International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 12, No.1 (Oct-2022)

https://dx.doi.org/10.12785/ijcds/120173

Convolutional Neural Network-based Marine Cetaceans Detection around the Swatch of No Ground in the Bay of Bengal

Md. Ariful Islam[1] **and Mosa. Tania Alim Shampa**[2]

[1]*Department of Robotics and Mechatronics Engineering, University of Dhaka, Dhaka-1000, Bangladesh*
[2]*Department of Oceanography, University of Dhaka, Dhaka-1000, Bangladesh*

Received 22 Jan. 2021, Revised 15 Jul. 2022, Accepted 23 Jul. 2022, Published 31 Oct. 2022

Abstract: The blue revolution of the blue economy on the way to build a golden Bangladesh is now the demand of the time. The blue economy is sea-based. The economy of exploiting the vast resources of the oceans and their bottoms. This means that whatever is extracted from the sea if it is added to the country's economy, will fall into the category of the blue economy. But the amount of resources of Bangladesh at the Bay of Bengal (BoB) has not yet been surveyed properly. If the number of marine cetaceans can be known, then proper steps of marine management can be taken to protect marine mammals. This paper deals with detecting marine cetaceans based on various machine learning classification algorithms such as Support Vector Machine (SVM), Decision Trees (DT) classifier, k-Nearest Neighbors (kNN) classifier, Artificial Neural Network (ANN) classifier, and Convolutional Neural Network (CNN) around the Swatch of No Ground (SoNG) in the BoB. At first, the possible marine cetaceans living around the BoB has been listed for the training purpose of classification algorithms. Then the dataset (both training and validation or test) being trained to classification algorithms have been created by extracting spectrogram images of the clicks, whistles or songs of listed marine cetaceans around the SoNG. Three types of test data such as original test data (OTD), synthetic test data (STD) and practical test data (PTD) have considered validating the proposed method. The test data retrieved from the original dataset is the OTD. The STD and PTD have been derived from the OTD. Then these algorithms have been trained with the training sets selected from created dataset for the detection and classification of marine cetaceans. After completing the training process, the proposed algorithm has been evaluated with three types of test data and recorded the output to analyze the performance in detection and classification of marine cetaceans. The detection process will be very challenging as there is a lot of noise in the sea. That's why we tested our model by generating synthetic and practical clicks, whistles or songs of marine cetaceans and comparatively satisfactory results have been obtained for CNN algorithm. This algorithm has been successfully detected and classified the species of marine cetaceans with the accuracy of 96.60% for OTD (Recall=0.94, F1-score=0.93), 93.38% for STD (Recall=0.91, F1-score=0.90) and 90.79% for PTD(Recall=0.91, F1-score=0.90).

Keywords: Bay of Bengal (BoB)Swatch of No Ground (SoNG)SpectrogramsConvolutional Neural Network (CNN)Marine Cetaceans

1. Introduction and Overview

Three parts of the earth are water. In this reality, the countries of the world are looking at the resources stored in the sea to meet their current and future needs. By 2050, the world's population will be about 900 million [1]. To provide food to this huge population, they have to depend on the sea. Throughout the twentieth century, various environmental movements and conferences have been brought. The Green Economy model [2] was at the centre of the discussion. In the twenty-first century, there was a need for further expansion of this model. The next step in the expansion of the green economy model [2] is known as the blue economy, which has already established a strong position around the world as an effective alternative to the challenges of the 21st century in achieving economic prosperity as well as maintaining environmental balance.

With the settlement of maritime disputes with Myanmar in 2012 and India in 2014 by the International Court of Justice, the total territorial sea area of Bangladesh is now more than 1 lakh 18 thousand 613 square kilometres [3]. It has a 200-nautical-mile [3] exclusive economic zone (EEZ) and sovereign rights over all kinds of animal and non-animal resources at the bottom of the continent from the coast of Chittagong to 354 nautical miles [3]. Bangladesh's economic zone has expanded since the maritime borders with Myanmar and India were demarcated. We have two types of resources at sea [4]. One is living resources and the other is non-living resources. Unfortunately, it is also true that the amount of resources we have at sea has not

E-mail address: arif.rme@du.ac.bd, taniaocndu@gmail.com

yet been surveyed.

Experts [4] said that the sea resources of the BoB can give Bangladesh future energy security, as well as change the overall look of the economy. They added that the efficient use of living and non-living resources in the sea can easily take the Gross Domestic Product (GDP) to double digits. According to a 2016 World Bank Group study [5], Bangladesh has so far failed to come up with a comprehensive policy plan on the maritime economy. But, we have huge sea resources. Now if we can utilize the huge resources of our sea area, we will go far ahead economically. In this case, short, medium and long term plans have to be taken. The conquest of the sea has created untapped opportunities and possibilities. We can link it with the Sustainable Development Goals (SDGs) number 14.

We do not know the number of marine resources in our EEZ. It is very important to know the number of marine resources to conserve. If we know the number of our marine resources, then we can take various steps to protect from extinction. For example, we can save these endangered mammals from extinction by declaring the areas where whales and dolphins have a high presence as marine protected areas (MPA). Whales or dolphins have enriched aquatic biodiversity. These are very important mammals in the aquatic environment. The ocean where whales and dolphins live increases the number of fish and keeps the ocean environment healthy. Their presence indicates a change in the quality or condition of the water. So if we know the presence of dolphins and whales in our BoB, we can know about the overall condition of the sea and take various initiatives to take action accordingly.

SoNG [3] is a protected area in the BoB in Bangladesh. The SoNG is a marine sanctuary. Known as one of the fish stocks of the BoB, the SoNG is home to a wide variety of marine fish as well as giant whales, dolphins, sharks, turtles and some rare species of aquatic animals [3]. The vast area of about one and a half thousand square miles is a safe breeding ground for rare biodiversity. It is also a breeding ground for dolphins and whales [3]. Scientists say this is the only swatch in the world where these three species of marine cetaceans can be seen together [3].

The vast area of about one and a half thousand square miles is a safe breeding ground for rare biodiversity, which could become especially important for the proposed blue economy.

The marine cetaceans can be divided into two suborders such as Mysticeti or baleen whales and Odontoceti or toothed whales [6].

Different types of cetaceans are available around the ocean in the world. But in this paper, those types of whales, dolphins and porpoises (Fig. 1, 2 and 3) of cetacean order [7] [8] [9] have been studied that may come around the SoNG in the BoB.

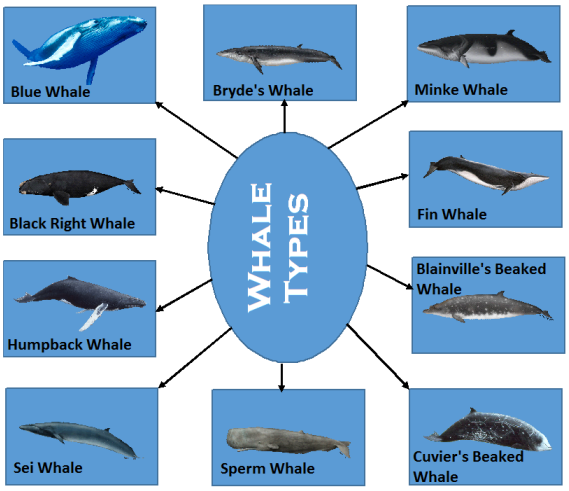

Figure 1. Probable Whale types around the swatch of no ground

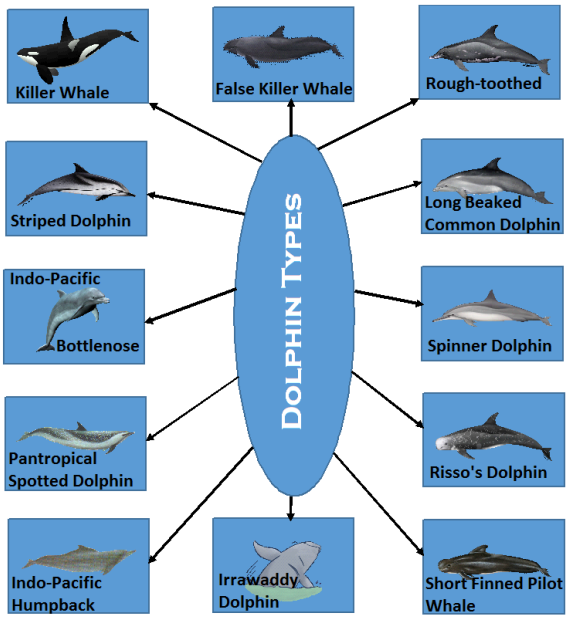

Figure 2. Probable Dolphin types around the swatch of no ground

2. Literature Review

Jiang, J. J., Bu [10] proposed a method based on CNN to detect and classify whistles of whales. They analyzed the denoised sound to estimate the target whistles and to classify the species of the detected whistles. The correction and classification rate of their proposed method was 87% and 85% respectively. They analyzed only two species of whales such as killer whales and long-finned pilot whales. They selected 15 sounds of these two species as raw data to generate dataset. Luo, W., Yang, W. [11] proposed a method based on CNN to detect odontocetes echolocation clicks by analyzing acoustic data. To distinguish between click and non-click clips, they trained the neural network.

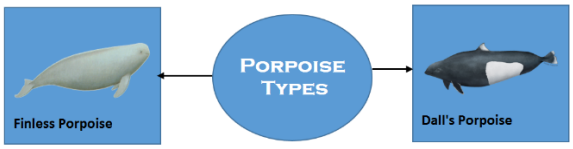

Figure 3. Probable Porpoise types around the swatch of no ground

They concluded that their proposed method worked accurately with different echolocation clicks. John R. Potter [12] developed a three-layer feed-forward artificial neural network (ANN) to detect bowhead whale (Balaena mysticetus) SoNG notes in the place of linear spectrogram correlator filter applied recently. They implemented the method on 1475 sounds. They used 54% and 46% of these sounds for training and test data respectively. They found an error rate of 1.5% of the trained Artificial Neural Network (ANN). Finally, they compared the spectrogram correlator filter with the ANN. Sue E. Moore [13] merged the broadband omnidirectional hydrophones having frequency ranges between 5 Hz to 30 kHz to the sea gliders. They experimentally showed that acoustic sea gliders (ASGs) were able to detect whale calls as well as various whistles and clicks produced by dolphins and small whales. They concluded that the ASGs can be expanded to detect marine mammal species in a broad range by which the habitats of cetaceans and their role in the marine ecosystem could be investigated. Yu Shiu [14] developed a deep neural network (DNN) to detect the vocalizations of North Atlantic right whales (Eubalena glacialis). They tried to compare the performance of the DNN with the traditional detection algorithm. They showed that the DNN based detection method was able to produce fewer false detection rates. They trained the DNN with recordings from a geographic region and the implementation of the network was easy with existing software. To recognize underwater acoustic targets, Xingmei Wang [15] combined a modified DNN with multi-dimensional fusion features. To extract these multi-dimensional features, they developed modified empirical mode decomposition (MEMD) and gammatone frequency cepstral coefficients (GFCCs) techniques. They modified the DNN using the Gaussian mixture model (GMM). They concluded that they obtained an accuracy of 94.3% with this underwater acoustic target identification method. Zhong, M., Castellote [16] developed a CNN model to classify and detect beluga whales. Also, the machine learning approach was tested in the detection of beluga whale acoustic signals. They concluded that their method successfully classified the beluga signals. In this paper, a CNN has been developed to detect and classify the marine cetaceans around the SoNG in the BoB, Bangladesh.Table II is provided to clarify the similarities and differences between the related works and the proposed method. At first, the species of cetaceans have listed that may come around the Indian Ocean and the BoB. Three types of cetaceans of 22 species have been identified such as whales, dolphins and porpoises for the detection and classification in this paper. Jiang, J. J., Bu [10], John

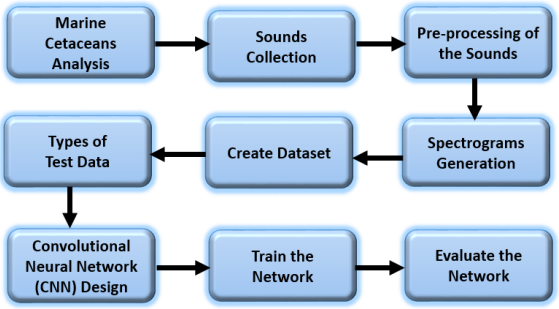

Figure 4. Flowchart of the proposed method

R. Potter [12] and Yu Shiu [14] did not provide any method for the detection of all whales or dolphin species. But in our work, we will consider a total of 22 species of these mammals. Then the songs, clicks or whistles of the 22 species have collected for creating the dataset. These songs, clicks or whistles from the ocean can be obtained using the hydrophone [13]. To record or listen to underwater sounds, a hydrophone which is a type of microphone can be designed [13]. The dataset has been used to train the neural network to implement the detection and classification model. The dataset contains training set and test data. After that three types of test data such as OTD, STD and PTD have been validated using the developed neural model and determined the accuracy in detecting and classifying the marine cetaceans that can come around the SoNG.

3. Methodology

A. *Work Plan*

The block diagram shown in Fig. 4 summarizes the whole process of the work. At first, the possible marine cetaceans that may be found around the SoNG have analyzed and made a list of these species. Then the songs, clicks or whistles of the corresponding cetaceans have been collected in waveform audio file extension (.wav) for pre-processing of these sounds for the features extraction.

After obtaining the feature images, a dataset has been created with the training data and test data. After that, maching learning based classification methods have been designed to detect and classify the marine cetaceans around the SoNG. As the paper deals with the detection and classification based on spectrogram images of cetaceans, SVM, DT classifier, kNN, ANN, and CNN have been implemented. Then the models have trained with the training data to fit with the models. After completing the training process, the trained models have been validated using the OTD, STD, and PTD generated during the creation of the dataset.

B. *Marine Cetaceans analysis around the swatch of no ground*

The BoB is an almost triangular bay located in the northern part of the Indian Ocean [17]. The Gulf is bordered by India and Sri Lanka to the west, India and Bangladesh to the north, and Myanmar and Thailand to the east [17].

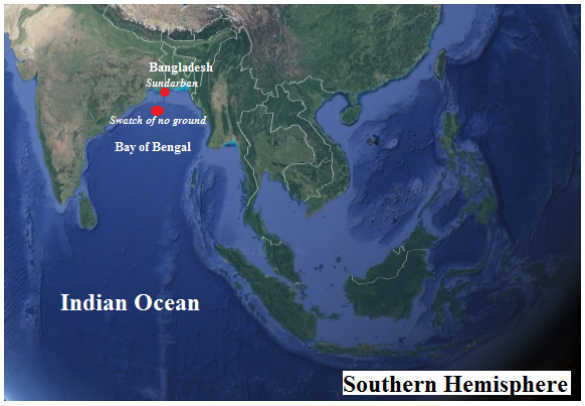

Figure 5. Location of BoB and SoNG[25]

Geographically, the BoB is located between 5° North and 22° South latitudes and 80° East and 100° East longitudes [17] shown in Fig. 5. The largest bay in the world is bounded on three sides by the east coast of India and Sri Lanka to the west, the delta formed by the Ganges-Brahmaputra-Meghna river system to the north, and from the eastern peninsula to the Andaman-Nicobar ridges [17]. Not all species of cetaceans are found in this marine region due to water temperature, environmental variation and food habitat. That is why it is important to know the natural habitat of each species to detect this class of animals in the SoNG. Then the detection work will be much easier due to the reduction of data being trained to the maching learning based models. The sounds, whistles or songs of cetaceans shown in Fig. 1, 2 and 3 have been considered to train and validate the neural network to detect and classify the marine cetaceans around the SoNG.

C. Sound Collection

Various types of sounds of the marine cetaceans that may be found around the SoNG have been collected from three websites [18] [19] [20]. The probability of finding the marine cetaceans shown in Fig. 1, 2 and 3 in SoNG is much higher. So a total of 6003 sounds of songs, whistles or songs of these 22 species from various websites such as Voices in the Sea, Discovery Sound, National Oceanic and Atmospheric Administration [18] [19] [20] and youtube have been considered as raw data for the generation of the dataset in implementing detection and classification models based on machine learning models.

D. Pre-processing of the sounds

For the excerpted sounds, the waveform audio file extension (.wav) has been considered in this work. The continuous signals of songs, whistles or songs of the marine cetaceans are reduced to discrete values for further analysis in the digital domain. All of these sounds of cetaceans are pre-processed according to the procedures described in Fig. 6.

Total 6003 sound samples of the 22 species of marine cetaceans are considered as raw data. For the analysis of

Figure 6. Pre-processing steps of the sound data

these sounds data of marine cetaceans, a python package library named librosa [21] has been used. The sampling rate of 22050 has been used and then the corresponding waveforms in the time-domain of these sounds have been generated. After generating the waveform of the Irrawaddy dolphin, Fast Fourier Transform (FFT) [22] has been applied to generate a spectrum of this waveform based on equation 1. The resultant spectrum is in the frequency domain.

$$F(u,v) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f[m,n] e^{-j2\Pi(umx_0+vny_0)} \quad (1)$$

Where, $F(u,v)$ is the two-dimensional spectrum of $f(x,y)$ defined over an x − y plane, u is the spatial frequency in the x-direction, v is the spatial frequency in the y-direction, x_0 and y_0 are the spatial intervals between consecutive sound signals in x and y direction, m and n are the number of points. Equation 1 describes the amplitude and phase characteristics of the sound signal as a function of frequency. Frequency analysis helps to analyze these sounds in noisy and variable parametric situations in the ocean. According to the procedures described in 6, the obtained waveforms of IrrawaddyDolphin.wav and MinkeWhale.wav (audio file) are shown in Fig. 7.

The FFT has been applied to convert time-domain sound signals into frequency domain signals referred to as spectrum shown in Fig. 8 and 9.

E. Spectrograms Generation

To make a visual representation of the spectrum of frequencies of these sound signals varying with time, a spectrogram has been obtained using equations 2 and 6. The spectrogram is obtained by applying Short-term Fourier Transform (STFT) [22] to compute the squared magnitude for a window width ω.

$$Spectrogram(\mathrm{t},\omega) = |STFT(t,\omega)|^2 \quad (2)$$

$$STFT\{x[n]\}(m,\omega) \equiv X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]\omega[n-m]e^{-j\omega n} \quad (3)$$

Where t is the time in second To train the neural network, spectrograms in portable network graphics (png)

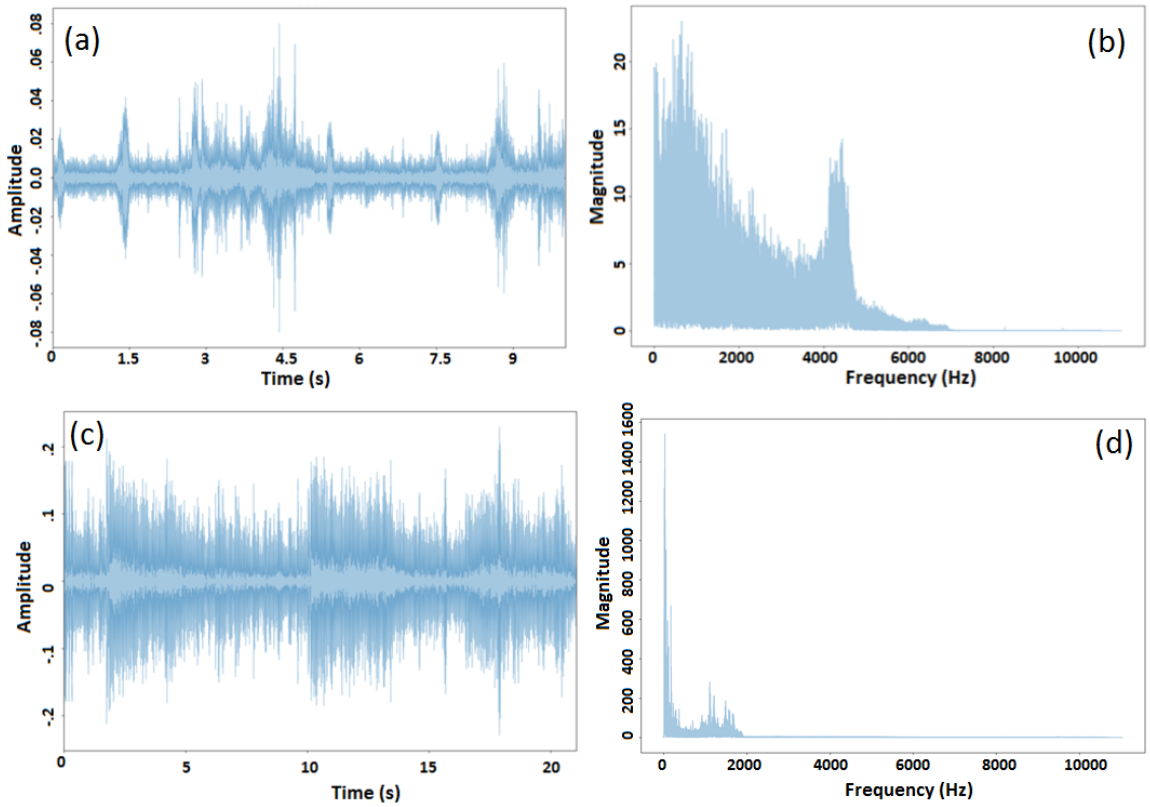

Figure 7. Generated waveforms of Irrawaddy dolphin's (a) sound, (b) power spectrum, and Common minke whale's (c) sound, (d) power spectrum

Figure 8. Spectrogram of original sound data generation

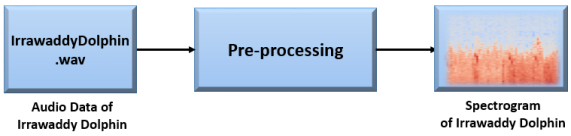

Figure 9. Spectrogram of original data

extension has to be obtained. The spectrogram of sounds of sounds, whistles or songs of each cetacean can be generated using the procedure shown in Fig. 8.

The sounds of cetaceans in wav extension are passed through the pre-processing stage to obtain the corresponding waveform, power spectrum and finally the spectrogram of each cetacean. Each spectrogram is of 64×64 pixel image. The obtained spectrogram is in png extension. For the sound of Irrawaddy Dolphin's clicks, the resultant output of the pre-processing stage is the spectrogram in png shown in Fig. 9.

With spectrogram, the variation of the energy level can be visualized over time. The spectrograms of Irrawaddy-Dolphin.wav and MinkeWhale.wav (wav) obtained from

Similarly, for all of the sounds of marine cetaceans, the spectrograms have been obtained using the procedures described in Fig. 8. The resultant spectrograms are image data that have been stored in .png format for creating a dataset (Will be described in next section of Create Dataset) to train and validate the network model in the detection and classification of marine cetaceans.

F. Create Dataset

The dataset of cetacean's clicks, whistles or songs consists of 22 species. A total dataset is about 6003 sounds (spectrogram images) of the species of marine cetaceans. In this section, the data has been divided into the training set (70%) and test data (30%). Among 6003 spectrograms, 4003 have been used to train the machine learning based models. The generalized training set can be formatted [23] by equation 7.

$$\{input, correctoutput\} \tag{4}$$

Here, the input contains about 4003 spectrograms of marine cetaceans and the correct output contains about 22 species of marine cetaceans. After completing the train-

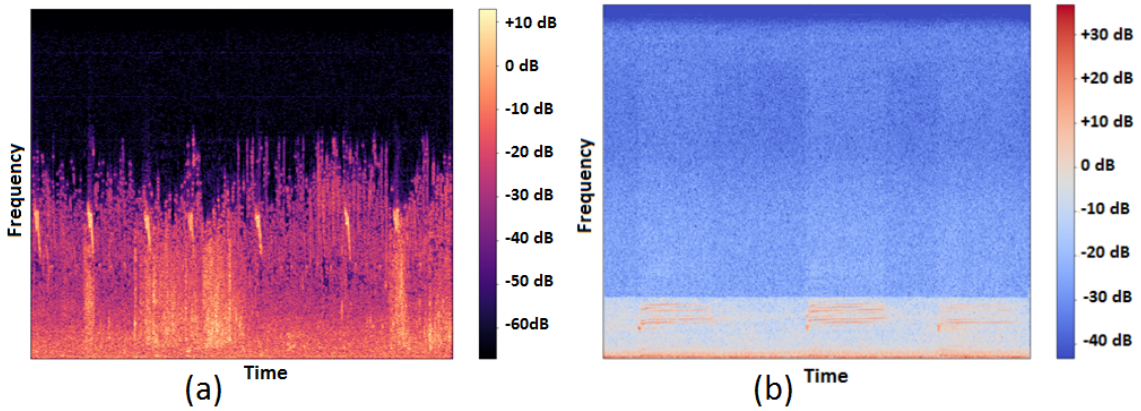

Figure 10. Spectrogram (dB) of (a) Irrawaddy dolphin and (b) Common minke whale's sound

ing procedure, the models have been validated using the remaining 2000 test data. The training set and test data have been used herein in image format (.png). To train the network with fewer images rather than with lots of images, image augmentation has been performed to create training images artificially through multiple processing including flips, shifts and random rotation etc.

G. Synthetic and Practical Test Data Generation

Three types of test data such as original, synthetic and practical spectrograms have been generated to evaluate the proposed method. The OTD has been obtained from the dataset obtained in section F. The TD derived from the dataset has been considered as OTD in this paper. From this OTD, the STD and PTD have been generated to evaluate the performance of the proposed detection and classification method.

1) Spectrograms of STD

The synthetic spectrogram of each marine cetacean can be generated using the procedure shown in Fig. 11. Using a synthetic data generator implemented in python, the STD has been generated based on the image augmentation process applied on the training set. This process converts the spectrograms of test data obtained in subsection 3.5 into a new format that may not be found in OTD. As the STD is artificially manufactured rather than created by real word incidents, these data are used as a simulator.

Here, the spectrogram of test data obtained in subsection D has considered as the original spectrogram that has been generated in subsection E. The procedure of creating an STD is described in Fig. 11. At first, the spectrogram of each test data has segmented (Fig. 11) to separate the foreground portion from the background. After successfully extracting the foreground, this separated portion has been used multiples times in multiple images with different backgrounds. In the augmentation process, the colour, background, rotation, a horizontal or vertical mirror of the spectrogram has changed to obtain the STD.

The spectrogram obtained from Fig. 11 of each STD has been obtained using the procedure described in Fig. 7.

For the original spectrogram of Irrawaddy Dolphin's click (Fig Generated waveforms of Irrawaddy dolphin's sound original sound.), the resultant output of the synthetic data generator stage is the synthetic spectrogram in the png extension shown in Fig. Generated waveforms of Irrawaddy dolphin's sound original sound.. Similarly, for all of the sounds of marine cetaceans, the spectrograms of STD have been obtained using the methodology described in Fig. 11. The resultant spectrograms are image data that have been stored in .png format.

These spectrograms obtained from Fig. 11 have been considered as STD to evaluate the network model in the detection and classification of marine cetaceans.

2) Spectrograms of PTD

Practical data refers to the data that are found in the real world. In this paper, PTD refers to the sounds data that are obtained by hydrophones from the ocean. The sounds obtained from the ocean contain the original clicks, whistles or songs of marine cetaceans as well as underwater background sounds and ship sounds in the ocean. So the proposed models have to be validated using this type of data to implement the models in real applications. In this paper, only the underwater background sounds (UWBS) in the deep ocean have been considered to mix up with the original clicks, whistles or songs of marine cetaceans.

The practical spectrogram of each marine cetacean can be generated using the procedure shown in Fig. 12. In this case, the UWBS in the deep ocean has mixed up with the original clicks, whistles or songs expressed in wav file. Then the resulting audio signal has passed through the pre-processing stage described in Fig. 3-D to obtain the practical spectrogram of each marine cetaceans. Windows movie maker of version 8.0.8.2 has been used to perform the mixture task shown in Fig. 12. Here the original sounds of clicks, whistles or songs are mixed up with the UWBS in

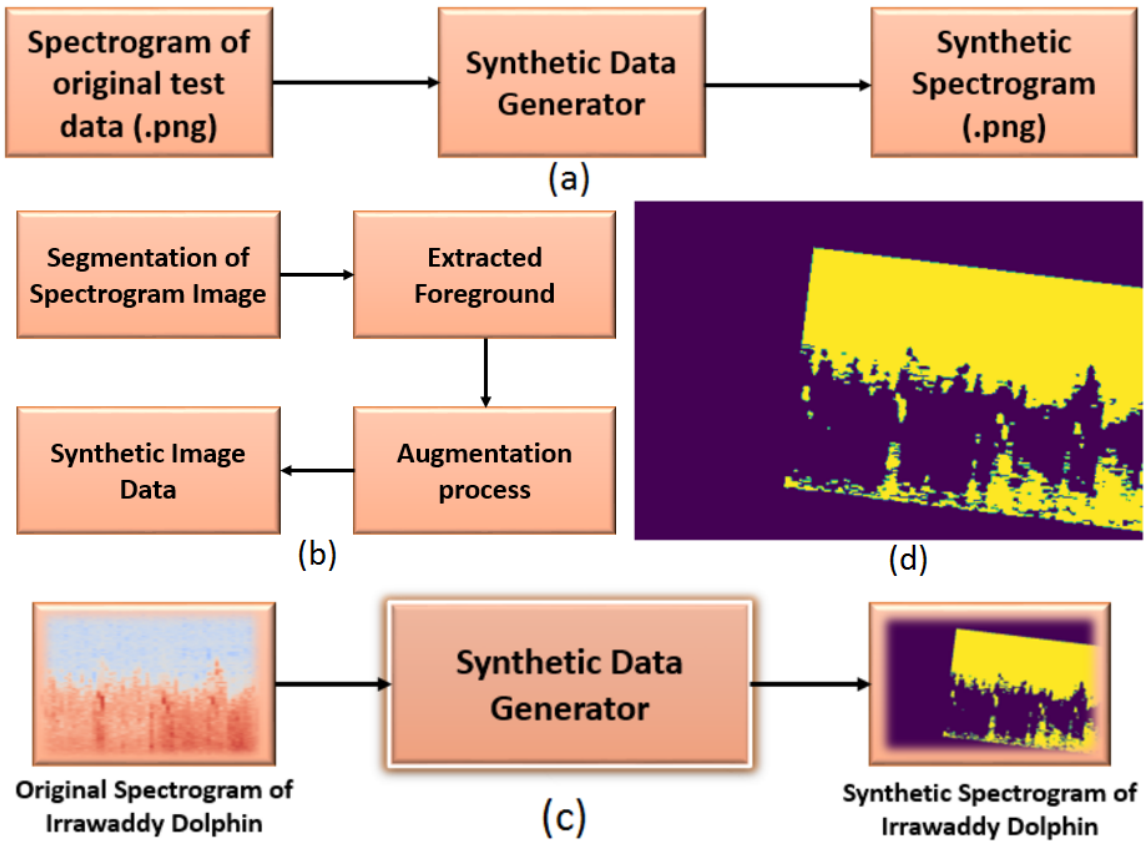

Figure 11. Synthetic Data Generator (SDG) (a) spectrogram generation, (b) generation process, (c) original to synthetic data generation, and (d) generated spectrogram of synthetic data

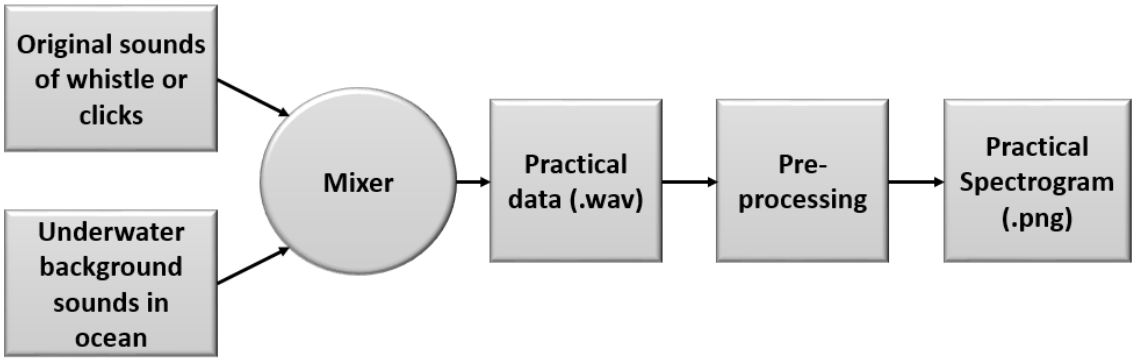

Figure 12. Synthetic spectrogram of Irrawaddy dolphin

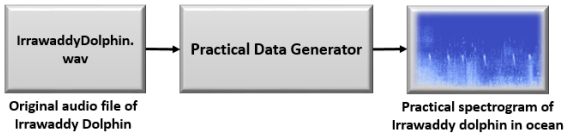

Figure 13. Spectrogram of real data

the deep ocean to develop the PTD that are available in the ocean. The obtained PTD is similar to the obtained data from the hydrophone. The obtained PTD is in waveform audio file (.wav) extension. Then the pre-processing step has performed shown in Fig. 8 to get the spectrogram in png extension. The spectrogram of each PTD has been obtained using the procedure described in Fig. 12. For the original sound of Irrawaddy Dolphin's click, the resultant output of the practical data generator stage is the practical spectrogram in the png extension shown in Fig. 13.

The waveforms of the original and practical audio file of IrrawaddyDolphin.wav are shown in Fig. (Fig:Generated waveforms of Irrawaddy dolphin's sound original sound). The FFT has been applied to convert time-domain sound signals into frequency domain signals referred to as spectrum shown in Fig.7. Then the python library librosa has used to obtain the spectrograms for every audio file containing marine cetaceans sounds of clicks, whistles or songs.

With spectrogram, the variation of the energy level can be visualized over time. The spectrogram of an original and practical audio file of IrrawaddyDolphin.wav are shown in Fig. 15.

Similarly, for all of the sounds of marine cetaceans, the spectrograms of PTD have been obtained using the procedures described in Fig. 13.

The resultant spectrograms are image data that have been stored in .png format. These power spectrums have been considered as the PTD to evaluate the neural network.

H. Convolutional Neural Network (CNN) Design

As the classification process has to perform using the spectrogram images of the corresponding marine cetaceans, the CNN has been implemented shown in Fig. 17. Because the CNN is a DNN that is specialized for image recognition [23]. After passing this network, the features have extracted from the feature extraction network and then enter into the classification network for further operation.

The classification network can operate based on the features of the spectrogram image extracted from the feature extraction network and generate output according to the features.

1) Feature Extraction Network

The feature extraction network implemented in Fig. 17 consists of assembles of the convolutional layer and pooling layer pairs. To convert the image, the convolution layer performs a convolution operation. The convolution layer produces new images referred to as feature maps which make the unique features of the input spectrograms more noticeable or prominent. This layer contains filters to convert images called convolution filters yields the feature maps. The feature map produced through the several convolution operations is now processed through the Rectified Linear Unit (ReLU) [23] activation function. As the method deals with multiclass applications, the ReLU function [23] has used rather a sigmoid function.

The neighbouring pixels are combined into a single pixel by the pooling layer which reduces the dimension of the image. The pooling process can be called a convolution process, but the only difference is the elements do not overlap in the pooling process. The CNN has been designed to implement the detection and classification model of marine cetaceans around the SoNG in the BoB. The design parameters are shown in Table 2. Sequential type model has been used in this network. To build a model layer by layer in Keras [21], a sequential model is the easiest way. Each layer of the sequential model has weight values that correspond to the layer the follows it [21]. The input image shape is the size of the spectrogram of clicks, whistles or songs of marine cetacean. The size of the spectrogram obtained in sub section 3.7 is a 64-by-64 pixel image. As the neural model have to deal with the colour image spectrogram, the input node of the neural network will be 12288 (=64 × 64 × 3).

Three hidden layers which are mathematical functions have been used in this network to produce an output definite to an intended result. Whether a neuron of the network should be activated or not by calculating weighted sum and further adding bias value with it, the activation function is used in the neural network [21]. As the ReLU function does not activate all the neurons at the same time, this activation function has been used for hidden layers. As the output deals with the multi-class classification problem of 22 marine cetacean species, the softmax activation function has been used for the output layer. The output of the softmax function from the i-th output node is determined by the expression [23] shown in equation 8.

$$y_i = \varphi(v_i) = \frac{e^{v_i}}{e^{v_1} + e^{v_2} + e^{v_3} + \cdots + e^{v_M}} = \frac{e^{v_i}}{\sum_{k=1}^{M} e^{v_k}} \qquad (5)$$

Where, y_i is the output from the i-th output node, i is the number of output node, v_i is the weighted sum of the i-th output node, φ is the activation function, k is the number of training data, and M is the total number of the output node.

To calculate the gradients to update the weights of the neural network, categorical cross-entropy has been used as the loss function of prediction error. The term categorical

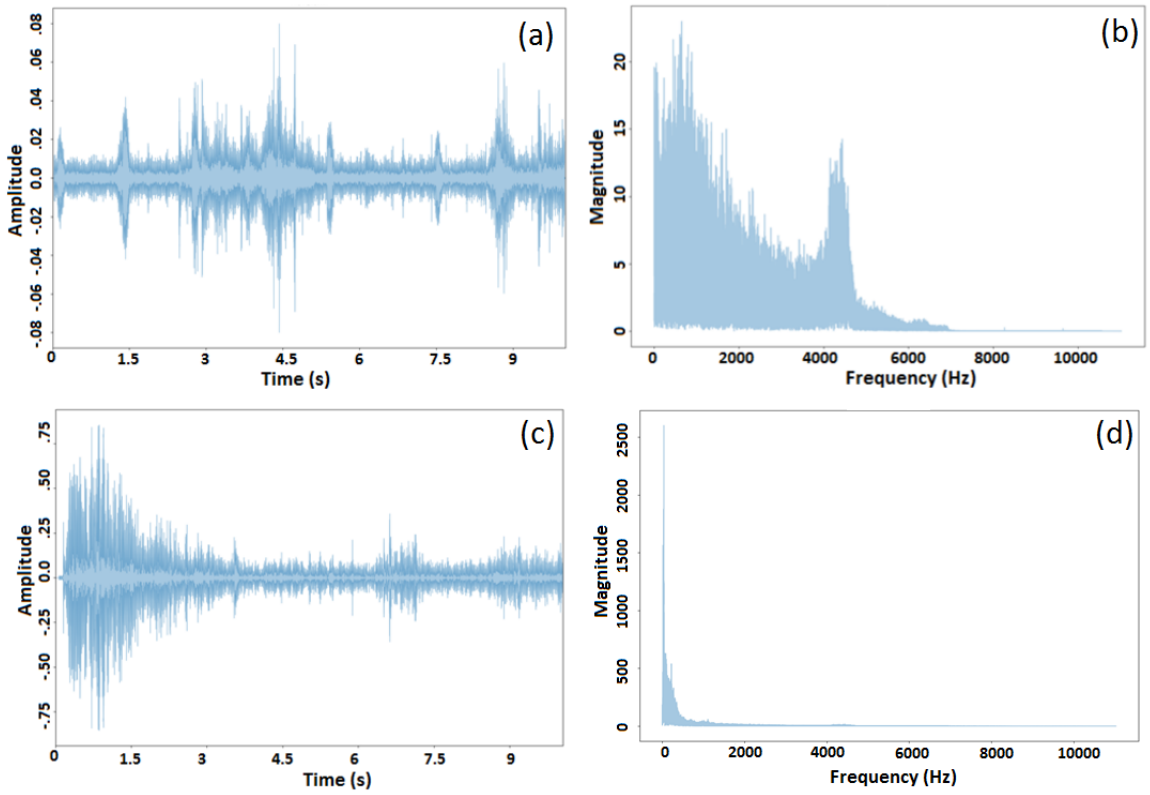

Figure 14. Generated waveforms of Irrawaddy dolphin's (a) original sound, (b) power spectrum of original sound, (c) practical sound, and (d) power spectrum of practical sound in the ocean

refers to the multi-class detection application [21]. Due to the faster operation of Adam than stochastic gradient descent (SGD) [21], the proposed neural network model has used the Adam optimizer to hold the model into its most accurate form with the handling capability on noisy problems.

I. Train the Network

A CNN has been designed to detect and classify the marine cetaceans around the SoNG in the BoB based on the spectrograms of the clicks, whistles, or songs of each species. Now the neural network has been implemented using a dataset created in subsection 3.6. It contains total of 6003 sounds of marine cetaceans that may be found around the SoNG. Among them, about 4003 sounds have been used for training and 2000 for validation. The species have been classified into 22 classes implemented in the output layer.

Where e is the error of the output (e.g difference between output and correct output), d is the correct output, y is the output, δ is the delta function, φ' is the derivative of the activation function, v is the weighted sum of the output node, α is the learning rate, w_{ij} is the updated weight, Δw_{ij} are the weights updates, i is the number of output node, j is the number of input node, k is the number of training data, and W^T is the transpose weighted sum of the hidden layer.

The input data (e.g. spectrogram of each marine cetacean) of the proposed neural network travels through the input layer, then the hidden layer and finally the output layer. In the backpropagation algorithm [24] [25] the output error starts from the output layer of the neural network and moves backwards. The backward movement continues until it reaches the right next hidden layer to the input layer. Also, the input signal flows forward to the connecting lines to get multiplied with weights [17]. The flowchart of the whole training process is shown in Fig. 16. The block diagram of CNN shown in Fig. 17 can recognize the dataset images. As the spectrogram of the input image is a 64-by-64 pixel colour image, a total of 12288 (=64×64×3) input nodes have been allowed. The feature extraction network consists of two convolution layers with 20 (dimension of 9×9) classification filters. When the neural network has successfully trained with the training set, the network has been subjected to get the output for different inputs of spectrograms of each marine cetacean.

The convolution layer's output is passed through the ReLU activation function followed by the pooling layer employing a max-pooling process of two by two sub-matrices. Table I summarizes the training parameters of the proposed neural network model.

This paper provides 4003 of sound data of spectrograms

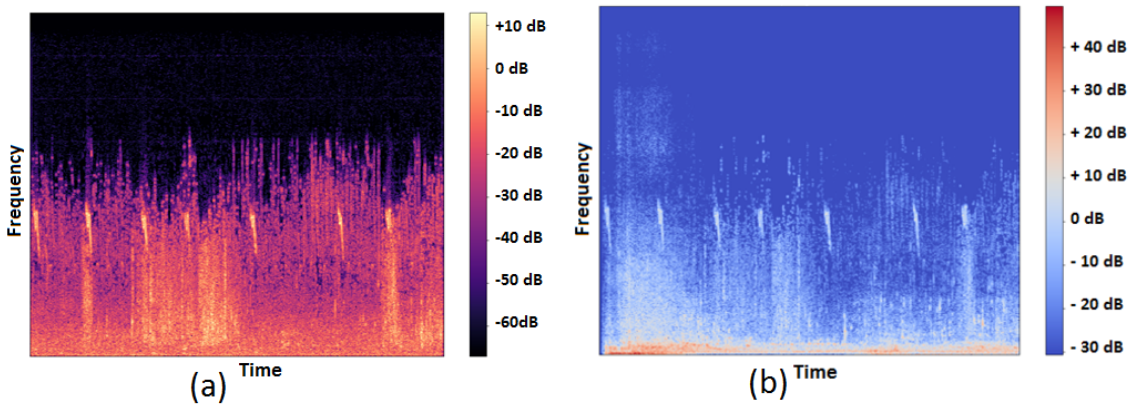

Figure 15. Irrawaddy dolphin's spectrograms of (a) original sound, (b) practical sound in the ocean

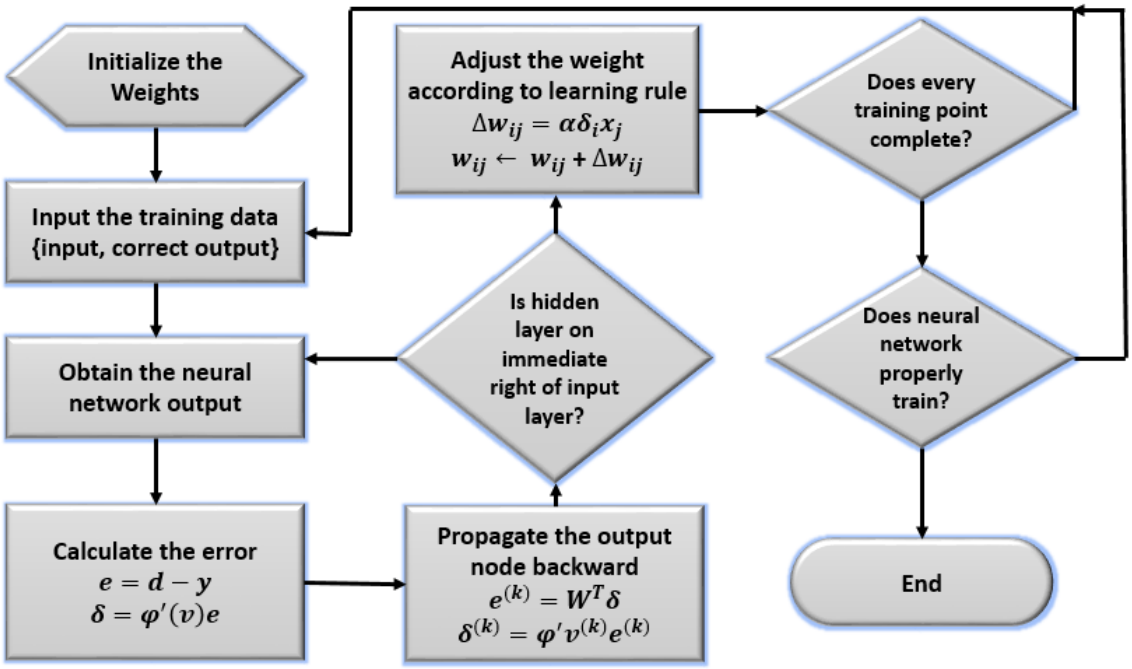

Figure 16. Training algorithm of the network

TABLE I. Training Parameters of the Neural Network

Layer	Remark	Activation Function
Input	64×64×3 nodes	-
Convolution	20 convolution filters (9×9)	ReLU
Pooling	2 max-pooling (2×2)	-
Hidden	100 nodes	ReLU
Output	22 nodes	Softmax

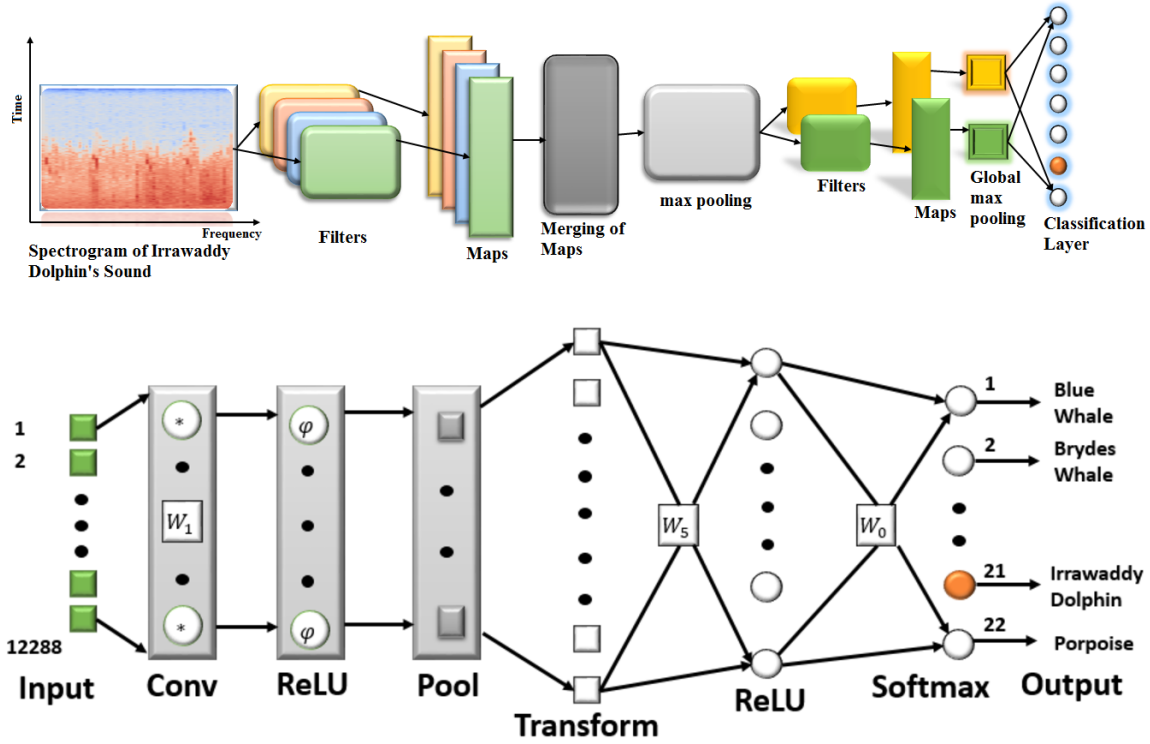

Figure 17. Convolutional Neural Network (CNN) architecture designed for taking spectrogram image as an input to generate the output of a classification

to train the network and the accuracy has displayed which was 93% in 120 seconds.

J. Evaluate the Network

A CNN has been implemented that takes the spectrograms of the species of marine cetaceans as input images and detects the species that it corresponds. The training set contains 4003 spectrogram images of the sounds of clicks, whistles or songs of marine cetaceans. Among 6003 spectrograms created to form a database, 2000 spectrogram images have been used for the test purpose shown in Table II. Each spectrogram image has a dimension of a 64-by-64 pixel.

This paper employs 6003 of spectrogram images with the training set and test data in an 7:3 ratio. Thus 4003 of spectrogram images have been used to train the neural network and 2000 of spectrogram images (OTD) have been used to generate STD and PTD to validate the performance of the neural network. The dataset problem is caused by the multi-class classification of the 64-by-64 pixel image into one of the 22 classes of species. The architecture and evaluation process of the designed neural network to detect and classify the species of marine cetaceans around the SoNG is shown in Fig. 18.

Then the class number of 22 shown in Fig. 17 (b) representing Irrawaddy dolphin has been predicted based on the designed neural network.

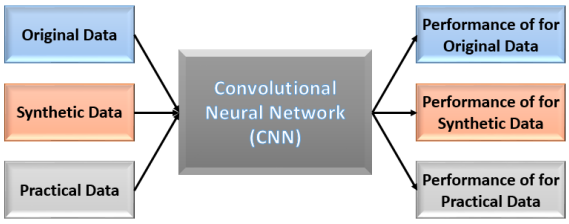

Figure 18. Performance evaluation for test data of original, synthetic and practical data

K. Classification metrics

Because classification techniques generate distinct outcome, designers require somewhat of a metric that can relate the specific categories that are generated. Classification Indicators assess a performance of the models and show us how right or wrong the categorization is; however, each of them analyzes it in an unique manner. This is because each concept has its own distinctive features. Therefore, in addition to assessing Classification techniques, we just go over the following metrics in considerable detail:

1) Accuracy

The classifier performance metric can be expressed as the ratio of correctly predicted divided by the total number

TABLE II. Validation or Test Data being used to Evaluate the Network

Training data	Validation data or test data		
	Original data	Synthetic data	Practical data
4003	2000	2000	2000

		Predicted Marine Cetaceans	
		Blue Whale	Not Blue Whale
Ground Truth	Blue Whale	TP	FP
	Not Blue Whale	FN	TN

Figure 19. Confusion matrix for null hypothesis

of assumptions, multiplied by 100. This is maybe the easiest measurement to use it and put into effect. Either we can enact this by repeatedly evaluating the current qualities with the experimental and predicted in a loop, and we'll just use the scikit-learn module to just let it do all the grunt work for us.

2) Confusion Matrix

The Confusion Matrix is a table-based graphical presentation of the ground-truth labels in comparison to the model's assumptions. The incidences that belong to a final prediction are represented across the rows of the confusion matrix, while the incidences that belong to an actual class are represented across the columns. While the Confusion Matrix isn't really technically a performance indicator, it does serve as the foundation upon which other measurements can assess the outcomes.

We ought to make assumptions about just the value of the null hypothesis in order to fully understand the confusion matrix. This value could have been anything. For instance, based on the information we have regarding marine cetaceans detection (Fig. 19), let's make the assumption that our null hypothesis H0 is "Blue Whale is detected !"

The term "True Positive" (TP) refers to the number of instances in which one's method accurately predicted true positive samples. The term "True Negative" (TN) refers to the number of correctly identified negative class samples for a given model. The term "false positive" (FP) refers to the number of wrongfully postulated negative class instances one's method generated. Within the realm of empirical designations, this aspect denotes a Type-I error. The placement of this error within the confusion matrix is determined by the null hypothesis that is selected. The term "false negative" (FN) refers to the number of instances in which one's method inaccurately prophesied true positive samples. Within the realm of analytical phraseology, this aspect exemplifies a Type-II error. The placement of such a error with in confusion matrix also is contingent on the particular null hypothesis that is chosen.

3) Precision

The ratio of the number of true positives to the total positives anticipated is known as precision.

$$P = \frac{TP}{TP + FP} \tag{6}$$

$$P = \frac{IdentifiedBlueWhaleCorrectly}{IdentifiedBlueWhaleCorrectly + IdentifiedBlueWhaleIncorrectly} \tag{7}$$

A precision score that's also closer to just one indicates that ones model did not overlook any true positives and is capable of classifying Blue Whale into groups that are correctly labeled and those that are not correctly labeled. Just what is unable to quantify, however, is the presence of Type-II error, also leading to false negatives – instances where a Killer Whale is incorrectly identified as Blue Whale. Ones discriminator may have an elevated false positive rate whether it has a low precision rating (0.5), which could be the result of an extremely unbalanced class or part of the production model hyper - parameters. In order to avoid false positives and false negatives when dealing with a class imbalanced conundrum, you will need to start preparing your information is crucial by either over-sampling that are under them.

4) Recall/Hit-Rate/Sensitivity

A recall can be thought of as the proportion of genuine positives relative to the total number of positives in ground truth.

$$R = \frac{TP}{TP + FN} \tag{8}$$

Recall closer to one indicates that ones model did not overlook any true positives and is capable of classifying Blue Whale effectively around those who have been correctly labeled and those who have been mislabeled. Just what is unable to assess is indeed the presence of type I error, also leading to false positives or instances in which a Killer Whale is incorrectly identified as Whale types. Ones discriminator may have an elevated lot of false negatives if its recall score is low (less than 0.5), which may be the result of an extremely unbalanced class or part of the production model hyper - parameters. In order to avoid false positives and false negatives when dealing with a class imbalanced problem, readers will need to prepare their data in advance by either over-sampling that are under them.

5) F1-score

The F1-score metric takes into account both the accuracy and the amount of information remembered. In reality, the F1 score is calculated by taking the harmonic mean of something like the two scores. In essence, the equation for both the two is as follows:

$$F_1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

A poor F1 score doesn't inform us (nearly) anything; all it does is inform us how well our accomplish at a certain criterion. A poor experiencing that we did not make an effort to perform very well with a significant portion of the a whole testing set. A poor level of precision indicates that we were incorrect about a significant number of the instances that we considered to be positive instances.

4. Result and Discussion

Among 6003 spectrograms of the sounds of clicks, whistles or songs of marine cetaceans, about 4003 of spectrograms have been used to train the neural network. The test data are of three types such as OTD retrieved from the created dataset, STD derived from the OTD using image augmentation process and PTD obtained from the mixture of original clicks and UWBS in deep ocean fed to CNN shown in Fig. 17. The outputs for OTD, STD and PTD have been recorded to evaluate the performance of the developed CNN. From Table II, a total of 2000 of spectrograms have been tested to detect and classify the marine cetaceans around the SoNG in the BoB.

Out of the three types of test data, 2000 of the OTD, STD and PTD of marine cetaceans have been tested the models based on machine learning with a total of 2000 test data. The CNN has successfully detected marine cetaceans with 96.60% accuracy for OTD, 93.38% for PTD and 90.79% for PTD shown in Fig. 21. The CNN model shows best performance compared to other models in terms of accuracy, confusion matrix, precision, recall and F1-score. The CNN model correctly predicted 1748 of marine cetaceans and 23 of marine cetaceans which were not considered as marine mammals. The CNN model predicted 31 of marine cetaceans that were not actually marine cetaceans and 198 of these detected as non-marine cetaceans that were marine cetaceans such as Killer Whale, Porpoise etc. The SVM algorithm performed worst performance to detect the marine cetaceans with OTD, STD, and PTD.

The model has successfully detected all the species of marine cetaceans with the highest accuracy of 96.60% for OTD which has been retrieved from the dataset. These test data contain only the clicks, whistles or songs of marine cetaceans. No noise or complex signal has mixed up with this test data. That's why the precision has reached to 0.97 with the false detection for 40 test samples. The model has successfully detected 1740 species of marine cetaceans (out of 2000) for STD which has derived from the OTD. This test data contains only a slightly modified version of the clicks, whistles or songs of marine cetaceans. No noise or complex signal has mixed up with this test data. In this case, the accuracy has reached 0.93 with 91 false detections.

The model has successfully detected 1846 species of marine cetaceans (out of 2000) for PTD which has been developed by mixing the UWBS in the deep ocean with the OTD. These test data contain both the clicks, whistles or songs of marine cetaceans and the UWBS of the deep ocean. So, noises or complex signals have mixed up with these test data. The CNN model showed 90.79% accuracy for PTD. The performance degraded due to the mixture of under water ocean sound with the songs, whistles or clicks of the marine cetaceans. That's why the precision drops to 0.89 with 154 false detection.

Table III, IV, and V show the overall performance of the neural network model with three different types of test data.

A comparison (Table VI) between the outcomes of proposed method and the related work done before. No work was found where synthetic data and practical data were generated to validate the neural network model. This paper deals with the neural model where both synthetic data and practical data have been generated to validate the proposed methodology to detect and classify marine cetaceans. With the newly generated synthetic and practical data the proposed model shows comparatively better results. Also, the ANN error rate decreases with the proposed method.

The overall accuracy of the models based on machine learning for three test data is shown in Fig. 20. For OTD, the SVM shows lesser accuracy where the CNN shows highest. For OTD, the CNN detected the species of marine cetaceans with 96.60% accuracy compared to STD and PTD. As the supervised learning method has been applied in this case, the outcome is in highest accuracy with only 40 false detections. But when the whistles, songs or clicks of marine cetaceans were mixed with the under water background sound, the performance degrades to 90.79% due to the adverse sound effect in the ocean region.

In this paper, 6003 spectrograms of 22 species marine cetaceans have been considered to create the dataset. So if the number of spectrograms can be increased to create the dataset, the number of training sets and test data will be increased and thus the performance can be evaluated satisfactorily in a broad sense. Moreover, the test sets that have been used to train the network contain only the sounds of clicks, whistles or songs of 22 species of marine cetaceans. So if the actual data are used to train the neural network, the accuracy will be increased. If any species of marine cetaceans are available in the area of SoNG in the BoB, the hydrophone will capture the clicks, whistles or songs and generate the sound wave. By extracting the features from the sound wave, the cetaceans can be detected and classified based on the designed CNN successfully.

TABLE III. Performance of various algorithms for original test data

Classifier	Accuracy	Confusion Matrix	Precision	Recall	F1-score
SVM	84.98%	1699 75 70 156	0.85	0.86	0.79
Decision Tree	83.67%	1700 74 71 155	0.84	0.83	0.82
kNN	87.13%	1750 65 80 15	0.87	0.86	0.87
ANN	89.13%	1803 40 33 124	0.90	0.88	0.88
CNN	96.60%	1753 18 22 207	0.97	0.94	0.93

TABLE IV. Performance of various algorithms for synthetic test data

Classifier	Accuracy	Confusion Matrix	Precision	Recall	F1-score
SVM	82.91%	1690 79 80 151	0.83	0.84	0.69
Decision Tree	81.76%	1688 70 89 153	0.82	0.81	0.72
kNN	86.47%	1738 50 199 13	0.86	0.80	0.67
ANN	87.94%	1800 50 40 110	0.88	0.87	0.88
CNN	93.38%	1740 23 48 189	0.93	0.91	0.90

TABLE V. Performance of various algorithms for practical test data in the ocean

Classifier	Accuracy	Confusion Matrix	Precision	Recall	F1-score
SVM	78.97%	1541 91 242 126	0.79	0.76	0.78
Decision Tree	81.37%	1603 97 191 109	0.81	0.79	0.84
kNN	82.41%	1654 99 228 19	0.82	0.83	0.86
ANN	86.38%	1683 73 110 134	0.88	0.85	0.89
CNN	90.79%	1748 73 81 98	0.89	0.91	0.90

If anyone wants to know their marine cetaceans around their exclusive economic zone of oceanic zone, this method can be used efficiently to detect and classify the marine mammals. This work can be further extended to the applications of detecting marine fishes also. The clicks, whistles or songs of marine cetaceans has been collected based on the recorded version. It may be created via an artificial process. The dataset has been generated based on these clicks, whistles or songs. So, the created dataset in the given detection task is not sufficient to generate intended outcomes for real-world applications. To make it more generalized, two cases were also introduced such as synthetic data and practical data approach. If the dataset can be generated considering these three types of data, it will be a better approach for real-world applications. Moreover, the number of samples to validate the model is comparatively smaller.

5. Conclusion

To detect and classify the marine cetaceans at the SoNG in the BoB, a CNN has been developed in this paper. Three types of marine cetaceans such as whales, dolphins and porpoises have been identified for the detection and classification applications. Then the sounds of the clicks, whistles or songs of 22 species marine cetaceans that are considered have been collected for creating the dataset. These clicks, whistles or songs can be obtained via the hydrophone. To create a dataset, the clicks, whistles or songs of 22 species of marine cetaceans have been collected as wav. The audio file contains only the original clicks, whistles or songs of the selected marine cetaceans. For 22 species of marine cetaceans, a total of 6003 waveform audio files have been

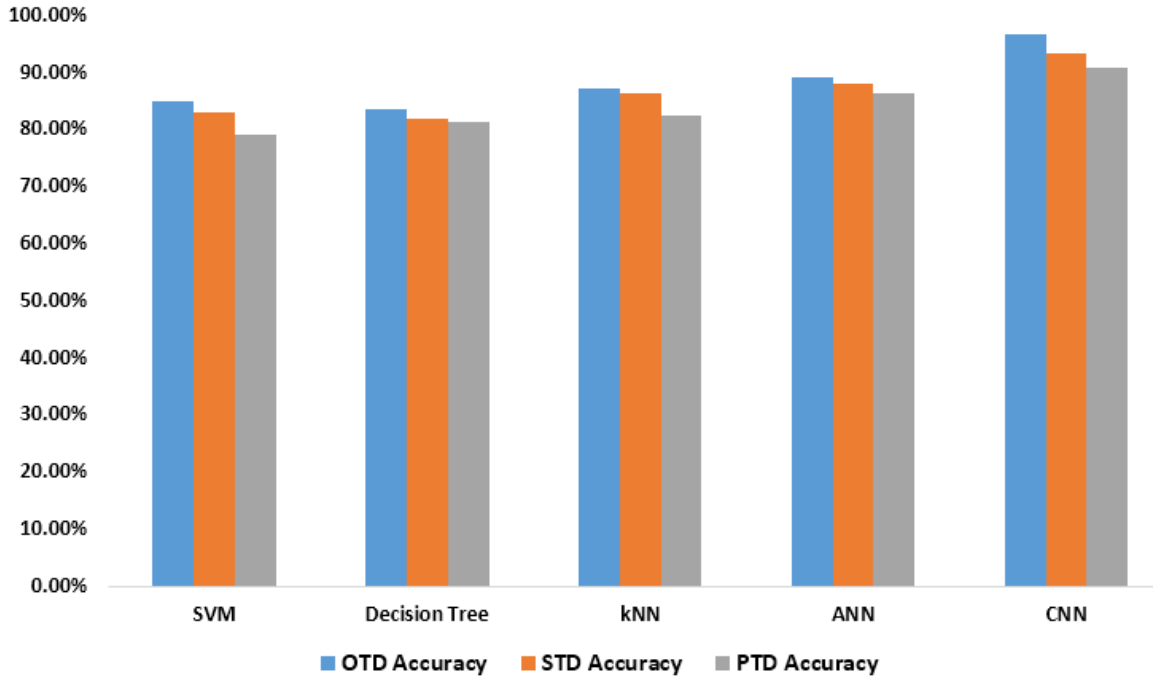

Figure 20. Accuracy comparison of various algorithms for original, synthetic and practical test data

TABLE VI. Related work versus proposed method at a glance

Method	**Data types**	**Function**	**Type of marine cetaceans considered**	**Accuracy**
CNN [10]	15 sounds	Classify the species of whistles of the whale	killer whales and long-finned pilot whales	87% (OTD)
CNN [11]	analyzing acoustic data	Detects echolocation clicks	odontocetes	91% (OTD)
ANN [12]	1475 sounds	Detect whale	bowhead whale	89% (OTD)
ASG [13]			Dolphins and small whales	82% (OTD)
DNN [14]	recordings from a geographic region	detect the vocalizations	North Atlantic right whales	92% (OTD)
MEMD and GFCCs techniques [15]	modified DNN with multi-dimensional fusion features	acoustic target identification	Underwater acoustic target	94.3% (OTD)
CNN [16]	Acoustic signals	to classify and detect whale	beluga whales	93% (OTD)
Proposed Method	84 sounds of whistles and songs. OTD, STD and PTD have been considered	Detect and classify marine cetaceans	22 species whales, dolphins and porpoises	96.60% (OTD), 93.38% (STD), 90.79% (PTD)

collected to generate the dataset. Among 6003 spectrograms of these audio files, a total of 2000 spectrograms (30% of total spectrograms) have been considered as OTD and then STD and PTD have derived from the OTD to validate the proposed method. As the supervised learning method has been implemented in detection and classification, the CNN has detected the species of marine cetaceans with 96.60% of accuracy and 0.97 of precision in this case. To make the case a little bit practical, the spectrograms of OTD have been slightly modified to obtain the STD using the image

augmentation process. This type of data may be appeared in hydrophones in the deep ocean due to ocean environmental effects. With the STD, the CNN has detected the species of marine cetaceans with 93.38% of accuracy and 0.93 of precision. To make the case more practical, the under water background noise of deep sounds have mixed up with the original waveform audio file (.wav). When the hydrophone detects the sounds of the ocean, this type of data has captured from the ocean. So, the CNN should respond to these types of test data to detect and classify the species of marine cetaceans. But the training sets that are used to train the network is the original sounds of clicks, whistles or songs of 22 species of marine cetaceans. So in this case, false detection may occur in the proposed CNN method. Simulation reveals that the performance of the proposed CNN method has reduced to 90.79% of accuracy and 0.89 of precision in the detection and classification of marine cetaceans. So if these practical data are used as well to train the neural network, the performance will increase in detecting the marine cetaceans with improved accuracy. But as there has no dataset of the sounds of clicks, whistles or songs of marine cetaceans, there will be some limitations to train the neural network with a huge amount of training sets. In future, the few-shot learning may be a solution by which the network can be trained with few training sets and can conduct better performance with unseen data.

References

[1] N. Jafrin, A. N. M. Saif, and M. I. Hossain, "Blue economy in bangladesh: proposed model and policy recommendations," *Journal of Economics and Sustainable Development*, vol. 7, no. 21, pp. 131–135, 2016.

[2] M. G. Hussain, P. Failler, A. A. Karim, and M. K. Alam, "Major opportunities of blue economy development in bangladesh," *Journal of the Indian Ocean Region*, vol. 14, no. 1, pp. 88–99, 2018.

[3] M. R. Rahman, "Blue economy and maritime cooperation in the bay of bengal: Role of bangladesh," *Procedia engineering*, vol. 194, pp. 356–361, 2017.

[4] A. Bari, "Our oceans and the blue economy: Opportunities and challenges," *Procedia engineering*, vol. 194, pp. 5–11, 2017.

[5] I. Paro, K. Dispur, P. B. Sallyan, R. Bhimphedi, D. Nepalganj, G. Prodrang, and B. Vishakhapatnam, "Toward a blue economy: A pathway for sustainable growth in bangladesh."

[6] A. Berta, *Whales, dolphins, and porpoises: A natural history and species guide*. University of Chicago Press, 2015.

[7] V. Afsal, K. Yousuf, B. Anoop, A. Anoop, P. Kannan, M. Rajagopalan, and E. Vivekanandan, "A note on cetacean distribution in the indian eez and contiguous seas during 2003-07," *Journal of Cetacean Research and Management*, vol. 10, no. 3, pp. 209–216, 2008.

[8] A. Alling, "Records of odontocetes in the northern indian ocean(1981-1982) and off the coast of sri lanka(1982-1984)." *Journal of the Bombay Natural History Society. Bombay*, vol. 83, no. 2, pp. 376–394, 1986.

[9] B. D. SMITH, B. AHMED, R. M. MOWGLI, and S. STRINDBERG, "45 species occurrence and distributional ecology of nearshore cetaceans in the bay of bengal, bangladesh, with abundance estimates for irrawaddy dolphins orcaella brevirostris and finless porpoises neophocaena phocaenoides," *J. Cetacean Res. Manage*, vol. 10, no. 1, pp. 45–58, 2008.

[10] J.-j. Jiang, L.-r. Bu, F.-j. Duan, X.-q. Wang, W. Liu, Z.-b. Sun, and C.-y. Li, "Whistle detection and classification for whales based on convolutional neural networks," *Applied Acoustics*, vol. 150, pp. 169–178, 2019.

[11] W. Luo, W. Yang, and Y. Zhang, "Convolutional neural network for detecting odontocete echolocation clicks," *The Journal of the Acoustical Society of America*, vol. 145, no. 1, pp. EL7–EL12, 2019.

[12] J. R. Potter, D. K. Mellinger, and C. W. Clark, "Marine mammal call discrimination using artificial neural networks," *The Journal of the Acoustical society of America*, vol. 96, no. 3, pp. 1255–1262, 1994.

[13] B. M. Howe, "Acoustic seaglider," WASHINGTON UNIV SEATTLE APPLIED PHYSICS LAB, Tech. Rep., 2008.

[14] Y. Shiu, K. Palmer, M. A. Roch, E. Fleishman, X. Liu, E.-M. Nosal, T. Helble, D. Cholewiak, D. Gillespie, and H. Klinck, "Deep neural networks for automated detection of marine mammal species," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[15] X. Wang, A. Liu, Y. Zhang, and F. Xue, "Underwater acoustic target recognition: a combination of multi-dimensional fusion features and modified deep neural network," *Remote Sensing*, vol. 11, no. 16, p. 1888, 2019.

[16] M. Zhong, M. Castellote, R. Dodhia, J. Lavista Ferres, M. Keogh, and A. Brewer, "Beluga whale acoustic signal classification using deep learning neural network models," *The Journal of the Acoustical Society of America*, vol. 147, no. 3, pp. 1834–1841, 2020.

[17] T. Akhter, M. A. Islam, and S. Islam, "Artificial neural network based covid-19 suspected area identification," *Journal of Engineering Advancements*, vol. 1, no. 04, pp. 188–194, 2020.

[18] T. P. L. Foundation, "Cetaceans, voices in the sea," 2007-2021. [Online]. Available: https://voicesinthesea.ucsd.edu/index.html

[19] U. of Rhode Island and I. S. Center, "Galleries: Marine mammals, discovery of sound in the sea," 2002-2021. [Online]. Available: https://dosits.org/galleries/audio-gallery/marine-mammals/

[20] N. Oceanic and U. Atmospheric Administration, "Noaa: Fisheries, find a species," 1871-2019. [Online]. Available: https://www.fisheries.noaa.gov/

[21] S. D. S. G. R. P. . Z. V. Vasilev, I., "Python deep learning: Exploring deep learning techniques and neural network architectures with pytorch, keras, and tensorflow." 2019.

[22] . M. D. G. Proakis, J. G., "Digital signal processing," 2004.

[23] ——, "Matlab deep learning with machine learning, neural networks and artificial intelligence," 2017.

[24] M. A. Islam, T. Akhter, A. Begum, M. R. Hasan, and F. S. Rafi, "Brain tumor detection from mri images using image processing."

[25] M. A. Islam, M. R. Hasan, and A. Begum, "Improvement of the

handover performance and channel allocation scheme using fuzzy logic, artificial neural network and neuro-fuzzy system to reduce call drop in cellular network," *Journal of Engineering Advancements*, vol. 1, no. 04, pp. 130–138, 2020.

Md. Ariful Islam Md. Ariful Islam earned his B.Sc.(2015) and M.Sc.(2017) degree in Electrical and Electronic Engineering at the University of Dhaka. Now he has been working as a faculty member at the Department of Robotics Mechatronics Engineering, University of Dhaka. His research interest includes Deep Learning, Fuzzy Logic Systems, Medical Robots, Robotics in Blue Economy and Hybrid Electric Vehicle Modeling.

Mosa. Tania Alim Shampa Mosa. Tania Alim Shampa earned his B.Sc (2019) degree in Oceanography at the University of Dhaka. Now she has been conducting the Master of Science program at the Department of Oceanography, University of Dhaka. Her research interest includes Climate Change, Satellite Oceanography and Coastal Processes.