DAFTAR PUSTAKA

Abdel-Hamid, O., Mohamed, A., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional Neural Networks for Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(10), 1533–1545.

Agga, A., Abbou, A., Labbadi, M., Houm, Y. El, & Ou Ali, I. H. (2022). CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. *Electric Power Systems Research*, *208*, 107908.

Ando, Y. (2019). Chapter 7 - Temporal and Spatial Primary Percepts of Sound and the Sound Field. In Y. Ando (Ed.), *Signal Processing in Auditory Neuroscience* (hlm. 51–66). Academic Press.

Ariyanti, W., Hussain, T., Wang, J.-C., Wang, C.-T., Fang, S.-H., & Tsao, Y. (2021). Ensemble and Multimodal Learning for Pathological Voice Classification. *IEEE Sensors Letters*, *5*(7), 1–4.

Ayache, M., Kanaan, H., Kassir, K., & Kassir, Y. (2021). Speech Command Recognition Using Deep Learning. *2021 Sixth International Conference on Advances in Biomedical Engineering (ICABME)*, 24–29.

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Fissore, L., Laface, P., Mertins, A., Ris, C., Rose, R., Tyagi, V., & Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication*, *49*(10), 763–786.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag.

Cantiabela, Z., Pardede, H. F., Zilvan, V., Sulandari, W., Yuwana, R. S., Supianto, A. A., & Krisnandi, D. (2023). Deep Learning for Robust Speech Command Recognition Using Convolutional Neural Networks (CNN). *Proceedings of the 2022 International Conference on Computer, Control, Informatics and Its Applications*, 101–105.

Cho, K., van Merrienboer, B., Gülçehre, Ç., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *CoRR*, *abs/1406.1*. http://arxiv.org/abs/1406.1078

Chu, W.-S., la Torre, F., & Cohn, J. F. (2017). Learning Spatial and Temporal Cues for Multi-Label Facial Action Unit Detection. *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 25–32.

Das, N., Chakraborty, S., Chaki, J., Padhy, N., & Dey, N. (2021). Fundamentals, present and future perspectives of speech enhancement. *International Journal of Speech Technology*, *24*(4), 883–901.

Deng, L., & Li, X. (2013). Machine Learning Paradigms for Speech Recognition: An Overview. *IEEE Transactions on Audio, Speech, and Language Processing*, *21*(5), 1060–1089.

Dongare, A. D., Kharde, R. R., & Kachare, A. D. (2012). Introduction to Artificial Neural Network (ANN) Methods. *International Journal of Engineering and Innovative Technology (IJEIT)*, *2*(1), 189–194.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org

Guo, Y., Xiong, X., Liu, Y., Xu, L., & Li, Q. (2022). A novel speech emotion recognition method based on feature construction and ensemble learning. *PLOS ONE*, *17*(8), 1–17.

Hahnioser, R. H. R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, *405*(6789), 947–951.

Halliday, D., Resnick, R., & Walker, J. (2013). *Fundamentals of Physics Extended, 10th Edition*. Wiley. https://books.google.co.id/books?id=DTccAAAAQBAJ

Haykin, S. (2009). *Neural Networks and Learning Machines* (Third). Pearson Education.

Hochreiter, S., & Schimdhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *8*(9), 1735–1780.

Hu, Y., & Lu, X. (2018). Learning spatial-temporal features for video copy detection by the combination of CNN and RNN. *Journal of Visual Communication and Image Representation*, *55*, 21–29.

Huang, X., Acero, A., Hon, H.-W., & Reddy, R. (2001). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development* (1st ed.). Prentice Hall PTR.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 4–37.

Johnson, D. H. (2006). Signal-to-noise ratio. *Scholarpedia*, *1*(12), 2088.

Kim, C., Kumar, K., & Stern, R. M. (2009). Robust speech recognition using a Small Power Boosting algorithm. *2009 IEEE Workshop on Automatic Speech Recognition & Understanding*, 243–248.

Kim, C., & Stern, R. M. (2008). Robust signal-to-noise ratio estimation based on waveform amplitude distribution analysis. *Proc. Interspeech 2008*, 2598–2601.

Kim, C., & Stern, R. M. (2009). Feature extraction for robust speech recognition using a power-law nonlinearity and power-bias subtraction. *INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association, Brighton, United Kingdom, September 6-10, 2009*, 28–31. http://www.isca-speech.org/archive/interspeech%5C_2009/i09%5C_0028.html

Kim, C., & Stern, R. M. (2010). Feature extraction for robust speech recognition based on maximizing the sharpness of the power distribution and on power flooring. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 4574–4577.

Kim, C., & Stern, R. M. (2016). Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *24*(7), 1315–1329.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Li, J., Deng, L., Gong, Y., & Haeb-Umbach, R. (2014). An Overview of Noise-Robust Automatic Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(4), 745–777.

Li, J., Deng, L., Haeb-Umbach, R., & Gong, Y. (2016). Chapter 1 - Introduction. In J. Li, L. Deng, R. Haeb-Umbach, & Y. Gong (Ed.), *Robust Automatic Speech Recognition* (hlm. 1–7). Academic Press.

Li, X., & Zhou, Z. (2016). Speech Command Recognition with Convolutional Neural Network. *Stanford Journal*.

Lim, J. (1979). Spectral root homomorphic deconvolution system. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *27*(3), 223–233.

Lockwood, P., & Alexandre, P. (1994). Root adaptive homomorphic deconvolution schemes for speech recognition in noise. *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*, *i*, I/441-I/444 vol.1.

Majumdar, S., & Ginsburg, B. (2020). MatchboxNet: 1D time-channel separable convolutional neural network architecture for speech commands recognition. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2020-Octob*, 3356–3360.

Martinez, J., Perez, H., Escamilla, E., & Suzuki, M. M. (2012). Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques. *CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers*, 248–251.

Naiborhu, A. P. F., & Endah, S. N. (2021). Indonesian Continuous Speech Recognition Using CNN and Bidirectional LSTM. *2021 5th International Conference on Informatics and Computational Sciences (ICICoS)*, 122–127.

Oppenheim, A. V, Willsky, A. S., Nawab, S. H., & Ding, J.-J. (1997). *Signals and systems* (Vol. 2). Prentice hall Upper Saddle River, NJ.

Orfanidis, S. J. (2023). *Introduction to Signal Processing, Second Edition*. Prentice-Hall, Inc.

Pardede, H. F. (2016). On noise robust feature for speech recognition based on power function family. *2015 International Symposium on Intelligent Signal Processing and Communication Systems, ISPACS 2015*, 386–391.

Pardede, H. F., Yuliani, A. R., & Sustika, R. (2018). Convolutional Neural Network and Feature Transformation for Distant Speech Recognition. *International Journal of Electrical and Computer Engineering (IJECE)*, *8*(6), 5381.

Pardede, H. F., Zilvan, V., Krisnandi, D., Heryana, A., & Kusumo, R. B. S. (2019). Generalized Filter-bank Features for Robust Speech Recognition Against Reverberation. *2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, 19–24.

Patterson, R. D., Robinson, K., Holdsworth, J. L., McKeown, D., Zhang, C. Q., & Allerhand, M. (1992). *Complex Sounds and Auditory Images*.

Powers, D. (2008). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *Mach. Learn. Technol.*, *2*.

Proakis, J. G., & Manolakis, D. G. (1996). *Digital Signal Processing (3rd Ed.): Principles, Algorithms, and Applications*. Prentice-Hall, Inc.

Raschka, S., & Mirjalili, V. (2019). *Python Machine Learning, 3rd Ed.* (3 ed.). Packt Publishing.

Rocchesso, D. (2003). *Introduction to sound processing*. Mondo estremo.

Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, *2*(6), 420.

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*(1), 77–89.

Wang, Y., & Nishizaki, H. (2022). *Combination of Time-domain, Frequency-*

domain, and Cepstral-domain Acoustic Features for Speech Commands Classification. arXiv.

Warden, P. (2018). Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition. *CoRR*, *abs/1804.0*. http://arxiv.org/abs/1804.03209

Yang, X., Yu, H., & Jia, L. (2020). Speech recognition of command words based on convolutional neural network. *Proceedings - 2020 International Conference on Computer Information and Big Data Applications, CIBDA 2020*, 465–469.

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J., & Li, Z. (2018). Deep Multi-View Spatial-Temporal Network for Taxi Demand Prediction. *CoRR*, *abs/1802.0*. http://arxiv.org/abs/1802.08714

Yuliani, A. R., Sustika, R., Yuwana, R. S., & Pardede, H. F. (2017). Feature transformations for robust speech recognition in reverberant conditions. *2017 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, 57–62.

Zhang, H., Wu, D., Wang, Z., Chen, Y., & Xu, J. (2021). Remaining Useful Life Prediction of Lithium Battery via Neural Network Ensemble. *2021 IEEE International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2021*, 392–397.

Zhang, J., Wang, P., Yan, R., & Gao, R. X. (2018). Long short-term memory for machine remaining life prediction. *Journal of Manufacturing Systems*, *48*(December 2017), 78–86.

Zhang, Z., Geiger, J., Pohjalainen, J., Mousa, A. E.-D., Jin, W., & Schuller, B. (2018). Deep Learning for Environmentally Robust Speech Recognition: An Overview of Recent Developments. *ACM Trans. Intell. Syst. Technol.*, *9*(5).