

References

- Allen, J. B. (1982). Applications of the short time Fourier transform to speech processing and spectral analysis. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. Vol. 1982-May. DOI: 10.1109/ICASSP.1982.1171703.
- Ayadi, Moataz El, Kamel, Mohamed S., and Karray, Fakhri (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. In: *Pattern Recognition* 44 (3). ISSN: 00313203. DOI: 10.1016/j.patcog.2010.09.020.
- Ba, Jimmy Lei, Kiros, Jamie Ryan, and Hinton, Geoffrey E. (July 2016). Layer Normalization. In: *Arxiv*.
- Bahdanau, Dzmitry, Cho, Kyung Hyun, and Bengio, Yoshua (2015). Neural machine translation by jointly learning to align and translate. In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dabbabi, Karim and Mars, Abdelkarim (Oct. 2024). Self-supervised Learning for Speech Emotion Recognition Task Using Audio-visual Features and Distil Hubert Model on BAVED and RAVDESS Databases. In: *Journal of Systems Science and Systems Engineering* 33 (5), pp. 576–606. ISSN: 1004-3756. DOI: 10.1007/s11518-024-5607-y.
- Davis, Steven B. and Mermelstein, Paul (1980). *Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences*. DOI: 10.1109/TASSP.1980.1163420.
- Deshpande, Kedar, Sodhi, Manjit Singh, Raniyer, Nidhi, and Rao, Madhav (Nov. 2024). A Time-Distributed CNN-LSTM with Attention Model for Speech Based Emotion Recognition. In: *Proceedings of the 2024 7th International Conference on Digital Medicine and Image Processing*. ACM, pp. 67–71. ISBN: 9798400709586. DOI: 10.1145/3705927.3705939.
- Dey, Rahul and Salemt, Fathi M. (2017). Gate-variants of Gated Recurrent Unit (GRU) neural networks. In: *Midwest Symposium on Circuits and Systems*. Vol. 2017-August. DOI: 10.1109/MWSCAS.2017.8053243.
- Escalante, Hugo Jair, Montes, Manuel, and ESUCAR, Luis Enrique Sucar (2009). *Particle Swarm Model Selection*. Tech. rep., pp. 405–440.

- Gad, Ahmed G. (2022). Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review. In: *Archives of Computational Methods in Engineering* 29 (5). ISSN: 18861784. DOI: 10.1007/s11831-021-09694-4.
- Guruvaya, Aruna Mogarala, Kollu, Archana, Divakarachari, Parameshachari Bidare, Falkowski-Gilski, Przemysław, and Praveena, Hiralid Dwaraka (July 2024). Bi-GRU-APSO: Bi-Directional Gated Recurrent Unit with Adaptive Particle Swarm Optimization Algorithm for Sales Forecasting in Multi-Channel Retail. In: *Telecom* 5 (3), pp. 537–555. ISSN: 2673-4001. DOI: 10.3390/telecom5030028.
- Habib, Maria, Faris, Mohammad, Qaddoura, Raneem, Alomari, Manal, Alomari, Alaa, and Faris, Hossam (2021). Toward an automatic quality assessment of voice-based telemedicine consultations: A deep learning approach. In: *Sensors* 21 (9). ISSN: 14248220. DOI: 10.3390/s21093279.
- Hashem, Ahlam, Arif, Muhammad, and Alghamdi, Manal (2023). *Speech emotion recognition approaches: A systematic review*. DOI: 10.1016/j.specom.2023.102974.
- Hendrycks, Dan and Gimpel, Kevin (2016). Bridging Nonlinearities and Stochastic Regularizers with Gaussian Error Linear Units. In: *CoRR* abs/1606.08415. arXiv: 1606.08415. URL: <http://arxiv.org/abs/1606.08415>.
- Huang, Chuan-Bao, Zhu, Kai, and Hu, Zhen (Jan. 2024). An improved speech emotion recognition method based on RepVGG network. In: *Proceedings of the 2024 3rd International Conference on Cryptography, Network Security and Communication Technology*. ACM, pp. 451–457. ISBN: 9798400716959. DOI: 10.1145/3673277.3673355.
- Ige, Ayokunle Olalekan and Sibiyi, Malusi (2024). State-of-the-Art in 1D Convolutional Neural Networks: A Survey. In: *IEEE Access* 12, pp. 144082–144105. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3433513.
- Issa, Dias, Demirci, M. Fatih, and Yazici, Adnan (2020). Speech emotion recognition with deep convolutional neural networks. In: *Biomedical Signal Processing and Control* 59. ISSN: 17468108. DOI: 10.1016/j.bspc.2020.101894.
- Kehtarnavaz, Nasser (2008). CHAPTER 7 - Frequency Domain Processing. In: *Digital Signal Processing System Design (Second Edition)*. Ed. by Kehtarnavaz, Nasser. Second Edition. Academic Press, pp. 175–196. ISBN: 978-0-12-374490-6. DOI: <https://doi.org/10.1016/B978-0-12-374490-6.00007-6>. URL: <https://www.sciencedirect.com/science/article/pii/B9780123744906000076>.

- Khurma, Manan (n.d.). *ANOVA Test*. URL: <https://www.cuemath.com/anova-formula/>.
- Kiranyaz, Serkan, Avci, Onur, Abdeljaber, Osama, Ince, Turker, Gabbouj, Moncef, and Inman, Daniel J. (2021). 1D convolutional neural networks and applications: A survey. In: *Mechanical Systems and Signal Processing* 151. ISSN: 10961216. DOI: 10.1016/j.ymssp.2020.107398.
- Kiranyaz, Serkan, Ince, Turker, and Gabbouj, Moncef (2016). Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. In: *IEEE Transactions on Biomedical Engineering* 63 (3). ISSN: 15582531. DOI: 10.1109/TBME.2015.2468589.
- Li, Dongdong, Liu, Jinlin, Yang, Zhuo, Sun, Linyu, and Wang, Zhe (2021). Speech emotion recognition using recurrent neural networks with directional self-attention. In: *Expert Systems with Applications* 173. ISSN: 09574174. DOI: 10.1016/j.eswa.2021.114683.
- Li, Gaoyun, Liu, Yong, and Wang, Xiong (2023). Speech Emotion Recognition Based on 1D CNN and MFCC. In: *Proceedings of 2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology, ICCASIT 2023*. DOI: 10.1109/ICCASIT58768.2023.10351697.
- Liu, Xinyu, Wang, Yongjun, Wang, Xishuo, Xu, Hui, Li, Chao, and Xin, Xiangjun (2021). Bi-directional gated recurrent unit neural network based nonlinear equalizer for coherent optical communication system. In: *Optics Express* 29 (4). ISSN: 10944087. DOI: 10.1364/oe.416672.
- Livingstone, Steven R. and Russo, Frank A. (May 2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. In: *PLOS ONE* 13 (5), e0196391. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0196391.
- Logan, Beth (2000). Mel Frequency Cepstral Coefficients for Music Modeling. In: *International Symposium on Music Information Retrieval* 28. ISSN: 1098-6596. DOI: 10.1.1.11.9216.
- López-Espejo, Iván, Joglekar, Aditya, Peinado, Antonio M, and Jensen, Jesper (2024). *On Speech Pre-emphasis as a Simple and Inexpensive Method to Boost Speech Enhancement*. URL: <https://arxiv.org/abs/2401.09315>.
- Mao, Jun Wei, He, Yong, and Liu, Zhen Tao (2018). Speech emotion recognition based on linear discriminant analysis and support vector machine decision tree. In: *Chinese Control Conference, CCC*. Vol. 2018-July. DOI: 10.23919/ChiCC.2018.8482931.

- Müller, Meinard (2021). *An Educational Guide through the FMP Notebooks for Teaching and Learning Fundamentals of Music Processing*. DOI: 10.3390/signals2020018.
- Mustaqeem and Kwon, Soonil (2020). A CNN-assisted enhanced audio signal processing for speech emotion recognition. In: *Sensors (Switzerland)* 20 (1). ISSN: 14248220. DOI: 10.3390/s20010183.
- Nanga, Salifu, Bawah, Ahmed Tijani, Acquaye, Benjamin Ansah, Billa, Mac-Issaka, Baeta, Francis Delali, Odai, Nii Afotey, Obeng, Samuel Kwaku, and Nsiah, Ampem Darko (2021). Review of Dimension Reduction Methods. In: *Journal of Data Analysis and Information Processing* 09 (03). ISSN: 2327-7211. DOI: 10.4236/jdaip.2021.93013.
- Nie, Feiping, Wang, Zheng, Wang, Rong, Wang, Zhen, and Li, Xuelong (2020). Adaptive local linear discriminant analysis. In: *ACM Transactions on Knowledge Discovery from Data* 14 (1). ISSN: 1556472X. DOI: 10.1145/3369870.
- Novais, Rui, Cardoso, Pedro J.S., and Rodrigues, João M.F. (2022). Emotion Classification from Speech by an Ensemble Strategy. In: *ACM International Conference Proceeding Series*. DOI: 10.1145/3563137.3563170.
- Paszke, Adam, Gross, Sam, Massa, Francisco, Lerer, Adam, Bradbury, James, Chanan, Gregory, Killeen, Trevor, Lin, Zeming, Gimeshein, Natalia, Antiga, Luca, Desmaison, Alban, Köpf, Andreas, Yang, Edward, DeVito, Zach, Raison, Martin, Tejani, Alykhan, Chilamkurthy, Sasank, Steiner, Benoit, Fang, Lu, Bai, Junjie, and Chintala, Soumith (2019). PyTorch: An imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems*. Vol. 32.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. In: *Journal of Machine Learning Research* 12, pp. 2825–2830.
- Pereira, Luis Manuel, Salazar, Addisson, and Vergara, Luis (2023). On Comparing Early and Late Fusion Methods. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 14134 LNCS. Springer Science and Business Media Deutschland GmbH, pp. 365–378. ISBN: 9783031430848. DOI: 10.1007/978-3-031-43085-5_29.

- Qu, Taiguo and Wang, Dong (Dec. 2023). A Fast Multidimensional Scaling Algorithm. In: *The 3rd International Conference on Electronic Information Technology and Smart Agriculture*. ACM, pp. 426–430. ISBN: 9798400716775. DOI: 10.1145/3641343.3641431.
- Saeed, Nasir, Nam, Haewoon, Haq, Mian Imtiaz Ul, and Bhatti, Dost Muhammad Saqib (2018). *A survey on multidimensional scaling*. DOI: 10.1145/3178155.
- Sawilowsky, S.S. (2009). New effect size rules of thumb. In: *Journal of Modern Applied Statistical Methods* 8.2, pp. 597–599. DOI: 10.22237/jmasm/1257035100.
- Singh, Prabhav, Srivastava, Ridam, Rana, K. P.S., and Kumar, Vineet (2021). A multimodal hierarchical approach to speech emotion recognition from audio and text[Formula presented]. In: *Knowledge-Based Systems* 229. ISSN: 09507051. DOI: 10.1016/j.knosys.2021.107316.
- Srivastava, Nitish, Hinton, Geoffrey, Krizhevsky, Alex, Sutskever, Ilya, and Salakhutdinov, Ruslan (2014). Dropout: A simple way to prevent neural networks from overfitting. In: *Journal of Machine Learning Research* 15. ISSN: 15337928.
- Stevens, S. S., Volkman, J., and Newman, E. B. (1937). A Scale for the Measurement of the Psychological Magnitude Pitch. In: *The Journal of the Acoustical Society of America* 8 (3). ISSN: 0001-4966. DOI: 10.1121/1.1915893.
- Strasser, Shane, Goodman, Rollie, Sheppard, John, and Butcher, Stephyn (2016). A new discrete Particle Swarm Optimization algorithm. In: *GECCO 2016 - Proceedings of the 2016 Genetic and Evolutionary Computation Conference*. DOI: 10.1145/2908812.2908935.
- Subbarao, M. Venkata, Terlapu, Sudheer Kumar, and Chowdary, Paladuga Satish Rama (2022). Emotion Recognition using BiLSTM Classifier. In: *Proceedings - 2022 International Conference on Computing, Communication and Power Technology, IC3P 2022*. DOI: 10.1109/IC3P52835.2022.00048.
- Torgerson, Warren S. (1952). Multidimensional scaling: I. Theory and method. In: *Psychometrika* 17 (4). ISSN: 18600980. DOI: 10.1007/BF02288916.
- Trigeorgis, George, Ringeval, Fabien, Brueckner, Raymond, Marchi, Erik, Nicolaou, Mihalis A., Schuller, Bjorn, and Zafeiriou, Stefanos (2016). Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. Vol. 2016-May. DOI: 10.1109/ICASSP.2016.7472669.

- Ullah, Rizwan, Asif, Muhammad, Shah, Wahab Ali, Anjam, Fakhar, Ullah, Ibrar, Khurshaid, Tahir, Wuttisittikulki, Lunchakorn, Shah, Shashi, Ali, Syed Mansoor, and Alibakhshikenari, Mohammad (2023). Speech Emotion Recognition Using Convolution Neural Networks and Multi-Head Convolutional Transformer. In: *Sensors* 23 (13). ISSN: 14248220. DOI: 10.3390/s23136212.
- Xie, Yue, Liang, Ruiyu, Liang, Zhenlin, and Zhao, Li (2019). Attention-based dense LSTM for speech emotion recognition. In: *IEICE Transactions on Information and Systems* E102D (7). ISSN: 17451361. DOI: 10.1587/transinf.2019EDL8019.
- Yadav, Om, Bastola, Laxmi, and Sharma, Jagdish (Oct. 2021). Speech Emotion Recognition using Convolutional Recurrent Neural Network. In.
- Zhao, Jianfeng, Mao, Xia, and Chen, Lijiang (2019). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. In: *Biomedical Signal Processing and Control* 47. ISSN: 17468108. DOI: 10.1016/j.bspc.2018.08.035.
- Zhu, Zijiang, Dai, Weihuang, Hu, Yi, and Li, Junshan (2020). Speech emotion recognition model based on Bi-GRU and Focal Loss. In: *Pattern Recognition Letters* 140. ISSN: 01678655. DOI: 10.1016/j.patrec.2020.11.009.