Journal of Scientific & Industrial Research
Vol. 84, March 2025, pp. 301-307
DOI: 10.56042/jsir.v84i03.12920

ISAVM: Improved Smart Avian Monitoring using FLANN-based Audio Activity Detection & Speech Enhancement

Jagannath Dayal Pradhan, Tusar Kanti Dash* & Ganapati Panda

Electronics and Communications Engineering, C. V. Raman Global University, Bhubaneswar 752 054, Odisha, India

Received 01 August 2024; revised 09 September 2024; accepted 17 September 2024

The Avian monitoring system is one of the challenging tasks that helps identify the environmental changes in the forest as well as the overall counts of specific species. Out of several methods available for avian monitoring, audio-based avian monitoring is one of the most efficient and cost-effective tools. By studying bird sounds, a smart society can be built for an enhanced avian surveillance system through the use of speech recognition algorithms. Conventionally, speech characteristics are employed for these tasks, which may not be appropriate given that these acoustic noises deviate from human speech and deteriorate the identification systems' performance. An application-specific audio activity identification technique is needed since the features of human voice and bird sound differ. As of now, few works have been reported mainly for the bird sound analysis with audio activity detection and speech enhancement schemes. This work has considered and implemented this problem in three steps. In the first stage, an improved voice activity algorithm is designed using a Functional Link Artificial Neural Network model. In the second stage, an effective AdaBoost classifier is used for training and testing. Finally, the developed model, ISAVM: Improved Smart Avian Monitoring System has been checked for improved performance in two standard bird datasets. The evaluation has been done for different preprocessing options with and without audio activity detection and speech enhancement schemes. It has been observed that the proposed model is performing consistently with more than 93% of classification accuracy which is better than the standard avian monitoring models.

Keywords: Acoustic signal analysis, Artificial intelligence, Bird recognition, Machine learning, Speech recognition

Introduction

Monitoring systems in avian are effective instruments in ornithology and conservation biology. They assist the researchers in counting birds, monitoring bird movements, habit utilization, and the effects of applying change of environment among others. Every field study of the avian population starts with field surveys; these are some of the earliest and most basic techniques.[1] In the transect surveys: observers follow a fixed line (transect) and note birds sighted within a given distance. This technique assists in the quantification and location of species populations.[2] In Nest Monitoring, the process of going through nests repeatedly in order to collect data on the breeding rate, number of eggs per nest, and growth rate of chicks.[3] Fortnight, monthly, or yearly sampling can be done in the field and it is used to record bird sounds for a long time using the automated acoustic recorders. Most of these devices are helpful in surveying bird species' abundance and their activities in a particular area without the presence of humans. These sounds can also be recorded and with the help of the latest technologies, experts would be able to accurately determine which species are around and their numbers.[4] Techniques including Global Positioning System (GPS) and radio telemetry help in understanding bird movement patterns including migration corridors, roosting, and foraging grounds as well as the habitats of the birds.[5] Birds as Biodiversity and Ecosystem Healers Biodiversity is the variety of life and its processes that are found in any ecosystem and birds act as indicators of the status of existing ecosystem health. Fluctuations in bird populations are believed to be a sign of physical changes in the environment which include; destruction of natural habitats, climate change, and pollution. Climate Change Studies Avian monitoring is a way of describing how the effects of climate change influence characteristics of the wildlife. Birds are acknowledged to be sensitive to temperature change and rainfall which makes them good bio-indicators.[6,7]

Avian monitoring using the Sound Recognition Technique is one of the effective methods used recently.[8] Audio-based monitoring has been boosted

*Author for Correspondence
E-mail: tusarkantidash@gmail.com

by improvements in technology through portable and cheap equipment like recorders and microphones.[9] To the list of difficulties of audio monitoring, the problem of noise is one of the main obstacles that prevents birds' calls from being heard. Sounds of wind, water, insects, and other people's interference can interfere with bird vocalizations hence decreasing an accurate estimate of bird species and density. This is due to the nature of the system that involves a large number of recordings and therefore a greatly increased number of data to be stored and analyzed. This remains a challenge because there is a need to design algorithms that can accurately differentiate between similar calls and filter the overlapping sounds.[10,11] Speech recognition is the application of techniques to analyze and understand spoken data to generate useful information like classifying bird species from their call. This technology has been integrated into avian monitoring to help in the classification of birds thus improving efficiency and effectiveness in the passive acquisition of data.[12] Feature extraction is an important step in the process of speech recognition; this is the process of extracting specific features from the acoustic signals that can be used as identification tools. In bird species monitoring, commonly used features include Mel-Frequency Cepstral Coefficients (MFCCs).[13,14]

Voice Activity Detection (VAD) is a very important factor in many of the applications involving speech signal processing because with it one can differentiate between portions of an audio stream containing speech and periods of no speech. On the basis of literature, many aspects that reveal the presence of speech have been provided.[15] The VAD is based on two calculations: the long-term spectral pattern and the degree of spectral deviation of speech with respect to noise. However, only in the case of low observed SNR a controlled hang-over is started, and the decision threshold is adjusted to the noise energy level. An interesting work shows an early perfect construction of a simple VAD algorithm that is resistant to noise, while some works do not have this advantageous impulse.[16] The proposed method takes advantage of short-term properties such as Short-term Energy and Spectral Flatness (SF). As a result, the method is more suited for online processing scenarios. The proposed method is evaluated on multiple speech datasets with additive noise and compared with some of the most recently proposed methods. The testing shows respectable results under a range of noise conditions.

The Functional Link Artificial Neural Network (FLANN), a neural network model without a hidden layer that functions at the single layer level, has a low computational complexity. When there are limited features and little data, it provides good prediction performance. It has been applied as a multi-objective strategy to the task of enhancing speech quality and intelligibility. In a different study, a bounded-input bounded-output stability criterion for the recursive FLANN filter was developed using trigonometric expansions. The key conclusion from the stability criterion is that instabilities have no influence on the recursive FLANN filter if the recursive linear element of the filter is stable.[17,18] Starting from recent years, there has been significant research work reported in the literature discussing various improvements and extensions of FLANN and some of the areas where FLANN has been applied are as follows: There have been publications presenting adapted FLANN systems integrated with other machine learning technologies with the aim at enhancing accuracy and resilience. For instance, there has been incorporation of FLANN with functional expansion with evolutionary algorithms hence improving the difficulty degrees it performs.[19] Although FLANN is usually considered a shallow network, some recent investigations have attempted to combine this system with deep learning environments.[20] Due to current developments in big data, FLANN has been extended to cater to a large number of inputs. There are several approaches that have been developed and proposed in order to enhance the scalability and performance of FLANN in distributed computing situations, and in turn, increase its relevance to BIG DATA processing.[21] FLANN has also been used in the biomedical area, for example in the diagnosis of diseases and in medical image processing. This makes it useful in those applications that involve modeling of complex non-linear relationships because of its capability to perform them in real-time.[22] FLANN has also been used for the basic speech enhancement task and for the prediction of the SNR levels.[23]

Bird sound recognition is the process of identifying and categorizing different bird species using their vocalizations. It is an essential component of avian monitoring and biodiversity assessment. The existence of background noise and other distortions that might deteriorate the quality of bird recordings is one of the main problems in this subject. In order to overcome these difficulties, speech enhancement techniques have been used to increase the comprehensibility and clarity of bird sounds prior to the application of recognition algorithms. These methods are customized to the special qualities of bird vocalizations, drawing inspiration from human speech processing.[24] It has been observed from the literature review that in the fields of ecology, conservation biology, and environmental research, avian monitoring the methodical observation and analysis of bird populations is essential. Scientists studying birds can learn a great deal about wider ecological trends and changes in the environment by gathering data on bird species, populations, and behaviors. The main justifications for the importance of bird monitoring are outlined in this overview of the literature, along with how it benefits ecological studies, environmental management, and biodiversity conservation. However, it has been observed that the major challenge in avian monitoring is audio activity detection and the effect of noise on the birds' sound. These two issues are being taken care of in this paper. To solve this problem, a low complexity-based FLANN model has been used for improved audio activity detection along with an adaptive spectral subtraction-based speech enhancement scheme.[25] The major research contributions are listed below-

- Development of improved audio activity detection specially designed for bird audio recognition with low complexity-based FLANN.
- Use of adaptive speech enhancement schemes for improved bird species recognition.
- Application of Spectral subtraction-based speech enhancement scheme for improved avian monitoring.
- Simulation-based experimentation in multiple datasets with more than 95% detection accuracy for the proposed model.

Materials & Methods

Datasets used

Bird Species Dataset (D-1)

The 14 categories in this dataset—Guinea Fowl, Barred Plymouth Rock, White Plymouth Rock, Red Codish, Black Rock, Naked Neck, Rhode Island Red, Kalinga Brown, Japanese Quail, White Leg Horn, AW Cross, Aseel Brown, Color Cross, and Vanaraja were gathered from the sounds made by birds at the Central Poultry Development Organization in Bhubaneswar, India.[26] Total of 641 samples are recorded in WAV format at a sampling frequency of 44.1 kHz and 16 bits per second. The time-frequency plot of the Aseel Brown category has been plotted in Fig. 1.

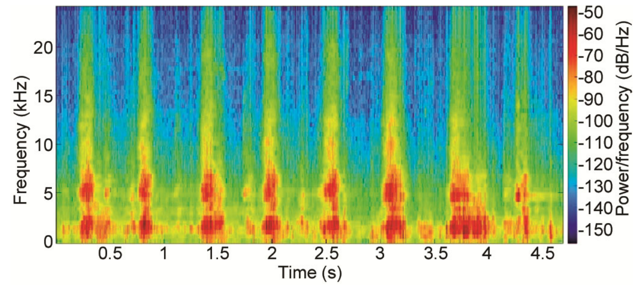

Fig. 1 — Time-frequency plot of Aseel Brown bird

Picidae Dataset (D-2)

An annotated dataset of bird calls from the Picidae family, separated from background noise and spanning 4984 seconds.[27] Sum of 1669 audio samples from 161 audio files collected from the Xeno-Canto Dataset are included in this dataset[28] in seven Picidae families that inhabit the Iberian Peninsula and emit up to three different sounds: call, drumming, or song. Though the number of families included is far fewer than in other general-purpose collections, the species in this dataset have been carefully selected for their importance to biodiversity quality and habitat protection. This information can be used as a classification problem to determine the song genre and identify the species of Picidae bird.

Improved Audio Activity Detection

In voice signal processing, framing and windowing are crucial preprocessing stages. In order to minimize spectral distortions and guarantee smoother frequency domain transitions, windowing applies a window function to each frame. Framing divides the continuous speech signal into manageable, quasi-stationary intervals.

By preparing the speech signal for additional processing, these actions improve the precision and potency of different speech-processing applications.[29] A key component of speech processing is Short-Time Energy (STE), which is used to examine the energy content of a speech signal over brief time intervals. It can be applied to a variety of applications, including feature extraction, segmentation, and speech identification, and it offers useful information about the signal's strength. This is a thorough explanation of short-time energy, covering its definition, importance, and uses. The energy of a signal within a brief window of time is referred to as STE. This feature aids in the identification and analysis of different components of the signal's structure by measuring the amount of energy present in each spoken signal frame. From the frame level, the STE has been extracted which is followed by

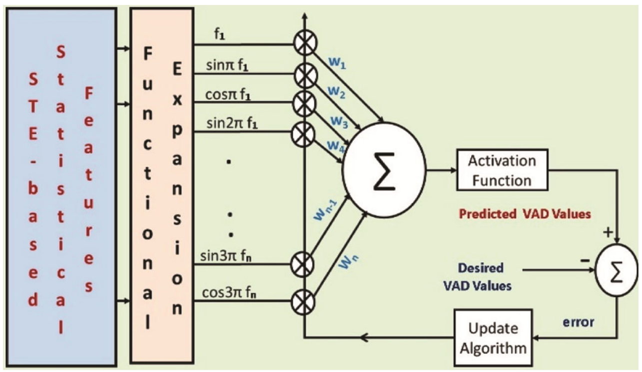

Fig. 2 — FLANN- based Audio Activity Detection

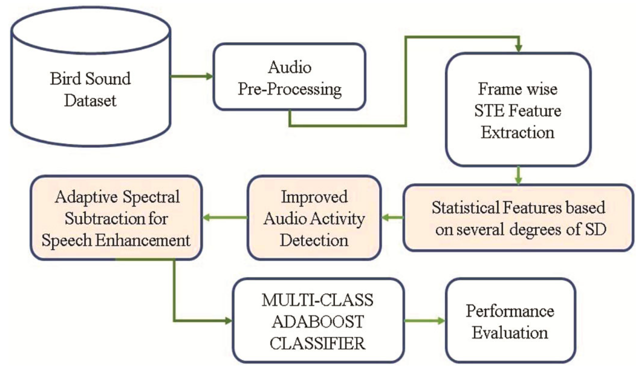

Fig. 3 — Block Diagram of the proposed ISAVM model

Table 1 — Symbols and their Interpretation

Symbols	Interpretation
$f_1, f_2, f_3, \ldots\ldots f_n$	Input features to the FLANN Model (Statistical Features using several degrees of SD)
$w_1, w_2, w_3, \ldots\ldots w_n$	Weights associated with each function expansion
VAD	Voice Activity Detection
Error	Predicted VAD values – Desired VAD values

the sample-level statistical feature extraction. For better extraction of relevant features, two Standard Deviations (SD) on either side of the mean, are used. However, for improved performance, the following additional degrees of deviation: 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 3.5, and 4 are used.[30]

FLANN-based Audio Activity Detection

The STE-based statistical feature vectors are given sample-wise to the functional expansion block of the FLANN. Each of the expanded feature sets is passed through tanh activation function and the predicted outputs are either 0 (unvoiced) or 1 (voiced). Here, the word voiced means with bird sound and unvoiced means without the bird sound. These values are being compared with the desired VAD values and the error is being updated using the least mean square algorithm. This process of weight updating continues for 10000 iterations and is stopped after the error has been reduced significantly lower value close to zero. The implementation of the improved VAD algorithm is shown in Fig. 2. The symbols used in the block diagram are listed in Table 1.

Adaptive Spectral Subtraction Algorithm

A well-known and frequently applied method in the field of speech enhancement, spectral subtraction is especially useful for noise reduction. By subtracting an estimate of the noise spectrum from the noisy speech spectrum, this method seeks to enhance the comprehensibility and quality of speech signals. During non-speech parts, the noise spectrum is estimated, and during speech segments, the noisy speech spectrum is subtracted from this estimate. The fundamental actions consist of Noise estimation: Calculate the noise spectrum, usually when the speech signal is pausing or silent. To obtain an estimate of the clean speech spectrum, subtract the noise estimate from the noisy speech spectrum. This process is known as spectrum subtraction. Reconstruction: Apply the inverse Fourier transform to reconstruct the time-domain signal.[31,32] The adaptability comes from the case where the audio activity is present or not. In the frames, where the audio activity is present, the noise level reduction is less as compared to the frames where no audio activity of bird sound has been detected.

Simulation Results and Discussion

The detailed implementation steps for the proposed model (ISAVM: Improved Smart Avian Monitoring using FLANN-based Audio Activity Detection & Speech Enhancement) are listed below and shown in Fig. 3.

- Improved audio activity detection is carried out once audio pre-processing techniques including windowing, framing, and silence removal are completed in the first step.
- The selected frames are those where the bird sound is present. From these frames, STE features are extracted along with several standard deviation-based statistical features. The FLANN model has been trained to predict the frames with Audio Activity and without Audio Activity.
- At the last step, multi-class AdaBoost classifier has been used. Boosting has proven to be an extremely effective method for solving the two- class classification

problem. When moving from two-class to multi-class classification, the majority of algorithms have only been able to split the multi-class classification problem down into several two-class problems. Rather than decomposing the problem into multiple two-class problems, the AdaBoost algorithm is a multi-class classifier that minimizes a unique exponential loss. It uses a progressive additive modeling methodology.[33] Due to its easy implementation and low misclassification error rate, it has been widely used for various multi-class problems including automatic speech recognition[34], emotion recognition[35], sign language recognition[36], EEG, vowel, silent speech signal discrimination.[37]

Performance Evaluation Measures

A classifier's efficacy is calculated using the training and testing procedure. The stratified five-fold cross-validation scheme is used to calculate the classifier's accuracy in this simulation, to test 20% of the data and use the remaining 80% for training. The data is tested five times, and the average validation accuracy is found. The performance of the proposed ISAVM model along with the baseline models are evaluated using standard evaluation parameters including classification accuracy (CA), F-1 score, Precision (Prc), and Recall (Rec).[38]

Comparative Analysis with baseline features and models

At the first step, a comparative analysis of simulation results has been carried out with baseline features. For each audio sample, the proposed audio features are extracted along with the basic cepstral features with melscale[39], ERB scale[40], and Bark scale.[41] After that a comparative analysis has been carried out with the proposed ISAVM Model with AdaBoost classifier along with baseline features. The results are listed in Table 2.

When applied to the same classifier for the two datasets, it is found that the proposed features perform better than the fundamental cepstral features. In terms of Classification Accuracy, the proposed ISAVM model outperforms the mel, bark, and ERB scales by an average of 4%. This displays the efficacy of the suggested features in comparison to others. In the second step, the performance of the proposed ISAVM model is evaluated by contrasting it with other popular machine learning-based recognition models. These techniques include the following Cepstral features: (MFCC and GTCC) features with Random Forest Classifier (RFC)[42], MFCC features with k-Nearest Neighbor Classifier (MKN)[43], MFCC with Naive Bayes Classifier (MNB)[44], MFCC with Support Vector Machines (SVM).[45] The training and test classification accuracies are listed in Table 3. The classification accuracy of these five models for the two datasets is shown in Fig. 4.

It can be observed that the proposed model consistently outperforms the baseline models. One of the reasons for the improved performance is the use of improved audio activity detection as well as an adaptive speech enhancement scheme. In the next section, the effects of these two additions on the proposed ISAVM model will be evaluated.

Effect of Improved Audio Activity Detection &Speech Enhancement

To analyze the effect of Voice/Audio Activity Detection (VAD) and Speech Enhancement (SE) on performance, the proposed model has been evaluated under three conditions. These conditions include the proposed ISAVM model with both VAD and SE, the ISAVM model with VAD, and the ISAVM model with SE. These three variations have been evaluated for the two standard data sets (D-1 and D-2). The simulation results are plotted in Fig. 4.

The baseline models are also simulated along with the ISAVM model. It has been observed that the VAD is having a higher impact on the working of the proposed model. Similarly, the effect of the SE varies from dataset to dataset. However, it has been observed for all five simulated models, that the VAD and SE have prominent effects to play on the overall performance. In the case of ISAVM, the performance

Table 2 — Comparison using basic cepstral feature sets

Dataset	Evaluation Measures	ISAVM	Mel	ERB	Bark
D-1	CA	0.96	0.91	0.88	0.86
	F-1	0.95	0.90	0.88	0.85
	Prc	0.95	0.91	0.88	0.84
	Rec	0.96	0.91	0.86	0.85
D-2	CA	0.96	0.89	0.84	0.85
	F-1	0.97	0.89	0.85	0.84
	Prc	0.95	0.88	0.84	0.85
	Rec	0.95	0.88	0.84	0.84

Table 3 — Comparative analysis with standard models for training and testing

Datasets	ISAVM	RFC	MKN	MNB	SVM
D1 (Train)	0.96	0.93	0.87	0.93	0.91
D1 (Test)	0.95	0.91	0.87	0.92	0.91
D2 (Train)	0.95	0.92	0.88	0.86	0.86
D2 (Test)	0.95	0.92	0.87	0.86	0.85

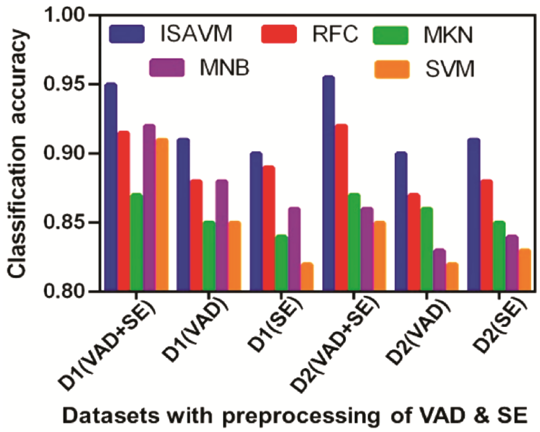

Fig. 4 — Block Performance Comparison of ISAVM with standard models using two datasets

has been improved with a minimum of 3–4% for both datasets. Similarly, for the other models, an average of 4% improvement has been noticed.

Conclusions

In this paper, an improved bird species monitoring system has been developed by using two modifications. Audio activity detection has been updated along with the speech enhancement scheme. From the simulation results, it has been observed that the proposed model is performing an average of 95% in two popular datasets of bird sounds. However, the results are to be tested and validated in multiple other bird species' data sets. Similarly, multiple other audio features as well as other speech enhancement schemes can be implemented to improve the detection accuracy further. In the future, the suggested ISAVM method may be used to monitor birds in other nations as well as other non-human sounds.

References

1 Bibby C J, Burgess N D, Hill D A & Mustoe S H, *Bird Census Techniques*, 2nd **edn** (Academic Press, London) 2000.

2 Buckland S T, Anderson D R, Burnham K P, Laake J L, Borchers D L & Thomas L, *Introduction to Distance Sampling: Estimating Abundance of Biological Populations* (Oxford Univ Press) 2001.

3 Martin T E & Geupel G R, Nest-monitoring plots: methods for locating nests and monitoring success, *J Field Ornithol*, **64(4)** (1993) 507–519.

4 Kirschel A N, Cody M L, Harlow Z T, Promponas V J, Vallejo E E & Taylor C E, Territorial dynamics of Mexicanant-thrushes formicarius moniliger revealed by individual recognition of their songs, *Ibis*, **153(2)** (2011) 255–268, https://doi.org/10.1111/j.1474-919X.2011.01102.x.

5 Bridge E S, Thorup K, Bowlin M S, Chilson P B, Diehl R H, Fléron R W, Hartl P, Kays R, Kelly J F & Robinson W D, Technology on the move: recent and forth coming innovations for tracking migratory birds, *Biosci Res*, **61(9)** (2011) 689–698, https://doi.org/10.1525/bio.2011.61.9.7.

6 Gregory R D & Vanstrien A, Wild bird indicators: Using composite population trends of birds asmeasures of environmental health, *Ornithol Sci*, **9(1)** (2010) 3–22, https://doi.org/10.2326/osj.9.3.

7 Robinson R A, Wilson J D & Crick H Q, The importance of arable habitat for farmland birds in grassland landscapes, *J Appl Ecol*, **38(5)** (2001) 1059–1069, https://doi.org/10.1046/j.1365-2664.2001.00654.x.

8 Marler P R & Slabbekoorn H, *Nature's Music: The Science of Birdsong* (Elsevier) 2004.

9 Mennill D J & Vehrencamp S L, Context-dependent functions of avian duets revealed by microphonearray recordings and multispeaker playback, *Curr Biol*, **18(17)** (2008) 1314–1319, https://doi.org/10.1016/j.cub.2008.07.073.

10 Blumstein D T, Mennill D J, Clemins P, Girod L, Yao K, Patricelli G, Deppe J L, Krakauer A H, Clark C & Cortopassi K A, Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus, *J Appl Ecol*, **48(3)** (2011) 758–767, https://doi.org/10.1111/j.1365-2664.2011.01993.x.

11 Kershenbaum A, Blumstein D T, Roch M A, Akçay C, Backus G, Bee M A, Bohn K, Cao Y, Carter G & Cäsar C, Acoustic sequences in non-human animals: A tutorial review and prospectus, *Biol Rev Camb Philos Soc*, **91(1)** (2016) 13–52, https://doi.org/10.1111/brv.12160.

12 Briggs F, Lakshminarayanan B, Neal L, Fern X Z, Raich R, Hadley S J, Hadley A S & Betts M G, Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach, *J Acoustical Soc Am*, **131(6)** (2012) 4640–4650, https://doi.org/10.1121/1.4707424.

13 Ramashini M, Abas P E, Mohanchandra K & Silva L C D, Robust cepstral feature for bird sound classification, *Int J Elect Compt Eng*, **12(2)** (2022) 1477–1487, https://doi.org/10.11591/ijece.v12i2.pp1477-1487.

14 Heller J R & Pinezich J D, Automatic recognition of harmonic bird sounds using a frequency track extraction algorithm, *J Acoustical Soc Am*, **124(3)** (2008) 1830–1837, https://doi.org/10.1121/1.2950085.

15 Graf S, Herbig T, Buck M & Schmidt G, Features for voice activity detection: A comparative analysis, *Eurasip J Adv Sig PR*, **2015(1)** (2015) 1–15, https://doi.org/10.1186/s13634-015-0277-z.

16 Moattar M H & Homayounpour M M, A simple but efficient real-time voice activity detection algorithm, in *17th European Signal Proces Conf* (IEEE) 2009, 2549–2553.

17 Dash T K, Solanki S S & Panda G, Multi-objective approach to speech enhancement using tunable Q-factor-based wavelet transform and ANN techniques, *Circuits, Syst Signal Process*, **40(12)** (2021) 6067–6097, https://doi.org/10.1007/s12065-020-00446-0.

18 Sicuranza G L & Carini A, On the BIBO stability condition of adaptive recursive FLANN filters with application to nonlinear active noise control, *IEEE/ACM Trans Audio Speech Lang Process*, **20(1)** (2011) 234–245, https://doi.org/10.1109/TASL.2011.2159788.

19 Buongiorno D, Brunetti A & Longo N, Deep learning for automatic vision-based recognition of industrial surface defects: asurvey, *IEEE Access*, **(11)** (2023) 43370–43423, https://doi.org/10.1109/access.2023.3271748.

20 Shi Q, Katuwal R, Suganthan P N & Tanveer M, Random vector functional link neural network-based ensemble deep learning, *Pattern Recognit*, **117** (2021) 107978, https://doi.org/10.48550/arXiv.1907.00350.
21 Chiroma H, Abdullahi U A, Alarood A A, Gabralla L A, Rana N, Shuib L, Hashem I A T, Gbenga D E, Abubakar A I & Zeki A M, Progress on artificial neural networks for big data analytics: a survey, *IEEE Access*, **7** (2018) 70535–70551, https://doi.org/10.1109/ ACCESS. 2018.2880694.
22 Kumar M & Mishra S K, Particle swarm optimization-based functional link artificial neural network for medical image denoising, in *Comput Vision Robot: Proc ICCVR* (Springer), 2015, 105–111, https://doi.org/10.1007/978-81-322-2196-8_13.
23 Dash T K & Solanki S S, Development and use of a new speech quality evaluation parameter ESNR using ANN and Grey Wolf Optimizer, *J Sci Ind Res*, **79(3)** (2020) 197–200, http://op.niscair.res.in/index.php/JSIR/article/view/68638.
24 Kogan J A & Margoliash D, Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study, *J Acoustical Soc Am*, **103(4)** (1998) 2185–2196, https://doi.org/10.1121/1.421364.
25 Dash T K, Chakraborty C, Mahapatra S & Panda G, Mitigating information interruptions by COVID- 19 face masks: A three-stage speech enhancement scheme, *IEEE Trans Comput Social Syst*, **11(4)** (2022) 1–10, https://doi.org/10.1109/TCSS.2022.3210988.
26 Mohanty R, Mallik B K & Solanki S S, Recognition of bird species based on spike model using bird dataset, *Data Brief*, **29** (2020) 105301, https://doi.org/10.1016/ j.apacoust. 2019.107177.
27 Vidaña-Vila E, Navarro J & Alsina-Pagès R M, Towards automatic bird detection: An annotated and segmented acoustic dataset of seven picidaespecies, *Data*, **2(2)** (2017) 18, https://doi.org/10.3390/data2020018.
28 Vellinga W P & Planqué R, The Xeno-canto collection and its relation to sound recognition and classification, *CLEF*, (2015).
29 Rabiner L & Schafer R, *Theory and Applications of Digital Speech Processing* (1st. ed.) (*Prentice Hall Press, USA*) 2010.
30 Abdulmohsin H A, Abdul H B & Abdul A M, A new proposed statistical feature extraction method in speech emotion recognition, *Comput Electr Eng*, **93** (2021) 107172, https://doi.org/10.1016/j.compeleceng.2021.107172.
31 Benesty J, *Fundamentals of Speech Enhancement* (Springer) 2018.
32 Dash T K & Solanki S S, Insight on the utilization of the noise estimation techniques in phase aware modified multi band spectral subtraction, *Aust J Electr Electron Eng,* **16(4)** (2019) 250–255, https://doi.org/10.1080/1448837X. 2019.1651137.
33 Hastie T, Rosset S, Zhu J & Zou H, Multiclassada boost, *Stat Its Interface*, **2(3)** (2009) 349–360, https://dx. doi.org/10.4310/ SII.2009.v2.n3.a8.
34 Saon G & Soltau H, Boosting systems for large vocabulary continuous speech recognition, *Speech Commun*, **54(2)** (2012) 212–218, https://doi.org/10.1016/j.specom.2011. 07.011.
35 Alreshidi A & Ullah M, Facial emotion recognition using hybrid features, *Inform*, **7(1)** (2020) 6, https://doi.org/10.3390/ informatics7010006.
36 Rao G A & Kishore P V V, Selfie sign language recognition with multiple features on ada boost multilabel multiclass classifier, *J Eng Sci Tech*, **13(8)** (2018) 2352–2368.
37 Villamizar S I, Sarmiento L C, López O, Caballero J & Bacca J, EEG Vowel silent speech signal discrimination based on APIT-EMD and SVD, in *Recent Advances in Electrical Engineering Related Sciences: Theory and Application. AETA 2019. Lecture Notes in Electrical Engineering* edited by T D Cortes, D V Hoang and D T Trong (Springer, Cham) 2019, 74–83, https://doi.org/10.1007/978-3-030-53021-1_8.
38 Liu Y, Zhou Y, Wen S & Tang C, A strategy on selecting performance metrics for classifier evaluation, *Int J Mob Comput Multimed Commun*, **6(4)** (2014) 20–35, https://doi.org/10.4018/ IJMCMC.2014100102.
39 Cai J, Ee D, Pham B, Roe P & Zhang J, Sensor network for the monitoring of ecosystem: bird species recognition, in *3rd Int Conf Intell Sensors, Sensor Networks Inform* (Melbourne, VIC, Australia*)* 2007, 293–298, https://doi.org/ 10.1109/ISSNIP.2007.4496859.
40 Boulmaiz A, Messadeg D, Doghmane N & Taleb-Ahmed A, Robust acoustic bird recognition for habitat monitoring with wireless sensor networks, *Int J Speech Technol*, **19(3)** (2016) 631–645, https://doi.org/10.1007/s10772-016-9354-4.
41 Nanni L, Costa Y M, Lucio D R, Silla C N & Brahnam S, Combining visual and acoustic features for bird species classification, in *IEEE 28th Int Conf Tools Artificial Intell* (IEEE) 2016, 396–401, https://doi.org/ 10.1109/ICTAI.2016.0067.
42 Neal L, Briggs F, Raich R & Fern X Z, Time- frequency segmentation of bird song in noisy acoustic environments, in *Int Conf Acoustics, Speech Signal Process* (IEEE) 2011, 2012–2015, http://dx.doi.org/10.1109/ ICASSP. 2011. 5946906.
43 Fagerlund S & Harma A, Parametrization of in harmonic bird sounds for automatic recognition, in *13th European Signal Process Conf* (IEEE) 2005, 1–4.
44 Mehyadin A E, Abdulazeez A M, Hasan D A & Saeed J N, Birds sound classification based on machine learning algorithms, *Asian J Res Comput Sci*, **9(4)** (2021) 1–11, https://doi.org/10.9734/AJRCOS/2021/v9i430227.
45 Fagerlund S, Bird species recognition using support vector machines, *Eurasip J Adv Sig Pr*, **7(1)** (2007) 1–8, https://doi.org/10.1155/2007/38637.