IV. Conclusion

In this study, we addressed the problem of data imbalance in the ACS classification case by using the K-Means SMOTE algorithm to oversample the training data. Our simulations showed that all models built using K-Means SMOTE oversampling data increased F1 scores in all scenarios, with an average increase of 10.07%. We also compared the performance of other oversampling algorithms and found that K-Means SMOTE had the most significant increase in F1 scores.

Our study's findings suggest that oversampling algorithms can improve the output of machine learning models on imbalanced ACS datasets. However, we acknowledge that our research has some limitations, such as using only one dataset and an oversampling algorithm. Therefore, future research could explore other oversampling algorithms, feature engineering processes, and advanced machine learning algorithms to improve the output of these models further.

In conclusion, our research provides insight into the use of oversampling algorithms to address data imbalance in the ACS classification case. Our findings can be used as a foundation for future research to improve the output of machine-learning models on imbalanced ACS datasets.

Acknowledgment

This work was supported by the PPMI KSE Research Group grant, School of Electrical Engineering and Informatics, Institut Teknologi Bandung (grant number 799a/IT1.C12/KU/2022). We thank the reviewers for their valuable comments and suggestions to improve our manuscript.

References

[1] E. A. Amsterdam et al., “2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines,” Circulation, vol. 130, no. 25, pp. 2354–2394, 2014.

[2] P. Libby, G. Pasterkamp, F. Crea, and I. K. Jang, “Reassessing the Mechanisms of Acute Coronary Syndromes: The ‘vulnerable Plaque’ and Superficial Erosion,” Circulation Research, vol. 124, no. 1, pp. 150–160, 2019.

[3] N. Makki, T. M. Brennan, and S. Girotra, “Acute coronary syndrome,” Journal of Intensive Care Medicine, vol. 30, no. 4, pp. 186–200, 2015.

[4] E. A. Dziedzic, J. S. Gasiorą, A. Tuzimek, M. Dabrowskią, and P. Jankowski, “Neutrophil-to-Lymphocyte Ratio Is Not Associated with Severity of Coronary Artery Disease and Is Not Correlated with Vitamin D Level in Patients with a History of an Acute Coronary Syndrome,” Biology, vol. 11, no. 7, pp. 1–12, 2022.

[5] P. A. Iannattone, X. Zhao, J. VanHouten, A. Garg, and T. Huynh, “Artificial Intelligence for Diagnosis of Acute Coronary Syndromes: A Meta-analysis of Machine Learning Approaches,” Canadian Journal of Cardiology, vol. 36, no. 4, pp. 577–583, 2020.

[6] M. F. Muzakki, J. A. Utama, R. Priyatikanto, and L. S. Riza, “Detection System of Solar Flare Occurrence in PROBA2 SWAP Images Using Seeded Region Growing and Machine Learning,” vol. 62, no. 07, pp. 3329–3342, 2020.

[7] W. G. Baxt, F. S. Shofer, F. D. Sites, and J. E. Hollander, “A neural network aid for the early diagnosis of cardiac ischemia in patients presenting to the emergency department with chest pain,” Annals of Emergency Medicine, vol. 40, no. 6, pp. 575–583, 2002.

[8] A. M. Bulgiba and M. Razaz, “How well can signs and symptoms predict AMI in the Malaysian population?,” International Journal of Cardiology, vol. 102, no. 1, pp. 87–93, 2005.

[9] E. P. Cynthia, M. Afif Rizky A., A. Nazir, and F. Syafria, “Random Forest Algorithm to Investigate the Case of Acute Coronary Syndrome,” Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 2, pp. 369–378, 2021.

[10] S. Calderon-ramirez et al., “Correcting data imbalance for semi-supervised COVID-19 detection using X-ray chest images,” Applied Soft Computing, vol. 111, p. 107692, 2021.

[11] V. Karia, W. Zhang, A. Naeim, and R. Ramezani, “Gensample: A genetic algorithm for oversampling in imbalanced datasets,” arXiv preprint arXiv:1910.10806, 2019.

[12] R. Mohammed, J. Rawashdeh, and M. Abdullah, “Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results,” 2020 11th International Conference on Information and Communication Systems, ICICS 2020, pp. 243–248, 2020.

[13] F. Last, G. Douzas, and F. Bacao, “Oversampling for Imbalanced Learning Based on K-Means and SMOTE,” pp. 1–19, 2017.

[14] X. W. Liang, A. P. Jiang, T. Li, Y. Y. Xue, and G. T. Wang, “LR-SMOTE—An improved unbalanced data set oversampling based on K-means and SVM,” Knowledge-Based Systems, vol. 196, p. 105845, 2020.

[15] Q. Wang, L. Li, B. Jiang, Z. Lu, J. Liu, and S. Jian, “Malicious domain detection based on k-means and smote,” in International Conference on Computational Science, 2020, pp. 468–481.

[16] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” Journal of Artificial Intelligence, vol. 16, pp. 321–357, 2002.

[17] L. Breiman, “Random forests,” Machine learning, vol. 45, no. 1, pp. 5–32, 2001.

[18] K. Zhang, X. Wu, R. Niu, K. Yang, and L. Zhao, “The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China,” Environmental Earth Sciences, vol. 76, no. 11, 2017.

[19] R. G. Leiva, A. F. Anta, V. Mancuso, and P. Casari, “A novel hyperparameter-free approach to decision tree construction that avoids overfitting by design,” IEEE Access, vol. 7, pp. 99978–99987, 2019.

[20] M. Tschannen, O. Bachem, and M. Lucic, “Recent Advances in Autoencoder-Based Representation Learning,” no. NeurIPS, pp. 1–25, 2018.

[21] W. Xu and Y. Tan, “Semisupervised Text Classification by Variational Autoencoder,” IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 1, pp. 295–308, 2020.

[22] S. Nembrini, I. R. König, and M. N. Wright, “The revival of the Gini importance?,” Bioinformatics, vol. 34, no. 21, pp. 3711–3718, 2018.

[23] H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning,” in 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), 2008, pp. 1322–1328.

[24] H. Lee, J. Kim, and S. Kim, “Gaussian-based SMOTE algorithm for solving skewed class distributions,” International Journal of Fuzzy Logic and Intelligent Systems, vol. 17, no. 4, pp. 229–234, 2017.

[25] L. Ma and S. Fan, “CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests,” BMC bioinformatics, vol. 18, no. 1, pp. 1–18, 2017.

[26] F. R. Torres, J. A. Carrasco-Ochoa, and J. F. Mart\'\inez-Trinidad, “SMOTE-D a deterministic version of SMOTE,” in Mexican Conference on Pattern Recognition, 2016, pp. 177–188.

[27] H. Han, W.-Y. Wang, and B.-H. Mao, “Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning,” in international conference on intelligent computing, pp. 878–887, 2005.

[28] S. H. Ha and S. H. Joo, “A hybrid data mining method for the medical classification of chest pain,” International Journal of Computer and Information Engineering, vol. 4, no. 1, pp. 99–104, 2010.

[29] G. B. Berikol, O. Yildiz, and T. Özcan, “Diagnosis of Acute Coronary Syndrome with a Support Vector Machine,” Journal of Medical Systems, vol. 40, no. 4, pp. 1–8, 2016.

[30] R. D. Prayogo and S. A. Karimah, “Feature Selection and Adaptive Synthetic Sampling Approach for Optimizing Online Shopper Purchase Intent Prediction,” 2021.

[31] M. P. Perme and D. Manevski, “Confidence intervals for the Mann–Whitney test,” Statistical Methods in Medical Research, vol. 28, no. 12, pp. 3755–3768, 2019.