

DAFTAR PUSTAKA

- Al Mascaty, A., Khakim, Z., & Dianesti, N. P. (2026). Beyond emotion: exploring image memorability with eye tracking. *Cogent Psychology*, 13(1), 2677960. <https://doi.org/10.1080/23311908.2026.2677960>
- Al Qohar, B., Dullah, A. U., Utami, P., & Unjung, J. (2025). Vehicle CO2 Emission Predictive Analytics Using HistGradientBoosting Regression Algorithms. *Jurnal Media Informasi Teknologi*, 2(1), 1-12. <https://doi.org/10.69616/mit.v2i1.198>
- Al Saadi, N. (2025). Children's Drawings: An Expression of the Psychological States Experienced by the Child. *Open Journal of Social Sciences*, 13(10), 705-717. <https://doi.org/10.4236/jss.2025.1310040>
- Almulla, M. A. (2025). A multimodal emotion recognition system using deep convolution neural networks. *Journal of Engineering Research*, 13(2), 721–729. <https://doi.org/10.1016/j.jer.2024.03.021>
- Alvarado, J. (2025). Text-to-image models reveal specific color-emotion associations. *Frontiers in Psychology*, 16, 1593928. <https://doi.org/10.3389/fpsyg.2025.1593928>
- Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., & Zhang, L. (2018). Bottom-up and top-down attention for image captioning and visual question answering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6077–6086. <https://doi.org/10.1109/CVPR.2018.00636>
- Bahdanau, D., Cho, K., & Bengio, Y. (2016). Neural machine translation by jointly learning to align and translate. *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*. <https://doi.org/10.48550/arXiv.1409.0473>
- Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423-443. <https://doi.org/10.1109/TPAMI.2018.2798607>
- Barradas, I., Tschiesner, R., & Peer, A. (2025). Dynamic emotion intensity estimation from physiological signals facilitating interpretation via appraisal theory. *PLoS ONE*, 20(1), Article e0315929. <https://doi.org/10.1371/journal.pone.0315929>
- Bautista, J. L., & Shin, H. S. (2025). Speech emotion recognition model based on joint modeling of discrete and dimensional emotion representation. *Applied Sciences*, 15(2), 623. <https://doi.org/10.3390/app15020623>
- Bin Saeed, Q., & Cha, Y. (2025). Multi-modal deep-attention-BiLSTM based early detection of mental health issues using social media posts. *Scientific Reports*, 15(1), 35152. <https://doi.org/10.1038/s41598-025-19141-0>
- Boateng, G., Santhanam, P., Fleisch, E., Lüscher, J., Pauly, T., Scholz, U., Bodenmann, G., & Kowatsch, T. (2022). “Are you okay, honey?”: Recognizing emotions among couples managing diabetes in daily life using multimodal real-world smartwatch data. *arXiv*. <https://doi.org/10.48550/arXiv.2208.08909>

- Bose, D., Somandepalli, K., Kundu, S., Lahiri, R., Gratch, J., & Narayanan, S. (2021). *Understanding of emotion perception from art*. arXiv. <https://doi.org/10.48550/arXiv.2110.06486>
- Buechel, S., & Hahn, U. (2017). EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (pp. 578–585). Association for Computational Linguistics. <https://aclanthology.org/E17-2092/>
- Burkitt, E., Barrett, M., & Davis, A. (2007). Effects of different emotion terms on the size and colour of children's drawings. *International Journal of Art Therapy*, 14(2), 74–84. <https://doi.org/10.1080/17454830701529567>
- Cannoni, E., Pinto, G., & Bombi, A. S. (2023). Typical emotional expression in children's drawings of the human face. *Current Psychology*, 42, 2762–2768. <https://doi.org/10.1007/s12144-021-01558-1>
- Chen, Y., Yuan, J., Zhao, L., Luo, R., Davis, L., & Metaxas, D. N. (2021). More Than Just Attention: Learning Cross-Modal Attentions with Contrastive Constraints. arXiv preprint arXiv:2105.09597. <https://doi.org/10.48550/arXiv.2105.09597>
- Choi, H., Kang, M., Kwon, Y., & Yoon, S.-E. (2019). *An objectness score for accurate and fast detection during navigation*. arXiv. <https://doi.org/10.48550/arXiv.1909.05626>
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv. <https://doi.org/10.48550/arXiv.1412.3555>
- Çiftçi, S., Karci, H., Karaca, N., Söylemez, B., & Koçakoğlu, H. (2025). Emotion analysis of children's drawings. *IEEE Access*, 13, 159730–159748. <https://doi.org/10.1109/ACCESS.2025.3606359>
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734). Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1179>
- Crawford, E., Gross, J., Patterson, T., & Hayne, H. (2012). Does children's colour use reflect the emotional content of their drawings? *Infant and Child Development*, 21(2), 198–215. <https://doi.org/10.1002/icd.742>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An image is worth 16×16 words: Transformers for image recognition at scale*. arXiv. <https://doi.org/10.48550/arXiv.2010.11929>

- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3-4), 169-200. <https://doi.org/10.1080/02699939208411068>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fabris, M. A., Lange-Küttner, C., Shiakou, M., & Longobardi, C. (2023). Editorial: Children's drawings: Evidence-based research and practice. *Frontiers in Psychology*, 14, Article 1250556. <https://doi.org/10.3389/fpsyg.2023.1250556>
- Fahmi, M. I., & Nababan, A. A. (2025). A comparative study of BERT and RoBERTa for sentiment analysis on Twitter data related to mental health. *Mathematical Modelling of Engineering Problems*, 12(9), 3289–3295. <https://doi.org/10.18280/mmep.120931>
- Farokhi, M., & Hashemi, M. (2011). The analysis of children's drawings: Social, emotional, physical, and psychological aspects. *Procedia - Social and Behavioral Sciences*, 30, 2219–2224. <https://doi.org/10.1016/j.sbspro.2011.10.433>
- Ferro, F., Gaspari, C., Manfrè, G., Cernigliaro, F., D'Agnano, D., Panzica, R., Correnti, E., Ruta, M. R., Marchese, F., Pitino, R., Capizzi, M., Santangelo, G., Versace, A., Sciruicchio, V., & Raieli, V. (2025). Children's Headache Through Drawings: A Narrative Review and a Portrait Gallery. *Life (Basel, Switzerland)*, 15(7), 996. <https://doi.org/10.3390/life15070996>
- Galvão, F., Alarcão, S. M., & Fonseca, M. J. (2021). Predicting Exact Valence and Arousal Values from EEG. *Sensors*, 21(10), 3414. <https://doi.org/10.3390/s21103414>
- Gao, T., Singh, S., & Mooney, R. (2022). Towards Automated Error Analysis: Learning to Characterize Errors. *The International FLAIRS Conference Proceedings*, 35. <https://doi.org/10.32473/flairs.v35i.130632>
- Guo, C., Pleiss, G., Sun, S., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning*, 70, 1321–1330. <https://proceedings.mlr.press/v70/guo17a.html>
- Ham S-M, Lee H-M, Lim J-H, Seo J. A Negative Emotion Recognition System with Internet of Things-Based Multimodal Biosignal Data. *Electronics*. 2023; 12(20):4321. <https://doi.org/10.3390/electronics12204321>
- Han, D., Tohti, T., & Hamdulla, A. (2022). Attention-Based Transformer-BiGRU for Question Classification. *Information*, 13(5), 214. <https://doi.org/10.3390/info13050214>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Huang, X., Zhu, S., Wang, Z., et al. (2025). An enhanced valence-arousal multimodal emotion dataset for emotion recognition. *Scientific Data*, 12, 1944. <https://doi.org/10.1038/s41597-025-06214-y>
- Huang, Z., Low, C., Teng, M., Zhang, H., Ho, D. E., Krass, M. S., & Grabmair, M. (2021). Context-aware legal citation recommendation using deep learning. In *Proceedings of the Eighteenth International Conference on Artificial*

- Intelligence and Law (ICAIL '21)* (pp. 79–88). ACM. <https://doi.org/10.1145/3462757.3466066>
- Hutson, A. D., & Yu, H. (2021). A robust permutation test for the concordance correlation coefficient. *Pharmaceutical statistics*, 20(4), 696–709. <https://doi.org/10.1002/pst.2101>
- John, V., & Kawanishi, Y. (2022). Audio and video-based emotion recognition using multimodal transformers. *Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR)*, 2582–2588. <https://doi.org/10.1109/ICPR56361.2022.9956730>
- Karim, M. A., Rafiuddin, S. M., Razin, M. J. I., & Alam, T. (2022). Isolated Bangla handwritten character classification using transfer learning. In *Proceedings of the International Conference on Computing Advancements (ICCA '22)* (pp. 1–7). ACM. <https://doi.org/10.1145/3542954.3542957>
- Khlaif, A. I., Naceur, M. S., & Kherallah, M. (2025). AI-driven classification of children's drawings for pediatric psychological evaluation: An ensemble deep learning approach. *Journal of Robotics and Control*, 6(1), 124–139. <https://doi.org/10.18196/jrc.v6i1.23302>
- Lee, L. H., Li, J. H., & Yu, L. C. (2022). Chinese EmoBank: Building Valence-Arousal Resources for Dimensional Sentiment Analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(4), 1-18. <https://doi.org/10.1145/3489141>
- Lee, T.-G., & Yoo, J.-H. (2025). Interpretable autism screening via object-based analysis of child-drawn sketches. *IEEE Access*, 13, 211138–211148. <https://doi.org/10.1109/ACCESS.2025.3642950>
- Li, B. (2025). Speech emotion recognition based on CNN-Transformer with different loss function. *Journal of Computer and Communications*, 13(3), 103–115. <https://doi.org/10.4236/jcc.2025.133008>
- Liang, Y., Maeda, K., Ogawa, T., & Haseyama, M. (2021). Deep metric network via heterogeneous semantics for image sentiment analysis. *Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP)*, 1039–1043. <https://doi.org/10.1109/ICIP42928.2021.9506701>
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv. <https://doi.org/10.48550/arXiv.1907.11692>
- Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. In *7th International Conference on Learning Representations (ICLR)*. <https://doi.org/10.48550/arXiv.1711.05101>
- Mahadevkar, S., Patil, S., & Kotecha, K. (2024). Enhancement of handwritten text recognition using AI-based hybrid approach. *MethodsX*, 12, 102654. <https://doi.org/10.1016/j.mex.2024.102654>
- Manomenidis, G., Liamopoulou, P., Vlotinou, P., Tsiakiri, A., Cassimos, D., Amanatidou, M., Koutelekos, I., & Georgousopoulou, V. (2025). Evaluating Children's Drawings as a Means of Expression in Children with Chronic Diseases. *Medical archives (Sarajevo, Bosnia and Herzegovina)*, 79(2), 159–163. <https://doi.org/10.5455/medarh.2025.79.159-163>
- Mendes, G. A., & Martins, B. (2023). Quantifying valence and arousal in text with multilingual pre-trained transformers. In J. Kamps dkk. (Eds.), *Advances in*

- Information Retrieval (ECIR 2023)* (Lecture Notes in Computer Science, Vol. 13980). Springer. https://doi.org/10.1007/978-3-031-28244-7_6
- Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In H. L. Meiselman (Ed.), *Emotion measurement* (pp. 201–237). Woodhead Publishing. <https://doi.org/10.1016/B978-0-08-100508-8.00009-6>
- Moula, Z., Walshe, N., & Lee, E. (2021). Making nature explicit in children's drawings of wellbeing and happy spaces. *Child Indicators Research*, 14(4), 1653–1675. <https://doi.org/10.1007/s12187-021-09811-6>
- Nath, S. S., del Cuavillo y Schröder, G., & Stevenson, C. E. (2025). *Pencils to pixels: A systematic study of creative drawings across children, adults and AI*. arXiv. <https://doi.org/10.48550/arXiv.2502.05999>
- Pinto, J., Fred, A., & Silva, H. P. (2019). Biosignal-Based Multimodal Emotion Recognition in a Valence-Arousal Affective Framework Applied to Immersive Video Visualization. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3577–3583. <https://doi.org/10.1109/EMBC.2019.8856950>
- Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37, 98–125. <https://doi.org/10.1016/j.inffus.2017.02.003>
- Quaglia, R., Longobardi, C., Iotti, N. O., & Prino, L. E. (2015). A new theory on children's drawings: Analyzing the role of emotion and movement in graphical development. *Infant Behavior and Development*, 39, 81–91. <https://doi.org/10.1016/j.infbeh.2015.02.009>
- Restoy, S., Martinet, L., Sueur, C., & Pelé, M. (2022). Draw yourself: How culture influences drawings by children between the ages of two and fifteen. *Frontiers in Psychology*, 13, 940617. <https://doi.org/10.3389/fpsyg.2022.940617>
- Rezatofghi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized Intersection over Union: A metric and a loss for bounding box regression. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 658–666. <https://doi.org/10.1109/CVPR.2019.00075>
- Scherer, K. R. (2022). Theory convergence in emotion science is timely and realistic. *Cognition and Emotion*, 36(2), 154–170. <https://doi.org/10.1080/02699931.2021.1973378>
- Shoer, I., & Erzin, E. (2025). Learning annotation consensus for continuous emotion recognition. arXiv preprint arXiv:2505.21196. <https://doi.org/10.48550/arXiv.2505.21196>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Stanojevic, S., Kaminsky, D. A., Miller, M. R., Thompson, B., Aliverti, A., Barjaktarevic, I., Cooper, B. G., Culver, B., Derom, E., Hall, G. L., Hallstrand, T. S., Leuppi, J. D., MacIntyre, N., McCormack, M., Rosenfeld, M., & Swenson, E. R. (2022). ERS/ATS technical standard on interpretive strategies for routine lung function tests. *European Respiratory Journal*, 60(1), Article 2101499. <https://doi.org/10.1183/13993003.01499-2021>

- Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to fine-tune BERT for text classification? In M. Sun, X. Huang, H. Ji, Z. Liu, & Y. Liu (Eds.), *Chinese computational linguistics (CCL 2019)* (Lecture Notes in Computer Science, Vol. 11856, pp. 194–206). Springer. https://doi.org/10.1007/978-3-030-32381-3_16
- Sudanayasa, A. T., Lagandesa, Y. R., Hariana, K., Rizal, R., & Kasmawati, K. (2026). Creativity Characteristics in Indonesian Elementary Students' Drawings: A Cross-Cultural Validation of Lowenfeld's Theory. *Journal of Innovation and Research in Primary Education*, 5(1), 1856–1858. <https://doi.org/10.56916/jirpe.v5i1.3246>
- Sugiyono. (2019). *Metode penelitian kuantitatif kualitatif dan R&D* (Cet. 1). Alfabeta.
- Touhami, S., & Lakhal, M. (2025). The psychology of drawing in children. *Psychology and Education Journal*, 62(2), 436–447. <http://psychologyandeducation.net/pae/index.php/pae/article/view/9973>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2023). Attention is all you need. arXiv. <https://doi.org/10.48550/arXiv.1706.03762>
- Wang, W., Liu, P., Yang, S., & Zhang, W. (2020). Dynamic interaction networks for image-text multimodal learning. *Neurocomputing*, 379, 262–272. <https://doi.org/10.1016/j.neucom.2019.10.103>
- Wei, Y., Li, S., Feng, R., & Hu, D. (2024). Diagnosing and re-learning for balanced multimodal learning. arXiv. <https://doi.org/10.48550/arXiv.2407.09705>
- Wongoutong, C. (2024). The impact of neglecting feature scaling in k-means clustering. *PLOS ONE*, 19(12), 1–19. <https://doi.org/10.1371/journal.pone.0310839>
- Yang, X., Feng, S., Wang, D., & Zhang, Y. (2021). Image-text multimodal emotion classification via multi-view attentional network. *IEEE Transactions on Multimedia*, 23, 4014–4026. <https://doi.org/10.1109/TMM.2020.3035277>
- Ye, L. (2025). Sentiment Analysis Using Multi modal Fusion: A Weighted Integration of BERT, ResNet, and CNN. *Informatica*, 49(24). <https://doi.org/10.31449/inf.v49i24.8315>
- Yu, Q., Zhao, H., & Wang, Z. (2019). Attention-based bidirectional gated recurrent unit neural networks for sentiment analysis. *Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition*, 116–119. <https://doi.org/10.1145/3357254.3357262>
- Zhang, Z., Wang, W., Zhu, L., & Tang, Z. (2025). TAG-fusion: Two-stage attention guided multi-modal fusion network for semantic segmentation. *Digital Signal Processing*, 156(Part A), 104807. <https://doi.org/10.1016/j.dsp.2024.104807>
- Zhao, J., Wang, Y., Bao, J., Wu, Y., & He, X. (2022). Fine- and coarse-granularity hybrid self-attention for efficient BERT. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, 4811–4820. <https://doi.org/10.18653/v1/2022.acl-long.330>
- Zheng, F., Li, W., Wang, X., Wang, L., Zhang, X., & Zhang, H. (2022). A Cross-Attention Mechanism Based on Regional-Level Semantic Features of Images for Cross-Modal Text-Image Retrieval in Remote Sensing. *Applied Sciences*, 12(23), 12221. <https://doi.org/10.3390/app122312221>

Zhu, T., Li, L., Yang, J., Zhao, S., Liu, H., & Qian, J. (2022). Multimodal Sentiment Analysis With Image-Text Interaction Network. *IEEE Transactions on Multimedia*, 25, 3375–3387. <https://doi.org/10.1109/TMM.2022.3160060>