AUTOMATED BIRD DETECTION IN AUDIO RECORDINGS BY A SIGNAL PROCESSING PERSPECTIVE

Raja Shekar Kadurka
Department of Computer Science and Engineering,
University of California, USA
rajashekar562@gmail.com

Harish Kanakalla
Department of Information System Technology,
Alliant International University, California, USA
kanakallaharish@gmail.com

Abstract: In this study, an effective automated technique for the detection of bird sounds is presented in a signal processing perspective. The detection of bird sound by examining the sound patterns is the basic step for wildlife monitoring. An Automated Bird Detection (ABD) system based on Dual-tree M-band Wavelet transform (DMWT) is designed. The more intrinsic content of the audio is extracted as features by DMWT and this is the crucial stage as the extracted features directly influence the efficiency of the ABD system. It classifies the given audio signals into two classes; birds are present or not present. The sounds in the audio signals are modeled by Gaussian Mixture Model (GMM) with the help of DMWT features. The ABD system is analyzed by changing the DMWT decomposition level, and Gaussian components used to model each fault. Results show that the ABD system achieves 97.82% accuracy by 3rd level DMWT features when modeled by 16 Gaussian components.

Keywords: Bird detection, wavelet transform, dual tree m-band wavelets, Gaussian mixture model.

I. INTRODUCTION

In recent years, bird sound classification is the main research area in ecology, signal processing and wildlife monitoring. The significant advances in machine learning algorithms make the bird sound classification as a pattern recognition or computer vision systems. An active learning based ABD system is discussed in [1]. It uses extreme learning machine and Support Vector Machine (SVM) classifiers to classify the bird sounds. A multivariate statistics based ABD system is described in [2]. At first, the bird sounds are preprocessed using linear predictive coding and windowed Fourier transform. Then back propagation learning is employed for the classification in two-layer perceptrons.

Wavelets based recognition of bird sounds in discussed in [3]. The temporal and spectral information of bird sounds are extracted by wavelets and then they are classified using self organizing map and MultiLayer Perceptron (MLP). Before extracting bird sound features, the audio signals are normalized with zero mean. A multi-label approach is discussed in [4] for bird sound classification by an acoustic classification system. It represents the audio signal into a bag-of-instances such as spectrum bin index, values of spectrogram and its variance. Random Forest (RF) classifier is employed and also 2D time-frequency segmentation is used as a preprocessing technique before feature extraction.

Linear Discriminant Analysis (LDA) based ABD system is discussed in [5]. Initially, thirty-five features are extracted from the bird sounds and dominant features are selected by LDA. Then, nearest centroid classifier is used for classifying the bird sounds. Artificial Neural Network (ANN) based ABD system is discussed in [6]. After preprocessing, the power spectral density of bird signal is modeled for the classification using ANN.

Wavelet based features are discussed in [7] for bird sound classification. It uses two different classifiers; SVM and Convolutional Neural Network (CNN) for classifying the sounds. It uses not only the wavelet features from 3 different Daubechies and Symlet filters but also Fourier based features from mel spectrogram. A bird classifier system is described in [8] using machine learning approach. It uses audio features from the mel spectrogram for classifying bird sounds using K-nearest neighbour, MLP, Bayes and RF algorithms.

An investigation is made in [9] using different CNNs for bird sound classification. Three different representation from the spectrogram; Mel, harmonic and percussive component based are employed to extract features. Then the prediction outputs from two different CNNs are fused to form the final prediction. Spectrograms are directly fed to CNN for bird sound classification in [10]. A series of preprocessing steps before extracting spectrograms is employed such as pre-emphasis, framing the audio signal, removal of silence.

An unsupervised model by Hidden Markov Model (HMM) is discussed in [11] to classify bird sounds. At first each bird sound is modeled by the HMM and then a hybrid deep network with HMM is designed for the classification. CNN based system in [12] classifies the bird sounds using transfer learning. It uses pretrained MobileNet for the classification. The spectro-temporal features from the preprocessed signal are extracted and the obtained spectrogram images are rendered to colour map for effective classification.

In this study, an efficient ABD system is designed for the classification of bird sound. The salient feature of the ABD system is the use of advanced wavelet for the extraction of dominant features from different sounds. Also, a probabilistic classifier is used as a machine learning approach for the classification. The rest of the paper is as follows: The various methods and materials used for the ABD system is discussed in section 2. The performance of the ABD system to identify the bird sounds is discussed in section 3, and finally, conclusion is made from the results.

II. METHODS AND MATERIALS

The main aim of the ABD system is to identify whether the bird sound is present in the given signal in a signal processing perspective. It consists of three stages; preprocessing, DMWT feature extraction and modelling the bird sound by GMM for classification. Fig. 1 shows the framework of the ABD system using DMWT and GMM.

A. Pre-processing

To improve the performance of any pattern recognition or machine learning approaches, preprocessing is employed if required. This study employs two steps; explosion and pre-emphasis filtration to preprocess the sound signals. The former one is used to remove the ambiguity in the fault signals, and the later one is used to suppress the lower frequency components in the fault signals.
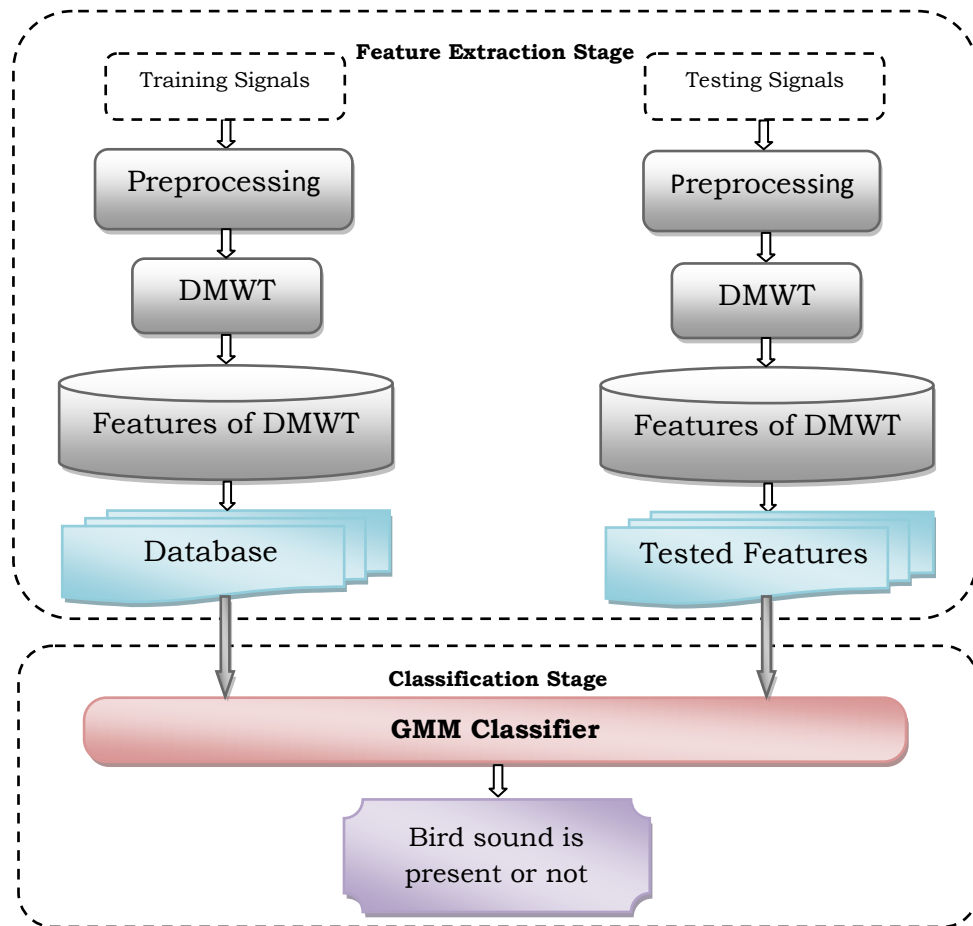
Fig. 1 Framework of the ABD system using DMWT and GMM classifier

In order to consider the small amplitudes in the fault signals, which may play a vital role in the classification stage, explosion is employed by representing the samples in fault signals to 16 bit. Then, the pre-emphasis filter is applied to the exploded signal which is given by

$$y_t = \alpha x_t + (1 - \alpha)x_{t-1} \tag{1}$$

where α is the coefficient of pre-emphasis and x is the fault signal. The application of the pre-emphasis filter increases the amplitude of high frequency component and suppresses the amplitude of the lower frequency component of the given fault signal. Fig. 2 shows the air filter fault signals and preprocessed signals.

B. *DMWT Features*

The more intrinsic content of the signal is extracted as features in this stage. This is the crucial stage as the extracted features directly influence the efficiency of the ABD system. This study uses DMWT as a feature extraction technique as it has more representation power than wavelets [13-14]. It uses a Hilbert pair of wavelets for decomposition signals. The decomposition procedure of DMWT consists of three steps; pre-filtration, DWT decomposition using M-band

13

filters and direction extraction. Fig. 3 shows the DMWT decomposition procedure. The local directions are extracted by the linear combination of primal and dual tree filters. These filters are obtained by parallel processing of M-band filters. More information about DMWT can be found in [13-14].
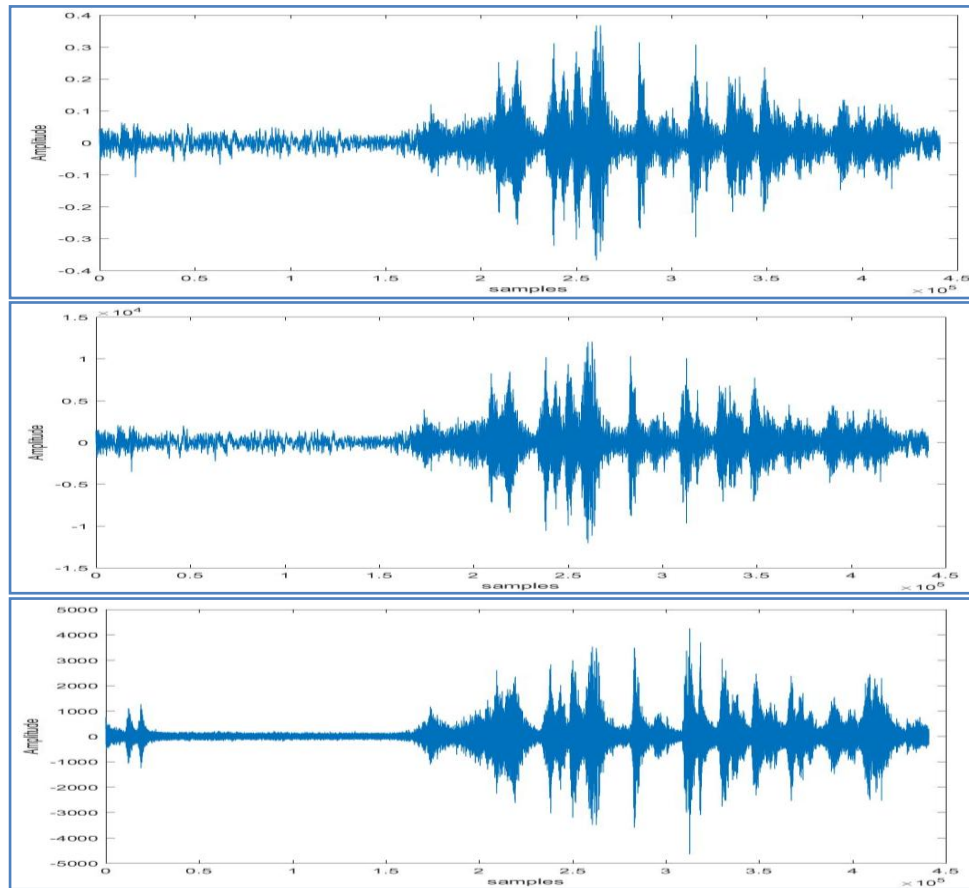

Fig. 2 Input signal (top row), Exploded signal (middle row), Pre-emphasis filtered signal (bottom row)

The DMWT decomposition is not directly applied to the fault signal, instead hamming window concept is applied to divide the whole signal into samples of predefined time frame (25 milliseconds) with frame shift of 10 milliseconds. Let us consider the preprocessed signal Ps, q is the sample point of window applied, and k is the window length, then the resulting speech signal in a single frame is defined by,

$$P_s(n, q) = P_s(n)w(q - n) \tag{2}$$

where w is the windowing function which is a hamming window defined, by

$$w(n) = 0.54 - 0.46\,cos\frac{2\pi(n-1)}{k-1}\,n = 0,1,\dots\dots\dots k = -1 \tag{3}$$

```
┌──────────────┐         ┌──────────┐         ┌──────────┐
│    Pre-      │ c_{0,0,0}│  M-band  │ c_{1,0,0}│  M-band  │
│  filtering   │─────────▶│  Filter  │─────────▶│  Filter  │
│    (F₁)      │          │   bank   │          │   bank   │
└──────────────┘          └──────────┘          └──────────┘
```

$c_{0,0,0}$ $c_{1,0,0}$

Pre-filtering (F_1) — M-band Filter bank — M-band Filter bank — Linear combination of the sub bands

$c_{0,0,0}^{H}$ $c_{1,0,0}$

Pre-filtering (F_2) — Dual M-band Filter bank — Dual M-band Filter bank — Linear combination of the sub bands
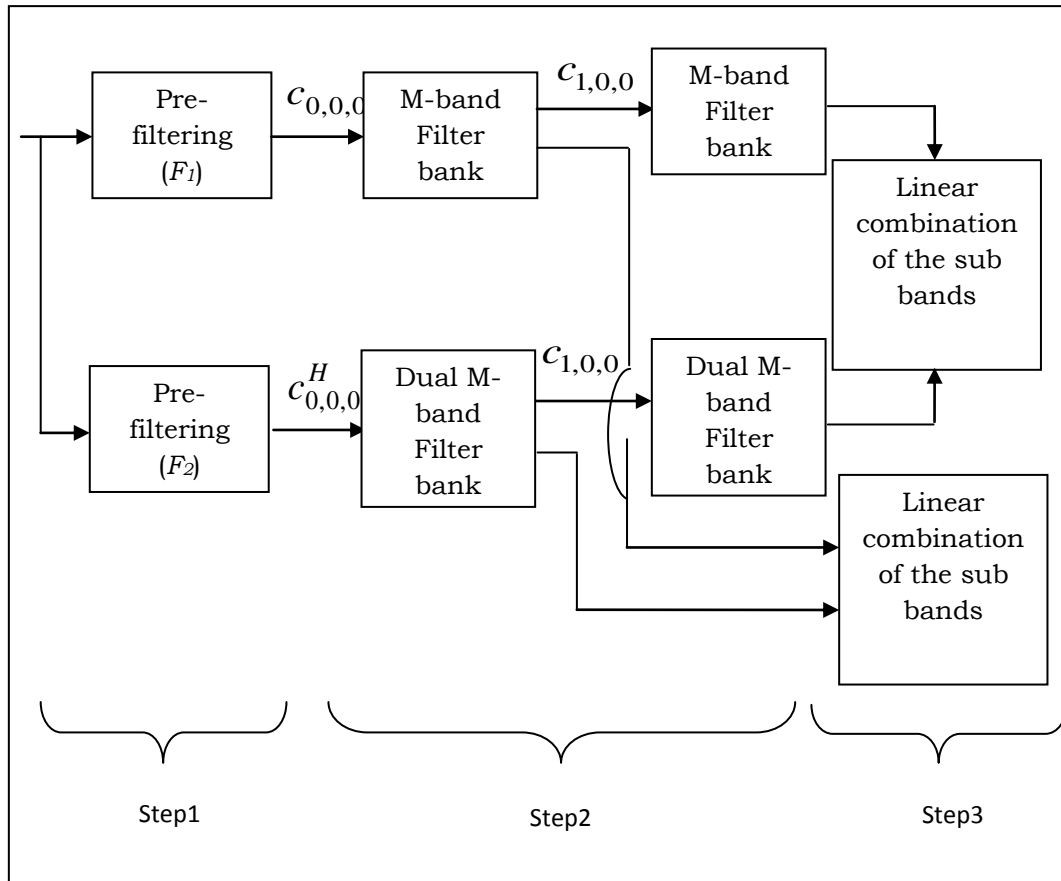
Step1   Step2   Step3

Fig. 3 DMWT decomposition steps

This process generates many samples for a particular input signal. Fig. 4 shows the first two samples of pre-emphasis filtered signal in Fig. 4 obtained by the hamming window concept.
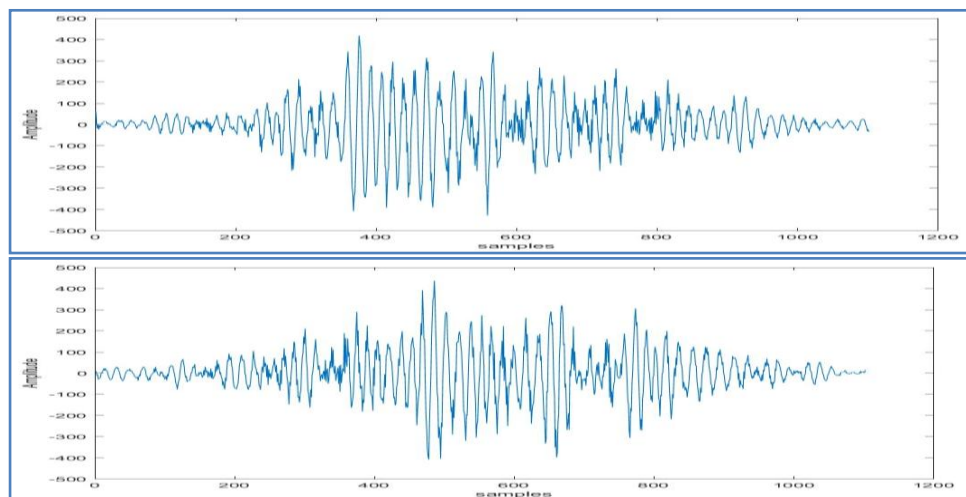
Fig. 4 First two samples of the input signal by Hamming window

To extract features form each sample, DMWT decomposition is applied, and energy features are computed from the sub-bands of the decomposed signal. It is defined by,

$$e_k = \frac{1}{M}\sum_{i=1}^{M}|x_k(i)| \tag{4}$$

where M is the length of the sample k and $x_k(.)$ is the DMWT sub-band coefficients. The features from samples are used to model the bird sounds by GMM. Fig. 5 shows the sub-bands of 1st level DMWT for the first samples in Fig. 4.
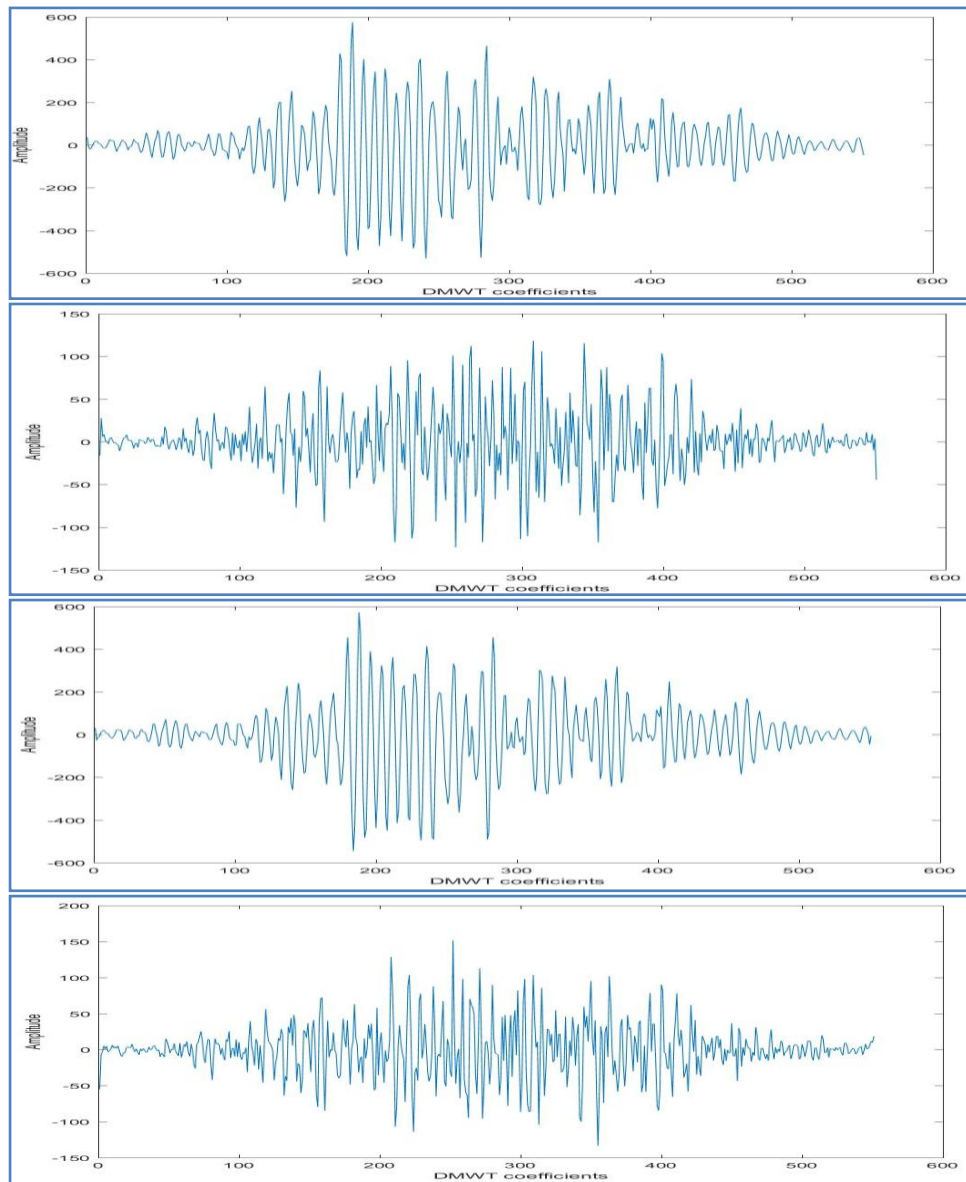

Fig. 5 DMWT sub-bands at 1st level – Primal Tree (top two rows) and Dual tree (bottom two rows)

C.   *GMM for Bird Sound Classification*

GMM classifier classifies the signal based on the probability theory [15]. It finds the posterior probability of a testing signal with the known signals in the training database. Generally, a class of signals is modeled using M number of Gaussians ($G_m = \{\gamma_1, \gamma_2, \gamma_3, \ldots \ldots \gamma_M\}$). The relative weights of each model are computed using Expectation and Maximization (EM) algorithm [16]. It fits each Gaussian model to the training data points. Then, the conditional probability is given by,

$$p(T|\overline{V}) = \sum_{i=1}^{M} c_i \cdot \gamma_i(T) \tag{5}$$

where $\gamma_i(T)$ and c_i are the N-variate Gaussian function and mixture weights respectively. More information about GMM can be found in [15]. The best fit for fault signal is identified from the posterior probability of the fault signal with the training samples by Bayes rule.

III. RESULTS AND DISCUSSIONS

The performance of the ABD system using DMWT and GMM is discussed in this section. The signals from freefield1010 database [17] are used for the evaluation. These signals are collected from different locations and environments and each signal is annotated as whether the bird sound is present or not. It has 5755 signals without bird sounds and 1935 signals with bird sounds. As the ABD system is treated as a classification system, it requires a vast number of training and testing samples. To create such samples, 10-fold cross-validation is employed. Figure 6 shows the *K*-fold cross-validation approach.
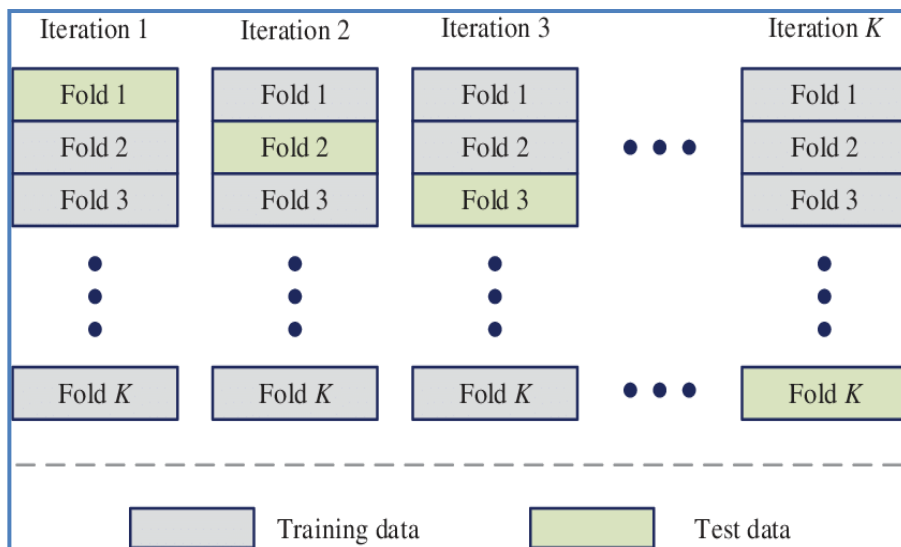

Fig. 6 k-fold cross-validation

The performance is analyzed in terms of average classification accuracy which is defined by,

$$\text{Classification Accuracy} = \frac{\#\text{signals correctly classified with bird sounds or normal}}{\text{Total number of signals tested}} \times 100 \quad (1)$$

Tab. 1 shows the classification accuracy of ABD system using DMWT and GMM by varying DMWT decomposition level and number of Gaussians model used to model the bird sounds.

TABLE. 1 Average classification accuracy of ABD system using DMWT and GMM

DMWT decomposition level	#Gaussian Models used to model			
	G4	G8	G16	G32
1	71.55	78.24	84.67	80.90
2	75.62	81.27	91.00	87.09
3	85.85	92.61	**97.82**	94.90
4	79.39	87.31	94.47	90.56
5	74.17	80.46	90.48	85.80

It is inferred that the ABD system gives an average classification accuracy of 97.82% when the features of 3rd level DMWT are modeled by 16 Gaussian components. Also, it is observed that the increase in DMWT increases the average recognition accuracy of ABD system. After 3rd level, the accuracy of the ABD system reduces by the redundant information of DMWT coefficients. Each fault is modeled by 4 Gaussian components at first and the ABD system provides below 90% for all DMWT features. Hence, the number of Gaussian components is increased to model the bird sounds. It is found that the ABD system provides a better result when the bird sounds are modeled by 16 Gaussian components.

IV. CONCLUSION

In this study, an efficient ABD system to detect the bird sounds is presented using DMWT and GMM. It uses DMWT, an extension of wavelet as it representation provides more information in the signal than conventional wavelet. Due to its multi-scale nature, the signals obtained from outdoors are represented up to 5 level of DMWT decomposition. GMM is used to model the features from DMWT to classify the bird sounds in the signal. The ABD system is analyzed using DMWT features at various decomposition levels and varying Gaussian components used to model the bird sounds. The ABD system with DMWT and GMM provides promising results with a maximum accuracy of 97.82%.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1]. K. Qian, Z. Zhang, A. Baird and 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, 2017, pp.1796-1804.

[2]. A.L. McIlraith and H.C. Card, "Birdsong recognition using backpropagation and multivariate statistics," IEEE Transaction on Signal Processing, Vol. 45, No. 11, 1997, 2740–2748.

[3]. A. Selin, J. Turunen, and J.T. Tanttu, "Wavelets in recognition of bird sounds," EURASIP Journal of Advances in Signal Processing, Vol. 2007, 2007, pp.1-9.

[4]. F. Briggs, B. Lakshminarayanan, L. Neal, X. Z. Fern, R. Raich, S. J. Hadley, A.S. Hadley, and M.G. Betts, "Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach," J The Journal of the Acoustical Society of America, Vol. 131, No. 6, 2012, 4640–4650.

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

[6]. M.M.M. Sukri, U. Fadlilah, S. Saon, A.K. Mahamad, M.M. Som and A. Sidek, "Bird Sound Identification based on Artificial Neural Network," 2020 IEEE Student Conference on Research and Development, 2020, pp. 342-345.

[7]. B.N. Suhas et al., "Automatic bird sound detection in long range field recordings using Wavelets & Mel filter bank features," 2nd International Conference on Cognitive Machine Intelligence, 2020, pp. 218-226.

[8]. Y. Jadhav, V. Patil and D. Parasar, "Machine Learning Approach to Classify Birds on the Basis of Their Sound," International Conference on Inventive Computation Technologies , 2020, pp. 69-73.

[9]. J. Xie, K. Hu, M. Zhu, J. Yu and Q. Zhu, "Investigation of Different CNN-Based Models for Improved Bird Sound Classification," IEEE Access, Vol. 7, 2019, pp. 175353-175361.

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

[11]. P. Jancovic and M. Kokuer, "Bird Species Recognition Using Unsupervised Modeling of Individual Vocalization Elements," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 27, No. 5, 2019, pp. 932-947.

[12]. Á. Incze, H. Jancsó, Z. Szilágyi, A. Farkas and C. Sulyok, "Bird Sound Recognition Using a Convolutional Neural Network," 16th International Symposium on Intelligent Systems and Informatics, 2018, pp. 000295-000300.

[13]. C. Chaux, L. Duval, and J.C. Pesquet, "Hilbert pairs of M-band orthonormal wavelet bases," 12th European Signal and Image Processing Conference, 2004, pp. 1187–1190.

[14]. I. Bayram, and I.W. Selesnick, "A simple construction for the m-band dual-tree complex wavelet transform," 12th Signal Processing Education Workshop, 2006, pp. 596-601.

[15]. D.A. Reynolds and R.C. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," IEEE Transactions on Speech and Audio Processing, Vol. 3, No. 1, 1995, pp. 72–83.

[16]. C.M Bishop, "Pattern recognition and machine learning", Springer, Chapter 9, Vol. 1, 2006, pp.43-50.

[17]. D. Stowell and M.D. Plumbley, "An open dataset for research on audio field recording archives: freefield1010," 2013, pp.1-10.