
BIBLIOGRAPHY

- [1] W. Pei, B. Xue, M. Zhang, L. Shang, X. Yao, and Q. Zhang, “A Survey on Unbalanced Classification: How Can Evolutionary Computation Help?” *IEEE Transactions on Evolutionary Computation*, vol. 28, no. 2, pp. 353–373, Apr. 2024.
- [2] J. Gong, “A Novel Oversampling Technique for Imbalanced Learning Based on SMOTE and Genetic Algorithm,” in *Neural Information Processing*, T. Mantoro, M. Lee, M. A. Ayu, K. W. Wong, and A. N. Hidayanto, Eds. Cham: Springer International Publishing, 2021, vol. 13110, pp. 201–212.
- [3] F. Ferreira, N. Lourenco, B. Cabral, and J. P. Fernandes, “When Two are Better Than One: Synthesizing Heavily Unbalanced Data,” *IEEE Access*, vol. 9, pp. 150 459–150 469, 2021.
- [4] Nurul Chamidah, Mayanda Mega Santoni, and Nurhafifah Matondang, “Oversampling Method on Classifying Hypertension Using Naive Bayes, Decision Tree, and Artificial Neural Network,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 4, pp. 635–641, Aug. 2020.
- [5] R. Obiedat, R. Qaddoura, A. M. Al-Zoubi, L. Al-Qaisi, O. Harfoushi, M. Alrefai, and H. Faris, “Sentiment Analysis of Customers’ Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution,” *IEEE Access*, vol. 10, pp. 22 260–22 273, 2022.
- [6] Y. Pang, Z. Chen, L. Peng, K. Ma, C. Zhao, and K. Ji, “A Signature-Based Assistant Random Oversampling Method for Malware Detection,” in *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. Rotorua, New Zealand: IEEE, Aug. 2019, pp. 256–263.
- [7] D. Javale, P. Pillai, P. Patel, and S. Jagtap, “A Comparison of Oversampling and Undersampling Methods for Predicting Air Quality in Metropolitan Region,” in *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*. Salem, India: IEEE, May 2022, pp. 1615–1620.
- [8] L.-S. Lin, Z.-Y. Chen, Y. Wang, and L.-W. Jiang, “Improving Anomaly Detection in IoT-Based Solar Energy System Using SMOTE-PSO and SVM Model,” in *Frontiers in Artificial Intelligence and Applications*, J.-L. Kim, Ed. IOS Press, Nov. 2022.
- [9] C. Yang, E. A. Fridgeirsson, J. A. Kors, J. M. Reys, and P. R. Rijnbeek, “Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data,” *Journal of Big Data*, vol. 11, no. 1, p. 7, Jan. 2024.

- [10] H. Shamsudin, U. K. Yusof, Y. Haijie, and I. S. Isa, “AN OPTIMIZED SUPPORT VECTOR MACHINE WITH GENETIC ALGORITHM FOR IMBALANCED DATA CLASSIFICATION,” *Jurnal Teknologi*, vol. 85, no. 4, pp. 67–74, Jun. 2023.
- [11] Rifky, P. D. Pamungkasari, and E. T. E. Handayani, “Performance Analysis of Facial Expression Recognition by Using Geometry Augmentation and Random Oversampling for CNN Model,” in *2024 10th International Conference on Smart Computing and Communication (ICSCC)*. Bali, Indonesia: IEEE, Jul. 2024, pp. 614–618.
- [12] D. Devi, S. K. Biswas, and B. Purkayastha, “A Review on Solution to Class Imbalance Problem: Undersampling Approaches,” in *2020 International Conference on Computational Performance Evaluation (ComPE)*. Shillong, India: IEEE, Jul. 2020, pp. 626–631.
- [13] Y. Zhu, Y. Yan, Y. Zhang, and Y. Zhang, “EHSO: Evolutionary Hybrid Sampling in overlapping scenarios for imbalanced learning,” *Neurocomputing*, vol. 417, pp. 333–346, Dec. 2020.
- [14] Z. S. Rubaidi, B. B. Ammar, and M. B. Aouicha, “Fraud Detection Using Large-scale Imbalance Dataset,” *International Journal on Artificial Intelligence Tools*, vol. 31, no. 08, p. 2250037, Dec. 2022.
- [15] A. Martinez-Velasco, L. Martínez -Villaseñor, and L. Miralles-Pechuán, “Addressing Class Imbalance in Healthcare Data: Machine Learning Solutions for Age-Related Macular Degeneration and Preeclampsia,” *IEEE Latin America Transactions*, vol. 22, no. 10, pp. 806–820, Oct. 2024.
- [16] M. H. Kotb and R. Ming, “Comparing SMOTE Family Techniques in Predicting Insurance Premium Defaulting using Machine Learning Models,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, 2021.
- [17] S. N. Almuayqil, M. Humayun, N. Z. Jhanjhi, M. F. Almufareh, and D. Javed, “Framework for Improved Sentiment Analysis via Random Minority Oversampling for User Tweet Review Classification,” *Electronics*, vol. 11, no. 19, p. 3058, Sep. 2022.
- [18] B. Muhibullaev and J. Kim, “Accurate Information Type Classification for Software Issue Discussions With Random Oversampling,” *IEEE Access*, vol. 12, pp. 65 373–65 385, 2024.
- [19] M. Saripuddin, A. Suliman, S. Syarmila Sameon, and B. N. Jorgensen, “Random Undersampling on Imbalance Time Series Data for Anomaly Detection,” in *2021 The 4th International Conference on Machine Learning and Machine Intelligence*. Hangzhou China: ACM, Sep. 2021, pp. 151–156.

- [20] I. Binanto, R. A. Kumalasanti, and N. F. Sianipar, "Handling of unbalanced LC-MS medicinal plant data using Near-Miss Undersampling tested with Gaussian Naive Bayes and K-Nearest Neighbors," in *2023 2nd International Conference on Computational Modelling, Simulation and Optimization (ICCMSO)*. Bali, Indonesia: IEEE, Jun. 2023, pp. 120–124.
- [21] P. Mittal, H. S. Lallie, and E. Titis, "Impact of imbalanced datasets on ML algorithms for malware classification," *Information Security Journal: A Global Perspective*, vol. 34, no. 3, pp. 251–264, May 2025.
- [22] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection," *IEEE Access*, vol. 10, pp. 16 400–16 407, 2022.
- [23] M. Masum, H. Shahriar, and H. Haddad, "Analysis of Sampling Techniques Towards Epileptic Seizure Detection from Imbalanced Dataset," in *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. Madrid, Spain: IEEE, Jul. 2020, pp. 684–692.
- [24] R. Kannan, H. Ng, T. T. V. Yap, L. K. Wong, F. F. Chua, V. T. Goh, Y. L. Lee, and H. L. Wong, "Handling class imbalance in education using data-level and deep learning methods," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 15, no. 1, p. 741, Feb. 2025.
- [25] W. Yee Wong, K. Hasikin, A. Salwa Mohd Khairuddin, S. Abdul Razak, H. Farzana Hizaddin, M. Istajib Mokhtar, and M. Mokhzaini Azizan, "A Stacked Ensemble Deep Learning Approach for Imbalanced Multi-Class Water Quality Index Prediction," *Computers, Materials & Continua*, vol. 76, no. 2, pp. 1361–1384, 2023.
- [26] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, Aug. 2016.
- [27] H. L. Le, D. Landa-Silva, M. Galar, S. Garcia, and I. Triguero, "A Hybrid Surrogate Model for Evolutionary Undersampling in Imbalanced Classification," in *2020 IEEE Congress on Evolutionary Computation (CEC)*. Glasgow, United Kingdom: IEEE, Jul. 2020, pp. 1–8.
- [28] G. Yang and L. Qicheng, "An Over Sampling Method of Unbalanced Data Based on Ant Colony Clustering," *IEEE Access*, vol. 9, pp. 130 990–130 996, 2021.
- [29] F. Sağlam, M. SÖZEN, and M. A. CENGİZ, "Optimization Based Undersampling for Imbalanced Classes," *Adıyaman University Journal of Science*, pp. 385–409, Oct. 2021.

- [30] S. Suyanto, A. A. Ariyanto, and A. F. Ariyanto, “Komodo Mlipir Algorithm,” *Applied Soft Computing*, vol. 114, p. 108043, Jan. 2022.
- [31] H. Alsawalqah, N. Hijazi, M. Eshtay, H. Faris, A. A. Radaideh, I. Aljarah, and Y. Alshamaileh, “Software Defect Prediction Using Heterogeneous Ensemble Classification Based on Segmented Patterns,” *Applied Sciences*, vol. 10, no. 5, p. 1745, Mar. 2020.
- [32] K. Saengtabtim, N. Leelawat, J. Tang, W. Treeranurat, N. Wisittiwong, A. Suppasri, K. Pakoksung, F. Imamura, N. Takahashi, and I. Charvet, “Predictive Analysis of the Building Damage From the 2011 Great East Japan Tsunami Using Decision Tree Classification Related Algorithms,” *IEEE Access*, vol. 9, pp. 31 065–31 077, 2021.
- [33] N. A. A. Khleel, K. Nehéz, M. Fadulalla, and A. Hisaen, “Ensemble-Based Machine Learning Algorithms Combined with Near Miss Method for Software Bug Prediction,” *International Journal of Networked and Distributed Computing*, vol. 13, no. 1, p. 11, Jun. 2025.
- [34] V. Geetha, A. Punitha, M. Abarna, M. Akshaya, S. Illakiya, and A. Janani, “An Effective Crop Prediction Using Random Forest Algorithm,” in *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*. Pondicherry, India: IEEE, Jul. 2020, pp. 1–5.
- [35] G. Ryan, P. Katarina, and D. Suhartono, “MBTI Personality Prediction Using Machine Learning and SMOTE for Balancing Data Based on Statement Sentences,” *Information*, vol. 14, no. 4, p. 217, Apr. 2023.
- [36] N. M. Nayan, A. Islam, M. U. Islam, E. Ahmed, M. M. Hossain, and M. Z. Alam, “SMOTE Oversampling and Near Miss Undersampling Based Diabetes Diagnosis from Imbalanced Dataset with XAI Visualization,” in *2023 IEEE Symposium on Computers and Communications (ISCC)*. Gammarth, Tunisia: IEEE, Jul. 2023, pp. 1–6.
- [37] H.-T. Duong and T.-A. Nguyen-Thi, “A review: Preprocessing techniques and data augmentation for sentiment analysis,” *Computational Social Networks*, vol. 8, no. 1, p. 1, Dec. 2021.
- [38] S. Ouf, K. T. Mahmoud, and M. A. Abdel-Fattah, “A proposed hybrid framework to improve the accuracy of customer churn prediction in telecom industry,” *Journal of Big Data*, vol. 11, no. 1, p. 70, May 2024.
- [39] H. Tsai, T.-W. Yang, W.-M. Wong, H.-Y. Kao, and C.-F. Chou, “A Hybrid Approach for Binary Classification of Imbalanced Data,” *International Journal of Computational Intelligence and Applications*, vol. 23, no. 03, p. 2450013, Sep. 2024.

- [40] B. Arkok, A. M. Zeki, R. Othman, and A. Aborujilah, “An Extended Oversampling Method for Imbalanced Quranic Text Classification based on A Genetic Algorithm,” no. 4, 2023.
- [41] S. J. J. Guesné, T. Hanser, S. Werner, S. Boobier, and S. Scott, “Mind your prevalence!” *Journal of Cheminformatics*, vol. 16, no. 1, p. 43, Apr. 2024.
- [42] J. Alcalá-Fdez, L. Sánchez, S. García, M. J. Del Jesus, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas, J. C. Fernández, and F. Herrera, “KEEL: A software tool to assess evolutionary algorithms for data mining problems,” vol. 13, no. 3, pp. 307–318. [Online]. Available: <http://link.springer.com/10.1007/s00500-008-0323-y>
- [43] W. Nugraha, M. S. Maulana, and A. Sasongko, “Clustering Based Undersampling for Handling Class Imbalance in C4.5 Classification Algorithm,” *Journal of Physics*, 2020.
- [44] M. Diallo, S. Xiong, E. D. Emiru, A. Fesseha, A. O. Abdulsalami, and M. A. Elaziz, “A Hybrid MultiLayer Perceptron Under-Sampling with Bagging Dealing with a Real-Life Imbalanced Rice Dataset,” *information*, vol. 12, pp. 1–21, 2021.
- [45] Z. Yu, T. D. Nguyen, L. Ju, Y. Gal, M. Sashindranath, P. Bonnington, L. Zhang, V. Mar, and Z. Ge, “Hierarchical skin lesion image classification with prototypical decision tree,” *Digital Medicine*, vol. 8, pp. 1–15, Jan. 2025.
- [46] M. S. Sawah, H. Elmannai, A. A. El-Bary, K. Lotfy, and O. E. Sheta, “Distributed denial of service (DDoS),” *Scientific Reports*, vol. 15, pp. 1–20, May 2025.
- [47] A. Chatterjee, M. A. Riegler, K. Ganesh, and P. Halvorsen, “Stress management with HRV following AI, semantic ontology, genetic algorithm and tree explainer,” *Scientific Reports*, vol. 15, pp. 1–25, Feb. 2015.
- [48] J. Lee, B. Bartholmai, T. Peikert, J. Chun, H. Kim, J. S. Kim, and S. Y. Park, “Evaluation of Computer-Aided Nodule Assessment and Risk Yield (CANARY) in Korean patients for prediction of invasiveness of ground-glass opacity nodule,” *PLOS ONE*, vol. 16, no. 6, p. e0253204, Jun. 2021.
- [49] S. Alfasly, G. Alabtah, S. Hemati, K. R. Kalari, J. J. Garcia, and H. R. Tizhoosh, “Validation of histopathology foundation models through whole slide image retrieval,” *Scientific Reports*, vol. 15, no. 1, Feb. 2025.
- [50] T. Hasanin and T. Khoshgoftaar, “The Effects of Random Undersampling with Simulated Class Imbalance for Big Data,” in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*. IEEE, pp. 70–79. [Online]. Available: <https://ieeexplore.ieee.org/document/8424689/>

-
- [51] E. F. Swana, W. Doorsamy, and P. Bokoro, "Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset," *Sensors*, vol. 22, no. 9, p. 3246, Apr. 2022.
- [52] D. Sisodia and D. S. Sisodia, "Quad division prototype selection-based k-nearest neighbor classifier for click fraud detection from highly skewed user click dataset," vol. 28, p. 101011. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2215098621001233>
- [53] F. Yang, K. Wang, L. Sun, M. Zhai, J. Song, and H. Wang, "A hybrid sampling algorithm combining synthetic minority over-sampling technique and edited nearest neighbor for missed abortion diagnosis," vol. 22, no. 1, p. 344. [Online]. Available: <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-022-02075-2>