Research Inventy: International Journal of Engineering And Science
Vol.14, Issue 1 (January 2024), PP 102-108
Issn (e): 2278-4721, Issn (p):2319-6483, www.researchinventy.com

Integration of Random Undersampling Ensemble and SP-SMOTE Based on Density Difference in Handling Class Imbalance

Yeni Risyani, Susi Japit
Department of Computer Science, Universitas IBBI, Medan, Indonesia

Abstract: *Class imbalance is one of the main problems in classification in machine learning. This problem is characterized by the existence of one class with a larger number of instances compared to the other classes. Approaches to dealing with this class imbalance include using sampling methods and cost sensitive learning methods. The sampling approach is specifically carried out in the majority class using the oversampling method in the minority class and in the majority class using the undersampling method. The cost sensitive approach is usually carried out by providing a larger training error penalty for the minority class. The main problem with the sampling method and cost sensitive learning method is the density problem where errors often occur in determining the minority class and majority class. This research will combine the application of sampling methods and cost sensitive learning methods and at the same time overcome density differences. Space Partitioning SMOTE (SP-SMOTE) is a fairly good oversampling method because it is based on the difference in density between the majority class and minority class so that the oversampling process can be carried out well only on the minority class. The cost sensitive learning method that will be used in this research is using Random Undersampling Ensemble (RUE) which will determine the location of the undersampling majority class based on the density produced through error feedback. The proposed method achieved average accuracy, precision, and recall values of 0.872, 0.86, and 0.864, respectively, compared to the accuracy, precision, and recall values of 0.838. 0.828. and 0.832 obtained bv the SMOTE method.*

Keywords: *Class Imbalance, Density Difference, Randoum Undersampling Ensemble, SP-SMOTE*

Date of Submission: 08-01-2024 Date of acceptance: 20-01-2024

I. INTRODUCTION

Class imbalance refers to an uneven distribution of instances among different classes, where one class has much more instances than the others(Sun et al., 2015). Class imbalance issues frequently arise in machine learning classification challenges. This issue is garnering increased attention due to class imbalance, which makes it increasingly challenging to obtain accurate classification results for the minority class (positive class). There is a tendency to achieve high accuracy in classification by simply grouping all instances into the majority class (negative class)(Thölke et al., 2023). Class imbalance is a challenging issue to circumvent as real-world data frequently exhibits an imbalanced distribution, with positive classes typically having a lesser number of instances. However, these positive classes are of particular importance for observation(Vuttipittayamongkol et al., 2021). Class imbalance commonly presents challenges in various domains, such as medical image categorization(Park et al., 2024), network intrusion detection(Ren et al., 2023), clinical data(Kuo et al., 2023), computer vision tasks(Dai et al., 2022), fraud detection(Z. Li et al., 2021), and other topics pertaining to the categorization using machine learning.

The sampling method and the cost sensitive learning method are two regularly employed approaches for addressing class imbalance(Mienye & Sun, 2021). Sampling method dilakukan dengan menggunakan oversampling dan undersampling. Oversampling is performed on the minority class to augment the number of instances by generating synthetic cases, whereas undersampling is performed on the majority class(Grina et al., 2023). Nevertheless, both oversampling and undersampling encounter challenges related to the density disparities among instances in the majority and minority classes. This is caused by the presence of overlapping instances, which can lead to an overemphasis on the learning model(Yan et al., 2022). Cost-sensitive learning addresses class imbalance by assigning a greater penalty for errors made on the minority class(Kamalov, 2020). Cost sensitive learning is not effective in circumstances when there are density discrepancies between the minority and majority classes, resulting in overlapping conditions for the instances(Steininger et al., 2021). Both the sample approach and the cost sensitive learning method encounter the issue of density disparity(Shi et al., 2023).

Through various initiatives, including the following, some researchers have focused on density in cost-sensitive learning approaches as well as sampling techniques: Remove the majority of data in overlapping regions and transform the high-density data of the majority class into data with appropriate density (Mayabadi & Saadatfar, 2022), Utilize clustering to group the minority instances, as it has the ability to effectively detect sub-clusters with varying sizes and densities (Tao et al., 2020), and The local density of samples in both the majority and minority classes is calculated to assess their quality within each cluster. The roulette wheel selection operator is then utilized to choose the most appropriate samples based on their probabilities (Mirzaei et al., 2021). Numerous researchers have endeavored to ascertain the optimal technique for calculating the density of both the majority and minority groups. Determining density differences relies on the quality of clusters; sampling methods and cost-sensitive learning methods should both account for high-density samples.

The Random Undersampling Ensemble (RUE), which proposed by (Zhou et al., 2023), operates under the premise that an instance exhibiting a high error rate is typically indicative of noise, while an instance with a medium error rate signifies proximity to the classification boundary. Conversely, an instance with a low error rate signifies ease of learning and is situated far from the classification boundary. Using the error rate feedback in RUE, it is possible to provide reasonably precise estimations of instance location and significance, according to this hypothesis. This approach is renowned for effectively segregating a sample from noise and overlapping, and its error penalty calculation for the minority class is determined through error feedback.

On the other hand, SP-SMOTE, proposed by(Y. Li et al., 2021), aims to create distinct partitions in areas with a high concentration of data density. It achieves this by constructing a binary tree for oversampling. Subsequently, the minority class within the dataset may be readily identified. This method is deemed sufficiently effective for conducting the oversampling procedure for minority groups due to its capacity to generate precise density data.

This study aims to integrate sampling techniques and cost-sensitive learning approaches to effectively address variations in data density. Space Partitioning SMOTE (SP-SMOTE) is an effective oversampling technique that focuses on the disparity in density between the majority and minority classes. This allows for a targeted oversampling approach that specifically addresses the minority class.

II. EXPERIMENTAL PROCEDURE

2.1. Datasets

The research was carried out utilizing a dataset obtained from the KEEL Repository (Alcalá-Fdez et al., 2009). The data set that was used in this study is shown in Table 1.

Table 1. Detailed information of the datasets utilized in this study

Dataset	Number of Attributes	Number of Instances	Minority Class	Majority Class	IR
Haberman	3	306	Positive	Negative	2.78
Ecoli1	7	336	Class 1	Reminder Class	3.36
Yeast-2_vs_4	8	514	Class 2	Class 4	9.06
Shuttle-c0-vs-c4	9	1829	Class 0	Class 4	13.87
Yeast4	8	1484	Class4	Reminder Class	28.10

The dataset utilized in this study comprises a multitude of attributes, instances, and an Imbalance Ratio (IR), as shown in Table 1.

2.2. Methodology

This research was carried out by comparing the results obtained by Integration of Random Undersampling Ensemble and SP-SMOTE Based on Density Difference when compared with the results obtained using SMOTE. The performance measurement is carried out using accuracy, precision and recall. Figure 1 illustrates the phases of this study.

Figure 1. Stages of research

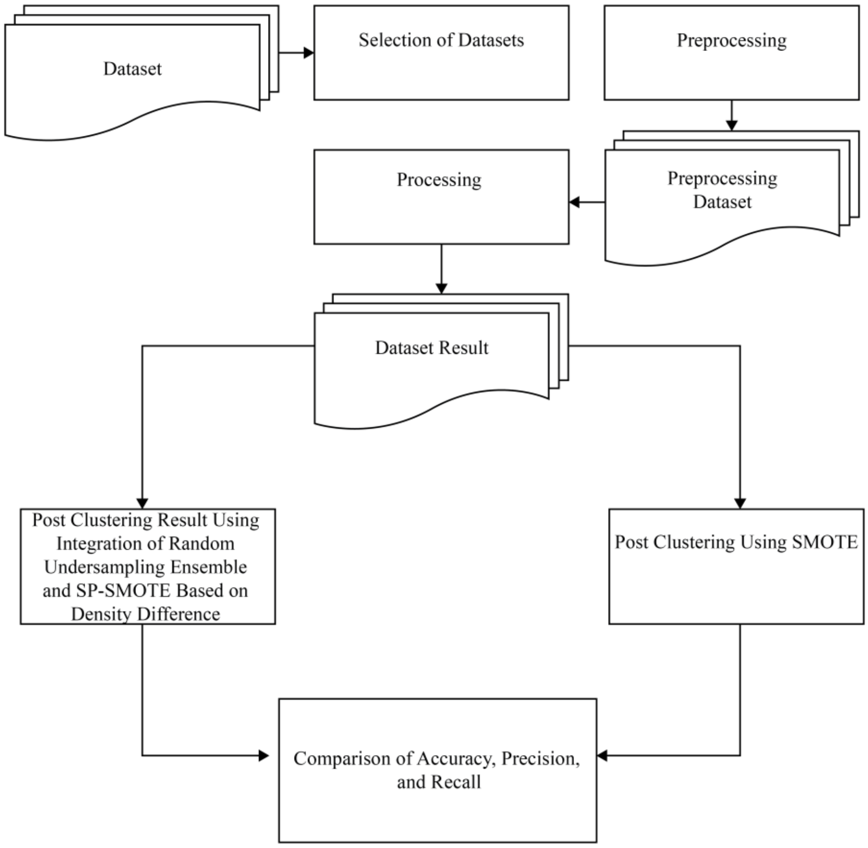

Figure 1 illustrates that the investigation commenced with the identification of the dataset. The dataset utilized in this study is sourced from the KEEL Repository. After obtaining a dataset, the subsequent procedure involves preprocessing the dataset. After preprocessing, the dataset will be subjected to a comparative analysis between the proposed method and the commonly used Synthetic Minority Oversampling Technique (SMOTE) for addressing class imbalances. Performance testing relies on the evaluation of accuracy, precision, and recall metrics.

2.3. SMOTE

The SMOTE method is used as a benchmark to test the performance of the proposed method. The reason for using SMOTE is that it is a widely used method for handling class imbalance. This method operates by oversampling the minority class. The pseudocode for SMOTE is as follows.

```
Input: X_minor, N_percent, K
Function SMOTE (X_minor, N_percent, K)
X_SMOTE ← {   }
for i ← 1 to len(X_minor)do
   nn ← K Nearest Neighbors (X_i, N_Percent, K)
   p ← [Npercent/100
   while p! = do
      X_neighbour ← select random (nn)
      X_SMOTE ← X_i + rand(0,1)*|X_neighbour - X_i|
      p ← p-1
   end while
end for
return ← X_SMOTE
```

The above pseudocode demonstrates that the process of generating synthetic samples in SMOTE is based on determining the number of synthetic samples to be generated. The synthetic samples are generated based on their proximity to the minority class through oversampling.

2.4. SP-SMOTE

The Space Partitioning SMOTE (SP-SMOTE) proposed by (Y. Li et al., 2021) is based on A unique algorithm called Dannoy has been devised for density-adaptive space partitioning. Dannoy is constructed using the Annoy architecture. It generates substantial divisions in highly populated regions and builds a binary tree to augment the quantity of samples. Consequently, it is easy to identify the minority class of the dataset. The pseudocode for SP SMOTE is as follows:

$Input$: $Dataset\ D$; $Number\ of\ Nearest\ Neighbor\ K$; $Max\ Split\ Number\ Max_{Splits}$; $The\ Size\ of\ Leaf\ Leaf\_Size$
$Output$: $New\ Dataset\ ND$+D

$Tree = Dannoy\ (D, K, Max_{Splits}, Leaf\_Size)$
$For\ each\ node\ n_i\ in\ tree\ do$
  $N = the\ number\ of\ n_i^{'}s\ Child\ nodes$
  $if\ N \geq 2\ then$
    $Children = n_i^{'}s\ Child\ nodes$
    $for\ each\ \overrightarrow{p1}\ in\ children\ and\ \overrightarrow{p1}\ in\ minority\ class\ do$ D
      $\overrightarrow{p2} = the\ nearest\ neighbor\ of\ \overrightarrow{p1}$
      $\vec{m} = the\ midpoint\ of\ \overrightarrow{p1}\ and\ \overrightarrow{p2}$
      $d = the\ distance\ of\ \overrightarrow{p1}\ and\ \overrightarrow{p2}$
      $While\ N! = 0\ do$
        $\vec{x} = Choose\ value\ from\ range\ [\vec{m}\text{-}d, \vec{m} + d\ ]\ random$
        $ND \leftarrow \bar{X}$
        $N = N\text{-}1$
      $End\ While$
    $End\ for$
  $End\ if$
$End\ For$
$dif = the\ quantity\ difference\ between\ minority\ samples\ and\ majority\ samples$
$ND = filter(ND, dif)$
$return\ NDF$

By utilizing the pseudocode provided, it is possible to discern that the Dannoy Tree process comprises a single node derived from Dataset D. In order to identify a specific node within a tree or minority class, it is possible to determine its nearest neighbor p1 and displace it as follows: obtain a perhitungan jarak from the jarak node p2 to node p1, and establish a midpoint between node p2 and node p1. Initiate the procedure to obtain values by extracting them from parameters m-2 and placing them in the new dataset ND.

2.5. Random Undersampling Ensemble (RUE)

The Random Undersampling Ensemble proposed by (Zhou et al., 2023) is based on the following principles: an instance with a high error rate is typically considered noise, an instance with a medium error rate suggests it is close to the classification boundary, and an instance with a low error rate signifies it is an easily learnable instance that is far from the classification boundary. According to this concept, the error rate feedback in RUE can be utilized to obtain reasonably precise estimations for both the location and significance of instances. The pseudocode for RUE is as follows:

1: $\boldsymbol{Input}$: $A\ binary\text{-}class\ imbalanced\ dataset\ \psi, the\ threshold\ \lambda, a$
$learning\ algorithm\ j, and\ the\ number\ of\ undersampled\ sets\ k$
2: $\boldsymbol{Output}$: $A\ filtered\ dataset\ \psi', and\ a\ corresponding\ cost\ sequence\ \{c_1, c_2, \ldots, c_N\}$
3: $\boldsymbol{Procedure}$:
4: $Randomly\ undersampling\ \psi\ for\ k\ times, and\ acquiring\ k\ corresponding\ training\ subsets$;
5: $on\ each\ subset, training\ a\ learner\ by\ learning\ algorithm\ J$
6: $Calculating\ error\ rate\ for\ each\ instance\ in\ \psi\ by\ combining\ k\ learners$;
7: $Removing\ noisy\ instances\ in\ \psi\ by\ comparing\ their\ error\ rates\ with\ threshold\ \lambda$,
$further\ acquiring\ the\ filtered\ dataset\ \psi'$;
8: $For\ each\ instance\ in\ \psi^{'}, regulating\ its\ error\ rate\ by\ equation\ 1$;

$$\xi_1^{'} = \xi_1 + \Delta \tag{1}$$

$Where\ \Delta\ is\ a\ small\ positive\ value\ that\ is\ designated\ as\ 10^{-2}\ by\ default$
9: $For\ each\ instance\ in\ \psi^{'}, calculating\ its\ cost\ by\ equation\ 2$;

$$C_i = \frac{\xi_i^{'}}{\Sigma_{i=1}^{s} \xi_i^{'}} \tag{2}$$

$where\ C_i\ denotes\ the\ cost\ assigned\ for\ the\ instance\ x_i$
10: $Outputting\ the\ filtered\ dataset\ \psi^{'} and\ the\ cost\ sequence\ \{c_1, c_2, \ldots, c_N\}$

2.6. Proposed Method

The proposed method is based on the application of oversampling on the minority class using SP-SMOTE and undersampling on the majority class using RUE. The pseudocode for the proposed method is as follows.

$Input$: $Dataset\ D$; X_{minor}; X_{major}
$Output$: $New\ Dataset\ ND$
1: $for\ each\ instances\ in\ X_{minor}\ do$
2: $Apply\ SP\text{-}SMOTE$
3: $ND \leftarrow ND + X'$
4: $end\ for$
5: $for\ each\ instances\ in\ X_{major}\ do$
6: $Apply\ RUE$
7: $ND \leftarrow ND + X'$
8: $end\ for$

2.7. Confusion Matrix

The measurement of accuracy, precision, and recall is based on the confusion matrix(Ruuska et al., 2018). The confusion matrix can be shown in Table 2.

Table 2. Confusion Matrix

	Predictive Positive Class	**Predictive Negative Class**
Actual Positive Class	True Positive (TP)	False Negative (FN)
Actual Negative Class	False Positive (FP)	True Negative (TN)

2.8. Accuracy, Precision, and Recall

The equations for calculating accuracy, precision, and recall can be found in Equation 3-5(Antariksa et al., 2022).

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

III. RESULTS AND DISCUSSIONS

3.1. Results

The testing was conducted on the accuracy, precision, and recall values obtained by the proposed method compared to the SMOTE method. The results of the testing can be seen in Table 3.

Table 3. Experimental results

Dataset	**Integration of Random Undersampling Ensemble and SP-SMOTE Based on Density Difference in Handling Class Imbalance**			**SMOTE**		
	Accuracy	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**
Haberman	0.91	0.89	0.87	0.89	0.87	0.88
Ecoli1	0.89	0.90	0.88	0.88	0.9	0.87
Yeast-2_vs_4	0.85	0.88	0.90	0.83	0.81	0.82
Shuttle-c0-vs-c4	0.85	0.82	0.88	0.84	0.83	0.84
Yeast4	0.86	0.81	0.79	0.75	0.73	0.75

Based on table 3, it can be observed that the proposed method has advantages over the SMOTE method in terms of accuracy, precision, and recall.

3.2. Discussion

Based on the test results, it was found that the Imbalance Ratio (IR) significantly influences the obtained results in terms of accuracy, precision, and recall. The proposed method achieved average accuracy, precision, and recall values of 0.872, 0.86, and 0.864, respectively, compared to the accuracy, precision, and recall values of 0.838, 0.828, and 0.832 obtained by the SMOTE method. The test results indicate that when the IR increases, the obtained results decrease in terms of accuracy, precision, and recall. Another factor that influences performance is the number of attributes and instances. The larger the number of attributes and instances, the lower the performance in terms of accuracy, precision, and recall.

IV. CONCLUSION

Based on the test results, it was found that the proposed method achieved higher accuracy, precision, and recall values compared to the SMOTE method. This implies that the sampling method and cost-sensitive learning method heavily rely on determining the density difference due to the occasional overlapping of instances in both the majority and minority classes. The research findings also indicate that the values of IR, number of attributes, and number of instances have a significant impact on the obtained results. For optimal results, future research can utilize the feature selection method to reduce the number of attributes.

Conflict of interest

The authors certify that no personal relationships or known competing financial interests could have compromised the integrity of the research presented in this article.

ACKNOWLEDGEMENT

The authors wish to express their gratitude to the editor and reviewers for their insightful feedback and recommendations. The research was funded in part by Universitas IBBI.

REFERENCES

[1]. Alcalá-Fdez, J., Sánchez, L., García, S., Jesus, M. J. del, Ventura, S., Garrell, J. M., Otero, J., Romero, C., Bacardit, J., Rivas, V. M., Fernández, J. C., & Herrera, F. (2009). KEEL: A software tool to assess evolutionary algorithms for data mining problems. Soft Computing, 13(3), 307–318. https://doi.org/10.1007/s00500-008-0323-y

[2]. Antariksa, G., Muammar, R., & Lee, J. (2022). Performance evaluation of machine learning-based classification with rock-physics analysis of geological lithofacies in Tarakan Basin, Indonesia. Journal of Petroleum Science and Engineering, 208, 109250. https://doi.org/10.1016/j.petrol.2021.109250

[3]. Azevedo, F. R. de, Leitão, A. C. L., Lima, M. A. A., & Guimarães, J. A. (2007). Efficacy of natural products to control Callosobruchus maculatus (Fab) in stored Material e Métodos. Revista Ciência Agronômica, 38, 182–187.

[4]. Dai, W., Li, D., Tang, D., Wang, H., & Peng, Y. (2022). Deep learning approach for defective spot welds classification using small and class-imbalanced datasets. Neurocomputing, 477, 46–60. https://doi.org/10.1016/j.neucom.2022.01.004

[5]. Grina, F., Elouedi, Z., & Lefevre, E. (2023). Re-sampling of multi-class imbalanced data using belief function theory and ensemble learning. International Journal of Approximate Reasoning, 156, 1–15. https://doi.org/10.1016/j.ijar.2023.02.006

[6]. Kamalov, F. (2020). Kernel density estimation based sampling for imbalanced class distribution. Information Sciences, 512, 1192–1201. https://doi.org/10.1016/j.ins.2019.10.017

[7]. Kuo, N. I.-H., Garcia, F., Sönnerborg, A., Böhm, M., Kaiser, R., Zazzi, M., Polizzotto, M., Jorm, L., & Barbieri, S. (2023). Generating synthetic clinical data that capture class imbalanced distributions with generative adversarial networks: Example using antiretroviral therapy for HIV. Journal of Biomedical Informatics, 144, 104436. https://doi.org/10.1016/j.jbi.2023.104436

[8]. Law-Ogbomo, K. E., & Egharevba, R. K. A. (2006). The Use of Vegetable Oils in the Control of Callosobruchus maculatus (F) (Coleoptera- Bruchidae) in Three Cowpea Varieties.pdf. Asian Journal of Plant Sciences, 5(3), 547–552.

[9]. Li, Y., Wang, Y., Li, T., Li, B., & Lan, X. (2021). SP-SMOTE: A novel space partitioning based synthetic minority oversampling technique. Knowledge-Based Systems, 228, 107269. https://doi.org/10.1016/j.knosys.2021.107269

[10]. Li, Z., Huang, M., Liu, G., & Jiang, C. (2021). A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. Expert Systems with Applications, 175, 114750. https://doi.org/10.1016/j.eswa.2021.114750

[11]. Loni, A., & Panahi, O. (2015). Control of stored grain pest , Callosobruchus maculatus (F .) (Coleoptera: Bruchidae), using the essential oils isolated from Zingiber officinale (L .) and Mentha pulegium (L .) in laboratory condition. Archives Of Phytopathology And Plant Protection, 48, 434–440.

[12]. Mahdneshin, Z., Vojoudi, S., Ghosta, Y., Safaralizadae, M. H., & Saber, M. (2011). Laboratory evaluation of the entomopathogenic fungi , Iranian isolates of Beauveria bassiana (Balsamo) Vuillemin and Metarhizium anisopliae (Metsch) Sorokin against the control of the cowpea weevil , Callosobruchus maculatus F. (Coleoptera: Bruchid. African Journal of Microbiology Research, 5(29), 5215–5220.

[13]. Mayabadi, S., & Saadatfar, H. (2022). Two density-based sampling approaches for imbalanced and overlapping data. Knowledge-Based Systems, 241, 108217. https://doi.org/10.1016/j.knosys.2022.108217

[14]. Mienye, I. D., & Sun, Y. (2021). Performance analysis of cost-sensitive learning methods with application to imbalanced medical data. Informatics in Medicine Unlocked, 25, 100690. https://doi.org/10.1016/j.imu.2021.100690

[15]. Mirzaei, B., Nikpour, B., & Nezamabadi-pour, H. (2021). CDBH: A clustering and density-based hybrid approach for imbalanced data classification. Expert Systems with Applications, 164, 114035. https://doi.org/10.1016/j.eswa.2020.114035

[16]. Nagawa, C., Böhmdorfer, S., & Rosenau, T. (2015). Chemical composition and anti-termitic activity of essential oil from Canarium schweinfurthii Engl. Industrial Crops & Products, 71, 75–79. https://doi.org/10.1016/j.indcrop.2015.03.078

[17]. Nenaah, G. E. (2013). Potential of using flavonoids , latex and extracts from Calotropis procera (Ait .) as grain protectant s against two coleopteran pests of stored rice. Industrial Crops & Products, 45, 327–334.

[18]. Park, I., Kim, W. H., & Ryu, J. (2024). Style-KD: Class-imbalanced medical image classification via style knowledge distillation. Biomedical Signal Processing and Control, 91, 105928. https://doi.org/10.1016/j.bspc.2023.105928

[19]. Ren, H., Tang, Y., Dong, W., Ren, S., & Jiang, L. (2023). DUEN: Dynamic ensemble handling class imbalance in network intrusion detection. Expert Systems with Applications, 229, 120420. https://doi.org/10.1016/j.eswa.2023.120420

[20]. Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P., & Mononen, J. (2018). Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. Behavioural Processes, 148, 56–62. https://doi.org/10.1016/j.beproc.2018.01.004

[21]. Shi, S., Li, J., Zhu, D., Yang, F., & Xu, Y. (2023). A hybrid imbalanced classification model based on data density. Information Sciences, 624, 50–67. https://doi.org/10.1016/j.ins.2022.12.046

[22]. Steininger, M., Kobs, K., Davidson, P., Krause, A., & Hotho, A. (2021). Density-based weighting for imbalanced regression. Machine Learning, 110(8), 2187–2211. https://doi.org/10.1007/s10994-021-06023-5

[23]. Sun, Z., Song, Q., Zhu, X., Sun, H., Xu, B., & Zhou, Y. (2015). A novel ensemble method for classifying imbalanced data. Pattern Recognition, 48(5), 1623–1637. https://doi.org/10.1016/j.patcog.2014.11.014

[24]. Tao, X., Li, Q., Guo, W., Ren, C., He, Q., Liu, R., & Zou, J. (2020). Adaptive weighted over-sampling for imbalanced datasets based on density peaks clustering with heuristic filtering. Information Sciences, 519, 43–73. https://doi.org/10.1016/j.ins.2020.01.032

[25]. Thölke, P., Mantilla-Ramos, Y.-J., Abdelhedi, H., Maschke, C., Dehgan, A., Harel, Y., Kemtur, A., Mekki Berrada, L., Sahraoui, M., Young, T., Bellemare Pépin, A., El Khantour, C., Landry, M., Pascarella, A., Hadid, V., Combrisson, E., O'Byrne, J., & Jerbi, K. (2023). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. NeuroImage, 277, 120253. https://doi.org/10.1016/j.neuroimage.2023.120253

[26]. Tofel, K. H., Kosma, P., Stähler, M., Adler, C., & Nukenine, E. N. (2017). Insecticidal products from Azadirachta indica and Plectranthus glandulosus growing in Cameroon for the protection of stored cowpea and maize against their major insect pests. Industrial Crops & Products, 110(September), 58–64.

[27]. Vuttipittayamongkol, P., Elyan, E., & Petrovski, A. (2021). On the class overlap problem in imbalanced data classification. Knowledge-Based Systems, 212, 106631. https://doi.org/10.1016/j.knosys.2020.106631

[28]. Yan, Y., Jiang, Y., Zheng, Z., Yu, C., Zhang, Y., & Zhang, Y. (2022). LDAS: Local density-based adaptive sampling for imbalanced data classification. Expert Systems with Applications, 191, 116213. https://doi.org/10.1016/j.eswa.2021.116213

[29]. Zhou, S., Gu, Y., Yu, H., Yang, X., & Gao, S. (2023). RUE: A robust personalized cost assignment strategy for class imbalance cost-sensitive learning. Journal of King Saud University - Computer and Information Sciences, 35(4), 36–49. https://doi.org/10.1016/j.jksuci.2023.03.001