International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

CONVENTIONAL AND HYBRID RESAMPLING TECHNIQUES IN MACHINE LEARNING FOR HEALTHCARE DATA

Ghayth AlMahadin

Assistant Professor, Dpartment of Data Science, College of Information
IDhttps://www.scopus.com/authid/detail.uri?authorId=57219945588

Abstract

In order to better understand and diagnose disease, several healthcare technologies now make use of large amounts of patient health data collected in the past. The performance of a classification model is directly affected by data imbalance with regard to the class labels, making this a challenging topic for machine learning techniques. The aim of the research is to categorize all of the prominent resampling techniques reported to overcome the data imbalance in the automated analysis techniques under healthcare sector. The purpose of this study is to perform a review analysis of the reported contemporary literature to identify a recently effective resampling technique. Finally, the review study is discussed and comprehended that SMOTE and its hybridized method are identified as a finest resampling technique foreseeing, the advancements, and applications in healthcare.

Keywords: *Resampling techniques, Healthcare, Medical data, Machine Learning*

Introduction

Data imbalance in relation to class labels has been recognized as a formidable challenge for machine learning methods due to its direct impact on the effectiveness of the classification model. This is due to the influence that it has on the accuracy of the model [1]. The vast majority of machine learning algorithms contain a bias that causes them to favour the class that has the most members and ignore classes that have less members [2], [3]. This can lead to the construction of classification models that cannot be generalized and is a limitation of machine learning. There are extensive machine learning techniques developed and reported [4]. In order to rectify the data imbalance, resampling either eliminates instances from the validation set (under-sampling) or introduces new records to the minority group (over-sampling). Cross-validation and bootstrapping are two examples of Resampling techniques [5].

The several classes of resampling techniques are discussed followingly. Model performance can be assessed by Cross-Validation by estimating the test error connected to it. The most fundamental method of resampling is the *validation set method* which splits the dataset at random into a training set and a validation set or hold-out set as required. The model is trained using the data in the training set, and then predictions are made using the data in the validation set [6]. The *k-fold cross-validation method* requires the data set to be split into k almost equal-sized folds at random. To fit the model, the remaining folds are used as training data, while the initial fold is used as a validation set. Then, k iterations are performed and implemented as 5-fold, 10-fold cross validation, with a distinct set of data serving as the validation set each time [7]. Comparatively, the *leave-one-out-cross-validation (LOOCV)* method is superior to the validation set method. Instead of using half of the data for validation and the other half for model fitting, a single observation is used for both purposes.

Received: August 13,2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

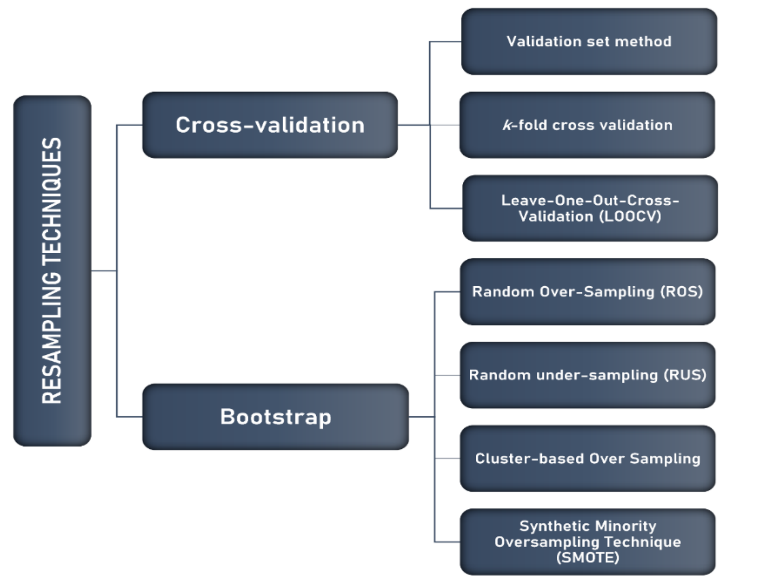

Figure 1. Classification of Resampling techniques in Machine Learning

The ambiguity of a model can be determined with the aid of Bootstrap, a robust statistical tool. Nevertheless, bootstrap's real strength lies in the fact it can be used for a diverse range of models regardless of whether the variability is hard to collect or is not supplied automatically [8]. The purpose of *Random Over-Sampling (ROS)* is to achieve statistical parity between classes by artificially and arbitrarily increasing the number of instances from underrepresented classes. In order to achieve statistical parity between classes, *random under-sampling (RUS)* often involves deleting cases from the majority class at random. The instances of the majority class are removed to increase the distances between the two classes when they are very near to each other which aids in determining proper categorization. The *Synthetic Minority Oversampling Technique (SMOTE)* is a common oversampling technique that creates "synthetic" observations in the dataset rather than repeating data [9], [10]. In order to uncover groups in datasets, *cluster-based Over Sampling* or *K-means clustering technique* can be employed on each class instance separately [11], [12].

This decade has seen the continued development of extremely effective, diverse resampling approaches by software developers aiming to address the limitations brought to light by the rapid time expectation with massive amounts of work input. This research work aims to provide a definitive category of all of the prominent resampling methods that have been reported in the past few years that pertain to the healthcare industry. The objective of this research is to identify and discuss the efficient resampling technique foreseeing future, elements of its growth and application in the field of healthcare.

Literature Review

The use of machine learning in healthcare is growing as digital health information become more widely available and computational power improves. In regard with this, [13]stated that the classifiers are developed in machine learning to reduce misclassification errors and increase prediction accuracy of the datasets considered [14]. These classification techniques work under the presumption that there are nearly equal numbers of samples in each class of considered data set. Extensive research reported by Belarouci [15] showed that class imbalance is particularly common in classification models that are used to detect significant diseases and diagnosis data. According to the study by Han et al.[16], resampling technique have been applied to a number of high-profile datasets, including the lung cancer, thyroid, dermatology [17], hepatitis [18], skin lesions [19], and other medical diagnosis [20], [21] in order to examine class imbalance strategies.

Received: August 13,2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Medical Data Classification

In order to determine the optimal data resampling technique to stabilize classification performance, Alahmari [22] investigated class imbalance for a medical application pertaining to autistic spectrum disorder (ASD) screening. The goal was accomplished by experimental studies conducted on a genuine imbalanced ASD dataset consisting of 10 people. Experiments comparing the oversampling algorithms SMOTE, ROS, and RUS indicate that class imbalance can lead to distorted findings if not handled. Random Forest and Naive Bayes models fared best on the considered dataset when ROS resampled the data. The research by Ganie and Malik [23] introduced a novel ensemble learning-based approach for an early prediction of type II diabetes mellitus utilizing the lifestyle indicators. In the study, *SMOTE* is used to ensure that each population class is represented equally, and *K-fold cross-validation* has been employed to ensure that the results produced are reliable. Out of the classification methods available, bagged decision tree had the best accuracy (99.56%), precision, recall, specificity and misclassification rate (MCR) (0.86%). Concatenated resampling(CR) has also been tested experimentally for diagnosis of vertebral column pathologies by Reshi et al., [24]. Combining Adaptive Synthetic (ADASYN) Sampling, SMOTE, and Cluster Centroids (CC) improves learning model outcomes. ADASYN is a viable resampling model which can generate 147 entries for the typical patients’ class with default parameter settings. Experiments demonstrate that the Random Forest classifier with integrated ML tools performs better without resampling and with under sampling. Oversampling the same training dataset improved Extra Tree Classifier (ETC) ML technique’s accuracy to 0.90 for SMOTE and 0.95 for ADASYN.

Shi et al., [25] used resampling to fix the prognostic model's uneven data structure and improve forecasts. Authors employed a resampling technique called SMOTE-Edited Nearest Neighbour (SMOTE-ENN), which generates synthetic instances of End Stage Kidney Disease (ESKD), to appropriately identify high-risk individuals. Despite a high rate of misclassification for the negative cases (45%), author identified ESKD high-risk people much better to that of the logistic regression and survival models. Varatto et al., [26] analysed Stereoelectroencephalography (SEEG) recordings from surgically-treated individuals with focal epilepsy. RUS performed best despite being the simplest classification method, since under sampling was more robust than oversampling. SMOTE Resampling is used to equalize positive and negative classes in a dynamic population with mostly negative results.

Barbieri et al., [27] assessed the usefulness of a data mining approach in predicting electrocardiogram (ECG) outcomes and whether resampling an unbalanced dataset may help clinical decision making. The proposed technique can improve clinical decision making and reduce needless cardiovascular reactivity (CVR) tests. SMOTE boosts Decision Tree (DT)'s area under curve (AUC), sensitivity and generated useful knowledge that reduced cardiovascular disease (CVD) predicting guesswork.

Healthcare professionals have discovered tremor classification data to be subjective, inaccurate, and unreliable. In respect to overcome the limitation, Almahadin et al., [28] suggested resampling and classifier integrated approaches like artificial neural network based on multi-layer perceptron (ANN-MLP) and random forest (RF) to improve tremor severity classification. Hybrid and over-sampling strategies outperformed conventional resampling methods. The proposed methodology enhanced tremor severity classification substantially and demonstrated that ANN-MLP with Borderline SMOTE, exhibits 93.81% overall accuracy.

Enhancement of Dataset and Model Evaluation

Mohammed et al., [29] conducted an experiment with three machine learning algorithms DT (J48), sequential minimal optimization (SMO) and NB to evaluate which ML strategies are most effective. The experiment has been conducted using two of the popular datasets, namely the white blood cells (WBC) and the datasets of breast cancer. Incorporating resampling data labelling techniques to mitigate the effects of the imbalance problem is just one of the many reasons why this study is significant. The SMO algorithms outperformed the other two classifiers, with an accuracy of above 95% in both datasets. To lessen the disparity ratio, researchers resampled the data several times, which could have influenced the results. It was comprehended that due to this, the performance of these three machine learning methods may endure when applied to an uneven or irregularly distributed dataset.

Received: August 13,2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The accuracy of Appearance, Pulse, Grimace, Activity and Respiration combinedly known as Apgar score can be forecasted using a variety of resampling and pre-processing strategies. Tarimo et al., [30] narrowed attention on the SMOTE, the Borderline SMOTE (BSMOTE), and the RUS. XGBoost improved after being subjected to BSMOTE resampling approaches, allowing the model to accurately identify 93% (an increase from 20% baseline performance) of new-borns with a low Apgar score, while missing only 7% of the time. By comparison, this is a 20% improvement above the norm. Potentially improved outcomes can be obtained by applying the borderline-SMOTE method to the selected ensemble classifiers which was concluded effective.

Murphy et al., [31] determined the values for AUC and the receiver operator characteristic (ROC) curve for over 10 years of collective training data and evaluated the classifier's ability to distinguish between stomach cancer cases. Estimates of the capacity to forecast the affected and unaffected case status of observations often shows unduly optimistic response when based on an evaluation of discrimination within the training data. Hence, LOOCV was incorporated to check and validate the model internally. Under LOOCV the elements like Urea, Pepsinogen identification models were validated, and AUC of the models were determined.

Healthcare Database

Francesco and team reported the best model based on plausibility of observed correlations and fixed effects model accuracy on simulated holdout data using LOOCV[32]. Humanitarian indivituals should routinely make critical datasets publicly present in curated and accessible form to encourage research of this nature and other studies that benefit the public, such as anthropometric surveys among the pregnant women and outpatient children, admissions and the exit outcomes for the management of acute malnutrition. The dataset that was accessible both before and after it was processed was subjected to the application of a number of different ensemble classifiers. These classifiers used several sampling techniques to classify intensive-care units (ICU) based infection dataset [33]. RUS ratio should be high enough to boost minority class classification but low enough to prevent information loss owing to imbalanced data set. 6.7% of patients showed signs of disease, whereas 93.3% did not get an infection. The authors state that creating classifiers from unbalanced samples is difficult. The methodology and process of discussed various clustering based resampling approach is shown in Figure 2.

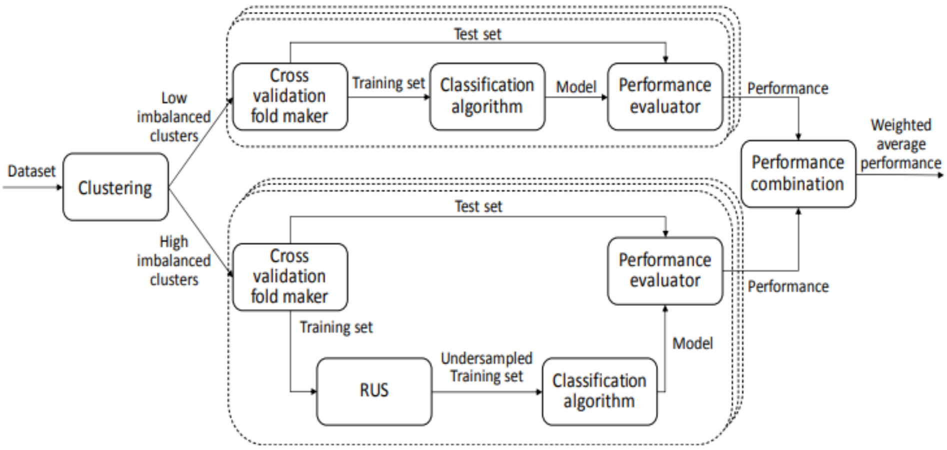

Figure 2. Structure of integrated cluster-based sampling method [33]

Jiaxing Liu's [34] article analyses fall incidence severity levels by data imbalance and structured clinical factors. Author created an incident report classification (IRC) framework that compares structured feature extraction, resampling, and ML approaches. Using this IRC framework, users can compare the performance of SMOTE and its variant structural features for improved classification and unusual class identification. Zhao et al., [35] analyses oversampling, under sampling, SMOTE, and *cost-sensitive learning* on classifier performance. Parameters used to train a classifier might affect its effectiveness, thus they are assessed extensively during implementation. SMOTE improves recall by 45.3% and accuracy by 1.5% when applied to the base

Received: August 13,2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

classifier. Classifiers' accuracy and recall vary and all rebalancing strategies improve recall over the base classifier, despite a loss in detection accuracy of healthcare data.

Synthetic Resampling Approaches for Diverse Implementations

Using resampling techniques in the field of identity management in healthcare helps to prevent adversarial learning by projecting data in partitions. By integrating a synthetic data creation technique with a deconstruction strategy, Ilavarsi et al., [36] assessed the efficacy of learning models in relation to a collection of data balancing approaches. Non-parametric analysis results presented in the experiments justify the use of *SMOTE-TL (Tomek Links)* for this purpose. The number of simulated minority groups is set by the required level of oversampling.

Li et al., [37] used a class imbalance learning method to determine the likelihood of in-hospital mortality for intensive care unit patients. The bootstrapping approach was used to determine the mean and standard deviation of the assessment methods, to calculate statistically meaningful differences between the different models, and to construct bootstrap error bars for the models with a confidence level of 95%.

SMOTE oversamples minorities when resampling imbalanced data. SMOTE's samples may be confusing, low-quality, and similar to the rest. This model finds "visible" nearest neighbours adaptively. The proposed Self-inspected Adaptive SMOTE (SASMOTE) model by Kosolwattana et al., [38] incorporates uncertainty removal via self-inspection to improve sample quality. Both techniques perform better around a threshold of uncertainty score before declining. The optimal uncertainty threshold for SASMOTE with visible neighbours and SASMOTE alone is 50%.

Reeves et al., [39] developed machine learning and statistical methods for suicide death prediction using race and ethnicity. Blind, Separate, or Equity resampling can be used to sensitive categories. By using one of these resampling methods, users can create predictive models with less heterogeneity between data sets. Resampling and three ensemble approaches increase model performance. ADASYN improves classifier performance on an uneven dataset. The method proposed by Mohammed et al., [40] focused on synthesizing minority classes based on the asymmetry between difficult-to-learn and easy-to-learn data samples in the training dataset. The ADASYN approach was reported to use distribution probability as a criterion to synthesize minority data samples and obtain a new dataset sample. The results from the researchers suggest that the proposed approach is the best way to anticipate healthcare software issues.

Data generalizability issues hinder oversampling development. The proposed method is based on pruning, which involves looking for specific minority areas while keeping the generalization of the data. Wilbowo et al., [41]

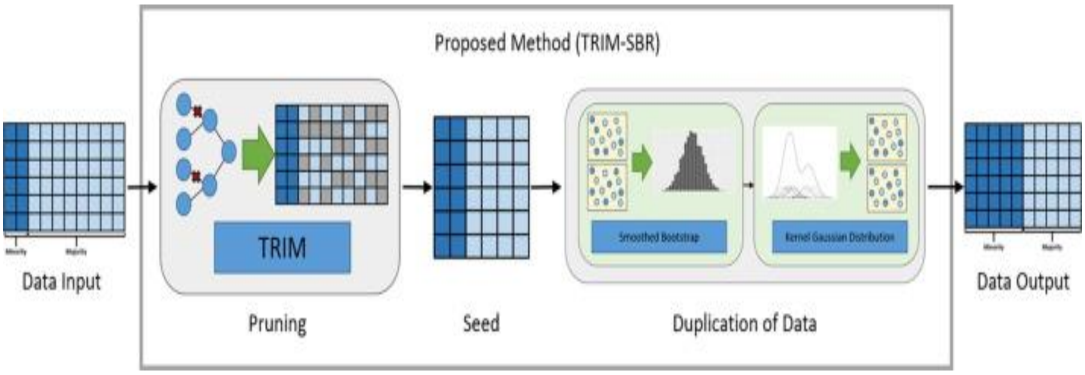

Figure 3. Process of TRIM-SBR resampling method [41]

proposed TRIM-Smoothed Bootstrap Resampling TRIM-SBR to create COVID-19 minor class data as shown in figure 3. This search for specific minority areas while keeping the generalization of the data leads to minority data seeds, which are used as standards to create new synthetic data using bootstrap resampling techniques. When training data is unevenly distributed, accuracy, specificity, sensitivity, F-measure, and AUC are used to evaluate a classifier's performance. It was discovered by the authors that TRIM-SBR is superior to alternative oversampling methods.

Received: August 13,2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Discussion On Significant Methods

The use of machine learning in healthcare is growing as digital health information become more widely available and computational power improves. The use of machine learning to the healthcare business has been criticized for raising equity concerns, such as the unequal allocation of scarce healthcare resources or the disproportionate risk of illness among particular populations. Many classifiers showed higher levels of specificity, sensitivity, and precision after being pre-processed with oversampling. In order to construct the training set, data from underrepresented groups is combined with samples from the dominant group using a random sampling procedure. It was determined that under sampling techniques were the least effective of these resampling approaches, probably because they are left out crucial concepts from the main body of data.

Furthermore, MI-MOTE wouldn't perform well when the rate of missing data is very low because it just duplicates the minority of cases that are complete and have no missing values. To duplicate data from a minority group, the ROS approach, for example, adds new examples to the training set at random [42], [43]. Under-sampling reduces the number of samples in the majority class to establish statistical parity between classes. Most commonly, this is accomplished by combining samples from the two groups, with the training set consisting of a subset of majority-class data that is selected at random. Many oversampling algorithms were developed to address this issue by enhancing coverage of the minority class in imbalanced datasets and enlarging the decision region [44], [45].

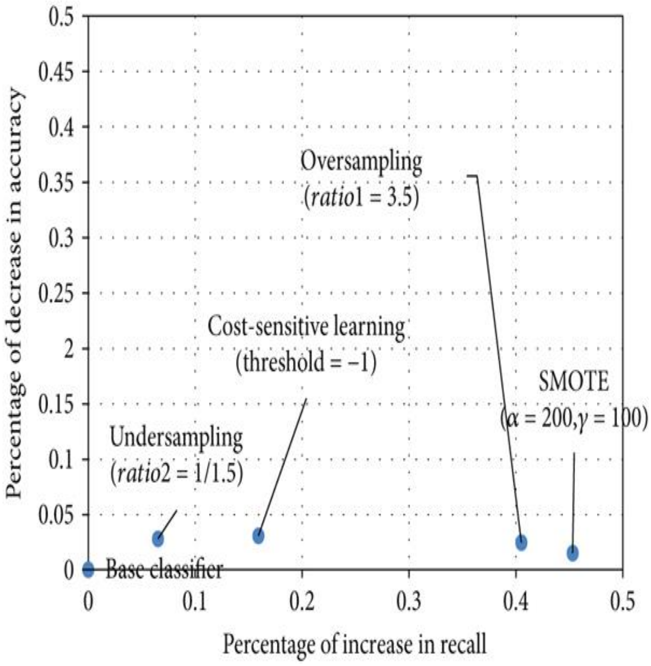

Figure 4. Accuracy and recall Percentage representation of resampling techniques[35]

The end purpose of SMOTE methods is to increase classification precision. As a result, they are biased toward the dominant class and display insensitivity to the concerns of the minority class. As a result of the analysis of the contextual section, it is clear that the SMOTE resampling method is one of the most popular options because to the versatility it offers [35]. Figure 4 from Zhao et al., and figure 5 from Alexandre et al., [46] depicts the efficient performance of the SMOTE resampling method with accuracy values. Integrated SMOTE discussed in contextual section describes how the Tomek Link technique, which uses guided under-sampling, can be used to remove examples from the majority class. T-link, a method of data compression, and SMOTE, a method of sampling, are combined to achieve the required resampling rate. Tomek-Link technique supplies the SMOTE algorithm with the optimal resampling rate n for interpolating minority class instances of k-neighbours.

Received: August 13,2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

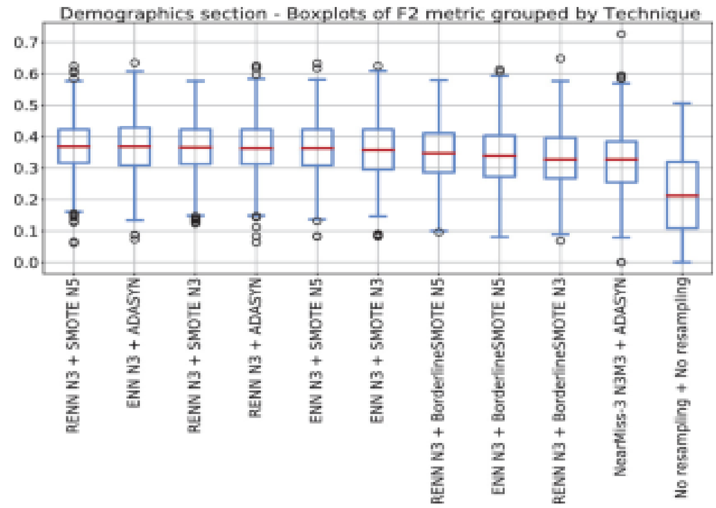

Figure 5. Comparison of Resampling methods accuracy scores [46]

Integrated method affirms an advancement in overcoming imbalanced datasets as the SMOTE version has restrictions [47]. For complex techniques, implementation of SMOTE method and hybridised resampling algorithms has been identified to aid increased effectiveness and minimized inaccurate sampling. The Boxplot and ROC curve used to assess the optimum resampling method showed a definite significance of the SENN method in a study reported by Sajad Khodabandelu et al., [48] . Since the models trained using the original data had inadequate predictive power for the positive class and had poor calibration compared to the balanced dataset using the Stomek and SENN technique, this factor resulted in the elimination of the other methods in favour of the Stomek resampling method. On consideration of generic approach irrespective of specific classification, under sampling techniques performed superior to over sampling. In individual resampling technique category, SMOTE excelled in over sampling and cluster centroid excelled in under sampling, according to the significant insights obtained. A SMOTE variant technique was proposed by Daochen Zha et al., [49] for 1000 iterations and the values were observed to excel the current state of art methods, which was also evident from other studies [50] that SMOTE has efficient performance over 1000 iterations. Furthermore, it is undeniable that there isn't a single machine learning method that can be ranked as the most effective at the top of the hierarchy. In applications with a defined domain, they can produce the best outcomes and hybrid techniques were highly recommend for advanced data balancing.

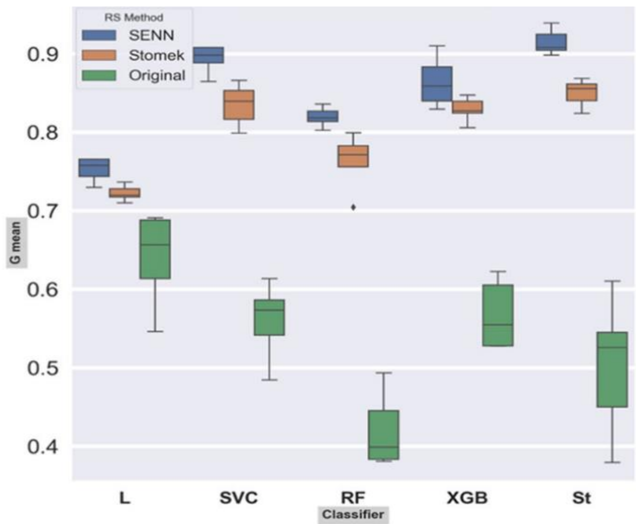

Figure 6. Box plot showing the performance of resampling technique implemented imbalanced data classification [48]

Received: August 13,2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Conclusion

The above discussed review study has paved way to establish the insights on the resampling techniques for health applications. The proliferation of datasets in recent years has led many industries to recognize them as valuable resources for gaining insights that can be used in strategic decision-making. In fields of healthcare involving screening and diagnosis, it is rarely apparent for the majority of instances to be identified with a single class, prompting the machine learning algorithm to develop classifiers that favour the majority class while overlooking the minority class labels. On consideration of generic approach irrespective of specific classification, under sampling techniques performed superior to over sampling. In individual resampling technique category, SMOTE excelled in over sampling and cluster centroid excelled in under sampling, according to the significant insights obtained.

Furthermore, it is undeniable that there isn't a single machine learning method that can be ranked as the most effective at the top of the hierarchy. In applications with a defined domain, they can produce the best outcomes and hybrid techniques were highly recommend for advanced data balancing. In such a consideration a remarkable inference could indeed be drawn, seeing that integrated/hybrid SMOTE possesses impressive adaptability for resampling techniques and nonetheless, due to the fact that it is known for certain that the core SMOTE version possesses limitations, and the integrated techniques serves to establish an improvisation.

Reference

1. N. Liu, X. Li, E. Qi, M. Xu, L. Li, and B. Gao, "A novel ensemble learning paradigm for medical diagnosis with imbalanced data," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3014362.
2. J. Piri and P. Mohapatra, "Imbalanced Cardiotocography Data Classification Using Re-sampling Techniques," 2021. doi: 10.1007/978-981-33-4087-9_55.
3. B. Song *et al.*, "Classification of imbalanced oral cancer image data from high-risk population," *J Biomed Opt*, vol. 26, no. 10, 2021, doi: 10.1117/1.jbo.26.10.105001.
4. K. Upreti *et al.*, "IoT-Based Control System to Measure, Analyze, and Track Basic Vital Indicators in Patient Healthcare Monitoring System," in *Communications in Computer and Information Science*, 2023. doi: 10.1007/978-3-031-25088-0_63.
5. K. Upreti, N. Kumar, M. S. Alam, A. Verma, M. Nandan, and A. K. Gupta, "Machine Learning-based Congestion Control Routing Strategy for Healthcare IoT Enabled Wireless Sensor Networks," in *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, IEEE, Sep. 2021, pp. 1–6. doi: 10.1109/ICECCT52121.2021.9616864.
6. Q. Feng, M. Du, N. Zou, and X. Hu, "Fair Machine Learning in Healthcare: A Review," Jun. 2022, doi: 10.48550/arxiv.2206.14397.
7. A. Singh, D. Singh, K. Upreti, V. Sharma, B. S. Rathore, and J. Raikwal, "Investigating New Patterns in Symptoms of COVID-19 Patients by Association Rule Mining (ARM)," *Journal of Mobile Multimedia*, Aug. 2022, doi: 10.13052/jmm1550-4646.1911.
8. S. F. Chong and R. Choo, "Introduction to Bootstrap," *StatiSticS Proceedings of Singapore Healthcare* ♣, vol. 20, no. 3, 2011.
9. N. S. Rahmi, N. W. S. Wardhani, M. B. Mitakda, R. S. Fauztina, and I. Salsabila, "SMOTE Classification and Random Oversampling Naive Bayes in Imbalanced Data: (Case Study of Early Detection of Cervical Cancer in Indonesia)," in *Proceedings of the 2022 IEEE 7th International Conference on Information Technology and Digital Applications, ICITDA 2022*, 2022. doi: 10.1109/ICITDA55840.2022.9971421.
10. A. Singh, D. Singh, K. Upreti, V. Sharma, B. S. Rathore, and J. Raikwal, "Investigating new patterns in symptoms of covid-19 patients by association rule mining (arm)," *Journal of Mobile Multimedia*, vol. 19, no. 1, 2023, doi: 10.13052/jmm1550-4646.1911.
11. M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review," *Healthcare (Basel)*, vol. 10, no. 3, Mar. 2022, doi: 10.3390/HEALTHCARE10030541.

Received: August 13,2025

12. K. Shin, J. Han, and S. Kang, "MI-MOTE: Multiple imputation-based minority oversampling technique for imbalanced and incomplete data classification," *Inf Sci (N Y)*, vol. 575, 2021, doi: 10.1016/j.ins.2021.06.043.
13. Y. Zeng and F. Cheng, "Medical and Health Data Classification Method Based on Machine Learning," *J Healthc Eng*, vol. 2021, 2021, doi: 10.1155/2021/2722854.
14. H. M. Safhi, B. Frikh, B. Hirchoua, B. Ouhbi, and I. Khalil, "Data intelligence in the context of big data: A survey," *Journal of Mobile Multimedia*, vol. 13, no. 1–2, 2017.
15. S. Belarouci and M. A. Chikh, "Medical imbalanced data classification," *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, no. 3, pp. 116–124, 2017, [Online]. Available: http://astesj.com/archive/volume-2/volume-2-issue-3/medical-imbalanced-data-classification/
16. W. Han, Z. Huang, S. Li, and Y. Jia, "Distribution-Sensitive Unbalanced Data Oversampling Method for Medical Diagnosis," *J Med Syst*, vol. 43, no. 2, 2019, doi: 10.1007/s10916-018-1154-8.
17. E. Elyan, C. F. Moreno-Garcia, and C. Jayne, "CDSMOTE: class decomposition and synthetic minority class oversampling technique for imbalanced-data classification," *Neural Comput Appl*, vol. 33, no. 7, 2021, doi: 10.1007/s00521-020-05130-z.
18. M. Kamaladevi and V. Venkatraman, "Tversky Similarity based Under Sampling with Gaussian Kernelized Decision Stump Adaboost Algorithm for Imbalanced Medical Data Classification," *International Journal of Computers, Communications and Control*, vol. 16, no. 6, 2021, doi: 10.15837/IJCCC.2021.6.4291.
19. Z. Yang *et al.*, "ProCo: Prototype-Aware Contrastive Learning for Long-Tailed Medical Image Classification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13438 LNCS, pp. 173–182, 2022, doi: 10.1007/978-3-031-16452-1_17/COVER.
20. N. Anuradha, G. Partha, and S. Varma, "PBCCUT- Priority based Class Clustered under Sampling Technique Approaches for Imbalanced Data Classification," *Indian J Sci Technol*, vol. 10, no. 18, pp. 1–9, May 2017, doi: 10.17485/IJST/2017/V10I18/107590.
21. S. Mittal, C. Monga, K. Upreti, N. Kumar, R. D. Raut, and M. S. Alam, "Light Weight Cryptography for Cloud-Based E-Health Records," in *2022 7th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, Jun. 2022, pp. 690–696. doi: 10.1109/ICCES54183.2022.9835827.
22. F. Alahmari, "A Comparison of Resampling Techniques for Medical Data Using Machine Learning," *https://doi.org/10.1142/S021964922040016X*, vol. 19, no. 1, Apr. 2020, doi: 10.1142/S021964922040016X.
23. S. M. Ganie and M. B. Malik, "An ensemble Machine Learning approach for predicting Type-II diabetes mellitus based on lifestyle indicators," *Healthcare Analytics*, vol. 2, p. 100092, Nov. 2022, doi: 10.1016/J.HEALTH.2022.100092.
24. A. A. Reshi, I. Ashraf, F. Rustam, H. F. Shahzad, A. Mehmood, and G. S. Choi, "Diagnosis of vertebral column pathologies using concatenated resampling with machine learning algorithms," *PeerJ Comput Sci*, vol. 7, 2021, doi: 10.7717/PEERJ-CS.547.
25. X. Shi, T. Qu, G. Van Pottelbergh, M. van den Akker, and B. De Moor, "A Resampling Method to Improve the Prognostic Model of End-Stage Kidney Disease: A Better Strategy for Imbalanced Data," *Front Med (Lausanne)*, vol. 9, p. 730748, Mar. 2022, doi: 10.3389/FMED.2022.730748/FULL.
26. G. Varotto, G. Susi, L. Tassi, F. Gozzo, S. Franceschetti, and F. Panzica, "Comparison of Resampling Techniques for Imbalanced Datasets in Machine Learning: Application to Epileptogenic Zone Localization From Interictal Intracranial EEG Recordings in Patients With Focal Epilepsy," *Front Neuroinform*, vol. 15, 2021, doi: 10.3389/fninf.2021.715421.
27. D. Barbieri *et al.*, "Predicting cardiovascular risk in athletes: Resampling improves classification performance," *Int J Environ Res Public Health*, vol. 17, no. 21, 2020, doi: 10.3390/ijerph17217923.
28. G. AlMahadin, A. Lotfi, M. M. Carthy, and P. Breedon, "Enhanced Parkinson's Disease Tremor Severity Classification by Combining Signal Processing with Resampling Techniques," *SN Computer Science 2021 3:1*, vol. 3, no. 1, pp. 1–21, Nov. 2021, doi: 10.1007/S42979-021-00953-6.

Received: August 13,2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

29. S. A. Mohammed, S. Darrab, S. A. Noaman, and G. Saake, “Analysis of Breast Cancer Detection Using Different Machine Learning Techniques,” *Data Mining and Big Data*, vol. 1234, p. 108, 2020, doi: 10.1007/978-981-15-7205-0_10.
30. C. S. Tarimo, S. S. Bhuyan, Q. Li, W. Ren, M. J. Mahande, and J. Wu, “Combining resampling strategies and ensemble machine learning methods to enhance prediction of neonates with a low apgar score after induction of labor in Northern Tanzania,” *Risk Manag Healthc Policy*, vol. 14, 2021, doi: 10.2147/RMHP.S331077.
31. J. D. Murphy *et al.*, “A Predictive Model of Noncardia Gastric Adenocarcinoma Risk Using Antibody Response to Helicobacter pylori Proteins and Pepsinogen,” *Cancer Epidemiology Biomarkers and Prevention*, vol. 31, no. 4, pp. 811–820, 2022, doi: 10.1158/1055-9965.EPI-21-0869.
32. F. Checchi *et al.*, “Can we predict the burden of acute malnutrition in crisis-affected countries? Findings from Somalia and South Sudan,” *BMC Nutr*, vol. 8, no. 1, pp. 1–20, 2022, doi: 10.1186/s40795-022-00563-2.
33. F. Sánchez-Hernández, J. C. Ballesteros-Herráez, M. S. Kraiem, M. Sánchez-Barba, and M. N. Moreno-García, “Predictive modeling of ICU healthcare-associated infections from imbalanced data. Using ensembles and a clustering-based undersampling approach,” *Applied Sciences (Switzerland)*, vol. 9, no. 24, 2019, doi: 10.3390/app9245287.
34. J. Liu, Z. S. Y. Wong, H. Y. So, and K. L. Tsui, “Evaluating resampling methods and structured features to improve fall incident report identification by the severity level,” *Journal of the American Medical Informatics Association*, vol. 28, no. 8, 2021, doi: 10.1093/jamia/ocab048.
35. Y. Zhao, Z. S. Y. Wong, and K. L. Tsui, “A Framework of Rebalancing Imbalanced Healthcare Data for Rare Events’ Classification: A Case of Look-Alike Sound-Alike Mix-Up Incident Detection,” *J Healthc Eng*, vol. 2018, 2018, doi: 10.1155/2018/6275435.
36. A. K. Ilavarasi, “Class imbalance learning for identity management in healthcare,” *Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC 2020*, pp. 995–1000, 2020, doi: 10.1109/I-SMAC49090.2020.9243420.
37. L. Li and G. Liu, “In-hospital Mortality Prediction for ICU Patients on Large Healthcare MIMIC Datasets Using Class Imbalance Learning,” *2020 5th IEEE International Conference on Big Data Analytics, ICBDA 2020*, pp. 90–93, 2020, doi: 10.1109/ICBDA49040.2020.9101272.
38. T. Kosolwattana *et al.*, “A Self-inspected Adaptive SMOTE Algorithm (SASMOTE) for Highly Imbalanced Data Classi cation in Healthcare A Self-inspected Adaptive SMOTE Algorithm (SASMOTE) for Highly Imbalanced Data Classification in Healthcare,” 2022.
39. M. Reeves, H. S. Bhat, and S. Goldman-Mellor, “Resampling to address inequities in predictive modeling of suicide deaths,” *BMJ Health Care Inform*, vol. 29, no. 1, p. e100456, Apr. 2022, doi: 10.1136/BMJHCI-2021-100456.
40. U. G. Mohammad, S. Imtiaz, M. Shakya, A. Almadhor, and F. Anwar, “An Optimized Feature Selection Method Using Ensemble Classifiers in Software Defect Prediction for Healthcare Systems,” *Wirel Commun Mob Comput*, vol. 2022, 2022, doi: 10.1155/2022/1028175.
41. P. Wibowo and C. Fatichah, “Pruning-based oversampling technique with smoothed bootstrap resampling for imbalanced clinical dataset of Covid-19,” *Journal of King Saud University - Computer and Information Sciences*, 2021, doi: 10.1016/j.jksuci.2021.09.021.
42. M. Khushi *et al.*, “A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data,” *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3102399.
43. M. Lamari *et al.*, “SMOTE–ENN-Based Data Sampling and Improved Dynamic Ensemble Selection for Imbalanced Medical Data Classification,” in *Advances in Intelligent Systems and Computing*, 2021. doi: 10.1007/978-981-15-6048-4_4.
44. M. Qian and Y. F. Li, “A Weakly Supervised Learning-Based Oversampling Framework for Class-Imbalanced Fault Diagnosis,” *IEEE Trans Reliab*, vol. 71, no. 1, 2022, doi: 10.1109/TR.2021.3138448.

Received: August 13,2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

45. R. Zhou *et al.*, "Prediction Model for Infectious Disease Health Literacy Based on Synthetic Minority Oversampling Technique Algorithm," *Comput Math Methods Med*, vol. 2022, 2022, doi: 10.1155/2022/8498159.
46. A. B. Fonseca, D. C. Martins-Jr, Z. Wicik, M. Postula, and S. N. Simões, "Addressing Classification on Highly Imbalanced Clinical Datasets," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2022. doi: 10.1007/978-3-031-17531-2_9.
47. A. S. Tarawneh, A. B. Hassanat, G. A. Altarawneh, and A. Almuhaimeed, "Stop Oversampling for Class Imbalance Learning: A Review," *IEEE Access*, vol. 10, pp. 47643–47660, 2022, doi: 10.1109/ACCESS.2022.3169512.
48. S. Khodabandelu, Z. Basirat, S. Khaleghi, S. Khafri, H. Montazery Kordy, and M. Golsorkhtabaramiri, "Developing machine learning-based models to predict intrauterine insemination (IUI) success by address modeling challenges in imbalanced data and providing modification solutions for them," *BMC Med Inform Decis Mak*, vol. 22, no. 1, 2022, doi: 10.1186/s12911-022-01974-8.
49. D. Zha, K. H. Lai, Q. Tan, S. Ding, N. Zou, and X. Ben Hu, "Towards Automated Imbalanced Learning with Deep Hierarchical Reinforcement Learning," in *International Conference on Information and Knowledge Management, Proceedings*, 2022. doi: 10.1145/3511808.3557474.
50. R. T. Nakatsu, "An Evaluation of Four Resampling Methods Used in Machine Learning Classification," *IEEE Intell Syst*, vol. 36, no. 3, 2021, doi: 10.1109/MIS.2020.2978066.

Received: August 13,2025