DAFTAR PUSTAKA

Aji, A. F., Winata, G. I., Koto, F., Cahyawijaya, S., Romadhony, A., Mahendra, R., Kurniawan, K., Moeljadi, D., Prasojo, R. E., Baldwin, T., Lau, J. H., & Ruder, S. (2022). One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, *1*, 7226–7249. https://doi.org/10.18653/v1/2022.acl-long.500

Akçay, M. B., & Oğuz, K. (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*, *116*, 56–76. https://doi.org/10.1016/J.SPECOM.2019.12.001

Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., & Weber, G. (2020). Common voice: A massively-multilingual speech corpus. *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, 4218–4222.

Arquitectura, E. Y., Introducci, T. I., 赫暁霞, Iv, T., Teatinas, L. A. S., Conclusiones, T. V. I. I., Contemporáneo, P. D. E. U. S. O., Evaluaci, T. V, Ai, F., Jakubiec, J. A., Weeks, D. P. C. C. L. E. Y. N. to K. in 20, Mu, A., Inan, T., Sierra Garriga, C., Library, P. Y., Hom, H., Kong, H., Castilla, N., Uzaimi, A., … Waldenström, L. (2015). LIBRISPEECH: AN ASR CORPUS BASED ON PUBLIC DOMAIN AUDIO BOOKS Vassil. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *53*(9), 1689–1699. http://publications.lib.chalmers.se/records/fulltext/245180/245180.pdf%0Ahttps://hdl.handle.net/20.500.12380/245180%0Ahttp://dx.doi.org/10.1016/j.jsames.2011.03.003%0Ahttps://doi.org/10.1016/j.gr.2017.08.001%0Ahttp://dx.doi.org/10.1016/j.precamres.2014.12

Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, *2020-Decem*(Figure 1), 1–19.

Barrett, L. F. (2017). How emotions are made: The secret life of the brain. In *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Bgn, J. (2021). *An Illustrated Tour of Wav2vec 2.0*. Jonathan Bgn.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–347. https://www.researchgate.net/publication/208032992_PRAAT_a_system_for_doing_phonetics_by_computer

Bryant, G. A., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, *8*(1–2), 135–148. https://doi.org/10.1163/156770908X289242

Cao, H., Cooper, D. G., Keutmann, M. K., Gur, R. C., Nenkova, A., & Verma, R. (2014). CREMA-D: Crowd-sourced emotional multimodal actors dataset. *IEEE Transactions on Affective Computing*, *5*(4). https://doi.org/10.1109/TAFFC.2014.2336244

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *37th International Conference on Machine Learning, ICML 2020*, *PartF16814*(Figure 1), 1575–1585.

Chen, W., Xing, X., Xu, X., & Yang, J. (2021). *Key-Sparse Transformer with Cascaded Cross-Attention Block for Multimodal Speech Emotion Recognition*. http://arxiv.org/abs/2106.11532

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*(Mlm), 4171–4186.

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Géron, A. (2023). Hands-on machine learning with Scikit-Learn, Keras and TensorFlow : concepts, tools, and techniques to build intelligent systems. In *TA - TT* - (Third edit). O'Reilly Media, Inc. https://doi.org/LK - https://worldcat.org/title/1346503549

Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M. (2020). Spleeter: a fast and efficient music source separation tool with pre-trained models. *Journal of Open Source Software*, *5*(50), 2154. https://doi.org/10.21105/JOSS.02154

Lasiman, J. J., & Lestari, D. P. (2018). Speech Emotion Recognition for Indonesian Language Using Long Short-Term Memory. *2018 International Conference on Computer, Control, Informatics and Its Applications: Recent Challenges in Machine Learning for Computing Applications, IC3INA 2018 - Proceeding*, 40–43. https://doi.org/10.1109/IC3INA.2018.8629525

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. In *American Psychologist* (Vol. 46, Issue 8, pp. 819–834). American Psychological Association. https://doi.org/10.1037/0003-066X.46.8.819

Li, Y., Bell, P., & Lai, C. (2023). Transfer Learning for Personality Perception via Speech Emotion Recognition. *Proceedings of the Annual Conference of the*

International Speech Communication Association, INTERSPEECH, *2023-Augus*, 5197–5201. https://doi.org/10.21437/Interspeech.2023-2061

Livingstone, S. R., & Russo, F. A. (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. *PLoS ONE*, *13*(5). https://doi.org/10.1371/journal.pone.0196391

Ma, J., Pan, S., Chandran, D., Fanelli, A., & Cartwright, R. (2023). *Low latency transformers for speech processing*. http://arxiv.org/abs/2302.13451

Mehrish, A., Majumder, N., Bharadwaj, R., Mihalcea, R., & Poria, S. (2023). A review of deep learning techniques for speech processing. *Information Fusion*, *99*. https://doi.org/10.1016/j.inffus.2023.101869

Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive, Affective & Behavioral Neuroscience*, *3*(3), 207–233. https://doi.org/10.3758/cabn.3.3.207

Nagraniy, A., Chungy, J. S., & Zisserman, A. (2017). VoxCeleb: A large-scale speaker identification dataset. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2017-Augus*, 2616–2620. https://doi.org/10.21437/Interspeech.2017-950

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Pepino, L., Riera, P., & Ferrer, L. (2021). Emotion recognition from speech using wav2vec 2.0 embeddings. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *1*, 551–555. https://doi.org/10.21437/Interspeech.2021-703

Pichora-Fuller, M. K., & Dupuis, K. (2020). Toronto emotional speech set (TESS). *Scholars Portal Dataverse*, *1*.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Saganowski, S. (2022). Bringing Emotion Recognition out of the Lab into Real Life: Recent Advances in Sensors and Machine Learning. *Electronics (Switzerland)*, *11*(3), 1–19. https://doi.org/10.3390/electronics11030496

Samin, Y., Shanjidul, M., Sadhin, I., & Ifty, R. H. (2023). *Speech Emotion Recognition using Transfer Learning Approach and Real-Time Evaluation in English and Bengali Language*. *January*. https://doi.org/10.13140/RG.2.2.31324.87684

Scherer, K. R. (1999). Appraisal theory. In *Handbook of cognition and emotion.* (pp. 637–663). John Wiley & Sons Ltd. https://doi.org/10.1002/0470013494.ch30

Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., Chetouani, M., Weninger, F., Eyben, F., Marchi, E., Mortillaro, M., Salamin, H., Polychroniou, A., Valente, F., & Kim, S. (2013). The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *May 2014*, 148–152. https://doi.org/10.21437/interspeech.2013-56

Sen, S., Dutta, A., & Dey, N. (n.d.). *Audio Processing and Speech Recognition:Concepts, Techniques and Research Overviews*. Springer Singapore. https://doi.org/https://doi.org/10.1007/978-981-13-6098-5

Sharma, G., Umapathy, K., & Krishnan, S. (2020). Trends in audio signal feature extraction methods. *Applied Acoustics*, *158*, 107020. https://doi.org/10.1016/j.apacoust.2019.107020

Singh, A., & Gupta, A. (2023). *Decoding Emotions: A comprehensive Multilingual Study of Speech Models for Speech Emotion Recognition*. http://arxiv.org/abs/2308.08713

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*(1), 77–89. https://doi.org/https://doi.org/10.1016/S0034-4257(97)00083-7

Tang, D., Kuppens, P., Geurts, L., & van Waterschoot, T. (2023). *End-to-end Transfer Learning for Speaker-independent Cross-language Speech Emotion Recognition*. http://arxiv.org/abs/2311.13678

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). [Transformer] Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-Decem*(Nips), 5999–6009.

Wagner, J., Triantafyllopoulos, A., Wierstorf, H., Schmitt, M., Burkhardt, F., & Eyben, F. (2023). *DAWN OF THE TRANSFORMER ERA IN SPEECH EMOTION RECOGNITION: CLOSING THE VALENCE GAP*.

Wu, H., Liu, X., Hagan, C. C., & Mobbs, D. (2020). Mentalizing during social InterAction: A four component model. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *126*, 242–252. https://doi.org/10.1016/j.cortex.2019.12.031

Wunarso, N. B., & Soelistio, Y. E. (2017). Towards Indonesian speech-emotion automatic recognition (I-SpEAR). *Proceedings of 2017 4th International Conference on New Media Studies, CONMEDIA 2017*, *2018-Janua*, 98–101. https://doi.org/10.1109/CONMEDIA.2017.8266038

Yenduri, G., M, R., G, C. S., Y, S., Srivastava, G., Maddikunta, P. K. R., G, D. R., Jhaveri, R. H., B, P., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2023). *Generative Pre-trained Transformer: A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions*. 1–40. http://arxiv.org/abs/2305.10435

Zahra, H. N., Ibrohim, M. O., Fahmi, J., Adelia, R., Nur Febryanto, F. A., & Riandi, O. (2020). Speech emotion recognition on indonesian youtube web series using deep learning approach. *2020 5th International Conference on Informatics and Computing, ICIC 2020*. https://doi.org/10.1109/ICIC50835.2020.9288650

Zaidi, S. A. M., Latif, S., & Qadir, J. (2023). *Cross-Language Speech Emotion Recognition Using Multimodal Dual Attention Transformers*. 1–14. http://arxiv.org/abs/2306.13804