VUTTIPITTAYAMONGKOL, P. and ELYAN, E. 2020. Overlap-based undersampling method for classification of imbalanced medical datasets. In *Maglogiannis, I., Iliadis, L. and Pimenidis, E. (eds.) Artificial intelligence applications and innovations: AIAI 2020; proceedings of 16th International Federation for Information Processing working group (IFIP WG) 12.5 International artificial intelligence applications and innovations, 5-7 June 2020, Halkidiki, Greece.* IFIP advances in information and communication technology, 584. Cham: Springer [online], pages 358-369. Available from: https://doi.org/10.1007/978-3-030-49186-4_30.

Overlap-based undersampling method for classification of imbalanced medical datasets.

VUTTIPITTAYAMONGKOL, P. and ELYAN, E.

2020

Overlap-based Undersampling Method for Classification of Imbalanced Medical Datasets

Pattaramon Vuttipittayamongkol and Eyad Elyan

Robert Gordon University, UK
{p.vuttipittayamongkol,e.elyan}@rgu.ac.uk

Abstract. Early diagnosis of some life-threatening diseases such as cancers and heart is crucial for effective treatments. Supervised machine learning has proved to be a very useful tool to serve this purpose. Historical data of patients including clinical and demographic information is used for training learning algorithms. This builds predictive models that provide initial diagnoses. However, in the medical domain, it is common to have the positive class under-represented in a dataset. In such a scenario, a typical learning algorithm tends to be biased towards the negative class, which is the majority class, and misclassify positive cases. This is known as the class imbalance problem. In this paper, a framework for predictive diagnostics of diseases with imbalanced records is presented. To reduce the classification bias, we propose the usage of an overlap-based undersampling method to improve the visibility of minority class samples in the region where the two classes overlap. This is achieved by detecting and removing negative class instances from the overlapping region. This will improve class separability in the data space. Experimental results show achievement of high accuracy in the positive class, which is highly preferable in the medical domain, while good trade-offs between sensitivity and specificity were obtained. Results also show that the method often outperformed other state-of-the-art and well-established techniques.

Keywords: Imbalanced data · Medical diagnosis · Medical prediction · Class overlap · Classification · Undersampling · Nearest neighbour · Machine learning.

1 Introduction

In the past decade, machine learning has been widely used to aid medical diagnosis. Same as in other domains, hidden knowledge can be discovered based on previous information. This is often too complicated to be done by hand or through simple statistical techniques, especially when there are many related features and the data is large. In the medical domain, it is important that prevention and early diagnosis are carried out to avoid further complications and achieve better treatment outcomes [2]. Hence, detecting possible existence or occurrence of diseases is of high interest in supervised learning. This is achieved by training classification models to predict the patients' conditions based on the given symptoms and personal information.

It is common in the medical domain that a dataset has an uneven class distribution. In many situations, the class of interest rarely occurs, hence its samples are limited compared to the other classes. However, traditional learning algorithms are generally designed to maximise the overall prediction accuracy. Thus, on imbalanecd datasets, they tend to be biased in classification towards the majority class and fail to detect anomaly cases. A number of solutions have been proposed to handle classification of datasets with skewed class distributions, so-called imbalanced datasets. Many of them focused on a medical dataset of a specific disease [23,15,1] while others proved their performance on several medical-related datasets [24,10].

Learning from imbalanced medical datasets are seen in a wide range of problems. Besides classification of well-known public datasets such as breast cancer Wisconsin and Pima Indian diabetes, other types of classification tasks have also been carried out. These include classification of electrodiogram (ECG) heartbeats [13], image classification of breast cancer [15] and video classification of bowel cancer [23]. Regardless of problem types, a common objective is to achieve high prediction accuracy, especially on the positive class, which is under-represented.

Rebalancing class distributions seems to be a typical approach to handle imbalanced medical datasets. However, it was shown in the literature that solutions based on improving the visibility of positive samples in the overlapping region could produce significantly higher positive class accuracy (sensitivity) [19,20,4].

In this paper, we propose a framework for improving classification of imbalanced medical datasets. Aiming at high sensitivity on the diagnosis, an overlap-based undersampling method is used. Recursive searching of neighbouring instances is employed to identify instances in the overlapping region. Then, overlapped negative instances are removed to maximise the presence of positive instances to the learning algorithm.

The rest of this paper is organised as follows. In Section 2, we discuss existing methods for handling imbalanced medical datasets. Section 3 gives the details of the proposed framework. Section 4 contains experimental setup including brief descriptions of real-world medical datasets used in the experiment. In Section 5, results and discussion are provided. Finally, in Section 6, we conclude the paper and discuss potential future directions.

2 Related literature

Despite high interests in classification of medical data, the common issue of imbalanced class distributions is not often addressed [14]. This is evidenced by a review paper discussing existing methods used for medical datasets classification [14]. Only 1 out of 71 proposed solutions considered the class imbalanced issue.

To tackle class imbalance, long-established methods such as random undersampling, SMOTE [6], ENN[22] and ADASYN [12] were still used in many recent studies [2,7]. Although improvements in results were reported, they have been constantly outperformed by newer methods.

Novel methods for handling imbalanced medical datasets have also been proposed. In [10], the authors selectively chosen minority class instances for oversampling based on their nearest neighbours. Minority class instances were defined as noise, unstable or boundary samples. Then, noisy instances were removed and only boundary instances were oversampled using linear interpolation techniques. The method showed improvement over SMOTE and an extension of SMOTE. However, it has disadvantages of high parameters dependency and the risk of losing important information in eliminating minority class instances.

In [18], a new technique for determining the final outputs for medical datasets with multiple minority classes was used. Unlike the traditional majority voting approach, classes were assigned based on the highest weighted combination of accuracy, sensitivity, specificity and AUC. Results showed trivial improvement over the traditional method and the improvement might not be attributed to increases in the minority class accuracy, which is highly desirable in the medical domain.

Wan et al. designed a scoring function that assigned ranking to differentiate between minority class and majority class instances [21]. Boosting was adopted to carried out automatic scoring. The method could improved sensitivity on medical datasets further than a cost-sensitive approach and other well-known ensemble-based methods. Moreover, it has the benefit of no prior costs required, which is often unknown and hard to estimate.

One of the latest techniques, Generative Adversarial Net (GAN), was employed in [24] to synthesise minority class instances. It was combined with a multilayer extreme learning machine (ELM) algorithm and showed superior performance to other techniques used with ELM such as weighting and SMOTE. The method also consumed low computational time.

Rather than using a method to broadly handle datasets of multiple diseases, many studies focused on a specific disease such as cancers [7,23,15], polyps [3] and osteoporosis [2]. For instance, Yuan et al. proposed an ensemble-based deep learning approach for detecting bowel cancer [23]. They modified the loss function to penalise the classifier when missclassifying samples that have been correctly classified in the previous iteration. However, results showed that the method was comparable to a long-established ensemble, RUSBoost, in terms of sensitivity and computational time. Other methods for classification of cancer datasets were also proposed [15,16]. In [15], an evolutionary algorithm was used as an undersampling approach to select the most significant samples, then combined with Boosting. The method improved the classification of a breast cancer dataset over other ensemble-based techniques. Similarly, in [16], a cost-sensitive ensemble integrated with a genetic algorithm was proposed to handle an imbalanced breast thermogram dataset. The method provided higher sensitivity than other existing ones. Even so, a common drawback of these ensemble-based solutions is high computational costs.

ECG datasets of heartbeats are also of high interest, and they are generally highly imbalanced, where most heartbeats are normal. With complicated components and morphology of ECG, deep convolutional neural networks (CNN) are often employed for classification tasks [13,1]. CNN is used in combination with many other techniques to enhance results. These include Borderline-SMOTE, feature selection and two-phase training presented in [11]. Two-phase training, introduced by Havaei et al., is known as a promising training technique for imbalanced data. In the first stage, a balanced portion of the data is used to train so that CNN can distinguish different classes. Then in the second stage, the original imbalanced data is fed to fine-tune the output layer parameters.

3 The proposed framework

The proposed framework for improving prediction on imbalanced medical datasets is presented in Fig. 1. Firstly, the training data is preprocessed using normalisation and the overlap-based undersampling technique. Then, the preprocessed data is used to train a learning algorithm to build a predictive model. Finally, the model is evaluated with the testing data.

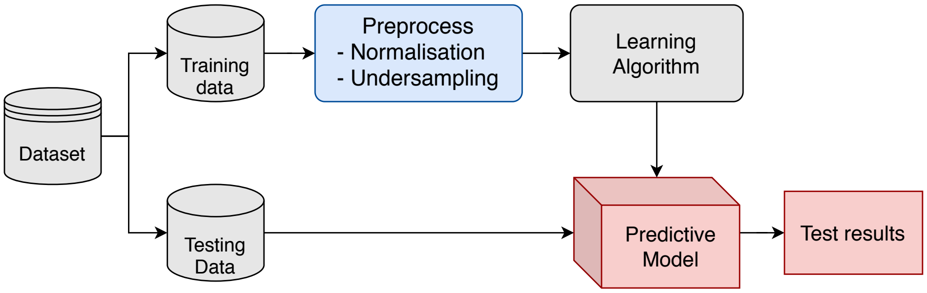


Fig. 1: The proposed framework for classification of imbalanced medical datasets

In the data preprocessing step, we aim at maximising the presence of minority class instances in the overlapping region. The undersampling method based on recursive neigbourhood searching [19] is used. For convenience, we refer to it as URNS. The method will perform a challenging task of identifying overlapped negative instances by considering their k nearest neighbours. These instances are then removed to reduce the complexity of the learning task and reduce the bias classification towards the negative class. Since URNS employs a distance-based technique, its sensitivity to noise has to be concerned. To address the issue, we propose that the data is normalised before URNS is applied. Here, we used standard scores (z-scores) as the normalisation method. The detailed discussion on URNS is provided as follows.

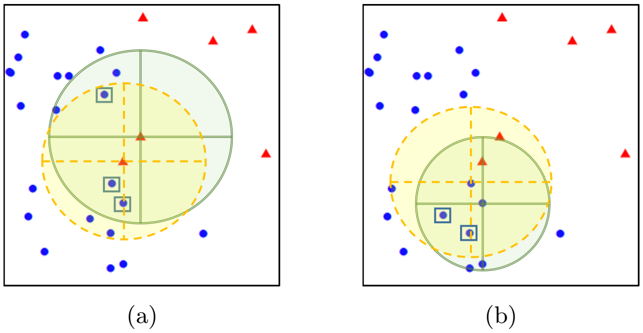
3.1 The URNS method

The main objective of URNS is to maximise the visibility of the minority class to the learning algorithm. This is achieved by eliminating majority class instances from the overlapping region. To identify potential overlapped instances, the k-Nearest Neighbours algorithm (kNN) is used. The local surroundings of each minority class instances are carefully explored. Majority class instances that are in close proximity, which are highly likely to weaken the appearance of minority class ones, are to be removed. However, to prevent excessive elimination, we consider removing only instances that each impacts more than one minority class sample. On the other hand, sufficient elimination is also ensured by the recursive searching. That is the search is carried out twice, where the output of the first round becomes the input of the second round of searching.

Algorithm 1: Recursive Neighbour Search Undersampling

```
Data: training set, $k$
Result: undersampled training set
begin
    $T \leftarrow training\ set$;
    $T_{pos} \leftarrow positive\ instances\ in\ T$;
    Function CommonNeighbour($T$, $Q$, $k$):
        $A \leftarrow frequency\ table$;
        foreach $q \in Q$ do
            $B \leftarrow kNN(q, T, k)$;
            $B_{neg} \leftarrow majority\ class\ members\ of\ B$;
            foreach $y \in B_{neg}$ do
                $A_y.freq \leftarrow A_y.freq + 1$;
        foreach $q \in A.instance$ do
            if $A_x.freq > 1$ then
                $X \leftarrow X \cup \{x\}$;
        return $X$;
    $R_1 \leftarrow CommonNeighbour(T, T_{pos}, k)$;
    $R_2 \leftarrow CommonNeighbour(T, R_1, k)$;
    $\hat{T} \leftarrow T - \{R_1 \cup R_2\}$;
    return $(\hat{T})$
```

Algorithm 1 describes the process of the recursive neighbour search undersampling method, and Fig. 2 illustrates the detection of potential overlapped instances. The method begins with searching for k nearest neighbours of all minority class instances (queries). A majority class neighbour that any two queries have in common is considered as an overlapped instance. This is shown in Fig. 2a,



(a) (b)

Fig. 2: Recursive neighbour Searching involves (a) detecting common majority class neighbours of positive instances followed by (b) searching further for the nearest neighbours they have in common. All majority class common neighbours (in blue boxes) from both steps are then eliminated. [19]

where the common neighbours are marked with blue boxes. Then, to ensure thorough detection, the search is repeated by using the common neighbours detected in the first step as the queries. Their common neighbours are then searched for as depicted in Fig. 2b. Finally, all common neighbours in the majority class from both steps are removed from the training data.

To allow generalisation of the method across any datasets, we present the use of an adaptive k value as shown in Eq. 1 in the neighbour searching algorithm. A rule of thumb where k is related to the square root of the data size (N) was considered. Then, the value of k was adjusted so that it is at the same time proportional to the imbalance ratio (IR). This add-on will also help enhance the discovery of overlapped majority class instances.

$$k = \sqrt{N} + \sqrt{IR} \tag{1}$$

4 Experiment

4.1 Setup

We carried out an experiment using five real-world binary-class datasets. Each dataset was partitioned into training and testing data at 70:30. Random Forest (RF) was chosen as the learning algorithm as it is one of the most-used classifiers for imbalanced datasets [8]. Also, it showed promising results on sensitivity with a better trade-off between sensitivity and specificity than other algorithms [2]. The performance of our method was compared against well-established and state-of-the-art algorithms. These were SMOTE [6], BLSMOTE [9], DBSMOTE [5] and k-means undersampling [17]. The parameters of these methods were set as in the original works. The methods were evaluated in terms of sensitivity, specificity, G-mean and F1-score. Except for KDD's breast cancer, where sufficient data was available, 10-fold cross-validation was used in the training phase for the purpose of model selection.

4.2 Datasets

Five datasets used in the experiment are presented in Table 1 in ascending order of IR with their general information. Wisconsin, Thoracic, Cleveland and Thyroid were obtained from the UCI repository[1]. Breast cancer was given as a challenge in the KDD Cup 2008[2]. We cleaned the datasets so that there were no missing values. In all datasets, the positive class is the minority class.

Table 1: Datasets

dataset	instances	features	IR	%neg
Wisconsin	683	9	1.86	65
Thoracic	470	17	5.71	85.11
Cleveland	173	13	12.31	92.49
Thyroid	7200	21	12.48	92.58
Breast cancer	102294	117	163.2	99.39

Wisconsin breast cancer The Wisconsin breast cancer dataset, widely-known as Wisconsin, was collected at the University of Wisconsin Hospitals, USA during 1989-1991. The class labels are diagnoses of malignant (positive) or benign (negative) breast mass. Other given information is cells characteristics.

Thoracic surgery The data was collected from patients who underwent major lung resections for primary lung cancer at Wroclaw Thoracic Surgery Centre, Poland during 2007-2011. The prediction labels are one-year survival period, which are died (positive) and survived (negative). Model training will be based on patients' personal information, conditions, behaviour and symptoms.

Cleveland heart disease The dataset consists of databases obtained from patients in different regions, namely Cleveland, Long Beach, Hungary and Switzerland. Patients with the presence of heart disease (positive) are to be distinguished from those with absence (negative).

Thyroid The records were collected by the Garavan Institute of Sydney, Australia. The problem is to determine whether a patient referred to the clinic is hypothyroid. The original dataset contains 3 classes: normal, hyperfunction

[1] https://archive.ics.uci.edu/ml/index.php
[2] https://www.kdd.org/kdd-cup/view/kdd-cup-2008

and subnormal function. The normal cases (negative) occupies over 92 % of the dataset and the last two classes are the minority groups. In our experiment, we combined hyperfunction and subnormal function and recognised both cases as hypothyroid (positive).

Breast cancer The dataset is composed of features computed from X-ray images of breasts for early detection of breast cancer. Each sample is labelled with malignant (positive) or benign (negative). This dataset is very large and extremely imbalanced with positive instances of less than 1%.

5 Results and Discussion

Experimental results show that our proposed framework were effective in handling classification of imbalanced medical datasets. URNS showed better results than the well-established and state-of-the-art methods by achieving the highest sensitivity and the highest G-mean on most datasets. Across all datasets, sensitivity and G-mean were significantly improved over the baseline (RF with no resampling). These results are presented in Table 2, 3, 4, 5 and 6, where the highest value in each evaluation metric is highlighted in **bold**.

Table 2: Results on Wisconsin

method	sensitivity	specificity	G-mean	F1-score
baseline	94.37	**96.97**	95.66	**94.37**
URNS	**98.59**	93.18	**95.85**	93.33
SMOTE	94.37[a]	96.97[a]	95.66[a]	**94.37**[a]
BLSMOTE	94.37[a]	96.97[a]	95.66[a]	**94.37**[a]
DBSMOTE	94.37[a]	96.97[a]	95.66[a]	**94.37**[a]
kmUnder	95.77[b]	95.45[b]	95.61[b]	93.79[b]

[a] No changes in the results after applying the method

[b] Results obtained with modified parameter setting

Table 2 shows the results on Wisconsin breast cancer dataset. Our URNS method provided the highest sensitivity of 98.59% and the highest G-mean of 95.85%. These were achieved with high specificity and F1-score. It should be noted that the other methods failed to work on this dataset. In particular, the SMOTE-based methods, i.e., SMOTE, BLSMOTE and DBSMOTE, had no effects on the classification results. This could have been because insufficient positive samples were synthesised, which was due to their objective to rebalance data. As a result, the presence of the positive class, especially around the boundary regions, could not be improved. As opposed, our method does not factor the imbalance ratio and the removal only depends on the amount of class overlap.

Lastly, kmUnder could not be carried out using the k value proposed in the original work since there were fewer distinct samples than k. Thus, we replaced it with $k = N_{minority}/2$. However, it did not give better results than URNS.

Table 3: Results on Thoracic

method	sensitivity	specificity	G-mean	F1-score
baseline	0	**99.17**	0	0
URNS	**95.24**	5.83	23.57	**25.97**
SMOTE	9.52	89.17	29.14	11.11
BLSMOTE	9.52	87.5	28.87	10.53
DBSMOTE	9.52	97.5	30.47	15.38
kmUnder	80.95	20.83	**41.07**	25.56

As shown in Table 3, URNS achieved the best sensitivity and F1-score on Thoracic surgery dataset. It is worth pointing out that this dataset is very hard to classify. This can be seen from the baseline results that none of the positive test cases were correctly identified. Moreover, none of the methods could produce high sensitivity and high specificity at the same time. This high trade-off between the accuracy of the two classes indicates that the dataset is likely to suffer from severe class overlap. Due to this trade-off, even though URNS achieved very high sensitivity of 95.24%, it had the lowest specificity. Thus, URNS is preferable when it is required that nearly all death cases are correctly predicted, otherwise an alternative method providing a more compromised result needs to be explored.

Table 4: Results on Cleveland

method	sensitivity	specificity	G-mean	F1-score
baseline	33.33	**100**	57.74	50
URNS	**100**	93.75	96.82	66.67
SMOTE	**100**	97.92	98.95	85.71
BLSMOTE	**100**	91.67	95.74	60
DBSMOTE	**100**	**100**	**100**	**100**
kmUnder	**100**	39.58	62.92	17.14

From Table 4, our method perfectly classified the positive test cases on the Cleveland heart disease dataset. Its specificity and G-mean were high and comparable to SMOTE, BLSMOTE and DBSMOTE. Due to the high class imbalance nature of the dataset, F1-score of URNS was much lower than those of SMOTE and DBSMOTE even though their specificity values were not far different. This is because F1-score considers true positives and false positives. Thus, in a highly class imbalanced situation, F1-score will be strongly negatively affected by high false positives, which could be misleading when considering the metric alone. Compared to kmUnder, our method provided a substantially higher trade-off between sensitivity and specificity. This could be attributed to less information loss of the URNS method.

Table 5: Results on Thyroid

method	sensitivity	specificity	G-mean	F1-score
baseline	98.74	99.75	99.24	**97.82**
URNS	**100**	99.2	**99.6**	95.21
SMOTE	98.74[a]	99.75[a]	99.24[a]	**97.82**[a]
BLSMOTE	98.11	98.15	98.13	88.64
DBSMOTE	98.74[a]	99.75[a]	99.24[a]	**97.82**[a]
kmUnder	0	**100**	0	0

[a] No changes in the results after applying the method

As can be seen from Table 5, our URNS method provided the best trade-off between sensitivity and specificity on the Thyroid dataset. This is evidenced by the highest G-mean of 99.60%. With the highest sensitivity of 100% achieved, it also yielded high values of specificity and F1-score, which were competitive with the other methods except kmUnder. Note that SMOTE and DBSMOTE led to no changes in the classification results. BLSMOTE had lower performance than the baseline. Lastly, kmUnder completely failed to handle the dataset.

Table 6: Results on Breast cancer

method	sensitivity	specificity	G-mean	F1-score
baseline	29.57	**99.98**	54.37	44.72
URNS	74.73	93.49	**83.59**	12.03
SMOTE	45.16	99.75	67.12	**48.55**
BLSMOTE	33.33	99.89	57.7	44.13
DBSMOTE	36.02	99.84	59.97	44.37
kmUnder	**93.01**	40.27	61.2	1.86

Finally, results on the large and extremely imbalanced dataset of breast cancer are presented in Table 6. Our method achieved the second highest sensitivity, which was lower than kmUnder but significantly higher than the other methods. Essentially higher specificity, G-mean and F1-score indicate that URNS had a better trade-off than kmUnder. URNS showed high specificity and the highest

G-mean of 83.59%. Its low F1-score was due to the bias caused by very high class imbalance as discussed above.

6 Conclusions

In this paper, we handled imbalanced medical datasets using an overlap-based undersampling method. By recursively exploring the neighbourhood of instances, majority class instances potentially in the overlapping region were identified. Then, removal of these instances led to better acknowledgement of minority class instances. Results on real-world datasets showed that the URNS method provided high sensitivity, which is highly desirable in the medical domain, while offering good trade-offs between the accuracy rates of the positive class and the negative class. Moreover, these results were competitive with those of other state-of-the-art and well-established solutions. This can be attributed to some advantages of URNS over other methods. First, the resampling rate is independent of class imbalance and based on the amount of class overlap. Second, the method specifically addresses the problem of class overlap, which often causes errors in classification. Furthermore, this method was implemented with an adaptive k value and no parameter setting is needed. These enable generalisation of the method across any medical datasets. A potential future direction will include improving the framework. A method for setting k value to also be adaptive to the local surroundings of instances such as data density and regional class distribution may improve identification of overlapped instances and hence classification results. To allow wider applicability on real-world medical problems, a framework for multi-class datasets will be developed.

References

1. Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A., San Tan, R.: A deep convolutional neural network model to classify heartbeats. Computers in biology and medicine **89**, 389–396 (2017)
2. Bach, M., Werner, A., Żywiec, J., Pluskiewicz, W.: The study of under-and over-sampling methods' utility in analysis of highly imbalanced data on osteoporosis. Information Sciences **384**, 174–190 (2017)
3. Bae, S.H., Yoon, K.J.: Polyp detection via imbalanced learning and discriminative feature learning. IEEE transactions on medical imaging **34**(11), 2379–2393 (2015)
4. Bunkhumpornpat, C., Sinapiromsaran, K.: Dbmute: density-based majority under-sampling technique. Knowledge and Information Systems **50**(3), 827–850 (2017)
5. Bunkhumpornpat, C., Sinapiromsaran, K., Lursinsap, C.: Dbsmote: density-based synthetic minority over-sampling technique. Applied Intelligence **36**(3), 664–684 (2012)
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research **16**, 321–357 (2002)
7. Fotouhi, S., Asadi, S., Kattan, M.W.: A comprehensive data level analysis for cancer diagnosis on imbalanced data. Journal of biomedical informatics (2019)

8. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G.: Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications **73**, 220–239 (2017)
9. Han, H., Wang, W.Y., Mao, B.H.: Borderline-smote: a new over-sampling method in imbalanced data sets learning. In: International Conference on Intelligent Computing. pp. 878–887. Springer (2005)
10. Han, W., Huang, Z., Li, S., Jia, Y.: Distribution-sensitive unbalanced data oversampling method for medical diagnosis. Journal of medical Systems **43**(2), 39 (2019)
11. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. Medical image analysis **35**, 18–31 (2017)
12. He, H., Bai, Y., Garcia, E.A., Li, S.: Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on. pp. 1322–1328. IEEE (2008)
13. Jiang, J., Zhang, H., Pi, D., Dai, C.: A novel multi-module neural network system for imbalanced heartbeats classification. Expert Systems with Applications: X **1**, 100003 (2019)
14. Kalantari, A., Kamsin, A., Shamshirband, S., Gani, A., Alinejad-Rokny, H., Chronopoulos, A.T.: Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions. Neurocomputing **276**, 2–22 (2018)
15. Krawczyk, B., Galar, M., Jeleń, Ł., Herrera, F.: Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy. Applied Soft Computing **38**, 714–726 (2016)
16. Krawczyk, B., Schaefer, G., Woźniak, M.: A hybrid cost-sensitive ensemble for imbalanced breast thermogram classification. Artificial intelligence in medicine **65**(3), 219–227 (2015)
17. Lin, W.C., Tsai, C.F., Hu, Y.H., Jhang, J.S.: Clustering-based undersampling in class-imbalanced data. Information Sciences **409**, 17–26 (2017)
18. Shilaskar, S., Ghatol, A.: Diagnosis system for imbalanced multi-minority medical dataset. Soft Computing **23**(13), 4789–4799 (2019)
19. Vuttipittayamongkol, P., Elyan, E.: Neighbourhood-based undersampling approach for handling imbalanced and overlapped data. Information Sciences **509**, 47–70 (2020)
20. Vuttipittayamongkol, P., Elyan, E., Petrovski, A., Jayne, C.: Overlap-based undersampling for improving imbalanced data classification. In: International Conference on Intelligent Data Engineering and Automated Learning. pp. 689–697. Springer (2018)
21. Wan, X., Liu, J., Cheung, W.K., Tong, T.: Learning to improve medical decision making from imbalanced data without a priori cost. BMC medical informatics and decision making **14**(1), 111 (2014)
22. Wilson, D.L.: Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics (3), 408–421 (1972)
23. Yuan, X., Xie, L., Abouelenien, M.: A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. Pattern Recognition **77**, 160–172 (2018)
24. Zhang, L., Yang, H., Jiang, Z.: Imbalanced biomedical data classification using self-adaptive multilayer elm combined with dynamic gan. Biomedical engineering online **17**(1), 181 (2018)