

DAFTAR PUSTAKA

- [1] K. Kesehatan, “Hasil Riset Kesehatan Dasar 2018,” *Repository BKKP Kemkes*, 30 Agustus 2018. [Online]. Available: <http://repository.bkpk.kemkes.go.id/3514/1/Laporan%20Riskesdas%202018%20Nasional.pdf>. [Accessed: 12 Desember 2022].
- [2] Kementerian Kesehatan Republik Indonesia, “Survei Kesehatan Indonesia Dalam Angka, Data Akurat Kebijakan Tepat,” pp. 223–279, 2023.
- [3] H. Yuk, J. Gim, J. K. Min, J. Yun, and T.-Y. Heo, “Artificial Intelligence-based Prediction of Diabetes and Prediabetes Using Health Checkup Data in Korea,” *Appl. Artif. Intell.*, vol. 36, no. 1, p. 2145644, Dec. 2022.
- [4] S. Peñafiel *et al.*, “Associating risks of getting strokes with data from health checkup records using Dempster-Shafer Theory,” in *2018 20th International Conference on Advanced Communication Technology (ICACT)*, 2018, pp. 239–246.
- [5] Z. Xie, O. Nikolayeva, J. Luo, and D. Li, “Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques,” *Prev. Chronic Dis.*, vol. 16, p. E130, Sep. 2019.
- [6] Y. Choi, J. An, S. Ryu, and J. Kim, “Development and Evaluation of Machine Learning-Based High-Cost Prediction Model Using Health Check-Up Data by the National Health Insurance Service of Korea,” *Int. J. Environ. Res. Public Health*, vol. 19, no. 20, Oct. 2022.
- [7] S. Lee, E. K. Choe, and B. Park, “Exploration of Machine Learning for Hyperuricemia Prediction Models Based on Basic Health Checkup Tests,” *J. Clin. Med. Res.*, vol. 8, no. 2, Feb. 2019.
- [8] M. Tsunekawa, N. Oka, M. Araki, M. Shintani, M. Yoshikawa, and T. Tanigawa, “Prediction of Onset of Lifestyle-Related Diseases Using Regular Health Checkup Data,” in *Advances in Artificial Intelligence*, 2020, pp. 14–26.
- [9] M. Takahara, N. Katakami, H. Kaneto, M. Noguchi, and I. Shimomura, “Prediction of the presence of insulin resistance using general health checkup data in Japanese employees with metabolic risk factors,” *J. Atheroscler. Thromb.*, vol. 21, no. 1, pp. 38–48, 2014.
- [10] M. Kawasoe *et al.*, “Development of a risk prediction score for hypertension incidence using Japanese health checkup data,” *Hypertens. Res.*, vol. 45, no. 4, pp. 730–740, Apr. 2022.
- [11] M. Morimoto, N. Kamiura, Y. Hata, and I. Yamamoto, “Incidence rate prediction of diabetes from medical checkup data,” *IEICE Trans. Inf. Syst.*, vol. E100.D, no. 8, pp. 1642–1646, 2017.
- [12] D. Ichikawa, T. Saito, and H. Oyama, “Impact of predicting health-guidance candidates using massive health check-up data: A data-driven analysis,” *Int. J. Med. Inform.*, vol. 106, pp. 32–36, Oct. 2017.
- [13] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, “Development of disease prediction model based on ensemble learning approach for diabetes and hypertension,” *IEEE Access*, vol. 7, pp. 144777–144789, 2019.
- [14] H. Qi *et al.*, “Prediction methods of common cancers in china using PCA-ANN and DBN-ELM-BP,” *IEEE Access*, vol. 10, pp. 113397–113409, 2022.
- [15] E. Dritsas and M. Trigka, “Stroke Risk Prediction with Machine Learning Techniques,” *Sensors*, vol. 22, no. 13, Jun. 2022.

- [16] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, and D. John, "A predictive analytics approach for stroke prediction using machine learning and neural networks," *Healthcare Analytics*, vol. 2, p. 100032, Nov. 2022.
- [17] I. Chourib, G. Guillard, I. R. Farah, and B. Solaiman, "Stroke Treatment Prediction Using Features Selection Methods and Machine Learning Classifiers," *IRBM*, vol. 43, no. 6, pp. 678–686, Dec. 2022.
- [18] K. M. Almustaafa, "Prediction of chronic kidney disease using different classification algorithms," *Informatics in Medicine Unlocked*, vol. 24, p. 100631, Jan. 2021.
- [19] S. P. Potharaju, Dept of CSE , K L Universit y, Guntur, Andhra Pradesh, India, M. Sreedevi, and Dept of CSE , K L Universit y, Guntur, Andhra Pradesh, India, "Ensembled rule based classification algorithms for predicting imbalanced kidney disease data," *J. Eng. Sci. Technol. Rev.*, vol. 9, no. 5, pp. 201–207, Oct. 2016.
- [20] M. Almasoud and T. E. Ward, "Detection of chronic kidney disease using machine learning algorithms with least number of predictors," *International Journal of Soft Computing and Its Applications*, vol. 10, no. 8, Sep. 2019.
- [21] P. Moghadam and A. Ahmadi, "A machine learning framework to predict kidney graft failure with class imbalance using Red Deer algorithm," *Expert Syst. Appl.*, vol. 210, p. 118515, Dec. 2022.
- [22] Z. Ullah *et al.*, "Detecting high-risk factors and early diagnosis of diabetes using machine learning methods," *Comput. Intell. Neurosci.*, vol. 2022, p. 2557795, Sep. 2022.
- [23] B. Kaur and G. Kaur, "Heart Disease Prediction Using Modified Machine Learning Algorithm," in *International Conference on Innovative Computing and Communications*, 2023, pp. 189–201.
- [24] M. Rizwan, S. Arshad, H. Aijaz, R. A. Khan, and M. Zeeshan Ul Haque, "Heart Attack Prediction using Machine Learning Approach," in *2022 Third International Conference on Latest trends in Electrical Engineering and Computing Technologies (INTELLECT)*, 2022, pp. 1–8.
- [25] C. Sirocchi, A. Bogliolo, and S. Montagna, "Medical-informed machine learning: integrating prior knowledge into medical decision systems," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. Suppl 4, p. 186, Jun. 2024.
- [26] Y. Su, C. Huang, W. Yin, X. Lyu, L. Ma, and Z. Tao, "Diabetes mellitus risk prediction using age adaptation models," *Biomed. Signal Process. Control*, vol. 80, p. 104381, Feb. 2023.
- [27] J. A. Castellanos-Garzón, E. Costa, J. L. Jaimes S., and J. M. Corchado, "An evolutionary framework for machine learning applied to medical data," *Knowledge-Based Systems*, vol. 185, p. 104982, Dec. 2019.
- [28] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques," *Front. Genet.*, vol. 9, p. 515, Nov. 2018.
- [29] A. Amina, Y. Ali, M. Awais, and K. Zaheer, "Data mining models comparison for diabetes prediction," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, pp. 320–323, 2018.
- [30] H. Wu, S. Yang, Z. Huang, J. He, and X. Wang, "Type 2 diabetes mellitus prediction model based on data mining," *Informatics in Medicine Unlocked*, vol. 10, pp. 100–107, Jan. 2018.

- [31] M. Maniruzzaman *et al.*, “Accurate Diabetes Risk Stratification Using Machine Learning: Role of Missing Value and Outliers,” *J. Med. Syst.*, vol. 42, no. 5, p. 92, Apr. 2018.
- [32] N. G. Ramadhan, Adiwijaya, and A. Romadhony, “Preprocessing handling to enhance detection of type 2 diabetes mellitus based on random forest,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, pp. 223–228, 2021.
- [33] F. López-Martínez, E. R. Núñez-Valdez, R. G. Crespo, and V. García-Díaz, “An artificial neural network approach for predicting hypertension using NHANES data,” *Sci. Rep.*, vol. 10, no. 1, p. 10620, Jun. 2020.
- [34] Q. Wang, W. Cao, J. Guo, J. Ren, Y. Cheng, and D. N. Davis, “DMP_MI: An Effective Diabetes Mellitus Classification Algorithm on Imbalanced Data With Missing Values,” *IEEE Access*, vol. 7, pp. 102232–102238, 2019.
- [35] T. Cai, “Breast cancer diagnosis using imbalanced learning and ensemble method,” *Appl. Comput. Math.*, vol. 7, no. 3, p. 146, 2018.
- [36] Y. Xiao, J. Wu, and Z. Lin, “Cancer diagnosis using generative adversarial networks based on deep learning from imbalanced data,” *Comput. Biol. Med.*, vol. 135, p. 104540, Aug. 2021.
- [37] S. Nazah, S. Huda, J. Abawajy, and M. M. Hassan, “Evolution of dark web threat analysis and detection: A systematic approach,” *IEEE Access*, vol. 8, pp. 171796–171819, 2020.
- [38] P. Saint-Louis, M. C. Morency, and J. Lapalme, “Defining enterprise architecture: A systematic literature review,” in *2017 IEEE 21st International Enterprise Distributed Object Computing Workshop (EDOCW)*, 2017, pp. 41–49.
- [39] C. Fisch and J. Block, “Six tips for your (systematic) literature review in business and management research,” *Manag. Rev. Q.*, vol. 68, no. 2, pp. 103–106, Apr. 2018.
- [40] N. G. Ramadhan, Adiwijaya, W. Maharani, and A. Akbar Gozali, “Chronic diseases prediction using machine learning with data preprocessing handling: a critical review,” *IEEE Access*, vol. 12, pp. 80698–80730, 2024.
- [41] F. Shokraneh and C. E. Adams, “Increasing value and reducing waste in data extraction for systematic reviews: tracking data in data extraction forms,” *Syst. Rev.*, vol. 6, no. 1, p. 153, Aug. 2017.
- [42] L. Yang *et al.*, “Quality assessment in systematic literature reviews: A software engineering perspective,” *Inf. Softw. Technol.*, vol. 130, no. 106397, p. 106397, Feb. 2021.
- [43] K. Mahboob, S. A. Ali, and U.-E. Laila, “Investigating learning outcomes in engineering education with data mining,” *Comput. Appl. Eng. Educ.*, vol. 28, no. 6, pp. 1652–1670, Nov. 2020.
- [44] V. H. Bhat, P. G. Rao, S. Krishna, P. D. Shenoy, K. R. Venugopal, and L. M. Patnaik, “An efficient framework for prediction in healthcare data using soft computing techniques,” in *Advances in Computing and Communications*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 522–532.
- [45] I. Kadi, A. Idri, and J. L. Fernandez-Aleman, “Knowledge discovery in cardiology: A systematic literature review,” *Int. J. Med. Inform.*, vol. 97, pp. 12–32, Jan. 2017.
- [46] M. Albayrak, K. Turhan, and B. Kurt, “A missing data imputation approach using clustering and maximum likelihood estimation,” in *2017 Medical Technologies National Congress (TIPTEKNO)*, 2017, pp. 1–4.

- [47] W.-C. Lin and C.-F. Tsai, "Missing value imputation: a review and analysis of the literature (2006–2017)," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 1487–1509, Feb. 2020.
- [48] T. Aittokallio, "Dealing with missing values in large-scale studies: microarray data imputation and beyond," *Brief. Bioinform.*, vol. 11, no. 2, pp. 253–264, Mar. 2010.
- [49] P. J. García-Laencina, J.-L. Sancho-Gómez, and A. R. Figueiras-Vidal, "Pattern classification with missing data: a review," *Neural Comput. Appl.*, vol. 19, no. 2, pp. 263–282, Mar. 2010.
- [50] M. Alice, "Imputing missing data with R," *MICE Package 2015*. [Online]. Available: <https://datascienceplus.com/imputing-missing-data-with-r-mice-package>. [Accessed: 29-Jan-2018].
- [51] P. Madley-Dowd, R. Hughes, K. Tilling, and J. Heron, "The proportion of missing data should not be used to guide decisions on multiple imputation," *J. Clin. Epidemiol.*, vol. 110, pp. 63–73, Jun. 2019.
- [52] L. Ren, T. Wang, A. Sekhari Seklouli, H. Zhang, and A. Bouras, "A review on missing values for main challenges and methods," *Inf. Syst.*, vol. 119, p. 102268, Oct. 2023.
- [53] N. Rout, D. Mishra, and M. K. Mallick, "Handling imbalanced data: A survey," in *Advances in Intelligent Systems and Computing*, Singapore: Springer Singapore, 2018, pp. 431–443.
- [54] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012.
- [55] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, Nov. 2016.
- [56] V. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial intelligence review*, 2004.
- [57] S. G. K. Patro and K. K. Sahu, "Normalization: A Preprocessing Stage," *arXiv [cs.OH]*, 19-Mar-2015.
- [58] L. A. Al Shalabi, Z. Shaaban, and B. Kasasbeh, "Data Mining: A Preprocessing Engine," *Journal of Computer Science*, vol. 2, pp. 735–739, Sep. 2006.
- [59] S. G. K. Patro, P. P. Sahoo, I. Panda, and K. K. Sahu, "Technical analysis on financial forecasting," *arXiv [cs.NE]*, 10-Mar-2015.
- [60] M. Fang *et al.*, "A hybrid machine learning approach for hypertension risk prediction," *Neural Comput. Appl.*, vol. 35, no. 20, pp. 14487–14497, Jul. 2023.
- [61] J. B. Raja and S. C. Pandian, "PSO-FCM based data mining model to predict diabetic disease," *Comput. Methods Programs Biomed.*, vol. 196, no. 105659, p. 105659, Nov. 2020.
- [62] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, Dec. 2021.
- [63] P. Ghosh, S. Azam, A. Karim, M. Hassan, K. Roy, and M. Jonkman, "A comparative study of different machine learning tools in detecting diabetes," *Procedia Comput. Sci.*, vol. 192, pp. 467–477, 2021.
- [64] S. Mundra *et al.*, "Classification of imbalanced medical data: An empirical study of machine learning approaches," *J. Intell. Fuzzy Syst.*, vol. 43, no. 2, pp. 1933–1946, Jun. 2022.

- [65] S. Balasubramanian, R. Kashyap, S. T. Cvn, and M. Anuradha, "Hybrid prediction model for type-2 diabetes with class imbalance," in *2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)*, Hyderabad, India, 2020, pp. 1–6.
- [66] A. Basant, M. Menem, M. Nagham Elsayed, and E. Ibrahim Mahmoud, "Heart Disease Prediction Using Genetic Algorithm with Machine Learning Classifiers," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 12, no. 5, pp. 1887–1894, 2022.
- [67] H. Zhang, C.-L. Hung, W. C.-C. Chu, P.-F. Chiu, and C. Y. Tang, "Chronic kidney disease survival prediction with artificial neural networks," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018, pp. 1351–1356.
- [68] C. Azad, B. Bhushan, R. Sharma, A. Shankar, K. K. Singh, and A. Khamparia, "Prediction model using SMOTE, genetic algorithm and decision tree (PMSGD) for classification of diabetes mellitus," *Multimedia Systems*, vol. 28, no. 4, pp. 1289–1307, Aug. 2022.
- [69] M. S. Roobini and M. Lakshmi, "Autonomous prediction of Type 2 Diabetes with high impact of glucose level," *Comput. Electr. Eng.*, vol. 101, no. 108082, p. 108082, Jul. 2022.
- [70] S. Tyagi and S. Mittal, "Sampling approaches for imbalanced data classification problem in machine learning," in *Proceedings of ICRIC 2019*, 2020, pp. 209–221.
- [71] N. G. Ramadhan, "Comparative analysis of ADASYN-SVM and SMOTE-SVM methods on the detection of type 2 diabetes mellitus," *Sci. J. Inform.*, vol. 8, no. 2, pp. 276–282, Nov. 2021.
- [72] B. S. Ahamed, M. S. Arya, and A. O. V. Nancy, "Diabetes mellitus disease prediction using machine learning classifiers with oversampling and feature augmentation," *Adv. Hum.-Comput. Interact.*, vol. 2022, pp. 1–14, Sep. 2022.
- [73] S. Sreejith, H. Khanna Nehemiah, and A. Kannan, "Clinical data classification using an enhanced SMOTE and chaotic evolutionary feature selection," *Comput. Biol. Med.*, vol. 126, p. 103991, Nov. 2020.
- [74] K. Roy *et al.*, "An Enhanced Machine Learning Framework for Type 2 Diabetes Classification Using Imbalanced Data with Missing Values," *Complexity*, vol. 2021, Jul. 2021.
- [75] Z. Ning, Z. Ye, Z. Jiang, and D. Zhang, "BESS: Balanced evolutionary semi-stacking for disease detection using partially labeled imbalanced data," *Inf. Sci. (Ny)*, vol. 594, pp. 233–248, May 2022.
- [76] L. Yousefi and A. Tucker, "Identifying latent variables in dynamic bayesian networks with bootstrapping applied to type 2 diabetes complication prediction," *Intell. Data Anal.*, 2022.
- [77] M. Kamaladevi and V. Venkatraman, "Tversky similarity based undersampling with gaussian kernelized decision stump adaboost algorithm for imbalanced medical data classification," *Int. J. Comput. Commun. Control*, vol. 16, no. 6, Nov. 2021.
- [78] Y. Shaikh, V. Parvati, and S. R. Biradar, "Early disease prediction algorithm for hypertension-based diseases using data aware algorithms," *Indonesian Journal of Electrical*, 2022.

- [79] Y. Zhang, J. Li, P. Qu, W. Yao, and C. Lu, "Analysis and prediction of diabetes incidence based on big data," in *2022 5th International Conference on Big Data Technologies (ICBDT)*, Qingdao China, 2022.
- [80] K. Swathi and S. Kodukula, "XGBoost classifier with hyperband optimization for cancer prediction based on geneselection by using machine learning techniques," *Rev. D Intell. Artif.*, vol. 36, no. 5, 2022.
- [81] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, 2006.
- [82] T. G. Dietterich, "Ensemble learning," *The handbook of brain theory and neural networks*, vol. 2, no. 1, pp. 110–125, 2002.
- [83] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan, and J. Haider, "An Ensemble Approach for the Prediction of Diabetes Mellitus Using a Soft Voting Classifier with an Explainable AI," *Sensors*, vol. 22, no. 19, Sep. 2022.
- [84] A. Ashfaq, A. Imran, I. Ullah, A. Alzahrani, K. M. Ali Alheeti, and A. Yasin, "Multi-model ensemble based approach for heart disease diagnosis," in *2022 International Conference on Recent Advances in Electrical Engineering & Computer Sciences (RAEE & CS)*, 2022, pp. 1–8.
- [85] M. Z. Alom *et al.*, "A state-of-the-art survey on Deep Learning theory and architectures," *Electronics (Basel)*, vol. 8, no. 3, p. 292, Mar. 2019.
- [86] K. Mridha, "Early prediction of breast cancer by using artificial neural network and machine learning techniques," in *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, Bhopal, India, 2021, pp. 582–587.
- [87] S. Patidar, D. Kumar, and D. Rukwal, "Comparative analysis of Machine Learning algorithms for heart disease prediction," in *Advances in Transdisciplinary Engineering*, IOS Press, 2022.
- [88] P. S Phani, J. Veerapaneni Esther, A. Chokka, V. K, S. Vahiduddin, and S. S, "Chronic Kidney Disease prediction using ML-based Neuro-Fuzzy model," *Int. J. Image Graph.*, vol. 24, no. 06, Nov. 2024.
- [89] R. D. Joshi and C. K. Dhakal, "Predicting type 2 diabetes using logistic regression and machine learning approaches," *Int. J. Environ. Res. Public Health*, vol. 18, no. 14, p. 7346, Jul. 2021.
- [90] I. D. Mienye and Y. Sun, "Performance analysis of cost-sensitive learning methods with application to imbalanced medical data," *Informatics in Medicine Unlocked*, vol. 25, p. 100690, Jan. 2021.
- [91] T. A. Assegie, P. K. Rangarajan, N. K. Kumar, and D. Vigneswari, "An empirical study on machine learning algorithms for heart disease prediction," *IAES Int. J. Artif. Intell. (IJ-AI)*, vol. 11, no. 3, p. 1066, Sep. 2022.
- [92] A. Kishor and C. Chakraborty, "Early and accurate prediction of diabetics based on FCBF feature selection and SMOTE," *Int. J. Syst. Assur. Eng. Manag.*, Jun. 2021.
- [93] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," *J. Big Data*, vol. 9, no. 1, Nov. 2022.
- [94] A. R. Pasha Syed *et al.*, "Implementation of ensemble machine learning algorithms on exome datasets for predicting early diagnosis of cancers," *BMC Bioinformatics*, vol. 23, no. 1, p. 496, Nov. 2022.
- [95] P. K. Syriopoulos, N. G. Kalampalikis, S. B. Kotsiantis, and M. N. Vrahatis, "KNN classification: A review," *Ann. Math. Artif. Intell.*, Sep. 2023.

- [96] A. M. Majid and W. H. Utomo, "Application of discretization and adaboost method to improve accuracy of classification algorithms in predicting diabetes mellitus," *ICIC Exp. Lett.*, vol. 12, no. 12, pp. 1177–1184, 2021.
- [97] N. Komal Kumar, D. Vigneswari, M. Vamsi Krishna, and G. V. Phanindra Reddy, "An optimized random forest classifier for diabetes mellitus," in *Advances in Intelligent Systems and Computing*, Singapore: Springer Singapore, 2019, pp. 765–773.
- [98] M. G. El-Shafiey, A. Hagag, E.-S. A. El-Dahshan, and M. A. Ismail, "A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest," *Multimed. Tools Appl.*, vol. 81, no. 13, pp. 18155–18179, May 2022.
- [99] M. T. García-Ordás, C. Benavides, J. A. Benítez-Andrades, H. Alaiz-Moretón, and I. García-Rodríguez, "Diabetes detection using deep learning techniques with oversampling and feature augmentation," *arXiv [cs.CV]*, 03-Feb-2024.
- [100] R. Rajkamal, A. Karthi, and X.-Z. Gao, "Diabetes prediction using derived features and ensembling of boosting classifiers," *Comput. Mater. Contin.*, vol. 73, no. 1, pp. 2013–2033, 2022.
- [101] S. Penafiel, N. Baloian, H. Sanson, and J. A. Pino, "Predicting stroke risk with an interpretable classifier," *IEEE Access*, vol. 9, pp. 1154–1166, 2021.
- [102] P. Sornsuwit, "Enhance Weak Learner Model of AdaBoost (EWDm) for Diabetes Mellitus Classification," *International Journal of Innovative Computing, Information and Control*, vol. 18, no. 4, pp. 1117–1132, 2022.
- [103] R. Rajendran and A. Karthi, "Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers," *Expert Syst. Appl.*, vol. 207, p. 117882, Nov. 2022.
- [104] J. Yang and J. Guan, "A heart disease prediction model based on feature optimization and smote-xgboost algorithm," *Inf.*, vol. 13, p. 475, Oct. 2022.
- [105] D. LaFreniere, F. Zulkernine, D. Barber, and K. Martin, "Using machine learning to predict hypertension from a clinical dataset," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2016, pp. 1–7.
- [106] L. Antony *et al.*, "A comprehensive unsupervised framework for chronic kidney disease prediction," *IEEE Access*, vol. 9, pp. 126481–126501, 2021.
- [107] W. Chang, Y. Liu, X. Wu, Y. Xiao, S. Zhou, and W. Cao, "A new hybrid XGBSVM model: Application for hypertensive heart disease," *IEEE Access*, vol. 7, pp. 175248–175258, 2019.
- [108] C. Song and X. Li, "Cost-sensitive KNN algorithm for cancer prediction based on entropy analysis," *Entropy*, vol. 24, no. 2, Feb. 2022.
- [109] C. Kokkotis *et al.*, "An explainable machine learning pipeline for stroke prediction on imbalanced data," *Diagnostics (Basel)*, vol. 12, no. 10, Oct. 2022.
- [110] H. Wu, X. Song, L. Zhu, X. Feng, Y. Li, and J. Chang, "A hypertension risk prediction model based on improve random forest," in *2022 2nd International Conference on Bioinformatics and Intelligent Computing*, 2022, pp. 429–436.
- [111] A. Ramezankhani, O. Pournik, J. Shahrabi, F. Azizi, F. Hadaegh, and D. Khalili, "The impact of oversampling with SMOTE on the performance of 3 classifiers in prediction of type 2 diabetes," *Med. Decis. Making*, vol. 36, no. 1, pp. 137–144, Jan. 2016.
- [112] N. G. Ramadhan, Adiwijaya, W. Maharani, and A. Akbar Gozali, "Prediction of cardiovascular disease (CVD) in the upcoming year using tree-based ensemble

- model,” in *12th International Conference on Information and Communication Technology (ICOICT)*, 2024, pp. 210–216.
- [113] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *jair*, vol. 16, pp. 321–357, Jun. 2002.
 - [114] T. Kosolwattana, C. Liu, R. Hu, S. Han, H. Chen, and Y. Lin, “A self-inspected adaptive SMOTE algorithm (SASMOTE) for highly imbalanced data classification in healthcare,” *BioData Min.*, vol. 16, no. 1, p. 15, Apr. 2023.
 - [115] T. Fahrudin, J. L. Buliali, and C. Fatichah, “Enhancing the performance of smote algorithm by using attribute weighting scheme and new selective sampling method for imbalanced data set,” *Int. J. Innov. Comput. Inf. Control*, vol. 15, no. 2, pp. 423–444, 2019.
 - [116] M. Li, H. Zhou, Q. Liu, and G. Wang, “SW: A weighted space division framework for imbalanced problems with label noise,” *Knowledge-Based Systems*, vol. 251, p. 109233, Sep. 2022.
 - [117] S. Maldonado, C. Vairetti, A. Fernandez, and F. Herrera, “FW-SMOTE: A feature-weighted oversampling approach for imbalanced classification,” *Pattern Recognit.*, vol. 124, p. 108511, Apr. 2022.
 - [118] S. Maldonado, J. López, and C. Vairetti, “An alternative SMOTE oversampling strategy for high-dimensional datasets,” *Appl. Soft Comput.*, vol. 76, pp. 380–389, Mar. 2019.
 - [119] X.-F. Song, Y. Zhang, D.-W. Gong, and X.-Y. Sun, “Feature selection using bare-bones particle swarm optimization with mutual information,” *Pattern Recognit.*, vol. 112, p. 107804, Apr. 2021.
 - [120] H. Singh, M. Kaur, and B. Singh, “A hybrid feature weighting and selection-based strategy to classify the high-dimensional and imbalanced medical data,” *Neural Comput. Appl.*, vol. 36, no. 20, pp. 12299–12316, Jul. 2024.
 - [121] A. Bruhin, M. Manai, and L. Santos-Pinto, “Risk and rationality: The relative importance of probability weighting and choice set dependence,” *J. Risk Uncertain.*, vol. 65, no. 2, pp. 139–184, Oct. 2022.
 - [122] J. Shanteau, “Competence in experts: The role of task characteristics,” *Organ. Behav. Hum. Decis. Process.*, vol. 53, no. 2, pp. 252–266, Nov. 1992.
 - [123] B. Kalaivani and A. Ranichitra, “Unveiling the impact of outliers: an improved feature engineering technique for heart disease prediction,” in *IoT Based Control Networks and Intelligent Systems*, 2024, pp. 469–478.
 - [124] A. ur Rehman and S. B. Belhaouari, “Unsupervised outlier detection in multidimensional data,” *Journal of Big Data*, vol. 8, no. 1, p. 80, Jun. 2021.
 - [125] L. I. Santos *et al.*, “Decision tree and artificial immune systems for stroke prediction in imbalanced data,” *Expert Syst. Appl.*, vol. 191, p. 116221, Apr. 2022.
 - [126] N. Nnamoko and I. Korkontzelos, “Efficient treatment of outliers and class imbalance for diabetes prediction,” *Artif. Intell. Med.*, vol. 104, p. 101815, Apr. 2020.
 - [127] N. G. Ramadhan, Adiwijaya, W. Maharani, and A. Akbar Gozali, “Prediction of hypertension in the upcoming year: feature correlation analysis and handling imbalanced based on random forest,” in *2023 Eighth International Conference on Informatics and Computing (ICIC)*, 2023, pp. 1–6.
 - [128] A. Géron, *Hands-on machine learning with scikit-learn, keras, and tensorflow*. “O’Reilly Media, Inc.,” 2022.

- [129] R. V. McCarthy, M. M. McCarthy, and W. Ceccucci, *Applying predictive analytics*. Springer International Publishing, 2022.
- [130] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.
- [131] N. G. Ramadhan, Adiwijaya, W. Maharani, and A. Akbar Gozali, “Prediction of diabetes mellitus in the upcoming year using SMOTE and random forest,” in *2023 International Conference on Data Science and Its Applications (ICoDSA)*, 2023, pp. 316–321.
- [132] G. Sailasya and G. L. A. Kumari, “Analyzing the performance of stroke prediction using ML classification algorithms,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, 2021.
- [133] O. Sagi and L. Rokach, “Ensemble learning: A survey,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, p. e1249, Jul. 2018.
- [134] A. Purwar and S. K. Singh, “Hybrid prediction model with missing value imputation for medical data,” *Expert Syst. Appl.*, vol. 42, no. 13, pp. 5621–5631, Aug. 2015.
- [135] H. Aydadenta and A. Adiwijaya, “A clustering approach for feature selection in microarray data classification using random forest,” *Journal of Information Processing Systems*, vol. 14, no. 5, pp. 1167–1175, 2018.
- [136] A. I. Pratiwi and Adiwijaya, “On the feature selection and classification based on information gain for document sentiment analysis,” *Appl. Comput. Intell. Soft Comput.*, vol. 2018, pp. 1–5, 2018.
- [137] U. N. Wisesty, Adiwijaya, and W. Astuti, “Feature extraction analysis on Indonesian speech recognition system,” in *2015 3rd International Conference on Information and Communication Technology (ICoICT)*, 2015, pp. 54–58.
- [138] C.-S. Yu *et al.*, “Clustering heatmap for visualizing and exploring complex and high-dimensional data related to chronic kidney disease,” *J. Clin. Med.*, vol. 9, no. 2, p. 403, Feb. 2020.
- [139] B. Jonathan, P. H. Putra, and Y. Ruldeviyani, “Observation Imbalanced Data Text to Predict Users Selling Products on Female Daily with SMOTE, Tomek, and SMOTE-Tomek,” in *2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 2020, pp. 81–85.
- [140] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, “Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis,” *Informatics*, vol. 8, no. 4, p. 79, Nov. 2021.
- [141] G. N. Ahmad, H. Fatima, S. Ullah, A. S. Saidi, and Imdadullah, “Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV,” *IEEE Access*, vol. 10, pp. 80151–80173, 2022.
- [142] N. G. Ramadhan, W. Maharani, A. A. Gozali, and A. Adiwijaya, “Modified SMOTE and Ensemble Learning Based on Expert Judgment for Chronic Diseases Prediction,” *International Journal of Innovative Computing, Information and Control (IJICIC)*, vol. 21, no. 4, 2025.
- [143] B. Butcher and B. J. Smith, “Feature engineering and selection: A practical approach for predictive models,” *Am. Stat.*, vol. 74, no. 3, pp. 308–309, Jul. 2020.
- [144] A. Amro, M. Al-Akhras, K. E. Hindi, M. Habib, and B. A. Shawar, “Instance reduction for avoiding overfitting in decision trees,” *J. Intell. Syst.*, vol. 30, no. 1, pp. 438–459, Jan. 2021.

- [145] J. Derrac, S. Garcia, L. Sanchez, and F. Herrera, “Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework,” *J. Mult. Valued Logic Soft Comput*, vol. 17, pp. 255–287, 2015.
- [146] Y. Miao and Y. Xu, “Random forest-based analysis of variability in feature impacts,” in *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA)*, Shenyang, China, 2024, pp. 1130–1135.
- [147] Y. Nam and S. Han, “Random Forest variable importance-based selection algorithm in class imbalance problem,” *arXiv [stat.ML]*, 16-Dec-2023.
- [148] J. Van Hulse, T. M. Khoshgoftaar, A. Napolitano, and R. Wald, “Threshold-based feature selection techniques for high-dimensional bioinformatics data,” *Netw. Model. Anal. Health Inform. Bioinform.*, vol. 1, no. 1–2, pp. 47–61, Jun. 2012.
- [149] V. Hamer and P. Dupont, “An importance weighted feature selection stability measure,” *J. Mach. Learn. Res.*, vol. 22, no. 116, p. 116:1-116:57, 2021.