DAFTAR PUSTAKA

[1] V. Bhardwaj, V. Kukreja, M. T. B. Othman, Y. Belkhier, M. Bajaj, B. S. Goud, A. U. Rehman, M. Shafiq, and H. Hamam, "Automatic speech recognition (asr) systems for children: A systematic literature review," *Applied Sciences 2022, Vol. 12, Page 4419*, vol. 12, p. 4419, 4 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/9/4419/htmhttps://www.mdpi.com/2076-3417/12/9/4419

[2] A. V. K. Shanbhogue, R. Xue, S. Saha, D. Y. Zhang, and A. Ganesan, "Improving low resource speech translation with data augmentation and ensemble strategies," *20th International Conference on Spoken Language Translation, IWSLT 2023 - Proceedings of the Conference*, pp. 241–250, 2023. [Online]. Available: https://aclanthology.org/2023.iwslt-1.21

[3] R. Coto-Solano, S. A. Nicholas, S. Datta, V. Quint, P. Wills, E. N. Powell, L. Koka'ua, S. Tanveer, and I. Feldman, "Development of automatic speech recognition for the documentation of cook islands m⁻ aori," pp. 20–25, 2022.

[4] M. Bartelds, N. San, B. McDonnell, D. Jurafsky, and M. Wieling, "Making more of little data: Improving low-resource automatic speech recognition using data augmentation," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 715–729, 2023. [Online]. Available: https://aclanthology.org/2023.acl-long.42

[5] M. P. Lewis, Ed., *Ethnologue: Languages of the World*, 25th ed. Dallas, TX, USA: SIL International, 2022.

[6] A. O. Anindyatri and I. Mufidah, "Gambaran kondisi vitalitas bahasa daerah di indonesia: berdasarkan data tahun 2018 – 2019," p. 43, 2020.

[7] A. Soeharno and H. A. A. Faruq, "A semantic analysis on the javanese word "jajan" as used by "suara akbar" radio listener in jember," *MEDIAKOM*, 2022.

[8] P. Arisaputra, A. T. Handoyo, and A. Zahra, "Xls-r deep learning model for multilingual asr on low-resource languages: Indonesian, javanese, and sundanese," 2020.

[9] K. Azizah and M. Adriani, "Hierarchical transfer learning for text-to-speech in indonesian, javanese, and sundanese languages," *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 421–428, 2020.

[10] S. Cahyawijaya, G. I. Winata, B. Wilie, K. Vincentio, X. Li, A. Kuncoro, S. Ruder, Z. Y. Lim, S. Bahar, M. L. Khodra, A. Purwarianti, and P. Fung, "Indonlg: Benchmark and resources for evaluating indonesian natural language generation," pp. 8875–8898, 2021.

[11] L. Susanto, R. Diandaru, A. A. Krisnadhi, A. Purwarianti, and D. Wijaya, "Replicable benchmarking of neural machine translation (nmt) on low-resource local languages in indonesia," *ArXiv*, vol. abs/2311.00998, 2023.

[12] C.-C. Chiu, T. N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen, Z. Chen, A. Kannan, R. J. Weiss, K. Rao, E. Gonina, N. Jaitly, B. Li, J. Chorowski, and M. Bacchiani, "State-of-the-art speech recognition with sequence-to-sequence models," in *ICASSP*, 2018.

[13] N. Rossenbach, A. Zeyer, R. Schluter, and H. Ney, "Generating synthetic audio data for attention-based speech recognition systems," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2020-May, pp. 7069–7073, 5 2020.

[14] J. Mira, L. F. Seoane, J. J. Nieto, L. Hong, and Y. Murakami, "Indonesia language sphere: an ecosystem for dictionary development for low-resource languages the importance of interlinguistic similarity and stable bilingualism when two languages compete design of networked intelligent translation system based on machine learning algorithm indonesia language sphere: an ecosystem for dictionary development for low-resource languages," p. 12001, 2019. [Online]. Available: https://catalog.ldc.upenn.edu

[15] R. Joel, Z. Salazar, D. Núria, and B. L. T. I. Aplicada, "Text-to-speech data augmentation for low resource speech recognition," 4 2022. [Online]. Available: https://arxiv.org/abs/2204.00291v1

[16] A. Renduchintala, S. Ding, M. Wiesner, and S. Watanabe, "Multi-modal data augmentation for end-to-end asr," 2018. [Online]. Available: https://dumps.wikimedia.org/backup-index.html

[17] X. Cui, V. Goel, and B. Kingsbury, "Data augmentation for deep neural network acoustic modeling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, pp. 1469–1477, 2015.

[18] C. Bhat, A. Panda, and H. Strik, "Improved asr performance for dysarthric speech using two-stage dataaugmentation," pp. 46–50, 2022.

[19] R. Nahar, S. Miwa, and A. Kai, "Domain adaptation with augmented data by deep neural network based method using re-recorded speech for automatic speech recognition in real environment," *Sensors 2022, Vol. 22, Page 9945*, vol. 22, p. 9945, 12 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/24/9945/htmhttps://www.mdpi.com/1424-8220/22/24/9945

[20] U. Shrawankar and V. Thakare, "Adverse conditions and asr techniques for robust speech user interface," 3 2013. [Online]. Available: https://arxiv.org/abs/1303.5515v1

[21] L. Wang, M. Fazel-Zarandi, A. Tiwari, S. Matsoukas, and L. Polymenakos, "Data augmentation for training dialog models robust to speech recognition errors," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 63–70, 2020. [Online]. Available: https://aclanthology.org/2020.nlp4convai-1.8

[22] Z. Jin, M. Geng, X. Xie, J. Yu, S. Liu, X. Liu, and H. Meng, "Adversarial data augmentation for disordered speech recognition," 8 2021. [Online]. Available: https://www.researchgate.net/publication/353653671_Adversarial_Data_Augmentation_for_Disordered_Speech_Recognition

[23] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, and A. Ng, "Deep speech: Scaling up end-to-end speech recognition," *arXiv*, 2014.

[24] T. Ko, V. Peddinti, D. Povey, and S. Khudanpur, "Audio augmentation for speech recognition," in *INTERSPEECH*, 2015.

[25] N. Jaitly and G. Hinton, "Vocal tract length perturbation (vtlp) improves speech recognition," in *ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.

[26] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "Specaugment: A simple data augmentation method for automatic speech recognition," *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2019-September, pp. 2613–2617, 4 2019. [Online]. Available: http://arxiv.org/abs/1904.08779http://dx.doi.org/10.21437/Interspeech.2019-2680

[27] J. Hwang, M. Hira, C. Chen, X. Zhang, Z. Ni, G. Sun, P. Ma, R. Huang, V. Pratap, Y. Zhang, A. Kumar, C.-Y. Yu, C. Zhu, C. Liu, J. Kahn, M. Ravanelli, P. Sun, S. Watanabe, Y. Shi, Y. Tao, R. Scheibler, S. Cornell, S. Kim, and S. Petridis, "Torchaudio 2.1: Advancing speech recognition, self-supervised learning, and audio processing components for pytorch," 10 2023. [Online]. Available: https://arxiv.org/abs/2310.17864v1

[28] S. Cahyawijaya, H. Lovenia, A. F. Aji, G. I. Winata, B. Wilie, F. Koto, R. Mahendra, C. Wibisono, A. Romadhony, K. Vincentio, J. Santoso, D. Moeljadi, C. Wirawan, F. Hudi, M. S. Wicaksono, I. H. Parmonangan, I. Alfina, I. F. Putra, S. Rahmadani, Y. Oenang, A. A. Septiandri, J. Jaya, K. D. Dhole, A. A. Suryani, R. A. Putri, D. Su, K. Stevens, M. N. Nityasya, M. F. Adilazuarda, R. Ignatius, R. Diandaru, V. Ghifari, T. Yu, W. Dai, Y. Xu, D. Damapuspita, H. A. Wibowo, C. Tho, I. M. K. Karo, T. N. Fatyanosa, Z. Ji, G. Neubig, T. Baldwin, S. Ruder, P. Fung, H. Sujaini, S. Sakti, A. Purwarianti, and M. Authors, "Nusacrowd: Open source initiative for indonesian nlp resources." [Online]. Available: https://indonlp.github.io/

Implementasi Augmentasi Audio..., Raphael Gregorius Hakim, Universitas Multimedia Nusantara

[29] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[30] B. H. Juang and L. R. Rabiner, "Automatic speech recognition-a brief history of the technology development," 2004. [Online]. Available: http://www.recording-history.org/

[31] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel, and Y. Bengio, "End-to-end attention-based large vocabulary speech recognition," *arXiv preprint arXiv:1508.04395*, 2016.

[32] D. Yu and L. Deng, *Automatic Speech Recognition: A Deep Learning Approach*. Springer, 2015.

[33] S. Benkerzaz, Y. Elmir, and A. Dennai, "A study on automatic speech recognition," 2019. [Online]. Available: https://www.researchgate.net/publication/337155654_A_Study_on_Automatic_Speech_Recognition

[34] C. V. Budijanto, "Sistem transkripsi percakapan bahasa indonesia dengan metode pengenalan suara bahasa sumber daya rendah menggunakan wav2vec2," 2022.

[35] N. Jain and S. Rastogi, "Speech recognition systems – a comprehensive study of concepts and mechanism," *Acta Informatica Malaysia*, vol. 3, pp. 01–03, 1 2019.

[36] P. Favaro, "Beyond supervised learning," 2018. [Online]. Available: https://consensus.app/papers/beyond-supervised-learning-favaro/b298dfdc9fdf53aea97908f6a1230aa9/?utm_source=chatgpt

[37] E. Zheltonozhskii, C. Baskin, A. M. Bronstein, and A. Mendelson, "Self-supervised learning for large-scale unsupervised image clustering," *ArXiv*, vol. abs/2008.10312, 2020. [Online]. Available: https://consensus.app/papers/selfsupervised-learning-largescale-unsupervised-image-zheltonozhskii/8cb146c728245a499efaee3dbc3a533e/?utm_source=chatgpt

[38] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp. 4037–4058, 2019. [Online]. Available: https://consensus.app/papers/selfsupervised-feature-learning-with-deep-neural-jing/b9a00f9d790154a395bf4a9ad1915e4d/?utm_source=chatgpt

[39] A. Saeed, T. Ozcelebi, and J. Lukkien, "Multi-task self-supervised learning for human activity detection," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, pp. 1–30, 2019. [Online]. Available: https://consensus.

app/papers/multitask-selfsupervised-learning-human-activity-saeed/
0ff064dec10d5297b49c6efd0a5cf16d/?utm_source=chatgpt

[40] G.-J. Qi and J. Luo, "Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, pp. 2168–2187, 2019. [Online]. Available: https://consensus.app/papers/small-data-challenges-data-survey-recent-progress-qi/9f38c21e4df0508c8b01bf9b7f2695ed/?utm_source=chatgpt

[41] D. Greene, P. Cunningham, and R. Mayer, "Unsupervised learning and clustering," in *Machine Learning Techniques for Multimedia*. Springer, 2008, pp. 51–90. [Online]. Available: https://consensus.app/papers/unsupervised-learning-clustering-greene/11c3012d28c5506f9a7939308c58c3af/?utm_source=chatgpt

[42] C. Reed, S. L. Metzger, A. Srinivas, T. Darrell, and K. Keutzer, "Selfaugment: Automatic augmentation policies for self-supervised learning," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2673–2682, 2020. [Online]. Available: https://consensus.app/papers/selfaugment-automatic-augmentation-policies-reed/70c6f6098b08566a9598a8d7ba856934/?utm_source=chatgpt

[43] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.

[44] A. Abdusalomov, F. Safarov, M. Rakhimov, B. Turaev, and T. Whangbo, "Improved feature parameter extraction from speech signals using machine learning algorithm," *Sensors (Basel, Switzerland)*, vol. 22, 2022.

[45] A. Aggarwal, A. Srivastava, A. Agarwal, N. Chahal, D. Singh, A. Alnuaim, A. Alhadlaq, and H.-N. Lee, "Two-way feature extraction for speech emotion recognition using deep learning," *Sensors (Basel, Switzerland)*, vol. 22, 2022.

[46] A. Tamara, A. Rusli, and S. Hansun, "Utilizing e-learning with pattern recognition and mnemonic to enhance japanese characters learning experience," *2019 International Conference on Engineering, Science, and Industrial Applications, ICESI 2019*, 8 2019.

[47] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Łukasz Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 2017-December, pp. 5999–6009, 6 2017. [Online]. Available: https://arxiv.org/abs/1706.03762v7

[48] H. Li and W. Wang, "Reinterpreting ctc training as iterative fitting," 2020. [Online]. Available: https://github.com/hzli-ucas/caffe/tree/ctc.

[49] K. Sodimana, K. Pipatsrisawat, L. Ha, M. Jansche, O. Kjartansson, P. D. Silva, and S. Sarin, "A Step-by-Step Process for Building TTS Voices Using Open Source Data and Framework for Bangla, Javanese, Khmer, Nepali, Sinhala, and Sundanese," in *Proc. The 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU)*, Gurugram, India, Aug. 2018.

[50] "Streamlit." [Online]. Available: https:// phrasenmaeher-audio-transformat-visualize-transformation-5s1n4t.streamlit. app/

[51] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.

[52] P. Gamper, I. Korshunova, and T. Kemp, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.

[53] M. Müller, F. Kurth, and M. Clausen, "Audio-based music structure analysis," in *Computer Music Modeling and Retrieval*, ser. Lecture Notes in Computer Science, vol. 6684. Springer, 2011, pp. 138–149.

[54] T. V. Neumann, C. Boeddeker, K. Kinoshita, M. Delcroix, and R. Haeb-Umbach, "On word error rate definitions and their efficient computation for multi-speaker speech recognition systems." [Online]. Available: https: //github.com/fgnt/meeteval

[55] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12 449–12 460, 2020.