Machine Intelligence Research
www.mi-research.net

19(1), February 2022, 52-62
DOI: 10.1007/s11633-022-1318-3

Probability Enhanced Entropy (PEE) Novel Feature for Improved Bird Sound Classification

Ramashini Murugaiya[1,2] Pg Emeroylariffion Abas[1] Liyanage Chandratilak De Silva[1]

[1] Faculty of Integrated Technologies, Universiti Brunei Darussalam, Bandar Seri Begawan BE1410, Brunei Darussalam

[2] Department of Computer Science and Informatics, Uva Wellassa University, Badulla 90000, Sri Lanka

Abstract: Identification of bird species from their sounds has become an important area in biodiversity-related research due to the relative ease of capturing bird sounds in the commonly challenging habitat. Audio features have a massive impact on the classification task since they are the fundamental elements used to differentiate classes. As such, the extraction of informative properties of the data is a crucial stage of any classification-based application. Therefore, it is vital to identify the most significant feature to represent the actual bird sounds. In this paper, we propose a novel feature that can advance classification accuracy with modified features, which are most suitable for classifying birds from its audio sounds. Modified Gammatone frequency cepstral coefficient (GTCC) features have been extracted with their frequency banks adjusted to suit bird sounds. The features are then used to train and test a support vector machine (SVM) classifier. It has been shown that the modified GTCC features are able to give 86% accuracy with twenty Bornean birds. Furthermore, in this paper, we are proposing a novel probability enhanced entropy (PEE) feature, which, when combined with the modified GTCC features, is able to improve accuracy further to 89.5%. These results are significant as the relatively low-resource intensive SVM with the proposed modified GTCC, and the proposed novel PEE feature can be implemented in a real-time system to assist researchers, scientists, conservationists, and even eco-tourists in identifying bird species in the dense forest.

Keywords: Bird sounds, classification, Gammatone frequency cepstral coefficient (GTCC), probability enhanced entropy (PEE), support vector machine (SVM).

Citation: R. Murugaiya, P. E. Abas, L. C. De Silva. Probability enhanced entropy (PEE) novel feature for improved bird sound classification. *Machine Intelligence Research*, vol.19, no.1, pp.52–62, 2022. http://doi.org/10.1007/s11633-022-1318-3

1 Introduction

The bird population worldwide has been declining rapidly due to deforestation and urban developments, and consequently, bird diversity monitoring projects have received considerable attention in recent years[1]. When considering the diversity of birds, there is an incredible number of bird species and varieties spread worldwide in various habitats. Due to this and other factors, such as the harsh environment these birds live in, bird diversity monitoring has become one of the most challenging biodiversity projects. Visual identification of birds is a daunting task due to the mobile nature and size of these airborne creatures and their habitats. Recent technological developments have given a new leap of life to these very time-consuming and tedious projects; advancements in audio signal processing techniques provide an edge over birds′ visual identification, even in dense environments[2]. Indeed, it has been shown that bird sound classification is an essential stage of bird biodiversity monitoring[3]. Generally, due to bird habitats and environments, recordings of these environments may contain multiple birds′ sounds, which need to be classified into different classes for species identification purposes.

General bird sound classification research commonly follows several stages, including data collection, pre-processing, feature extraction, and training of the chosen model, and then followed by testing[4]. Only few researchers have collected their own data. Instead, many researchers have preferred to utilize the volume of freely available data on online repositories for research purposes, in order to focus on the classification task rather than the data collection process. In the pre-processing stage, unwanted noise needs to be removed using various filtering techniques. Some researchers have also implemented segmentation of bird sounds by considering silent intervals between bird vocalization present in the recordings.

The next stage is the most critical stage of any audio-based application, which is the feature extraction stage. Extracting the appropriate feature that can capture the essence of the bird sounds is crucial, as including too many redundant features would increase the complexity

Research Article
Manuscript received June 28, 2021; accepted October 28, 2021
Recommended by Associate Editor Jian-Hua Tao
Colored figures are available in the online version at https://link.springer.com/journal/11633

of the model, while excluding some features may result in reduced accuracy during classification. Researchers have used various time, frequency, cepstral domain, and image/texture-based features[5] for their works. Some standard features extracted from bird sounds in the time domain include zero crossing rate (ZCR), energy, and autocorrelation harmonic ratio, while in the frequency domain, these include spectral centroid, spectral bandwidth, mean frequency, spectral flux, spectral roll-off, and spectral flatness[4, 5]. The most used feature in the literature has been mel spectral cepstrum coefficient (MFCC)[4, 6–9]. Apart from MFCC, linear predictive cepstral coefficient (LPCC) has also been used by a few researchers for bird sound-related works[10]. Kogan and Margoliash[11], Lee et al.[12] extract both average MFCC and average LPCC features, and conclude that classification results by using MFCC features are better than LPCC features.

Even though most researchers have used the standard features, some have also focused on determining novel features suitable for bird sound classification. Tyagi et al.[13] introduce prints from computing on the spectrum by frame-wise averaging fast Fourier transform (FFT) coefficients as features to automatically identify birds with good recognition performance. Stowel and Plumbley[14] use an unsupervised feature learning method, which can learn features from four different large databases without any classifier knowledge or even the training labels. This proposed method strongly outperformed MFCC. On the other hand, Digby et al.[15] use species-specific features to recognize bird calls. The authors extract five features from each putative call event: the change in syllable periodicity throughout the call, the degree of frequency modulation within each syllable, the consistency of amplitude across the syllables, correlated acoustic energy outside the bandwidth, and a weighted combination of the five feature scores namely period score, chirp score, consistency score, bandwidth score, and combined score. These are then fed into a suitable classification algorithm for classification. Graciarena et al.[16] compute note-loop lattices from bird song waveform and extract expected note n-gram statistics from lattices. Rank normalization is later used to normalize the feature before training the model.

Ulla et al.[17] derive a template from ten standardized samples to perform template matching using the spectrogram cross-correlation method. Furthermore, Bastas et al.[18] propose a novel feature extraction algorithm called spectrogram based image frequency statistics (SIFS) to classify bird fight calls. To detect vanellus chilensis lampronotus bird species using their sounds, Ganchev et al.[19] propose a log-likelihood ratio estimator based on a Gaussian mixture model–universal background model (GMM–UBM) with 85.6% recognition accuracy. For the NIPS4B 2013 competition, Lasseck[20] uses statistical features such as file statistics, segment statistics, and segment probabilities. Later, he improves the model by adding more features and consequently, wins the LifeCLEF 2014 Bird task[21].

For the classification stage, a wide range of supervised machine learning (ML) and deep learning algorithms have been used, such as the hidden Markov model (HMM), dynamic time warping (DTW)[10], Gaussian mixture model (GMM), K-nearest neighbor (KNN)[9], support vector machine (SVM)[3, 7], artificial neural network (ANN) and convolution neural network (CNN)[1, 2, 22].

Most of the existing works have approached the birds sound classification research from machine learning or deep learning approaches[2], by considering different classification models. However, it is also vital to handle this from the signal processing domain′s view by looking into identifying appropriate features of bird sounds that are able to give better classification results[13, 15–17, 19]. The feature must be easily extractable from the bird sounds and should be representative of the actual sound, such that it will provide high accuracy during the classification process despite using a simple classification method. Moreover, it needs to be easily extractable as it is envisaged that the classification task will be performed on a simple portable device with minimal processing power. This paper proposes a novel audio feature that can be used for such a purpose.

2 Procedure

Generally, the bird sound classification process follows several general stages, including data collection, preprocessing, feature extraction, and finally, the actual classification[4]. Fig. 1 depicts a general bird sound classification process composed of two important processes: training and testing. As such, the collected data samples have been divided into two sets: training and testing datasets. Naturally, all stages of the bird sound classification process would affect the overall classification results. It is important to choose the optimum method at each stage, by holistically considering the whole classification process. However, this paper focuses on the feature extraction stage by finding suitable features from the bird sounds that would give the optimum classification results. In this regard, identical segmentation and classification methods are used in order to make the performance comparison of different features easier.

Automated energy-based segmentation is used to remove unwanted and silent intervals from the collected bird sound data. Features from these segmented samples are then extracted to form the feature matrix, which is then used to train and test a chosen classification model for the prediction of bird classes.

2.1 Data collection and segmentation

Data collection is an important but time-consuming process when considering the nature of the bird and its habitats. New technologies such as wireless acoustic

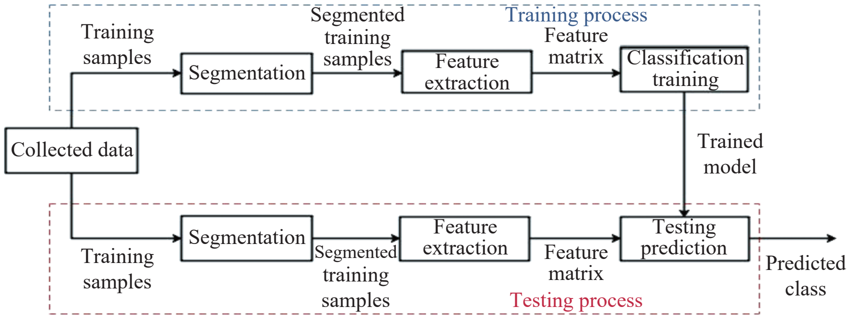

Fig. 1 Methodology used for bird sound classification

sensor networks and autonomous recording units allow researchers to get more data, even within the dense jungles of the Amazon[2]. To allow researchers to focus on classification tasks, online sound repositories, such as Xeno-Canto[23], have been made freely available to share their data and download necessary sounds for their work. Segmentation may be a necessary step in this process in order to obtain the required bird sounds and remove unwanted information[24]. Metric-based, energy-based, model-based, and hybrid segmentation methods are the available types of segmentation methods and can either be performed automatically or manually[4].

Xie et al.[24] conveniently categorize segmentation methods, albeit for frog calls, into syllable-based and sliding-window based segmentation. Sliding-window based segmentation necessitates sliding a window across the audio signal, and hence, is relatively more complex. As this paper focuses on the extraction features, a simpler, automated energy-based segmentation method, which comes under the syllable-segmentation method[24], has been adopted. The method automatically identifies silence gaps in the input audio recording by setting a threshold on the audio energy and segments bird sounds by truncating the input signal with energy exceeding the set threshold[25].

2.2 Feature extraction

The next step in both the training and testing processes is feature extraction. This paper focuses on extracting appropriate features, specifically suitable for differentiating bird sounds. In general, audio features can be extracted from several domains such as time, frequency, and cepstral domains. Many researchers have used MFCC, which is one of the most widely used cepstral features, as inputs to their classification model to give significant results, despite the fact that it has been initially developed for speech processing[26]. In recent years, researchers working in speech-related applications such as speaker identification have used Gammatone frequency cepstral coefficient (GTCC). They have shown that GTCC outperforms MFCC features to give higher accuracy results even with noisy data[27]. This paper attempts to improve GTCC filter banks, specifically for use in bird sound classification tasks. Furthermore, a novel time-domain feature is also introduced, as an additional feature, in order to enhance prediction accuracy during the classification stage.

Gammatone frequency cepstral coefficients. In recent years, GTCCs have been shown to be more robust to noise in many automatic speech recognition (ASR) systems[28, 29] and noisy environmental sound-related research[30]. GTCCs are based on Gammatone filter banks; these filter banks give a cochleagram as output, which is the frequency-time representation of the sound signal. The extraction process for GTCCs is similar to that of MFCCs, except for the mel filter bank, which has been replaced by a Gammatone filter bank[5]. GTCCs are biologically inspired modifications employing Gammatone filters with equivalent rectangular bandwidth (ERB) bands[28]. These filter bands are designed to simulate the process of a human auditory system with frequency resolution features and filtering characteristics of the cochlear basilar membrane. Fig. 2 depicts the GTCC feature extraction process.

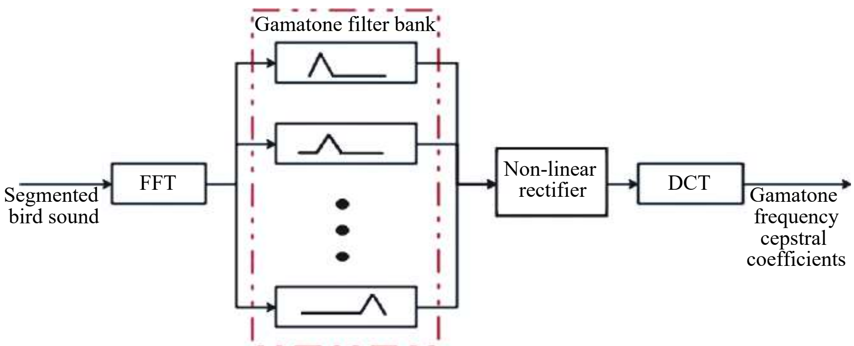

Fig. 2 GTCC feature extraction process

The segmented bird sound sample is passed as input to the GTCC feature extraction process. The segmented bird sound sample is passed as input to the GTCC feature extraction process, and it is converted into its frequency domain representation using fast Fourier transform (FTT). Gammatone filter banks are then applied. Gammatone filters are linear approximations of the filtering function performed by the cochlea in the inner ear, whereby the ear analyzes the sub-bands more delicately at higher frequencies. These linear filters are used to simulate the motion of the basilar membrane within the cochlea as a function of time. They are described by an impulse response that produces a gamma distribution and sinusoidal tone. The output of each filter models the frequency response of the basilar membrane at a particular place. The filter bank is defined such that the filter center frequencies are distributed across frequencies in proportion to their bandwidth, known as the ERB scale, with the ERB scale approximately logarithmic[31]. A Gammatone filter with a center frequency f_c is defined as

$$g(t) = at^{o-1}\mathrm{e}^{-2\pi bt}\cos(2\pi f_c + \varphi) \tag{1}$$

where t refers to the time, φ is the phase (usually set to zero), constant a controls the gain, and o defines the order of the filter.

The attenuation factor of the i-th filter is represented by the factor b, which is defined in (2). The factor b determines the decay rate of the impulse response across the filter bandwidth, with the bandwidth of each filter related to the human auditory critical band.

$$b = 25.17\left(\frac{4.37f_c}{1\,000} + 1\right). \tag{2}$$

To obtain a representation similar to spectrograms, a set of Gammatone filters, often referred to as channels, with different center frequencies, are used to create a Gammatone filter bank. The Gammatone filter bank emulates human hearing by simulating the impulse response of the auditory nerve fiber, with its shape resembling a tone $\cos(2\pi \mathrm{f_c} + \varphi)$ modulated with a gamma function $\mathrm{e}^{-2\pi \mathrm{bt}}$[27]. Finally, the discrete cosine transform (DCT) of the signal is taken to produce cepstral coefficients. Equations (1) and (2) define the Gammatone filter with a center frequency f_c and the bandwidth calculation, respectively. In this paper, the GTCC filter bands have been modified according to the bird sounds to retrieve the sound′s factual information.

Probability enhanced entropy (PEE) feature. This paper proposes a novel feature called PEE to advance the birds′ sound classification task. When analyzing the sound signals of different species, it has been observed that different bird species have different characteristics in terms of the randomness of the sound. In order to capture this difference in characteristics, the PEE feature has been introduced, which takes into account the probability of the sound residing at different energy intensities. In general, the more random the input signal, the higher the entropy. This feature has five significant steps, as shown in Fig. 3. The segmented i-th bird sample $x_i(n)$ is taken as input and normalised between −1 to 1 as follows:

$$x_{i_{\mathrm{norm}}}(n) = \frac{x_i(n) - x_{\min}}{x_{\max} - x_{\min}}. \tag{3}$$

It is then quantized into L decision levels ($A_1, A_2, \cdots, A_L$) by dividing the amplitude range for $x_{i_{\mathrm{norm}}}(n)$, with the size of each interval as $^2\!/\!_L$ intervals. The number of occurrences in each level $n_i(A)$ is counted, followed by probability calculation of each level as in (4).

$$P(A_i) = \frac{n(A_i)}{\sum\limits_{j=1}^{L} n(A_j)} \tag{4}$$

where $P(A_i)$ is the probability of occurrence of the i-th level, and $n(A_i)$ is the number of occurrences on the i-th level. Finally, based on the probability of each level, entropy is calculated using

$$F_i(S) = -\sum_{j=1}^{L} P(A_j)\log_2(P(A_j)) \tag{5}$$

where $F_i(S)$ is the probability enhanced entropy feature of the sampled bird sound $x_i(n)$.

2.3 Classification

Support vector machine (SVM), developed in the early 1990s as a non-linear solution for classification tasks, has been used as the classifier model for the classification of bird sounds. Admittedly, SVM is not the most advanced classification method. However, its robustness against error, its ability to learn well even with fewer features, and lower computational complexity compared to other ML methods, such as neural networks, are just some of the reasons for choosing SVM for the classification task. Support vector classification (SVC) attempts to

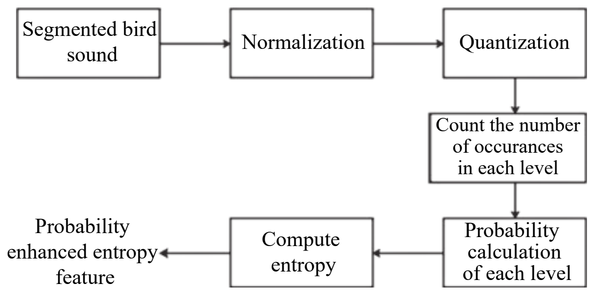

Fig. 3 Probability enhanced entropy feature extraction process

find the decision function that can adequately separate data with a perfect generalization.

SVM algorithms use a set of mathematical functions defined as their kernel, which takes data as input and transforms it into the required form[32]. The kernel functions then return the inner product between two points in a suitable feature space. Equation (6) defines the kernel $K(\bar{x})$,

$$K(\bar{x})=\begin{cases}1, & \text{if } |\bar{x}|\leq 1\\ 0, & \text{otherwise.}\end{cases} \tag{6}$$

Different SVM algorithms use different types of kernel functions, including linear, non-linear, polynomial, radial basis function (RBF), and sigmoid. The well-known polynomial kernel function has been used in the proposed methodology[33] and formally defined in (7).

$$K_f(x_i,c)=\left(x_i^{\mathrm{T}}\times c_i+1\right)^d \tag{7}$$

where x_i is a feature matrix and c_i is the class vector of two input spaces, and 1 is the constant that allows trade-offs to influence the higher order and lower order. Since the polynomial kernel function of order $d=3$ is used in

Table 1 List of Bornean birds considered for this paper

Class number	Bird name	Abbreviation used
1	Bornean blue flycatcher	BBF
2	Bushy crested hornbill	BCH
3	Black copped white-eye	BCW
4	Blue-headed pitta	BHP
5	Bornean spider hunter	BSH
6	Bornean tree pie	BTP
7	Bornean whistler	BW
8	Eagle (Crested serpent)	EAGLE
9	Golden naped barbet	GNB
10	Green pitta	GP
11	Golden whiskered barbet	GWB
12	Rhinocerous hornbill	RH
13	Kingfisher (collared)	KING
14	Malaysian banded pitta	MBP
15	Malaysian partridge	MP
16	Malaysian pied fantail	MPF
17	Savanna nightjar	SN
18	Owlet (Collared)	OWLET
19	Pitta (Hooded)	PITTA
20	White-crowned forktail	WCF

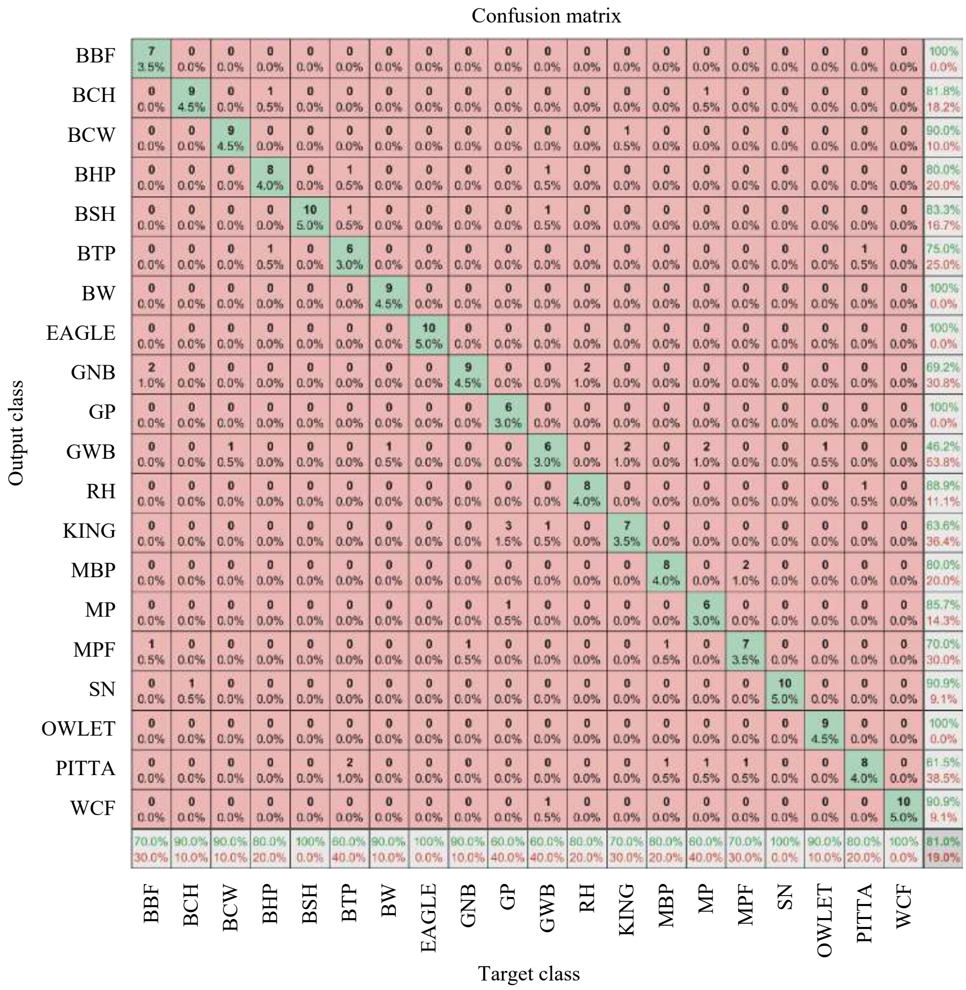

Fig. 4 Confusion matrix of testing results based on GTCC features for twenty birds

this paper, the classifier can be known as an SVM classifier with a cubic kernel function and can be defined as (8).

$$K_f(x_i, c) = \left(x_i^{\mathrm{T}} \times c_i + 1\right)^3. \quad (8)$$

3 Results and discussions

Twenty ($C = 20$) Bornean bird species, listed in Table 1, have been chosen. Bird sound recordings were collected from an online Xeno-Canto database, with the collected data separated into training and testing datasets based on an 80:20 ratio. Forty and ten samples of each bird species have been used to train and test the model, respectively, giving a total of 1 000 samples considered.

GTCC features[34] have been extracted from the segmented bird sample and then used to train and test the SVM classifier. As a result, the SVM classifier produces 81% accuracy, with some low-class wise prediction accuracy, as shown in Fig. 4. Consequently, the covered frequency range has been analyzed to study in-depth the GTCC feature, whereby it has been identified that the frequency range covered by the GTCC feature does not cover the bird sounds′ frequency range since the fundamental frequencies (f_0) observed in birds are comparatively high compared to the fundamental frequencies of a human voice[35]. Furthermore, the center frequency (F_c) plays a major role in the coverage of the frequency range of the signal. Past research works have highlighted that bird sound has a wide range of fundamental frequencies f_0 between approximately 100−12 000 Hz[36]. Therefore, it is vital to consider the entire bird frequency range when considering more birds, to represent the bird sounds.

Accordingly, the GTCC filter bank range has been improved along with the center frequency. The bandwidth (BW) has changed based on the improvement in GTCC, and the comparison is shown in Table 2. When F_c increases, BW also increases considerably. The original GTCC has 32 filter bands, while the modified GTCC has 34 filter bands, to extract useful information from the bird sounds. In general, the first few coefficients cover most of the information of the signal. Thirteen ($M = 13$) coefficients have been extracted as the first step from all the segmented samples, to make the training feature matrix X with size 800 × 13 and testing feature matrix Y with size 200 × 13. This training matrix is used to train the SVM classifier, while the testing matrix Y is used to test the trained model.

Fig. 5 shows the confusion matrix based on the modified GTCC features, giving overall prediction accuracy of 86% prediction for the twenty bird species. Of the twenty bird species, six bird species: BSH, GP, GWB, RH, SN, and WCF, have been predicted with 100% accuracy. Other bird species, including BSH, BCW, BSH, KING, MBP, and OWLET birds, reported 90% accuracy. Two samples of birds for GNB, MP, and PITTA have been predicted wrongly, while three samples of birds BBF, BHP, BW, and MPF have also been wrongly predicted. The lowest prediction accuracy is for EAGLE bird species, with a prediction accuracy of 60%.

The novel feature, probability enhanced entropy proposed in this paper, is also extracted from both the training and testing datasets and then added as an additional

Table 2 Comparison of the center frequency (F_c) and bandwidth (BW) change between GTCC and modified GTCC

Number of filter bands	GTCC		Modified GTCC	
	F_c	BW	F_c	BW
1	50	30.6	100	36.2
2	82.2	34.2	138	40.4
3	118.1	38.1	181	45.0
4	158.1	42.55	228	50.2
5	202.7	47.46	281	56.1
6	252.5	52.9	340	62.6
7	308.1	59	406	69.8
8	370	65.8	479	77.9
9	439.1	73.4	561	86.9
10	516.2	81.9	653	97
11	602.2	91.3	455	108.2
12	698	101.9	869	120.7
13	805	113.7	996	134.7
14	924.2	126.8	1 138	150.3
15	1 057.3	141.4	1 296	167.7
16	1 205.6	157.7	1 473	187.2
17	1 371.1	175.9	1 670	208.8
18	1 555.7	196.2	1890	233.0
19	1 761.6	218.9	2 135	260.0
20	1 991.2	244.1	2 409	290.1
21	2 247.3	272.3	2 714	323.7
22	2 532.9	303.7	3 055	361.2
23	2 851.5	338.8	3 435	403
24	3 206.9	377.8	3 859	449.7
25	3 603.2	421.4	4 333	501.7
26	4 045.3	470.1	4 861	559.8
27	4 538.4	524.3	5 450	624.7
28	5 088.3	584.8	6 108	697
29	5 701.7	652.2	6 842	777.7
30	6 385.8	727.5	7 661	867.8
31	7 148.9	811.4	8 574	968.2
32	8 000	905	9 594	1 080.4
33	–	–	10 731	1 205.5
34	–	–	12 000	1 345

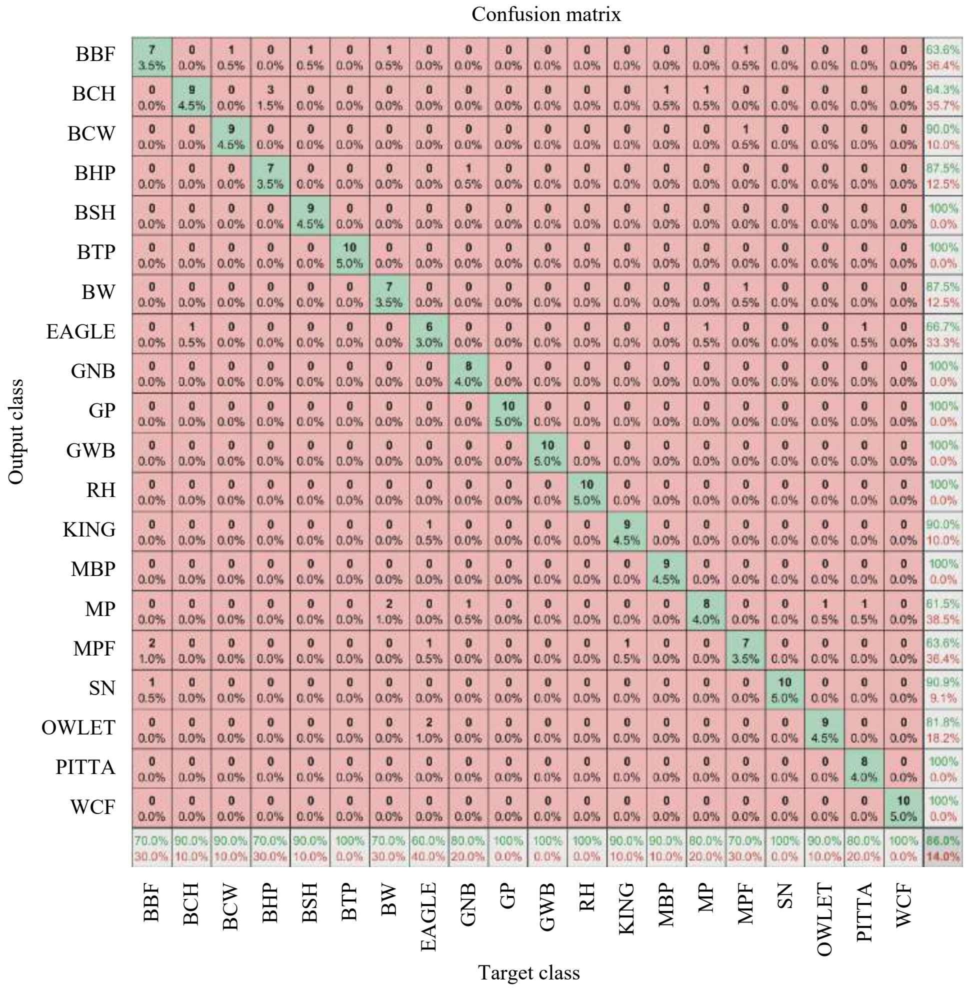

Fig. 5 Confusion matrix of testing results based on modified GTCC features for twenty birds

feature into the feature matrix to give a new training feature matrix (X_n) with size 800 × 14 and testing feature matrix (Y_n) with size 200 × 14. Fig. 6 shows the prediction results of twenty birds with the modified GTCC feature combined with the proposed PEE feature.

As demonstrated in Table 3, the novel PEE feature combined with the modified GTCC features improved the overall prediction accuracy from 86% to 89.5%. Particularly, the prediction accuracy of EAGLE bird is improved significantly from 60% to 80%. Also, seven bird classes are predicted with 100% accuracy, and seven more classes are predicted with 90% accuracy. BBF, GNB, and MPF also show significant improvements in terms of prediction accuracy.

GTCC[34] and the other two well-known cepstral features, namely, MFCC[4, 6–9] and LPCC[10–12], have also been modified based on the characteristics of bird sounds and used to classify birds separately. Table 4 compares prediction results with and without improvements to cepstral features, along with the novel PEE feature. It clearly shows that the cepstral feature′s modification positively impacts classification results with all three cepstral features. Also, the proposed probability enhanced entropy feature clearly improves the prediction results significantly. Modifying GTCC, MFFC, and LPCC using the appropriate frequency bands specific to bird sounds improved prediction accuracies from 81%, 76% and 72.5% to 86%, 77.5% and 74.5%, respectively. Appending the novel PEE feature improves prediction accuracy even further to 89.5%, 81.5%, and 79%, respectively. It is highlighted that using the modified GTCC combined with the PEE as input features to the classifier gives the highest accuracy of 89.5%.

4 Conclusions

Twenty bird sounds of Bornean species have been collected from an online repository and segmented based on their energy signals. Once unwanted components of the sounds have been removed, the samples have been divided into training and testing datasets, in the ratio of 80: 20. An extra focus has been given to the feature extraction process, as it is a vital step in the classification task′s success and has a huge impact on prediction accuracy. In recent audio-based research works, researchers have provided evidence that GTCC outperforms MFCC

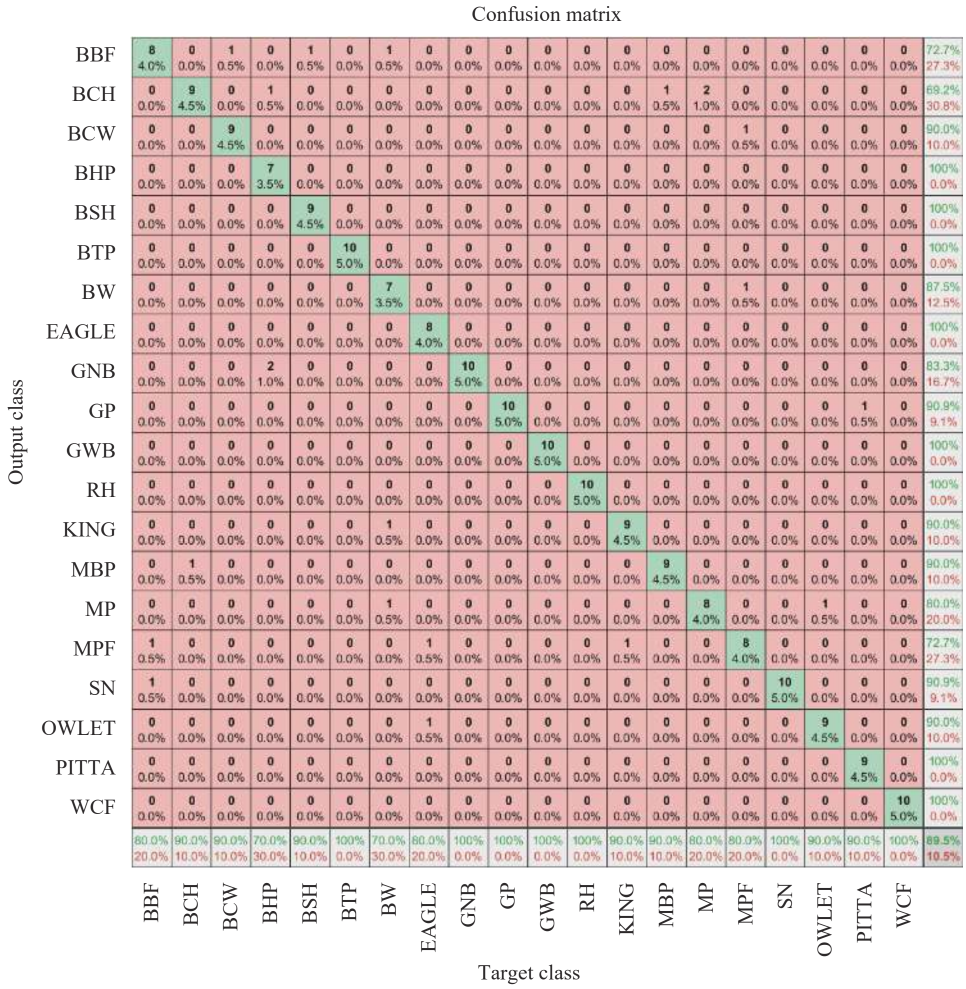

Fig. 6 Confusion matrix of testing results based on modified GTCC features combined with the novel PEE feature for twenty birds

Table 3 Comparison of class-wise prediction accuracy of twenty birds

Bird name	Prediction accuracy with modified GTCC (%)	Prediction accuracy with modified GTCC and PEE (%)
BBF	70	80
BCH	90	90
BCW	90	90
BHP	70	70
BSH	90	90
BTP	100	100
BW	70	70
EAGLE	60	80
GNB	80	100
GP	100	100
GWB	100	100
RH	100	100
KING	90	90
MBP	90	90
MP	80	80
MPF	70	80
SN	100	100
OWLET	90	90
PITTA	80	90
WCF	100	100
Total accuracy	86	89.5

Table 4 Comparison of prediction accuracy of features considered in this paper

Features used	Prediction accuracy (%)
MFCC[4, 6–9]	76
Modified MFFC	77.5
Modified MFCC with PEE	81.5
LPCC[10–12]	72.5
Modified LPCC	74.5
Modified LPCC with PEE	79
GTCC[34]	81
Modified GTCC	86
Modified GTCC with PEE	89.5

in audio speech processing applications. For this paper, the GTTC features have been modified by considering the audio characteristics of bird sound. After extracting the modified GTCC features from the bird sound samples, an SVM classifier has been used to train and then predict unknown bird sounds. For twenty bird species, 86% accuracy has been achieved. Furthermore, this work advances further the classification accuracy by introducing a novel PEE feature. Using the modified GTCC together with the novel feature, prediction accuracy increases to 89.5%.

Similar modifications to MFCC and LPCC filter-bank bandwidth, considering the frequency range of bird sounds, have also been shown to improve performance results. This is significant not only for bird sound classification but also for any audio feature extraction process, as tuning fundamental frequency range to that of the audio signal under consideration can potentially improve the process. Furthermore, using the novel PEE feature also has the potential to improve performance even further.

A public repository with the PEE feature extraction implementation can be found in the following link: https://github.com/ramashini/PEE_acoustic_feature_extraction.git.

References

[1] J. Xie, K. Hu, M. Y. Zhu, J. H. Yu, Q. B. Zhu. Investigation of different CNN-based models for improved bird sound classification. *IEEE Access*, vol. 7, pp. 175353–175361, 2019. DOI: 10.1109/ACCESS.2019.2957572.

[2] Y. Qiao, K. Qian, Z. P. Zhao. Learning higher representations from bioacoustics: A sequence-to-sequence deep learning approach for bird sound classification. In *Proceedings of the 27th International Conference on Neural Information Processing*, Springer, Bangkok, Thailand, pp. 130–138, 2020. DOI: 10.1007/978-3-030-63823-8_16.

[3] K. Qian, Z. X. Zhang, A. Baird, B. Schuller. Active learning for bird sound classification via a kernel-based extreme learning machine. *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 1796–1804, 2017. DOI: 10.1121/1.5004570.

[4] M. Ramashini, P. E. Abas, U. Grafe, L. C. De Silva. Bird sounds classification using linear discriminant analysis. In *Proceedings of the 4th International Conference and Workshops on Recent Advances and Innovations in Engineering*, IEEE, Kedah, Malaysia, 2019. DOI: 10.1109/ICRAIE47735.2019.9037645.

[5] G. Sharma, K. Umapathy, S. Krishnan. Trends in audio signal feature extraction methods. *Applied Acoustics*, vol. 158, Article number 107020, 2020. DOI: 10.1016/j.apacoust.2019.107020.

[6] O. Küçüktopcu, E. Masazade, C. Ünsalan, P. K. Varshney. A real-time bird sound recognition system using a low-cost microcontroller. *Applied Acoustics*, vol. 148, pp. 194–201, 2019. DOI: 10.1016/j.apacoust.2018.12.028.

[7] J. Ludeña-Choez, R. Quispe-Soncco, A. Gallardo-Antolín. Bird sound spectrogram decomposition through non-negative matrix factorization for the acoustic classification of bird species. *PLoS One*, vol. 12, no. 6, Article number e0179403, 2017. DOI: 10.1371/journal.pone.0179403.

[8] Y. R. Leng, H. Dat Tran. Multi-label bird classification using an ensemble classifier with simple features. In *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, IEEE, Siem Reap, Cambodia, 2014. DOI: 10.1109/APSIPA.2014.7041649.

[9] S. Fagerlund, U. K. Laine. New parametric representations of bird sounds for automatic classification. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, Florence, Italy, pp. 8247–8251, 2014. DOI: 10.1109/ICASSP.2014.6855209.

[10] Z. X. Chen, R. C. Maher. Semi-automatic classification of bird vocalizations using spectral peak tracks. *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2974–2984, 2006. DOI: 10.1121/1.2345831.

[11] J. A. Kogan, D. Margoliash. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *The Journal of the Acoustical Society of America*, vol. 103, no. 4, pp. 2185–2196, 1998. DOI: 10.1121/1.421364.

[12] C. H. Lee, C. C. Lien, R. Z. Huang. Automatic recognition of birdsongs using mel-frequency cepstral coefficients and vector quantization. In *Proceedings of International Muti Conference of Engineering and Computer Scientists*, Hong Kong, China, pp. 331–335, 2006.

[13] H. Tyagi, R. M. Hegde, H. A. Murthy, A. Prabhakar. Automatic identification of bird calls using spectral ensemble average voice prints. In *Proceedings of the 14th European Signal Processing Conference*, IEEE, Florence, Italy, pp. 1–5, 2006.

[14] D. Stowell, M. D. Plumbley. Audio-only bird classification using unsupervised feature learning. In *Proceedings of Working Notes of CLEF 2014 Conference*, Sheffield, UK, pp. 673–684, 2014.

[15] A. Digby, M. Towsey, B. D. Bell, P. D. Teal. A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, vol. 4, no. 7, pp. 675–683, 2013. DOI: 10.1111/2041-210X.12060.

[16] M. Graciarena, M. Delplanche, E. Shriberg, A. Stolcke. Bird species recognition combining acoustic and sequence modeling. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, pp. 341–344, 2011. DOI: 10.1109/ICASSP.2011.5946410.

[17] J. S. Ulloa, A. Gasc, P. Gaucher, T. Aubin, M. Réjou-Méchain, J. Sueur. Screening large audio datasets to determine the time and space distribution of Screaming Piha birds in a tropical forest. *Ecological Informatics*, vol. 31, pp. 91–99, 2016. DOI: 10.1016/j.ecoinf.2015.11.012.

[18] S. Bastas, M. W. Majid, G. Mirzaei, J. Ross, M. M. Jamali, P. V. Gorsevski, J. Frizado, V. P. Bingman. A novel feature extraction algorithm for classification of bird flight calls. In *Proceedings of IEEE International Symposium on Circuits and Systems*, IEEE, Seoul, Korea, pp. 1676–1679,

2012. DOI: 10.1109/ISCAS.2012.6271580.

[19] T. D. Ganchev, O. Jahn, M. I. Marques, J. M. De Figueiredo, K. L. Schuchmann. Automated acoustic detection of Vanellus chilensis lampronotus. *Expert Systems with Applications*, vol. 42, no. 15-16, pp. 6098–6111, 2015. DOI: 10.1016/j.eswa.2015.03.036.

[20] M. Lasseck. Bird song classification in field recordings: Winning solution for NIPS4B 2013 competition. In *Proceedings of 'Neural Information Scaled for Bioacoustics' Joint to NIPS*, Nevada, USA, pp. 176–181, 2013.

[21] M. Lasseck. Large-scale identification of birds in audio recordings notes on the winning solution of the LifeCLEF 2014 Bird Task. *In Proceedings of CEUR Workshop*, vol. 1180, pp. 643–653, 2014.

[22] J. Salamon, J. P. Bello, A. Farnsworth, S. Kelling. Fusing shallow and deep learning for Bioacoustic bird species classification. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, New Orleans, USA, pp. 141–145, 2017. DOI: 10.1109/ICASSP.2017.7952134.

[23] E. Znidersic. Audio-based bird species identification with deep convolutional neural networks. *The Journal of the Acoustical Society of America*, vol. 5, no. 4, pp. 4640–4650, 2017.

[24] J. Xie, K. Hu, M. Y. Zhu, Y. Guo. Bioacoustic signal classification in continuous recordings: Syllable-segmentation vs sliding-window. *Expert Systems with Applications*, vol. 152, Article number 113390, 2020. DOI: 10.1016/j.eswa.2020.113390.

[25] T. Kemp, M. Schmidt, M. Westphal, A. Waibel. Strategies for automatic segmentation of audio data. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal*, IEEE, Istanbul, Turkey, pp. 1423–1426, 2000. DOI: 10.1109/ICASSP.2000.861862.

[26] M. Ramashini, P. E. Abas, L. C. De Silva. A novel approach of audio based feature optimisation for bird classification. *Pertanika Journal of Science and Technology*, vol. 29, no. 4, pp. 2383–2407, 2021. DOI: 10.47836/pjst.29.4.08.

[27] H. P. Wang, C. L. Zhang. The application of Gammatone frequency cepstral coefficients for forensic voice comparison under noisy conditions. *Australian Journal of Forensic Sciences*, vol. 52, no. 5, pp. 553–568, 2020. DOI: 10.1080/00450618.2019.1584830.

[28] R. Fathima, P. E. Raseena. Gammatone cepstral coefficient for speaker identification. *International Journal of Scientific & Engineering Research*, vol. 4, no. 10, pp. 795–798, 2013.

[29] Y. Shao, Z. Z. Jin, D. L. Wang, S. Srinivasan. An auditory-based feature for robust speech recognition. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal*, IEEE, Taipei, China, pp. 4625–4628, 2009. DOI: 10.1109/ICASSP.2009.4960661.

[30] X. J. Zhao, D. L. Wang. Analyzing noise robustness of MFCC and GFCC features in speaker identification. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, Vancouver, Canada, pp. 7204–7208, 2013. DOI: 10.1109/ICASSP.2013.6639061.

[31] X. Valero, F. Alías. Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification. *IEEE Transactions on Multimedia*, vol. 14, no. 6, pp. 1684–1689, 2012. DOI: 10.1109/TMM.2012.2199972.

[32] S. Singh, R. Kumar. Histopathological image analysis for breast cancer detection using cubic SVM. In *Proceedings of the 7th International Conference on Signal Processing and Integrated Networks*, IEEE, Noida, India, pp. 498–503, 2020. DOI: 10.1109/SPIN48934.2020.9071218.

[33] R. Gholami, N. Fakhari. Support vector machine: Principles, parameters, and applications. *Handbook of Neural Computation*, P. Samui, S. Sekhar, V. E. Balas, Eds., Amsterdam, Netherlands: Elsevier, pp. 515–535, 2017. DOI: 10.1016/B978-0-12-811318-9.00027-2.

[34] M. Ramashini, P. E. Abas, K. Mohanchandra, L. C. De Silva. Robust cepstral feature for bird sound classification. *International Journal of Electrical and Computer Engineering*, vol. 12, Article number 2, 2022. DOI: 10.11591/IJECE.V12I2.pp%25p.

[35] S. Nowicki, P. Marler. How do birds sing? *Music Perception*, vol. 5, no. 4, pp. 391–426, 1998. DOI: 10.2307/40285408.

[36] F. Goller, Riede. Integrative physiology of fundamental frequency control in birds. *Journal of Physiology-Paris*, vol. 107, no. 3, pp. 230–242, 2013. DOI: 10.1016/j.jphysparis.2012.11.001.

Ramashini Murugaiya received the B. Tech. degree in information technology from Anna University, India in 2012, the M. Eng. degree in computer and communication engineering from Anna University, India in 2014, and the Ph. D. degree in systems engineering from Universiti Brunei Darussalam, Brunei Darussalam in 2021. She is now a lecturer at Uva Wellassa University, Sri Lanka.

Her research interests include audio signal processing, biomedical signal processing, bio and environmental acoustics.

E-mail: ramashini@uwu.ac.lk (Corresponding author)
ORCID iD: 0000-0001-5651-4674

Pg Emeroylariffion Abas received the B. Eng. degree in information systems engineering from Imperial College, UK in 2001, and received the Ph. D. degree in communication systems from the same institution in 2005. He is now working as an assistant professor in system engineering in Faculty of Integrated Technologies, Universiti Brunei Darussalam, Brunei Darussalam.

His research interests include data analysis, security of info-

communication systems and design of photonic crystal fiber in fiber optics communication.

E-mail: emeroylariffion.abas@ubd.edu.bn

ORCID iD: 0000-0002-7006-3838

Liyanage Chandratilak De Silva received the B. Sc. (Hons) degree from the University of Moratuwa, Sri Lanka in 1985, the M. Phil. degree in electrical and computer engineering from The Open University of Sri Lanka, Sri Lanka in 1989, the M. Eng. degree in information engineering and Ph. D degree in electrical engineering from University of Tokyo, Japan in 1992 and 1995 respectively. He was with the University of Tokyo, Japan from 1989 to 1995. From April 1995 to March 1997, he pursued his postdoctoral research as a post-doctoral researcher at ATR (Advanced Telecommunication Research) Laboratories, Japan. In March 1997, he joined National University of Singapore as a lecturer, where he was an assistant professor till June 2003. He was with the Massey University, New Zealand from 2003 to 2007 as a senior lecturer. He joined University of Brunei Darussalam in 2007 as an associate professor. Currently, he is a professor of engineering and the dean of Faculty of Integrated Technologies, University of Brunei Darussalam, Brunei Darussalam. He has over 30 years of postgraduate experience in various levels in his career in the Asia Pacific region. He has published over 190 technical papers in these areas in international conferences, journals and Japanese national conventions and holds one Japanese national patent, which was successfully sold to Sony Corporation Japan for commercial utilization. He also holds 3 U. S., and 1 Brunei patents with several patents pending. His works have been cited as one of the pioneering works in the bimodal (audio and video signal based) emotion recognition by many researchers. His papers so far have been cited by more than 3500 times (according to scholar.google.com) with an H-index of 25. He is a senior member of IEEE (USA).

His research interests include internet of things, image and speech signal processing, information theory, computer vision, data analytics pattern recognition and understanding, smart homes and smart sensors, multimedia signal processing, digital electronics.

E-mail: liyanage.silva@ubd.edu.bn

ORCID iD: 0000-0001-7128-5945