

- Computational Technologies (ICICCT), Apr. 2018, doi:10.1109/icicct.2018.8473291.
- [7] P. Jeevan, K. Viswanathan, and A. Sethi, "WaveMix-Lite: A Resource-efficient Neural Network for Image Analysis," May 2022.
- [8] Z. Dai, H. Liu, Q. V. Le, and M. Tan, "CoAtNet: Marrying Convolution and Attention for All Data Sizes," Jun. 2021.
- [9] T. S. Gunawan, A. F. R. Mohd Noor, and M. Kartiwi, "Development of English Handwritten Recognition Using Deep Neural Network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 2, p. 562, May 2018, doi: 10.11591/ijeecs.v10.i2.pp562-568.
- [10] P. Pad, S. Narduzzi, C. Kundig, E. Turetken, S. A. Bigdeli, and L. A. Dunbar, "Efficient Neural Vision Systems Based on Convolutional Image Acquisition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020.
- [11] A. Gesmundo and J. Dean, "An Evolutionary Approach to Dynamic Introduction of Tasks in Large-scale Multitask Learning Systems," May 2022.
- [12] A. Gesmundo, "A Continual Development Methodology for Large-scale Multitask Dynamic ML Systems," Sep. 2022.
- [13] B. Ma, X. Li, Y. Xia, and Y. Zhang, "Autonomous deep learning: A genetic DCNN designer for image classification," *Neurocomputing*, vol. 379, pp. 152–161, Feb. 2020, doi: 10.1016/j.neucom.2019.10.007.
- [14] H. M. D. Kabir *et al.*, "SpinalNet: Deep Neural Network with Gradual Input," Jul. 2020.
- [15] H. Taud and J. F. Mas, "Multilayer Perceptron (MLP)," *Lecture Notes in Geoinformation and Cartography*, pp. 451–455, Oct. 2017, doi:10.1007/978-3-319-60801-3_27.
- [16] Z.-H. Huang, W.-J. Li, J. Shang, J. Wang, and T. Zhang, "Non-uniform patch based face recognition via 2D-DWT," *Image and Vision Computing*, vol. 37, pp. 12–19, May 2015, doi:10.1016/j.imavis.2014.12.005.
- [17] A. Ghosh, B. Bhattacharya, and S. Basu Roy Chowdhury, "AdGAP: Advanced Global Average Pooling," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, Apr. 2018, doi:10.1609/aaai.v32i1.12154.
- [18] J. Ul Rahman, A. Ali, M. Ur Rehman, and R. Kazmi, "A Unit Softmax with Laplacian Smoothing Stochastic Gradient Descent for Deep Convolutional Neural Networks," *Intelligent Technologies and Applications*, pp. 162–174, 2020, doi: 10.1007/978-981-15-5232-8_14.
- [19] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–13, 2018, doi: 10.1155/2018/7068349.
- [20] K. Han *et al.*, "A Survey on Vision Transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 87–110, Jan. 2023, doi: 10.1109/tpami.2022.3152247.
- [21] K. Islam, "Recent Advances in Vision Transformer: A Survey and Outlook of Recent Work," Mar. 2022.
- [22] J. Wolfe, X. Jin, T. Bahr, and N. Holzer, "Application of Softmax Regression and Its Validation for Spectral-Based Land Cover Mapping," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-1/W1, pp. 455–459, May 2017, doi: 10.5194/isprs-archives-xlii-1-w1-455-2017.
- [23] X. Qi, T. Wang, and J. Liu, "Comparison of Support Vector Machine and Softmax Classifiers in Computer Vision," 2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Dec. 2017, doi: 10.1109/icmce.2017.49.
- [24] Y. Cheng, W. Liu, P. Duan, J. Liu, and T. Mei, "PyAnomaly: A Pytorch-based Toolkit for Video Anomaly Detection," *Proceedings of the 28th ACM International Conference on Multimedia*, Oct. 2020, doi: 10.1145/3394171.3414540.
- [25] R. L. Kumar, J. Kakarla, B. V. Isunuri, and M. Singh, "Multi-class brain tumor classification using residual network and global average pooling," *Multimedia Tools and Applications*, vol. 80, no. 9, pp. 13429–13438, Jan. 2021, doi: 10.1007/s11042-020-10335-4.
- [26] V. K. Chauhan, K. Dahiya, and A. Sharma, "Problem formulations and solvers in linear SVM: a review," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 803–855, Jan. 2018, doi: 10.1007/s10462-018-9614-6.
- [27] L. Breiman, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi:10.1023/a:1010933404324.
- [28] T. Chen and C. Guestrin, "XGBoost," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, doi: 10.1145/2939672.2939785.
- [29] W. Liu, J. Wei, and Q. Meng, "Comparisons on KNN, SVM, BP and the CNN for Handwritten Digit Recognition," 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), Aug. 2020, doi:10.1109/aeeca49918.2020.9213482.
- [30] P. C. Vashist, A. Pandey, and A. Tripathi, "A Comparative Study of Handwriting Recognition Techniques," 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), Jan. 2020, doi: 10.1109/iccakm46823.2020.9051464.
- [31] B. K. Nyaupane, R. K. Sah, and K. C. Dahal, "SVM, KNN, Random Forest, and Neural Network based Handwritten Nepali Barnamala Recognition," *Journal of Innovations in Engineering Education*, vol. 4, no. 2, pp. 64–70, Dec. 2021, doi: 10.3126/jiee.v4i2.38254.
- [32] B. Vamsi, D. Bhattacharyya, and D. Midhunchakkaravarthy, "Detection of Brain Stroke Based on the Family History Using Machine Learning Techniques," *Lecture Notes in Networks and Systems*, pp. 17–31, 2021, doi: 10.1007/978-981-16-1773-7_2.
- [33] O. M. Khandy and S. Dadvandipour, "Analysis of machine learning algorithms for character recognition: a case study on handwritten digit recognition," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, pp. 574–581, 2021.
- [34] A. Ibrahim Ahmed Osman, A. Najah Ahmed, M. F. Chow, Y. Feng Huang, and A. El-Shafie, "Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1545–1556, Jun. 2021, doi:10.1016/j.asej.2020.11.011.
- [35] A. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," *Interactive Learning Environments*, vol. 31, no. 6, pp. 3360–3379, May 2021, doi: 10.1080/10494820.2021.1928235.
- [36] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, "EMNIST: Extending MNIST to handwritten letters," 2017 International Joint Conference on Neural Networks (IJCNN), May 2017, doi:10.1109/ijcnn.2017.7966217.
- [37] R. M. Gower, N. Loizou, X. Qian, A. Sailanbayev, E. Shulgin, and P. Richtárik, "SGD: General Analysis and Improved Rates," *PMLR*, pp. 5200–5209, May 24, 2019. Accessed: Apr. 16, 2023. [Online]. Available: <https://proceedings.mlr.press/v97/qian19b.html>
- [38] S. Bock and M. Weis, "A Proof of Local Convergence for the Adam Optimizer," 2019 International Joint Conference on Neural Networks (IJCNN), Jul. 2019, doi: 10.1109/ijcnn.2019.8852239.
- [39] T.-T.-H. Le, J. Kim, and H. Kim, "An Effective Intrusion Detection Classifier Using Long Short-Term Memory with Gradient Descent Optimization," 2017 International Conference on Platform Technology and Service (PlatCon), Feb. 2017, doi:10.1109/platcon.2017.7883684.
- [40] N. S. Keskar and R. Socher, "Improving Generalization Performance by Switching from Adam to SGD," Dec. 2017.