Class-Balancing Deep Active Learning with Auto-Feature Mixing and Minority Push-Pull Sampling

Hongxin Lin[1,2] ⋆, Chu Zhang[1,2⋆], Mingyu Wang[1,4], Bin Huang[1,2], Jingjing Shao[3], Jinxiang Zhang[3], Zhenhua Gao[3(✉)], Xianfen Diao[2(✉)], and Bingsheng Huang[1,2(✉)]

[1] Medical AI Lab, School of Biomedical Engineering, Medical School, Shenzhen University, China
huangb@szu.edu.cn
[2] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Medical School, Shenzhen University, China.
laodiao@szu.edu.cn
[3] Department of Radiology, The First Affiliated Hospital, Sun Yat-Sen University, China
gaozhh@mail.sysu.edu.cn
[4] Ainimal Animal Medical Technology (Shenzhen) Co., Ltd,

Abstract. Deep neural networks demand large-scale labeled dataset for optimal performance, yet the cost of annotation remains high. Deep active learning (DAL) offers a promising approach to reduce annotation cost while maintaining performance. However, traditional DAL methods often fail to balance performance and computational efficiency, and overlook the challenge posed by class imbalance. To address these challenges, we propose a novel framework, named **C**lass-**B**alancing **D**eep **A**ctive **L**earning(**CB-DAL**), comprising two key modules: auto-mode feature mixing(Auto-FM) and minority push-pull sampling(MPPS). Auto-FM identifies informative samples by simply detecting in inconsistencies in predicted labels after feature mixing, while MPPS mitigates the class imbalance within the selected training pool by selecting candidates whose features close to the minority class centroid while distant from features of the labelled majority class. Evaluated across varying class imbalance ratios and dataset scales, CB-DAL outperforms traditional DAL methods and the counterparts designed for imbalanced dataset. Our method provides a simple yet effective solution to the class imbalance problem in DAL ,with broad potential applications.

Keywords: Active learning · Class imbalance · Medical image classification.

⋆ Hongxin Lin, Chu Zhang contribute equally to this work

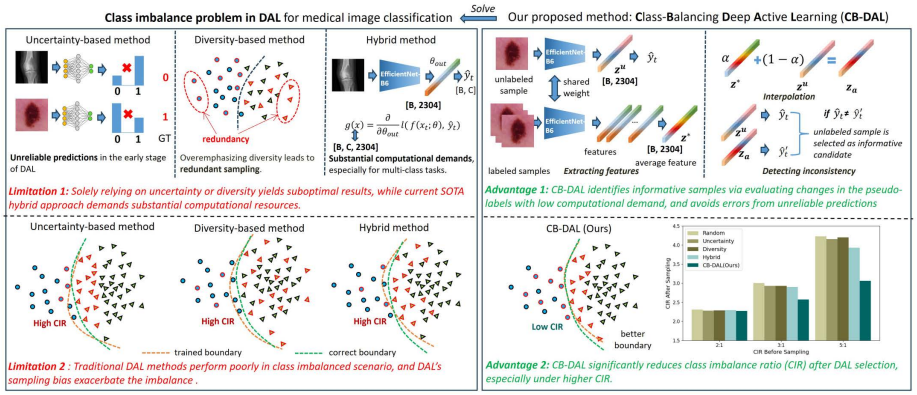

Fig. 1: Limitations and solutions for the class imbalance problem in DAL of medical image classification

1 Introduction

Deep neural networks (DNNs) have achieved great success on medical image analysis [13]. However, their training heavily relies on large-scale labeled datasets to achieve high performance. Annotating medical image is not only laborious and time-consuming, but requires costly expert effort. To tackle this challenge, deep active learning(DAL) [17] is devised to identify and annotate valuable samples from unlabeled data pool. By iteratively training on updated labeled data pool, DAL achieves comparable performance to models trained on fully labelled dataset while reducing annotation costs. Typically, traditional DAL methods utilize criteria such as uncertainty, diversity, or a hybrid approach combining both for sample selection. Nevertheless, they often struggle to strike a balance between optimal performance and computational cost. Although methods solely relying on uncertainty [9, 11] or diversity [6, 19] offer simplicity in implementation, they often yield suboptimal results due to their inherent limitations. Conversely, the state-of-the-art(SOTA) hybrid approach [2] exhibits superior performance, but demands substantial computational resources.

However, traditional DAL methods overlook the class distribution bias in datasets, especially for medical image datasets. Furthermore, the sampling bias issue inherent in DAL [8, 14] leads to a more complex optimization process and suboptimal performance [3, 21]. Although some studies [1, 7, 10] have begun to address the class imbalance issue in DAL. Nevertheless, VaB-AL [7] and BAL [10] resort to difficult-to-train variational autoencoders (VAEs) to model the overall data distribution, substantially increasing computational demands.Moreover, among these works, only BAL [10] has been validated on a real-world imbalanced medical dataset.

Rethinking these limitations of existing DAL methods, as depicted in Fig.1, we propose a framework, named Class-Balancing Deep Active Learning (CB-DAL), and validate our method on two real-world imbalanced dataset. Our con-

tributions are: 1) We propose an effective DAL framework to consider uncertainty and class imbalance at the latent feature space. Employing Auto-mode Feature Mixing(Auto-FM) between labeled feature centroids and unlabeled feature representations, we efficiently identify informative samples via evaluating pseudo-label changes. 2) We propose Minority Push-Pull Sampling (MPPS) based on Euclidean distance to select candidates whose features are close to the centroid of the minority class while distant from features of labelled majority class. 3) Combining Auto-FM with MPPS, CB-DAL achieves superior performance than traditional DAL methods and the counterparts designed for imbalanced dataset.

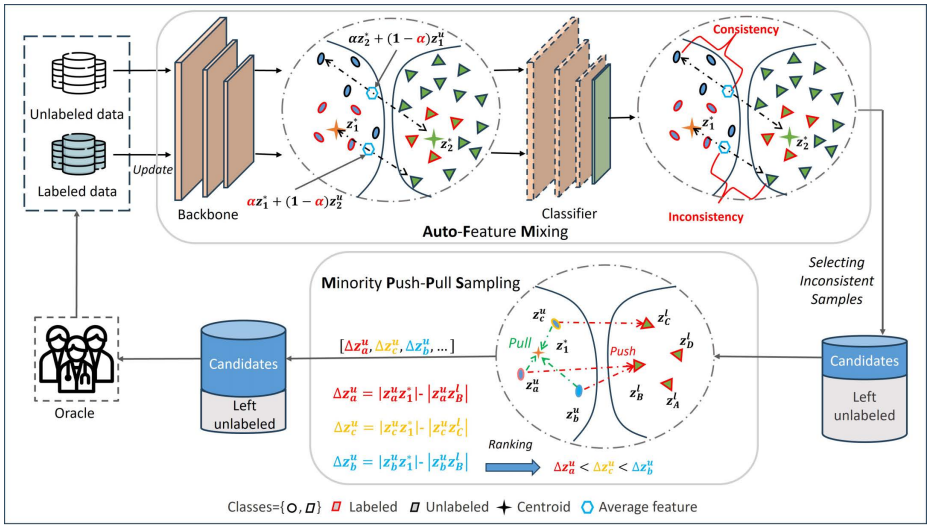

Fig. 2: Overview of the proposed CB-DAL framework for imbalanced medical datasets

2 Methodology

Inspired by BAL [10] and ALFA-Mix [15], our framework enhances active learning efficacy by explicitly addressing the challenges posed by class imbalance (Fig. 2). It contains two modules: Auto-FM for more effectively selecting informative unlabelled samples , and MPPS for mitigating the impact of class imbalance and sampling bias introducing by traditional active learning.

2.1 Framework Formulation

In the framework, given a data pool P, we allocate a labeled data pool P_L and an unlabeled data pool P_U, $P = P_L \cup P_U$ and $\oslash = P_L \cap P_U$. We first use the initial P_L

to pretrain the model $f(\theta)$ for K classes, which parameterized by $\theta = \{\theta_e, \theta_c\}$. Here, $f_e : \mathcal{X} \to \mathbb{R}^D$ is the backbone encoding the input into a D-dimensional representation in a latent space, *i.e.* $z = f_e(x; \theta_e)$. We compute the average representation z^* of the labelled samples per class. Furthermore, $f_c : \mathbb{R}^D \to \mathbb{R}^K$ acts as a classifier, mapping unseen samples from their representations to their corresponding logits. For each round, we first maximize the performance of model with P_L. The subsequent query process can be divided into two stages: In the Auto-FM stage, we select P_{S1} for informative samples. After that, we introduce MPPS to select P_{S2} for class balancing. Finally, we get $P_S = P_{S1} + P_{S2}$, and then P_S is annotated by the oracle and moved from P_U to P_L. When the performance of $f_c(\theta)$ reaches a plateau or the budget is exhausted, active learning can be terminated.

2.2 Auto-Feature Mixing

Feature Mixing in Auto Mode The feature mixing process involves interpolating between the representations of the unlabelled and labelled samples. This is based on the assumption that features from samples near the decision boundary can be easily influenced to predict differently after interpolating.

Firstly, we utilizes f_e to extract features from all samples, and aggregate the features from labelled data pool for each class by computing the average. Secondly, we mix features extracting from unlabeled samples Z_u with each average feature Z^* to obtain new features Z_a. Z_a is denoted below:

$$Z_a = \alpha Z^* + (1 - \alpha) Z^u, \quad \alpha \in [0, 1)^D \tag{1}$$

Finally, Z_a is fed into f_c to obtain updated predictions, if the predictions differ from the original ones, the corresponding unlabeled samples are considered as informative candidates. To optimize the feature mixing process more effectively, we set α as a learnable Gaussian-like matrix, L1 norm of α is added into the loss function to mitigate model overfitting. Thereby, Auto-FM simplifies the process of solving for α and reduces computational demands, while maintaining the high performance advantages of the original approach

Variation in The Loss Function after Feature Mixing Applying Taylor expansion to calculate the loss for mixed features yields:

$$\ell\left(f_c\left(Z_a\right), y^*\right) \approx \ell\left(f_c\left(Z^u\right), y^*\right) + \left(\alpha\left(Z^* - Z^u\right)^T \cdot \bigtriangledown \ell\left(f_c\left(Z^u\right), y^*\right)\right) \tag{2}$$

$$\max\left[\ell\left(f_c\left(Z_a\right), y^*\right) - \ell\left(f_c\left(Z^u\right), y^*\right)\right] \approx \max[\left(\alpha\left(Z^* - Z^u\right)^T \cdot \bigtriangledown \ell\left(f_c\left(Z^u\right), y^*\right)\right)] \tag{3}$$

The variation in the loss function after feature mixing is influenced by two conditions: 1) the disparity between Z^* and Z_u; 2) the gradient of the loss. The former indicates the distinctiveness of the features and the degree of feature discrepancy between labeled and unlabeled data, while the latter affects the

model's sensitivity to features. If there is a difference in labeled and unlabeled data yet this results in only minimal changes in the model's loss gradient, these features may not be distinct enough for the model. Therefore, during training, the final loss function is denoted below:

$$Loss = 1 - \ell\left(f\left(Z^a, y^*\right)\right) + \|\alpha\| \tag{4}$$

2.3 Minority Push-Pull Sampling

To address the challenges posed by class imbalance and mitigate the negative impact of sampling bias of AL, we introduce Minority Push-Pull Sampling (MPPS). Unlike label-based methods for imbalance, MPPS is a feature-based approach that identifies samples belonging to minority class in latent feature space. By "pulling" features from unlabeled samples while "pushing" them away from the majority class features, MPPS is denoted below:

$$X_s = argmin\left\{\ell\left(e\left(x_u\right), \mu\left(c_t^{minor}\right)\right) - \min_{\forall c_t^{maj} \in C_t^{MAJ}} \ell\left(e\left(x_u\right), \mu\left(c_t^{maj}\right)\right)\right\} \tag{5}$$

X_s represents the selected samples, $\ell\left(\cdot\right)$ represents the Euclidean distance between two sets of features, $e\left(x_u\right)$ represents the encoding vectors of an unlabeled sample, and μ represents the average feature values for a specific class within the labeled samples. c_t^{minor} denotes features of minority classes, c_t^{maj} denotes features of majority classes, and t indicates the current t-th round of iteration.

The expression aims to minimize the difference between the features of the current sample and those of the minority class while maximizing the difference with the majority class in the labeled data pool.

3 Experiments and Results

3.1 Dataset and Evaluation

Bone Tumor Dataset This private dataset comprises radiographs obtained from four centers, totaling 333 patients(osteolytic OS:136, GCT:197). Two radiologists, each with over 10 years of experience in reading musculoskeletal radiographs and blinded to the study, independently reviewed all radiographs and selected the patients included in the study. All included bone tumors were pathologically confirmed. We designate the primary center A as the internal dataset(osteolytic OS:124, GCT:155), while the remaining three centers(osteolytic OS:12, GCT:42) serve as the external test set.

ISIC2020 Dataset This public dataset [18] contains 33,126 images (benign:32542, malignant:584) of benign and malignant skin lesions from 2,056 patients. The dataset was curated by the International Skin Imaging Collaboration (ISIC) and includes images from six different sources. All malignant diagnoses have

been confirmed via histopathology, while benign diagnoses have been confirmed using using expert agreement, longitudinal follow-up, or histopathology.

We primarily utilize the area under the receiver operating characteristic (AUC) for evaluation.

3.2 Implementation Details

For Bone Tumor Dataset, all X-ray images are resized to 1080x1080. We randomly split the internal dataset into a training set and an internal test set. Considering its small data scale, we adopt a partial annotation setting at different percentages(50%, 60%, 70%, 80%) of the training set. We randomly selected 30% of each class from the training set to initialize the model, while the remaining data is used for the DAL process. As the class imbalance in the original dataset is not severe enough, We resample the original training set to three imbalanced ratios (2:1, 3:1, 5:1) to validate our method. The batch size is set to 2.

For ISIC2020 Dataset, we follow the setting in [22]. We randomly divide this dataset into a training set, a validation set and a test set at a ratio of 8: 1: 1. Considering its large data scale, we adopt a partial annotation setting at at different percentages(10%, 20%, 50%) of training set. We randomly selected 5% of each class from the training set to initialize the model, while the remaining data is used for AL process. The batch size is set to 128.

For both datasets, in each round, we allocate an equal budget to select the most informative samples and minority class samples. We adopt EfficientNet-B6 with weights pretrained on ImageNet, using its backbone as the feature extractor and fully connected layers as the classifier. Binary Cross Entropy (BCE) loss is employed as the loss function. We employ the Adam optimizer and utilize cosine annealing to reduce the learning rate from 1e-4 to 1e-12 over all 300 epochs. Our method is implemented in Pytorch, using an NVIDIA RTX TITAN GPU with 24GB memory. The weights used for testing were selected based on the best-performing AUC on the internal test set or validation set.

3.3 Comparison and Ablation Experiments

To analyze of the effectiveness of Auto-FM, we compare our CB-DAL* (CB-DAL without MPPS) with other traditional DAL methods, including Entropy [20], Core-Set [19], BALD [11], BADGE [2], and ALFA-Mix($\alpha = 0.5$) [15]. The results of these comparison experiments on the original Bone Tumor Dataset are reported in Table 1. CB-DAL* outperforms all traditional DAL methods and random sampling method at each annotation budget, particularly under low budget conditions.

Table 1: Comparison experiments on internal/external test set of Bone Tumor Dataset, CB-DAL* is our proposed DAL without MPPS for class imbalance

	Annotation Ratio				
Method	50%	60%	70%	80%	100%
Full	/	/	/	/	**0.935/0.925**
Random	0.521/0.581	0.669/0.658	0.769/0.776	0.834/0.839	/
Entropy [20]	0.710/0.679	0.766/0.806	0.802/0.828	0.855/0.853	/
Core-Set [19]	0.748/0.702	0.815/0.790	0.866/0.857	0.877/0.877	/
BALD [11]	0.753/0.754	0.839/0.802	0.869/0.827	0.893/0.849	/
BADGE [2]	0.755/0.738	0.838/0.816	0.862/0.849	0.890/0.887	/
ALFA-Mix [15]	0.783/0.772	0.860/0.845	0.899/0.889	0.918/0.899	/
CB-DAL*	**0.820/0.800**	**0.872/0.865**	**0.902/0.899**	**0.923/0.903**	/

We also analyze the effectiveness of auto-learned α. As shown in Fig 3, CB-DAL* with auto-learned α demonstrates optimal performance and avoids disadvantages introduced by manual settings.

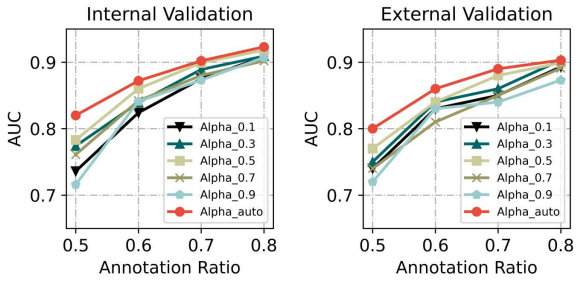

Fig. 3: The effectiveness of auto-learned α on the Bone Tumor Dataset.

To address class imbalance problem in DAL, we combine Auto-FM with MPPS, and validate the effectiveness of MPPS on three class imbalance ratios (2:1, 3:1, 5:1) of the Bone Tumor Dataset. With MPPS, CB-DAL demonstrates a significant improvement over CB-DAL*. To further validate the effectiveness of MPPS, we compare MPPS with other counterparts for class imbalance,such as over-sampling [4, 16] and Focal loss [12] and LADM [5] . CB-DAL maintains its advantage, especially under high class imbalance ratio. All results of these experiments are shown in Fig. 4.

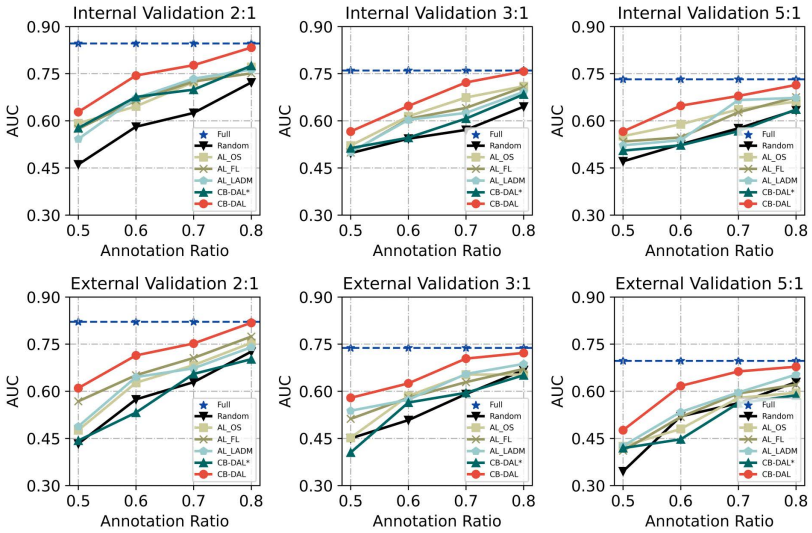

Fig. 4: The comparison experiments with other class-balancing counterparts and the ablation experiments of CB-BAL on the Bone Tumor Dataset

Table 2: Comparison experiments on ISIC2020 Dataset

	Annotation Ratio			
Method	10%	30%	50%	100%
Full	/	/	/	**0.904**
Random	0.793	0.803	0.830	/
BAL	0.834	0.873	0.876	/
CB-DAL	**0.852**	**0.883**	**0.896**	/

4 Conclusion

In this paper, we proposed a simple yet effective framework, CB-DAL, to strike a balance between performance and computational efficiency and solve class imbalance problem in DAL for medical image classification. Based on Auto-FM and MPPS, CB-DAL achieves optimal performance and high efficiency through simple arithmetic operations between features. CB-DAL outperforms traditional DAL methods and counterparts designed for imbalanced datasets across varying class imbalance ratios and data scales, especially when faced with limited resources and high class imbalance.

Acknowledgments. This work was funded by Natural Science Foundation of Guangdong Province, China (Grant No.2022A1515011593) and Medical Science, Technology Foun- dation of Guangdong Province, China (Grant No.A2021010) and Shenzhen Municipal Scheme for Basic Research (JCYJ20210324100208022).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aggarwal, U., Popescu, A., Hudelot, C.: Active Learning for Imbalanced Datasets. In: IEEE Winter Conference on Applications of Computer Vision, pp. 1417–1426 (2020)
2. Ash, J.T., Zhang, C., Krishnamurthy, A., Langford, J., Agarwal, A.: Deep Batch Active Learning by Diverse, Uncertain Gradient Lower Bounds. In: International Conference on Learning Representations (2020)
3. Bengar, J.Z., van de Weijer, J., Fuentes, L.L., Raducanu, B.: Class-Balancing active learning for image classification. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1536–1545 (2022)
4. Bria, A., Marrocco, C., Tortorella, F.: Addressing class imbalance in deep learning for small lesion detection on medical images. Computers in Biology and Medicine **120**, 103735 (2020)
5. Cao, K., Wei, C., Gaidon, A., Arechiga, N., Ma, T.: Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss, in: Advances in Neural Information Processing Systems **32**, (2019)
6. Cao, X.: A divide-and-conquer approach to geometric sampling for active learning. Expert Systems with Applications **140**, 112907.(2020)
7. Choi, J., Yi, K.M., Kim, J., Choo, J., Kim, B., Chang, J., Gwon, Y., Chang, H.J.: VaB-AL: Incorporating Class Imbalance and Difficulty With Variational Bayes for Active Learning. In: Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6749–6758 (2021)
8. Dasgupta, S., Hsu, D. Hierarchical sampling for active learning, In: Proceedings of the 25th International Conference on Machine Learning. pp. 208–215. (2008)
9. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning, In: International Conference on Machine Learning. pp. 1050–1059. (2016)
10. Jin, Q., Yuan, M., Wang, H., Wang, M., Song, Z.: Deep active learning models for imbalanced image classification. Knowledge-Based Systems **257**, 109817. (2022)
11. Kirsch A, Van Amersfoort J, Gal Y. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning[J]. Advances in neural information processing systems 32. pp. 7026-7037. (2019)
12. Lin, T.-Y. et al.: Focal Loss for Dense Object Detection, In. Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988.(2017)
13. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88. (2017)
14. MacKay, D.J.: Information-based objective functions for active data selection. Neural computation **4**, 590–604. (1992)

15. Parvaneh, A., Abbasnejad, E., Teney, D., Haffari, G. (Reza), van den Hengel, A., Shi, J.Q.: Active Learning by Feature Mixing. In: Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12237–12246. (2022)
16. Qu, W., Balki, I., Mendez, M., Valen, J., Levman, J., Tyrrell, P.N.: Assessing and mitigating the effects of class imbalance in machine learning with application to X-ray imaging. Int J Comput Assist Radiol Surg **15**, 2041–2048. (2020)
17. Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B.B., Chen, X., Wang, X.: A Survey of Deep Active Learning. ACM Computing Surveys **54**, 180:1-180:40. (2021)
18. Rotemberg, V., Kurtansky, N., Betz-Stablein, B., et al.: A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Scientific data, **8**(1), 34. (2021)
19. Sener, O., Savarese, S.: Active Learning for Convolutional Neural Networks: A Core-Set Approach. arXiv preprint arXiv:1708.00489 (2018)
20. Shannon, C. E.: A Mathematical Theory of Communication. SIGMOBILE Mob. Comput. Commun. Rev. **5**(1), 3–55. (2001)
21. Yang, Y., Xu, Z.: Rethinking the value of labels for improving class-imbalanced learning. Advances in neural information processing systems 33, pp. 19290–19301. (2020)
22. Öztürk, Ş. and Çukur, T.: Deep Clustering via Center-Oriented Margin Free-Triplet Loss for Skin Lesion Detection in Highly Imbalanced Datasets. IEEE Journal of Biomedical and Health Informatics, **26**(9), pp. 4679–4690.(2022)