

- [21] I. Trowitzsch, J. Taghia, Y. Kashef, and K. Obermayer, "The NIGENS General Sound Events Database," Feb. 2019, doi: 10.5281/zenodo.2535878.
- [22] Q. Zhou *et al.*, "Cough Recognition Based on Mel-Spectrogram and Convolutional Neural Network," *Front. Robot. AI*, vol. 8, no. May, pp. 1–7, 2021, doi: 10.3389/frobt.2021.580080.
- [23] I. Tsiamas, G. I. Gállego, J. A. R. Fonollosa, and M. R. Costa-Jussà, "SHAS: Approaching optimal Segmentation for End-to-End Speech Translation," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2022-September, pp. 106–110, 2022, doi: 10.21437/Interspeech.2022-59.
- [24] S. Sahoo, P. Kumar, B. Raman, and P. P. Roy, "A Segment Level Approach to Speech Emotion Recognition Using Transfer Learning," 2020, doi: 10.1007/978-3-030-41299-9_34.
- [25] H. Yakura and J. Sakuma, "Robust audio adversarial example for a physical attack," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2019-August, pp. 5334–5341, 2019, doi: 10.24963/ijcai.2019/741.
- [26] J. B. Allen, "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform," *IEEE Trans. Acoust.*, vol. 25, no. 3, pp. 235–238, 1977, doi: 10.1109/TASSP.1977.1162950.
- [27] G. Rybak and K. Strzecha, "Short-time fourier transform based on metaprogramming and the stockham optimization method," *Sensors*, vol. 21, no. 12, 2021, doi: 10.3390/s21124123.
- [28] T. Tran and J. Lundgren, "Drill fault diagnosis based on the scalogram and MEL spectrogram of sound signals using artificial intelligence," *IEEE Access*, vol. 8, pp. 203655–203666, 2020, doi: 10.1109/ACCESS.2020.3036769.
- [29] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [30] L. H. Iksan, M. I. Awal, R. Z. Fhamy, A. A. Pratama, D. K. Basuki, and S. Sukaridhoto, "Implementation of Cloud Based Action Recognition Backend Platform," *AIMS 2021 - Int. Conf. Artif. Intell. Mechatronics Syst.*, 2021, doi: 10.1109/AIMS2415.2021.9466068.