

# Optimizing Machine Learning Models for Class Imbalance in Heart Disease Prediction

**Majid Rahardi**

Department of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, 55283, Indonesia  
majid@amikom.ac.id (corresponding author)

**Bima Pramudya Asaddulloh**

Department of Informatics, Postgraduate Program, Universitas Amikom Yogyakarta, Sleman, 55283, Indonesia  
bima@students.amikom.ac.id

**Afrig Aminuddin**

Department of Information Systems, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, 55283, Indonesia  
afrig@amikom.ac.id

**Ferian Fauzi Abdulloh**

Department of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, 55283, Indonesia  
ferian@amikom.ac.id

**Ilham Saifudin**

Department of Informatics Engineering, Faculty of Engineering, Universitas Muhammadiyah Jember, Jember, 68121, Indonesia  
ilham.saifudin@unmuhjember.ac.id

**Faqih Putra Kusumawijaya**

School of Information Systems, Faculty of Science, Queensland University of Technology, Brisbane, 4001, Australia  
faqih.kusumawijaya@connect.qut.edu.au

Received: 31 January 2025 | Revised: 27 February 2025 and 15 April 2025 | Accepted: 23 April 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.10407>

## ABSTRACT

Machine learning models serve as a potent instrument for forecasting heart diseases, nevertheless, class imbalance in datasets—characterized by a disproportionate number of healthy individuals compared to those with heart disease—can markedly affect the efficacy of these models. This study presents a machine learning pipeline that incorporates resampling methods, including SMOTE, ADASYN, and Random Oversampling (ROS), with commonly utilized classifiers, such as Random Forest (RF), k-Nearest Neighbors (kNN), Gradient Boosting, and Adaboost. Utilizing the 2022 CDC's Indicators of Heart Disease dataset, we examine the efficacy of these methodologies considering prediction accuracy, precision, recall, F1-score, and AUC. Compared to various previous studies, the findings show that RF with ROS achieves the highest overall performance, showing 95.75% accuracy, 99.84% recall, 95.91% F1-score, and 99.59% AUC. The findings illustrate the efficacy of oversampling approaches to rectify class imbalance and enhance heart disease prediction.

*Keywords-ML; heart disease classification; resampling methods*

## I. INTRODUCTION

Heart diseases are a leading cause of death globally [1]. Early diagnosis and treatment improve outcomes and prevent serious consequences. Clinical diagnostics predict heart disease prognoses, but Machine Learning (ML) can automate and improve them, by identifying trends in large datasets that doctors miss, enabling early and more accurate heart disease detection [2]. Medical ML has many benefits, but often face the class imbalance issue. Medical datasets with more healthy people than heart disease patients may bias model predictions toward the majority class. The model may perform well overall but poorly in recognizing heart disease patients [3].

Authors in [4] studied early heart disease with ML using Random Forest (RF), Support Vector Machines (SVM), k-Nearest Neighbors (kNN), and Decision Trees (DTs) to train and validate the model after patient data analysis with RF being the most accurate (94.958%), while model accuracy increased to 95.4% after changing random search hyperparameters. Authors in [5] developed an ML system using SVM, Logistic Regression (LR), Artificial Neural Networks (ANNs), kNN, Naïve Bayes (NB), DT Relief, MRMR, LASSO, LLBFS, and FCMIM to detect cardiac illness. The study found that feature selection improves model performance. Authors in [6] predicted heart disease with a Hybrid RF and Linear Model (HRFLM). Feature selection and categorization improved prediction accuracy to 88.7%. Authors in [7] developed the UCI Heart Disease Cleveland dataset-based MIFH for heart disease detection. Factor Analysis of Mixed Ddata (FAMD) was utilized and fed the proposed MIFH method along with various established models. The proposed MIFH method surpassed the other models in accuracy, sensitivity, and specificity. Authors in [8] developed an intelligent heart disease diagnostic system employing RSA feature selection and an upgraded RF model. Quality training/testing data reduced overfitting. The heart failure data recognition was 93.33% surpassing 11 other ML systems.

Existing studies on heart disease prediction focus on improving classification accuracy through feature selection, model optimization, and hybrid approaches but usually overlook class imbalance. Several studies employed machine learning techniques such as RF, SVM, kNN, and DTs, achieving high accuracy by tuning hyperparameters, integrating feature selection methods, and combining multiple datasets [9]. While some approaches demonstrated improved classification performance, they often relied on small or single datasets, raising concerns about overfitting and generalizability [10]. Additionally, methods optimizing accuracy and efficiency often neglected the impact of imbalanced data, potentially limiting their ability to detect minority class cases [11]. Despite advancements in predictive modeling, a research gap in addressing class imbalance remains through resampling techniques while maintaining high classification performance. This gap highlights the need for robust ML frameworks that integrate effective resampling methods to enhance heart disease detection.

This study addresses class imbalance in heart disease prediction using SMOTE, ADASYN, ROS, and robust ML

classifiers like RF, kNN, Gradient Boosting (GB), and Adaboost (AB). The 2022 CDC Indicators of Heart Disease dataset [12] was preprocessed using several methods such as data encoding, handling missing values, handling duplicated data, various resampling methods, feature selection and normalization. Through comprehensive resampling method evaluation, the models' heart disease recall and F1-score, which are critical for correct diagnosis, improved. RF with ROS surpassed the other approaches with 95.75% accuracy and 99.84% recall.

## II. PROPOSED METHOD

This chapter describes how an ML pipeline forecasts heart diseases on the heart disease dataset. The methodology includes data loading and preprocessing, exploratory data analysis, balancing, feature selection, normalization, splitting, model training, and evaluation. Figure 1 illustrates the research flow diagram.

### A. Dataset

This study used a dataset [12] with 5,110 occurrences and 12 attributes per patient. The collection includes demographic data like age and gender, medical history like hypertension, lifestyle hazards like smoking status, and other relevant variables that aid heart disease prediction. This dataset targets the heart disease column with two values: "Yes" and "No." The dataset has 289,653 "No" and 12,064 "Yes." instances. Table I shows some dataset samples.

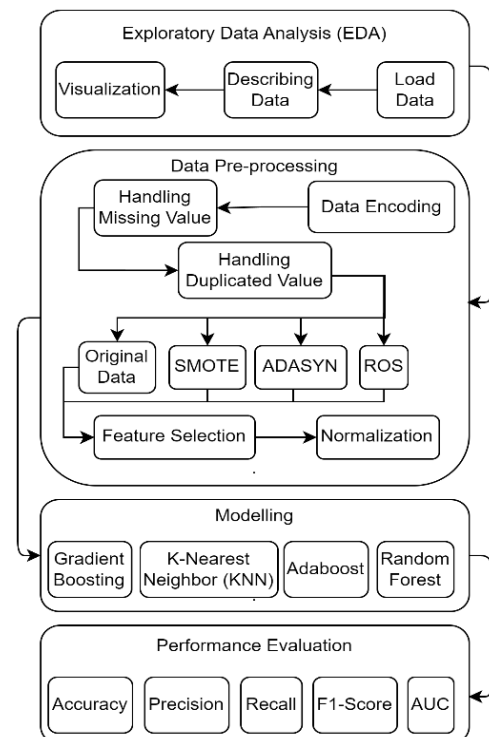


Fig. 1. The research flow diagram.

TABLE I. HEART DISEASE CLASSIFICATION DATASET SAMPLE

No	Class	Sex	...	SleepTime	Smoking	GenHealth
1	No	Female	...	5.0	Yes	Very good
2	No	Female	...	7.0	No	Very good
3	No	Male	...	8.0	Yes	Fair
4	No	Female	...	6.0	No	Good
5	No	Female	...	8.0	No	Very good
...	...	...	...	...	...	...
319790	Yes	Male	...	6.0	Yes	Fair
319791	No	Male	...	5.0	Yes	Very good
319792	No	Female	...	6.0	No	Good
319793	No	Female	...	12.0	No	Good
319794	No	Female	...	8.0	No	Good

### B. Exploratory Data Analysis (EDA)

ML begins with data collection and investigation. The following steps were used to manage and evaluate the data. EDA shows dataset structure and dispersion [13]. Patterns, anomalies, and correlations between characteristics and the desired variable were visualized. The considered total of 319,795 entries had 8.56% heart disease. The dataset is gender-balanced and diverse in age and race. These traits, lifestyle, and health markers affect heart disease risk. Heart disease contrasts positive (heart disease) and negative (no illness) classes. Uneven samples favor the dominant class, biasing model performance.

### C. Data Preprocessing

Most ML approaches numericalize the dataset's categorical attributes. Each class column utilizes LabelEncoder. Label encoding gives feature categories integer values [13]. Represent categories inside a feature  $X$  with  $C$ . In label encoding, the count of non-missing items in column  $X$  determines the integer value ( $ci$ ) for each  $ci \in C$ :

$$\rho = \frac{\bar{E}}{J_C(T=\text{const.}) \cdot \left( P \cdot \left( \frac{\bar{E}}{E_C} \right)^m + (1-P) \right)} \quad (1)$$

After missing value correction, no attributes had any missing values. Deficient data may have been imputed or partially deleted [14]. Hence, all features, including "BMI," "PhysicalHealth," "MentalHealth," and categorical variables like "Smoking," "Sex," and "Race," have been finalized, assuring that the dataset is ready for ML modeling. A total of 18,078 duplicates was detected and removed to avoid model training distortion [14]. Data cleaning left 283,639 data entries.

While class dominance can increase the likelihood of overfitting, datasets processed using various resampling methods produce balanced datasets, allowing the model to predict each class more accurately. SMOTE, ADASYN, ROS [3] oversampling methods were utilized to fix the dataset's class imbalance.

SMOTE interpolates data to construct minority class instances [3]. Equation (2) demonstrates how SMOTE uses a nearest neighbor  $x_j$  to create a new synthetic sample  $x_{new}$  for a minority class sample  $x_i$ , where  $\lambda$  is a stochastic variable. Without copying samples, this method creates a more egalitarian dataset.

$$x_{new} = x_i + \lambda \cdot (x_j - x_i), \lambda \in [0,1] \quad (2)$$

ADASYN creates synthetic samples for minority class occurrences that are harder to identify [15]. The minority class density determines the amount of synthetic samples produced, ensuring that harder samples receive more synthetic instances. ROS is used to replicate minority class instances until the dataset is balanced [16]. It is simpler than SMOTE and ADASYN, but it can overfit if not maintained properly.

In the original dataset, only 27,261 persons had heart disease and 274,456 were healthy. Oversampling ensured that models trained on these data were not majority-class biased. Table II compares SMOTE, ADASYN, and ROS data balancing before and after target classes (0: no heart disease, 1: heart disease).

TABLE II. AMOUNT OF DATA FOR EACH RESAMPLING METHOD

Balancing Method	0	1	Total
Original Data	274456	27261	301717
SMOTE	274456	274456	548912
ADASYN	274456	273062	547518
ROS	274456	274456	548912

ML model efficiency and performance depend on feature selection. The most important heart disease predictors are identified using correlation analysis. The Pearson correlation coefficient was calculated for each attribute versus the target variable [17]. The Pearson correlation coefficient for two variables  $X$  (features) and  $Y$  (target) is shown in (3), where  $\text{cov}(X, Y)$  represents the covariance and  $\sigma_X$  and  $\sigma_Y$  indicate the standard deviations. Model training uses five features with the highest absolute correlation to the target variable.

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (3)$$

The resampling method causes changes in the distribution of data, so the correlation between variable  $X$  (features) and variable  $Y$  (target) will be slightly different for each resampling method. Table III shows the top five attributes, the original data, and SMOTE, ADASYN, and ROS data.

TABLE III. FIVE BEST FEATURES OF EACH RESAMPLING TECHNIQUE

Balancing Method	Five best features selected
Original data	AgeCategory, DiffWalking, Stroke, PhysicalHealth, Diabetic
SMOTE	AgeCategory, PhysicalActivity, PhysicalHealth, Diabetic, AlcoholDrinking
ADASYN	AgeCategory, PhysicalActivity, PhysicalHealth, Diabetic, AlcoholDrinking
ROS	AgeCategory, DiffWalking, Diabetic, PhysicalHealth, Stroke

If class imbalance is corrected, these traits may help diagnose cardiac disease. Balanced techniques vary feature choices by showing data in different situations. Standard scaling ensures that model predictions include all selected features uniformly. Standard-scaled data's mean and standard deviation are 0 and 1 [18]. In (4),  $\mu$  is the mean and  $\sigma$  is the standard deviation for a typical scaling adjustment for feature

X. This transformation gives each feature a 0 mean and 1 variance, preventing larger features from overshadowing the model.

$$X' = \frac{X-\mu}{\sigma} \quad (4)$$

#### D. Modeling

The dataset was divided into 80% training and 20% testing sets. Four ML classifiers were considered: RF, kNN, GB, and AB. Training RF generates DTs and class modes. RF reduces overfitting and handles high-dimensional data. Feature space dominant kNN works with non-linear feature correlations on computationally intensive large datasets. GB fixes DT errors, but may overfit without correction. AB uses weak DTs to create robust classifiers. kNN is simpler and more distance-oriented than RF, GB, and AB for complex patterns.

#### E. Performance Evaluation

Each model's accuracy, precision, recall, F1-score, and confusion matrix were calculated. These measures thoroughly assess the accuracy of the models' heart disease and non-heart disease categorization. The metrics are described as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

$$\text{AUC} = \int_0^1 \text{ROC}(t) dt \quad (9)$$

where TP, TN, FP, and FN stand for the classified instances of True Positive, True Negative, False Positive, and False Negative, respectively.

### III. RESULTS AND ANALYSIS

#### A. Model Performance on Original Data

This section assesses the efficacy of several classifiers on the original, imbalanced dataset without implementing any resampling techniques. The objective is to create a benchmark for comparison and ascertain the impact of class imbalance on classification efficacy. Table IV shows the original dataset classifiers' efficacy. The classifiers have great accuracy but a significant imbalance in recognizing the minority class (heart disease). GB had the highest accuracy, 91.13%. The recall was poor (8.72%), showing that the model had trouble detecting cardiac disease. With an F1-score of 15.07 to 15.26, imbalanced data management was poor. kNN, AB, and RF performed poorly in recall, with values between 8% and 11%. These findings demonstrate the limitations of classifiers on imbalanced datasets without resampling. The bias toward the majority class hinders models' capacity to recognize heart disease patients. The models were accurate but misclassified heart disease cases, highlighting the need for data balance to correct the dataset's skew.

TABLE IV. PERFORMANCE METRICS ON THE ORIGINAL DATASET

Classifier	Accuracy	Precision	Recall	F1-Score	AUC
GB	91.13	55.56	08.72	15.07	83.10
kNN	90.08	31.71	08.48	13.38	66.86
AB	91.02	51.28	11.39	18.65	82.81
RF	89.98	33.31	10.88	16.40	77.86

#### B. Performance with Oversampling

This section examines how SMOTE, ADASYN, and ROS oversampling approaches improve classification performance. Each method's impact on model performance was evaluated using consistent metrics and compared to the original dataset and the three oversampling strategies to discover the best class imbalance solution.

TABLE V. PERFORMANCE METRICS OF MODELS TRAINED ON VARIOUS METHODS

Method	Classifier	Acc.	Prec.	Rec.	F1	AUC
SMOTE	GB	80.91	79.34	83.58	81.41	90.15
	kNN	83.85	77.30	95.82	85.57	91.86
	AB	77.57	76.34	79.89	78.08	86.56
	RF	89.27	87.75	91.29	89.48	95.69
ADASYN	GB	91.13	55.56	08.72	15.07	83.10
	kNN	90.08	31.71	08.48	13.38	66.86
	AB	91.02	51.28	11.39	18.65	82.81
	RF	89.98	33.31	10.94	16.47	78.05
ROS	GB	76.19	74.23	80.22	77.11	83.90
	kNN	86.45	79.88	97.41	87.78	93.11
	AB	75.86	74.82	77.94	76.35	83.48
	RF	95.75	92.28	99.84	95.91	99.59

Table V compares classifier performance using SMOTE, ADASYN, and ROS. ROS performed best, especially with the Because it supported high-dimensional data without overfitting and kept data distribution, RF with ROS performed best, with the highest recall (99.84%), F1-score (95.91%), and AUC (99.59%). ROS minimized class imbalance while maintaining good classification performance with kNN at 97.41% recall, 87.78% F1-score, and 93.11% AUC. kNN performed worse because distance learning renders it prone to ROS data duplication, which biases predictions. GB failed due to overfitting on noisy synthetic data.

SMOTE also improved model performance, especially for RF, which achieved an 89.48% F1-score, 91.29% recall, and a 95.69% AUC, demonstrating its effectiveness in generating synthetic minority samples. kNN with SMOTE also showed strong performance, with 95.82% recall and an 85.57% F1-score, surpassing GB and AB. However, some classifiers under SMOTE, such as GB and AB, showed moderate performance, with F1 scores of 81.41% and 78.08%, respectively.

In contrast, ADASYN failed to improve classifier performance significantly. GB and kNN under ADASYN had extremely low recall values (8.72% and 8.48%, respectively), and RF only achieved a 16.47% F1-score with a recall of 10.94%, highlighting ADASYN's ineffectiveness in balancing the dataset. Despite maintaining high accuracy values (over 89%), the poor recall and F1-scores indicate that models trained with ADASYN struggled to detect minority class instances effectively. Synthetic samples introduced noise rather

than meaningful patterns and increased data anomalies, reducing model effectiveness for ADASYN.

Overall, ROS outperformed SMOTE and ADASYN, particularly for the RF classifier, which consistently achieved the highest recall, F1-score, and AUC. ROS improves RF by preserving data distribution without synthetic noise. SMOTE may create unlikely data points, while ROS stabilizes tiny samples. SMOTE also provided considerable improvements despite some limitations in minority class detection. Meanwhile, ADASYN was the least effective method, as its models retained poor recall and F1 scores despite the high accuracy, making it unsuitable for datasets with similar characteristics.

### C. Comparison between the Proposed Method and Existing Research

This section contrasts the outcomes of the proposed method with established heart disease classification systems from the literature. Table VI presents a comparative analysis between the proposed method and existing heart disease classification schemes. The proposed method, utilizing RF with ROS, outperforms all prior approaches across most evaluation metrics. It achieves the highest recall (99.84%) and F1-score (95.91%), indicating a superior ability to detect heart disease cases while maintaining balanced precision (92.28%). Additionally, it records the highest AUC (99.59%), demonstrating its excellent discriminatory power.

TABLE VI. COMPARISON BETWEEN THE PROPOSED METHOD AND THE EXISTING SCHEMES IN HEART DISEASE CLASSIFICATIONS

Scheme	Best Model	Acc.	Prec.	Rec.	F1	AUC
[4]	RF+Tuning	95.4	94	-	95	-
[5]	FCMIM+SVM	92.37	-	89	-	-
[6]	RF+LM	88.7	87.5	92.8	90	-
[7]	FAMD+RF	93.44	-	89.28	-	93.1
[8]	RSA+RF	93.33	-	95.12	-	-
Proposed	RF+ROS	95.75	92.28	99.84	95.91	99.59

Among compared methods, the one in [4] achieved 95.4% accuracy with a tuned RF model, comparable to the proposed method but lacking recall and AUC metrics for direct comparison. The RSA+RF model in [8] attained 95.12% recall, which is slightly lower than the proposed method, but did not report other key metrics. FCMIM+SVM [5] and FAMD+RF [7] achieved recall values of 89% and 89.28%, respectively, notably lower than the proposed method's 99.84%. Similarly, the RF+LM [6] approach was outperformed by the proposed method in all metrics. Overall, the proposed RF with ROS is the most effective technique, achieving the highest recall, AUC, and F1 score, making it the most suitable for heart disease classification.

## IV. CONCLUSION

In this study, machine-learning algorithms were used to improve heart disease prediction by addressing class imbalance in medical datasets. Resampling methods like SMOTE, ADASYN, and Random Oversampling improved classifier performance, especially for Random Forest, in minority class recognition, precision, and accuracy. Random Forest and

Random Oversampling produced the best results for heart disease categorization, with a 99.84% recall rate and 95.75% accuracy. This method balanced recall and precision better than other methods and past studies, making it more reliable for early heart disease diagnosis. Class imbalance in healthcare datasets must be addressed, and this study recommends oversampling approaches, particularly Random Oversampling, for clinical predictive model development. Future research may use improved machine-learning methods or hybrid models to improve heart disease prediction.

## REFERENCES

- [1] J. Rwebembera *et al.*, "2023 World Heart Federation guidelines for the echocardiographic diagnosis of rheumatic heart disease," *Nature Reviews Cardiology*, vol. 21, no. 4, pp. 250–263, Apr. 2024, <https://doi.org/10.1038/s41569-023-00940-9>.
- [2] G. S. Bhavakar, A. Das Goswami, C. P. Vasantryo, A. K. Gaikwad, A. V. Zade, and H. Vyawahare, "Heart disease prediction using machine learning, deep Learning and optimization techniques-A semantic review," *Multimedia Tools and Applications*, vol. 83, no. 39, pp. 86895–86922, Nov. 2024, <https://doi.org/10.1007/s11042-024-19680-0>.
- [3] M. Dubey, J. Tembhurne, and R. Makhijani, "Improving coronary heart disease prediction with real-life dataset: a stacked generalization framework with maximum clinical attributes and SMOTE balancing for imbalanced data," *Multimedia Tools and Applications*, vol. 83, no. 37, pp. 85139–85168, Nov. 2024, <https://doi.org/10.1007/s11042-024-19429-9>.
- [4] M. Kadhim and A. Radhi, "Heart disease classification using optimized Machine learning algorithms," *Iraqi Journal for Computer Science and Mathematics*, vol. 4, no. 2, Jan. 2023, <https://doi.org/10.52866/ijcs.2023.02.02.004>.
- [5] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020, <https://doi.org/10.1109/ACCESS.2020.3001149>.
- [6] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, <https://doi.org/10.1109/ACCESS.2019.2923707>.
- [7] A. Gupta, R. Kumar, H. Singh Arora, and B. Raman, "MIFH: A Machine Intelligence Framework for Heart Disease Diagnosis," *IEEE Access*, vol. 8, pp. 14659–14674, 2020, <https://doi.org/10.1109/ACCESS.2019.2962755>.
- [8] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," *IEEE Access*, vol. 7, pp. 180235–180243, 2019, <https://doi.org/10.1109/ACCESS.2019.2952107>.
- [9] S. M. Alanazi and G. S. M. Khamis, "Optimizing Machine Learning Classifiers for Enhanced Cardiovascular Disease Prediction," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12911–12917, Feb. 2024, <https://doi.org/10.48084/etasr.6684>.
- [10] M. J. J. Ghrabat *et al.*, "Utilizing Machine Learning for the Early Detection of Coronary Heart Disease," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17363–17375, Oct. 2024, <https://doi.org/10.48084/etasr.8171>.
- [11] A. K. Dubey, A. K. Sinhal, and R. Sharma, "An Improved Auto Categorical PSO with ML for Heart Disease Prediction," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8567–8573, Jun. 2022, <https://doi.org/10.48084/etasr.4854>.
- [12] K. Pytlak, "Indicators of Heart Disease (2022 UPDATE)." kaggle, 2022, [Online]. Available: <https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>.
- [13] A. Thakur, T. Zhu, V. Abrol, J. Armstrong, Y. Wang, and D. A. Clifton, "Data encoding for healthcare data democratization and information leakage prevention," *Nature Communications*, vol. 15, no. 1, Feb. 2024, Art. no. 1582, <https://doi.org/10.1038/s41467-024-45777-z>.

- [14] H. Ameer, H. Njah, and S. Jamoussi, "Merits of Bayesian networks in overcoming small data challenges: a meta-model for handling missing data," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 1, pp. 229–251, Jan. 2023, <https://doi.org/10.1007/s13042-022-01577-9>.
- [15] H. Park and H. Kim, "AR-ADASYN: angle radius-adaptive synthetic data generation approach for imbalanced learning," *Statistics and Computing*, vol. 34, no. 5, Aug. 2024, Art. no. 166, <https://doi.org/10.1007/s11222-024-10479-5>.
- [16] F. Kamalov, H.-H. Leung, and A. K. Cherukuri, "Keep it simple: random oversampling for imbalanced data," in *2023 Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, United Arab Emirates, Oct. 2023, pp. 1–4, <https://doi.org/10.1109/ASET56582.2023.10180891>.
- [17] L. da Silva Renato, R. M. de Souza, E. Aparecido Vieira, J. R. de Oliveira, and P. V. Morbach Dixini, "Statistical Inference Modeling Using Pearson Correlation Filters and Best Subset Selection Applied to Steel Desulfurization Preliminary to Ladle Furnace Treatment," *JOM*, vol. 75, no. 4, pp. 1284–1293, Apr. 2023, <https://doi.org/10.1007/s11837-022-05663-9>.
- [18] B. Biswas, N. Kumar, Md. A. Hoque, and Md. A. Alam, "Weighted scaling approach for metabolomics data analysis," *Japanese Journal of Statistics and Data Science*, vol. 6, no. 2, pp. 785–802, Nov. 2023, <https://doi.org/10.1007/s42081-023-00205-2>.