IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 2, April 2026, pp. 1402~1408
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i2.pp1402-1408

The effects of data imbalance on fraud detection model accuracy

Rusma Anieza Ruslan[1], Nureize Arbaiy[1], Pei-Chun Lin[2]
[1]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Malaysia
[2]Department of Information Engineering and Computer Science, Feng Chia University, Taichung, Taiwan

Article Info

Article history:

Received Aug 27, 2024
Revised Feb 9, 2026
Accepted Mar 5, 2026

Keywords:

Data augmentation
Imbalanced dataset
Machine learning
Resampling
SMOTE

ABSTRACT

Machine learning (ML) model performance is often assessed by accuracy, but the quality and balance of data also play crucial roles. Imbalanced datasets, where the minority class has fewer samples than the majority class, can lead to biased predictions favoring the majority class. This study addresses the issue of class imbalance through resampling techniques, including random undersampling (RUS) and random oversampling (ROS), specifically applied to a fraud detection dataset. We classify the resampled datasets using random forest (RF) and gradient boosting (GB) models. Our findings indicate that the RF model, when combined with ROS, achieves an accuracy of 97.4%, surpassing the 96.1% accuracy of the GB model with RUS. This approach demonstrates the importance of addressing class imbalance to improve prediction accuracy in ML.

Corresponding Author:

Nureize Arbaiy
Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia
Parit Raja, Batu Pahat, Malaysia
Email: nureize@uthm.edu.my

1. INTRODUCTION

In artificial intelligence (AI), machine learning (ML) focuses on performing tasks without explicit human programming by developing models that identify patterns in data, enabling predictions and decision-making [1]. ML enhances computational capabilities, allowing for faster and more accurate processing of large datasets. Key categories of ML include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [2]. Supervised learning uses labeled datasets for training, enabling models to predict outcomes for unseen data [3]. This category further divides into regression and classification tasks. Conversely, unsupervised learning works with unlabeled datasets, allowing models to find patterns independently, employing techniques like clustering and association [4].

A robust dataset is essential for effective ML, comprising structured and unstructured data that serves as input for training models [5]. Datasets can be classified as balanced or imbalanced based on class representation [6]. Balanced datasets have equal representation across classes, while imbalanced datasets may lead to underperforming models on minority classes. Addressing this imbalance is vital for accurate predictions, with methods such as resampling, where minority classes are oversampled, and majority classes undersampled [7]. This study evaluates class imbalance mitigation techniques in credit card fraud detection. It analyzes performance metrics—accuracy, precision, recall, and F1-score—across various resampling ratios, while addressing the risks of overfitting associated with synthetic oversampling. The goal is to improve ML model reliability in imbalanced scenarios, particularly in fraud detection applications.

The structure of this article is as follows. Section 1 presents background on imbalanced datasets. Section 2 covers foundational concepts in ML. Section 3 reviews studies on addressing class imbalances. Section 4 presents experimental results of the classification models under different resampling techniques. Lastly, section 5 provides a conclusion.

2. RELATED WORK
2.1. Datasets in machine learning
In ML, datasets are structured collections of information used for testing, validating, and training models. They enable ML algorithms to identify patterns, leading to accurate predictions. Typically, datasets are organized in tabular format, where rows represent individual samples and columns signify attributes associated with those samples [8]. The quality of datasets is crucial for performance and generalizability of ML models; well-structured datasets allow for meaningful pattern extraction and reliable predictions [9], [10].

Datasets can be categorized as labeled or unlabeled. Labeled datasets contain instances linked to target values, which are essential in supervised learning for classifying data and generating precise predictions [11]. In contrast, unlabeled datasets lack predefined labels, allowing models to learn patterns independently [12]. Additionally, datasets can be structured, organized in tables (e.g., customer ID, age, and gender), or unstructured, including formats like text, images, and audio.

Datasets are further divided into three main categories: training, validation, and testing sets [13]. The training set contains labeled samples to educate the model, while the validation set helps adjust parameters during training, and the testing set evaluates the model's performance on unseen data. Common metrics for assessment include accuracy and error rate [14]. Preprocessing is vital for preparing datasets before model training. This includes normalization, transformation, and cleaning to address outliers, missing values, or noise, all of which can negatively impact model performance [15]. Ensuring data integrity is essential for accurate and efficient analysis and modeling.

2.2. Imbalanced datasets
Imbalances in datasets refer to variations in the number of input samples across different output classes. An imbalanced dataset exhibits a skewed distribution of classes, posing significant challenges for ML. When the number of instances in the minority class is substantially lower than in the majority class, the model tends to bias predictions toward the majority class. This can lead to inaccurate predictions and poor performance for the minority class [16], which is critical in real-world applications like fraud detection and medical diagnostics. For example, fraudulent transactions often constitute a small fraction of total transactions, making accurate identification essential to mitigate financial losses. Similarly, diagnosing rare medical conditions requires precision to ensure effective patient care, emphasizing the need to address dataset imbalances to avoid adverse outcomes.

Imbalanced datasets are characterized by unequal class representation, resulting in skewed predictions by algorithms. To address these disparities, various techniques have been developed, including resampling methods, advanced ensemble approaches and cost-sensitive learning strategies. These techniques are instrumental in improving classification performance in scenarios where class distribution is skewed, ensuring that models are trained effectively even with limited representation of minority classes.

2.2.1. Resampling method
Resampling is a technique used to adjust the number of instances in majority and minority classes to address data imbalances in ML [17]. There are two primary resampling methods: oversampling and undersampling. Oversampling increases the representation of the minority class by duplicating instances. While this method is useful, it may lead to overfitting as it does not introduce new information. Alternatives to simple oversampling include data augmentation, adaptive synthetic sampling (ADASYN), and synthetic minority oversampling technique (SMOTE). SMOTE generates new synthetic samples for the minority class, ensuring a distribution across the feature space; however, it can create dense clusters of synthetic samples [18]. Conversely, ADASYN specifically generates synthetic samples for hard-to-classify minority instances, which helps enhance decision-making by situating new samples closer to the decision boundary [19].

Undersampling reduces the number of instances in the majority class, which can pose a risk of losing valuable data. Techniques like near-miss undersampling select majority class instances based on their proximity to minority class examples, using Euclidean distance to improve the balance. Resampling techniques have been effectively applied across various domains, particularly in classification tasks, demonstrating their capability to enhance classification outcomes. Table 1 summarizes notable applications of resampling techniques, highlighting their contributions and key findings.

2.2.2. Ensemble methods
Ensemble methods provide effective strategy for addressing class imbalance by combining multiple baseline models to create more robust overall model [20]. This approach can enhance classification accuracy, especially in datasets with imbalanced classes. Key ensemble techniques include boosting and bagging. Boosting involves training weak models sequentially, with each successive model focusing on the errors made by its predecessor. An example of this technique is AdaBoost, which adjusts weights of misclassified instances to improve future predictions [21]. Bagging, generates multiple subsets for the training data by sampling with

replacement, which helps reduce prediction variance. This technique aggregates predictions through majority voting to produce a final output [22]. An example of ensemble methods in action is presented by utilizing an ensemble convolutional neural network (EnCNN) approach for intelligent fault diagnosis in machines under imbalanced conditions. This method combines several classifiers trained on balanced subsets of the imbalanced dataset, employing weighted voting for final predictions to improve accuracy and robustness [23].

2.2.3. Cost-sensitive learning

Cost-sensitive learning techniques adjust the costs associated with misclassifying instances from majority and minority classes to address class imbalances effectively [24]. In this framework, the majority class usually incurs lower costs, while the minority class is associated with higher costs. This approach can be categorized into four types: i) false positive (FP): the cost of incorrectly classifying a positive instance as negative, ii) false negative (FN): the cost of incorrectly classifying a negative instance as positive, iii) true positive (TP): the accurate classification of a positive instance, and iv) true negative (TN): the accurate classification of a negative instance.

In fraud detection, for instance, mislabeling legitimate transactions as fraudulent (FP) can lead to financial losses, while misclassifying fraudulent transactions as legitimate (FN) can result in customer dissatisfaction. The typically small proportion of fraudulent transactions contributes to highly skewed dataset. An illustrative example of cost-sensitive learning is the development of cost-sensitive feature selection general vector machine (CFGVM) algorithm. This algorithm integrates general vector machine with binary ant lion optimizer to enhance the performance of imbalanced classification tasks. This approach underscores the importance of adjusting costs to improve predictive accuracy in scenarios with imbalanced data [25].

Table 1. Application of resampling techniques

Reference	Contribution	Findings
[26]	Addresses class imbalance in credit card fraud detection, impacting the effectiveness of classification techniques	Classification models demonstrated statistically significant improvements over the initial imbalanced dataset
[27]	Introduces a unique approach combining undersampling and oversampling methods	Significant improvement in classifier performance, particularly in detecting rare events
[28]	Develops models for automatic determination of effective resampling techniques based on dataset properties	The efficiency of oversampling and undersampling methods varies based on the imbalance ratio and the dataset characteristics
[29]	Studies the impact of resampling on ANN classifier performance in network intrusion detection	Enhanced identification of minority-class data through resampling
[30]	Investigates distinctions in classification efficacy using various resampling strategies	RF classifier outperformed others when enhanced with the SVM-SMOTE method

2.3. Machine learning models

Classifiers are models that assign specific classifications to input data, with applications in areas like fraud detection. Standard classifiers often struggle with imbalanced datasets, prompting researchers to develop techniques to enhance model robustness and accuracy. This research focuses on two effective classification techniques: random forest (RF) and gradient boosting (GB). Both are adept at handling complex datasets and imbalanced class distributions. RF employs an ensemble approach to improve accuracy and reduce overfitting, while GB enhances predictive performance through iterative learning. By focusing on these classifiers, we aim to leverage their strengths to tackle the challenges posed by skewed target class distribution.

3. METHOD

3.1. Data gathering and data splitting

This study analyzes a dataset of credit card transactions from European cardholders in September 2013, comprising 284,807 transactions, of which 492 were fraudulent. Due to the significant imbalance between non-fraudulent and fraudulent transactions, the study focuses on resampling the classes at ratios of 20:80, 30:70, and 40:60. These ratios enable a comprehensive evaluation of how different levels of resampling affect model performance.

Data preprocessing included several critical steps to ensure input quality. Features were standardized to have a mean of zero and a standard deviation of one, enhancing the performance of algorithms sensitive to feature scaling. Temporal patterns were captured by engineering time-related features, and anonymized V-features, created through principal component analysis (PCA) for confidentiality, were also included. The analysis was conducted using Python, utilizing the scikit-learn library (version 0.24.2) for implementing ML models and resampling techniques.

3.2. Resampling methodology and performance assessment

This study employed three prominent resampling techniques: random undersampling (RUS), random oversampling (ROS), and SMOTE. Performance evaluations were conducted for each technique to assess the effectiveness of the classification methods used, focusing on key classification errors: FP and FN, as well as TP and TN. Classification model accuracy was compared using (1).

$$Accuracy = \frac{TP+TN}{\Sigma(TP+FP+TN+FN)} \quad (1)$$

To provide a comprehensive assessment of model performance, we also calculated the (2) to (4) based on the confusion matrix.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

While metrics like receiver operating characteristic (ROC) - area under the curve (AUC) and Matthews correlation coefficient (MCC) are valuable, this study focused on accuracy, precision, recall, and F1-score to provide a clear assessment of model performance in fraud detection. Accuracy offers a straightforward performance measure, while precision and recall help understand the balance between correctly identified fraudulent cases and misclassifications. The F1-score, as the harmonic mean of precision and recall, effectively captures the trade-off between these two metrics, which is especially relevant in context of imbalanced datasets.

4. RESULTS AND DISCUSSION

This study analyzes the performance of various resampling techniques—RUS, ROS, and SMOTE—across three ratios: 20:80, 30:70, and 40:60. In 20:80 ratio as shown in Table 2, RF achieved 97.36% accuracy but had low precision (0.0579) and F1-score (0.1091), while GB had a similar accuracy of 96.15% with a precision of 0.0403. The ROS approach showed strong performance, with RF achieving 99.96% accuracy and 97.44% precision, though its recall dropped to 77.55%, resulting in an F1-score of 0.8636; GB showed 99.34% accuracy but lower precision (19.61%). With SMOTE, RF maintained high accuracy (99.95%) and good precision (86.60%) alongside a recall of 85.71%, leading to an F1-score of 0.8615, while GB's performance was adequate, achieving 21.31% precision. In 30:70 ratio as shown in Table 3, RF again demonstrated 97.41% accuracy yet low precision (0.0545) and an F1-score of 0.1029, while GB's precision fell to 0.0347. ROS resulted in RF achieving 99.96% accuracy with a precision of 94.83% and an F1-score of 0.8730, while GB performed lower with a precision of 15.14%. SMOTE yielded 99.95% accuracy and 83.10% precision for RF, leading to an F1-score of 0.8489; GB recorded 99.35% accuracy with lower metrics. In 40:60 ratio as shown in Table 4, RF reached 98.05% accuracy but reported low precision (0.0727) and an F1-score of 0.1346, and GB showed similar low precision at 0.0420. However, with ROS, RF maintained high accuracy (99.95%) and precision (93.75%), achieving an F1-score of 0.8547, while GB's performance lagged. When using SMOTE, RF achieved 99.95% accuracy with 84.46% precision, culminating in an F1-score of 0.8490.

Table 2. Comparison performance results of 20:80 ratio

Sampling technique	Model	Accuracy	Precision	Recall	F1-score
Random undersampling	Random forest	0.9736	0.0579	0.9388	0.1091
	Gradient boosting	0.9615	0.0403	0.9388	0.0774
Random oversampling	Random forest	0.9996	0.9744	0.7755	0.8636
	Gradient boosting	0.9934	0.1961	0.9184	0.3232
SMOTE	Random forest	0.9995	0.8660	0.8571	0.8615
	Gradient boosting	0.9941	0.2131	0.8980	0.3444

Table 3. Comparison performance results of 30:70 ratio

Sampling technique	Model	Accuracy	Precision	Recall	F1-score
Random undersampling	Random forest	0.9741	0.0545	0.9338	0.1029
	Gradient boosting	0.9586	0.0347	0.9338	0.0670
Random oversampling	Random forest	0.9996	0.9483	0.8088	0.8730
	Gradient boosting	0.9916	0.1514	0.9265	0.2603
SMOTE	Random forest	0.9995	0.8310	0.8676	0.8489
	Gradient boosting	0.9935	0.1856	0.9118	0.3085

Table 4. Comparison performance results of 40:60 ratio

Sampling technique	Model	Accuracy	Precision	Recall	F1-score
Random undersampling	Random forest	0.9805	0.0727	0.9058	0.1346
	Gradient boosting	0.9648	0.0420	0.9162	0.0802
Random oversampling	Random forest	0.9995	0.9375	0.7853	0.8547
	Gradient boosting	0.9939	0.2012	0.8901	0.3282
SMOTE	Random forest	0.9995	0.8446	0.8534	0.8490
	Gradient boosting	0.9943	0.2127	0.8796	0.3425

RF consistently outperformed GB across all metrics and ratios, particularly with ROS and SMOTE, which significantly enhanced RF's ability to accurately classify fraudulent transactions. While SMOTE offers advantages, it also carries risks such as overfitting, especially with limited diversity in the minority class [31]. Therefore, mitigation strategies, including applying SMOTE only to the training set or using modified versions like Borderline-SMOTE, can help improve model robustness. While RF with ROS and SMOTE demonstrated strong performance, exploring alternative methods and tuning for GB may also yield favorable results, emphasizing the need to tailor the choice of resampling technique and classifier to the specific dataset and context for optimal outcomes.

While this study demonstrates the effectiveness of RF with ROS and SMOTE, other research has explored different approaches. For instance, some studies have found GB to be more effective for handling imbalanced data, particularly when carefully tuned. This is because GB focuses more on difficult examples, potentially outperforming RF in certain scenarios. Additionally, research indicates that combining SMOTE with RF can achieve high accuracy and F1-scores in credit card fraud detection [32], [33]. Other studies suggest Borderline-SMOTE can further enhance accuracy compared to other oversampling methods [34], [35]. These varying results highlight the importance of considering the specific dataset and problem context when selecting the most appropriate resampling technique and classification algorithm.

5. CONCLUSION

This study highlights the critical issue of class imbalance in ML and its impact on model performance. The RF classifier consistently achieved the highest accuracy across three resampling techniques (RUS, ROS and SMOTE). Notably, while SMOTE demonstrated strong efficacy, ROS yielded even more compelling results. These findings emphasize the necessity of carefully selecting resampling techniques and algorithms to optimize performance, particularly in applications like fraud detection and medical diagnosis. Despite the promising results, there are limitations to this study. The performance of the classifiers may vary with different datasets or in real-world scenarios outside of the scope of this research. Additionally, the study focused solely on RF and GB models, leaving other potentially effective algorithms unexplored. Future work should explore a wider range of classification algorithms and consider additional evaluation metrics, such as ROC-AUC and MCC, to provide a more comprehensive assessment of model performance. Furthermore, expanding the dataset to include more diverse cases can help validate the robustness of the proposed methodologies. Continued research into effective methodologies for managing class imbalance is vital for enhancing accuracy and reliability in real-world scenarios.

FUNDING INFORMATION

This research was supported by Universiti Tun Hussein Onn Malaysia through GPPS (Vot Q662).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	**C**	**M**	**So**	**Va**	**Fo**	**I**	**R**	**D**	**O**	**E**	**Vi**	**Su**	**P**	**Fu**
Rusma Anieza Ruslan	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	
Nureize Arbaiy	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	
Pei-Chun Lin	✓		✓	✓						✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/code/janiobachmann/credit-fraud-dealing-with-imbalanced-datasets.

REFERENCES

[1] B. Mahesh, "Machine learning algorithms - a review," *International Journal of Science and Research*, vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/ART20203995.
[2] T. T. Khoei and N. Kaabouch, "Machine learning: models, challenges, and research directions," *Future Internet*, vol. 15, no. 10, 2023, doi: 10.3390/fi15100332.
[3] S. Roy, T. Meena, and S.-J. Lim, "Demystifying supervised learning in healthcare 4.0: a new reality of transforming diagnostic medicine," *Diagnostics*, vol. 12, no. 10, 2022, doi: 10.3390/diagnostics12102549.
[4] M. T. Almuqati, F. Sidi, S. N. M. Rum, M. Zolkepli, and I. Ishak, "Challenges in supervised and unsupervised learning: a comprehensive overview," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 14, no. 4, pp. 1449–1455, 2024, doi: 10.18517/ijaseit.14.4.20191.
[5] L. Zhao, "Event prediction in the big data era: a systematic survey," *ACM Computing Surveys*, vol. 54, no. 5, pp. 1–37, 2022, doi: 10.1145/3450287.
[6] T. P. Pagano *et al.*, "Bias and unfairness in machine learning models: a systematic review on datasets, tools, fairness metrics, and identification and mitigation methods," *Big Data and Cognitive Computing*, vol. 7, no. 1, 2023, doi: 10.3390/bdcc7010015.
[7] F. Gurcan and A. Soylu, "Learning from imbalanced data: integration of advanced resampling techniques and machine learning models for enhanced cancer diagnosis and prognosis," *Cancers*, vol. 16, no. 19, 2024, doi: 10.3390/cancers16193417.
[8] N. Hollmann *et al.*, "Accurate predictions on small data with a tabular foundation model," *Nature*, vol. 637, no. 8045, pp. 319–326, 2025, doi: 10.1038/s41586-024-08328-6.
[9] A. O. Kehinde, "Leveraging machine learning for predictive models in healthcare to enhance patient outcome management," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 1, pp. 1465–1482, 2025, doi: 10.56726/IRJMETS66198.
[10] S. Friedrich and T. Friede, "On the role of benchmarking data sets and simulations in method comparison studies," *Biometrical Journal*, vol. 66, no. 1, 2024, doi: 10.1002/bimj.202200212.
[11] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison," *Computers in Biology and Medicine*, vol. 136, 2021, doi: 10.1016/j.compbiomed.2021.104672.
[12] E. Oliveira, M. Rodrigues, J. P. Pereira, A. M. Lopes, I. I. Mestric, and S. Bjelogrlic, "Unlabeled learning algorithms and operations: overview and future trends in defense sector," *Artificial Intelligence Review*, vol. 57, no. 3, 2024, doi: 10.1007/s10462-023-10692-0.
[13] K. Alomar, H. I. Aysel, and X. Cai, "Data augmentation in classification and segmentation: a survey and new strategies," *Journal of Imaging*, vol. 9, no. 2, 2023, doi: 10.3390/jimaging9020046.
[14] M. O. -Adjei, J. B. H. -Acquah, T. Frimpong, and G. A. -Salaam, "Imbalanced class distribution and performance evaluation metrics: a systematic review of prediction accuracy for determining model performance in healthcare systems," *PLOS Digital Health*, vol. 2, no. 11, 2023, doi: 10.1371/journal.pdig.0000290.
[15] K. Maharana, S. Mondal, and B. Nemade, "A review: data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, 2022, doi: 10.1016/j.gltp.2022.04.020.
[16] M. S. Milosevic and V. M. Ciric, "Extreme minority class detection in imbalanced data for network intrusion," *Computers & Security*, vol. 123, 2022, doi: 10.1016/j.cose.2022.102940.
[17] M. Carvalho, A. J. Pinho, and S. Brás, "Resampling approaches to handle class imbalance: a review from a data perspective," *Journal of Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01119-4.
[18] H. Hairani, T. Widiyaningtyas, and D. D. Prasetya, "Addressing class imbalance of health data: a systematic literature review on modified synthetic minority oversampling technique (SMOTE) strategies," *JOIV : International Journal on Informatics Visualization*, vol. 8, no. 3, pp. 1310–1318, 2024, doi: 10.62527/joiv.8.3.2283.
[19] M. Zakariah, S. A. AlQahtani, and M. S. Al-Rakhami, "Machine learning-based adaptive synthetic sampling technique for intrusion detection," *Applied Sciences*, vol. 13, no. 11, 2023, doi: 10.3390/app13116504.
[20] T.-T.-H. Le, Y. Shin, M. Kim, and H. Kim, "Towards unbalanced multiclass intrusion detection with hybrid sampling methods and ensemble classification," *Applied Soft Computing*, vol. 157, 2024, doi: 10.1016/j.asoc.2024.111517.
[21] Y. Ding, H. Zhu, R. Chen, and R. Li, "An efficient AdaBoost algorithm with the multiple thresholds classification," *Applied Sciences*, vol. 12, no. 12, 2022, doi: 10.3390/app12125872.
[22] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble learning for disease prediction: a review," *Healthcare*, vol. 11, no. 12, 2023, doi: 10.3390/healthcare11121808.
[23] F. Jia, S. Li, H. Zuo, and J. Shen, "Deep neural network ensemble for the intelligent fault diagnosis of machines under imbalanced data," *IEEE Access*, vol. 8, pp. 120974–120982, 2020, doi: 10.1109/ACCESS.2020.3006895.
[24] O. Volk and G. Singer, "An adaptive cost-sensitive learning approach in neural networks to minimize local training–test class distributions mismatch," *Intelligent Systems with Applications*, vol. 21, 2024, doi: 10.1016/j.iswa.2023.200316.
[25] F. Feng, K.-C. Li, J. Shen, Q. Zhou, and X. Yang, "Using cost-sensitive learning and feature selection algorithms to improve the performance of imbalanced classification," *IEEE Access*, vol. 8, pp. 69979–69996, 2020, doi: 10.1109/ACCESS.2020.2987364.
[26] N. F. Hordri, S. Sophiayati, N. Firdaus, and S. Mariyam, "Handling class imbalance in credit card fraud using resampling methods," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 11, 2018, doi: 10.14569/IJACSA.2018.091155.
[27] S. Cateni, V. Colla, and M. Vannucci, "A method for resampling imbalanced datasets in binary classification tasks for real-world problems," *Neurocomputing*, vol. 135, pp. 32–41, 2014, doi: 10.1016/j.neucom.2013.05.059.

[28] M. S. Kraiem, F. S. -Hernández, and M. N. M. -García, “Selecting the suitable resampling strategy for imbalanced data classification regarding dataset properties: an approach based on association models,” *Applied Sciences*, vol. 11, no. 18, 2021, doi: 10.3390/app11188546.
[29] S. Bagui and K. Li, “Resampling imbalanced data for network intrusion detection datasets,” *Journal of Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-020-00390-x.
[30] R. Ghorbani and R. Ghousi, “Comparing different resampling methods in predicting students' performance using machine learning techniques,” *IEEE Access*, vol. 8, pp. 67899–67911, 2020, doi: 10.1109/ACCESS.2020.2986809.
[31] M. Sopiyan, F. Fauziah, and Y. F. Wijaya, “Fraud detection using random forest classifier, logistic regression, and gradient boosting classifier algorithms on credit cards,” *JUITA: Jurnal Informatika*, vol. 10, no. 1, pp. 77–88, 2022, doi: 10.30595/juita.v10i1.12050.
[32] N. S. S. Pranavi, T. K. S. S. Sruthi, B. J. N. Sirisha, M. S. Nayak, and V. S. G. Thadikemalla, “Credit card fraud detection using minority oversampling and random forest technique,” in *2022 3rd International Conference for Emerging Technology*, 2022, pp. 1–6, doi: 10.1109/INCET54531.2022.9824146.
[33] A. M. Aburbeian and H. I. Ashqar, “Credit card fraud detection using enhanced random forest classifier for imbalanced data,” in *Proceedings of the 2023 International Conference on Advances in Computing Research*, 2023, pp. 605–616, doi: 10.1007/978-3-031-33743-7_48.
[34] F. Yagin, I. Cicek, and Z. Kucukakcali, “Classification of stroke with gradient boosting tree using smote-based oversampling method,” *Medicine Science: International Medical Journal*, vol. 10, no. 4, pp. 1510–1515, 2021, doi: 10.5455/medscience.2021.09.322.
[35] Y. Pristyanto, Z. Mukarabiman, and A. F. Nugraha, “Extreme gradient boosting algorithm to improve machine learning model performance on multiclass imbalanced dataset,” *JOIV : International Journal on Informatics Visualization*, vol. 7, no. 3, pp. 710–715, 2023, doi: 10.30630/joiv.7.3.1102.

BIOGRAPHIES OF AUTHORS

Rusma Anieza Ruslan holds a B.S. degree in Information Technology from Universiti Tun Hussein Onn Malaysia, Johor, in 2024. She is pursuing an M.S. in Information Technology and working on a project titled "Addressing imbalance datasets in machine learning". She can be contacted at email: hi230027@student.uthm.edu.my.

Nureize Arbaiy is currently in the Department of Software Engineering at the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM). Her research interests are multi-criteria decision-making, fuzzy regression, fuzzy auto-regression, information systems, and deep learning. She has supervised several Ph.D. and master's students. She published articles in various international journals and conference proceedings. She has acted as a reviewer for multiple journals and conferences. She can be contacted at email: nureize@uthm.edu.my.

Pei-Chun Lin received a B.Sc. degree in the Department of Mathematics from National Kaohsiung Normal University, Kaohsiung, Taiwan and received an M.Sc. degree in the Department of Mathematical Sciences from National Chengchi University, Taipei, Taiwan. Moreover, she received a Ph.D. degree from graduate school of Information, Production and Systems, Waseda University, Fukuoka, Japan. She is currently an assistant professor in the Department of Information Engineering and Computer Science, Feng Chia University. Her research interests include soft computing, artificial intelligent, robot computing, statistical modeling, and big data analysis. She can be contacted at email: peiclin@fcu.edu.tw.