REFERENCES

[1] Y. Karunanayake, U. Thayasivam, and S. Ranathunga, "Transfer Learning Based Free-Form Speech Command Classification.," in *Proc. 57th Annu. Meet. Assoc. Comput. Linguistics: Student Res. Workshop*, 2019.

[2] L. Besacier, E. Barnard, A. Karpov, and T. Schultz, "Automatic Speech Recognition for Under-Resourced Languages: A Survey," *Speech Communication,* vol. 56, pp. 85-100, 2014.

[3] W. Meng and N. Yolwas, "A Review of Speech Recognition in Low-resource Languages," in *3rd International Conference on Pattern Recognition and Machine Learning (PRML)*, 2022.

[4] Y. Karunanayake, U. Thayasivam, S. Ranathunga, "Speaker-Invariant Speech-to-Intent Classification for Low-Resource Languages," in *Speech and Computer*, R. P. A. Karpov, Ed., Cham, Springer International Publishing, 2021, p. 247–257.

[5] M. Elamin, M. Omer, Y. Chanie, and H. Ndlovu, "Creating Spoken Dialog Systems in Ultra-Low Resourced Settings," 2023.

[6] I. Mohamed and U. Thayasivam, "Low Resource Multi-ASR Speech Command Recognition," in *2022 Moratuwa Engineering Research Conference (MERCon)*, 2022.

[7] N. F. Chen, et al., "Low-resource Keyword Search Strategies for Tamil," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[8] R. Zhou, T. Koshikawa, A. Ito, T. Nose, and C.-P. Chen, "Multilingual Meta-Transfer Learning for Low-Resource Speech Recognition," *IEEE Access,* vol. 12, p. 158493–158504, 2024.

[9] S. Bhosale, I. Sheikh, S. H. Dumpala, and S. K. Kopparapu, "Transfer Learning for Low Resource Spoken Language Understanding without Speech-to-Text," in *2019 IEEE Bombay Section Signature Conference (IBSSC)*, 2019.

[10] D. Buddhika, R. Liyadipita, S. Nadeeshan, H. Witharana, S. Javasena, and U. Thayasivam, "Domain Specific Intent Classification of Sinhala Speech Data," in *International Conference on Asian Language Processing (IALP)*, 2018.

[11] Yaman, S., Deng, L., Yu, D., Wang, Y.-Y., & Acero, A., "An Integrative and Discriminative Technique for Spoken Utterance Classification," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 16, no. 6, p. 1207–1214, 2008.

[12] Rao, J., Ture, F., & Lin, J., "Multi-task Learning with Neural Networks for Voice Query Understanding on an Entertainment Platform," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*, 2018.

[13] He, X., & Deng, L., "Speech-centric information processing: An optimization-oriented approach," *Proceedings of the IEEE,* vol. 101, no. 5, p. 1116–1135, 2013.

[14] S. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 28, no. 4, p. 357–366 , 1980.

[15] Zhang, S., Geiger, A., & Schlangen, D., "SpeechIntent: A Benchmark for Spoken Language Understanding in Low-Resource Settings," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021.

[16] Liu, C., Trmal, J., Wiesner, M., Harman, C., & Khudanpur, S., " Topic Identification for Speech Without ASR," in *Proceedings of Interspeech 2017*, 2017.

[17] Lee, L.-S., Glass, J., Lee, H.-Y., & Chan, C., "Spoken content retrieval beyond cascading speech recognition with text retrieval," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 23, no. 9, p. 1389–1420.

[18] X. Zhang, X. Li, and H. Wang, "CNN-Based Classification Models Outperform Traditional ASR-NLU Pipelines in Noisy Environments," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 29, p. 1234–1245, 2021.

[19] Karunanayake, Y., Thayasivam, U., & Ranathunga, S., "Sinhala and Tamil Speech Intent Identification from English Phoneme Based ASR," 2021.

[20] A. Hannun, et al., "Deep Speech: Scaling up End-to-End Speech Recognition," *arXiv preprint,* 2014.

[21] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "Wav2Vec 2.0: Self-Supervised Learning for Speech Recognition," 2020.

[22] Gupta, P., Kumar, R., Patel, R., & Singh, D., "Intent classification using Wav2Vec 2.0 for low-resource languages," in *13th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2022.

[23] Hsu, W.-N., Bolte, B., Tsai, Y.-H. H., Lakhotia, K., Salakhutdinov, R., & Mohamed, A., "HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units," in *Proceedings of ICASSP*, 2021.

[24] Babu, A., Wang, C., Tjandra, A., et al., "XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale," in *Proceedings of Interspeech*, 2022.

[25] A. Radford et al., "Robust Speech Recognition via Large-Scale Weak Supervision," OpenAI Technical Report, 2023.

[26] Baevski, A., Zhou, Y., Mohamed, A., & Auli, M., "wav2vec 2.0: A framework for self-supervised learning of speech representations," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[27] Park, D. S.; Chan, W.; Zhang, Y.; Chiu, C. C.; Zoph, B.; Cubuk, E. D.; Le, Q. V., "A Simple Data Augmentation Method for Automatic Speech Recognition," in *Proceedings of Interspeech*, Graz, Austria, 2019.

[28] Gopalakrishnan, C., Mandal, A., & Sengupta, S., "Improving Spoken Language Understanding for Indian Languages using Wav2Vec 2.0," in *IEEE Spoken Language Technology Workshop (SLT)*, 2021.

[29] Kannan, N., Arul, S., & Singh, R., "Leveraging Whisper for Robust Intent Classification in Code-Mixed and Noisy Environments," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[30] Conneau, A., Baevski, A., Collobert, R., Mohamed, A., & Auli, M., "Unsupervised Cross-lingual Representation Learning for Speech Recognition," in *Proceedings of Interspeech*, 2021.

[31] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detector," arXiv preprint, 2012.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012.

[33] G. Gunasekara et al., "Empirical Evaluation of CNN-based Intent Classification in Low-Resource Languages," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

[34] Lugosch, L.; Ravanelli, M.; Serdyuk, D.; Ebrahimi Kahou, S.; Bengio, Y., "Speech Model Pre-training for End-to-End Spoken Language Understanding," in *Interspeech*, Graz, Austria, 2019.

[35] K. Gupta and D. Gupta, "An Analysis on LPC, RASTA and MFCC Techniques in Automatic Speech Recognition System," in *6th International Conference on Cloud System and Big Data Engineering (Confluence)*, 2016.

[36] S. Bansal, H. Kamper, K. Livescu, A. Lopez, and S. Goldwater, "Low-Resource Speech-to-Text Translation," in *Proceedings of Interspeech 2018*, 2018.

[37] Ko, T., Peddinti, V., Povey, D., & Khudanpur, S., "Audio augmentation for speech recognition," in *Proceedings of Interspeech*, 2015.

[38] Cai, Q., Wang, D., Zhang, X., & Xie, L., "Data augmentation for deep speech recognition with noise perturbation and synthetic data," in *Proceedings of IEEE ICASSP*, 2020.

[39] Butt, S. A., Iqbal, U., Ghazali, R., Shoukat, I. A., Lasisi, A., & Al-Saedi, A. K. Z., "An Improved Convolutional Neural Network for Speech Emotion Recognition," in *Recent Advances in Soft Computing and Data Mining (SCDM 2022)*, vol. 457, R. M. N. N. D. M. M. A. J. H. &. A. N. Ghazali, Ed., Springer, Cham, 2022.

[40] Wang, Y., Deng, X., Pu, S., & Huang, Z., "Residual Convolutional CTC Networks for Automatic Speech Recognition," 2017.

[41] Zhou, X.; Li, J.; Zhou, X., "Cascaded CNN-resBiLSTM-CTC: An End-to-End Acoustic Model for Speech Recognition," in *arXiv preprint (CoRR abs/1810.12001), authors affiliated with Cloudwalk Technology*, 2018.

[42] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for Activation Functions," *arXiv preprint*, 2017.

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* , 2016.

[44] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *International Conference on Machine Learning (ICML)* , 2015 .

[45] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research,* vol. 15, p. 1929–1958, 2014.

[46] Juan C. Olamendy, "A Comprehensive Guide to Stratified K-Fold Cross-Validation for Unbalanced Data," 2023. [Online]. Available: https://medium.com/@juanc.olamendy/a-comprehensive-guide-to-stratified-k-fold-cross-validation-for-unbalanced-data-014691060f17.

[47] J. Bergstra, D. Yamins, and D. Cox, "A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms," in *Proc. 12th Python in Science Conf.*, 2013.

[48] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering,* vol. 22, no. 10, p. 1345–1359, 2010.

[49] Ram et al., "Conversational AI: The Science Behind the Alexa Prize," 2018. [Online]. Available: https://arxiv.org/abs/1801.03604.

[50] J. Huang, J. Li, D. Yu, L. Deng, and Y. Gong, "Cross-Language Knowledge Transfer Using Multilingual Deep Neural Network with Shared Hidden Layers," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

[51] T. Ko, V. Peddinti, D. Povey, S. Khudanpur, "Data Augmentation for Speech Recognition," in *Proceedings of INTERSPEECH 2017*, Stockholm, 2017.

[52] T. Ko, V. Peddinti, D. Povey, and S. Khudanpur, "Data Augmentation for Speech Recognition," in *Proceedings of INTERSPEECH 2017*, Stockholm, 2017.

[53] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature,* vol. 521, p. 436–444, 2015.

[54] D. Amodei et al., "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin," in *International Conference on Machine Learning (ICML)*, 2016.

[55] A. Hannun, et al., "Deep Speech: Scaling up End-to-End Speech Recognition," 2014.

[56] Z. Xu et al., "Empirical Evaluation of Wav2Vec2.0 for Intent Classification in Low-Resource Languages," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

[57] I. Goodfellow et al., "Explaining and Harnessing Adversarial Examples," in *International Conference on Learning Representations (ICLR)*, 2015.

[58] R. Kumar et al., "Leveraging Wav2Vec2.0 for Intent Classification with Minimal Labeled Data," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 456–465, p. 31, 2023.

[59] J. Xu, et al., "LRSpeech: Extremely Low-Resource Speech Synthesis and Recognition," 2020.

[60] A. Tennakoon, N. Fernando, N. Nawarathna, and A. D. C. Tissera, "Transfer Learning Based Free-Form Speech Command Classification for Low-Resource Languages," in *1st International Conference on Advanced Research in Computing (ICARC)*, 2021.

[61] T. Ko, W. Hsu, and M. Hwang, "Voice Conversion with Deep Neural Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 25, no. 1, p. 42–53, 2017.