird species identification using its image. Various methods and their limitation have been discussed in the following subsections.

1) Using Deep CNN Technique

A brief discussion on this work [*3*], A very efficient technique, the Residential Network model as a pre-trained Convolutional Neural Network, has been used for feature extraction has been used to train and test the input image to identify the single bird at a time. The following Figure 17 delivers the base deep CNN network model used to extract input image features. Figure 18 presents pre-trained model of RestNet18, ResNet34, ResNet50, ResNet101 for transfer learning sequentially.

The limitation of this research work, the collected dataset is limited to have greater accuracy. The weights for the base model have been selected randomly, which draws degraded performance.

A discussion on this research work [*4*] firstly, input bird image has been converted into a grey-scale format, and autograph produced then examining this autograph, the score sheet was calculated and after analyzing the score sheet through Deep Convolutional Neural Network (DCNN) on Google-Net framework bird species has been identified.

The limitation of this paper is this, sometimes cannot compare among the different types of bird species or misclassify the bird species due to the visual similarities of color and size between and among the bird species. The dataset has to restructure with more features of the birds to identify bird species.

In this paper [*10*], a very low-resolution bird image, the small object has been detected by deep learning methods. A new dataset was collected from an imagery dataset constructed with real-life bird images.

The limitation of this paper is, the dataset used here is an imagery dataset. To distinguish the irrelevant background from the bird image only is sometimes problematic because the resolution is poor. So, the accuracy of the model can show poorly.

A brief discussion on the research work [*16*], a bird image with the image taken by the Radar has been trained and tested by Deep Convolutional Neural Network for its identification. This system is applicable when the training dataset is low. Here, data expansion has been applied, which improves the performance of the work. The limitation of this work is, here combination of radar-produced image and digital image data has been used as input. But the data provided by the radar generates additional information which can identify alone a misclassified species. In association, the dataset is inadequate from a deep learning perspective.

1) Using VGG-16, SVM and miscellaneous technique

In the research paper [*8*], Bangladeshi bird image has been trained and tested using the VGG-16 network for classification. A comparative study has also been made to other classification methods such as Random Forest, K-Nearest Neighbor, and Support Vector Machine. The flow diagram of the system is in

Figure 19. The main limitation is an insufficient dataset. It may not give an accurate result or may show erroneous results sometimes.

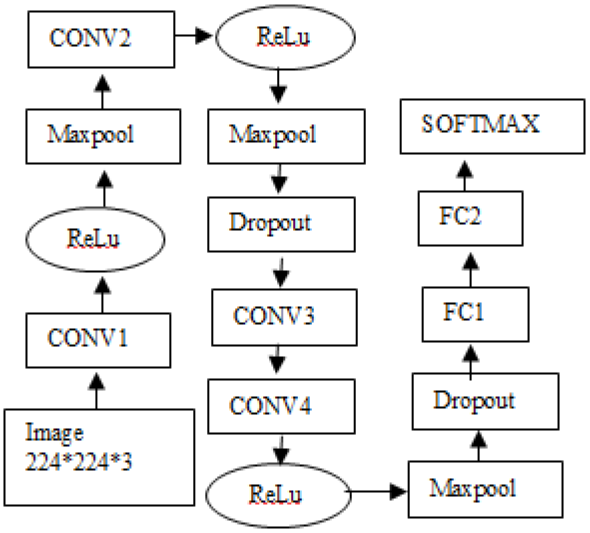


Figure. 17. Base model of paper [*3*]

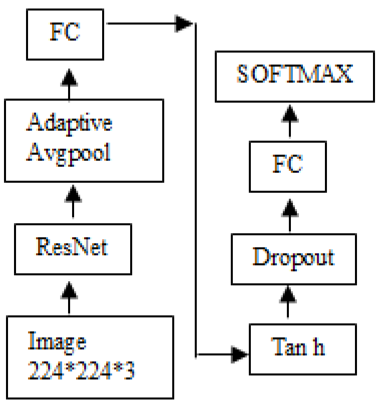


Figure. 18. Pre-trained CNN model of paper [*3*]

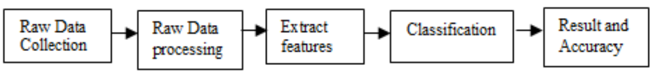


Figure. 19. Flow diagram of the system of paper [*8*]

In this work [*15*], hundred images of snowy owl and toucan have been segmented and examined nine color-related features like mean, standard deviation, and skewness of red, green, and blue (RGB) and then classified using a Support Vector Machine algorithm.

The limitation of this research work, the color-based feature has been extracted from only two types of bird species. The total no of images of snowy owl and toucan is utilized as the test images, which is not up to the mark. The result may be error-prone. The extraction of the color feature from an image will depend on the background because the same color background can lead to an inaccurate result.

In this paper [*22*], one-vs-most classifiers have been used for fine-grained image classification and clustering from a large dataset. This work is mainly suitable when there are highly similar bird species in the classification problem. Similar bird species are rejected during training.

The limitation of this work in this research is that the accuracy of the result can be increased by using manually labeled parts detection only.

In this work [*25*], visual features of the bird body, color characteristics, shape, and size have been examined for bird species identification. Bird species have been identified using similarity comparison by visual feature analysis.

The limitation of this work is that a single feature value has been adopted, which may cause performance degradation.

A brief discussion on this paper [*26*], the color feature has been extracted from the bird image. Then segmentation has been done to eliminate background elements next, splitting the image into component planes, normalized color histograms are computed, which has been used by the learning algorithm to distinguish bird species.

In the limitation of this research work, RGB and HSV color spaces have been used. But for bird species classification, this technique is not showing an impressive result. So, when the numbers of classes are high, then segmentation is not required in such a problem.

C. Bird species identification using image and call

In this approach, both bird images and bird calls have been used to identify and classify birds. The reason for using both the bird image and call is the diversity of nature creates an obstacle to have data in proper form and format.

A brief description of work in [*21*], bird species have been identified based on both image and sound, where visual features taken out using Scale Invariant Feature Transform (SIFT) method from unconstrained bird images, then trained by a support vector machine classifier and also acoustic features taken out using Mel-frequency Cepstral coefficients (MFCCs) from bird vocalizations. In this approach, visual classification has been applied, and if the visual classification has been rejected or failed, then acoustic monitoring has been examined to identify the bird.

In this work, whether a bird image will be accepted or rejected is calculated on the output vector

$[P(\omega_1|\hat{x}), \ldots., P(\omega_c|\hat{x})\,]$ From [*21*]

Case 1: the bird species image hypothesis is accepted whenever $P(\omega'|\hat{x}) \gg \lambda$, From [*21*]

Case 2: the bird species image hypothesis is rejected whenever $P(\omega'|\hat{x}) < \lambda$, From [*21*]

If an image is rejected then it is re-classified using the acoustic data if available on the verification stage.

The limitation of this work [*21*] is that a smaller part of the total instances has both pictorial and auditory material which is very less to train and test in a deep learning environment. The rate of accuracy is very low.

V. Comparative Analysis and Outcome

In this section, a detail comparative analysis of different research works has been presented in Table 4. It will help researchers to gain knowledge about the different models of bird species identification. Here, for every individual research work, a cluster-wise summary has been delivered as shown in

Table 5. In this cluster, different subsections are there and every subsection has been accommodated with several works of similar logic.

A bioacoustics data analysis on different species has been established in the research paper [*41*]. Here, several methods of species identification using its vocal data have been discussed. A part of the bird species bioacoustics model has been analyzed. In comparison with this paper, the proposed system has some significant contributions as follows:

Delivers a complete review of state-of-the-art research works in the Bird Species Identification domain. It will help to figure out the amount of work done in the last few years.

Precisely organize into the different clusters: Bird species identification by visual monitoring, Bird species identification by bioacoustics monitoring, Bird species identification by audio-visual monitoring. It will tell the researchers the different ways out in this domain specifically.

Cluster-wise comprehensively summarizes the workflow and delivers relevant detailed studies about the methodology used, feature used, the dataset used, the accuracy of every research work. It will help to gather accurate bits of knowledge. It will also help to analyze and compare the different models used so far. Presents a comparative and comprehensive analysis of the research studies used for this work, which will show the summary of the investigation at a glance. Collectively describes different knowledge gaps and research targets for the future, which will help the researchers to work further in this domain.

VI. Discussion

In this section, the scope of the research work on bird species identification has been discussed in detail. Three different clusters have been elaborated, and a lot of future work is there in these clusters where more attention has to take to improve the accuracy of the result.

1. The dataset is very poor for some of the works, so the result may sometimes be inaccurate when trained and tested in a machine learning model.
2. More care has to take for fine-tuning performance parameters.
3. The weights of the hidden layer of the neural network have to initialize and update to improve the performance.
4. Input features should have to be optimized to have a better result.
5. More improvement should have to be included in strategies to have more success rate.

VII. Conclusion

Bird species identification is a topic of interest for global researchers recently. The objective of this study is to open a research window for new researchers intended to work in this domain. In this communication, the various models of bird species identification and classification systems have been investigated systematically by reviewing past and recent approaches used in the last ten years in search of the scope of further research work on this topic. The various knowledge gaps of these existing models have been summarized in this paper after deep meticulous scrutiny. Three clusters are about the categorization of published works intended for a particular scope of bird species identification. It will assist the researcher by providing a section-wise review in this domain. In this paper, a total of thirty-eight communications were considered for this survey. From this detailed study, it is seen that very few works have been done yet in clusters 2 and 3. If there are different types of birds in a group, then clustering those birds is yet not unfolded. This review has identified several targets for future research in this area. This paper can serve as a reference point for future work in this area to provide insight into work to date, pointing towards the key aspects that merit further investigation.

Acknowledgment

We hereby acknowledge the Department of Computer Science and Engineering, Guru Nanak Institute of Technology, Kolkata, India for providing the guidance and infrastructure for conducting the research.

Reference s	**Method or Algorithm Used**	**Features**	**Dataset**	**Accuracy**
[*1*]	Convolutional Neural Networks	Acoustic monitoring	Aras River Bird Sanctuary	Accuracy rates between 81.47% to 92.08%
[*2*]	Power spectral density, Artificial Neural Network (ANN)	Bird call feature by preprocessing technique.	Bird sound collected from multiple sources.	The test accuracy 95%
[*3*]	Pretrained ResNet model as pretrained CNN networks with base model	The bird images transformed into 224*224*3 to ensure the same dimensions	Asian countries western dataset	The test accuracy 97.98%
[*4*]	CNN model with skip connections	Feature vector of bird image	27 bird species images of Taiwan region collected using the Internet of Birds (IoB) mobile app.	The sensitivity, specificity, and accuracy were 93.79%, 96.11%, and 95.37%.
[*5*]	Convolutional Neural Network (CNN) classifies	Spectrogram from preprocessed sound clip	Acquire 400 samples of 4 species of bird sound	Accuracy rate - 97%.

			recordings	
[6]	Linear Discriminant Analysis (LDA), Nearest Centroid (NC)	Extracted sound features like Zero Crossing Rate (ZCR), Energy (E), Entropy of energy, Spectral Centroid, Spectral Spread, Spectral Entropy, Spectral Flux, Spectral Roll-off, Mel Frequency Cepstral Coefficients (MFCCs), Chroma Vectors, Harmonic Ratio, Fundamental Period	Borneo regional birds from their sounds	Accuracy rate - 96%.
[7]	Zero-frequency filter, Hilbert envelope, Support vector machine, Deep neural network.	Sound feature	Bird calls recordings collected at real time	Accuracy: Linear kernel-based SVM - 80.93% RBF kernel-based SVM- 98.41%
[8]	VGG-16 network	1600 bird images of dimensions 550 x 600	Bangladesh region bird species images collected from the internet.	The accuracy is 89%.
[9]	Hidden Markov models (HMMs), hybrid Deep Neural Network-hidden Markov model.	Bird calls features.	Bird call recorded in real time.	Identification accuracy-98.7% , error rate - 2.7%, multiple bird species recognition accuracy- 97.3% and 95.4%
[10]	Deep learning, such as YOLOv2, SSH, and Tiny Face, U-Net and Mask, R-CNN.	Birds color, shapes, poses, resolutions, scales, and backgrounds	Little Birds in Aerial Imagery (LBAI), created from real-life aerial imagery data.	SSH attained the highest F1 score of 92.5%.
[11]	K-means clustering, SVD, matrix factorization	Bioacoustics features extracted by SVD	The Great Himalayan National Park (GHNP) dataset	F-Score – 84%
[12]	Support Vector Machine classifier	Sound features	Real world bird audio capturing	Accuracy rate - 93.85%
[13]	Gaussian mixture model (GMM) based energy detector, ReliefF-based feature selector, Support Vector Machine	Spectral pattern and Texture features of bird audio recording	The Xeno-canto Archive (http://www.xeno-canto.org/)	Accuracy rate - 96.7%
[14]	Mel- Frequency Cepstral Coefficients, Linear Discriminant Analysis (LDA) with probabilistic approach.	Syllable feature vector	The database created with common bird species songs found in geographic region of India	Average accuracy rate: Detection – 97.15% Classification- 97.12%
[15]	Support Vector Machine algorithm	Color based features of bird images.	Dataset downloaded from Computer Vision research	For training - 97.14% accuracy for testing - 98.33%.
[16]	Convolutional Neural Network (CNN) for feature extraction	Segmented image, size evaluation taking parameter from radar.	Dataset collected from offshore wind farms in Finland.	TPR - between 0.86–0.99
[17]	MFCC features extractor, classifier Support Vector Machine.	Sound based feature matrix	Bird calls collected native to India	The classifier accuracy- 89.4%

[18]	Random Forest Regressor based Classification.	Spectroscope image of bird audio file.	BIOTOPE society (https://www.advantest.com/csr/biotope-and-birdpia)wildlife recordings of birds in Europe.	Accuracy rate - 96.2%
[19]	Mean Square Error (MSE) approach, Correlation Analysis, Mel Frequency Cepstral Coefficients approach	Bird call spectrum	The Xeno-canto Archive (http://www.xeno-canto.org/)	Mean square error- 0.000
[20]	Mel-Frequency Cepstral Coefficients, 2D Visualization, Visualization Map, Principal Component Analysis.	Distance matrix	The Xeno-canto Archive (http://www.xeno-canto.org/)	Performance has been substantially enhanced.
[21]	Scale Invariant Feature Transform, Mel-frequency cepstral coefficients, Support Vector Machine classifier	Visual feature, audio features	Bird Images - CUB200-2011 database Bird audio- Xeno-Canto database	Improvements 1.2 - 15.7 percentage points
[22]	One-vs-Most Classifiers	Large no of fine-grained bird images	Birdsnap dataset available at birdsnap.com.	Accuracy at rank 1- 79.9% at rank 5 - 95.1%
[23]	SVM classifier, MARSYAS framework	Feature vector	75 species bird songs from Southern Atlantic Coast of South America	Overall accuracy- 52.78%
[24]	K-NN-classifier, MFCC	Feature vector	The bird sounds collected from numerous sources.	Recognition rates - 61% and 59%.
[25]	Similarity comparison algorithm	Standard images feature for bird species like bird body, color, characteristics and shape	Image collected from different sources.	Performance has been substantially enhanced.
[26]	HSV and RGB color segmentation approach	color features	CUB-200 dataset.	Accuracy rate- 90%
[27]	Hierarchical Naive Bayes classifier, MARSYAS framework	64 feature set using Mel-Frequency Cepstral coefficients.	Xeno-canto web site, the South Atlantic Coast of Brazil bird species.	Accuracy rate- 98.39%
[28]	FFT with frequency band network. CNN	Audio features	The bird sounds collected from numerous sources.	Accuracy- 96.3%
[29]	Wavelet transform, Neural Network.	Audio features	Nocturnal wild bird species	Average identification rate - 90%
[30]	Classifiers -KNN and SOM	Audio features	Bird calls collected from natural environment	Performance has been enhanced.
[31]	Wavelet transform, neural network	Modulation spectrum, mean value	Bird calls collected from natural environment	The recognition result- 0.9776772- 0.9821246
[32]	MARSYAS framework, MLP and SMO classifiers.	Audio features	Southern Atlantic Brazilian Coast bird species audio song collection	F-measure- 95.1%
[33]	Probabilistic, Instance-based, decision trees, Neural Networks and Support Vector Machines	The MARSYAS feature The IOIHC feature set The Sound Ruler feature set	The bird sounds collected from numerous sources.	Best accuracy rate- 98.39%
[34]	Random Forest Training.	Audio features	The bird sounds collected from numerous sources.	Accuracy rate-93.6% False positive rate- 8.6%
[35]	Mel-frequency cepstrum coefficients, Principal Component Analysis,	Audio features	The bird sounds collected from numerous sources.	Accuracy rate- 86%- 90%

	k-Nearest Neighbor			
[36]	Deep Convolutional Neural Network	Bird images of resolution 256 × 256	1.2 million high-resolution images in the ImageNet LSVRC-2010 dataset.	Error rates of 37.5% and 17.0%
[37]	Deep residual learning framework	Bird image preprocessed and normalised	ImageNet dataset.	3.57% top-5 error rate on the test dataset
[38]	A novel graph-based clustering algorithm	Bird image normalized by the pose.	CUB-200-2011 dataset.	Accuracy – 75.7%

Table. 4. Comparative analysis.

Cluster #	**Cluster Name**	**Sub clusters**	**References**	**Summary of each paper**
Cluster 1	Bird species identification using call	Using Deep CNN technique	[1]	This is a comparative analysis of raw bird calls with processed bird calls trained and tested in the Deep CNN model. In this evaluation, it has been shown that pre-processed calls can generate better accuracy than raw call datasets.
			[2]	Individual bird call has been recorded and pre-processed with Power Spectral Density and then trained and tested using ANN.
			[5]	Here, in this work, bird call recorded, pre-processed, framed, silently removed, reconstructed, and spectrogram generated and then trained and tested through CNN and classified
			[20]	In this paper, bird calls have been recorded, pre-processed, feature extracted, then 2D visualization technique has been applied, after that trained and tested by machine learning technique to identify bird species.
			[31]	Wild bird call has been recorded and de-noised using wavelet transformation. Then mean value has been calculated for that frequency which is trained in the neural network for classification.
			[32]	In this paper, bird calls are recoded pre-processed, feature extracted, and classified using a neural network. It has experimented with whether performance accuracy is improving by dividing the call into pulses.
			[33]	A comparative analysis of probabilistic, instance-based decision trees, neural networks, and support vector machines has been experimented with using bird audio signal features.
		Using Mel-Frequency Cepstrum Coefficients (MFCC)KNN and K-means clustering technique	[11]	Birdcall has been recorded, segmented, and features extracted by matrix factorization method than using k-means clustering algorithm the birds sound has been classified.
			[24]	In this work, two different sets of bird calls have been recorded, and acoustic features have been extracted using the permutation feature and have been classified using a k-NN-classifier.
			[30]	Here, three groups of bird's audio input have been experimented with for identification employing two classifiers-KNN (k Nearest Neighbor) and SOM (Self Organizing Maps).
			[35]	Birdcall has been recorded, segmented using the time-domain technique, and using Mel-frequency Cepstrum coefficients (MFCC) features have been extracted. These features are evaluated by principal components analysis (PCA) and K-Nearest Neighbours (k-NN) to identify the same.
		Using SVM technique	[12]	Bioacoustics monitoring has been applied on five bird species by recording the calls, then those calls have been pre-processed, segmented, trained, and tested with an SVM classifier to classify the bird species.
			[13]	In this paper, through Gaussian mixture model (GMM) energy detector recorded bird calls have been passed to detect auditory parameterization in birdcall for which spectral pattern and texture features have been taken out, then this feature has been run in a Relief F-based algorithm to find out distinctive features which have been passed to support vector machine for classification the call.
			[17]	In this work, four classes of bird calls have been recorded, and for each recording, a Cepstral feature matrix on Mel scale has been generated, which has been trained and tested on support vector machines for classification.
			[19]	A comparative study of different frequency domain analysis techniques has

				been examined for bird species call identification.
			[23]	Different birds found near the airport have been identified using its calls. Birdcall has been recorded and segmented in pulses using high amplitude. These pulses have been trained and tested through the SVM classifier and obtained the required result.
		Using miscellaneous technique	[6]	Borneo regional birds' calls have been recorded and features have been extracted, then dimensionality reduced using Linear Discriminant Analysis (LDA) and then classified and identified using Nearest Centroid (NC) classifier.
			[7]	The average amount of epochs per second was calculated using zero-frequency filtering. Then it has been discriminated bird call made by a single bird or made by multiple birds.
			[9]	Acoustic signal from different bird species pre-processed and then using hidden Markov models (HMMs) and unsupervised modelling has been identified a particular bird species.
			[14]	In this paper, bird call has been recorded, pre-processed, frequency-based segmentation applied and syllable generated and using AMFCC, DELTA, and DELTA-DELTA features generated. Then it has been trained using machine learning algorithms and recognized by Linear Discriminant Analysis (LDA) with a probabilistic approach on four individual bird species.
			[18]	In this work, the ROC performance analysis uses a Random Decision Tree, which has been proved better for bird species or subspecies classification from its calls.
			[27]	It is also a comparative study between three approaches viz the flat classification approach, the local-model per parent node classifier approach employing a classic Naive Bayes algorithm, and the global-model hierarchical-classification approach by the Global Model Naive Bayes (GMNB) algorithm.
			[28]	Wild bird call has been recorded, pre-processed, segmented, and feature extracted, then these features have been trained and tested in FFT and to the 4-layer neural network and compared the result.
			[29]	On thirteen wild birds, the call has been recorded, pre-processed, and feature extracted. Then wavelet transform has been applied and means value, the strength of the frequency and modulation spectrum has been calculated then these features has been trained and tested through Neural Network and classified bird call.
			[34]	Birdcall has been recorded from a noisy environment. Then audio segmentation has been performed on that call. Then Random Forest classifier has been applied for classification.
Cluster 2	Bird species identification using image	Using Deep CNN technique	[3]	A very efficient technique, the Residential Network model as a pre-trained Convolutional Neural Network, which was used for feature extraction, has been used to train and test the input image to identify a single bird at a time.
			[4]	Input bird image has been converted into a greyscale format and autograph produced, then examining the autograph the score sheet has been calculated and after analyzing the score sheet through Deep Convolutional Neural Network (DCNN) on Google-Net framework bird species has been identified.
			[10]	In this paper, a small object low-resolution bird image has been detected by deep learning methods. In this paper, a new dataset has been created with the imagery of real-life bird images.
			[16]	In this paper, a bird image combining with the bird image taken by radar has been trained and tested by a Deep Convolutional Neural Network for its identification.
		Using VGG-16, SVM and miscellaneous technique	[8]	Bangladeshi bird image has been trained and tested using the VGG-16 network for classification. A comparative study has also been made to other classification methods such as Random Forest, K-Nearest Neighbour, and Support Vector Machine.
			[15]	Snowy owl and toucan bird images have been segmented and examined nine color-related features like mean, standard deviation, and skewness of red, green, and blue (RGB) and then classified using a Support Vector Machine algorithm.
			[22]	One-vs.-most classifiers have been used for fine-grained image classification and clustering from a large dataset.
			[25]	Bird species have been identified using similarity comparison by visual feature

				analysis.
			[26]	The color feature has been extracted from the bird image. Then segmentation has been done to eliminate background elements next, splitting the image into component planes, normalized color histograms are computed, which has been used by the learning algorithm to distinguish bird species.
Cluster 3	Bird species identification using image and call both	--	[21]	In this paper, bird species have been identified based on both image and sound. The visual features were taken out using Scale Invariant Feature Transform (SIFT) method from unconstrained bird images. Then it was trained by a support vector machine classifier. The acoustic features were taken out using Mel-frequency Cepstral coefficients (MFCCs) from bird vocalizations. The visual classification has been applied, and if the visual classification has been rejected or failed, then acoustic monitoring has been examined to identify the bird.

Table. 5. Summary of the Survey.

References

[1] S. Bayat and G. Işık, "Identification of Aras Birds with Convolutional Neural Networks," in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2020, pp. 1-5.

[2] M. M. M. Sukri et al., "Bird Sound Identification based on Artificial Neural Network," in *2020 IEEE Student Conference on Research and Development (SCOReD)*, 2020, pp. 342-345.

[3] K. M. Ragib et al., "PakhiChini: Automatic Bird Species Identification Using Deep Learning," in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, London, 2020, pp. 1-6.

[4] P. Gavali and J. S. Banu, "Bird Species Identification using Deep Learning on GPU platform," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, Vellore, 2020, pp. 1-6.

[5] B. Chandu, A. Munikoti, K. S. Murthy, G. Murthy V., and C. Nagaraj, "Automated Bird Species Identification using Audio Signal Processing and Neural Networks," in *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*, Amaravati, 2020, pp. 1-5.

[6] M. Ramashini, P. E. Abas, U. Grafe, and L. C. De Silva, "Bird Sounds Classification Using Linear Discriminant Analysis," in *2019 4th International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)*, Jaipur, 2019, pp. 1-6.

[7] R. Sinha, V. Vadluri, A. Arya, and P. Rajan, "Single versus Multi-Source Discrimination in Birdcalls using Zero-Frequency Filtering," in *2019 National Conference on Communications (NCC)*, Bangalore, 2019, pp. 1-6.

[8] S. Islam, S. Ibn Ali Khan, Md M. Abedin, K. M. Habibullah, and A. K. Das, "Bird Species Classification from an Image Using VGG-16 Network," in *ICCCM 2019*, Bangkok, Thailand, 2019, pp. 27-29.

[9] P. Jancovic and K. Münevver, "Bird Species Recognition Using Unsupervised Modeling of Individual Vocalization Elements," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 5, no. 27, pp. 932-947, 2019.

[10] P. Sun et al., "Performance Comparison of Deep Learning Techniques for Recognizing Birds in Aerial Images," in *2018 IEEE Third International Conference on Data Science in Cyberspace*, Guangzhou, China, 2018.

[11] H. Seth, R. Bhatia, and P. Rajan, "Feature Learning for Bird Call Clustering," in *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, Rupnagar, 2018, pp. 72-76.

[12] H. Weerasena et al., "Continuous Automatic Bioacoustics Monitoring of Bird Calls with Local Processing on Node Level," in *TENCON 2018 - 2018 IEEE Region 10 Conference*, Korea, 2018, pp. 235-239.

[13] S. Zhang, Z. Zhao, Z. Xu, K. Bellisario, and B. C. Pijanowski, "Automatic Bird Vocalization Identification Based on Fusion of Spectral Pattern and Texture Features," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Canada, 2018, pp. 271-275.

[14] R. P. Tivarekar and H. G. Virani, "Species recognition using audio processing algorithm," in *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, Tirunelveli, 2017, pp. 527-532.

[15] R. Roslan, N. A. Nazery, N. Jamil, and R. Hamzah, "Color-based bird image classification using Support Vector Machine," in *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, Las Vegas, 2017, pp. 1-5.

[16] J. Niemi and J. T. Tanttu, "Automatic bird identification for offshore wind farms: A case study for deep learning," in *2017 International Symposium ELMAR*, Zadar, Croatia, 2017, pp. 263-266.

[17] P. Rai, V. Golchha, A. Srivastava, G. Vyas, and S. Mishra, "An automatic classification of bird species using audio feature extraction and support vector machines," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2016, pp. 1-5.

[18] S. Debnath, P. P. Roy, A. A. Ali, and M. A. Amin, "Identification of bird species from their singing," in *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, Coimbatore, India, 2016, pp. 182-186.

[19] P. C. Aparna, "Automatic Recognition of Birds through Audio Spectral Analysis," in *2015 Fifth International Conference on Advances in Computing and Communications (ICACC)*, Kochi, India, 2015, pp. 395-398.

[20] A. K. Reyes and J. E. Camargo, "Visualization of audio records for automatic bird species identification," in *2015 20th Symposium on Signal Processing, Images and*

Computer Vision (STSIVA), Bogotá, Colombia, 2015, pp. 1-6.

[21] A. Marini, A. J. Turatti, A. S. Britto, and A. L. Koerich, "Visual and acoustic identification of bird species," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Queensland, Australia, 2015, pp. 2309-2313.

[22] T. Berg et al., "Birdsnap: Large-Scale Fine-Grained Visual Categorization of Birds," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA , 2014, pp. 2019-2026.

[23] T. L. F. Evangelista, T. M. Priolli, C. N. Silla, B. A. Angelico, and C. A. A. Kaestner, "Automatic Segmentation of Audio Signals for Bird Species Identification," in *2014 IEEE International Symposium on Multimedia*, Taichung, Taiwan, 2014, pp. 223-228.

[24] S. Fagerlund and U. K. Laine, "New parametric representations of bird sounds for automatic classification," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, 2014, pp. 8247-8251.

[25] L. Jian, Z. Lei, and Y. Baoping, "Research and Application of Bird Species Identification Algorithm Based on Image Features," in *2014 International Symposium on Computer, Consumer and Control*, Taichung City, Taiwan, 2014, pp. 139-142.

[26] J. Facon, A. M. Koerich, and A. Marini, "Bird Species Classification Based on Color Features," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, vol. IEEE, Manchester, United Kingdom, 2013.

[27] C. N. Silla and C. A. A. Kaestner, "Hierarchical Classification of Bird Species Using Their Audio Recorded Songs," in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, United Kingdom, 2013, pp. 1895-1900.

[28] R. Sun, Y. W. Marye, and H. Zhao, "FFT based automatic species identification improvement with 4-layer neural network," in *2013 13th International Symposium on Communications and Information Technologies (ISCIT)*, Samui Island, Thailand, 2013, pp. 513-516.

[29] R. Sun, Y. W. Marye, and H Zhao, "Wavelet transform digital sound processing to identify wild bird species," in *2013 International Conference on Wavelet Analysis and Pattern Recognition*, Tianjin, China, 2013, pp. 306-309.

[30] D. Kamińska and A. Gmerek, "Automatic identification of bird species: A comparison between kNN and SOM classifiers," in *2012 Joint Conference New Trends In Audio & Video And Signal Processing: Algorithms, Architectures, Arrangements And Applications (NTAV/SPA)*, Lodz, Poland, 2012, pp. 77-82.

[31] R. Sun and N. Mita, "Nocturnal wild bird species identification by sound information using wavelet," in *2012 International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP)*, Chengdu, China, 2012, pp. 1-3.

[32] M. T. Lopes et al., "Automatic Bird Species Identification for Large Number of Species," in *2011 IEEE International Symposium on Multimedia*, Dana Point, CA, USA, 2011, pp. 117-122.

[33] M. T. Lopes, A. L. Koerich, C. N. Silla, and C. A. A. Kaestner, "Feature set comparison for automatic bird species identification," in *2011 IEEE International Conference on Systems, Man, and Cybernetics*, Alaska, USA, 2011, pp. 965-970.

[34] L. Neal, F. Briggs, R. Raich, and X. Z. Fern, "Time-frequency segmentation of bird song in noisy acoustic environments," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011, pp. 2012-2015.

[35] V. Gonzalo and R. Domingo, "Using syllabic Mel cepstrum features and k-nearest neighbors to identify anurans and birds species," in *2010 IEEE Workshop On Signal Processing Systems*, San Francisco, USA, 2010, pp. 466-471.

[36] I. Sutskever, G. E. Hinton, and A. Krizhevsky, "ImageNet Classification with Deep Convolutional Neural Networks," *COMMUNICATIONS OF THE ACM*, pp. 84-90, 2017 June.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[38] S. Branson, G. V. Horn, S. Belongie, and P. Perona, "arXiv:1406.2952 [cs.CV],".

[39] (2021, January) IEEE Xplore. [Online]. https://ieeexplore.ieee.org/search/searchresult.jsp?newsearch=true&queryText=bird%20species%20identification.

[40] A. Salih, S. Mahammed, and Ajith Abraham, "A Review of Ambient Intelligence Assisted Healthcare Monitoring," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 5, pp. 741-750, 2013.

[41] R., K. Rama, M. James, S. Garg, and M. Charleston, "Bioacoustics Data Analysis – A Taxonomy,Survey and Open Challenges," *Digital Object Identifier*, vol. 8, pp. 57684-57708, March 2020.

Author Biographies

Nilanjana Adhikari has obtained her B. Tech ('07) from Maulana Abul Kalam Azad University of Technology (formerly known as West Bengal University of Technology), IN and her M. Tech ('11) from University of Calcutta, IN in the Department of Computer Science and Engineering. Presently she is associated with Techno International New Town (Techno India Group) in the capacity of Assistant Professor in the Department of Computer Science and Engineering. She is having a cumulative teaching experience of 14 years. Her research interests lie in Data Mining, Machine Learning and Deep Learning. Currently she is also doing research work and whose registration is due soon.

Suman Bhattacharya, PhD has obtained his B. E. ('04) from Vidyasagar University, IN and his M. Tech ('10) and Ph. D. ('18) from University of Calcutta, IN in the Department of School of Information Technology. Presently he is associated with Guru Nanak Institute of Technology (JIS Group) in the capacity of Associate Professor in the Department of Computer Science and Engineering. He is having a cumulative teaching experience of 15 years and have 02 years of industry experience. His research interests lie in Technology Enhanced Learning, Soft Computing, Machine Learning and Deep Learning, and has over fifteen publications in reputed International Conferences and Journals. Currently he is also supervising a Ph.D. scholar whose registration is due soon. he has also served as reviewer in multiple publication houses. He has also been awarded "Young Achievers Award" from Institute of Scholars.

Mahamuda Sultana, PhD has obtained her B. Tech ('04) from University of Kalyani, IN and her M. Tech ('10) and Ph. D. ('19) from Jadavpur University, IN in the Department of Computer Science and Engineering. Presently she is associated with Guru Nanak Institute of Technology (JIS Group) in the capacity of Associate Professor in the Department of Computer Science and Engineering. She is having a cumulative teaching experience of 14 years. Her research interests lie in Reversible Logic, Quantum Dot Cellular Automata, Machine Learning and Deep Learning, and has over twenty publications in reputed International Conferences and Journals. Currently she is also supervising a Ph.D. scholar whose registration is due soon. She is a Life member of the Institution of Engineers (India) and member of ACM, and has been associated with multiple International Conferences of IEEE and Springer in different roles. She has also served as reviewer in multiple publication houses. She has also been awarded "Young Achievers Award" from Institute of Scholars.