

## REFERENCES

- [1] “Welcome to ElephantVoices — elephantvoices.org,” <https://www.elephantvoices.org>, [Accessed 31-05-2024].
- [2] M. Mohaimenuzzaman, C. Bergmeir, I. West, and B. Meyer, “Environmental sound classification on the edge: A pipeline for deep acoustic networks on extremely resource-constrained devices,” *Pattern Recognition*, vol. 133, p. 109025, 2023.
- [3] U. Ranasinghe, S. Abeyrathne, L. Samaranayake, T. Weerakoon, N. Harischandra, and G. Dissanayake, “Enhanced frequency domain analysis for detecting wild elephants in asia using acoustics,” in *2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 2023, pp. 140–145.
- [4] Y. H. Lin and O. N. N. Fernando, “Animal hunt: Ai-based animal sound recognition application,” in *International Conference on Human-Computer Interaction*. Springer, 2023, pp. 465–474.
- [5] S. Li, Y. Yao, J. Hu, G. Liu, X. Yao, and J. Hu, “An ensemble stacked convolutional neural network model for environmental event sound recognition,” *Applied Sciences*, vol. 8, no. 7, p. 1152, 2018.
- [6] T. T. Leonid and R. Jayaparvathy, “Classification of elephant sounds using parallel convolutional neural network.” *intelligent automation & soft computing*, vol. 32, no. 3, 2022.
- [7] Y. Ren, M. T. Johnson, P. J. Clemins, M. Darre, S. S. Glaeser, T. S. Osiejuk, and E. Out-Nyarko, “A framework for bioacoustic vocalization analysis using hidden markov models,” *Algorithms*, vol. 2, no. 4, pp. 1410–1428, 2009.
- [8] “Department of Wildlife Conservation & Ministry of Wildlife and Forest Resources Conservation — mwfc.gov.lk,” <https://www.mwfc.gov.lk/department-of-wildlife-conservation/>, [Accessed 31-05-2024].
- [9] O. O. Abayomi-Alli, R. Damaševičius, A. Qazi, M. Adedoyin-Olowe, and S. Misra, “Data augmentation and deep learning methods in sound classification: A systematic review,” *Electronics*, vol. 11, no. 22, p. 3795, 2022.
- [10] M. Labied and A. Belangour, “Automatic speech recognition features extraction techniques: A multi-criteria comparison,” *International Journal of Advanced Computer Science and Applications*, 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:239089801>

- [11] L. Mutanu, J. Gohil, K. Gupta, P. Wagio, and G. Kotonya, “A review of automated bioacoustics and general acoustics classification research,” *Sensors*, vol. 22, no. 21, p. 8361, 2022.
- [12] M. Ravanelli and Y. Bengio, “Speaker recognition from raw waveform with sincnet,” in *2018 IEEE spoken language technology workshop (SLT)*. IEEE, 2018, pp. 1021–1028.
- [13] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu *et al.*, “Wavenet: A generative model for raw audio,” *arXiv preprint arXiv:1609.03499*, vol. 12, 2016.
- [14] “Transfer learning with YAMNet for environmental sound classification | TensorFlow Core — tensorflow.org,” [https://www.tensorflow.org/tutorials/audio/transfer\\_learning\\_audio](https://www.tensorflow.org/tutorials/audio/transfer_learning_audio), [Accessed 31-05-2024].
- [15] J.-w. Jung, H.-S. Heo, J.-h. Kim, H.-j. Shim, and H.-J. Yu, “Rawnet: Advanced end-to-end deep neural network using raw waveforms for text-independent speaker verification,” *arXiv preprint arXiv:1904.08104*, 2019.
- [16] A. Polukhin, “MobileNet Architectures — pandrii000,” <https://medium.com/@pandrii000/mobilenet-architectures-17fe7406d794>, [Accessed 31-05-2024].
- [17] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan *et al.*, “Searching for mobilenetv3,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314–1324.
- [18] P. J. Clemins, M. T. Johnson, K. M. Leong, and A. Savage, “Automatic classification and speaker identification of african elephant (*loxodonta africana*) vocalizations,” *The Journal of the Acoustical Society of America*, vol. 117, no. 2, pp. 956–963, 2005.
- [19] P. S. Duth, N. Manohar, R. Suresha, M. Priyanka, and T. Jipeng, “Wild animal recognition: A vocal analysis,” in *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2024, pp. 1478–1484.
- [20] J. C. Brown and P. Smaragdis, “Hidden markov and gaussian mixture models for automatic call classification,” *The Journal of the Acoustical Society of America*, vol. 125, no. 6, pp. EL221–EL224, 2009.
- [21] L. Vithakshana and W. Samankula, “Iot based animal classification system using convolutional neural network,” in *2020 International Research Conference on Smart Computing and Systems Engineering (SCSE)*. IEEE, 2020, pp. 90–95.

- [22] P. J. Clemins and M. T. Johnson, “Automatic type classification and speaker identification of african elephant vocalizations,” 2002.
- [23] J. Bjorck, B. H. Rappazzo, D. Chen, R. Bernstein, P. H. Wrege, and C. P. Gomes, “Automatic detection and compression for passive acoustic monitoring of the african forest elephant,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 476–484.
- [24] J. Xie, K. Hu, M. Zhu, and Y. Guo, “Bioacoustic signal classification in continuous recordings: Syllable-segmentation vs sliding-window,” *Expert Systems with Applications*, vol. 152, p. 113390, 2020.
- [25] A. S. Stoeger, A. Baotic, and G. Heilmann, “Vocal creativity in elephant sound production,” *Biology*, vol. 10, no. 8, p. 750, 2021.
- [26] “Research reveals the four sounds of Asian elephants | Whipsnade Zoo — whipsnadezoo.org,” <https://www.whipsnadezoo.org/zoo-stories/blog/research-reveals-four-sounds-of-asian-elephants>, [Accessed 31-05-2024].
- [27] F. J. B. Sanchez, M. R. Hossain, N. B. English, and S. T. Moore, “Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture,” *Scientific Reports*, vol. 11, 2021.
- [28] R. Mahum, A. Irtaza, and A. Javed, “Edl-det: A robust tts synthesis detector using vgg19-based yamnet and ensemble learning block,” *IEEE Access*, vol. 11, pp. 134 701–134 716, 2023.
- [29] R. Mahum, A. Irtaza, A. Javed, H. A. Mahmoud, and H. Hassan, “Deepdet: Yamnet with bottleneck attention module (bam) tts synthesis detection,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, no. 1, p. 18, 2024.
- [30] J. A. C. Nunes, D. Macêdo, and C. Zanchettin, “Am-mobilenet1d: A portable model for speaker recognition,” in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.
- [31] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.
- [32] X. Zhang, W. Fu, and M. Liang, “Multimodal emotion recognition from raw audio with sinc-convolution,” *arXiv preprint arXiv:2402.11954*, 2024.
- [33] R. R. Kvsn, J. Montgomery, S. Garg, and M. Charleston, “Bioacoustics data analysis—a taxonomy, survey and open challenges,” *IEEE Access*, vol. 8, pp. 57 684–57 708, 2020.

- [34] “Asian Elephant Vocalizations - Linguistic Data Consortium — catalog.ldc.upenn.edu,” <https://catalog.ldc.upenn.edu/LDC2010S05>, [Accessed 31-05-2024].
- [35] “What Elephant Calls Mean: A User’s Guide — nationalgeographic.com,” <https://www.nationalgeographic.com/animals/article/what-elephant-calls-mean>, [Accessed 02-06-2024].
- [36] S. Nair, R. Balakrishnan, C. S. Seelamantula, and R. Sukumar, “Vocalizations of wild asian elephants (*elephas maximus*): structural classification and social context,” *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2768–2778, 2009.
- [37] J. D. Wood, B. McCowan, W. R. Langbauer Jr, J. J. Viljoen, and L. A. Hart, “Classification of african elephant *loxodonta africana* rumbles using acoustic parameters and cluster analysis,” *Bioacoustics*, vol. 15, no. 2, pp. 143–161, 2005.
- [38] anaverahrami, “Deep into infrasound- elephantlisteningproject.org,” <https://www.elephantlisteningproject.org/all-about-infrasound/>, [Accessed 02-06-2024].
- [39] M. Pardo, “Call Combinations Differ Among Living Elephants,” <https://maximus.trunksnleaves.org/2019/05/01/call-combinations-differ/>, [Accessed 02-06-2024].
- [40] C. Y. Yeo, S. Al-Haddad, and C. K. Ng, “Animal voice recognition for identification (id) detection system,” in *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*. IEEE, 2011, pp. 198–201.
- [41] wpengine, “Congo Soundscapes - Public Database — elephantlisteningproject.org,” <https://www.elephantlisteningproject.org/congo-soundscapes-public-database/>, [Accessed 31-05-2024].
- [42] “ETHOGRAM TABLE — elephantvoices.org,” <https://www.elephantvoices.org/elephant-ethogram/ethogram-table/overview.html>, [Accessed 31-05-2024].
- [43] H.-C. Chu, Y.-L. Zhang, and H.-C. Chiang, “A cnn sound classification mechanism using data augmentation,” *Sensors*, vol. 23, no. 15, p. 6972, 2023.
- [44] “WO2012136380A1 - Time-stretching of an audio signal - Google Patents — patents.google.com,” <https://patents.google.com/patent/WO2012136380A1/en>, [Accessed 31-05-2024].

- [45] N. Juillerat and B. Hirsbrunner, “Low latency audio pitch shifting in the frequency domain,” in *2010 International Conference on Audio, Language and Image Processing*. IEEE, 2010, pp. 16–24.
- [46] S. Sharma, K. Sato, and B. P. Gautam, “A methodological literature review of acoustic wildlife monitoring using artificial intelligence tools and techniques,” *Sustainability*, vol. 15, no. 9, p. 7128, 2023.
- [47] “Spatial Filters - Gaussian Smoothing — homepages.inf.ed.ac.uk,” <https://homepages.inf.ed.ac.uk/rbf/HIPR2/gsmooth.htm>, [Accessed 31-05-2024].
- [48] A. Nshimiyimana, “Acoustic data augmentation for small passive acoustic monitoring datasets,” *Multimedia Tools and Applications*, pp. 1–19, 2024.
- [49] S. Wei, S. Zou, F. Liao *et al.*, “A comparison on data augmentation methods based on deep learning for audio classification,” in *Journal of physics: Conference series*, vol. 1453, no. 1. IOP Publishing, 2020, p. 012085.
- [50] A. Jain, P. R. Samala, D. Mittal, P. Jyoti, and M. Singh, “Spliceout: A simple and efficient audio augmentation method,” *arXiv preprint arXiv:2110.00046*, 2021.
- [51] T. Necciarri, P. Balazs, R. Kronland-Martinet, S. Ystad, B. Laback, S. Savel, and S. Meunier, “Auditory time-frequency masking: psychoacoustical data and application to audio representations,” in *International Symposium on Computer Music Modeling and Retrieval*. Springer, 2011, pp. 146–171.
- [52] L. Tacioli, L. Toledo, and C. Medeiros, “An architecture for animal sound identification based on multiple feature extraction and classification algorithms,” in *Anais do XI Brazilian e-Science Workshop*. SBC, 2017, pp. 29–36.
- [53] F. J. Bravo Sanchez, M. R. Hossain, N. B. English, and S. T. Moore, “Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture,” *Scientific Reports*, vol. 11, no. 1, p. 15733, 2021.
- [54] “Mel-frequency cepstrum - Wikipedia — en.wikipedia.org,” [https://en.wikipedia.org/w/index.php?title=Mel-frequency\\_cepstrum&oldid=1210791515](https://en.wikipedia.org/w/index.php?title=Mel-frequency_cepstrum&oldid=1210791515), [Accessed 02-06-2024].
- [55] B. Tracey, D. Volfson, J. Glass, R. Haulcy, M. Kostrzebski, J. Adams, T. Kangaroo, A. Brodtmann, E. R. Dorsey, and A. Vogel, “Towards interpretable speech biomarkers: exploring mfccs,” *Scientific Reports*, vol. 13, no. 1, p. 22787, 2023.
- [56] “Spectral Centroid - an overview | ScienceDirect Topics — sciencedirect.com,” <https://www.sciencedirect.com/topics/engineering/spectral-centroid>, [Accessed 31-05-2024].

- [57] “Zero Crossing Rate - an overview | ScienceDirect Topics — sciencedirect.com,” <https://www.sciencedirect.com/topics/engineering/zero-crossing-rate>, [Accessed 31-05-2024].
- [58] “3.10. Fundamental frequency (F0) &x2014; Introduction to Speech Processing — speechprocessingbook.aalto.fi,” [https://speechprocessingbook.aalto.fi/Representations/Fundamental\\_frequency\\_F0.html](https://speechprocessingbook.aalto.fi/Representations/Fundamental_frequency_F0.html), [Accessed 31-05-2024].
- [59] “3.5. Signal energy, loudness and decibel &x2014; Introduction to Speech Processing — speechprocessingbook.aalto.fi,” [https://speechprocessingbook.aalto.fi/Representations/Signal\\_energy\\_loudness\\_and\\_decibel.html](https://speechprocessingbook.aalto.fi/Representations/Signal_energy_loudness_and_decibel.html), [Accessed 31-05-2024].
- [60] J. Makhoul, “Linear prediction: A tutorial review,” *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.
- [61] C. Silva, V. Piyathilake, C. Keppitiyagama, A. Sayakkara, P. Kumarasinghe, N. Jayasuriya, and U. Sampath, “A wavelet transform-based feature extraction pipeline for elephant rumble detection,” *The International Journal on Advances in ICT for Emerging Regions*, vol. 16, no. 3, 2023.
- [62] M. P. Mcloughlin, R. Stewart, and A. G. McElligott, “Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring,” *Journal of the Royal Society Interface*, vol. 16, no. 155, p. 20190225, 2019.
- [63] R. Bajric, N. Zuber, G. A. Skrimpas, N. Mijatovic *et al.*, “Feature extraction using discrete wavelet transform for gear fault diagnosis of wind turbine gearbox,” *Shock and Vibration*, vol. 2016, 2016.
- [64] H. K. Turesson, S. Ribeiro, D. R. Pereira, J. P. Papa, and V. H. C. de Albuquerque, “Machine learning algorithms for automatic classification of marmoset vocalizations,” *PLoS one*, vol. 11, no. 9, p. e0163041, 2016.
- [65] J. V. Wijayakulasooriya, “Automatic recognition of elephant infrasound calls using formant analysis and hidden markov model,” in *2011 6th International Conference on Industrial and Information Systems*. IEEE, 2011, pp. 244–248.
- [66] Y.-J. Zhang, J.-F. Huang, N. Gong, Z.-H. Ling, and Y. Hu, “Automatic detection and classification of marmoset vocalizations using deep and recurrent neural networks,” *The Journal of the Acoustical Society of America*, vol. 144, no. 1, pp. 478–487, 2018.

- [67] “ORCA-SPOT: An Automatic Killer Whale Sound Detection Toolkit Using Deep Learning — ncbi.nlm.nih.gov,” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6662697/>, [Accessed 31-05-2024].
- [68] R. Mohanty, B. K. Mallik, and S. S. Solanki, “Automatic bird species recognition system using neural network based on spike,” *Applied Acoustics*, vol. 161, p. 107177, 2020.
- [69] M. A. Pardo, K. Fristrup, D. S. Lolchuragi, J. Poole, P. Granli, C. Moss, I. Douglas-Hamilton, and G. Wittemyer, “African elephants address one another with individually specific calls,” *bioRxiv*, pp. 2023–08, 2023.
- [70] G. Wassie Geremew, J. Ding *et al.*, “Elephant flows detection using deep neural network, convolutional neural network, long short-term memory, and auto-encoder,” *Journal of Computer Networks and Communications*, vol. 2023, 2023.
- [71] M. Zeppelzauer and A. S. Stoeger, “Establishing the fundamentals for an elephant early warning and monitoring system,” *BMC research notes*, vol. 8, pp. 1–15, 2015.
- [72] H. Dubey, A. Sangwan, and J. H. Hansen, “Transfer learning using raw waveform sincnet for robust speaker diarization,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6296–6300.
- [73] B. Kaur and J. Singh, “Audio classification: Environmental sounds classification,” 2021.
- [74] “YAMNet - CodiMD — codimd.mcl.math.ncu.edu.tw,” <https://codimd.mcl.math.ncu.edu.tw/s/hoOqEgBSf>, [Accessed 31-05-2024].
- [75] Y. He and Y. Wang, “Rawnet: Fast end-to-end neural vocoder,” *arXiv preprint arXiv:1904.05351*, 2019.
- [76] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
- [77] T. Paranayapa, P. Ranasinghe, D. Ranmal, D. Meedeniya, and C. Perera, “A comparative study of preprocessing and model compression techniques in deep learning for forest sound classification,” *Sensors*, vol. 24, no. 4, p. 1149, 2024.
- [78] A. Bajaj, “Performance Metrics in Machine Learning [Complete Guide],” <https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide>, [Accessed 31-05-2024].

- [79] A. Al Bashit and D. Valles, "A mel-filterbank and mfcc-based neural network approach to train the houston toad call detection system design," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2018, pp. 438–443.
- [80] P. J. Clemins and M. T. Johnson, "Application of speech recognition to african elephant (*loxodonta africana*) vocalizations," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, vol. 1. IEEE, 2003, pp. I–I.
- [81] K. P. V. S. A. Chintada Chandu Sai, Ayush Dewangan, "AUTOMATIC ELEPHANT DETECTION SYSTEM USING MACHINE LEARNING AND IOT TECHNIQUES," <https://www.ijnrd.org/papers/IJNRD2306612.pdf>, [Accessed 31-05-2024].
- [82] L. Muda, M. Begam, and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques," *arXiv preprint arXiv:1003.4083*, 2010.
- [83] J. Xie, M. Zhu, K. Hu, J. Zhang, H. Hines, and Y. Guo, "Frog calling activity detection using lightweight cnn with multi-view spectrogram: A case study on kroombit tinker frog," *Machine Learning with Applications*, vol. 7, p. 100202, 2022.
- [84] J. Liao, H. Li, A. Feng, X. Wu, Y. Luo, X. Duan, M. Ni, and J. Li, "Domestic pig sound classification based on transformercnn," *Applied Intelligence*, vol. 53, no. 5, pp. 4907–4923, 2023.
- [85] Z. Mushtaq and S.-F. Su, "Efficient classification of environmental sounds through multiple features aggregation and data enhancement techniques for spectrogram images," *Symmetry*, vol. 12, no. 11, p. 1822, 2020.
- [86] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.