UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

Ahmad, I. S., Dai, J., Xie, Y., and Liang, X. (2025). Deep learning models for ct image classification: a comprehensive literature review. *Quantitative Imaging in Medicine and Surgery*, 15(1):962–1011. doi:10.21037/qims-24-1400

Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*. doi:10.48550/arXiv.1607.06450

Battleday, R. M., Peterson, J. C., and Griffiths, T. L. (2021). From convolutional neural networks to models of higher-level cognition (and back again). *Annals of the New York Academy of Sciences*, 1505(1):55–78. doi:10.1111/nyas.14593

Belkin, M., Hsu, D., Ma, S., and Mandal, S. (2019). Reconciling modern machine learning practice and the bias-variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854. doi:10.1073/pnas.1903070116

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Physica-Verlag HD. doi:10.1007/978-3-7908-2604-3_16

Boulouis, G., Morotti, A., Charidimou, A., Dowlatshahi, D., and Goldstein, J. N. (2022). Hematoma expansion in intracerebral hemorrhage: Predictors and outcome. *Stroke*, 53(1):e28–e34. doi:10.1016/S1474-4422(22)00338-6

Bousquet, O. and Elisseeff, A. (2002). Stability and generalization. *Journal of Machine Learning Research*, 2:499–526. doi:10.1162/153244302760200704

Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K. J., Brogi, E., Reuter, V. E., Klimstra, D. S., and Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, 25(8):1301–1309. doi:10.1038/s41591-019-0508-1

Castro, E., Cardoso, J. S., and Pereira, J. C. (2019). Convolutional neural networks for detection of intracranial hemorrhage in CT images. *arXiv preprint*. Evaluasi pada dataset CQ500 (491 studi) dengan split *patient-level* 85/15; VGG16 AUC 0,783.

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Cernekova, Z., Sisik, V., and Jafari, F. (2024). Detection of Vascular Leukoencephalopathy in CT Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 15446 LNAI:158–172. doi:10.1007/978-3-031-77915-2_12

Cherkassky, V., Shao, X., Mulier, F. M., and Vapnik, V. N. (1999). Model complexity control for regression using vc generalization bounds. *IEEE transactions on Neural Networks*, 10(5):1075–1089. doi:10.1109/72.788648

Chicco, D. and Jurman, G. (2020). The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21:6. doi:10.1186/s12864-019-6413-7

Chilamkurthy, S. et al. (2018). Deep learning algorithms for detection of critical findings in head ct scans: A retrospective study. *The Lancet*, 392(10162):2388–2396. doi:10.1016/S0140-6736(18)31645-3

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3):837–845. doi:10.2307/2531595

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1923. doi:10.1162/089976698300017197

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87. doi:10.1145/2347736.2347755

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Desai, M., Minderer, M., Heigold, G., Gelly, S., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*. doi:10.48550/arXiv.2010.11929

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., et al. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1):24–29. doi:10.1038/s41591-018-0316-z

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Flanders, A. E., Prevedello, L. M., Shih, G., Halabi, S. S., Kalpathy-Cramer, J., Ball, R., Mongan, J. T., Stein, A., Kitamura, F. C., Lungren, M. P., et al. (2020). Construction of a machine learning dataset through collaboration: the RSNA 2019 brain CT hemorrhage challenge. *Radiology: Artificial Intelligence*, 2(3):e190074. doi:10.1148/ryai.2020190074

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58. doi:10.1162/neco.1992.4.1.1

Glocker, B., Robinson, R., Castro, D. C., Dou, Q., and Konukoglu, E. (2019). Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects. *Medical Image Analysis*, 63:101692. doi:10.48550/arXiv.1910.04597

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y., and He, K. (2017). Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv preprint arXiv:1706.02677*. doi:10.48550/arXiv.1706.02677

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36. doi:10.1148/radiology.143.1.7063747

Hardt, M., Recht, B., and Singer, Y. (2016). Train faster, generalize better: Stability of stochastic gradient descent. In *International Conference on Machine Learning*, pages 1225–1234. doi:10.48550/arXiv.1509.01240

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. pages 770–778. doi:10.1109/CVPR.2016.90

Hendrycks, D. and Gimpel, K. (2016a). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*. doi:10.48550/arXiv.1606.08415

Hendrycks, D. and Gimpel, K. (2016b). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*. doi:10.48550/arXiv.1606.08415

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Howard, J. and Ruder, S. (2018). Universal language model fine-tuning for text classification. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 328–339. doi:10.18653/v1/P18-1031

Huang, C. C., Chiang, H. F., Hsieh, C. C., Zhu, B. R., Wu, W. J., and Shaw, J. S. (2025). Impact of Dataset Size on 3D CNN Performance in Intracranial Hemorrhage Classification. *Diagnostics*, 15(2):1–14. doi:10.3390/diagnostics15020216

Huang, G., Sun, Y., Liu, Z., Sedra, D., and Weinberger, K. Q. (2016). Deep networks with stochastic depth. pages 646–661. doi:10.1007/978-3-319-46493-0_39

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning (ICML)*, pages 448–456. doi:10.48550/arXiv.1502.03167

Kaufman, S., Rosset, S., Perlich, C., and Stitelman, O. (2012). Leakage in data mining: Formulation, detection, and avoidance. *ACM Transactions on Knowledge Discovery from Data*, 6(4):1–21. doi:10.1145/2382577.2382579

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., and King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17:195. doi:10.1186/s12916-019-1382-x

Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., and Tang, P. T. P. (2017). On large-batch training for deep learning: Generalization gap and sharp minima. In *International Conference on Learning Representations (ICLR)*. doi:10.48550/arXiv.1609.04836

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. doi:10.48550/arXiv.1412.6980

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, volume 14, pages 1137–1145.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444. doi:10.1038/nature14539

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Lemaître, G., Nogueira, F., and Aridas, C. K. (2017). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5.

Li, S., Fevens, T., and Krzyzak, A. (2019). Multi-instance multi-scale CNN for medical image classification. *arXiv preprint arXiv:1907.02413*. doi:10.1007/978-3-030-32251-9_58

Li, W., Ruan, X., Yang, H., Zhang, S., Rui, F., and Xiong, J. (2025). Global, regional and national trends in the burden of intracranial hemorrhage, 1990–2021: Results from the global burden of disease study. *Heliyon*, 11. doi:10.1016/j.heliyon.2025.e42608

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988. doi:10.1109/ICCV.2017.324

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017a). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88. doi:10.1016/j.media.2017.07.005

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., and Sánchez, C. I. (2017b). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88. doi:10.1016/j.media.2017.07.005

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986. doi:10.1109/CVPR52688.2022.01167

Loshchilov, I. and Hutter, F. (2017). Sgdr: Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations (ICLR)*, pages 1436–1446. doi:10.48550/arXiv.1608.03983

Loshchilov, I. and Hutter, F. (2019). Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*. doi:10.48550/arXiv.1711.05101

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Masters, D. and Luschi, C. (2018). Revisiting small batch training for deep neural networks. *arXiv preprint arXiv:1804.07612*. doi:10.48550/arXiv.1804.07612

McGurgan, I. J., Ziai, W. C., Werring, D. J., Salman, R. A.-S., and Parry-Jones, A. R. (2021). Acute intracerebral haemorrhage: diagnosis and management. *Practical Neurology*, 21(2):128–136. doi:10.1136/practneurol-2020-002763

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157. doi:10.1007/BF02295996

Müller, R., Kornblith, S., and Hinton, G. E. (2019). When does label smoothing help? In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 4694–4703. doi:10.48550/arXiv.1906.02629

Mushtaq, M. F., Shahroz, M., Aseere, A. M., Shah, H., Majeed, R., Shehzad, D., and Samad, A. (2021). BHCNet: Neural Network-Based Brain Hemorrhage Classification Using Head CT Scan. *IEEE Access*, 9:113901–113916. doi:10.1109/ACCESS.2021.3102740

Neethi, A. S., Kannath, S. K., Kumar, A. A., Mathew, J., and Rajan, J. (2024). A comprehensive review and experimental comparison of deep learning methods for automated hemorrhage detection. *Engineering Applications of Artificial Intelligence*, 133(PC):108192. doi:10.1016/j.engappai.2024.108192

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359. doi:10.1109/TKDE.2009.191

Parry-Jones, A., Krishnamurthi, R., Ziai, W. C., Shoamanesh, A., Wu, S., Martins, S., and Anderson, C. S. (2024). World stroke organization (wso): Global intracerebral hemorrhage factsheet 2025. *International Journal of Stroke*, 20:145–150. doi:10.1177/17474930241307876

Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International Conference on Machine Learning (ICML)*, pages 1310–1318. doi:10.48550/arXiv.1211.5063

Pianykh, O. S. (2012). *Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide*. Springer, 2nd edition.

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Pranav, B. et al. (2022). Pooling in convolutional neural networks for medical image analysis: A survey and an empirical study. *Neural Computing and Applications*, 34:14437–14470. doi:10.1007/s00521-022-06953-8

Price, W. N. and Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, 25(1):37–43. doi:10.1038/s41591-018-0272-7

Radenović, F., Tolias, G., and Chum, O. (2018). Fine-tuning CNN image retrieval with no human annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1655–1668. doi:10.1109/TPAMI.2018.2846566

Raghu, M., Zhang, C., Kleinberg, J., and Bengio, S. (2019a). Transfusion: Understanding transfer learning for medical imaging. 32. doi:10.48550/arXiv.1902.07208

Raghu, M., Zhang, C., Kleinberg, J., and Bengio, S. (2019b). Transfusion: Understanding transfer learning for medical imaging. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, pages 3342–3352. doi:10.48550/arXiv.1902.07208

Romans, L. E. (2021). *Computed Tomography for Technologists: A Comprehensive Text*. Wolters Kluwer, 2nd edition.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. pages 4510–4520. doi:10.1109/CVPR.2018.00474

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48. doi:10.1186/s40537-019-0197-0

Shwartz-Ziv, R. and Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81:84–90. doi:10.1016/j.inffus.2021.11.011

Smith, L. N. and Topin, N. (2017). Super-convergence: Very fast training of neural networks using large learning rates. *arXiv preprint arXiv:1708.07747*. doi:10.48550/arXiv.1708.07747

Soun, J. e., Chow, D., Nagamine, M., Takhtawala, R., Filippi, C., Yu, W., and Chang, P. (2021). Artificial intelligence and acute stroke imaging. *American Journal of Neuroradiology*, 42(1):2–11. doi:10.3174/ajnr.A6883

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958. doi:10.5555/2627435.2670313

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312. doi:10.1109/TMI.2016.2535302

Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning (ICML)*, pages 6105–6114. doi:10.48550/arXiv.1905.11946

Tran, A. T., Zeevi, T., and Payabvash, S. (2025). Strategies to Improve the Robustness and Generalizability of Deep Learning Segmentation and Classification in Neuroimaging. pages 1–26. doi:10.3390/biomedinformatics5020020

Vakanski, A. and Xian, M. (2021). Generalization ability of deep learning models in medical image analysis. *Medical Image Computing*. doi:10.1109/MLSP52302.2021.9596501

Varoquaux, G. and Cheplygina, V. (2022). Machine learning for medical imaging: Methodological failures and recommendations for the future. *npj Digital Medicine*, 5:48. doi:10.1038/s41746-022-00592-y

Williams, K. S. (2024). Evaluations of artificial intelligence and machine learning algorithms in neurodiagnostics. doi:10.1152/jn.00404.2023

Wu, Y., Castro-Macías, F. M., Morales-Álvarez, P., et al. (2023). Smooth attention for deep multiple instance learning: Application to CT intracranial hemorrhage detection. *MICCAI*, pages 327–337.

Xia, J., Yin, Y., and Li, X. (2025). An Efficient Medical Image Classification Method Based on a Lightweight Improved ConvNeXt-Tiny Architecture. pages 1–12. doi:10.48550/arXiv.2508.11532

Yosinski, J., Clune, J., Bengio, Y., and Liphardt, H. (2014). How transferable are features in deep neural networks? *arXiv preprint arXiv:1411.1792*. doi:10.48550/arXiv.1411.1792

UNIVERSITAS
GADJAH MADA
**Evaluasi Performa dan Stabilitas Pembelajaran Arsitektur ConvNeXt untuk Klasifikasi Perdarahan Otak
pada Dataset CT Scan Terbatas**
Tengku Syaid Farhan, Afiahayati, S.Kom., M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Yunus, R. E. (2024). Radiology loading and coverage hours in indonesia. *Korean Journal of Radiology*, 25(7):597. doi:10.3348/kjr.2024.0267

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., and Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Medicine*, 15(11):e1002683. doi:10.1371/journal.pmed.1002683

Zhang, C., Bengio, S., Hardt, M., Recht, B., and Vinyals, O. (2017). Understanding deep learning requires rethinking generalization. doi:10.48550/arXiv.1611.03530

Zhang, H. and Chung, A. C. S. (2024). Depth-aware networks for multi-organ lesion detection in chest ct scans. *Bioengineering*, 11(10):998. doi:10.3390/bioengineering11100998