Technical Journal, University of Engineering and Technology (UET) Taxila, Pakistan Vol. 29 No. 2-2024
ISSN:1813-1786 (Print) 2313-7770 (Online)

Birds' Sound Classification Using Acoustic Signals

S. H. Amjad [1], D. Shahid [2], I. Mahmood [3], W. Ali [4], A. Ghaffar [5]

[1, 2, 4] *Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan*
[3] *Department of CSMIS, Oman College of Management and Technology, Oman*
[5] *Department of Software Engineering, College of Computing, Umm Al-Qura University, Saudi Arabia*

[1] syed.haseeb@students.uettaxila.edu.pk

Abstract- The bird sound classification is a significant aspect of bioacoustics research, wildlife conservation, and an advanced system for analyzing bird populations. In this article, a framework is suggested to categorize bird species based on their sound vocals through acoustic signals. The framework comprises of fusion of features extracted from Acoustic Local Ternary Patterns (ALTPs), Mel-Frequency Cepstral Coefficients (MFCCs), and Linear Predictive Coding (LPC). After the signal representation is performed, we categorize the signal by different classifiers. We used a publicly available research-oriented Bird Sound dataset comprising 781 sound specimens from 18 diverse bird species. This research not only plays an important role in the field of bioacoustics but can also act as a valuable tool supporting wildlife conservation. Moreover, the vocalization detection method can also give an intuition of bird's behavior into their ecological roles, habitat inclination, and adaptation to environmental changes. The proposed sound classification technique achieved a higher accuracy rate of 97.4% using the Bagged Trees Ensemble method, in comparison with the research conducted on the same 18 classes of Bird Sound dataset. The experimental results depicted the reliability of a novel approach for the recognition of bird species based on their sounds.

Keywords- Birds Sound Classification, Acoustic-LTP; Acoustic-LPC; MFCC, Feature Fusion

I. INTRODUCTION

Birds are natural predictors of animal diversity and ecosystem preservation in the world. Birds are not only necessary for the functioning of ecosystems but also provide several benefits to human societies, making their conservation and protection crucial for the well-being of both natural environments and human communities. Bird sound detection and classification is a process that involves the identification and marking of the sound produced by distinct bird species [1]. The diverse variety of existing bird sounds and many other factors affecting the way of a particular sound production make this process complex and challenging. [2-3] There are many applications including wildlife conservation, bioacoustics research, and automated system development to classify bird sounds for tracking the bird population. [4] For instance, classification through bird sound detection can help monitor the endangered species, the migration pattern, and the fluctuation in the bird population over time [5].

This enables an evident focus on automating species detection in soundscape recording [6]. This study is an initiative of a set of data processing steps that integrates a convolutional neural network trained on Mel-spectrogram to depict the set of species present in the recording. The incorporation of transfer learning and unique loss functions primarily contributes to efficient training and highlights the challenges related to the limited labeled datasets [3]. In birds sound classification researcher presents an alternative solution using the spectrograms for transfer learning, the fine-tuning a pre-trained network for the visual representation of sound [7]. The process starts with the audio recording from the environment a capturing variety of birds' songs and calls. The audio data is transformed into visual representations such as spectrogram. The method applied to these spectrograms are machine learning models for the classification of different bird's species based on their acoustic patterns. In Fig. 1 there are some species of birds from Birds Sound dataset.

Fig. 1. Bird Images

The [8] study emphasizes the significance of ensemble classifiers in addressing challenges posed by limited environmental audio datasets. Through the strategic utilization of data augmentation techniques and multiple signal representations, this research retrains Convolutional Neural Networks (CNNs) on three standard datasets named Bird Calls, Environmental Sound Classification database and Cat Sounds dataset. The ensembles exhibit accuracy, reaching impressive up to 97% on bird datasets, 90.51% on cat datasets, and 88.65% on the challenging ESC-50 dataset [9]. In the exploration of Indonesia's avian soundscape, the [10-11] delve into the classification of Indonesian scop owls' vocal sounds. Both studies employ a four-layer CNN and compare the model's performance using different acoustic signal representations. Remarkably, the dual-input network emerges as the top-performing model, achieving a Mean Average Precision (MAP) of 97.55% in both experiments.

Researcher presented their research by focusing on the environmental sound identification smart cities, [4] introduces MosAIc a machine learning classifier and a lighter CNN model. These models compete in accuracy with deep learning solutions emphasizing the fact that classical machine learning classifiers can carry competitive outcomes with decreased computational cost. This research underscores the importance of Contemplating the resource limitation when disposing the sound classifiers in the smart city. In [12] researchers present the bird sound classification technique based on constant frame sequences and spectrogram-frame linear network (SFLN). This attains high mean average precision (MAP) values up to 0.97, underlining the effectiveness of continuous frame sequences in picking the intricate frequency distribution and time-changing characteristics of bird sounds.

The bird classification field has brought about an innovative acoustic approach [13]. This also addresses numerous challenges, including confined labeled data, computational complexity, and the necessary robust effective classification models as this field continues to enhance, these contributions offer valuable intuitions and methodologies or techniques for researchers directing the convergence of ornithology machine learning and environmental studies.

Bird sound detection classically depends upon various methods such as MFCC and linear Predictive Coding. MFCC comprises low-dimension features with efficient accuracy that make it favorable for sound detection. However. The other environmental noises interrupt the sound access during the collection of data. Affecting the efficiency of MFCC features. In addition to this. Various initiating conditions suppress the quality of MFCC features resulting in misidentification of sound recognitions. For the correction of these errors, this research brings about the latest feature-fusion through the acoustics-LTP, MFCC, and LPC using 20,13, and 10 defined characteristics from the signal received from each respective technique. This extensive feature vector provides robust characteristics of vast bird sounds with high accuracy according to the respective techniques.

II. Literature Review

The research [1] contributed to bird sound classification using acoustic signals and delved into the application of multileveled ternary pattern (TP) feature generation. The authors introduced Iterative Relief F (IRF) for environmental sound recognition, which is an improved version of Relief F. TP feature generation is used in the presented automated bird sound classification model whereas IRF works as a feature selector which selects the most reliable feature automatically operated on linear discriminant (LP).

This study contributed to bird sound classification using CNN and achieved an accuracy of 96.45%. Emphasizing the necessity of adaptable models across diverse datasets, the study introduces an ensemble of classifiers utilizing various data augmentation techniques. Five pre-trained CNNs are retrained and tested on three benchmark datasets [14].

Several related studies [15] contribute to the understanding of acoustic signal processing and convolutional neural networks focus on identifying bird and frog species in tropical soundscapes, introducing a custom training loss and false-positive detections for multi-label learning.

To explore fault diagnosis in industrial machines using acoustic signals, the researchers propose drill fault diagnosis based on sound signal scalograms and Mel spectrograms. Various visual representations and texture extraction techniques are also investigated for audio classification, enriching the literature [16]. The research [17] commonly performs machine learning techniques like traditional machine learning and deep learning techniques followed by signal modification techniques such as time-frequency for bird sound categorization. The respective techniques' efficiency varies depending upon the special data set used.

By using deep learning models, some studies have reported a high accuracy rate for sound classification whereas others have found that traditional machine learning can also be effective [18]. The notable advancements and emerging trends in bird sound classification using acoustic signals include the use of deep learning techniques such as CNNs and recurrent neural networks (RNNs) for feature selection and classification. Besides this, the use of unsupervised learning for clustering bird sounds, based on their acoustic features is also recognized. There is also a trend towards developing portable field recorders for the remote acoustic monitoring of birds and other animals [19-20]. The main findings underscore the promising application of convolutional neural networks for species identification in soundscape recordings. Notably, the custom training loss function contributes to stable learning and improved performance, with the false-positive detection algorithm reducing manual efforts in collecting multi-label training data. The outcomes demonstrate the potential of CNNs for multi-label audio classification, providing a valuable pipeline [3] for researchers interested in applying this approach to different datasets.

Mel-frequency cepstral Coefficients (MFCCs) are generally used as audio features. These coefficients constitute discrete values that add up to a unit vector and are used to catch the vocal tract form in time-continuum. Other discrete features that may be gained to analyze the bird sounds consist of prosody, note-like and sound elements, word series, and design expression patterns. Through the analysis of respective features, the researchers can get plenty of useful information regarding bird sounds such as their special sound types (e.g. High, low, sad, cheerful, gentle, quiet). These acoustics features can be fruitful for the categorization and detection of bird species contributing to environmental health and biodiversity preservation [12, 21-22].

The major challenges in bird sound classification include variation frequency in bird sounds, the lack of annotated datasets, background noise, and interruptions, confined computational resources [16]. It is noted that these challenges result in difficulty in differentiating among distinct species vocals, restricting the ability to train and evaluate machine learning models, and to withdraw suitable characteristics of bird sounds. To figure out these challenges, the researchers have designed more innovative feature extraction techniques, new datasets, and innovative tools for noise reduction and source separation. They have also found more effective algorithms and hardware architectures. The authors proposed an upgraded spectrogram representation that applies a logarithmic frequency scale and modified time-frequency resolution. They also use a Markov renewal process model to enhance the capture of the temporal structure of bird sounds [23]. A perspective of bird sound classification implies deep learning techniques such as neural networks. A recent study by [24-25] proposed a novel bird sound classification framework based on a CNN integrated with the Grad-CAM algorithm. They proposed a feature augmentation technique named the Gaussian Mixture Model (GMM) via Principal Component Analysis (PCA). Researchers used the feature fusion network MFF-ScSEnet and achieved high classification performance using the dataset Birdsdata. Their results showed an accuracy of 96.66%, which is a significant improvement over other methods [26]. Another recent study by [27] proposed a method using wavelet packet decomposition feature extraction and a deep neural network classifier to classify bird sounds. Their proposed method was tested on a dataset of 10 bird species and achieved an accuracy of 90.94%, which outperformed other classifiers such as k-NN, SVM, and RF [28].

III. Proposed Method

The proposed bird sound classification framework consists of various stages that are used to classify bird sounds using acoustic signals. Fig. 2 shows the different stages involved in the architecture of the proposed framework. It includes data acquisition, here we used the Birds Sound Dataset [15]. Sounds for this dataset have been collected by the researcher from sources such as the Xeno-Canto website and YouTube, maintaining a sampling frequency of 48 kHz. In the file format conversion stage, the bird sound signals are converted into wave format that is suitable for further analysis. Blank areas of sound signals are also handled using the windows overlap technique to get more active features for sound signals. As the dataset also includes sound from internet sources, it has been made sure that all sound signals are at the same sample rate. Using jAudio software Mel Frequency Cepstrum Coefficients (MFCC) number of coefficients are 13 as features dimensions. And linear predicted coding (LPC) technique is used for acoustic signal processing representing the spectral envelope of the digital signal using the autocorrelation method and extracted LPC 10 features. The sound analysis output is set as ARFF. Windows size of samples 512 and window overlap fraction is set to 0.5. For acoustic local patterns, Ternary patterns are used to extract 20 features for each signal using MATLAB. The fusion of these 43 features is used to create a comprehensive feature vector.

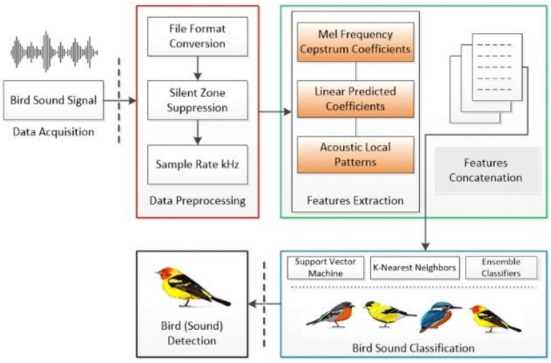

Fig. 2. Architecture of proposed Bird Sound Classification

This vector is then passed as an input to the sound classification module where the different state-of-the-art and already established classifiers are used for further classification of sounds of 18 classes of different birds.

Finally, the proposed system identifies the presence of a particular bird sound in the acoustic signal.
This robust fusion of features provides a comprehensive feature vector that supports machine learning algorithms to achieve acceptable accuracies. The classifiers used include ensemble methods of Bagged and RUSBoosted. Other classifiers are K-Nearest Neighbors (kNN) and Support Vector Machine (SVM). The proposed framework is designed to effectively classify bird sounds using acoustic signals which can be applicable in wildlife monitoring and conservation. The framework is a systematic approach that comprises various steps, such as file format conservation, data access, silent zone suppression, MFCC extraction, LPC modeling, acoustics local patterns extraction and fusion, and finally the bird sound categorization and recognition through machine learning algorithms.

A. Acoustic Local Binary Patterns (Acoustic-LBP)
Acoustic-LBP [30] is a fast and computationally efficient method for encoding signals that effectively emphasizes specific signal properties. These features, known as linear LBP codes, can be used to segment signals and generate signal thumbprints. LBP assigns a unique code to each center sample by analyzing the nearby data samples of a signal and applying thresholding.
Let $V\_s\hat{}((j))$ [n] be the value of the central sample in the samples window with P + 1 elements in audio signal Y for $j = [P : N - P]$.

$j = [\frac{P}{2}: N_s - \frac{P}{2}]$. The acoustic-LBP is stated as:

$$LBP_p\ (V_s[j]) = \sum_{k-0}^{\frac{p}{2}-1} \begin{Bmatrix} S\left[V_s\left[j+k-\frac{p}{2}\right]-V_s[j]\right]2^k + \\ \\ \ldots S\left[V_s\ [j+k+1]-V_s\ [j]\right]2^{k+\frac{P}{2}} \end{Bmatrix} \quad (1)$$

Where sign function S[.] is given by:

$$S[Vs] = \begin{Bmatrix} 1, & for\ V_s\ \geq 0 \\ 0, & for\ V_s\ \geq 0 \end{Bmatrix}$$

In LBP, illustration Vs [j] will be the threshold value for the vicinal samples. The function S[.] gives the difference of Vs [j] and its neighboring samples as a binary code representing as P-bit code. Now the LBP code is multiplied by the binomial weights and finally summation is performed to get the LBP value for the sample Vs [j]. LBP codes are employed to represent local patterns, characterized by;

$$H_k = \sum_{\frac{p}{2} < j \leq N - \frac{p}{2}} \delta\left(LBP_p(V_s[j]), k\right) \quad (2)$$

Where k=1…n, n illustrates histogram bins related to each LBP code, and $\delta(i,j)$ is the Kronecker delta function.
Because Acoustic-LBP features employ a threshold precisely at the main sample proving to be highly susceptible to noise. This sensitivity becomes especially pronounced at edges where differences in certain directions exceed those in others, as noted in references [28]. Even minimal exposure to noise can render the results and make acoustic-LBP descriptor unreliable.

B. Acoustic-Local Ternary Patterns (Acoustic-LTP)
In ternary patterns as sound signal descriptor the Vs [j] is generated, and a code of three values for -1,1 and 0 used in acoustic-LTP is defined in [29]. Here the magnitude difference of signals is calculated between Vs [j] and its eight neighbors UP. Signal values in the range of width $\pm th$ around Vs [j] are quantized to zero. The three valued numbers are attained 1 for the values above Vs [j] + th *and -1 is* quantized for the value below Vs [j] + th while 0 is the value between the above and below the line. The function calculating these three values is given by:

$$S'(U, V[j], t_h) = \begin{cases} +1, & [U_p - (V_s[j]+t_h)] \geq 0 \\ 0, & [V_s[j] - t_h] < U_p < [V_s[j] + t_h \\ -1, & [U_p[j] - t_h] \leq 0 \end{cases} \quad (3)$$

$S'(U_p, V_s\ [j], t_h)$ in equation (4) represents the audio signal by three-valued ternary patterns. This acoustic signal is then divided into S$'upper$ and $S'lower$ values. In S$'upper$ +1 is retained as 1 and all other -1 and 0 will be considered 0s in equation (4).

$$S'_{upper}\left(U_p, V[j], t_h\right) = \begin{cases} \frac{1, for[S'(U_p, V_s[j], t_h = +1]}{0,\quad Otherwise} \end{cases} \quad (4)$$

Similarly, in $S'_{lower}\left(U_p, V[j], t_h\right)$ -1 is retained as 1 while +1 and 0s will be considered as 0s in equation (5):

$$S'_{lower}\left(U_p, V[j], t_h\right) = \begin{cases} \frac{1, for[S'(U_p, V_s[j], t_h = -1]}{0,\quad Otherwise} \end{cases} \quad (5)$$

Now the acoustic-LTP is represented in equation (6):

$$Acoustic - LTP_p\ (V_s[j]) = \begin{cases} \sum_{i=-1}^{i=1} \sum_{k=-1}^{k=1} S'_{upper}\ [V_s\ [i,k] - V_s\ [j]]2^l \\ \\ \ldots \sum_{i=-1}^{i=1} \sum_{k=-1}^{k=1} S'_{lower}\ [V_s\ [i,k] - V_s\ [j]]2^l \end{cases} \quad (6)$$

Fig. 3. Acoustic-LTP for Bird Sound Classification [28]

In figure, the blue spikes represent the acoustic signal, with the variations in the line heights and patterns representing the magnitude differences between the signal and its surrounding neighbors. The three-valued code, or ternary pattern, of the Acoustic-LTP, could be represented by the positive (blue) and negative (white) regions, with zero represented by the absence of a line.

IV. RESULTS DISCUSSION

The experiments for the evaluation of the system are performed on a publicly available research-based Bird Sound dataset [30]. As shown in Table I this dataset contains recordings of 18 different bird species, each with an average sound length of 5 seconds. The research represents the dataset of a total of 781 sound samples, ranging from 25 to 51 recordings per species. Data has been collected by the researcher from sources such as the Xeno-Canto website and YouTube, maintaining a sampling frequency of 48 kHz. The curation process ensured that only bird sounds were retained, alongside various environmental noises, with a focus on prominent bird species' sounds.

Table II shows the classification results and comparison with base paper using different classifiers. Here our results using Bagged Trees and RUSBoosted Trees ensemble methods are 97.40% and 96.70% accuracy respectively which are higher than the reported research. [1] Authors used the combination of multileveled and handcrafted features as the input for machine learning classifiers. The model accuracy is achieved by these multileveled ternary patterns for feature extraction and then selecting the best features for classification. We also mentioned the other accuracies achieved by our proposed method 92.00% and 83.60% using kNN and SVM classifiers respectively. To achieve a better accuracy rate from our features fusion technique a set of base classifiers are used in Ensemble methods. The Bagged Trees Ensemble method is an ensemble learning technique mainly used to improve the stability of machine learning models in terms of accuracy, particularly decision trees. The bagging comprises four major steps bootstrap sampling, base learner training, ensemble aggregation, and the final prediction. The randomness in the training phase supports the reduction of variance and overfitting.

Each base learner (decision tree) is trained on a slightly different subset of the data, which leads to diverse models. By combining these diverse models, bagging reduces the risk of overfitting and improves the overall performance of the ensemble model. The Bagged Trees Ensemble method is a powerful technique for improving the accuracy and stability of machine learning models and in this research, this fulfills and results with a higher accuracy of 97.40%. Other classifiers including SVM and kNN are also used in this experiment analysis. Table III shows the summary of various performance metrics, for the best accuracy achieved by the Bagged Trees classification task, including True Negatives, True Positives, False Positives, False Negatives, Recall, Precision, F-1 Score, and False Discovery Rates (FDR).

TABLE I. Birds Sound Dataset (Adapted from [30])

No.	Sound class	Sound IDs	No. of files
1	Brown Cuckoo-Dove	BCD	25
2	Brown Honeyeate	BH	25
3	Bush Stone-curlew	BSC	26
4	Eastern Whipbird	EW	25
5	Eastern Yellow Robin	EYR	60
6	Grey Fantail	GF	40
7	Rainbow Lorikeet	RL	51
8	Rufous Whistler	RW	50
9	Shining Bronze Cuckoo	SBC	50
10	Silvereye	SI	50
11	Striated Pardalote	SP	50
12	Sulphur-crested Cockatoo	SCC	50
13	Torresian Crow	TC	49
14	White-throated Honeyeater	WTH	50
15	Budgerigar	BD	50
16	The Atlantic Canary	TAC	30
17	Goldfinch Carduel	GC	50
18	Zebra Finch	ZF	50

1). Evaluation Criteria

To evaluate performance Accuracy, Precision, Recall rate, F-1 Score, and False Discovery Rates have been measured using the following formulas in equations (7-11).

In equation (7) here the measurement of the Precision is made by dividing the number of true positive predictions by the sum of true and false positive predictions. Precision quantifies the ability of our model to avoid false positives showing its reliability when it predicts positive instances. A high Precision indicates that the model has a low rate of falsely predicting negative instances as positive. In equation (8) the Recall Rate is calculated, which evaluates the proportion of true positive predictions among all actual positive instances. Here the values in Table III show that the Recall quantifies the ability of the model to capture positive instances.

For accuracy, the calculations are made as mentioned in equation (9) measuring the overall correctness of the model's predictions across all available classes of the dataset. Accuracy is calculated by dividing the sum of true positive predictions and true negative predictions by the total number of instances in the dataset.

For F-1Score equation (10) is calculated by combining the Precision and Recall into a single value. It is calculated by taking the harmonic mean of Precision and Recall. This ensures that the F1 Score penalizes extreme values of Precision or Recall when a class imbalance exists.

Finally, in equation (11) we calculated the False Discovery Rate as it is a statistical measure used to control the proportion of false positives among all positive predictions made by a model. It is the ratio of false positive predictions to the total number of positive. False Discovery Rate represents the proportion of positive predictions that are false positives. Its lower values indicate that the model has fewer false positive predictions relative to the total number of positive predictions it makes.

$$Precision\ Rate = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (7)$$

$$Recall\ Rate = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (8)$$

$$Accuracy\ Rate = \frac{True\ Positive + True\ Negative}{Total\ Positive + Total\ Negative} \quad (9)$$

$$F\text{-}1Score = 2 * \frac{(Precision * Recall)}{Precision + Recall} \quad (10)$$

$$False\ Discovery\ Rate = \frac{False\ Positive}{True\ Positive + False\ Positive} \quad (11)$$

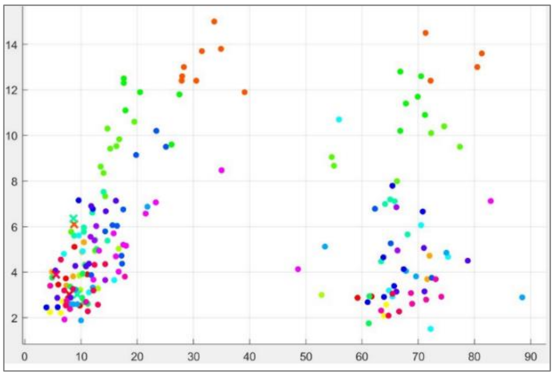

Fig. 4. Scatter plot of the Fusion Features over the Birds Sound dataset

Table II Accuracy Comparison with Different Classifiers

Features Fusion	Accuracy (%) Achieved			
	Bagged Trees	**RUSBoosted Trees**	**kNN**	**SVM**
Acoustic LTP20 + MFCC10 + LPC10	97.40	96.70	92.00	83.60
[1]	93.85	--	95.65	96.67

Table III Performance Metrics Summary for Classification Result

Class ID	TN	TP	FP	FN	Precision %	Recall %	F-1 Score%	FDR
BCD	190	7.00	0.00	0.00	100.00	100.00	100.00	0.00
BH	185	12.00	2.00	0.00	85.71	100.00	92.31	14.29
BSC	188	7.00	1.00	1.00	87.50	87.50	87.50	12.50
EW	188	5.00	1.00	3.00	83.33	62.50	71.43	16.67
EYR	192	5.00	0.00	0.00	100.00	100.00	100.00	0.00
GF	181	16.00	0.00	0.00	100.00	100.00	100.00	0.00
RL	185	12.00	0.00	0.00	100.00	100.00	100.00	0.00
RW	187	10.00	0.00	0.00	100.00	100.00	100.00	0.00
SBC	183	12.00	0.00	0.00	100.00	100.00	100.00	0.00
SI	185	11.00	0.00	1.00	100.00	91.67	95.65	0.00
SP	184	13.00	0.00	0.00	100.00	100.00	100.00	0.00
SCC	184	13.00	0.00	0.00	100.00	100.00	100.00	0.00
TC	185	11.00	0.00	1.00	100.00	91.67	95.65	0.00
WTH	184	13.00	0.00	0.00	100.00	100.00	100.00	0.00
BD	190	6.00	0.00	1.00	100.00	85.71	92.31	0.00
TAC	184	13.00	0.00	0.00	100.00	100.00	100.00	0.00
GC	185	12.00	0.00	0.00	100.00	100.00	100.00	0.00
ZF	197	12.00	1.00	0.00	92.31	100.00	96.00	7.69

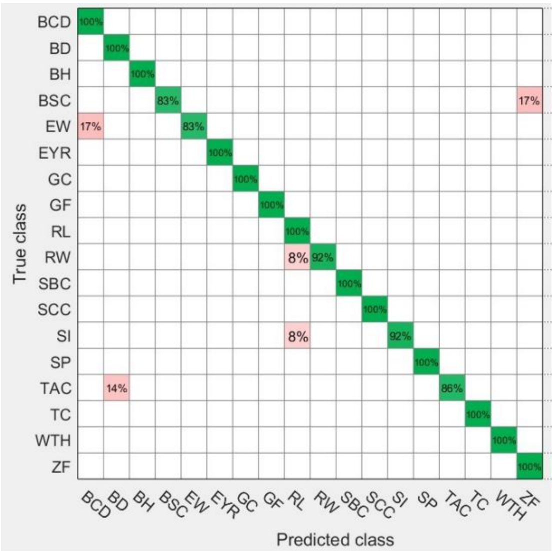

Fig. 5. Multiclass evaluation of Feature Fusion over Birds Sound dataset

V. CONCLUSION

Bird Sound Classification using Acoustic Signals presents a novel and effective approach to identifying bird species based on their vocalizations. By the fusion of features from Acoustic LTP, MFCC, and LPC coefficients representation, the proposed method demonstrates a high accuracy of 97.4% in classifying bird sounds. Utilizing a publicly available dataset of bird sounds, comprising recordings of 18 different bird species, the research showcases the practical application of the proposed model of feature fusion in bioacoustics research.

The findings underscore the importance of advanced technologies in automating species identification, wildlife conservation, and environmental monitoring. Moving forward, further research and development in this area can lead to more efficient and reliable systems for bird sound classification, benefiting ornithologists, conservationists, and ecosystem management efforts.

REFERENCES

[1] T. Tuncer, E. Akbal, and S. Dogan, "Multileveled ternary pattern and iterative ReliefF based bird sound classification," *Applied Acoustics,* vol. 176, p. 107866, 2021.

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

[3] J. LeBien *et al.*, "A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network," *Ecological Informatics,* vol. 59, p. 101113, 2020.

[4] M. Zhong *et al.*, "Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling," *Applied Acoustics,* vol. 166, p. 107375, 2020.

[5] X. Han and J. Peng, "Bird sound classification based on ECOC-SVM," *Applied Acoustics,* vol. 204, p. 109245, 2023.

[6] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal processing letters,* vol. 24, no. 3, pp. 279-283, 2017.

[7] A. Incze, H.-B. Jancsó, Z. Szilágyi, A. Farkas, and C. Sulyok, "Bird sound recognition using a convolutional neural network," in *2018 IEEE 16th international symposium on intelligent systems and informatics (SISY),* 2018: IEEE, pp. 000295-000300.

[8] L. Nanni, G. Maguolo, S. Brahnam, and M. Paci, "An ensemble of convolutional neural networks for audio classification," *Applied Sciences,* vol. 11, no. 13, p. 5796, 2021.

[9] L. Nanni, Y. M. Costa, R. L. Aguiar, R. B. Mangolin, S. Brahnam, and C. N. Silla, "Ensemble of convolutional neural networks to improve animal audio classification," *EURASIP Journal on Audio, Speech, and Music Processing,* vol. 2020, pp. 1-14, 2020.

[10] A. A. Hidayat, T. W. Cenggoro, and B. Pardamean, "Convolutional neural networks for scops owl sound classification," *Procedia Computer Science,* vol. 179, pp. 81-87, 2021.

[11] L. Lhoest *et al.*, "MosAIc: a classical machine learning multi-classifier based approach against deep learning classifiers for embedded sound classification," *Applied Sciences,* vol. 11, no. 18, p. 8394, 2021.

[12] X. Ji, K. Jiang, and J. Xie, "LBP-based bird sound classification using improved feature selection algorithm," *International Journal of Speech Technology,* vol. 24, pp. 1033-1045, 2021.

[13] A. E. Mehyadin, A. M. Abdulazeez, D. A. Hasan, and J. N. Saeed, "Birds sound classification based on machine learning algorithms," *Asian Journal of Research in Computer Science,* vol. 9, no. 4, pp. 1-11, 2021.

[14] S. D. H. Permana, G. Saputra, B. Arifitama, W. Caesarendra, and R. Rahim, "Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm," *Journal of King Saud University-Computer and Information Sciences,* vol. 34, no. 7, pp. 4345-4357, 2022.

[15] 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, pp. 175353-175361, 2019.

[16] J. Xie and M. Zhu, "Handcrafted features and late fusion with deep learning for bird sound classification," *Ecological Informatics,* vol. 52, pp. 74-81, 2019.

[17] X. Zhang, A. Chen, G. Zhou, Z. Zhang, X. Huang, and X. Qiang, "Spectrogram-frame linear network and continuous frame sequence for bird sound classification," *Ecological Informatics,* vol. 54, p. 101009, 2019.

[18] M. Li and Y. Li, "Ecological environmental sounds classification based on genetic algorithm and matching pursuit sparse decomposition," in *2012 5th International Congress on Image and Signal Processing*, 2012: IEEE, pp. 1439-1443.

[19] G. You and Y. Li, "Environmental sounds recognition using tespar," in *2012 5th International Congress on Image and Signal Processing*, 2012: IEEE, pp. 1796-1800.

[20] A. E. Mehyadin, A. M. Abdulazeez, D. A. Hasan, and J. N. Saeed, "Birds sound classification based on machine learning algorithms," *Asian J. Res. Comput. Sci,* vol. 9, p. 68530, 2021.

[21] R. Pahuja and A. Kumar, "Sound-spectrogram based automatic bird species recognition using MLP classifier," *Applied Acoustics,* vol. 180, p. 108077, 2021.

[22] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck, "BirdNET: A deep learning solution for avian

diversity monitoring," *Ecological Informatics,* vol. 61, p. 101236, 2021.

[23] J. Xie and M. Zhu, "Acoustic Classification of Bird Species Using an Early Fusion of Deep Features," *Birds,* vol. 4, no. 1, pp. 138-147, 2023.

[24] S. Hu, Y. Chu, Z. Wen, G. Zhou, Y. Sun, and A. Chen, "Deep learning bird song recognition based on MFF-ScSEnet," *Ecological Indicators,* vol. 154, p. 110844, 2023.

[25] A. Selin, J. Turunen, and J. T. Tanttu, "Wavelets in recognition of bird sounds," *EURASIP Journal on Advances in Signal Processing,* vol. 2007, pp. 1-9, 2006.

[26] Y. Mustaqim, E. Utami, and S. Raharjo, "Analysis of Daubechies Wavelet and Neural Network for Audio Classification," in *2019 International Conference on Information and Communications Technology (ICOIACT)*, 2019: IEEE, pp. 322-326.

[27] I. Potamitis, S. Ntalampiras, O. Jahn, and K. Riede, "Automatic bird sound detection in long real-field recordings: Applications and tools," *Applied Acoustics,* vol. 80, pp. 1-9, 2014.

[28] F. Behloul, F. Tafinine, and O. Yaman, "Induction Motor Fault Diagnosis with Local Ternary Pattern and AI Approaches," *Journal of Failure Analysis and Prevention,* vol. 23, no. 6, pp. 2533-2541, 2023.

[29] A. Irtaza, S. M. Adnan, S. Aziz, A. Javed, M. O. Ullah, and M. T. Mahmood, "A framework for fall detection of elderly people by analyzing environmental sounds through acoustic local ternary patterns," in *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, 2017: IEEE, pp. 1558-1563.

[30] *Bird Sound Dataset*. [Online]. Available: http://web.firat.edu.tr/turkertuncer/BIRDS.rar