UNIVERSITAS
GADJAH MADA
Model Paralel CNN dan BLIP-2 Untuk Identifikasi Jenis Subject pertanyaan Matematika Berbahasa Inggris
Muhammad Ashabul Kahfi, Prof. Dr. Drs. Azhari, MT
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

Abbasi, M. M., Iqbal, S., Aurangzeb, K., Alhussein, M., & Khan, T. M. (2024). Lmbis-net: A lightweight bidirectional skip connection based multipath cnn for retinal blood vessel segmentation. *Scientific Reports*, 14(1), 15219. https://doi.org/10.1038/s41598-024-63496-9

Ahmad, M. A., Ahmed, T., Aslam, M., Rehman, A., Alamri, F. S., Bahaj, S. A., & Saba, T. (2025). Mathvision: An accessible intelligent agent for visually impaired people to understand mathematical equations. *IEEE Access*, 13, 6155–6165. https://doi.org/10.1109/ACCESS.2024.3514079

AlSaad, R., Abd-alrazaq, A., Boughorbel, S., Ahmed, A., Renault, M.-A., Damseh, R., & Sheikh, J. (2024). Multimodal large language models in health care: Applications, challenges, and future outlook. *J Med Internet Res*, 26, e59505. https://doi.org/10.2196/59505

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), 53. https://doi.org/10.1186/s40537-021-00444-8

Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). 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

Batool, A. & Byun, Y.-C. (2025). A lightweight multi-path convolutional neural network architecture using optimal features selection for multiclass classification of brain tumor using magnetic resonance images. *Results in Engineering*, 25, 104327. https://doi.org/10.1016/j.rineng.2025.104327

Çetin, V. & Yıldız, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 28(2), 299–312. https://doi.org/10.5505/pajes.2021.62687

UNIVERSITAS
GADJAH MADA
Model Paralel CNN dan BLIP-2 Untuk Identifikasi Jenis Subject pertanyaan Matematika Berbahasa Inggris
Muhammad Ashabul Kahfi, Prof. Dr. Drs. Azhari, MT
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321–357. https://doi.org/10.1613/jair.953

Chen, F.-L., Zhang, D.-Z., Han, M.-L., Chen, X.-Y., Shi, J., Xu, S., & Xu, B. (2023). Vlp: A survey on vision-language pre-training. *Machine Intelligence Research*, 20(1), 38–56. https://doi.org/10.1007/s11633-022-1369-5

Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: latest research, applications and future directions. *Artificial Intelligence Review*, 57(6), 137. https://doi.org/10.1007/s10462-024-10759-6

Cubuk, E. D., Zoph, B., Shlens, J., & Le, Q. V. (2020). Randaugment: Practical automated data augmentation with a reduced search space. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 3008–3017. https://doi.org/10.1109/CVPRW50498.2020.00359

Deng, Y., Bansal, H., Yin, F., Peng, N., Wang, W., & Chang, K.-W. (2025). *Openvlthinker: Complex vision-language reasoning via iterative sft-rl cycles*. https://arxiv.org/abs/2503.17352

Ghosal, D., Majumder, N., Lee, R., Mihalcea, R., & Poria, S. (2023). Language guided visual question answering: Elevate your multimodal language model using knowledge-enriched prompts. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 12096–12102. https://doi.org/10.18653/v1/2023.findings-emnlp.809

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

He, H. & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Ishmam, M. F., Shovon, M. S. H., Mridha, M., & Dey, N. (2024). From image to language: A critical analysis of visual question answering (vqa) approaches, challenges, and opportunities. *Information Fusion*, 106, 102270. https://doi.org/https://doi.org/10.1016/j.inffus.2024.102270

UNIVERSITAS
GADJAH MADA
Model Paralel CNN dan BLIP-2 Untuk Identifikasi Jenis Subject pertanyaan Matematika Berbahasa Inggris
Muhammad Ashabul Kahfi, Prof. Dr. Drs. Azhari, MT
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Jamil, S., Ali, T., Nawaz, A., & Shahid, A. (2026). Eac-agent: A deep learning framework for multimodal emotion-aware conversational agent with contextual response generation. *PLOS ONE*, 21(4), 1–24. https://doi.org/10.1371/journal.pone.0346770

Johnson, J. M. & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1), 27. https://doi.org/10.1186/s40537-019-0192-5

Kuang, J., Shen, Y., Xie, J., Luo, H., Xu, Z., Li, R., Li, Y., Cheng, X., Lin, X., & Han, Y. (2025). Natural language understanding and inference with mllm in visual question answering: A survey. *ACM Comput. Surv.*, 57(8). https://doi.org/10.1145/3711680

Lan, Y., Li, X., Du, H., Lu, X., Gao, M., Qian, W., & Zhou, A. (2024). Survey of natural language processing for education: Taxonomy, systematic review, and future trends. *IEEE Transactions on Knowledge and Data Engineering*, 38, 659–678. https://api.semanticscholar.org/CorpusID:266999439

Lee, H. S., Choi, S., Lee, Y., Moon, H., Oh, S., Jeong, M., Go, H., & Wallraven, C. (2023). Cross encoding as augmentation: Towards effective educational text classification. *Findings of the Association for Computational Linguistics: ACL 2023*, 2184–2195. https://doi.org/10.18653/v1/2023.findings-acl.137

Li, C., Gan, Z., Yang, Z., Yang, J., Li, L., Wang, L., & Gao, J. (2024). Multimodal foundation models: From specialists to general-purpose assistants. *Found. Trends. Comput. Graph. Vis.*, 16(1–2), 1–214. https://doi.org/10.1561/0600000110

Li, J., Li, D., Savarese, S., & Hoi, S. (2023). BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 19730–19742. https://proceedings.mlr.press/v202/li23q.html

Lu, P., Bansal, H., Xia, T., Liu, J., Li, C., Hajishirzi, H., Cheng, H., Chang, K.-W., Galley, M., & Gao, J. (2024). *Mathvista: Evaluating mathematical reasoning*

UNIVERSITAS
GADJAH MADA
Model Paralel CNN dan BLIP-2 Untuk Identifikasi Jenis Subject pertanyaan Matematika Berbahasa Inggris
Muhammad Ashabul Kahfi, Prof. Dr. Drs. Azhari, MT
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

of foundation models in visual contexts. https://arxiv.org/abs/2310.02255

Pawłowski, M., Wróblewska, A., & Sysko-Romańczuk, S. (2023). Effective techniques for multimodal data fusion: A comparative analysis. *Sensors*, 23(5). https://doi.org/10.3390/s23052381

Qiao, R., Tan, Q., Dong, G., MinhuiWu, M., Sun, C., Song, X., Wang, J., GongQue, Z., Lei, S., Zhang, Y., Wei, Z., Zhang, M., Qiao, R., Zong, X., Xu, Y., Yang, P., Bao, Z., Diao, M., Li, C., & Zhang, H. (2025). We-math: Does your large multimodal model achieve human-like mathematical reasoning? *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 20023–20070. https://doi.org/10.18653/v1/2025.acl-long.983

Rabiee, A., Goel, A., D'Souza, J., & Khanwalkar, S. (2022). *Question-type identification for academic questions in online learning platform*. https://doi.org/10.48550/arXiv.2211.13727

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, 8748–8763. https://proceedings.mlr.press/v139/radford21a.html

Raeder, T., Forman, G., & Chawla, N. V. (2012). *Learning from Imbalanced Data: Evaluation Matters*, 315–331. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23166-7_12

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Wang, K., Pan, J., Shi, W., Lu, Z., Ren, H., Zhou, A., Zhan, M., & Li, H. (2024). Measuring multimodal mathematical reasoning with math-vision dataset. *Advances in Neural Information Processing Systems*, volume 37, 95095–95169. https://doi.org/10.52202/079017-3014

UNIVERSITAS
GADJAH MADA
Model Paralel CNN dan BLIP-2 Untuk Identifikasi Jenis Subject pertanyaan Matematika Berbahasa Inggris
Muhammad Ashabul Kahfi, Prof. Dr. Drs. Azhari, MT
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Wang, P., Li, D., Hu, X., Wang, Y., & Zhang, Y. (2025). Clipmulti: Explore the performance of multimodal enhanced clip for zero-shot text classification. *Computer Speech & Language*, 90, 101748. https://doi.org/10.1016/j.csl.2024.101748

Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2024a). A survey on multimodal large language models. *National Science Review*, 11(12), nwae403. https://doi.org/10.1093/nsr/nwae403

Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2024b). A survey on multimodal large language models. *National Science Review*, 11(12), nwae403. https://doi.org/10.1093/nsr/nwae403

Yuhana, U. L., Yuniarno, E. M., Rahayu, W., & Pardede, E. (2024). A context-based question selection model to support the adaptive assessment of learning: A study of online learning assessment in elementary schools in indonesia. *Education and Information Technologies*, 29(8), 9517 – 9540. Cited by: 1. https://doi.org/10.1007/s10639-023-12184-8

Zhang, J., Huang, J., Jin, S., & Lu, S. (2024). Vision-Language Models for Vision Tasks: A Survey . *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 46(08), 5625–5644. https://doi.org/10.1109/TPAMI.2024.3369699

Zhang, J., Khayatkhoei, M., Chhikara, P., & Ilievski, F. (2023). *Using visual cropping to enhance fine-detail question answering of blip-family models*. https://arxiv.org/abs/2306.00228

Zhang, R., Wei, X., Jiang, D., Guo, Z., Zhang, Y., Tong, C., Liu, J., Zhou, A., Zhang, S., Peng, G., & Li, H. (2025). Mavis: Mathematical visual instruction tuning with an automatic data engine. *International Conference on Learning Representations*, volume 2025, 87955–87989. https://proceedings.iclr.cc/paper_files/paper/2025/file/db36dcad6baee298a34ffca324b84b09-Paper-Conference.pdf