## 7 Acknowledgments

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education (ORISE) for the DOE under contract number DE-SC0014664. This work was also supported by the Applied Mathematics activity within the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research, under Contract DE-AC02-06CH11357. We also acknowledge the support by the U.S. Department of Energy for the SciDAC 5 RAPIDS institute

Additionally, this research was supported in part by an appointment with the National Science Foundation (NSF) Mathematical Sciences Graduate Internship (MSGI) Program. This program is administered by ORISE through an interagency agreement between the DOE and NSF. ORISE is managed for the DOE by Oak Ridge Associated Universities (ORAU). All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of NSF, ORAU/ORISE or DOE.


## References


Ahsan, M. M.; Ali, M. S.; and Siddique, Z. 2024. Enhancing and improving the performance of imbalanced class data using novel GBO and SSG: A comparative analysis. *Neural Networks: The Official Journal of the International Neural Network Society*, 173: 106157.

Badawy, M.; El-Fishawy, N.; Radad, M.; and H.Arafa, A. 2022. RN-SMOTE: Reduced Noise SMOTE based on DBSCAN for enhancing imbalanced data classification. *Journal of King Saud University - Computer and Information Sciences*, 34.

Bae, S.-Y.; Lee, J.; Jeong, J.; Lim, C.; and Choi, J. 2021. Effective data-balancing methods for class-imbalanced genotoxicity datasets using machine learning algorithms and molecular fingerprints. *Computational Toxicology*, 20: 100178.

Batista, G. E. A. P. A.; Prati, R. C.; and Monard, M. C. 2004. A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor. Newsl.*, 6(1): 20–29.

Bishop, C. M. 1996. *Neural Networks for Pattern Recognition*. Oxford, England: Clarendon Press.

Bria, A.; Marrocco, C.; and Tortorella, F. 2020. Addressing class imbalance in deep learning for small lesion detection on medical images. *Computers in biology and medicine*, 120.

Brownlee, J. 2021. Tour of Evaluation Metrics for Imbalanced Classification. https://machinelearningmastery.com.

Cao, K.; Chen, Y.; Lu, J.; Arechiga, N.; Gaidon, A.; and Ma, T. 2021. Heteroskedastic and Imbalanced Deep Learning with Adaptive Regularization. ArXiv:2006.15766 [cs, stat].

Chabbouh, M.; Bechikh, S.; Mezura-Montes, E.; and Ben Said, L. 2023. Imbalanced multi-label data classification as a bi-level optimization problem: application to miRNA-related diseases diagnosis. *Neural Computing and Applications*, 35: 1–19.

Chang, N.; Yu, Z.; Wang, Y.-X.; Anandkumar, A.; Fidler, S.; and Alvarez, J. M. 2021. Image-Level or Object-Level? A Tale of Two Resampling Strategies for Long-Tailed Detection. ArXiv:2104.05702 [cs].

Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; and Kegelmeyer, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 321–357.

Cheah, P. C. Y.; Yang, Y.; and Lee, B. G. 2023. Enhancing Financial Fraud Detection through Addressing Class Imbalance Using Hybrid SMOTE-GAN Techniques. *International Journal of Financial Studies*, 11(3): 110. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Dang, W.; Yang, Z.; Dong, W.; Li, X.; and Shi, G. 2024. Inverse Weight-Balancing for Deep Long-Tailed Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(10): 11713–11721. Number: 10.

Das, S.; Mullick, S. S.; and Zelinka, I. 2022. On Supervised Class-Imbalanced Learning: An Updated Perspective and Some Key Challenges. *IEEE Transactions on Artificial Intelligence*, 3(6): 973–993. Conference Name: IEEE Transactions on Artificial Intelligence.

Fan, J.; Han, F.; and Liu, H. 2014. Challenges of big data analysis. *National science review*, 1(2): 293–314.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Networks. ArXiv:1406.2661 [cs, stat].

Guo, L.-Z.; and Li, Y.-F. 2022. Class-Imbalanced Semi-Supervised Learning with Adaptive Thresholding. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, 8082–8094. PMLR.

Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; and Bing, G. 2017. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73: 220–239.

Hammami, M.; Bechikh, S.; Louati, A.; Makhlouf, M.; and Said, L. B. 2020. Feature construction as a bi-level optimization problem. *Neural Computing and Applications*, 32(17): 13783–13804.

Han, H.; Wang, W.-Y.; and Mao, B.-H. 2005. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. In *Advances in Intelligent Computing*, volume 3644, 878–887. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-540-28226-6 978-3-540-31902-3. Series Title: Lecture Notes in Computer Science.

Johnson, J. M.; and Khoshgoftaar, T. M. 2019. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1): 27.

Kanika; and Singla, J. 2020. A Survey of Deep Learning based Online Transactions Fraud Detection Systems. In *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, 130–136.

Khan, S. H.; Hayat, M.; Bennamoun, M.; Sohel, F. A.; and Togneri, R. 2018. Cost-Sensitive Learning of Deep Feature Representations From Imbalanced Data. *IEEE Transactions on Neural Networks and Learning Systems*, 29(8): 3573–3587. Conference Name: IEEE Transactions on Neural Networks and Learning Systems.

Korkmaz, S. 2020. Deep Learning-Based Imbalanced Data Classification for Drug Discovery. *Journal of Chemical Information and Modeling*, 4180–4190.

Kovács, G. 2019. An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets. *Applied Soft Computing*, 83: 105662.

Kumar, P.; Bhatnagar, R.; Gaur, K.; and Bhatnagar, A. 2021. Classification of Imbalanced Data:Review of Methods and Applications. *IOP Conference Series: Materials Science and Engineering*, 1099(1): 012077.

Li, M.; Zhang, X.; Thrampoulidis, C.; Chen, J.; and Oymak, S. 2022. AutoBalance: Optimized Loss Functions for Imbalanced Data. ArXiv:2201.01212 [cs].

Liu, H.; Simonyan, K.; and Yang, Y. 2019. DARTS: Differentiable Architecture Search. ArXiv:1806.09055 [cs, stat].

Luketina, J.; Berglund, M.; Greff, K.; and Raiko, T. 2016. Scalable Gradient-Based Tuning of Continuous Regularization Hyperparameters. ArXiv:1511.06727 [cs].

Majeed, A.; and Hwang, S. O. 2023. CTGAN-MOS: Conditional Generative Adversarial Network Based Minority-Class-Augmented Oversampling Scheme for Imbalanced Problems. *IEEE Access*, 11: 85878–85899.

Makki, S.; Assaghir, Z.; Taher, Y.; Haque, R.; Hacid, M.-S.; and Zeineddine, H. 2019. An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection. *IEEE Access*, 7: 93010–93022.

Nguyen, H. M.; Cooper, E. W.; and Kamei, K. 2011. Borderline over-sampling for imbalanced data classification. *Int. J. Knowl. Eng. Soft Data Paradigm.*, 3(1): 4–21.

Pradipta, G. A.; Wardoyo, R.; Musdholifah, A.; and Sanjaya, I. N. H. 2021. Radius-SMOTE: A New Oversampling Technique of Minority Samples Based on Radius Distance for Learning From Imbalanced Data. *IEEE Access*, 9: 74763–74777.

Rosales-Pérez, A.; García, S.; and Herrera, F. 2023. Handling Imbalanced Classification Problems With Support Vector Machines via Evolutionary Bilevel Optimization. *IEEE Transactions on Cybernetics*, 53(8): 4735–4747. ArXiv:2204.10231 [cs].

Sharma, A.; Singh, P. K.; and Chandra, R. 2022. SMOTified-GAN for Class Imbalanced Pattern Classification Problems. *IEEE Access*, 10: 30655–30665. Conference Name: IEEE Access.

Singh, A.; Ranjan, R. K.; and Tiwari, A. 2022. Credit Card Fraud Detection under Extreme Imbalanced Data: A Comparative Study of Data-level Algorithms. *Journal of Experimental & Theoretical Artificial Intelligence*, 34(4): 571–598.

Song, H.; Kim, M.; and Lee, J.-G. 2023. Toward Robustness in Multi-label Classification: A Data Augmentation Strategy against Imbalance and Noise. ArXiv:2312.07087 [cs].

Strelcenia, E.; and Prakoonwit, S. 2023. A Survey on GAN Techniques for Data Augmentation to Address the Imbalanced Data Issues in Credit Card Fraud Detection. *Machine Learning and Knowledge Extraction*, 5(1): 304–329.

Yang, Y.; Akbarzadeh Khorshidi, H.; and Aickelin, U. 2023. A Diversity-Based Synthetic Oversampling Using Clustering for Handling Extreme Imbalance. *SN Computer Science*, 4(6): 848.

Zhang, H.; Cisse, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2018. mixup: Beyond Empirical Risk Minimization. ArXiv:1710.09412 [cs, stat].

Zhou, P.; Xiong, C.; Socher, R.; and Hoi, S. C. H. 2020. Theory-Inspired Path-Regularized Differential Network Architecture Search. In *Advances in Neural Information Processing Systems*, volume 33, 8296–8307. Curran Associates, Inc.