Scientific Journal of Informatics
Vol. 11, No. 4, Nov 2024
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Evaluating Ensemble Learning Techniques for Class Imbalance in Machine Learning: A Comparative Analysis of Balanced Random Forest, SMOTE-RF, SMOTEBoost, and RUSBoost

Tahira Fulazzaky[1], Asep Saefuddin[2*], Agus Mohamad Soleh[3]
[1, 2, 3]Department of Statistics and Data Science, IPB University, Indonesia

Abstract.
Purpose: This research aims to identify the optimal ensemble learning method for mitigating class imbalance in datasets utilizing various advanced techniques which include balanced random forest (BRF), SMOTE-random forest (SMOTE-RF), RUSBoost, and SMOTEBoost. The methods were systematically evaluated against conventional algorithms, including random forest and AdaBoost, across heterogeneous datasets with varying class imbalance ratios.
Methods: This study utilized 13 secondary datasets from diverse sources, each with binary class outputs. The datasets exhibited varying degrees of class imbalance, offering scenarios to assess the effectiveness of ensemble learning techniques and traditional machine learning approaches in managing class imbalance issues. Study data were split into training (80%) and testing (20%), with stratified sampling applied to maintain consistent class proportions across both sets. Each method underwent hyperparameter optimization with distinct settings with repetition over 10 iterations. The optimal method was evaluated based on balanced accuracy, recall, and computation time.
Result: Based on the evaluation, the BRF method exhibited the highest performance in balanced accuracy and recall when compared to SMOTE-RF, RUSBoost, SMOTEBoost, random forest, and AdaBoost. Conversely, the classical random forest method outperformed other techniques in terms of computational efficiency.
Novelty: This study presents an innovative analysis of advanced ensemble learning techniques, including BRF, SMOTE-random forest, SMOTEBoost, and RUSBoost, which demonstrate significant effectiveness in addressing class imbalance across various datasets. By systematically optimizing hyperparameters and applying stratified sampling, this research produces findings that redefine the benchmarks of balanced accuracy, recall and computational efficiency in machine learning.

Keywords: Machine learning, Balanced random forest, SMOTE-RF, SMOTEBoost, RUSBoost, Random forest, Adaboost, Imbalaced data, Ensemble learning
Received November 2024 / **Revised** December 2024 / **Accepted** December 2024

INTRODUCTION

In real-world scenarios, class imbalance is a pervasive challenge when working with data, particularly in domains where the distribution of classes is significantly skewed. Imbalanced datasets often lead to machine learning models that prioritize the majority class, reducing their ability to make accurate predictions for the minority class [1]. This is a critical issue in high-stake fields such as healthcare, financial fraud detection, and disaster forecasting, where errors in identifying minority class instances, such as rare diseases, fraudulent transactions, or severe weather events, have severe consequences. For instance, failure to correctly identify patients with rare conditions could delay essential treatments, or overlooking fraudulent activities could lead to substantial financial losses [2], [3]. Consequently, addressing this imbalance is not merely a technical challenge but an ethical and operational necessity in ensuring fairness, reliability, and accuracy in decision-making systems [4].

While traditional approaches to addressing class imbalance often focus on data-level techniques such as under-sampling or over-sampling, these methods inherently alter the dataset by either removing instances from the majority class or artificially generating new ones for the minority class [5]. Such approaches, while useful, may not fully resolve the inherent difficulty of imbalanced classification tasks. Instead, this research emphasizes the development and evaluation of machine learning models specifically designed to adapt to

*Corresponding author.
Email addresses: tahirafulazzaky@apps.ipb.ac.id (Fulazzaky), asaefuddin@apps.ipb.ac.id (Saefuddin)*, agusms@apps.ipb.ac.id (Soleh)
DOI: 10.15294/sji.v11i4.15937

imbalanced datasets. This focus ensures that the models themselves are robust to skewed distributions without over-reliance on dataset modifications, thereby preserving the original data characteristics and minimizing information loss [4].

Machine learning techniques capable of adapting to class imbalance include ensemble learning methods such as SMOTEBoost, RUSBoost, and balanced random forest (BRF), which combine advanced sampling strategies with robust model architectures. These methods enhance predictive performance for minority classes by leveraging ensemble techniques like boosting and bagging, while addressing bias toward the majority class [6]. Among the methods, SMOTEBoost integrates the synthetic minority oversampling technique (SMOTE) with boosting, targeting improvements in both minority class recall and overall model accuracy [7]. Conversely, RUSBoost combines random under-sampling with boosting, offering computational efficiency without sacrificing predictive accuracy, as demonstrated in applications like machinery fault identification [8], [9]. A variant of random forest, BRF incorporates balanced subsets of the majority class during model training, ensuring fair treatment of all classes and enhancing model performance on imbalanced data [10].

This study evaluates the performance of advanced machine learning models on datasets from diverse domains, each characterized by varying degrees of class imbalance. The datasets include critical cases in healthcare, such as cerebral stroke, thyroid cancer, and diabetes, as well as datasets from other fields such as divorce prediction and cervical cancer. These real-world datasets present unique challenges in identifying minority class instances, underscoring the need for robust methods to ensure fairness and reliability in predictive modelling. By covering a broad spectrum of dataset sizes and imbalance levels, this study highlights the versatility of the proposed approaches in tackling imbalanced classification tasks. Through hyperparameter optimization, stratified sampling, and extensive experimentation over 10 iterations, this study aims to identify the optimal methods for imbalanced classification tasks. By prioritizing machine learning models that inherently adapt to class imbalance, this research seeks to advance the reliability and applicability of predictive systems in critical, real-world scenarios. The evaluation metrics utilized in this study include balanced accuracy, recall, and computation time, ensuring a comprehensive assessment of model performance.

METHODS

Flowchart Analysis

The framework displayed in Figure 1 depicts the stages of data processing involving 13 datasets with varying levels of class imbalance. The initial stage consisted of data pre-processing, which included data cleaning and label encoding. The data were subsequently divided into training (80%) and testing (20%). The data training process was separated into two main categories: data processed with class balancing techniques and data without class balancing. For class balancing, the methods included balanced random forest (BRF), SMOTE-random forest (SMOTE-RF), SMOTEBoost, and RUSBoost. For comparison, unbalanced data were modeled using the random forest (RF) and AdaBoost algorithms. During the training phase, hyperparameter tuning was performed using randomized search cross-validation (CV) with 10 iterations (see Table 1). Following the training, testing data were utilized to evaluate the performance of the models based on balanced accuracy, recall, and computation time.

Table 1. Hyperparameter settings for ensemble algorithms in imbalanced data classification

Algorithm	Hyperparameter
Balanced Random Forest, SMOTE-RF, and RF	N Estimators: 50,100, 200 Max Features: sqrt Max Depth: 10, 20, 30 Min Samples Split: 2, 5, 10 Min Samples Leaf: 1, 2, 4 Bootstrap: True
SMOTEBoost, AdaBoost	N Estimators: 50,100, 150 Learning Rate: 0.01, 0.1, 0.5 Estimator Max Depth: 1, 2, 3
RUSBoost	N Estimators: 50, 100, 200 Learning Rate: 0.01, 0.1, 1 Estimator Max Depth: 3, 5, 10 Estimator Min Samples Split: 2, 5, 10 Estimator Min Samples Leaf: 1, 2, 4 Estimator Max Features: sqrt

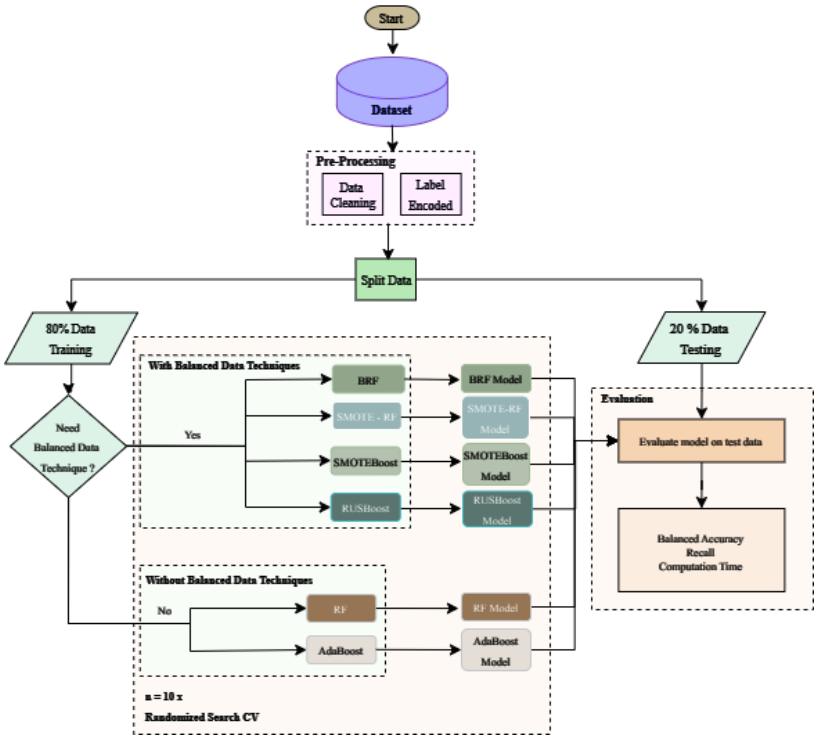

Figure 1. Flowchart analysis

Meanwhile, model performance was measured using balanced accuracy, recall and computation time. The performance value was calculated as follows:

1. Balanced accuracy is the average of sensitivity (true positive rate) and specificity (true negative rate) [11], which is defined using the following equation:

$$Balanced\ Accuracy = \frac{1}{2}\left(\frac{TP}{(TP+FN)} + \frac{TN}{(TN+FP)}\right) \quad (1)$$

2. Recall, also known as sensitivity or true positive rate, measures the proportion of actual positives correctly identified by the model [11], which is defined using the following equation:

$$Recall = \left(\frac{TP}{TP+FN}\right) \quad (2)$$

Dataset

The research utilized a diverse collection of 13 datasets spanning across various medical and behavioral domains to evaluate the performance of the proposed models (see Table 2). The datasets were selected to represent a wide range of class imbalance levels and data complexities, with key characteristics including diverse medical conditions, varying sample sizes, and imbalanced class distributions. By leveraging such diverse dataset, the research aimed to provide a comprehensive evaluation of the proposed class balancing techniques and ensemble models in improving classification performance, particularly for minority classes.

Table 1. List of datasets

No	Dataset Name	Previous Research	Total Observation
1	Cerebral Stroke	Liu [12]	43400
2	Monkeypox	Ahmed [13]	25000
3	Brain Stroke	Pathan et al. [14]	4981
4	Diabetes	Smith et al. [15]	768
5	Early-Stage Diabetes	Islam et al. [16]	520
6	Thyroid Cancer	Borzooei et al. [17]	383
7	Heart Disease	Detrano et al. [18]	303
8	Alpha Thalassemia	Kolambage et al. [19]	203
9	Divorce Prediction	Yontem et al [20]	170
10	Breast Cancer	Patricio et al. [21]	116
11	Fertility	Gil et al. [22]	100
12	Cervical Cancer	Çakır et al [23]	69
13	Screen Time	Bailey [24]	28

Ensemble Learning

As described by Brown [6], ensemble learning combines multiple classifiers to enhance predictive accuracy by mitigating individual model limitations. Bagging, introduced by Breiman [25], reduces variance by training models on randomly sampled datasets and aggregating predictions via averaging or voting. Boosting, introduced by Freund and Schapire [26], sequentially refines models by focusing on correcting prior errors. AdaBoost, a key boosting algorithm, assigns greater weight to misclassified instances in subsequent iterations, improving classification accuracy by adapting to error patterns.

Random Forest

Introduced by Breiman [25], random forest improves bagging by reducing overfitting through random variable selection. Unlike bagging, which uses all variables for each decision tree, random forest selects a subset of predictors (m, where $m < d$, with d representing the total number of variables) at each node split, decreasing tree correlation and enhancing generalization. The algorithm uses bootstrap sampling to generate n samples, builds unpruned decision trees with m predictors per split, and repeats this k times to create a forest of k trees. The final prediction is generated by majority voting, resulting in a robust model with improved performance across datasets [25].

Balanced Random Forest

Balanced random forest (BRF) is an adaptation of Breiman's random forest [25] which addresses challenges in highly imbalanced datasets. While standard random forests may struggle with minority class prediction, BRF improves accuracy by equally sampling from both minority and majority classes during tree construction. Proposed by Chen [10], this technique incorporates cost-sensitive learning and follows the CART algorithm, selecting m_{try} random descriptors per split and avoiding tree pruning, thereby enhancing performance on imbalanced datasets.

SMOTE

Synthetic minority oversampling technique (SMOTE), developed by Chawla et al. [7], addresses imbalanced datasets by generating synthetic samples for the minority class, reducing overfitting and preserving information. For numerical variables, SMOTE calculates Euclidean distance to find k-nearest neighbors and scales differences by a random factor to create new samples, while for categorical variables, the value difference metric (VDM) is applied to compute distances, followed by majority voting to generate synthetic values [27]. This technique improves class balance and enhances model performance on imbalanced datasets.

SMOTEBoost

SMOTEBoost, introduced by Chawla et al. [7], combines SMOTE with boosting to address class imbalance while maintaining classification accuracy. Unlike conventional boosting, which increases weights for misclassified instances equally, SMOTEBoost generates synthetic samples from the minority class in each iteration, enhancing model focus on minority class correction. It adjusts weight distributions, trains weak learners using CART-based decision trees, and outputs a final hypothesis that maximizes the weighted sum of weak hypotheses, improving classification of minority classes and model robustness in imbalanced settings [7].

RUSBoost

RUSBoost, introduced by Seiffert et al. [28], combines the principles of random under-sampling (RUS) and boosting to enhance classification in imbalanced datasets. In each iteration, RUS reduces the majority class size to create a balanced dataset before applying boosting, which emphasizes misclassified instances by adjusting weights [29]. The iterative process generates hypotheses with updated weights to minimize errors, focusing on improved accuracy for the minority class. This approach effectively addresses severe imbalances by balancing learning and reducing classification errors [28].

Imbalanced Ratio

Imbalance ratio (IR) quantifies class imbalance in a dataset as the ratio of the majority class size (N_{maj}) to the minority class size (N_{min}), represented by the following formula:

$$IR = \frac{N_{maj}}{N_{min}} \tag{3}$$

An IR of 1 indicates a balanced dataset, while higher values reflect greater imbalance, with larger IRs signifying majority-class dominance. This imbalance challenges standard classifiers, which often favor the majority class, reducing accuracy for minority class predictions [30].

RESULTS AND DISCUSSIONS

Table 3 presents class distributions and imbalance ratios (IR) for 13 health-related datasets. The Cerebral Stroke dataset shows the highest imbalance (IR = 54.43), with 42,617 negative and 783 positive instances, while the Divorce Prediction dataset is nearly balanced (IR = 1.02). Other datasets, such Brain Stroke (IR = 19.08) and Fertility (IR = 7.33), also exhibit significant imbalances. Most datasets demonstrate varying degrees of imbalance, posing challenges for predictive modeling.

Table 2. Class distribution and imbalance ratios for various health datasets

Data	Class Distribution		IR	Category
	Negative	Positive		
Cerebral Stroke	42617	783	54.43	Highly Imbalanced
Brain Stroke	4733	248	19.08	Highly Imbalanced
Fertility	88	12	7.33	Moderately Imbalanced
Alpha Thalassemia	55	148	2.69	Moderately Imbalanced
Thyroid Cancer	275	108	2.55	Moderately Imbalanced
Cervical Cancer	48	21	2.29	Moderately Imbalanced
Diabetes	500	268	1.87	Slightly Imbalanced
Monkeypox	9091	15909	1.75	Slightly Imbalanced
Early-Stage Diabetes	200	320	1.6	Slightly Imbalanced
Screen Time	12	16	1.33	Balanced
Breast Cancer	52	64	1.23	Balanced
Heart Disease	138	165	1.2	Balanced
Divorce Prediction	86	84	1.02	Balanced

Highly Imbalanced

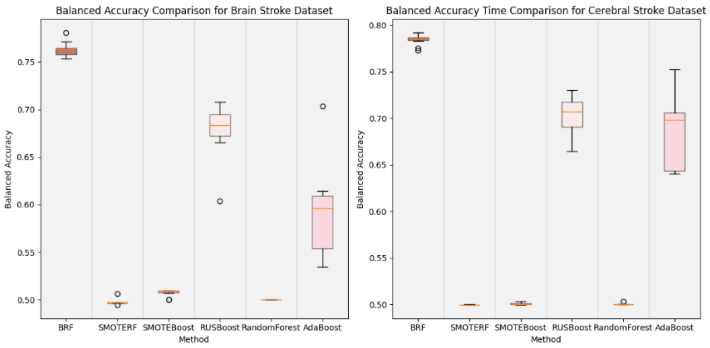

Figure 2. Balanced accuracy comparison across classification methods for highly imbalanced datasets

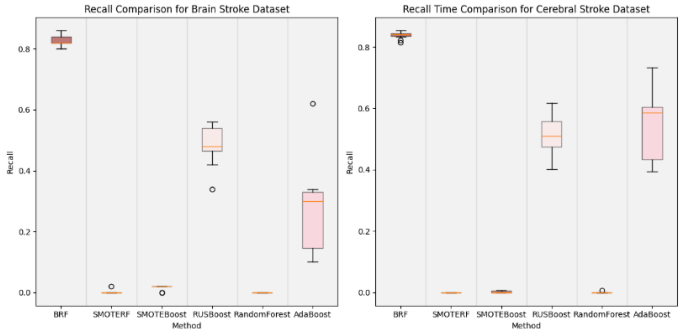

Figure 3. Recall comparison across classification methods for highly imbalanced datasets

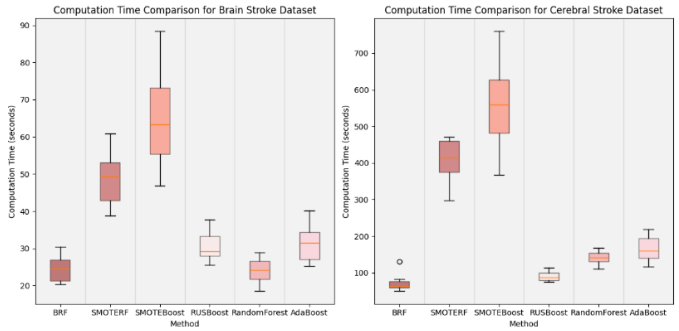

Figure 4. Computation time comparison across classification methods for highly imbalanced datasets

The balanced random forest (BRF) method consistently outperformed other classification techniques across multiple metrics for highly imbalanced datasets, as shown in Figures 2, 3, and 4. In terms of balanced accuracy (Figure 2), BRF achieved the highest scores on both the Brain Stroke and Cerebral Stroke datasets, surpassing SMOTE-RF, SMOTEBoost, random forest, RUSBoost, and AdaBoost. Similarly, BRF demonstrated superior recall performance (Figure 3), maintaining consistently high and stable values across both datasets, while RUSBoost and AdaBoost showed moderate performance, and other methods such as SMOTE-RF, SMOTEBoost, and random forest fell behind. In addition to its accuracy and recall advantages, BRF was also efficient in computation time (Figure 4), achieving lower runtimes compared to RUSBoost, SMOTEBoost, and SMOTE-RF, particularly on the Cerebral Stroke dataset. Overall, BRF emerged as the most effective and efficient method for handling highly imbalanced data.

Moderately Imbalanced

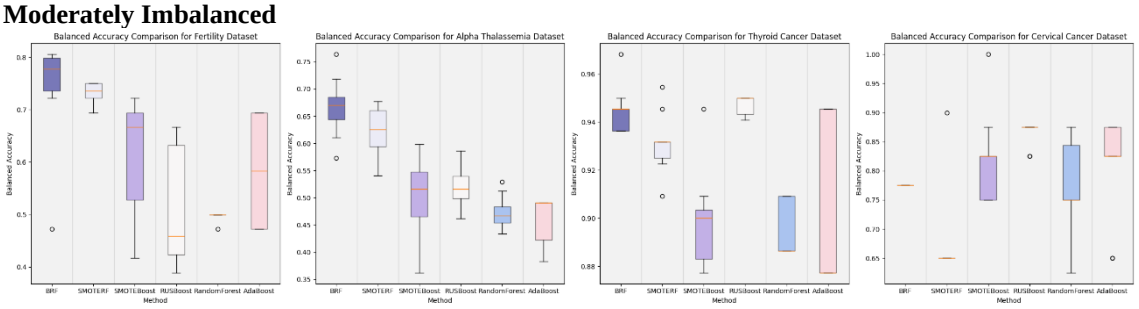

Figure 5. Balanced accuracy comparison across classification methods for moderately imbalanced datasets

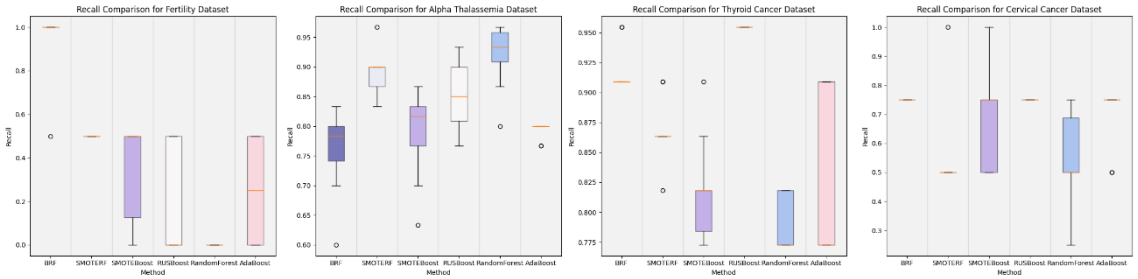

Figure 6. Recall comparison across classification methods for moderately imbalanced datasets

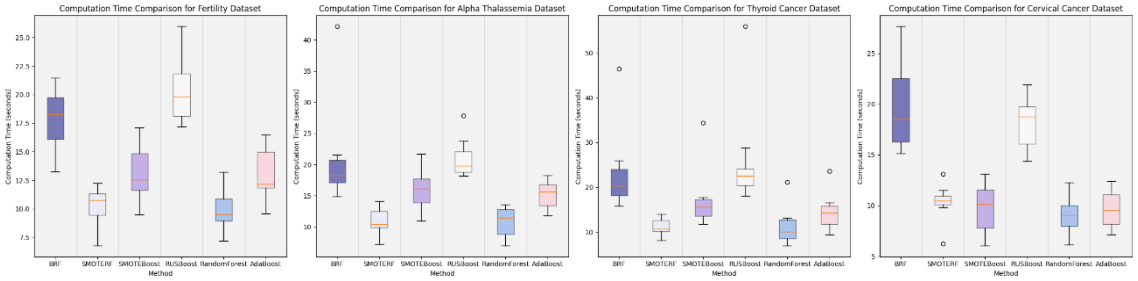

Figure 7. Computation time comparison across classification methods for moderately imbalanced datasets

For moderately imbalanced datasets, the balanced random forest (BRF) method consistently outperformed others in balanced accuracy, recall, and overall effectiveness, as shown in Figures 5, 6, and 7. BRF achieved high and stable balanced accuracy across datasets (Figure 5), surpassing methods such as SMOTE-RF, SMOTEBoost, AdaBoost, RUSBoost, and random forest, which displayed variable or lower performance. Similarly, BRF demonstrated superior recall (Figure 6), effectively capturing positive cases and minimizing false negatives, while SMOTE-RF and RUSBoost performed moderately well but less consistently, and random forest and AdaBoost lagged behind. In terms of computation time (Figure 7), BRF required more resources than certain methods, such as SMOTE-RF and random forest, but its performance advantages justified the trade-off. Overall, BRF proved to be the most reliable and effective choice for moderately imbalanced datasets, offering robust generalization and superior results across key metrics.

Slightly Imbalanced

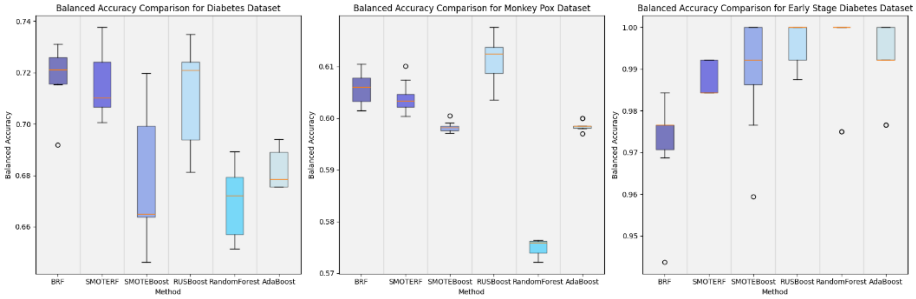

Figure 8. Balanced accuracy comparison across classification methods for slightly imbalanced datasets

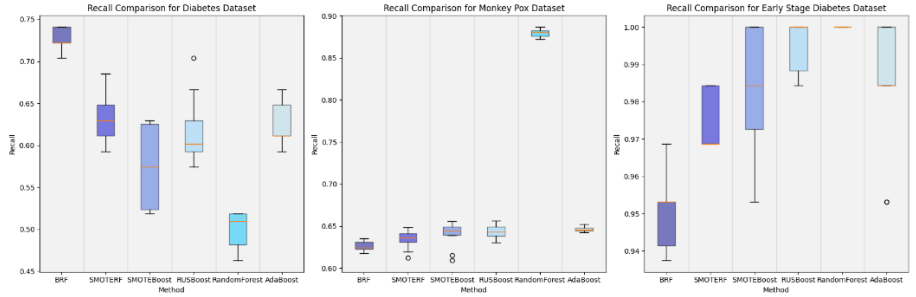

Figure 9. Recall comparison across classification methods for slightly imbalanced datasets

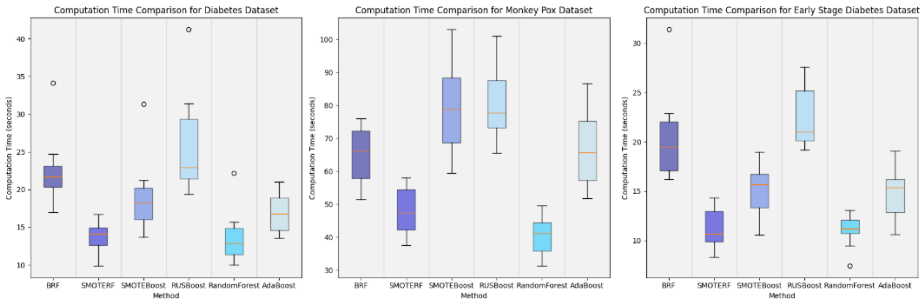

Figure 10. Computation time comparison across classification methods for slightly imbalanced datasets

For slightly imbalanced datasets, classification methods demonstrated varying performance across metrics, as shown in Figures 8, 9, and 10. In terms of balanced accuracy (Figure 8), BRF and SMOTERF performed consistently well across datasets, with random forest, AdaBoost, and RUSBoost excelling in the Early-Stage Diabetes dataset but showing mixed results in others. For recall (Figure 9), random forest and AdaBoost exhibited strong performance across datasets, while BRF and SMOTERF demonstrated the highest recall in the Diabetes dataset, and random forest excelled in the Monkeypox dataset. Computation time analysis (Figure 10) highlighted random forest as the most efficient method, making it favorable for time-sensitive applications. Overall, BRF and SMOTERF emerged as reliable choices for balanced accuracy, while RUSBoost, random forest, and AdaBoost offered strong recall and efficiency, providing flexibility based on dataset characteristics and computational constraints.

Balanced

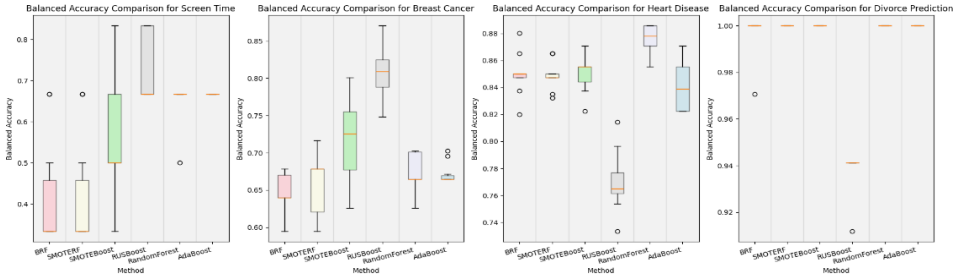

Figure 11. Balanced accuracy comparison across classification methods for balanced datasets

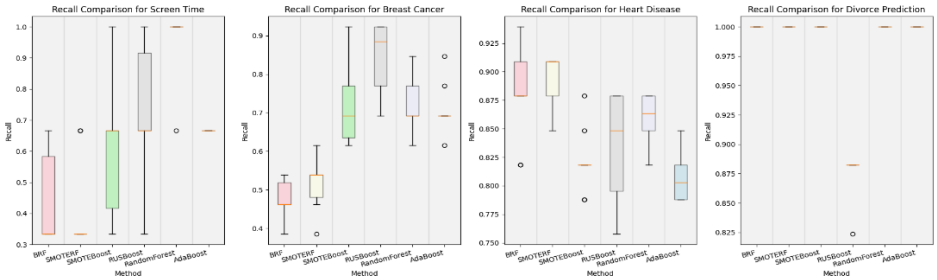

Figure 12. Recall comparison across classification methods for balanced datasets

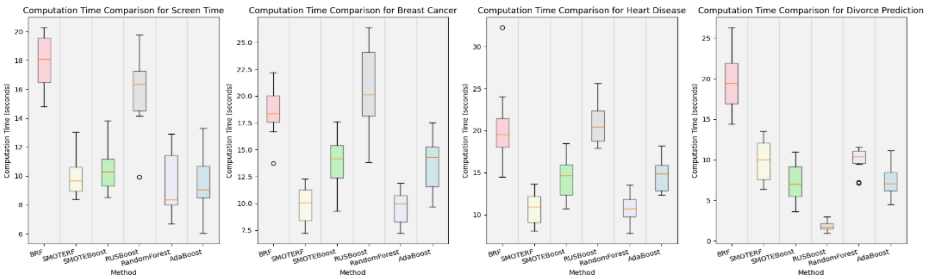

Figure 13. Computation time comparison across classification methods for balanced datasets

For balanced datasets, classification performance varied across domains, as shown in Figures 11, 12, and 13. In terms of balanced accuracy (Figure 11), RUSBoost and random forest performed well in the Screen Time and Breast Cancer datasets, with random forest and SMOTEBoost excelling in the Heart Disease dataset. All methods achieved near-perfect accuracy in the Divorce Prediction dataset, indicating minimal variation. Recall performance (Figure 12) was dataset-dependent, with RUSBoost and random forest excelling in the Screen Time dataset, RUSBoost leading in the Breast Cancer dataset, and BRF and SMOTERF performing strongly in the Heart Disease dataset. Computation time analysis (Figure 13) highlighted SMOTE-RF and RUSBoost as efficient options for most datasets, while all methods performed efficiently in the Divorce Prediction dataset. Overall, SMOTE-RF stood out as a versatile choice due to its balance between accuracy, recall, and efficiency, with RUSBoost and Random Forest also demonstrating strong, dataset-specific performance.

The study evaluated classification methods across datasets with varying levels of imbalance, revealing key insights into their performance. Balanced random forest (BRF) consistently outperformed other methods in highly and moderately imbalanced datasets, achieving the highest balanced accuracy and recall while maintaining efficient computation times. For example, BRF excelled on the Cerebral Stroke and Brain Stroke datasets with balanced accuracy of 0.767 and 0.752, respectively, demonstrating its ability to handle severe class imbalances and minimize false negatives.

For slightly imbalanced datasets, BRF remained competitive, though methods such as SMOTE-RF and random forest occasionally excelled in recall and efficiency. In balanced datasets, SMOTE-RF demonstrated strong performance, particularly in recall and runtime, making it a versatile choice.

Overall, the robust performance of BRF across imbalanced datasets makes it the optimal method for scenarios requiring high sensitivity, particularly in health-related applications. However, for balanced datasets or recall-specific tasks, SMOTE-RF and Random Forest are effective alternatives. These findings highlight the importance of adapting classification strategies to dataset characteristics for improved predictive accuracy.

Table 3. Average performance comparison of classification methods across datasets

Category	Dataset	Metric	BRF	SMOTE-RF	SMOTEBoost	RUSBoost	Random Forest	Boosting
Highly Imbalanced	Cerebral Stroke	Computation Time	26.68372064	21.11862824	25.70731282	23.40711932	**16.45948799**	19.90561016
		Balanced Accuracy	0.767064843	0.756481503	0.740996076	**0.791137439**	0.773031066	0.779992687
		Recall	0.751508835	0.699821466	0.620831907	0.753918585	**0.756229864**	0.710862144
	Brain Stroke	Computation Time	32.31407216	65.97960796	82.91335292	39.10716019	**31.00785823**	35.88805466
		Balanced Accuracy	**0.752296486**	0.692381818	0.676933156	0.711269142	0.64807569	0.67375541
		Recall	**0.785855411**	0.559574941	0.577039944	0.643774556	0.535176497	0.625185184
Moderately Imbalanced	Fertility	Computation Time	29.09703357	63.1871465	58.36156118	**23.73165534**	29.57558353	24.20393076
		Balanced Accuracy	**0.743508107**	0.695504418	0.659653902	0.725701921	0.682191938	0.713180081
		Recall	**0.770389107**	0.56994043	0.537814556	0.706059854	0.585037555	0.652574624
	Alpha Thalassemia	Computation Time	29.0389447	58.22939646	93.70887485	31.35031409	**26.59868901**	39.15483172
		Balanced Accuracy	**0.790523573**	0.712051131	0.746036486	0.757985181	0.711053578	0.707026654
		Recall	**0.825683339**	0.629243972	0.680391803	0.718120546	0.668070153	0.63877728
	Thyroid Cancer	Computation Time	38.11392124	56.94323404	101.9947126	32.10722396	**28.38992391**	40.39214439
		Balanced Accuracy	0.759270798	0.716584176	0.679079388	**0.764038081**	0.693392922	0.712773183
		Recall	**0.766827237**	0.586303856	0.55005677	0.740177897	0.550863766	0.632142416
	Cervical Cancer	Computation Time	**23.40567126**	50.70105472	79.76590195	23.70921955	23.5634732	33.84901135
		Balanced Accuracy	**0.810583012**	0.75340821	0.715047077	0.746745185	0.710275077	0.750932905
		Recall	**0.834671776**	0.634440398	0.58002051	0.700334575	0.576537167	0.704017416
Slightly Imbalanced	Diabetes	Computation Time	31.84821064	53.86345949	91.57508719	36.87196167	**25.86023686**	42.09002309
		Balanced Accuracy	**0.776416915**	0.684420735	0.663970552	0.750587945	0.685357706	0.734431913
		Recall	**0.791315545**	0.539247411	0.535706971	0.654229864	0.579952851	0.664678628
	Monkeypox	Computation Time	20.89568496	15.42915733	25.48141849	27.57898362	**12.56149824**	20.79321845
		Balanced Accuracy	0.739330399	0.695567492	0.7501962	**0.752329635**	0.712952025	0.743716163
		Recall	**0.743724544**	0.587539255	0.629174178	0.68623105	0.629025977	0.664554649

	Early-Stage Diabetes	Computation Time	27.53061454	56.98281481	74.07802916	28.8262861	**27.08085132**	34.11726229
		Balanced Accuracy	0.713690543	0.632584649	0.67561004	**0.719897742**	0.633069739	0.707721774
		Recall	**0.734423992**	0.518802802	0.574600999	0.690719822	0.552239767	0.678071323
Balanced	Screen Time	Computation Time	22.76719124	40.35046687	70.14733424	37.38072104	**20.90039985**	30.60396476
		Balanced Accuracy	0.768787305	0.720815368	0.735338876	**0.77686292**	0.737755836	0.757179392
		Recall	**0.788658848**	0.667201013	0.708707864	0.779714318	0.742324968	0.74273353
	Breast Cancer	Computation Time	33.20523789	66.5057796	66.41647193	30.77579589	32.07790792	**28.223875**
		Balanced Accuracy	**0.761352875**	0.711432334	0.713922512	0.75905335	0.686566955	0.713617511
		Recall	**0.80220525**	0.609903892	0.634304257	0.701614178	0.613177914	0.631765086
	Heart Disease	Computation Time	22.5519891	14.40934954	19.91664453	27.25697479	**13.00396631**	17.82963984
		Balanced Accuracy	0.757508687	0.738060663	0.777520646	**0.806116369**	0.746066947	0.776415559
		Recall	0.759591359	0.701606402	0.793864388	**0.812964354**	0.798530847	0.78759821
	Divorce Prediction	Computation Time	24.53392148	50.85495265	54.91413431	32.95188959	**23.96510623**	26.46778505
		Balanced Accuracy	**0.742768889**	0.68357563	0.712041084	0.739119182	0.663820916	0.714953368
		Recall	**0.778049717**	0.564026122	0.60116577	0.718128072	0.604489516	0.685883234

CONCLUSION

This study investigated the effectiveness of various classification methods across datasets with different levels of class imbalance, achieving the research objectives of identifying optimal techniques for handling imbalanced data. The balanced random forest (BRF) method consistently emerged as the top-performing model, particularly in highly and moderately imbalanced datasets, due to its superior balanced accuracy, recall, and efficiency. For balanced datasets, SMOTE-RF demonstrated strong recall and computational efficiency, providing a versatile alternative. The findings confirm the importance of model adaptability in achieving robust predictive performance.

The study's results underscore the significance of selecting methods tailored to dataset characteristics, particularly for health-related applications where accurate classification of minority classes is critical. By identifying BRF as the optimal model for imbalanced data, the research contributes to advancing predictive modeling in critical areas such as stroke diagnosis, cancer prediction, and chronic disease monitoring.

Future studies could focus on further improving computational efficiency for large-scale imbalanced datasets and exploring hybrid approaches that combine the strengths of multiple classification methods. Additionally, applying these techniques to real-world clinical data and incorporating advanced feature engineering or deep learning methods could provide valuable insights and enhance predictive accuracy. These directions offer exciting opportunities for advancing the field of imbalanced data classification and its applications in health and beyond.

REFERENCES

[1] C. X. Ling and V. S. Sheng, "Class Imbalance Problem BT - Encyclopedia of Machine Learning and Data Mining," C. Sammut and G. I. Webb, Eds., Boston, MA: Springer US, 2017, pp. 204–205. doi: 10.1007/978-1-4899-7687-1_110.

[2] S. Choi, Y. J. Kim, S. Briceno, and D. Mavris, "Prediction of weather-induced airline delays based on machine learning algorithms," in *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, 2016, pp. 1–6. doi: 10.1109/DASC.2016.7777956.

[3] B. Krawczyk, M. Galar, Ł. Jeleń, and F. Herrera, "Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy," *Appl. Soft Comput.*, vol. 38, pp. 714–726, 2016, doi: https://doi.org/10.1016/j.asoc.2015.08.060.

[4] V. Werner de Vargas, J. A. Schneider Aranda, R. dos Santos Costa, P. R. da Silva Pereira, and J. L. Victória Barbosa, "Imbalanced data preprocessing techniques for machine learning: a systematic mapping study," *Knowl. Inf. Syst.*, vol. 65, no. 1, pp. 31–57, 2023, doi: 10.1007/s10115-022-01772-8.

[5] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," in *2020 11th International Conference on Information and Communication Systems (ICICS)*, 2020, pp. 243–248. doi: 10.1109/ICICS49469.2020.239556.

[6] G. Brown, "Ensemble Learning BT - Encyclopedia of Machine Learning and Data Mining," C.

Sammut and G. I. Webb, Eds., Boston, MA: Springer US, 2017, pp. 393–402. doi: 10.1007/978-1-4899-7687-1_252.
[7] N. V Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving Prediction of the Minority Class in Boosting BT - Knowledge Discovery in Databases: PKDD 2003," N. Lavrač, D. Gamberger, L. Todorovski, and H. Blockeel, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 107–119.
[8] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: Improving classification performance when training data is skewed," in *2008 19th International Conference on Pattern Recognition*, 2008, pp. 1–4. doi: 10.1109/ICPR.2008.4761297.
[9] Q. Zhou, "Fault prediction of industrial machinery and equipment based on RUSBoost," in *Proceedings of the 2023 3rd International Conference on Bioinformatics and Intelligent Computing*, in BIC '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 22–26. doi: 10.1145/3592686.3592691.
[10] C. Chen, "Using Random Forest to Learn Imbalanced Data," 2004.
[11] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The Balanced Accuracy and Its Posterior Distribution," in *2010 20th International Conference on Pattern Recognition*, 2010, pp. 3121–3124. doi: 10.1109/ICPR.2010.764.
[12] T. Liu, W. Fan, and C. Wu, "A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset.," *Artif. Intell. Med.*, vol. 101, p. 101723, Nov. 2019, doi: 10.1016/j.artmed.2019.101723.
[13] M. Ahmed, "Monkey-Pox PATIENTS Dataset," *Kaggle*, 2022.
[14] M. S. Pathan, Z. Jianbiao, D. John, A. Nag, and S. Dev, "Identifying Stroke Indicators Using Rough Sets," *IEEE Access*, vol. 8, pp. 210318–210327, 2020, doi: 10.1109/ACCESS.2020.3039439.
[15] J. Smith, J. Everhart, W. Dickson, W. Knowler, and R. Johannes, "Using the ADAP Learning Algorithm to Forcast the Onset of Diabetes Mellitus," *Proc. - Annu. Symp. Comput. Appl. Med. Care*, vol. 10, Nov. 1988.
[16] M. M. F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, "Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques BT - Computer Vision and Machine Intelligence in Medical Image Analysis," M. Gupta, D. Konar, S. Bhattacharyya, and S. Biswas, Eds., Singapore: Springer Singapore, 2020, pp. 113–125.
[17] S. Borzooei, G. Briganti, M. Golparian, J. R. Lechien, and A. Tarokhian, "Machine learning for risk stratification of thyroid cancer patients: a 15-year cohort study.," *Eur. Arch. oto-rhino-laryngology Off. J. Eur. Fed. Oto-Rhino-Laryngological Soc. Affil. with Ger. Soc. Oto-Rhino-Laryngology - Head Neck Surg.*, vol. 281, no. 4, pp. 2095–2104, Apr. 2024, doi: 10.1007/s00405-023-08299-w.
[18] R. Detrano *et al.*, "International application of a new probability algorithm for the diagnosis of coronary artery disease.," *Am. J. Cardiol.*, vol. 64, no. 5, pp. 304–310, Aug. 1989, doi: 10.1016/0002-9149(89)90524-9.
[19] N. Kolambage, H. Goonasekara, and R. Hewapathirana, "Design, Development and Implementation of a Machine Learning-based Predictive Modelling Tool to Accurately Predict Thalassemia Carrier state using Full Blood Count Indices and Haemoglobin Variants," 2020.
[20] M. Yöntem, K. Adem, T. Ilhan, and S. Kılıçarslan, "DIVORCE PREDICTION USING CORRELATION BASED FEATURE SELECTION AND ARTIFICIAL NEURAL NETWORKS ," Jun. 2019.
[21] M. Patrício *et al.*, "Using Resistin, glucose, age and BMI to predict the presence of breast cancer," *BMC Cancer*, vol. 18, no. 1, p. 29, 2018, doi: 10.1186/s12885-017-3877-1.
[22] D. Gil, J. L. Girela, J. De Juan, M. J. Gomez-Torres, and M. Johnsson, "Predicting seminal quality with artificial intelligence methods," *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12564–12573, 2012, doi: https://doi.org/10.1016/j.eswa.2012.05.028.
[23] M. Çakır, A. Degirmenci, and O. Karal, "Exploring the Behavioural Factors of Cervical Cancer Using ANOVA and Machine Learning Techniques BT - Science, Engineering Management and Information Technology," A. Mirzazadeh, B. Erdebilli, E. Babaee Tirkolaee, G.-W. Weber, and A. K. Kar, Eds., Cham: Springer Nature Switzerland, 2023, pp. 249–260.
[24] T. L. Bailey, "Screen Time Data," *Kaggle*, 2019.
[25] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[26] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm," in *International Conference on Machine Learning*, 1996.

[27] S. Cost and S. Salzberg, "A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features," *Mach. Learn.*, vol. 10, no. 1, pp. 57–78, 1993, doi: 10.1023/A:1022664626993.
[28] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A Hybrid Approach to Alleviating Class Imbalance," *IEEE Trans. Syst. Man, Cybern. - Part A Syst. Humans*, vol. 40, no. 1, pp. 185–197, 2010, doi: 10.1109/TSMCA.2009.2029559.
[29] 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. Part C (Applications Rev.*, vol. 42, no. 4, pp. 463–484, 2012, doi: 10.1109/TSMCC.2011.2161285.
[30] R. Zhu, Y. Guo, and J.-H. Xue, "Adjusting the imbalance ratio by the dimensionality of imbalanced data," *Pattern Recognit. Lett.*, vol. 133, pp. 217–223, 2020, doi: https://doi.org/10.1016/j.patrec.2020.03.004.