DOI: 10.37943/18PUYJ4315

Zholdas Buribayev
PhD, Acting associate professor, Department of Computer Science
zhburibaev@gmail.com, orcid.org/0000-0002-3486-227X
Al-Farabi Kazakh National University, Kazakhstan

Saida Shaikalamova
BSc, Laboratory assistant, Department of Computer Science
shaikalamova02@gmail.com, orcid.org/0009-0002-9966-508X
Al-Farabi Kazakh National University, Kazakhstan

Ainur Yerkos
MSc, Lecturer, Department of Computer Science
yerkosova@gmail.com, orcid.org/0000-0001-5949-6942
Al-Farabi Kazakh National University, Kazakhstan

Rustem Imanbek
MSc, Software Engineer, Department of Computer Science
imanbek.rustem2000@gmail.com, orcid.org/0009-0008-7261-4382
Al-Farabi Kazakh National University, Kazakhstan

EKMGS: A HYBRID CLASS BALANCING METHOD FOR MEDICAL DATA PROCESSING

Abstract: The field of medicine is witnessing rapid development of AI, highlighting the importance of proper data processing. However, when working with medical data, there is a problem of class imbalance, where the amount of data about healthy patients significantly exceeds the amount of data about sick ones. This leads to incorrect classification of the minority class, resulting in inefficient operation of machine learning algorithms. In this study, a hybrid method was developed to address the problem of class imbalance, combining oversampling (GenSMOTE) and undersampling (ENN) algorithms. GenSMOTE used frequency oversampling optimization based on a genetic algorithm, selecting the optimal value using a fitness function. The next stage implemented an ensemble method based on stacking, consisting of three base (k-NN, SVM, LR) and one meta-model (Decision Tree). The hyperparameters of the meta-model were optimized using the GridSearchCV algorithm. During the study, datasets on diabetes, liver diseases, and brain glioma were used. The developed hybrid class balancing method significantly improved the quality of the model: the F1-score increased by 10-75%, and accuracy by 5-30%. Each stage of the hybrid algorithm was visualized using a nonlinear UMAP algorithm. The ensemble method based on stacking, in combination with the hybrid class balancing method, demonstrated high efficiency in solving classification tasks in medicine. This approach can be applied for diagnosing various diseases, which will increase the accuracy and reliability of forecasts. It is planned to expand the application of this approach to large volumes of data and improve the oversampling algorithm using additional capabilities of the genetic algorithm.

Keywords: imbalance; genetic algorithm (GA); oversampling; undersampling; hybrid; data analysis.

Received: 03.04.2024 Accepted: 27.06.2024 Published: 30.06.2024

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 18, JUNE 2024

Introduction (Literary review)

The rapid development of AI in medicine underscores the importance of proper medical data processing [1]. However, when working with such data, a problem of class imbalance arises [2]. The imbalance is due to the fact that the amount of data about healthy patients often exceeds the amount of data about the sick [3]. This problem leads to incorrect classification of the minority class, which can be accompanied by incorrect predictions and inefficient operation of machine learning algorithms [4].

There are many diverse methods to solve the problem of class imbalance, among which the following can be highlighted: oversampling algorithms, undersampling algorithms, and hybrid methods.

Among the first-class balancing algorithms are ROS (Random Oversampling) and RUS (Random Undersampling), which are based on random duplication/deletion. However, ROS contributes to overfitting due to a large number of homogeneous objects in the minority class [5], while RUS can lead to the loss of significant information [6]. It is important to note that there is a hybrid method ROS-RUS, which starts with random oversampling of data, then mixes data from the minority class with data from the majority class, reducing the dataset until a balance between classes is achieved [7].

The next known representative of oversampling methods is SMOTE (Synthetic Minority Over-sampling Technique), which increases the size of the minority class by generating synthetic samples based on the nearest neighbors of existing objects. However, SMOTE does not take into account hidden noise in the dataset when creating synthetic objects [8]. Nevertheless, there are currently over 85 different modifications of this algorithm, which demonstrate higher efficiency [9]. Some of them are listed below:

MSMOTE, which analyzes the nearest neighbors to classify objects of the minority class into safe, borderline, and noisy. New synthetic samples are created for safe and borderline objects. Synthetic data is not created for noisy samples [7]. Also, ADASYN, which uses the density distribution of weights to determine the number of synthetic samples for each object of the minority class, paying more attention to complex objects, and GASMOTE, (Genetic Algorithm-based SMOTE), which works by optimizing the sampling frequency for each instance of the minority class using a genetic algorithm [10].

The use of the previously considered oversampling methods may prove ineffective due to the need to remove noisy data in the majority class. However, to solve this problem, hybrid algorithms have been developed that demonstrate higher efficiency, which is confirmed by research results. One such study is [11], where the SMOTE+OSS algorithm was presented, which combines SMOTE with the undersampling algorithm OSS, which classifies objects of the majority class into four groups: noisy, borderline, redundant, and safe, among which noisy and borderline samples are removed from the dataset. Researchers Z. Xu, D. Shen, T. Nie, and Y. Kou in their work [8] proposed a new approach for the hybrid method RFMSE, which consists of the oversampling method MSMOTE, as well as the undersampling algorithm ENN, which removes objects from the majority class if their class labels do not match the labels of the majority of their nearest neighbors. Whereas in work [12], a completely different approach to solving the problem of class imbalance was presented. The hybrid balancing method NCL+A-SUWO, which includes an oversampling algorithm [13] using semi-self-learning hierarchical clustering for classifying minority data, adaptively determining the size for oversampling each subcluster, and the undersampling algorithm NCL [14], which combines the concepts of the Condensed Nearest Neighbor (CNN) rule, aimed at removing redundant objects, and the Edited Nearest Neighbors (ENN) rule, intended to eliminate noisy or ambiguous objects.

In this work, we propose a new hybrid method EKMGS, which includes the removal of noisy data from the majority class using the undersampling method ENN [15] and from the minority class using the k-means++ clustering method. The method also provides for the addition of

DOI: 10.37943/18PUYJ4315

synthetic data to the minority class using a modified version of the SMOTE method, based on a genetic algorithm.

Purpose and Objectives of Research

This work aims to develop a hybrid class balancing method for medical data processing. To achieve the goal, the following tasks were set:

- study existing methods for class balancing;
- develop a hybrid class balancing algorithm for processing medical data;
- implement machine learning methods and the ensemble stacking method.

Methods and Materials

This section describes the stages of the research, consisting of three main steps: data collection, hybrid class balancing method, implementation of an ensemble method based on stacking. Below is a description of the research stages shown in Figure 1.

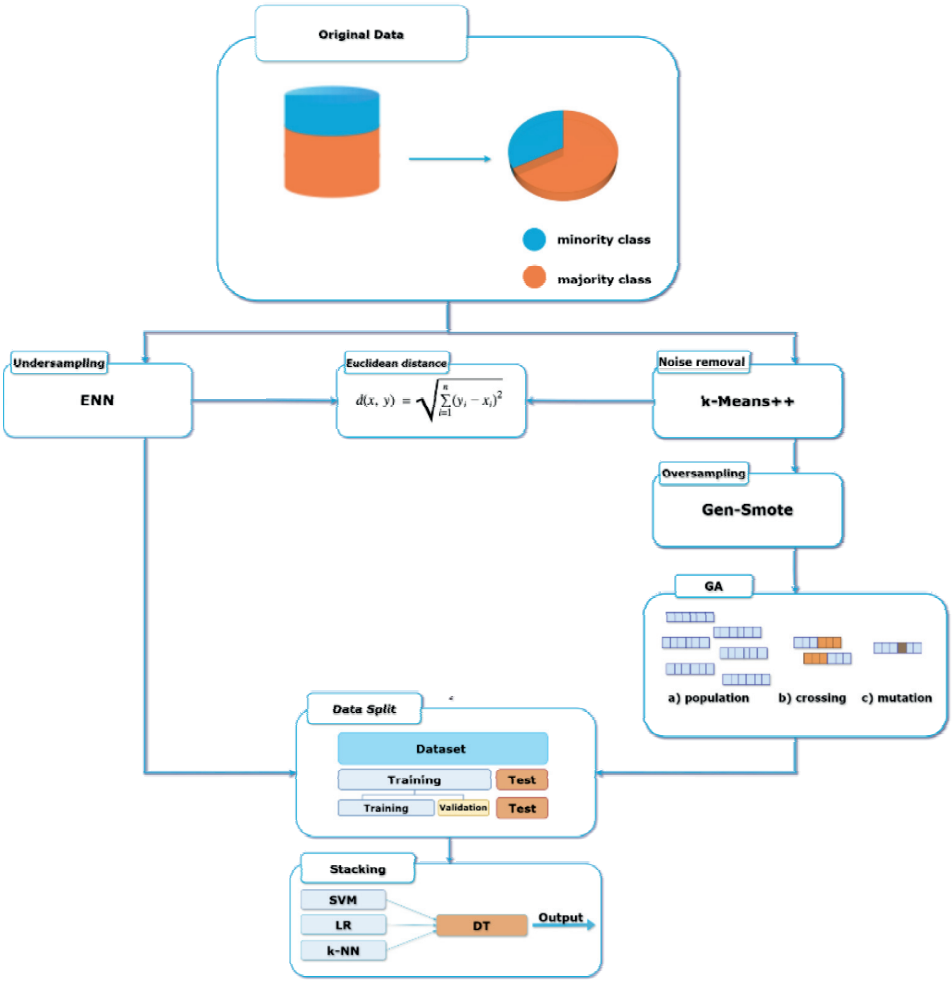


Figure 1. Stages of the study

A. Data collection

In this study, datasets on diabetes, liver diseases, and brain glioma were used [16]. These diseases were chosen due to their widespread prevalence. According to the World Health Organization, 422 million people worldwide suffer from diabetes [17]. And liver diseases cause more than two million deaths annually [18]. Whereas in the United States, six cases of glioma are diagnosed annually per 100,000 people [19]. The datasets were taken from the Kaggle and UCI Machine Learning Repository platforms. All ethical issues were observed. Table 1 presents a complete description of the datasets used.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 18, JUNE 2024

Table 1. The datasets used in the study

№	Name of dataframe	Count of objects	Count of features (with target)	Imbalance ratio
1	Pima Indians Diabetes Database	768	9	9:5
2	Indian Liver Patient Records	583	11	5:2
3	Glioma Grading Clinical and Mutation Features	839	23	13:10

B. Hybrid class balancing

This stage of the research consists of three parts: the undersampling method ENN, removal of noisy data from the minority class using the k-means++ clustering method, and the development of the GenSMOTE oversampling method.

1. ENN (Edited Nearest Neighbours)

ENN (Edited Nearest Neighbours) is one of the undersampling methods, which identifies objects from the majority class as noisy if their class labels do not match the labels of the majority of their nearest neighbors, the search for which is carried out by calculating the distance between objects.

2. k-means++

To remove noisy data from the minority class in this work, the k-means++ clustering algorithm is used. The main motivation for choosing this algorithm is the initialization of centroids. The data of the minority class were divided into K clusters, the number of which was chosen using the silhouette score. The arithmetic mean was calculated for each cluster.

3. GENSMOTE

3.1. SMOTE (Synthetic Minority Over-sampling Technique)

The idea of the SMOTE algorithm is to create new instances of the minority class by interpolating data. In particular, a synthetic instance is generated by calculating the difference between the original instance and its nearest neighbor, this difference is multiplied by a random number from 0 to 1. However, it should be noted that this algorithm has a number of disadvantages, one of which is the use of the same oversampling frequency for all instances of the minority class, which is inefficient, as different instances have different roles in the sample and classification. This work presents a modification of the algorithm that optimizes the oversampling frequency using a genetic algorithm.

3.2. Genetic Algorithm

As is known, the goal of genetic algorithms is to find the optimal solution to a specific problem. For this, the algorithm goes through the following stages: creation of an initial population, calculation of the fitness function, selection methods, crossover, and mutation. Below, the principle of the GenSMOTE oversampling algorithm, based on the listed principles of the genetic algorithm, will be discussed in detail.

3.2.1. Creating the initial population

At this stage, a population of size P is created, where each element, called an individual, is one of the solutions to the given problem. An individual in this case represents a set of genes, each of which corresponds to the oversampling frequency for objects of the minority class. The initialization of genes is done by selecting a random integer value between the upper and lower limits of the sampling frequency (the lower limit is 0, and the upper limit is the number of k nearest neighbors). This can be described by the following formula (1):

DOI: 10.37943/18PUYJ4315

$$F = random((range(minF, maxF)) \quad (1)$$

where $minF$ is the lower limit, and $maxF$ is the upper limit. Accordingly, an individual can then be represented as follows (2):

$$X = (F_1, F_2, F_3, F_4 \ldots, F_N) \quad (2)$$

where X is an individual, and N is the number of elements in the minority class. Then the population will have the following form (3):

$$P = (X_1, X_2, X_3, X_4, \ldots, X_j) \quad (3)$$

where P is the population, and j is its size.

3.2.2. Calculation of the fitness function

At each step of the genetic algorithm iteration, individuals are evaluated using a fitness function, which determines the fitness measure of each individual in the context of the target task that needs to be solved or optimized. In this work, the F1-score was used as the fitness function, which is described by equation (6). In the GenSMOTE algorithm, an individual in the population represents a combination of oversampling frequencies of minority class objects, based on which the SMOTE algorithm is performed for the original dataset. The obtained dataset is used to train the classifier, applying the Decision Tree algorithm. The F1-score value, based on the classification results, is the fitness function, where a higher F1-score corresponds to the best individual.

3.2.3. Selection method

It is known that selection is carried out at the beginning of each iteration of the genetic algorithm cycle to choose those individuals from the current population who will participate in the crossover and mutation methods in the next generation.

In this work, a tournament selection [20], was used, which implements N tournaments for the selection of N individuals. At the same time, the possibility of choosing identical individuals is taken into account. In each tournament, k individuals are randomly selected from the population, after which the best individual is selected from this sample. The best individual is the one who has the highest value of the fitness function.

3.2.4. The method of crossing

The crossover method is used to combine the genetic information of two individuals who act as parents in the process of creating new offspring. The two-point crossover method was used in the work, where two individuals X^l and X^m (4) and (5) are randomly selected.

$$X^l = (F_1^l, F_2^l, F_3^l, F_4^l, \ldots F_N^l) \quad (4)$$

$$X^m = (F_1^m, F_2^m, F_3^m, F_4^m, \ldots F_N^m) \quad (5)$$

Then, two crossover points are randomly selected, after which at these crossover points, genetic information is exchanged between two parental chromosomes to create two new individuals. Equations (6) and (7) present an example where crossover points were used in the interval [3, N-1].

$$X^{l\prime} = (F_1^l, F_2^l, F_3^m, F_4^m, \ldots, F_{N-1}^m, F_N^l) \quad (6)$$

$$X^{m'} = (F_1^m, F_2^m, F_3^l, F_4^l, \ldots, F_{N-1}^l, F_N^m) \quad (7)$$

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 18, JUNE 2024

3.2.5. The mutation method

Mutation in the context of a genetic algorithm represents a process of periodically random updating of the population by introducing new gene combinations. In this work, uniform mutation was used as a mutation method, where a random gene $t \in N$ is replaced with a random number in the range between the minimum and maximum value of the solution space. This method is described by equation (8).

$$F_t = swap(F_t, random(range(minF, maxF)) \quad (8)$$

C. Stacking method

1. Data separation

The data was divided as follows: 95% was used for training and validation, and the remaining 5% was allocated for testing. In addition, 3-fold cross-validation was applied.

2. Memoд stacking

In this work, an ensemble method based on stacking was implemented, consisting of three base models and one meta-model. The following methods were chosen as base models due to differences in their working principles: the k-nearest neighbors method (k-NN), the support vector machine method (SVM), and logistic regression (LR). The predicted values obtained were used as a new dataset for training the meta-model, which is based on a decision tree (Decision Tree). It is important to note that the greedy algorithm GridSearchCV was used to optimize the hyperparameters of the meta-model (max_depth, min_samples_split, min_samples_leaf, max_features), where the cross-validation equals 3.

Below is the pseudocode of the proposed hybrid algorithm EKMGS in Figure 2.

Algorithm 1 EKMGS

Require: ENN, Kmeans++, GenSMOTE
Ensure: optimal collection of sampling rates
  function ENN ($k = 5$, N_{maj}, N_{min}):
    N_{min} the number of nearest neighbors of a minority class
    N_{maj} the number of nearest neighbors of a majority class
    if $N_{min} > N_{maj}$:
      delete points from majority class
  function Kmeans++ (c_i, d_i)**:**
    K is the number of clusters, which was calculated using
silhouette score: $S_i = \frac{c_i - d_i}{\max(c_i, d_i)}$
    μ is arithmetic mean
    n is a number of points from minority class by K
    delete points from minority class $min(\mu(K_n))$
  function GenSMOTE (num_iterations):
    The goal is to find the optimal X (sampling rates) with the Genetic Algorithm
    X = (F_1, F_2, F_3, $F_4 \ldots F_N$)
    P = (X_1, X_2, X_3, $X_4 \ldots X_j$)
    F1_score(SMOTE(P))
    In the case of SMOTE, the synthetic data is generated as follows: $x_{new} \leftarrow EuclideanDistance(x_1, x_2 * random(0, 1))$
make the classification using *DecisionTree* and get train and test sets calculating the fitness function for each X_j(F1_score)
  while $i < num\_iterations$ **do**
    calculate fitness function
    tournament selection
    two point crossing operation and
    uniform mutation
    i++
  end while
  return optimal X

Figure 2. Pseudocode of the hybrid algorithm EKMGS

DOI: 10.37943/18PUYJ4315

Results

During the development of the hybrid class balancing algorithm, a significant increase in the quality assessment metrics of the model was achieved. For example, the F1-score increased in the range from 10% to 75%, while the accuracy increased in the range from 5% to 30%. Below are the results of the F1 and accuracy evaluations, presented in Tables 2 and 3.

Table 2. F1-score/Accuracy before balancing classes

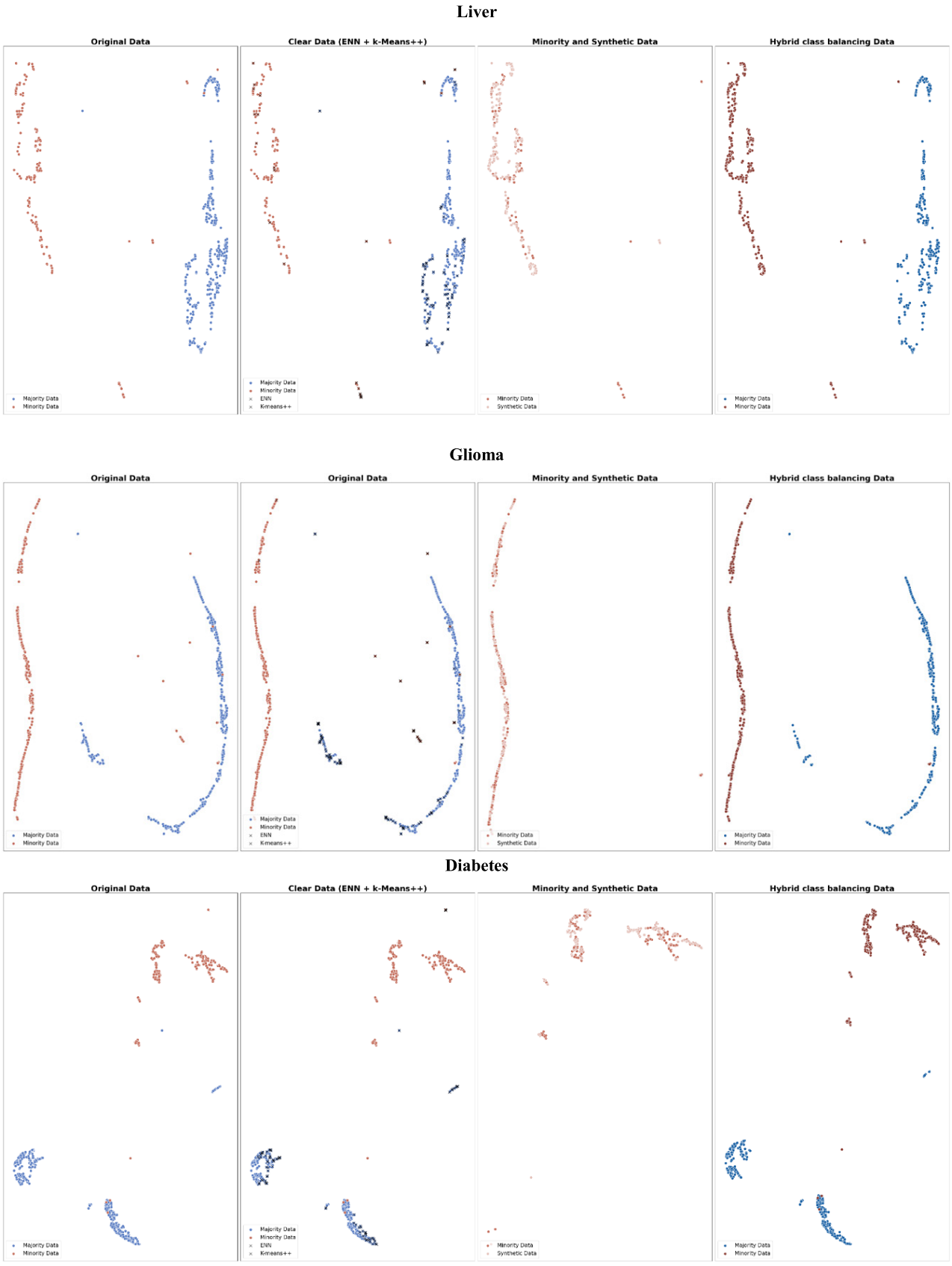
Data	Algorithm			
	LR	k-NN	SVM	Stacking
Diabetes	0.77/ 0.82	0.62/ 0.74	0.62/ 0.74	0.8/ 0.846
Liver	0.166/ 0.655	0.461/ 0.758	0.11/ 0.72	0.2/ 0.724
Glioma	0.878/ 0.88	0.871/ 0.88	0.685/ 0.738	0.878/ 0.88

Table 3. F1-score/Accuracy after balancing classes

Data	Algorithm			
	LR	k-NN	SVM	Stacking
Diabetes	0.864/ 0.871	0.75/ 0.76	0.79/ 0.804	1/1
Liver	0.888/ 0.882	0.857/ 0.853	0.864/ 0.852	1/1
Glioma	0.976/ 0.975	0.913/ 0.902	0.851/ 0.829	1/1

A nonlinear UMAP algorithm was applied for data visualization at each stage. The visualization results are presented in Figure 3.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 18, JUNE 2024



a) The original dataset

b) Data set after the insufficient sampling method (ANN+k-Means++)

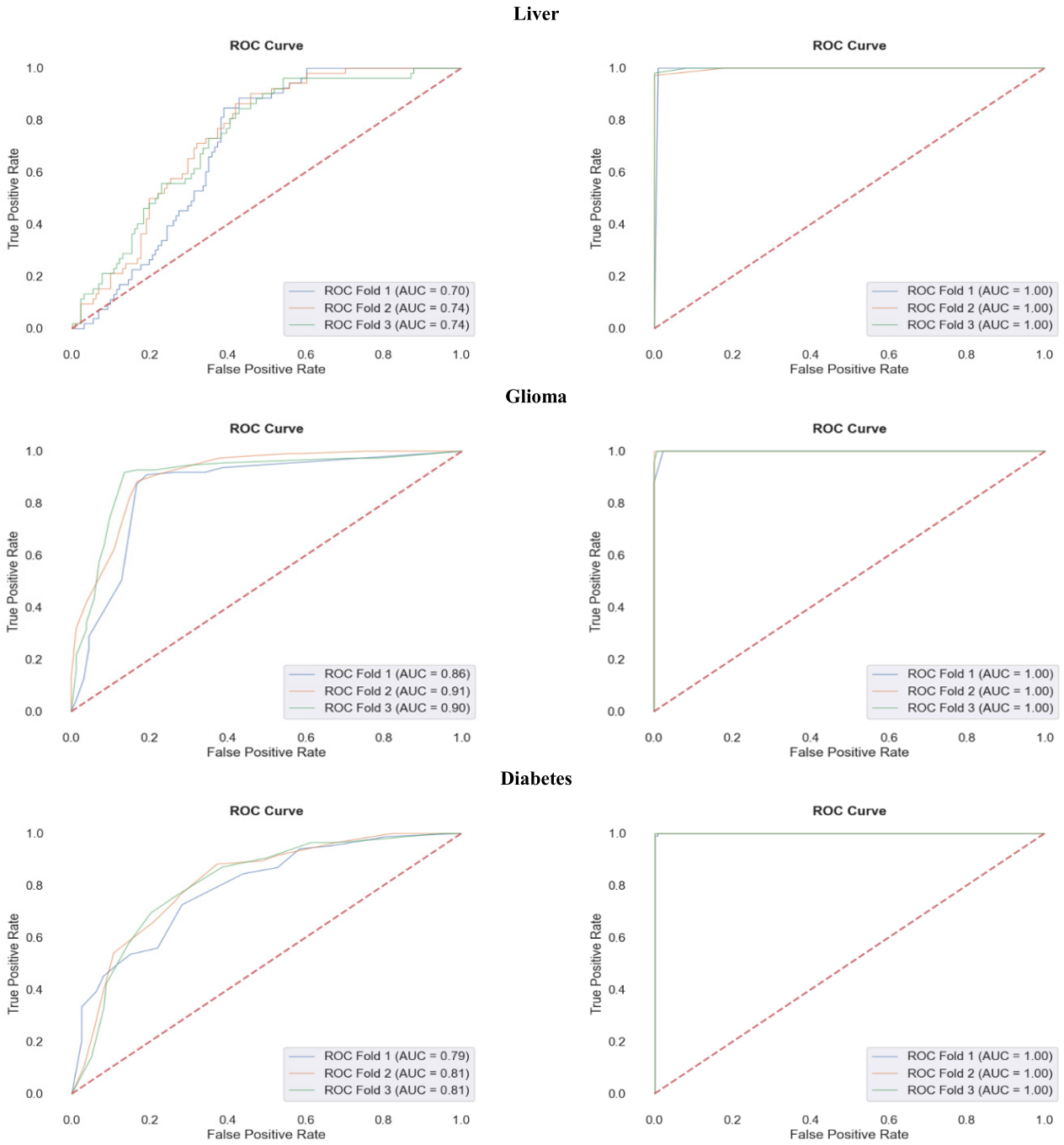
c) Minority class data after the oversampling method (GenSMOTE)

d) Data set after hybrid balancing

Figure 3. Visualization of data before and after hybrid balancing of liver, glioma and diabetes datasets using UMAP

DOI: 10.37943/18PUYJ4315

For the evaluation of classification quality, in addition to F1 and accuracy metrics, the ROC curve and confusion matrix were used, constructed using the sklearn library. Figure 4 shows a graphical representation of the ROC curve before and after hybrid class balancing.



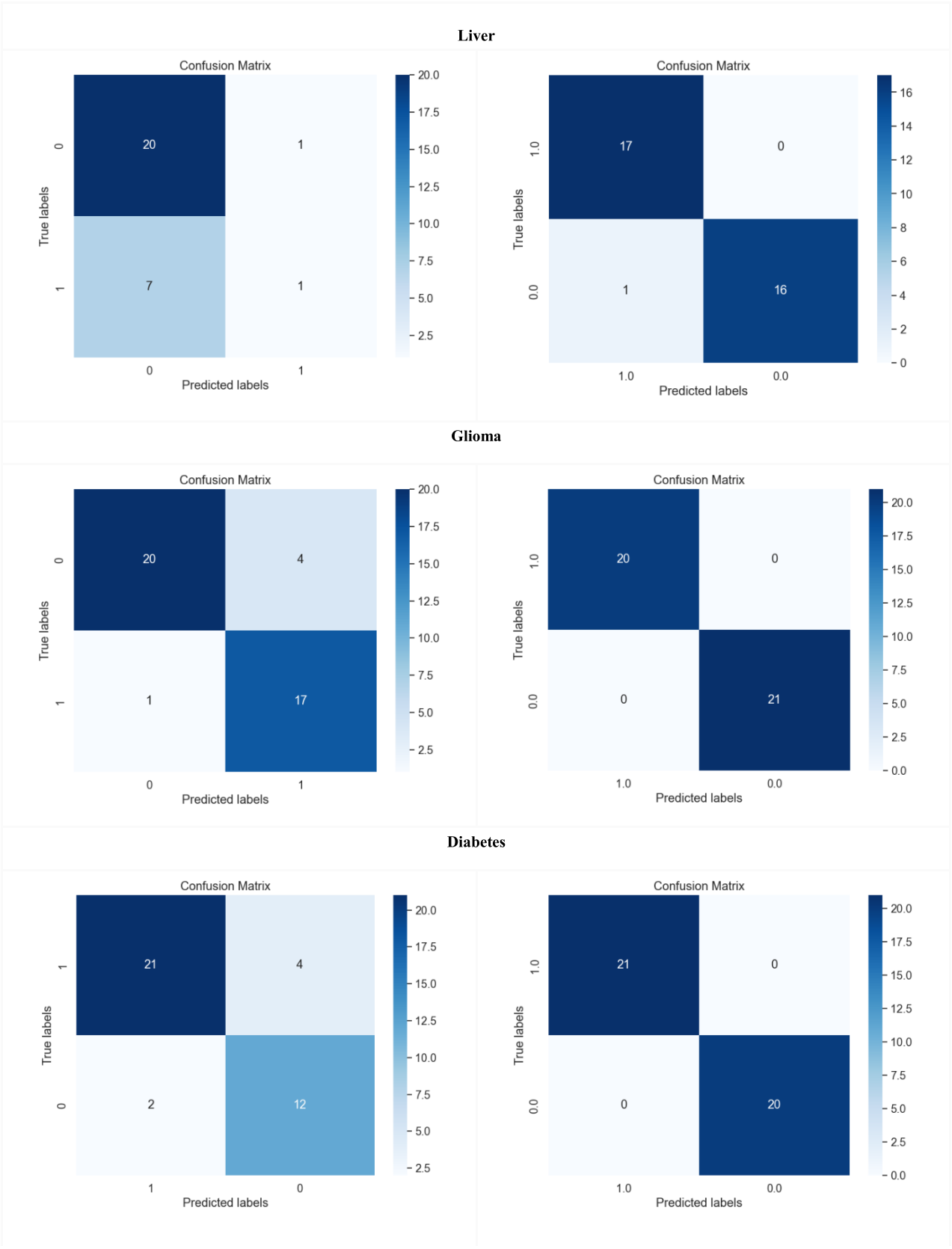
a) ROC curve of the Data Set Stacking method before Hybrid Class Balancing

b) ROC curve of the Data Set Stacking method after hybrid class balancing

Figure 4. ROC curve of the Stacking method before and after hybrid balancing of classes of liver, glioma and diabetes datasets

Whereas Figure 5 shows a visualization of the confusion matrix before and after the hybrid class balancing method of the corresponding datasets.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 18, JUNE 2024



a) The confusion matrix of the Data Set Stacking Method before Hybrid Class Balancing

b) The confusion matrix of the Data Set Stacking method after Hybrid Class Balancing

Figure 5. The matrix of confusion of the Stacking method before and after hybrid balancing of classes of datasets on liver, glioma and diabetes

Discussion

The study demonstrated that class imbalance in medical data can negatively affect the accuracy of machine learning algorithms. This is due to the fact that models trained on such data tend to overfit on the majority class (healthy patients) and poorly recognize the minority class (sick patients).

The hybrid class balancing method proposed in this work, combining GenSMOTE and ENN, demonstrated high efficiency. This is confirmed by the results on three different datasets (diabetes, liver diseases, brain glioma), it was shown that the proposed approach significantly improves the accuracy of classification, obtained before and after the application of hybrid balancing. As a result of applying this method, the F1-score increased in the range from 10% to 75%, and accuracy increased in the range from 5% to 30%. It is important to note that the implementation of the ensemble method based on stacking allowed to further increase the accuracy of classification. This is due to the fact that ensemble methods combine the predictions of several models, which allows to neutralize their shortcomings and improve the final result.

The results obtained are consistent with the conclusions of other studies dedicated to the problem of class imbalance in medical data.

Despite the results obtained, this study has a number of limitations. First, it was conducted on a limited set of data. Second, other methods of class balancing and ensemble learning were not investigated.

In the future, it is planned to conduct more extensive research, in which other datasets and methods of class balancing and ensemble learning will be used.

Conclusion

In the framework of this research, a hybrid class balancing algorithm was developed, based on the genetic algorithm GenSMOTE and ENN. It has a number of advantages such as efficiency, flexibility, and simplicity.

he contribution of the research lies in the development and testing of a new approach to solving the problem of class imbalance in medical data, which can be used to improve the quality of disease diagnosis systems.

It should be noted that this research has some limitations, such as the use of basic GA operators and limited testing of the algorithm on small volumes of data.

Further research will be aimed at improving the efficiency of the algorithm when working with large volumes of data and complex data structures.

References

[1] Xu, Z., Shen, D., Nie, T., Kou, Y., Yin, N., & Han, X. (2021). A cluster-based oversampling algorithm combining SMOTE and k-means for imbalanced medical data. Information Sciences, 572, 574-589. https://doi.org/10.1016/j.ins.2021.02.056

[2] Khushi, M., Shaukat, K., Alam, T. M., Hameed, I. A., Uddin, S., Luo, S., Yang, X., & Reyes, M. C. (2021). A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data. IEEE Access, 9, 109960–109975.

[3] 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

[4] Wang, Y.-C., & Cheng, C.-H. (2021). A multiple combined method for rebalancing medical data with class imbalances. Computers in Biology and Medicine, 134, 104527. https://doi.org/10.1016/j.compbiomed.2021.104527

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 18, JUNE 2024

[5] Lee, D., & Kim, K. (2021). An efficient method to determine sample size in oversampling based on classification complexity for imbalanced data. Expert Systems with Applications, 184, 115442. https://doi.org/10.1016/j.eswa.2021.115442
[6] Abdi, L., & Hashemi, S. (2016). To Combat Multi-Class Imbalanced Problems by Means of Over-Sampling Techniques. IEEE Transactions on Knowledge and Data Engineering, 28(1), 238–251. https://doi.org/10.1109/TKDE.2015.2458858
[7] Malek, N.H.A., Yaacob, W.F.W., Wah, Y.B., Md Nasir, S.A., Shaadan, N., & Indratno, S.W. (2022). Comparison of ensemble hybrid sampling with bagging and boosting machine learning approach for imbalanced data. Indonesian Journal of Electrical Engineering and Computer Science, 29(1), 598. https://doi.org/10.11591/ijeecs.v29.i1.pp598-608
[8] Xu, Z., Shen, D., Nie, T., & Kou, Y. (2020). A hybrid sampling algorithm combining MSMOTE and ENN based on Random forest for medical imbalanced data. Journal of Biomedical Informatics, 107, 103465. https://doi.org/10.1016/j.jbi.2020.103465
[9] Analyticalmindsltd/smote_variants. (2024). [Jupyter Notebook]. analyticalmindsltd. https://github.com/analyticalmindsltd/smote_variants (Original work published 2018)
[10] Jiang, K., Lu, J., & Xia, K. (2016). A Novel Algorithm for Imbalance Data Classification Based on Genetic Algorithm Improved SMOTE. Arabian Journal for Science and Engineering, 41(8), 3255–3266. https://doi.org/10.1007/s13369-016-2179-2
[11] Pristyanto, Y., Setiawan, N.A., & Ardiyanto, I. (2017). Hybrid resampling to handle imbalanced class on classification of student performance in classroom. 2017 1st International Conference on Informatics and Computational Sciences (ICICoS), 207–212. https://doi.org/10.1109/ICICOS.2017.8276363
[12] Choirunnisa, S., & Lianto, J. (2018). Hybrid Method of Undersampling and Oversampling for Handling Imbalanced Data. 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 276–280. https://doi.org/10.1109/ISRITI.2018.8864335
[13] Nekooeimehr, I., & Lai-Yuen, S. K. (2016). Adaptive semi-unsupervised weighted oversampling (A-SUWO) for imbalanced datasets. Expert Systems with Applications, 46, 405–416. https://doi.org/10.1016/j.eswa.2015.10.031
[14] Agustianto, K., & Destarianto, P. (2019). Imbalance Data Handling using Neighborhood Cleaning Rule (NCL) Sampling Method for Precision Student Modeling. 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), 86-89. https://doi.org/10.1109/ICOMITEE.2019.8921159
[15] Tang, B., & He, H. (2015). ENN: Extended Nearest Neighbor Method for Pattern Recognition [Research Frontier]. IEEE Computational Intelligence Magazine, 10(3), 52–60. https://doi.org/10.1109/MCI.2015.2437512
[16] Tasci, E., Zhuge, Y., Kaur, H., Camphausen, K., & Krauze, A.V. (2022). Hierarchical Voting-Based Feature Selection and Ensemble Learning Model Scheme for Glioma Grading with Clinical and Molecular Characteristics. International Journal of Molecular Sciences, 23(22), 14155. https://doi.org/10.3390/ijms232214155
[17] Diabetes. (n.d.). Retrieved 19 March 2024, from https://www.who.int/health-topics/diabetes
[18] Devarbhavi, H., Asrani, S.K., Arab, J.P., Nartey, Y.A., Pose, E., & Kamath, P.S. (2023). Global burden of liver disease: 2023 update. Journal of Hepatology, 79(2), 516–537. https://doi.org/10.1016/j.jhep.2023.03.017
[19] Mesfin, F.B., & Al-Dhahir, M.A. (2024). Gliomas. In StatPearls. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK441874/
[20] Siahaan, W.F.A., Sitompul, O.S., & Situmorang, Z. (2020). The Accuracies of ANFIS and Genetic Algorithm with Tournament Selection on Classifying Hepatitis Data. 1566(1). Scopus. https://doi.org/10.1088/1742-6596/1566/1/012121