DAFTAR PUSTAKA

Abdelhamid, A. A., El-Kenawy, E.-S. M., Alotaibi, B., Amer, G. M., Abdelkader, M. Y., Ibrahim, A., & Eid, M. M. (2022). Robust Speech Emotion Recognition Using CNN+LSTM Based on Stochastic Fractal Search Optimization Algorithm. *IEEE Access*, *10*, 49265–49284. https://doi.org/10.1109/ACCESS.2022.3172954

Abdel-Hamid, L. (2020). Egyptian Arabic speech emotion recognition using prosodic, spectral and wavelet features. *Speech Communication*, *122*, 19–30. https://doi.org/https://doi.org/10.1016/j.specom.2020.04.005

Aini, Y. K., Santoso, T. B., & Dutono, D. T. (2021). Pemodelan CNN Untuk Deteksi Emosi Berbasis Speech Bahasa Indonesia. *Jurnal Komputer Terapan*, *7*(1), 143–152. https://jurnal.pcr.ac.id/index.php/jkt/

Alluhaidan, A. S., Saidani, O., Jahangir, R., Nauman, M. A., & Neffati, O. S. (2023). Speech Emotion Recognition through Hybrid Features and Convolutional Neural Network. *Applied Sciences (Switzerland)*, *13*(8). https://doi.org/10.3390/app13084750

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00444-8

Ambar. (2017, November 27). *10 Fungsi Komunikasi Suara Secara Umum*. PakarKomunikasi.Com. https://pakarkomunikasi.com/fungsi-komunikasi-audio

Anagnostopoulos, C. N., Iliou, T., & Giannoukos, I. (2015). Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. *Artificial Intelligence Review*, *43*(2), 155–177. https://doi.org/10.1007/S10462-012-9368-5/METRICS

Berenzweig, A., Logan, B., Ellis, D. P. W., & Whitman, B. (2004). *A Large-Scale Evaluation of Acoustic and Subjective Music-Similarity Measures* (Vol. 28, Issue 2).

Burnwal, S. (2020). *Speech Emotion Recognition*. https://www.kaggle.com/code/shivamburnwal/speech-emotion-recognition

Bustamin, A., Rizky, A. M., Warni, E., Areni, I. S., & Indrabayu, I. (2024). IndoWaveSentiment: Indonesian Audio Dataset for Emotion Classification. In *Mendeley Data*. https://doi.org/10.17632/j9ytfdzy27.1

Dafit. (2023, October 11). *Teknologi Komunikasi: Dinamika Interaksi Sosial Online di Era Digital*. Kompasiana.Com. https://www.kompasiana.com/fitriawardani8538/6512d72e08a8b53bef550792/teknologi-komunikasi-dinamika-interaksi-sosial-online-di-era-digital

Das, J. K., Ghosh, A., Pal, A. K., Dutta, S., & Chakrabarty, A. (2020, October 21). Urban Sound Classification Using Convolutional Neural Network and Long Short Term Memory Based on Multiple Features. *4th International Conference on Intelligent Computing in Data Sciences, ICDS 2020*. https://doi.org/10.1109/ICDS50568.2020.9268723

Défossez, A. (2021). *Hybrid Spectrogram and Waveform Source Separation*. http://arxiv.org/abs/2111.03600

Ellis, D. P. W. (2007a). *Chroma Feature Analysis and Synthesis*. https://www.ee.columbia.edu/~dpwe/resources/matlab/chroma-ansyn/

Ellis, D. P. W. (2007b). Classifying Music Audio with Timbral and Chroma Features. *International Society for Music Information Retrieval Conference*. https://api.semanticscholar.org/CorpusID:802986

Ellis, D. P. W., & Poliner, G. E. (2007). Identifying `Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking. *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, *4*, IV-1429-IV–1432. https://doi.org/10.1109/ICASSP.2007.367348

Fahrudin, T. M., Laksana Aryananda, R., Gunawan, E. L., Belardo, V., Marcelia, F., & Halim, C. (2022). Analisis Video Keluhan Pelanggan Menggunakan Automatic Speech Recognition dan Analisis Polaritas Sentimen Video Analysis of Customer Complaints Using Automatic Speech Recognition and Sentiment Polarity Analysis. *Digital Business Intelligence, and Computer Engineering)*, *1*(1), 14–21.

Farhan. (2024, May 28). *Normalisasi dan Feature Scaling*. https://medium.com/@farhanopen/normalisasi-dan-feature-scaling-503deb9c0d3b

Frant, E., Ispas, I., Dragomir, V., Dascalu, M., Zoltan, E., & Stoica, I. C. (2017). Voice Based Emotion Recognition with Convolutional Neural Networks for Companion Robots. *ROMANIAN JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY*, *20*(3), 222–240.

George, S. M., & Muhamed Ilyas, P. (2024). A review on speech emotion recognition: A survey, recent advances, challenges, and the influence of noise. In *Neurocomputing* (Vol. 568). Elsevier B.V. https://doi.org/10.1016/j.neucom.2023.127015

Giannakopoulos, T., & Pikrakis, A. (2014). Chapter 4 - Audio Features. In T. Giannakopoulos & A. Pikrakis (Eds.), *Introduction to Audio Analysis* (pp. 59–103). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-08-099388-1.00004-2

Gorbunov, M. (2020). *The Influence of The Signal-to-Noise Ratio upon Radio Occultation Inversion Quality*. https://doi.org/10.5194/amt-2020-114

Hamid, M. A. (2023). *Speech emotion recognition 97.25% accuracy*. https://www.kaggle.com/code/mostafaabdlhamed/speech-emotion-recognition-97-25-accuracy/comments

Jackson, P., & ul haq, S. (2011). *Surrey Audio-Visual Expressed Emotion (SAVEE) database*.

John, G. (2024, June 3). *Apa itu Sample Rate dalam Musik Digital?* https://erudisi.com/apa-itu-sample-rate-dalam-musik-digital/

Kamus Besar Bahasa Indonesia (KBBI). (n.d.). *Analisis*. Retrieved May 17, 2024, from https://kbbi.kemdikbud.go.id/entri/analisis

Lambamo, W., Srinivasagan, R., & Jifara, W. (2023). Analyzing Noise Robustness of Cochleogram and Mel Spectrogram Features in Deep Learning Based Speaker Recognition. *Applied Sciences (Switzerland)*, *13*(1). https://doi.org/10.3390/app13010569

Leiber, M., Marnissi, Y., Barrau, A., & Badaoui, M. El. (2023). *Differentiable adaptive short-time Fourier transform with respect to the window length*. https://doi.org/10.1109/ICASSP49357.2023.10095245

Li, L.-Q., Xie, K., Guo, X.-L., Wen, C., & He, J.-B. (2022). Emotion recognition from speech with StarGAN and Dense-DCNN. *IET Signal Processing*, *16*(1), 62–79. https://doi.org/https://doi.org/10.1049/sil2.12078

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*. https://doi.org/10.5281/zenodo.1188976

Madanian, S., Chen, T., Adeleye, O., Templeton, J. M., Poellabauer, C., Parry, D., & Schneider, S. L. (2023). Speech emotion recognition using machine learning — A systematic review. *Intelligent Systems with Applications*, *20*. https://doi.org/10.1016/j.iswa.2023.200266

Mashhadi, M. M. R., & Osei-Bonsu, K. (2023). Speech emotion recognition using machine learning techniques: Feature extraction and comparison of convolutional neural network and random forest. *PLoS ONE*, *18*(11 November). https://doi.org/10.1371/journal.pone.0291500

Mazumder, M. A., & Salam, R. A. (2019). Feature extraction techniques for speech processing: A review. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(1.3 S1), 285–292. https://doi.org/10.30534/ijatcse/2019/5481.32019

McFee, B., Raffel, C., Liang, D., Ellis, D., Mcvicar, M., Battenberg, E., & Nieto, O. (2015). *librosa: Audio and Music Signal Analysis in Python*. https://doi.org/10.25080/Majora-7b98e3ed-003

More, Ms. Y. A. (2016). Effect Of Combination Of Different Features On Speech Recognition For Abnormal Speech. *International Journal Of Engineering And Computer Science*. https://doi.org/10.18535/ijecs/v5i8.28

Muller, M. (2021). *Fundamentals of Music Processing*.

Mustaqeem, & Kwon, S. (2020). A CNN-assisted enhanced audio signal processing for speech emotion recognition. *Sensors (Switzerland)*, *20*(1). https://doi.org/10.3390/s20010183

Nanos, G. (2024, March 18). *Neural Networks*. https://www.baeldung.com/cs/neural-networks-pooling-layers

Nantasri, P., Phaisangittisagul, E., Karnjana, J., Boonkla, S., Keerativittayanun, S., Rugchatjaroen, A., Usanavasin, S., & Shinozaki, T. (2020). A Light-Weight Artificial Neural Network for Speech Emotion Recognition using Average Values of MFCCs and Their Derivatives. *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and*

Information Technology (ECTI-CON), 41–44. https://doi.org/10.1109/ECTI-CON49241.2020.9158221

Nielsen. (2021, November 29). *Beyond martech: membangun kepercayaan dengan konsumen dan terlibat di mana sentimen tinggi*. Nielsen. https://www.nielsen.com/id/insights/2021/beyond-martech-building-trust-with-consumers-and-engaging-where-sentiment-is-high/

O'Shaughnessy, D. (2023). Review of analysis methods for speech applications. *Speech Communication*, *151*, 64–75. https://doi.org/10.1016/j.specom.2023.05.008

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*(4), 417–435. https://doi.org/10.1016/j.wocn.2009.07.005

Pichora-Fuller, M. K., & Dupuis, K. (2020). *Toronto emotional speech set (TESS)* (DRAFT VERSION). Borealis. https://doi.org/doi/10.5683/SP2/E8H2MF

Pickle, B. (2022, December 23). *Wafeform Definition*. https://techterms.com/definition/waveform

Piczak, K. J. (2015). ESC: Dataset for environmental sound classification. *MM 2015 - Proceedings of the 2015 ACM Multimedia Conference*, 1015–1018. https://doi.org/10.1145/2733373.2806390

Plutchik, R. (2001a). The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, *89*(4), 344–350. http://www.jstor.org/stable/27857503

Plutchik, R. (2001b). The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, *89*(4), 344–350. http://www.jstor.org/stable/27857503

Ponton, J. W. (2007). *Frequency Response*. https://www.homepages.ed.ac.uk/jwp/control06/controlcourse/restricted/course/advanced/index.html

Rayhan Ahmed, Md., Islam, S., Muzahidul Islam, A. K. M., & Shatabda, S. (2023). An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition. *Expert Systems with Applications*, *218*, 119633. https://doi.org/https://doi.org/10.1016/j.eswa.2023.119633

Russell, J. A., & Barrett, L. F. (1999a). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Russell, J. A., & Barrett, L. F. (1999b). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Saini, H. (2019, September 21). *What is Sound and How do we Hear it?* https://letstalkscience.ca/educational-resources/backgrounders/what-sound-and-how-do-we-hear-it

Scherer, D., Müller, A., & Behnke, S. (2010). *Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition*. http://www.ais.uni-bonn.de

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1–2), 227–256. www.elsevier.com/locate/specom

Scherer, K. R. (2005a). What are emotions? and how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Scherer, K. R. (2005b). What are emotions? and how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Shah, A., Kattel, M., Nepal, A., & Shrestha, D. (2019). *Chroma Feature Extraction*.

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

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerry-Ryan, R., Saurous, R. A., Agiomyrgiannakis, Y., & Wu, Y. (2017). *Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions*. http://arxiv.org/abs/1712.05884

Suarez-Perez, A., Gabriel, G., Rebollo, B., Illa, X., Guimerà-Brunet, A., Hernández-Ferrer, J., Martínez, M. T., Villa, R., & Sanchez-Vives, M. V. (2018). Quantification of signal-to-noise ratio in cerebral cortex recordings using flexible MEAs with co-localized platinum black, carbon nanotubes, and gold electrodes. *Frontiers in Neuroscience*, *12*(NOV). https://doi.org/10.3389/fnins.2018.00862

Taye, M. M. (2023). Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions. In *Computation* (Vol. 11, Issue 3). MDPI. https://doi.org/10.3390/computation11030052

Torres-García, A. A., Mendoza-Montoya, O., Molinas, M., Antelis, J. M., Moctezuma, L. A., & Hernández-Del-Toro, T. (2022). Chapter 4 - Pre-processing and feature extraction. In A. A. Torres-García, C. A. Reyes-García, L. Villaseñor-Pineda, & O. Mendoza-Montoya (Eds.), *Biosignal Processing and Classification Using Computational Learning and Intelligence* (pp. 59–91). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-820125-1.00014-2

Turab, M., Kumar, T., Bendechache, M., & Saber, T. (2022). *Investigating Multi-Feature Selection and Ensembling for Audio Classification*. http://arxiv.org/abs/2206.07511

Vásquez-Correa, J. C., García, N., Orozco-Arroyave, J. R., Arias-Londoño, J. D., Vargas-Bonilla, J. F., & Nöth, E. (2015). *Emotion Recognition from Speech under Environmental Noise Conditions using Wavelet Decomposition*.

Wang, F., & Shen, X. (2023). Research on Speech Emotion Recognition Based on Teager Energy Operator Coefficients and Inverted MFCC Feature Fusion. *Electronics (Switzerland)*, *12*(17). https://doi.org/10.3390/electronics12173599

Wei, S., Zou, S., Liao, F., & Lang, W. (2020). A Comparison on Data Augmentation Methods Based on Deep Learning for Audio Classification. *Journal of Physics: Conference Series*, *1453*(1). https://doi.org/10.1088/1742-6596/1453/1/012085

Widiyanti, E., & Endah, S. N. (2018). Feature Selection for Music Emotion Recognition. *2018 2nd International Conference on Informatics and Computational Sciences (ICICoS)*, 1–5. https://doi.org/10.1109/ICICOS.2018.8621783

Wijayasingha, L., & Stankovic, J. A. (2021). Robustness to noise for speech emotion classification using CNNs and attention mechanisms. *Smart Health*, *19*, 100165. https://doi.org/https://doi.org/10.1016/j.smhl.2020.100165

Xu, M., Zhang, F., & Khan, S. U. (2020). Improve Accuracy of Speech Emotion Recognition with Attention Head Fusion. *2020 10th Annual Computing and Communication Workshop and Conference, CCWC 2020*, 1058–1064. https://doi.org/10.1109/CCWC47524.2020.9031207

Xu, M., Zhang, F., Yang, J., & Khan, S. U. (2020). Exploring the Influence of Noise in Speech Emotion Recognition Devices for Internet of Thing. *Proceedings - 4th IEEE International Conference on Energy Internet, ICEI 2020*, 128–133. https://doi.org/10.1109/ICEI49372.2020.00031

Xu, M., Zhang, F., & Zhang, W. (2021). Head Fusion: Improving the Accuracy and Robustness of Speech Emotion Recognition on the IEMOCAP and RAVDESS Dataset. *IEEE Access*, *9*, 74539–74549. https://doi.org/10.1109/ACCESS.2021.3067460

Zhang, S., Zhang, S., Huang, T., & Gao, W. (2018). Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching. *IEEE Transactions on Multimedia*, *20*(6), 1576–1590. https://doi.org/10.1109/TMM.2017.2766843

Zhao, J., Mao, X., & Chen, L. (2019a). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomedical Signal Processing and Control*, *47*, 312–323. https://doi.org/10.1016/j.bspc.2018.08.035

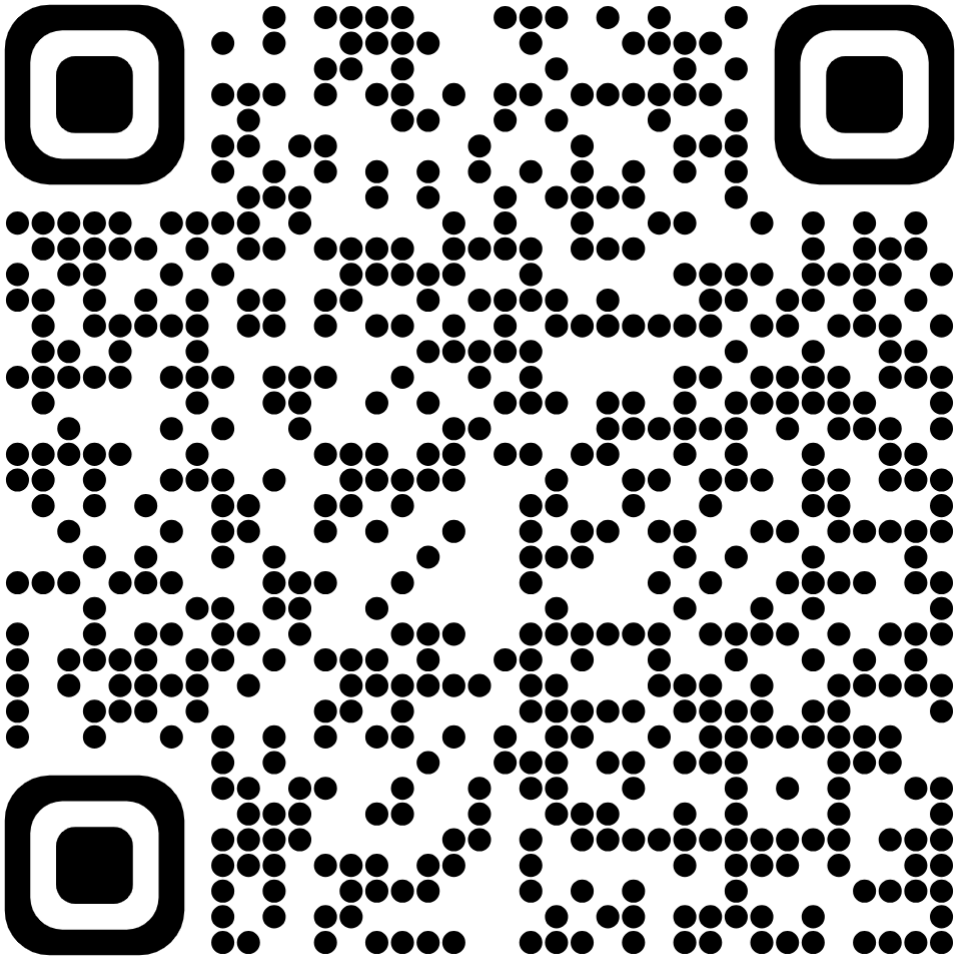
Zhao, J., Mao, X., & Chen, L. (2019b). Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomedical Signal Processing and Control*, *47*, 312–323. https://doi.org/10.1016/j.bspc.2018.08.035

Zhou, Q., Shan, J., Ding, W., Wang, C., Yuan, S., Sun, F., Li, H., & Fang, B. (2021). Cough Recognition Based on Mel-Spectrogram and Convolutional Neural Network. *Frontiers in Robotics and AI*, *8*. https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2021.580080

Zhou, T. (2013). Signal-to-Noise Ratio. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 1939–1940). Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_514

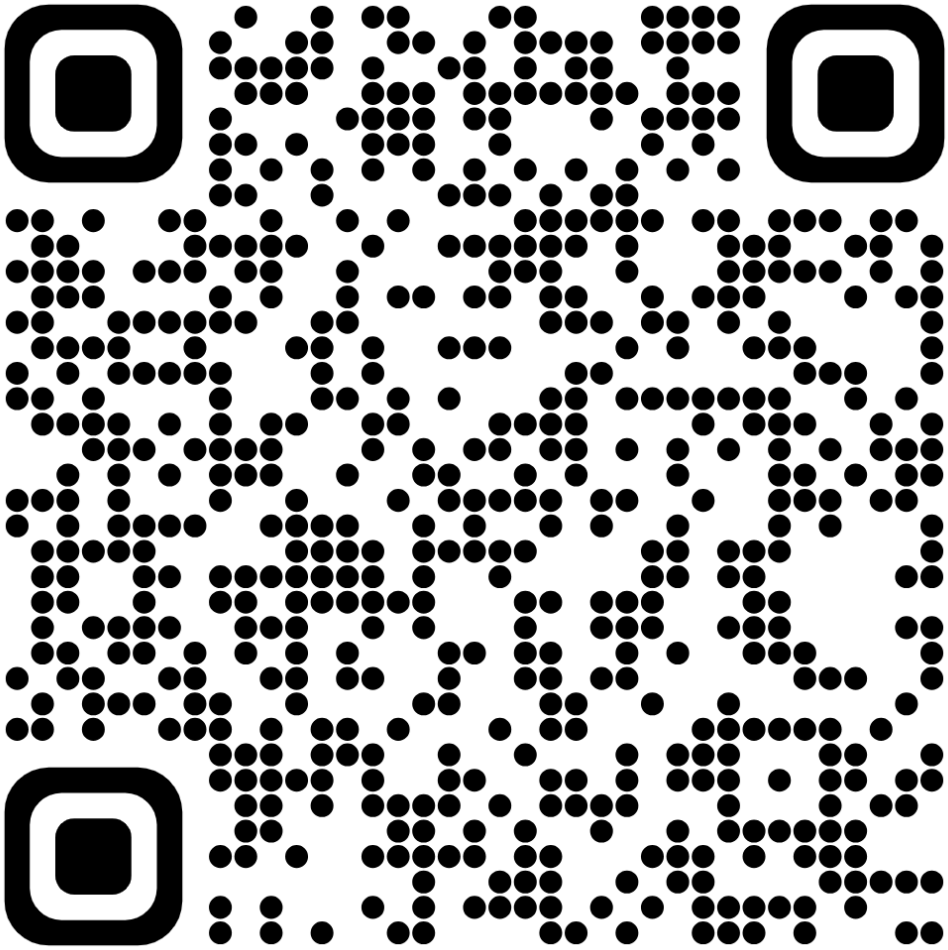
LAMPIRAN

Lampiran 1. *Source code*



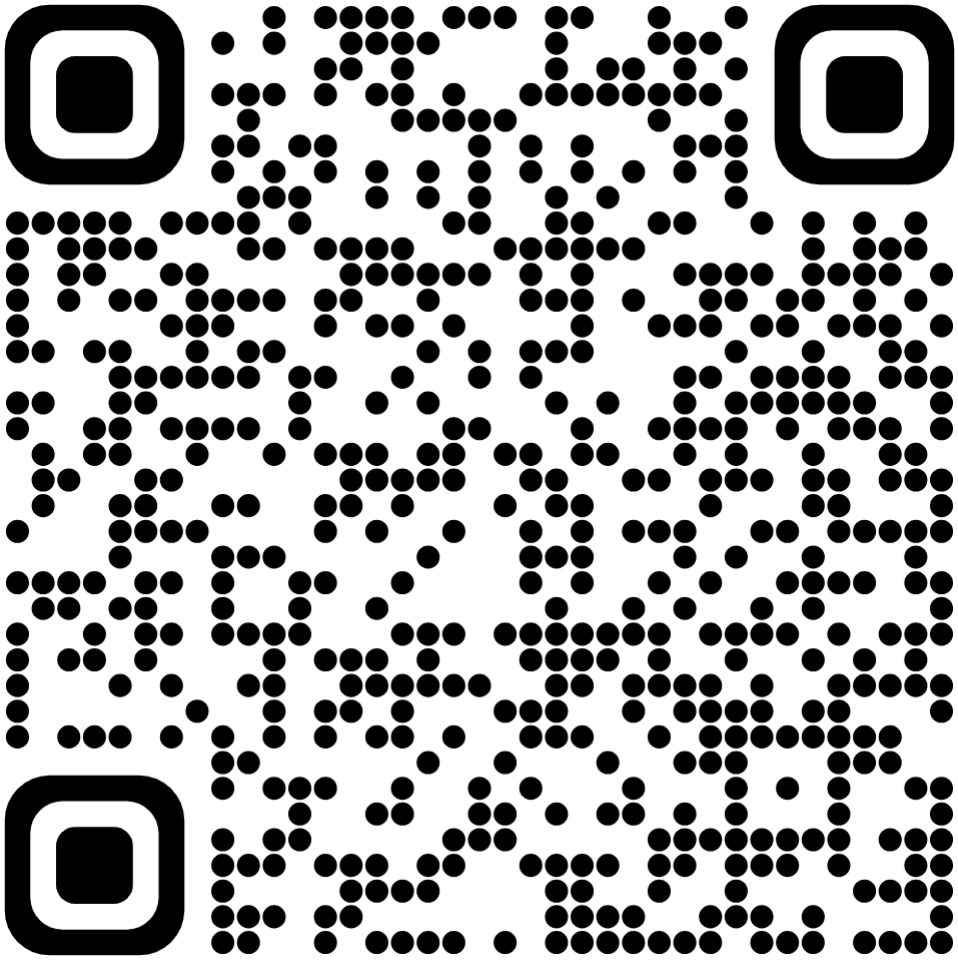
Link: https://s.unhas.ac.id/zidressourcecode

Lampiran 2. *Dataset*



Link: https://s.unhas.ac.id/zidresdataset

Lampiran 3. Hasil validasi data

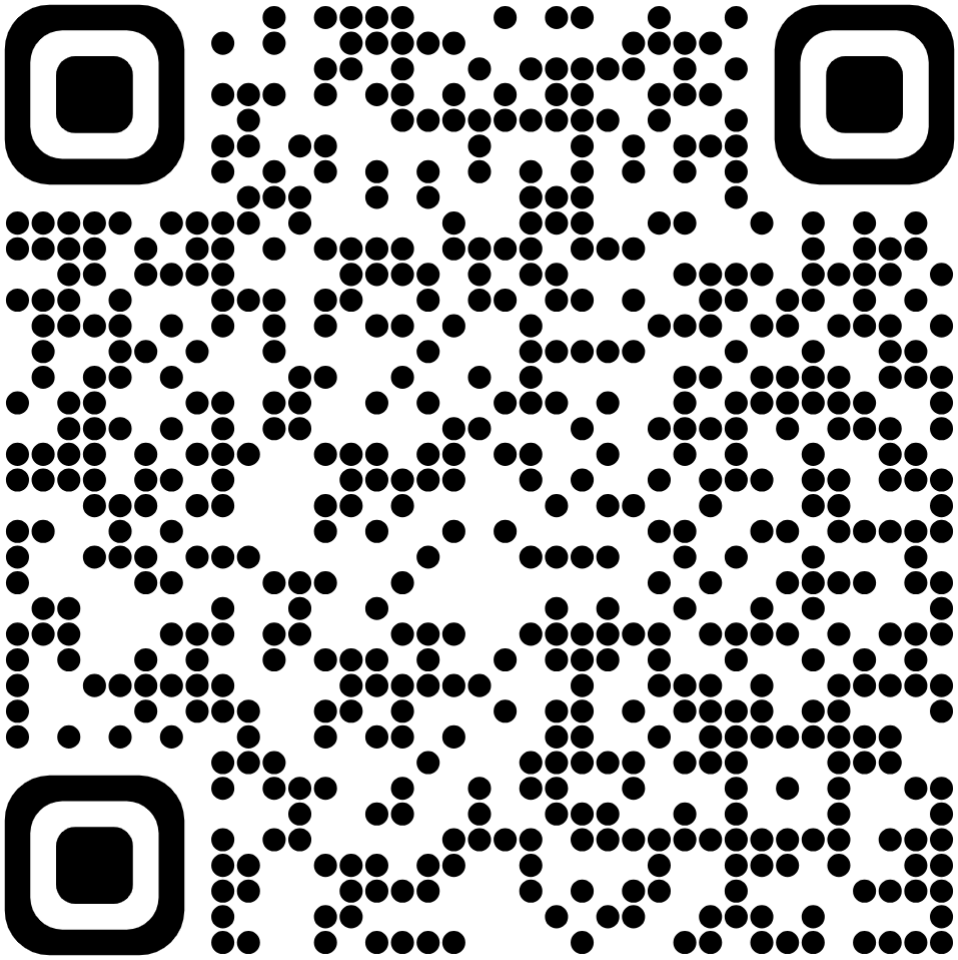


Link: https://s.unhas.ac.id/zidresvalidationdataset

Lampiran 4. Hasil data *path*

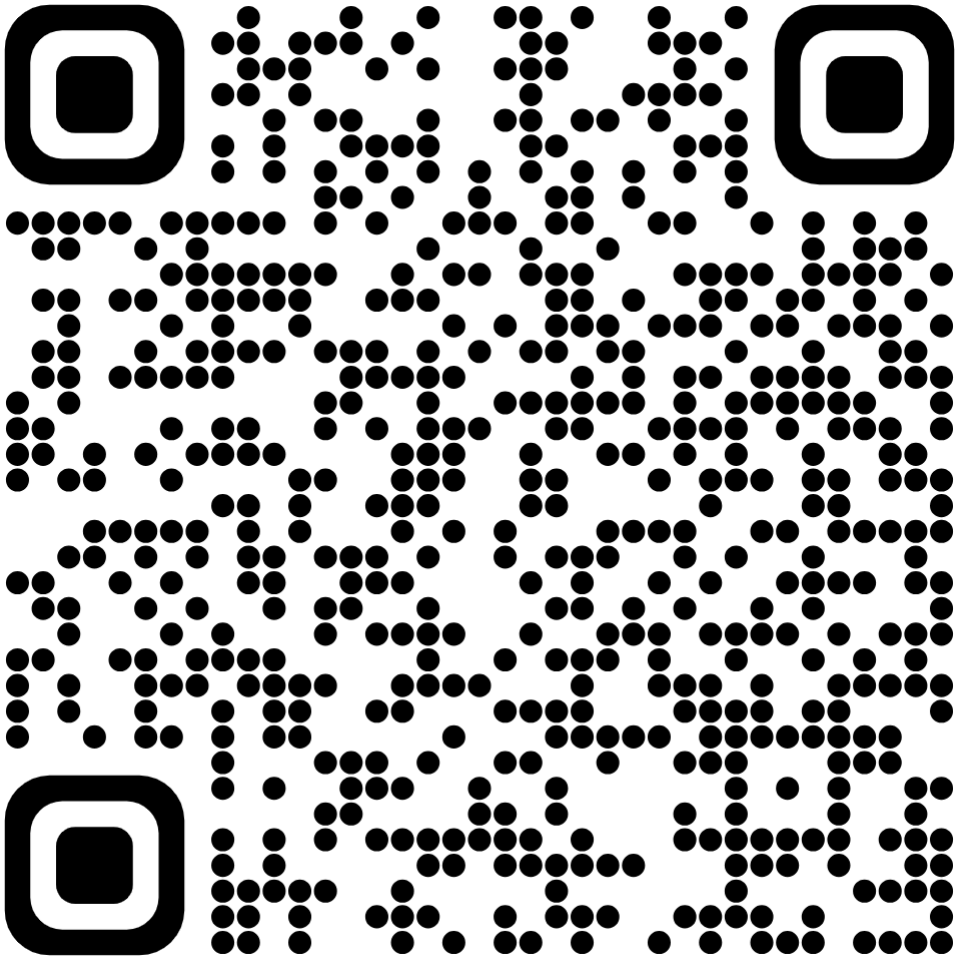
Link: https://s.unhas.ac.id/zidresdatapath

Lampiran 5. Hasil ekstraksi fitur



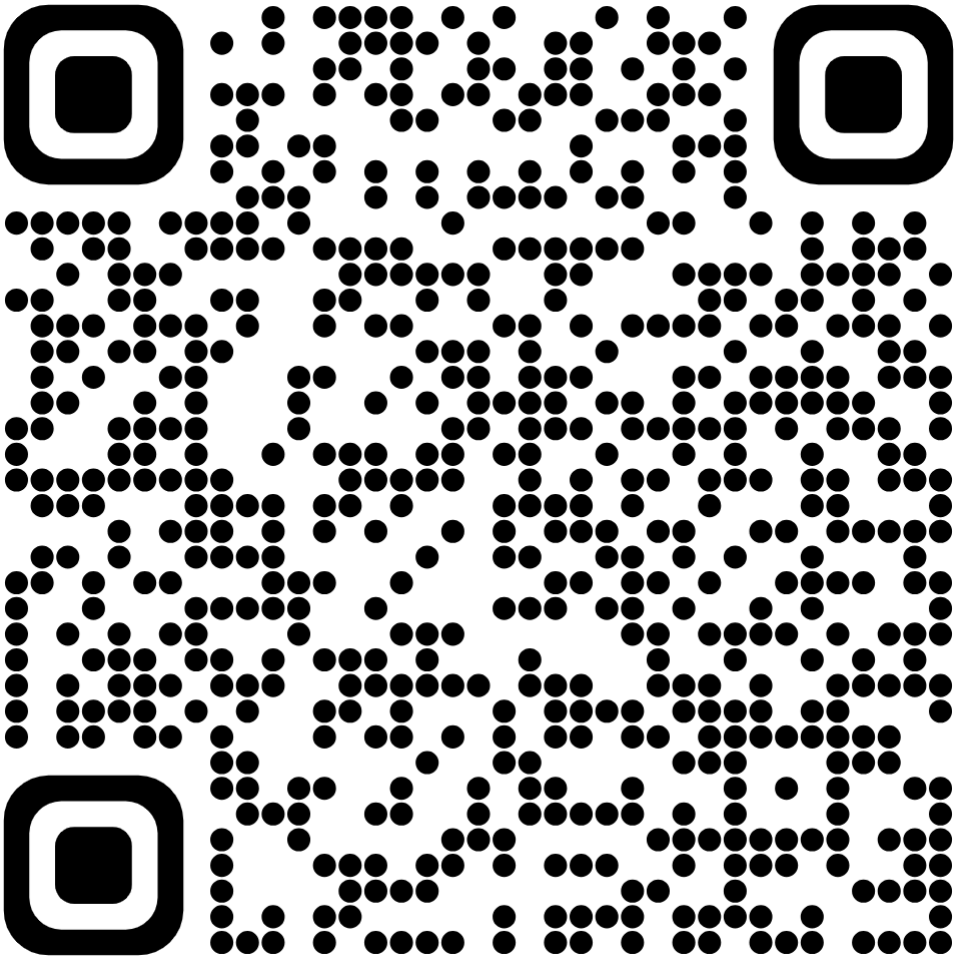
Link: https://s.unhas.ac.id/zidresfeature

Lampiran 6. Hasil *training*



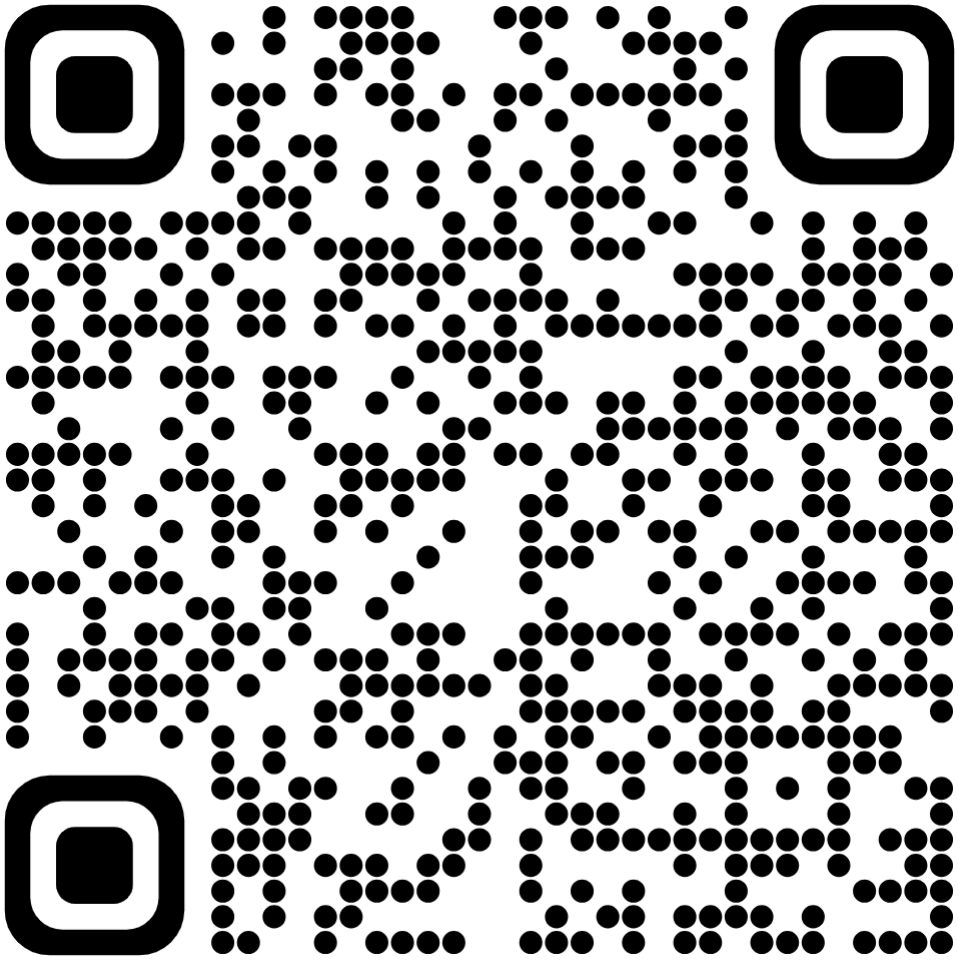
Link: https://s.unhas.ac.id/zidresthasiltraining

Lampiran 7. Hasil *testing*



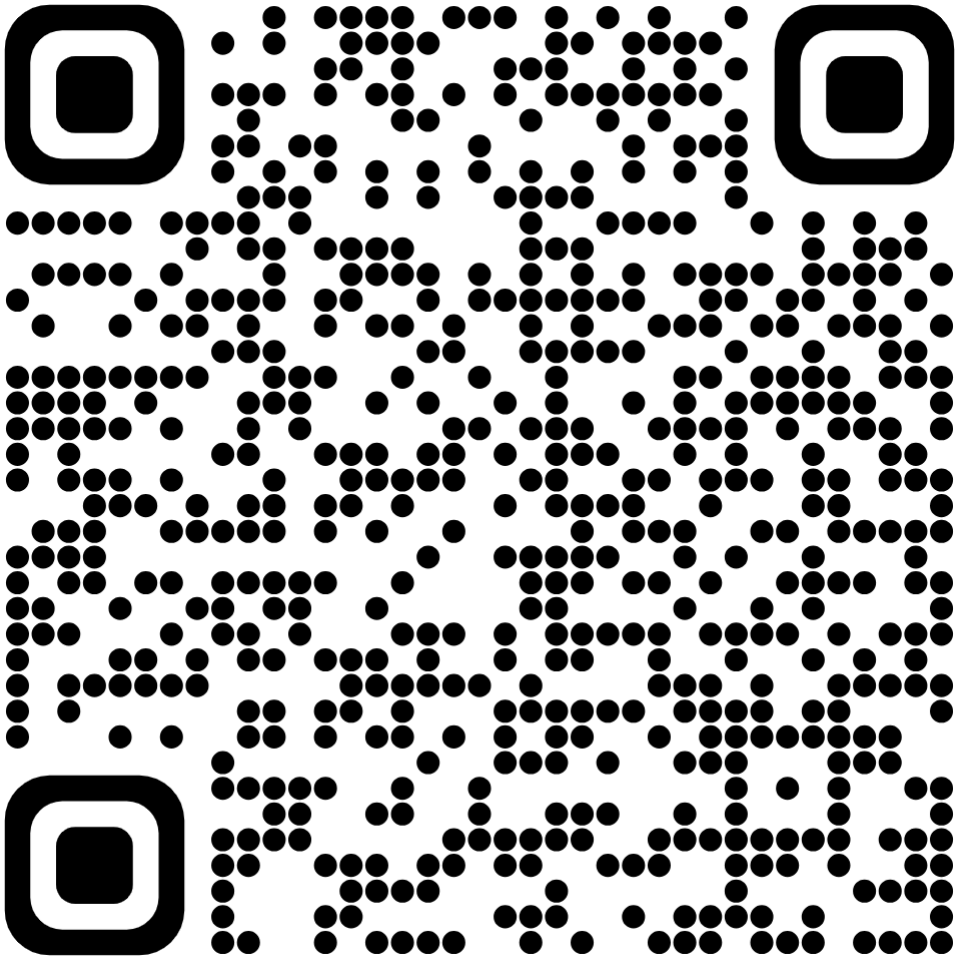
Link: https://s.unhas.ac.id/zidreshasiltesting

Lampiran 8. Hasil *confusion matrix testing*



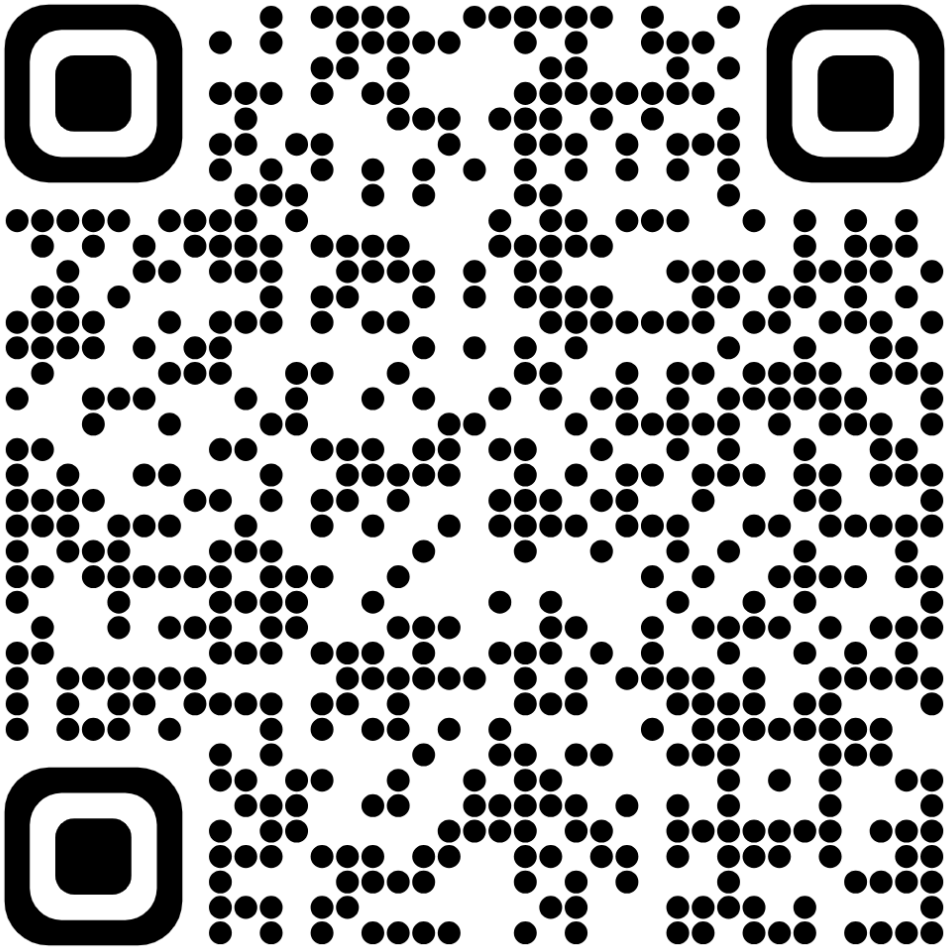
Link: https://s.unhas.ac.id/zidressourcecode

Lampiran 9. Model hasil *training*



Link: https://s.unhas.ac.id/zidresmodel

Lampiran 10. Dokumentasi



Link: https://s.unhas.ac.id/zidresdokumentasi