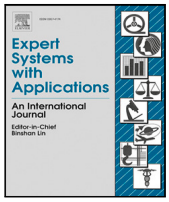
Expert Systems With Applications 251 (2024) 123783

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Adaptive unified contrastive learning with graph-based feature aggregator for imbalanced medical image classification

Cong Cong [a,*], Sidong Liu [b,c], Priyanka Rana [b], Maurice Pagnucco [a], Antonio Di Ieva [c], Shlomo Berkovsky [b], Yang Song [a]

[a] School of Computer Science and Engineering, University of New South Wales, Australia
[b] Centre for Health Informatics, Macquarie University, Australia
[c] Computational NeuroSurgery Lab, Macquarie University, Sydney, Australia

ARTICLE INFO

Keywords:
Imbalanced classification
Convolutional graph neural networks
Self-supervised learning

ABSTRACT

Medical image datasets are often imbalanced due to biases in data collection and limitations in acquiring data for rare conditions. Addressing class imbalance is crucial for developing reliable deep-learning algorithms capable of effectively handling all classes. Recent class imbalanced methods have investigated the effectiveness of self-supervised learning (SSL) and demonstrated that such learned features offer increased resilience to class imbalance issues and obtain much improved performances over other types of class imbalanced methods. However, existing SSL methods either lack end-to-end capabilities or require substantial memory resources, potentially resulting in sub-optimal features and classifiers and limiting their practical usage. Moreover, the conventional pooling operations (*e.g.*, max-pooling, or average-pooling) tend to generate less discriminative features when datasets pose high inter-class similarities. To alleviate the above issues, in this study, we present a novel end-to-end self-supervised learning framework tailored for imbalanced medical image datasets. Our framework constitutes an adaptive contrastive loss that can dynamically adjust the model's learning focus between feature learning and classifier learning and a feature aggregation mechanism based on Graph Neural Networks to further enhance feature discriminability. We evaluate the effectiveness of our framework on four medical datasets, and the experimental results highlight its superior performance in imbalanced image classification tasks.

1. Introduction

Deep learning has demonstrated remarkable advancements in medical image classification. Typically, a substantial amount of labelled samples across all classes is required to train deep learning deep-learning classifiers. However, data collection for biomedical tasks can be challenging, often due to the low incidence of certain diseases (Gao, Zhang, Liu, & Wu, 2020). Thus, medical image datasets are often imbalanced, where certain classes possess far higher numbers of samples compared to other classes in the datasets.

Training deep learning models on imbalanced datasets often results in biased models. As depicted in Fig. 1, we observe that a ResNet model (He et al., 2016) (base), trained on the imbalanced APTOS2019 dataset, demonstrates better performance on classes with larger sample sizes, while exhibiting relatively poorer performance on the minority classes. The inherent challenge lies in the fact that the minority classes, which have fewer samples, struggle to accurately represent the true distribution within the embedding space. Additionally, the majority classes, with their larger sample sizes, exert significant influence, thereby compressing the spatial span of the minority classes. Consequently, the predominance of gradients originating from the majority class samples biases the learning process in favour of these majority classes.

In order to address the class imbalance issue in medical datasets, existing studies have explored various techniques. Data re-sampling (Chai et al., 2022; Dai, Li, Tang, Wang, & Peng, 2022; Galdran, Carneiro, & González Ballester, 2021; Rana, Sowmya, Meijering, & Song, 2022b, 2023) is one the most widely explored approach. However, it is often observed that employing these techniques without taking into account the inherent characteristics of the task at hand causes over-fitting which in turn leads to poor classification performance. On the other hand, loss-weighting techniques (Ghorbani, Kazi, Baghshah, Rabiee, & Navab, 2022; Pan et al., 2023; Wei, Zhou, Li, & Xu, 2023; Yoon, Hamarneh,

* Corresponding author.
E-mail addresses: c.cong@student.unsw.edu.au (C. Cong), sidong.liu@mq.edu.au (S. Liu), priyanka.rana@mq.edu.au (P. Rana), morri@unsw.edu.au (M. Pagnucco), antonio.diieva@mq.edu.au (A. Di Ieva), shlomo.berkovsky@mq.edu.au (S. Berkovsky), yang.song1@unsw.edu.au (Y. Song).

https://doi.org/10.1016/j.eswa.2024.123783
Received 28 November 2023; Received in revised form 18 January 2024; Accepted 18 March 2024
Available online 18 April 2024

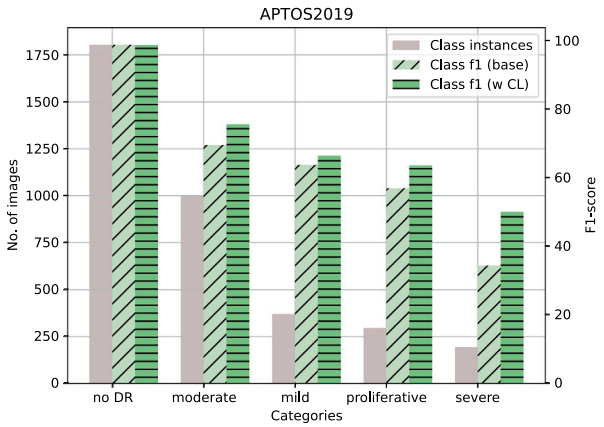


Fig. 1. The distribution of class instances in the APTOS2019 dataset and the corresponding per-class F1 scores of different models are depicted.

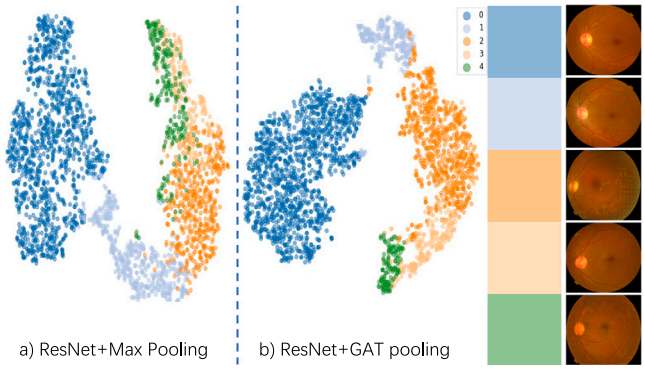


Fig. 2. The TSNE visualisation of the features obtained by a ResNet model trained with different feature aggregators on APTOS2019. In addition to the inherent challenge of class imbalance, the dataset presents significant inter-class similarities, further complicating the classification task in medical datasets.

& Garbi, 2019; Zhang, Tan, Li, & Hong, 2018) are also widely used. However, they typically achieve enhanced performance in minority classes at the expense of performance in majority classes and they often disregard the distinct properties of feature learning and classifier learning.

Accordingly, feature learning-based approaches have been proposed. These methods, which resort to contrastive Learning (CL) techniques (Cui, Zhong, Liu, Yu, & Jia, 2021; Marrakchi, Makansi, & Brox, 2021), often show state-of-the-art performances in alleviating the issue of class imbalance by acquiring more robust feature representations (Liu, HaoChen, Gaidon, & Ma, 2021). To further illustrate the effectiveness of CL, we conducted a two-stage training using a ResNet model, where we performed CL in the first stage followed by classifier training in the second stage. The results, shown in Fig. 1, reveal that the inclusion of CL (w CL) significantly enhances the representation of under-represented classes. This, in turn, facilitates more precise and robust classification. Although effective, the conventional two-stage learning scheme in CL often brings two drawbacks, namely, the model is not trained end-to-end, which leads to incompatible feature and classifier learning, and a large batch size which requires an extra memory queue to achieve a robust feature representation. These issues significantly increase the training costs and impact their practical applicability in real clinical scenarios.

Moreover, medical images typically pose high inter-class similarities. For example, as shown in Fig. 2, fundus images from the APTOS2019 dataset exhibit notable inter-class similarity. This characteristic poses additional challenges when dealing with imbalanced medical datasets. The existing methods described in the literature commonly utilise a CNN-based classifier with a pooling layer for feature aggregation, followed by a fully-connected classification layer. However, as illustrated in Fig. 2(a), while a conventional pooling-based feature aggregator can generate discriminative features for the majority classes, it often produces inseparable features for the minority classes, which leads to poor performance on minority classes. Hence, there is a need for improved feature aggregators to achieve effective classification results when dealing with datasets characterised by significant inter-class similarities.

To address the limitations of existing SSL-based methods and enhance their performance on medical image datasets, which frequently exhibit high inter-class similarities, we propose Adaptive Unified contrastive learning with a Graph-based feature aggregator (AdUniGraph). AdUniGraph is an end-to-end self-supervised learning framework that simultaneously performs feature and classifier learning. To combine two different learning stages, we employ an adaptive unified loss which effectively fuses the contrastive loss with the cross-entropy loss with an adapter parameter. Such a design allows us to obtain better feature quality and remove the requirement of a memory queue, leading to reduced training costs and better performance. Moreover, to obtain more discriminative features for minority classes, we include a Convolutional Graph Neural Networks (ConvGNN) algorithm as the feature extractor. Replacing the conventional pooling operations with ConvGNN-based aggregator better captures the inter-dependencies among different receptive fields, which leads to more discriminative features. As demonstrated in Fig. 2(b), ConvGNN-based pooling facilitates the formation of denser and more distinct feature clusters, thereby enhancing intra-class interactions while weakening inter-class interactions. Specifically, our contributions can be summarised as follows:

- We propose an end-to-end self-supervised learning framework designed to produce enhanced feature representations across all classes for alleviating the issue of class imbalanced medical image classification.
- Our framework employs an adaptive unified loss which effectively fuses the feature learning process and the classifier learning process for better compatibility between both learning processes and obtaining enhanced memory efficiency.
- We employ a ConvGNN-based aggregator that extracts more distinctive features, especially in datasets with significant inter-class similarities.
- Extensive experiments on the CAMELYON16 (Bejnordi et al., 2017), ISIC2018 (Codella et al., 2019), and APTOS2019 (APTOS, 2019), OCTMNIST (Yang, Shi, & Ni, 2021) datasets show that our method achieves superior performance over other class imbalanced methods with performance improvement in all classes.

Differences with previous works. This work is an extended version of our preliminary works (Cong et al., 2022a; Cong, Yang, Liu, Pagnucco, & Song, 2022). The single-stage contrastive learning framework (AdUni) proposed in Cong et al. (2022a) explores the effectiveness of end-to-end self-supervised learning under class imbalance. However, AdUni faces challenges with datasets having high inter-class similarities. On the other hand, DeepGAT (Cong et al., 2022) explores the usefulness of graph neural network feature representations in capturing complex relationships between different receptive fields. This inspires us to integrate their advantages.

In this work, we build AdUniGraph on the basis of AdUni and integrate the strengths of ConvGNN-based aggregators. Subsequently, we implement the following modifications:

- *ConvGNN-based feature representations for contrastive learning.* In conventional contrastive learning frameworks, a multi-layer perceptron (MLP) is used to combine features from pooling layers. However, replacing pooling layers with ConvGNN-based aggregators, as seen in DeepGAT, could lead to over-parameterisation.

Meanwhile, traditional pooling before aggregation may result in sub-optimal features, especially for minority classes. To address this, we introduce a novel approach by replacing MLP and fully connected layers with ConvGNN-based aggregators. These aggregators generate high-level features first, followed by global averaged pooling (GAP). This design reduces the risk of over-parameterisation (*e.g.*, cutting parameter count from 11.5M to 11.2M for a ResNet18 backbone) and improves performance.

- *An improved loss function with a sharp adapter parameter (α).* AdUni's transition function for calculating α is too smooth, unintentionally emphasising the feature learning process and potentially hindering classifier learning. We enhance the loss functions by introducing a more effective function to obtain the adapter, fostering a better integration between feature and classifier learning.
- *Extensive experiments on more diverse datasets.* In addition to the imbalanced datasets (ISIC2018 and APTOS2019) used previously, we conduct extensive experiments on a large histopathology dataset (CAMELYON16) and a retinal dataset (OCTMNIST). AdUniGraph's superior performance on these datasets showcases the effectiveness of ConvGNN-based feature aggregation and emphasises the advantages of unifying feature and classifier learning.

2. Related work

2.1. Class imbalanced classification

Class-imbalanced learning is important for achieving an unbiased model applicable in real-world settings. Accordingly, in the past, data imbalance has been studied extensively in the medical image domain and the general domain as well. This section provides a comprehensive overview of these methods.

Methods in the medical image domain. Data re-sampling methods, including up-sampling the minority classes or down-sampling the majority classes, have been widely used to address imbalanced datasets via constructing a more balanced data distribution (Bokhorst et al., 2018; Dong, Gong, & Zhu, 2018; Galdran et al., 2021; Rana et al., 2023; Reza & Ma, 2018). Furthermore, to alleviate the issue of low variances within minority classes, several methods employ generative networks (Chai et al., 2022; Dai et al., 2022; Park, Liu, Wang, & Zhu, 2019; Rana, Sowmya, Meijering, & Song, 2022a; Rana et al., 2022b) or mix-up strategies (Rana et al., 2022a, 2022b; Zhang, Cisse, Dauphin, & Lopez-Paz, 2018) to generate synthetic samples. For instance, Galdran et al. (2021) utilised mix-up for blending samples from two distributions to create synthetic samples for a balanced training distribution. Similarly, Bal-Mxp (Galdran et al., 2021) adopted mix-up to fuse samples from both regular and balanced distributions to promote effective model training without under-fitting the minority classes. Moreover, Zhao et al. Zhao, Chen, Chen, and Li (2022) perform feature space augmentation on the minority classes by distilling features from the sample-abundant majority classes. In addition, Zhuang et al. Zhuang, Cai, Zhang, Zheng, and Wang (2023) proposed to integrate an attention mechanism to help the model focus on learning the lesion regions of rare diseases. While data re-sampling techniques can be effective, they often ignore the specific characteristics of downstream tasks. This frequently leads to overfitting, potentially resulting in sub-optimal classification performance. Another approach to handle class imbalance is cost-sensitive training, which involves assigning different weights to classes based on their number of samples (Ghorbani et al., 2022; Lin, Wu, Wen, & Qin, 2021; Luo, Xu, Chen, Wong, & Heng, 2022; Pan et al., 2023; Wei et al., 2023; Yoon et al., 2019; Zhang, Tan, et al., 2018; Zhao et al., 2022). Typically, higher weights are assigned to the minority classes, indicating stronger penalties for their mis-classifications. For instance, Sivapuram et al. (2023) introduced a modified version of cross-entropy loss by integrating a modified angular margin alongside the Euclidean margin. Moreover, Ghorbani et al. Ghorbani et al. (2022) devised a Re-weighted Adversarial Graph Convolutional Network (RA-GCN). RA-GCN adjusts the weights assigned to class samples and alters the significance of each sample to help alleviate the class imbalanced issue on graph-structured datasets. Though effective, cost-sensitive trainings typically encounter challenges in determining suitable cost values for various classes and inaccurate cost assignments can heighten the risk of overfitting to the minority class. Additionally, recent studies have explored self-supervised learning (SSL) in class imbalance learning. For example, Marrakchi et al. (2021) introduced a two-stage supervised contrastive learning framework and demonstrated its effectiveness in improving the performance of tasks such as lesion diagnosis and blindness detection in imbalanced settings. Moreover, ProCo (Yang et al., 2022) proposed to generate generating contrastive pairs consisting of category prototype and adversarial proto-instance, and a proto-loss was used to enable single-stage training. Such SSL-based approaches tend to outperform re-sampling or cost-sensitive training at the cost of substantial computational resources. Additionally, the two-stage training mechanism inherent in these approaches is not end-to-end, potentially leading to a mismatch between feature learning and classifier training processes. Despite the demonstrated efficacy of these methods, there is still much space for improvement in the domain of class-imbalanced learning, particularly within the context of medical images.

Methods in the general image domain. Unlike the medical image domain, the issue of class imbalance in the general image domain is well-studied. These methods can be roughly divided into two main categories, the single-model-based methods, and the multi-expert-based methods.

Single-model-based methods perform learning without introducing extra models. These methods can be further divided into three subcategories. The first is the class re-balance method which aims to balance the class contributions to the learning by re-sampling (Zhang, Wu et al., 2021) or loss re-weighting (Deng et al., 2021; Wang, Zhang et al., 2021; Zhu, Niu, Hua, & Zhang, 2022). For example, Soltanzadeh, Feizi-Derakhshi, and Hashemzadeh (2023) recently proposed a under-sampling approach that simultaneously addresses both the imbalanced class distribution and the issue of class overlap. Moreover, model predictions are used to re-weight classes in Focal loss (Lin, Goyal, Girshick, He, & Dollár, 2017), greater weights are assigned to the more challenging tail classes while lower weights are assigned to the easier head classes. On the other hand, LADE (Hong et al., 2021) was proposed to adopt the test label distribution to post-adjust the model prediction. The second category relies on knowledge transfer. Since the majority classes have more samples with higher variances, it is feasible to transfer the knowledge from the data abundant majority classes to help the model learning on the minority classes. To achieve this, distribution calibration methods (Liu, Li, & Sun, 2022) and augmentation techniques (Chu, Bian, Liu, & Ling, 2020) have been proposed. These methods aim to align the distributions of minority and majority classes or augment the minority class samples to improve the model's performance on the underrepresented classes. The third category consists of multi-stage learning approaches based on self-supervised learning (SSL). Since SSL extracts more robust features than the standard supervised learning approach, it is believed to be more capable of handling class imbalance (Liu et al., 2021). Consequently, these methods (Kang et al., 2019; Zhang, Li, Yan, He & and Sun, 2021) initially employ SSL and subsequently finetune a classifier based on the learned features, leveraging the benefits of self-supervised learning for mitigating the effects of class imbalance. For instance, TSC (Li, Cao et al., 2022) increased the uniformity of the feature distribution by mapping the learned features to a set of class-balanced targets.

Current single-model strategies effectively diminish bias towards minority classes; however, they concurrently increase variance across all classes. This increase in variance results in a decrease in accuracy for

majority classes. Thus, various multi-expert-based methods have been introduced to enhance model variance by ensembling the predictions of multiple models or model branches. One such approach is RIDE (Wang, Lian, Miao, Liu, & Yu, 2020), which leverages a multi-expert design to capture complementary knowledge from individual models. Similarly, NCL (Li, Tan, Wan, Lei, & Guo, 2022) improves knowledge transfer between experts through an online distillation module. Additionally, SADE (Zhang, Hooi, Hong & Feng, 2022) adopts a distinct focus for each expert on different data distributions and incorporates a self-supervised test-time aggregation mechanism to fuse the outputs of the experts, and DO (Cong et al., 2024) proposes to effectively improve the interaction between sub-models by dynamically allocating model parameters into sub-groups based their importance to different classes. These techniques aim to diversify model predictions, exploit complementary expertise, and ultimately enhance the overall performance by effectively leveraging multiple experts. While multi-expert models have demonstrated impressive performance, it is important to note that their training requires significant computational resources.

2.2. Supervised contrastive learning

Contrastive learning serves as a proxy task within the realm of self-supervised learning, facilitating the acquisition of a more refined feature space (Chen, Kornblith, Norouzi, & Hinton, 2020; Oord, Li, & Vinyals, 2018). In recent years, contrastive learning has been extensively studied. The proposed methods use an instance discrimination task (Chen, Fan, Girshick & He, 2020; Chen et al., 2020) or clustering-based approaches (Caron, Bojanowski, Joulin, & Douze, 2018) to maximise mutual information between the positive image pairs and discrepancy among the negative pairs. Additionally, supervised contrastive learning has improved performance using label information (Khosla et al., 2020). However, these existing methods often require large memory queues or multi-stage training procedures. In contrast, our work presents a unified single-stage CL framework that eliminates the need for a memory queue and introduces a unified single-stage supervised contrastive learning framework tailored for imbalanced medical datasets.

2.3. Graph representation learning

Inspired by the astonishing progress of CNNs in the deep-learning era, Convolutional Graph Neural Networks (ConvGNNs) generalise convolution to the graph domain by stacking multiple graph convolutional layers to extract high-level graph representations. These methods can be divided into two categories, namely spectral approaches (Bruna, Zaremba, Szlam, & LeCun, 2013; Defferrard, Bresson, & Vandergheynst, 2016; Li, Wang, Zhu, & Huang, 2018), and spatial approaches (Huang, Zhang, Rong, & Huang, 2018; Niepert, Ahmed, & Kutzkov, 2016; Veličković et al., 2017). In particular, both ChebNet (Defferrard et al., 2016) and Graph Convolutional Networks (GCN) (Kipf & Welling, 2016) adopt the spectral method to parameterise the convolution kernel, resulting in significant reductions in both time and space complexity. These methods have introduced the concept of a weight matrix for each node from a spectral perspective. Inspired by these advancements, spatial methods (Huang et al., 2018; Veličković et al., 2017) were subsequently developed, incorporating attention mechanisms and serialisation models to consider the weight of modelling nodes. In recent years, numerous variants of ConvGNNs have emerged. Notably, with the advancement of CNN, researchers have proposed several studies to develop deeper networks that incorporate residual connections and dense connections (Li, Muller, Thabet, & Ghanem, 2019). Furthermore, Han, Wang, Guo, Tang, and Wu (2022) introduce a graph-based representation for images, enabling more flexible feature extraction and aggregation by incorporating a graph structure. However, training ConvGNNs can be memory extensive, since ConvGNNs usually require storing the entire graph and intermediate node features into memory during training. In our approach, we use the high-level feature representations as inputs for ConvGNNs, which tend to have smaller sizes. This strategy is designed to enhance the training efficiency of ConvGNNs.

Recently, Graph Contrastive Learning (GCL) has been studied to establish a new paradigm for learning effective graph representations. Specifically, the current GCL methods can be roughly catergorised into two main categories (Liu, Jin et al., 2022): (1) same-scale (Hafidi, Ghogho, Ciblat, & Swami, 2020; Jovanović, Meng, Faber, & Wattenhofer, 2021; Qiu et al., 2020; Wang, Liu, Han & Shi, 2021; Zhu, Xu et al., 2020) and (2) cross-scale (Hassani & Khasahmadi, 2020; Sun et al., 2021; Wang & Liu, 2021; Zhu, Yang et al., 2020) contrastive learning. Same-scale contrastive learning discriminates features on the same scale (*i.e.,* node to node, or graph to graph). For example, GRACE (Zhu, Xu et al., 2020) adopts node feature masking and edge dropping to obtain two contrastive views. This approach brings similar nodes closer together in two different views of the graph while pushing apart dissimilar ones. Moreover, to increase the robustness of node features, GROC (Jovanović et al., 2021) uses adversarial augmentation on graphs. GCC (Qiu et al., 2020) leverages the MoCo (He, Fan, Wu, Xie, & Girshick, 2020) framework and implements random work as augmentations to generate augmented views of each node. On the other hand, cross-scale discrimination spans various graph topologies. For example, EGI (Zhu, Yang et al., 2020) extracts high-level graph representations by maximising the mutual information between the node embeddings and its surrounding ego-graphs. Additionally, MVGRL (Hassani & Khasahmadi, 2020) uses graph diffusion (Gasteiger, Weißenberger, & Günnemann, 2019) and sub-graph sampling to generate graph views. It then enriches supervision signals by maximising mutual information between node embeddings in one view and the graph-level representation in another view.

Though our work also shares similarities with Graph Contrastive Learning (GCL) in which both use different data views to conduct contrastive learning to find enhanced, low-dimensional representations for graphs. We would like to highlight the fundamental differences between our work and GCL. Firstly, GCLs are typically limited to graph-structured datasets and the effectiveness of GCL methods on imbalanced image datasets is still under-explored. In contrast, our work indeed works on image datasets which do not always contain such structured samples. Secondly, GCLs usually generate data views from graphs, whereas, our work uses image data as different views. Thirdly, the level of imbalance is different. Most current integrated frameworks of GNN with contrastive learning are proposed for the node-level imbalance classification. Whereas, our focus is on image-level imbalance classification. In our context, the 'nodes' refer to features with distinct receptive fields that are extracted from the same input image.

3. Methods

Our proposed approach, AdUniGraph, is a variation of the supervised contrastive learning (SCL) framework. Unlike others, AdUniGraph conducts feature and classifier learning in a single stage and replaces the commonly used pooling operation with a ConvGNN-based feature aggregator. In the following sections, we first briefly describe the conventional two-stage SCL scheme in class imbalanced classification, then we present our modification to conduct SCL in a single stage and describe our ConvGNN-based feature aggregator.

3.1. Conventional two-stage supervised contrastive learning

A standard CNN is trained in a single stage with cross-entropy loss for the classification task. However, when classes are imbalanced, the gradients computed from the cross-entropy loss can be dominated by the majority class, which leads to a biased model. Cui et al. show

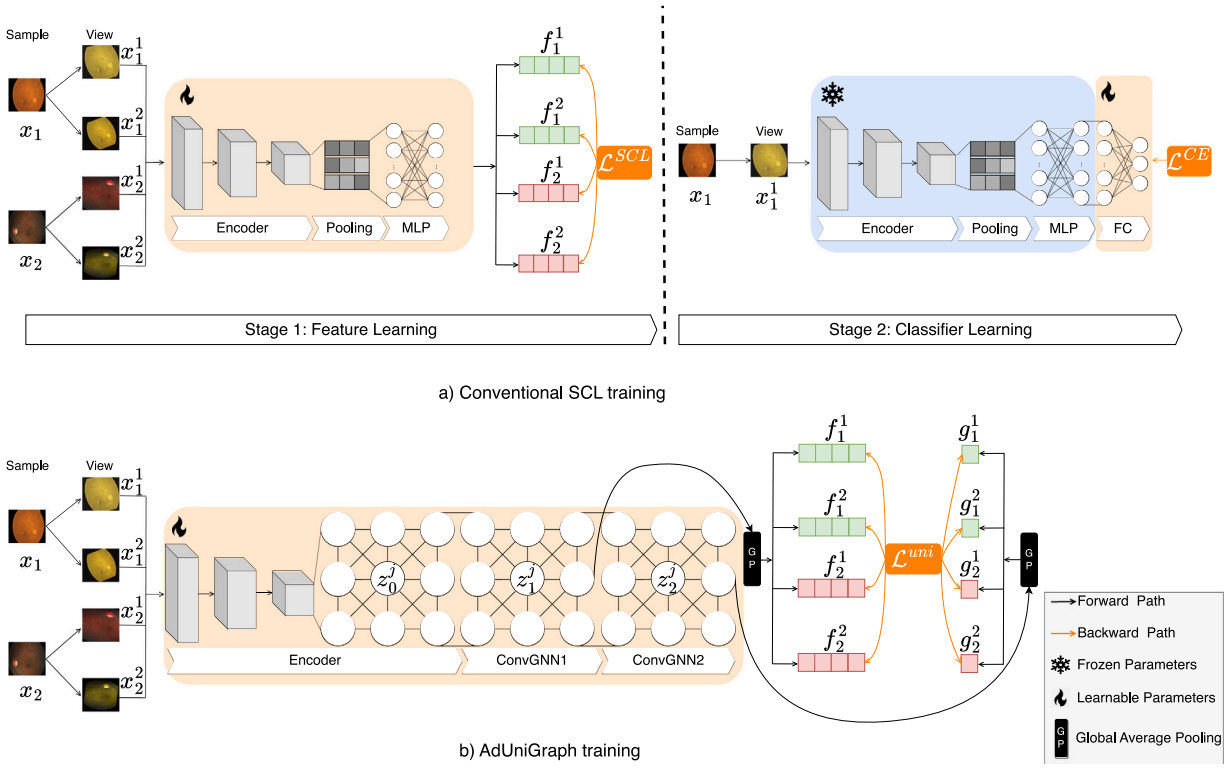


Fig. 3. Comparison between our AdUniGraph and conventional Supervised Contrastive Learning (SCL) training. (a) SCL usually adopts a two-stage training in which the feature learning is conducted at the first stage, then the features are fixed, and classifier learning is conducted at the second stage. In contrast, (b) our AdUniGraph adapts SCL by dynamically adjusting the focus between feature and classifier learning in an end-to-end fashion. Additionally, our novel ConvGNN-based feature aggregator replaces standard pooling and fully connected operations, enhancing intra-class interactions and reducing inter-class interactions.

Algorithm 1: Pseudocode of AdUniGraph training.

Data: Training set containing N image-label pairs, randomly initialised model M_θ with conventional CNN encoder $M_{\theta_{CNN}}$ and ConGNN feature aggregator $M_{\theta_{GNN}}$, learning rate η, # of epochs K.

for $k = 0, \cdots, K-1$ **do**
  for $i = 0, \cdots, N-1$ **do**
    /* Generate two views */
    $x_i^1 \leftarrow augment^1(x_i)$;
    $x_i^2 \leftarrow augmen^2(x_i)$;
    $x_i \leftarrow concatenate(x_i^1, x_i^2)$;
    /* Forward path */
    $z_i \leftarrow M_{\theta_{CNN}}(x_i)$;
    $f_i \leftarrow GP(M_{\theta_{GNN1}}(z_i))$;
    $g_i \leftarrow GP(M_{\theta_{GNN2}}(f_i))$;
    /* Backward path */
    $\theta \leftarrow \theta - \eta \times \nabla_\theta \mathcal{L}_i^{uni}(f_i,\ g_i)$;
  end
end

that robust features extracted using SCL can effectively alleviate this issue (Cui et al., 2021).

SCL methods typically follow a two-stage setting (Fig. 3(a)). Specifically, given a batch of N image-label pairs (x_i, y_i) where $i \in 1 \cdots N$, we generate two distinct views, namely x_i^1 and x_i^2, by applying different augmentation techniques to x_i. Subsequently, x_i^1 and x_i^2 are fed into an encoder network. The resulting outputs pass through a pooling layer and a two-layer multi-layer perception (MLP), yielding the respective feature representations f_i^1 and f_i^2. This ends up having $2N$ features for processing, and we use B to denote this feature set containing $2N$ features. Then the supervised contrastive loss ($\mathcal{L}^{SCL}$) can be calculated using the following equation:

$$\mathcal{L}_i^{SCL} = -\sum_{f_p \in P} \log \frac{\exp(f_i \cdot f_p)/\tau}{\sum_{f_j \in B} \exp(f_i \cdot f_j)/\tau}, \tag{1}$$

where $P = \{f_k \in B : y_k = y_i\}$ and τ is the temperature parameter. Eq. (1) explicitly pulls features with the same label closer and pushes away features from different classes, thus forming denser clusters.

Following the first stage of contrastive feature learning, the weights in the encoder and MLP layers are fixed, and classifier learning is conducted at the second stage. Specifically, a fully-connected classification layer is learned to map the feature f_i obtained from the first stage to the label space (l_i), and the cross-entropy (CE) loss can be used to update the classification layer:

$$\mathcal{L}_i^{CE} = -\sum_{k \in K} y_k \log \rho_{i,k}, \tag{2}$$

where K denotes the total number of classes and $\rho_{i,k} = \sigma(l_i)$ indicates the probability of the image i being classified as class k, and $\sigma(\cdot)$ is the softmax function.

Compared to standard CNN training with CE loss, since SCL involves all instances within a batch, it helps alleviate the gradient imbalances, SCL produces features that exhibit greater robustness to class imbalance. While SCL effectively enhances model performance under dataset imbalance, it is associated with some issues. The first one is a disjoint learning process, where features and classifiers are learned separately with two different targets, making the features learned from the first stage less compatible with the classifier at the second stage. The second issue is the requirement for a substantial batch size with an external memory queue in order to acquire diverse features that contribute to improved performance (Khosla et al., 2020). Moreover, the high inter-class similarity in the classification of medical datasets further poses further difficulties in learning discriminative features.

3.2. Adaptive unified supervised contrastive learning with ConvGNN-based aggregator

In order to tackle the above issues, we present an alternative solution, which incorporates additional supervision from off-the-shelf classifier learning into the feature learning process via an adaptive unified loss function ($\mathcal{L}_{uni}$). Furthermore, we adopt a more effective feature aggregator based on graph convolutional graph neural network (ConvGNN) (Cong et al., 2022) to obtain a more discriminative feature space.

ConvGNN-based aggregator. We argue that the conventional CNN structure's pooling layer cannot adequately capture the intricate relationships between features with varying receptive fields. We therefore adopt a more effective feature aggregator based on graph attention networks (GATs) (Veličković et al., 2017) to overcome this limitation. Our framework is also compatible with other ConvGNN networks, but GAT is chosen for its efficiency; please see the discussion in Section 5.2 for details. As shown in Fig. 3, we use GAT layers for feature aggregation instead of feeding the encoder outputs to a pooling layer. Specifically, we denote $z_0 \in R^{h\times w\times c} = (z_0^1, z_0^2, \cdots, z_0^{w\times h})$ as the feature map extracted by the encoder with width w, height h and expanding over c channels, and different z_0^m, indicates features with different receptive fields. Then, we construct a graph by treating each z_0^m as initial node features with dimension c; we set an edge e_{mn} connecting z_0^m and z_0^n, where $1 \le m, n \le w\times h$, if they are spatial neighbours. Once the graph is constructed, we process it using two consecutive GAT layers where node features are updated by:

$$\mathbf{z}_l^m = \alpha_{m,m}\mathbf{\Theta}\mathbf{z}_{l-1}^m + \sum_{n\in\mathcal{N}(m)} \alpha_{m,n}\mathbf{\Theta}\mathbf{z}_{l-1}^n \tag{3}$$

$$\alpha_{m,n} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top[\mathbf{\Theta}\mathbf{z}_{l-1}^m \parallel \mathbf{\Theta}\mathbf{z}_{l-1}^n]\right)\right)}{\sum_{u\in\mathcal{N}(m)\cup\{m\}}\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top[\mathbf{\Theta}\mathbf{z}_{l-1}^m \parallel \mathbf{\Theta}\mathbf{z}_{l-1}^u]\right)\right)} \tag{4}$$

Here l denotes the index of the lth GAT layer, $\mathcal{N}(m)$ indicates a set of nodes that are spatial neighbours to z^m and $\alpha_{m,n}$ is the attention weight indicating the importance for feature z^n to feature z^m; Θ is a linear transformation; and $\mathbf{a}$ is a single-layer feedforward neural network. GAT helps to capture feature-dependent relationships between z^1, z^2, …, $z^{w\times h}$, therefore enhancing intra-class interactions and weakening inter-class interactions.

Compared to the standard two-stage SCL, we replace the pooling and fully-connected layers with GAT layers. Specifically, in the first GAT layer, we condense the feature dimensions from c to c_{hidden}. Subsequently, we employ global average pooling (GAP) to derive the aggregated feature f_i for input x_i. Lastly, following the final GAT layer, we perform another dimension reduction from c_{hidden} to n_{class}, where n_{class} corresponds to the number of classes. This yields the ultimate output g_i.

Adaptive unified loss. In order to perform feature and classifier learning jointly in an end-to-end framework, the cross-entropy loss function (Eq. (2)) is modified as:

$$\mathcal{L}_i^{CE} = -\log\frac{\exp(g_i\cdot w_{y_i})}{\sum_{k\in K}\exp(g_i\cdot w_{y_k})}, \tag{5}$$

where g_i denotes the output produced by the final ConvGNN layer and w_{y_k} denotes the corresponding label y_k in one-hot format. Subsequently, Eqs. (1) and (5) are combined to obtain the unified loss function ($\mathcal{L}_{uni}$):

$$\mathcal{L}_i^{uni} = -\log\frac{\alpha\cdot\sum_{f_p\in P}\exp(f_i\cdot f_p)/\tau + \exp(g_i\cdot w_{y_i})/\tau}{\sum_{f_j\in B}\exp(f_i\cdot f_j)/\tau + \sum_{k\in K}\exp(g_i\cdot w_{y_k})/\tau}, \tag{6}$$

where α is an adapter parameter that controls the focus of the model between feature and classifier learning and τ is the temperature parameter.

Table 1
Dataset information used in this study.

Name	MaxIns	MinIns	Class count	Type
CAM16	57,942	6,866	2	Breast tumour
APTOS2019	1,805	193	5	Retina
ISIC2018	6,705	115	7	Skin Lesion
OCTMNIST	46,026	7,754	4	Retina

In AdUni (Cong et al., 2022a), α decreases smoothly throughout the training, which slowly transits model training from feature learning at the early stages of training to classifier learning at the later stages. However, such a smooth transition inevitably leads to insufficient classifier learning. Thus, to effectively capture the benefits of both the learning stages and maintain a balance between them, we are inspired by Zhou, Cui, Wei, and Chen (2020) and propose an adaptive approach that reduces α more sharply. In particular, we use the function below to get α_t linked to the current training epoch:

$$\alpha_t = \alpha_{min} + (\alpha_{t-1} - \alpha_{min})\times\delta(t) \tag{7}$$

$$\delta(t) = \frac{1 + cos(\frac{t}{t_{decay}}\pi)}{2} \tag{8}$$

Here, α_{min} is the minimum value of α, t is the current epoch number, $\delta(t)$ is a cosine updating function that moderates the decrease of α and t_{decay} controls the smoothness of $\delta(t)$. Compared to the transition function in AdUni, Eq. (7) adaptively reduces α more sharply which may cause more efficient classifier learning. We provide more detailed discussions in the ablation section.

Training pipeline. Given an input image x_i, we first generate two different views following the standard SCL setting. Both views are concatenated and fed to the encoder network to obtain a feature representation of x_i. Then, the obtained feature goes through two consecutive ConvGNN layers, where the output of the first ConvGNN layer is used as the aggregated feature f_i and the output of the final ConvGNN layer is used as the final output g_i. Both f_i and g_i are fed into $\mathcal{L}_{uni}$ for regularisation.

As can be seen, $\mathcal{L}_{uni}$ eliminates the disjoint learning process, enabling the simultaneous learning of features and classifiers within a single loss function. This approach is particularly advantageous in the context of medical datasets, which often exhibit limited intra-class variance. Simply increasing the batch size and incorporating an additional memory queue may not yield the optimal solution for enhancing feature quality. Instead, $\mathcal{L}_{uni}$ introduces task-specific learning signals to the feature learning process through classifier training. This approach has the potential to obtain more discriminative features without the need for excessively large batch sizes.

4. Experiments

4.1. Dataset details

We trained and evaluated our proposed method on three medical datasets with different image modalities, including CAMELYON16 (histopathology), ISIC2018 (skin lesion), APTOS2019 (retinal disease), and OCTMNIST (retinal disease). More information about each dataset is provided in Table 1 where we show the maximum and minimum number of instances within a class (**MaxIns** and **MinIns**).

CAMELYON16 (Bejnordi et al., 2017) is used for breast histopathology classification. In line with previous studies (Li & Ping, 2018; Shen & Ke, 2020), our approach involves partitioning Whole Slide Imaging (WSI) slides into patches and conducting patch-level classification. The objective is to accurately classify the patches as either normal or tumour, contributing to the overall classification of breast histopathology. It contains 399 WSI slides and each is provided with corresponding detailed pixel annotations. This allows us to perform patch-level classification. Specifically, we first apply the Otsu algorithm (Otsu, 1979)

to filter out the background regions of each slide and then densely extract patches of size 256 × 256 pixels at 5× magnification level. We label the patches containing cancer metastasis as positive (tumour) and other patches as negative (normal). This produces 64,828 patches, where 89.3% (57,942) are negative, and 10.7% (6866) are positive.

ISIC2018 (Codella et al., 2019) is a skin cancer dataset. It contains 10,015 skin session images of size 450 × 600 pixels which are categorised into 7 disease states, including 6705 melanocytic nevus ('nv', 67%), 1113 melanoma ('mel', 11%), 1099 benign keratosis ('bkl', 11%), 514 basal cell carcinoma ('bcc', 5%), 327 actinic keratosis ('akiec', 3%), 142 vascular lesion ('vasc', 1%) and 115 dermatofibroma ('df', 1%).

APTOS2019 (APTOS, 2019) refers to diabetic retinopathy (DR) which is a leading cause of adult blindness. The APTOS2019 dataset contains 3662 images with varying image sizes, which are categorised into 5 classes based on the severity of diabetic retinopathy: 1805 ('0', 49%) no DR, 999 ('1', 27%) moderate, 370 ('2', 10%) mild, 295 ('3', 8%) proliferative, and 193 ('4', 5%) severe DR.

OCTMNIST (Yang et al., 2021) contains 109,309 optical coherence tomography (OCT) images for retinal diseases. There are 4 classes in the dataset, including 37,455 choroidal neovascularisation ('0', 34.3%), 11,598 diabetic macular edema ('1', 10.6%), 8866 drusen ('2', 8.1%), 11,598 normal ('3', 47.0%). The officially provided images are greyscale and of size 28 × 28 pixels.

4.2. Experimental setup

For CAMELYON16, ISIC2018, and APTOS2019, we conduct 5-fold cross-validation and report the mean and standard deviation of the measurements, indicating that for each fold, we randomly select 80% of the samples for training and 20% for testing. In line with previous baseline methods (Cong et al., 2022a; Marrakchi et al., 2021), we resized images to 384 × 384 and applied random affine transformation, horizontal and vertical flip, and colour jittering for data augmentation. For OCTMNIST, following previous works (Chen et al., 2021; Yang et al., 2021), we used the images from the official train and validation set for training, and the images from the test-set are used for evaluation. Additionally, we resized the images to 32 × 32 and applied random horizontal flip for data augmentation. Moreover, we used ResNet18 (He et al., 2016) and GAT (Veličković et al., 2017) as the default encoder network and ConvGNN aggregator as they have shown promising performances in previous studies (Cong et al., 2022a). The model was trained for 400 epochs with a batch size 192 on 4 Nvidia V100 GPUs. The weight parameters were updated using the stochastic gradient descent method (SGD) (Ruder, 2016) with a learning rate of 0.1. The temperature parameter τ (Eqs. (1) and (6)) was set to 0.1 and the initial and minimum values of α were set to 1.0 and 0.1, respectively. We tried different values of t_{decay} (Fig. 10), and the best-performing value of 1000 was set.

For evaluation, we measured the accuracy and macro F1-score over all classes and the per-class accuracy and F1-score, where the macro F1-score is calculated using the averaged precision and recall across all classes. Moreover, inspired by Du and Wu (2023), we also reported the value of geometric mean (G-Mean) which is more sensitive to the lowest recall among all classes. Specifically, the G-Mean of n classes is calculated by $\sqrt[n]{x_1 \cdots x_n}$, where x_i is the accuracy of class i. Additionally, we conducted a Wilcoxon rank sum test at a significance level of 1% to assess whether AdUniGraph achieves statistically significant improvement over the compared approaches. Specifically, we recorded the averaged probability of the correct label for each test sample of 5 runs and our null hypothesis was that the computed probability using our proposed method was less than or equal to that of the other baseline approaches.

Table 2
Classification accuracy, geometric mean (G-Mean), macro F1-score and p-value on the test set of CAMELYON16.

Method	CAMELYON16			
	Acc	G-Mean	Macro F1	p–value
CE(Murphy, 2012)	$0.919_{\pm 1.10e-03}$	$0.917_{\pm 5.68e-03}$	$0.863_{\pm 5.27e-03}$	2.4_{e-7}
Focal Loss(Lin et al., 2017)	$0.920_{\pm 1.21e-03}$	$0.918_{\pm 3.37e-03}$	$0.863_{\pm 5.27e-03}$	2.6_{e-7}
BALMS(Ren et al., 2020)	$0.932_{\pm 3.61e-03}$	$0.932_{\pm 2.90e-03}$	$0.898_{\pm 1.47e-03}$	1.7_{e-4}
LDAM(Cao et al., 2019)	$0.943_{\pm 1.26e-03}$	$0.941_{\pm 1.20e-03}$	$0.912_{\pm 1.27e-03}$	2.3_{e-4}
LADE(Hong et al., 2021)	$0.951_{\pm 1.31e-03}$	$0.951_{\pm 1.30e-03}$	$0.926_{\pm 1.22e-03}$	2.1_{e-4}
Decouple(Kang et al., 2019)	$0.960_{\pm 1.34e-03}$	$0.958_{\pm 1.51e-03}$	$0.938_{\pm 1.21e-03}$	4.2_{e-4}
PaCo(Cui et al., 2021)	$0.970_{\pm 1.20e-03}$	$0.969_{\pm 1.21e-03}$	$0.940_{\pm 1.21e-03}$	1.2_{e-3}
TSC(Li, Cao et al., 2022)	$0.968_{\pm 1.19e-03}$	$0.964_{\pm 1.51e-03}$	$0.936_{\pm 1.19e-03}$	2.8_{e-4}
Bal-Mxp(Galdran et al., 2021)	$0.952_{\pm 1.75e-03}$	$0.952_{\pm 1.13e-03}$	$0.932_{\pm 1.20-03}$	4.2_{e-7}
ProCo(Yang et al., 2022)	$0.983_{\pm 1.21e-03}$	$0.982_{\pm 1.21e-03}$	$0.964_{\pm 1.28e-03}$	2.1_{e-3}
CICL(Marrakchi et al., 2021)	$0.968_{\pm 2.71e-03}$	$0.967_{\pm 2.70e-03}$	$0.942_{\pm 1.27e-03}$	1.3_{e-4}
DeepGAT(Cong et al., 2022a)	$0.948_{\pm 1.58e-03}$	$0.947_{\pm 2.27e-03}$	$0.942_{\pm 2.33e-03}$	5.4_{e-6}
AdUni(Cong et al., 2022a)	$0.972_{\pm 4.21e-03}$	$0.971_{\pm 3.41e-03}$	$0.943_{\pm 2.01e-03}$	8.0_{e-3}
AdUniGraph$_{woConvGNN}$	$0.975_{\pm 3.36e-03}$	$0.975_{\pm 3.24e-03}$	$0.950_{\pm 1.21e-03}$	1.0_{e-3}
AdUniGraph (Ours)	$\mathbf{0.992}_{\pm 1.74e-03}$	$\mathbf{0.992}_{\pm 2.63e-03}$	$\mathbf{0.965}_{\pm 1.10e-03}$	–

Table 3
Classification accuracy, geometric mean (G-Mean), macro F1-score and p-value on the test set of ISIC2018.

Method	ISIC2018			
	Acc	G-Mean	Macro F1	p-value
CE(Murphy, 2012)	$0.841_{\pm 5.37e-03}$	$0.692_{\pm 1.21e-03}$	$0.701_{\pm 1.21e-03}$	3.3_{e-5}
Focal Loss(Lin et al., 2017)	$0.848_{\pm 5.05e-03}$	$0.681_{\pm 1.64e-03}$	$0.722_{\pm 2.20e-03}$	3.8_{e-5}
BALMS(Ren et al., 2020)	$0.862_{\pm 2.28e-03}$	$0.740_{\pm 1.69e-03}$	$0.756_{\pm 1.94e-03}$	1.2_{e-7}
LDAM(Cao et al., 2019)	$0.851_{\pm 2.28e-03}$	$0.702_{\pm 3.24e-04}$	$0.728_{\pm 1.94e-03}$	2.5_{e-7}
LADE(Hong et al., 2021)	$0.856_{\pm 1.28e-03}$	$0.721_{\pm 1.28e-04}$	$0.732_{\pm 2.07e-03}$	3.8_{e-8}
Decouple(Kang et al., 2019)	$0.861_{\pm 2.32e-04}$	$0.710_{\pm 2.55e-03}$	$0.738_{\pm 4.01e-03}$	1.9_{e-6}
PaCo(Cui et al., 2021)	$0.864_{\pm 2.33e-03}$	$0.742_{\pm 4.33e-03}$	$0.768_{\pm 4.49e-03}$	4.9_{e-7}
TSC(Li, Cao et al., 2022)	$0.867_{\pm 3.75e-03}$	$0.748_{\pm 3.29e-03}$	$0.771_{\pm 1.20e-03}$	2.4_{e-7}
Bal-Mxp(Galdran et al., 2021)	$0.852_{\pm 4.63e-03}$	$0.728_{\pm 2.33e-03}$	$0.751_{\pm 4.33e-03}$	3.2_{e-4}
ProCo(Yang et al., 2022)	$0.883_{\pm 3.76e-03}$	$0.766_{\pm 2.32e-03}$	$0.764_{\pm 2.33e-03}$	5.4_{e-7}
CICL(Marrakchi et al., 2021)	$0.866_{\pm 7.39e-03}$	$0.739_{\pm 3.30e-03}$	$0.760_{\pm 9.71e-03}$	2.7_{e-5}
DeepGAT(Cong et al., 2022a)	$0.858_{\pm 2.28e-03}$	$0.732_{\pm 3.08e-03}$	$0.758_{\pm 1.28e-03}$	1.9_{e-4}
AdUni(Cong et al., 2022a)	$0.878_{\pm 4.36e-03}$	$0.784_{\pm 3.34e-03}$	$0.805_{\pm 1.18e-03}$	6.9_{e-3}
AdUniGraph$_{woConvGNN}$	$0.880_{\pm 1.58e-03}$	$0.786_{\pm 3.08e-03}$	$0.806_{\pm 2.33e-03}$	1.2_{e-3}
AdUniGraph (Ours)	$\mathbf{0.883}_{\pm 2.01e-03}$	$0.796_{\pm 3.08e-03}$	$\mathbf{0.808}_{\pm 8.01e-04}$	–

5. Results & discussion

5.1. Comparison to the state-of-the-art methods

We compare our AdUniGraph with previous state-of-the-art (SOTA) class-imbalanced approaches, including those proposed for medical image datasets (CICL (Marrakchi et al., 2021), DeepGAT (Cong et al., 2022), AdUni (Cong et al., 2022a), Bal-Mxp (Galdran et al., 2021) and ProCo (Yang et al., 2022)), and those proposed for general image datasets (Focal Loss (Lin et al., 2017), BALMS (Ren et al., 2020), LDAM (Cao, Wei, Gaidon, Arechiga, & Ma, 2019), LADE (Hong et al., 2021), Decouple (Kang et al., 2019), PaCo (Cui et al., 2021) and TSC (Li, Cao et al., 2022)). Since AdUniGraph is a single-model-based approach, we did not choose to compare it with multi-expert-based methods for fairness. On the other hand, the compared methods are chosen based on their significant impacts on class-imbalanced learning. We aim to cover a wide spectrum of methods for comprehensive comparison. This includes re-sampling approaches (Bal-Mxp), loss re-weighting approaches (Focal Loss, BALMS, LDAM, LADE), and contrastive learning-based approaches (CICL, DeepGAT, AdUni, ProCo, PaCo, and TSC). We have included more contrastive learning-based methods in our comparison due to their close relevance to our work. The reported results are re-implemented using the same backbone network (ResNet18).

The results of CAMELYON16 are shown in Table 2, the results of ISIC2018 are shown in Table 3, the results of APTOS2019 are

Table 4
Classification accuracy, geometric mean (G-Mean), macro F1-score and p-value on the test set of APTOS2019.

Method	APTOS2019			
	Acc	G-Mean	Macro F1	p-value
CE (Murphy, 2012)	$0.813_{\pm 6.09e-03}$	$0.541_{\pm 2.33e-03}$	$0.603_{\pm 7.81e-03}$	2.8_{e-6}
Focal Loss (Lin et al., 2017)	$0.820_{\pm 8.90e-04}$	$0.636_{\pm 1.69e-03}$	$0.635_{\pm 2.80e-03}$	1.5_{e-6}
BALMS (Ren et al., 2020)	$0.826_{\pm 3.46e-03}$	$0.629_{\pm 2.10e-03}$	$0.649_{\pm 3.19e-03}$	2.2_{e-4}
LDAM (Cao et al., 2019)	$0.815_{\pm 2.05e-03}$	$0.641_{\pm 1.66e-03}$	$0.636_{\pm 3.46e-03}$	1.6_{e-5}
LADE (Hong et al., 2021)	$0.825_{\pm 2.32e-03}$	$0.636_{\pm 3.24e-03}$	$0.642_{\pm 2.31e-03}$	1.4_{e-5}
Decouple (Kang et al., 2019)	$0.822_{\pm 2.83e-03}$	$0.619_{\pm 2.82e-03}$	$0.661_{\pm 3.31e-03}$	5.1_{e-7}
PaCo (Cui et al., 2021)	$0.825_{\pm 2.00e-03}$	$0.631_{\pm 4.33e-03}$	$0.680_{\pm 3.31e-03}$	4.4_{e-6}
TSC (Li, Cao et al., 2022)	$0.836_{\pm 1.29e-03}$	$0.642_{\pm 4.33e-03}$	$0.694_{\pm 3.45e-03}$	2.2_{e-3}
Bal-Mxp (Galdran et al., 2021)	$0.821_{\pm 1.37e-03}$	$0.622_{\pm 2.61e-03}$	$0.692_{\pm 2.33e-03}$	1.9_{e-3}
ProCo (Yang et al., 2022)	$0.826_{\pm 1.30e-03}$	$0.633_{\pm 1.30e-03}$	$0.674_{\pm 2.33e-03}$	4.4_{e-7}
CICL (Marrakchi et al., 2021)	$0.828_{\pm 2.05e-03}$	$0.636_{\pm 2.05e-03}$	$0.676_{\pm 1.24e-03}$	4.7_{e-8}
DeepGAT (Cong et al., 2022)	$0.821_{\pm 2.68e-03}$	$0.618_{\pm 2.55e-03}$	$0.640_{\pm 2.32e-03}$	4.9_{e-4}
AdUni (Cong et al., 2022a)	$0.839_{\pm 4.92e-03}$	$0.653_{\pm 2.82e-03}$	$0.695_{\pm 5.51e-03}$	7.1_{e-4}
AdUniGraph$_{woConvGNN}$	$0.843_{\pm 1.58e-03}$	$0.654_{\pm 1.92e-03}$	$0.699_{\pm 2.33e-03}$	1.7_{e-3}
AdUniGraph (Ours)	$\mathbf{0.845}_{\pm 4.21e-03}$	$0.661_{\pm 2.49e-03}$	$\mathbf{0.708}_{\pm 2.21e-03}$	–

Table 5
Classification accuracy, geometric mean (G-Mean), macro F1-score and p-value on the test set of OCTMNIST.

Method	OCTMNIST			
	Acc	G-Mean	Macro F1	p-value
CE (Murphy, 2012)	$0.754_{\pm 3.20e-03}$	$0.742_{\pm 2.50e-03}$	$0.724_{\pm 5.67e-03}$	1.6_{e-7}
Focal Loss (Lin et al., 2017)	$0.752_{\pm 2.00e-03}$	$0.728_{\pm 4.33e-03}$	$0.712_{\pm 3.19e-03}$	2.2_{e-6}
BALMS (Ren et al., 2020)	$0.784_{\pm 1.27e-03}$	$0.772_{\pm 2.33e-03}$	$0.764_{\pm 3.27e-03}$	3.2_{e-4}
LDAM (Cao et al., 2019)	$0.791_{\pm 3.82e-03}$	$0.785_{\pm 5.80e-03}$	$0.772_{\pm 3.84e-03}$	1.7_{e-4}
LADE (Hong et al., 2021)	$0.796_{\pm 1.58e-03}$	$0.793_{\pm 3.19e-03}$	$0.780_{\pm 2.33e-03}$	3.8_{e-6}
Decouple (Kang et al., 2019)	$0.828_{\pm 1.20e-03}$	$0.816_{\pm 1.22e-03}$	$0.814_{\pm 2.21e-03}$	1.9_{e-6}
PaCo (Cui et al., 2021)	$0.848_{\pm 1.33e-03}$	$0.835_{\pm 4.35e-03}$	$0.832_{\pm 1.33e-03}$	1.2_{e-4}
TSC (Li, Cao et al., 2022)	$0.843_{\pm 1.19e-03}$	$0.828_{\pm 1.20e-03}$	$0.832_{\pm 1.21e-03}$	1.5_{e-4}
Bal-Mxp (Galdran et al., 2021)	$0.783_{\pm 1.38e-03}$	$0.774_{\pm 2.64e-03}$	$0.769_{\pm 2.32e-03}$	3.7_{e-5}
ProCo (Yang et al., 2022)	$0.856_{\pm 1.33e-03}$	$0.843_{\pm 3.63e-03}$	$0.846_{\pm 1.37e-03}$	1.4_{e-4}
CICL (Marrakchi et al., 2021)	$0.843_{\pm 1.37e-03}$	$0.832_{\pm 4.33e-03}$	$0.835_{\pm 1.24e-03}$	5.4_{e-4}
DeepGAT (Cong et al., 2022)	$0.775_{\pm 3.20e-03}$	$0.762_{\pm 2.54e-03}$	$0.749_{\pm 4.33e-03}$	3.7_{e-6}
AdUni (Cong et al., 2022a)	$0.862_{\pm 1.20e-03}$	$0.862_{\pm 1.20e-03}$	$0.861_{\pm 2.02e-03}$	2.4_{e-3}
AdUniGraph$_{woConvGNN}$	$0.863_{\pm 4.33e-03}$	$0.864_{\pm 4.33e-03}$	$0.863_{\pm 3.13e-03}$	2.1_{e-3}
AdUniGraph (Ours)	$\mathbf{0.873}_{\pm 1.27e-03}$	$\mathbf{0.868}_{\pm 4.33e-03}$	$\mathbf{0.872}_{\pm 1.27e-03}$	–

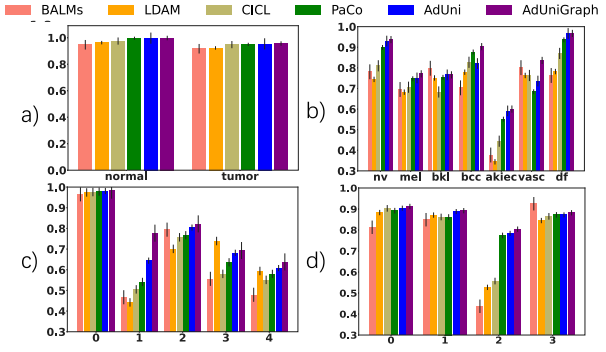


Fig. 4. Accuracy for each class on the test set on (a) CAMELYON16, (b) ISIC2018, (c) APTOS2019, and (d) OCTMNIST using different class-imbalanced studies.

shown in Table 4 and the results of OCTMNIST are shown in Table 5. As can be seen, our proposed method achieves state-of-the-art performance on all four datasets with an overall accuracy of 99.2%, 88.3%, 84.5%, 87.3% on CAMELYON16, ISIC2018, APTOS2019 and OCTMNIST, respectively. Meanwhile, it has the best macro-F1 score on four datasets, with 96.5%, 80.8%, 70.3% and 87.2%, respectively. In general, contrastive learning-based approaches demonstrate better performance than re-sampling or loss re-weighting approaches. This illustrates the usefulness of CL in increasing the robustness of features against class imbalance. Furthermore, our proposed method also has

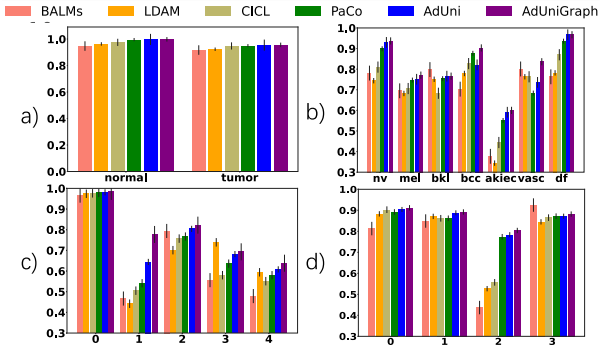


Fig. 5. F1-score for each class on the test set on (a) CAMELYON16, (b) ISIC2018, (c) APTOS2019, and (d) OCTMNIST using different class-imbalanced studies.

the best G-Mean over other comparing baselines. For example, its G-Mean values on CAMELYON16, ISIC2018, APTOS2019, and OCTMNIST are approximately 99.2%, 79.6%, 66.1%, 86.8%, respectively. On average, these scores are approximately 1.5% higher than those achieved by the previous SOTA method, AdUni. Since G-Mean is more sensitive to small values, the improvement in G-Mean indicates that our proposed method successfully improves the lowest recall across all categories. To further illustrate this statement, we show the accuracy and F1-score on every individual class for all four datasets in Figs. 4 and 5. The results clearly demonstrate that AdUniGraph not only exhibits the highest overall performance but also yields improvements in accuracies and F1-scores across all classes when compared to other methods.

Unlike previous SOTA methods with contrastive learning (PaCo (Cui et al., 2021)), TSC (Li, Cao et al., 2022), ProCo (Yang et al., 2022) and (CICL (Marrakchi et al., 2021)), both AdUni (Cong et al., 2022a) and AdUniGraph require only a single-stage of training, eliminate the need for a memory queue, and obtain better performance. This underscores the benefits of introducing learning signals from classifier training into the feature learning process, as opposed to relying solely on incorporating an external memory queue. Moreover, compared with our previous work AdUni (Cong et al., 2022a), using the new transition function for α (Eq. (7)) helps enhance the overall performance and adopting the graph convolutional-based aggregator from our previous work (Cong et al., 2022) further improves the model performance, which shows the importance of using a better feature aggregator in class imbalance classification. Additionally, the obtained p-values are smaller than 0.01, which indicates that AdUniGraph has statistically significantly better performance than the compared methods.

5.2. Ablation studies & discussions

We perform ablation studies to examine the effectiveness of employing single-stage contrastive learning, using the new transition function for α and using the graph convolution-based aggregator. The compared methods are as follows: (1) **baseline** method, which is the standard CNN trained in a single stage with a cross-entropy loss; (2) **two-stage CL**, which is the standard contrastive learning that uses supervised contrastive loss function for feature learning and cross-entropy for classifier learning; (3) **two-stage CL and ConvGNN**, which is standard contrastive learning with supervised contrastive loss function for feature learning and cross-entropy for classifier learning, and use ConvGNN as feature aggregator (Cong et al., 2022); (4) **unified CL**, which replaces supervised contrastive loss function and cross-entropy with unified loss function ($\mathcal{L}^{uni}$) and uses a smooth transition function for α (Cong et al., 2022a); (5) **unified CL with sharp** α, which uses the same structure and loss as unified CL, but with a sharp transition function (Eq. (7)) for α; and, (6) **unified CL with sharp** α **and ConvGNN** (AdUniGraph), which integrates both the sharp transition

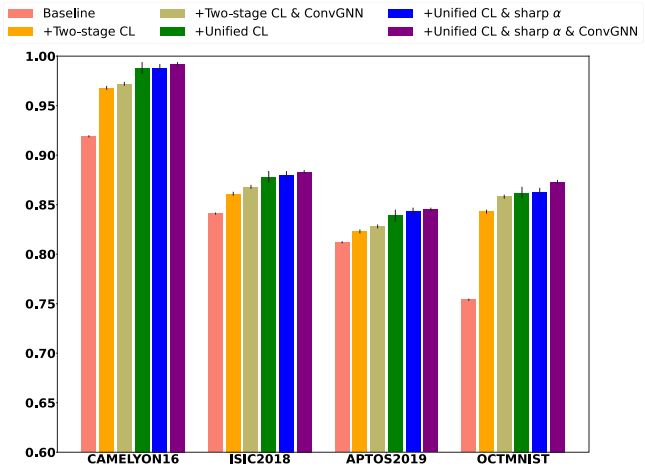


Fig. 6. Classification accuracy comparison with different training methods and model structures on CAMELYON16, ISIC2018, APTOS2019 and OCTMNIST datasets.

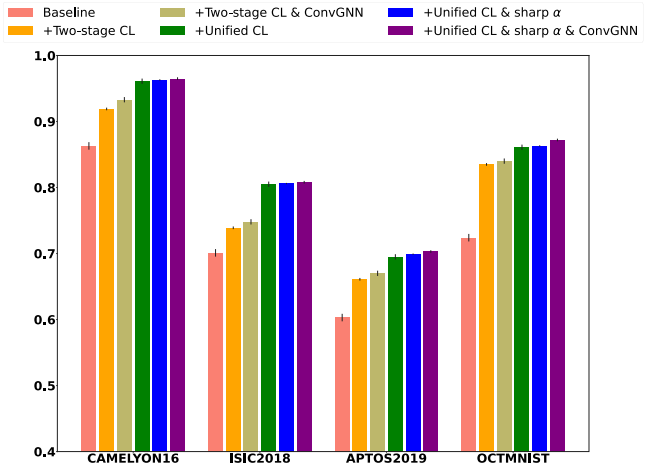


Fig. 7. F1-score comparison with different training methods and model structures on CAMELYON16, ISIC2018, APTOS2019 and OCTMNIST datasets.

function for α and the ConvGNN-based feature aggregator. We note that the model architecture, learning scheduler, optimiser, and other initial hyperparameter settings were the same for all the experiments. Additionally, we analyse the impact of several crucial components, such as using diverse values of t_{decay}, using different encoder backbones, and employing various convolutional graph neural networks for feature aggregation.

Ablation on all components. We perform a comprehensive ablation study on our proposed method, and the results are illustrated in Figs. 6 and 7. AdUniGraph, which includes unified CL, sharp transition function for α, and ConvGNN-based feature aggregator, achieves the best performance in the ablation study. In contrast, the Baseline method exhibits biased learning due to the imbalanced dataset, resulting in the poorest performance. Significant performance improvements can be observed by adopting either the conventional two-stage supervised contrastive learning (**two-stage CL**) or unified single-stage supervised contrastive learning (**Unified CL**). This observation aligns with the findings of Liu et al. (2021), suggesting that self-supervised learning generates more robust features in the presence of class imbalance. Furthermore, Unified CL notably demonstrates better performance than the conventional two-stage CL, highlighting the effectiveness of unifying feature and classifier learning.

We also report the classification performances on different datasets with different training parameters in Fig. 8, including learning rate,

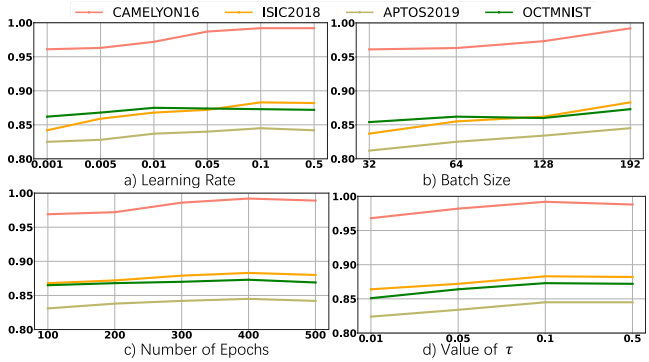


Fig. 8. Classification scores on different datasets with different training parameters.

batch size, number of epochs, and the value of τ. We notice that a larger learning rate is preferable for most datasets except for OCTMNIST where the best performance is obtained when the learning rate equals 0.01. However, increasing it to 0.1 only brings a marginal drop in classification accuracy. Consequently, we decided to use a learning rate of 0.01 for all four datasets. Moreover, larger batch sizes tend to improve performance, but they also demand more training resources. As a compromise, we settled on a batch size of 192. Furthermore, we found that longer training periods do not necessarily lead to better outcomes. For example, employing 500 epochs for training resulted in worse performance across the four datasets. Consequently, we chose to train for 400 epochs. Finally, we show the performance change versus the temperature parameter τ for (Eqs. (1) and (6)). Smaller τ values led to decreased performance, likely because they overly penalise the most similar negative samples while underemphasising other negative samples. Consequently, we determined that $\tau = 0.1$ yields the best performance.

We compare the change of the adapter parameter α during in AdUni (Cong et al., 2022a) and AdUniGraph, as visualised in Fig. 10. In AdUni, we notice a consistent and gradual decrease in α throughout training. This suggests that α remains relatively high for most of the training process, indicating a high contribution of feature learning to the overall learning. While this approach works well for AdUni, we propose that it might lead to insufficient learning of the classifier, which is crucial for capturing task-specific information. To address this concern, we introduce a modification to the way α changes during training. This change, shown in the right part of Fig. 10, leads to a more rapid decrease in α. As a result, more training time is dedicated to refining the classifier. The results we present in Figs. 6 and 7 demonstrate the positive effects of this modification.

We further highlight the usefulness of adopting the proposed ConvGNN aggregator. As shown in Figs. 6 and 7, further performance improvement is observed using the proposed ConvGNN aggregator. To further illustrate its usefulness, we visualise the feature space using TSNE in Fig. 11 and the Grad-CAM (Selvaraju et al., 2017) in Fig. 9 of AdUniGraph with and without the ConvGNN aggregator. It is observed that using the ConvGNN provides more compact intra-class feature clustering and enlarges the distance between inter-class clusters. In addition, we notice that replacing the standard pooling-based aggregator with ConvGNN aggregator helps the model to focus on more important regions, thus increasing the classification performance.

Influence of different t_{decay}. Note that, except for the hyperparameters for training (*e.g.*, learning rate, number of epochs, and batch size), our work only introduces one extra parameter α. More specifically, as shown in Eq. (8), the value of α purely depends on t_{decay}. Thus, in this section we explored the impact of different values of t_{decay} in Fig. 10. It is observed that increasing the value of t_{decay} leads to a smoother transition α, effectively assigning a higher importance to the feature learning stage. However, excessively emphasising the feature

Table 6
Experiment results using various values of t_{decay} on different datasets.

t_{decay}	CAMELYON16		ISIC2018		APTOS2019		OCTMNIST	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1
400	$0.988_{\pm 5.26e-03}$	$0.957_{\pm 3.95e-03}$	$0.873_{\pm 2.64e-03}$	$0.793_{\pm 3.69e-03}$	$0.838_{\pm 4.74e-03}$	$0.691_{\pm 1.63e-03}$	$0.856_{\pm 1.28e-03}$	$0.858_{\pm 2.64e-03}$
600	$0.990_{\pm 1.75e-03}$	$0.958_{\pm 1.89e-03}$	$0.879_{\pm 3.03e-03}$	$0.795_{\pm 1.02e-03}$	$0.840_{\pm 8.01e-04}$	$0.701_{\pm 1.17e-03}$	$0.858_{\pm 4.33e-03}$	$0.859_{\pm 1.63e-03}$
800	$0.992_{\pm 3.03e-03}$	$0.961_{\pm 1.54e-03}$	$0.883_{\pm 2.28e-03}$	$0.802_{\pm 1.21e-03}$	$0.841_{\pm 2.33e-03}$	$0.704_{\pm 2.69e-03}$	$0.864_{\pm 4.33e-03}$	$0.861_{\pm 4.33e-03}$
1000	$0.992_{\pm 1.74e-03}$	$\mathbf{0.965}_{\pm 1.10e-03}$	$\mathbf{0.883}_{\pm 2.01e-03}$	$\mathbf{0.808}_{\pm 8.01e-04}$	$\mathbf{0.845}_{\pm 4.21e-03}$	$\mathbf{0.708}_{\pm 2.21e-03}$	$\mathbf{0.873}_{\pm 1.27e-03}$	$\mathbf{0.872}_{\pm 1.27e-03}$
1200	$\mathbf{0.994}_{\pm 3.03e-03}$	$0.964_{\pm 2.20e-03}$	$0.881_{\pm 3.78e-03}$	$0.799_{\pm 2.37e-03}$	$0.843_{\pm 6.34e-03}$	$0.700_{\pm 1.06e-03}$	$0.873_{\pm 4.33e-03}$	$0.871_{\pm 1.27e-03}$

Table 7
Experiment results using various types of convolutional neural networks as encoders on four datasets.

Encoder type	CAMELYON16		ISIC2018		APTOS2019		OCTMNIST	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1
ResNet18	$\mathbf{0.992}_{\pm 1.74e-03}$	$\mathbf{0.965}_{\pm 1.10e-03}$	$\mathbf{0.883}_{\pm 2.01e-03}$	$\mathbf{0.808}_{\pm 8.01e-04}$	$\mathbf{0.845}_{\pm 4.21e-03}$	$\mathbf{0.708}_{\pm 2.21e-03}$	$\mathbf{0.873}_{\pm 1.27e-03}$	$\mathbf{0.872}_{\pm 1.27e-03}$
ResNet50	$0.986_{\pm 2.32e-03}$	$0.958_{\pm 6.52e-03}$	$0.887_{\pm 2.28e-03}$	$0.790_{\pm 1.21e-03}$	$0.837_{\pm 2.68e-03}$	$0.682_{\pm 2.21e-03}$	$0.870_{\pm 4.33e-03}$	$0.871_{\pm 2.32e-03}$
DenseNet121	$0.981_{\pm 2.37e-03}$	$0.960_{\pm 2.25e-03}$	$0.869_{\pm 1.96e-03}$	$0.758_{\pm 2.51e-03}$	$0.828_{\pm 1.35e-03}$	$0.679_{\pm 4.25e-03}$	$0.865_{\pm 2.24e-03}$	$0.858_{\pm 2.21e-03}$
DenseNet161	$0.975_{\pm 3.13e-03}$	$0.948_{\pm 4.25e-03}$	$0.862_{\pm 1.26e-03}$	$0.752_{\pm 2.57e-03}$	$0.827_{\pm 7.24e-03}$	$0.676_{\pm 3.54e-03}$	$0.862_{\pm 4.33e-03}$	$0.858_{\pm 2.21e-03}$

Table 8
Experiment results using various types of convolutional graph neural networks as feature aggregators on four datasets.

GNN type	CAMELYON16		ISIC2018		APTOS2019		OCTMNIST	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1
GCNConv	$0.973_{\pm 3.37e-03}$	$0.945_{\pm 4.49e-03}$	$0.872_{\pm 2.06e-03}$	$0.773_{\pm 2.46-03}$	$0.812_{\pm 1.37e-03}$	$0.664_{\pm 2.28-03}$	$0.858_{\pm 1.37e-03}$	$0.856_{\pm 4.33-03}$
ResConv	$0.981_{\pm 2.28e-03}$	$0.952_{\pm 2.37e-03}$	$0.880_{\pm 1.55e-03}$	$0.786_{\pm 1.81-03}$	$0.826_{\pm 1.28e-03}$	$0.678_{\pm 2.01-03}$	$0.864_{\pm 2.62e-03}$	$0.863_{\pm 1.28-03}$
GATConv	$0.992_{\pm 1.74e-03}$	$0.965_{\pm 1.10e-03}$	$0.883_{\pm 2.01e-03}$	$0.808_{\pm 8.01e-04}$	$0.845_{\pm 4.21e-03}$	$0.708_{\pm 2.21e-03}$	$\mathbf{0.873}_{\pm 1.27e-03}$	$0.872_{\pm 1.27e-03}$
GATv2Conv	$\mathbf{0.993}_{\pm 2.64e-03}$	$\mathbf{0.970}_{\pm 1.08e-03}$	$0.887_{\pm 2.20e-03}$	$\mathbf{0.812}_{\pm 2.01e-03}$	$\mathbf{0.848}_{\pm 4.92e-03}$	$\mathbf{0.710}_{\pm 1.20e-03}$	$0.873_{\pm 4.33e-03}$	$\mathbf{0.873}_{\pm 2.24-03}$

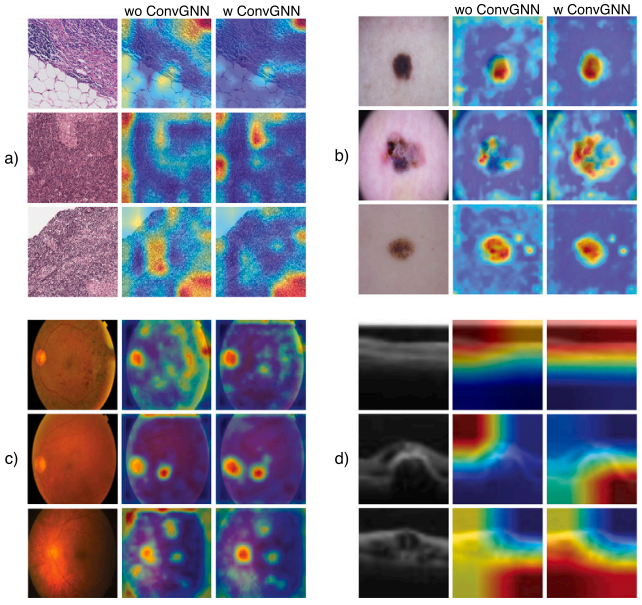


Fig. 9. The grad-CAM visualisation with (w) and without (wo) using the proposed ConvGNN aggregator on (a) CAMELYON16, (b) ISIC2018, (c) APTOS2019 and (d) OCTMNIST datasets.

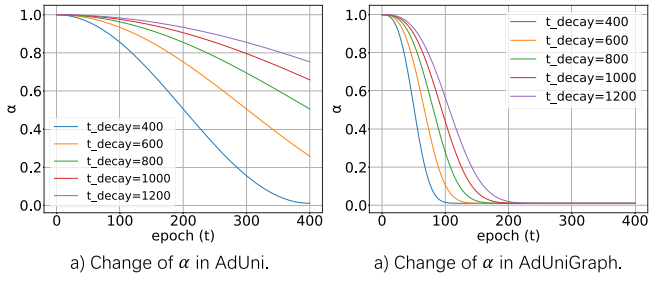


Fig. 10. Plots of α in AdUni (Top) and AdUniGraph (Bottom) with different values of t_{decay}. The X-axis denotes the epoch number.

learning stage (*i.e.,* enlarging t_{decay}) could lead to insufficient task-specific classifier learning, and excessively emphasising the classifier learning stage (*i.e.,* reducing t_{decay}) could lead to insufficient feature learning. According to the results shown in Fig. 10 and Table 6, since the overall best performance can be achieved with $t_{decay} = 1000$, we use this value as the default value of t_{decay}.

Influence of different encoder backbones. We explored the performance of various CNNs as encoder networks in our framework. As shown in Table 7, our framework is architecture agnostic, delivering commendable performance across different encoder networks. However, it is intriguing to note that employing deeper neural networks leads to a decline in performance, which could be due to the fact that medical datasets typically are of low scales and lack intricate visual features, potentially resulting in overfitting when using deeper networks.

Influence of different GCN aggregator. We conducted experiments utilising various types of convolutional graph neural networks as feature aggregators. Accordingly, we tried the graph convolutional operator (GCNConv) (Kipf & Welling, 2016), the residual gated graph convolutional (ResConv) operator (Bresson & Laurent, 2017), the graph attention operator (GATConv) (Veličković et al., 2017) and GATv2Conv (Brody, Alon, & Yahav, 2021). The results are shown in Table 8. Our framework works well with different choices of convolutional graph neural networks as feature aggregators, and more advanced convolutional graph neural networks steadily improve performance.

Limitation and Future works. While AdUniGraph effectively addresses class imbalance issues in medical image datasets, it does not take into account the valuable prior knowledge inherent in these medical datasets. For instance, factors such as complex shape orientation and H&E staining often play a crucial role in distinguishing tumour regions in histopathology images. Similarly, factors like blurriness and lighting conditions significantly impact retinal image classification (Qayyum, Sultani, Shamshad, Tufail, & Qadir, 2022).

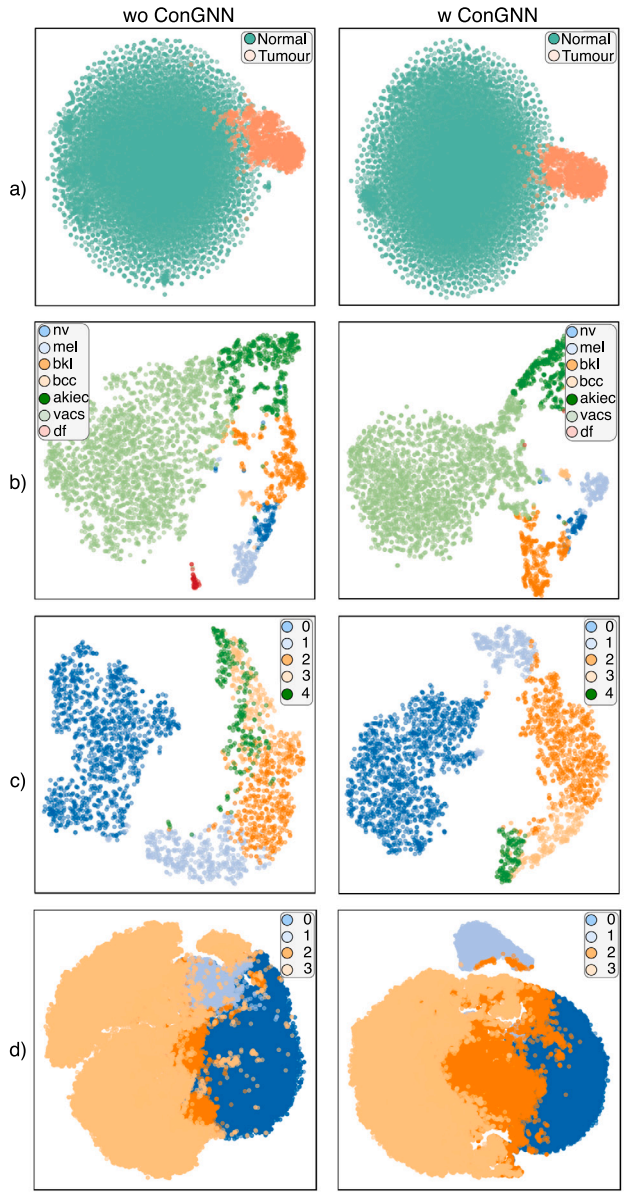


Fig. 11. We show the feature visualisation with (w) and without (wo) using the proposed ConvGNN aggregator on (a) CAMELYON16, (b) ISIC2018, (c) APTOS2019 and (d) OCTMNIST datasets.

Taking inspiration from recent work by Zhang, Ruan, Li, and Zhang (2023), Zhang, Zhao et al. (2022), which incorporates the physics laws governing muscle forces and joint kinematics into deep learning model training through a customised loss function, our future research will consider extracting meaningful stain colour features or performing atmospheric light estimation. We plan to integrate these features explicitly into the training process by designing advanced loss functions, with the ultimate goal of further improving the performance of AdUniGraph.

6. Conclusion

This paper addresses the challenge of imbalanced data distribution in medical image classification. Our objective is to develop an effective solution and to achieve this, we propose a novel end-to-end supervised contrastive training framework. Unlike traditional approaches that handle feature learning and classifier learning separately, our framework combines both stages by employing an adaptively unified loss function. Additionally, we introduce a novel feature aggregator based on convolutional graph neural networks, which replaces the conventional pooling layers and further enhances performance. Experimental studies conducted on CAMELYON16, ISIC2018, APTOS2019 and OCTMNIST datasets demonstrate that our proposed AdUniGraph framework achieves substantial performance improvements across all classes.

CRediT authorship contribution statement

Cong Cong: Conceptualization, Methodology, Software, Validation, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Sidong Liu:** Data curation, Resources, Project administration, Supervision, Writing – review & editing. **Priyanka Rana:** Writing – review & editing, Investigation. **Maurice Pagnucco:** Writing – review & editing, Supervision. **Antonio Di Ieva:** Writing – review & editing, Supervision. **Shlomo Berkovsky:** Writing – review & editing, Supervision. **Yang Song:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

APTOS (2019). APTOS 2019 blindness detection. https://www.kaggle.com/c/aptos2019-blindness-detection/data.

Bejnordi, B. E., Veta, M., Van Diest, P. J., Van Ginneken, B., Karssemeijer, N., Litjens, G., et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *The Journal of the American Medical Association*, *318*(22), 2199–2210.

Bokhorst, J. M., Pinckaers, H., van Zwam, P., Nagtegaal, I., van der Laak, J., & Ciompi, F. (2018). Learning from sparsely annotated data for semantic segmentation in histopathology images. In *Medical imaging with deep learning*.

Bresson, X., & Laurent, T. (2017). Residual gated graph convnets. arXiv preprint arXiv:1711.07553.

Brody, S., Alon, U., & Yahav, E. (2021). How attentive are graph attention networks? arXiv preprint arXiv:2105.14491.

Bruna, J., Zaremba, W., Szlam, A., & LeCun, Y. (2013). Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203.

Cao, K., Wei, C., Gaidon, A., Arechiga, N., & Ma, T. (2019). Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in Neural Information Processing Systems*, *32*.

Caron, M., Bojanowski, P., Joulin, A., & Douze, M. (2018). Deep clustering for unsupervised learning of visual features. In *Proceedings of the European conference on computer vision* (pp. 132–149).

Chai, L., Wang, Z., Chen, J., Zhang, G., Alsaadi, F. E., Alsaadi, F. E., et al. (2022). Synthetic augmentation for semantic segmentation of class imbalanced biomedical images: A data pair generative adversarial network approach. *Computers in Biology and Medicine*, *150*, Article 105985.

Chen, X., Fan, H., Girshick, R., & He, K. (2020). Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297.

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597–1607). PMLR.

Chen, K., Mao, Y., Lu, H., Zeng, C., Wang, R., & Zheng, W. S. (2021). Alleviating data imbalance issue with perturbed input during inference. In *Medical image computing and computer assisted intervention–MICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, proceedings, Part V 24* (pp. 407–417). Springer.

Chu, P., Bian, X., Liu, S., & Ling, H. (2020). Feature space augmentation for long-tailed data. In *European conference on computer vision* (pp. 694–710). Springer.

Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., et al. (2019). Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). arXiv preprint arXiv:1902.03368.

Cong, C., Xuan, S., Liu, S., Zhang, S., Pagnucco, M., & Song, Y. (2024). Decoupled optimisation for long-tailed visual recognition. In *Proceedings of the AAAI conference on artificial intelligence, vol. 38, no. 2* (pp. 1380–1388).

Cong, C., Yang, Y., Liu, S., Pagnucco, M., Di Ieva, A., Berkovsky, S., et al. (2022a). Adaptive unified contrastive learning for imbalanced classification. In *Machine learning in medical imaging* (pp. 348–357). Springer.

Cong, C., Yang, Y., Liu, S., Pagnucco, M., & Song, Y. (2022). Imbalanced histopathology image classification using deep feature graph attention network. In *2022 IEEE 19th international symposium on biomedical imaging* (pp. 1–4). IEEE.

Cui, J., Zhong, Z., Liu, S., Yu, B., & Jia, J. (2021). Parametric contrastive learning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 715–724).

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.

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in Neural Information Processing Systems*, *29*.

Deng, Z., Liu, H., Wang, Y., Wang, C., Yu, Z., & Sun, X. (2021). Pml: Progressive margin loss for long-tailed age classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10503–10512).

Dong, Q., Gong, S., & Zhu, X. (2018). Imbalanced deep learning by minority class incremental rectification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(6), 1367–1381.

Du, Y., & Wu, J. (2023). No one left behind: Improving the worst categories in long-tailed learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 15804–15813).

Galdran, A., Carneiro, G., & González Ballester, M. A. (2021). Balanced-mixup for highly imbalanced medical image classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 323–333). Springer.

Gao, L., Zhang, L., Liu, C., & Wu, S. (2020). Handling imbalanced medical image data: A deep-learning-based one-class classification approach. *Artificial Intelligence in Medicine*, *108*, Article 101935.

Gasteiger, J., Weißenberger, S., & Günnemann, S. (2019). Diffusion improves graph learning. In *Conference on neural information processing systems*.

Ghorbani, M., Kazi, A., Baghshah, M. S., Rabiee, H. R., & Navab, N. (2022). RA-GCN: Graph convolutional network for disease prediction problems with imbalanced data. *Medical Image Analysis*, *75*, Article 102272.

Hafidi, H., Ghogho, M., Ciblat, P., & Swami, A. (2020). GraphCL: Contrastive self-supervised learning of graph representations. ArXiv abs/2007.08025. URL https://api.semanticscholar.org/CorpusID:220546101.

Han, K., Wang, Y., Guo, J., Tang, Y., & Wu, E. (2022). Vision gnn: An image is worth graph of nodes. arXiv preprint arXiv:2206.00272.

Hassani, K., & Khasahmadi, A. H. (2020). Contrastive multi-view representation learning on graphs. In *Proceedings of international conference on machine learning* (pp. 3451–3461).

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *2020 IEEE/CVF conference on computer vision and pattern recognition* (pp. 9726–9735). http://dx.doi.org/10.1109/CVPR42600.2020.00975.

He, K., Zhang, X., et al. (2016). Deep residual learning for image recognition. In *IEEE/CVF conference on computer vision and pattern recognition* (pp. 770–778).

Hong, Y., Han, S., Choi, K., Seo, S., Kim, B., & Chang, B. (2021). Disentangling label distribution for long-tailed visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6626–6636).

Huang, W., Zhang, T., Rong, Y., & Huang, J. (2018). Adaptive sampling towards fast graph representation learning. *Advances in Neural Information Processing Systems*, *31*.

Jovanović, N., Meng, Z., Faber, L., & Wattenhofer, R. (2021). Towards robust graph contrastive learning. arXiv:2102.13085.

Kang, B., Xie, S., Rohrbach, M., Yan, Z., Gordo, A., Feng, J., et al. (2019). Decoupling representation and classifier for long-tailed recognition. arXiv preprint arXiv:1910.09217.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., et al. (2020). Supervised contrastive learning. *Advances in Neural Information Processing Systems*, *33*, 18661–18673.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Li, T., Cao, P., Yuan, Y., Fan, L., Yang, Y., Feris, R. S., et al. (2022). Targeted supervised contrastive learning for long-tailed recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6918–6928).

Li, G., Muller, M., Thabet, A., & Ghanem, B. (2019). Deepgcns: Can gcns go as deep as cnns? In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9267–9276).

Li, Y., & Ping, W. (2018). Cancer metastasis detection with neural conditional random field. arXiv preprint arXiv:1806.07064.

Li, J., Tan, Z., Wan, J., Lei, Z., & Guo, G. (2022). Nested collaborative learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6949–6958).

Li, R., Wang, S., Zhu, F., & Huang, J. (2018). Adaptive graph convolutional neural networks. In *Proceedings of the AAAI conference on artificial intelligence, vol. 32, no. 1*.

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980–2988).

Lin, C., Wu, H., Wen, Z., & Qin, J. (2021). Automated malaria cells detection from blood smears under severe class imbalance via importance-aware balanced group softmax. In *International conference on medical image computing and computer-assisted intervention* (pp. 455–465). Springer.

Liu, H., HaoChen, J. Z., Gaidon, A., & Ma, T. (2021). Self-supervised learning is more robust to dataset imbalance. arXiv preprint arXiv:2110.05025.

Liu, Y., Jin, M., Pan, S., Zhou, C., Zheng, Y., Xia, F., et al. (2022). Graph self-supervised learning: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *35*(6), 5879–5900.

Liu, J., Li, W., & Sun, Y. (2022). Memory-based jitter: Improving visual recognition on long-tailed data with diversity in memory. In *Proceedings of the AAAI conference on artificial intelligence, vol. 36* (pp. 1720–1728).

Luo, L., Xu, D., Chen, H., Wong, T. T., & Heng, P. A. (2022). Pseudo bias-balanced learning for debiased chest X-Ray classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 621–631). Springer.

Marrakchi, Y., Makansi, O., & Brox, T. (2021). Fighting class imbalance with contrastive learning. In *International conference on medical image computing and computer-assisted intervention* (pp. 466–476). Springer.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Niepert, M., Ahmed, M., & Kutzkov, K. (2016). Learning convolutional neural networks for graphs. In *International conference on machine learning* (pp. 2014–2023). PMLR.

Oord, A. v. d., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66.

Pan, X., Cheng, J., Hou, F., Lan, R., Lu, C., Li, L., et al. (2023). SMILE: Cost-sensitive multi-task learning for nuclear segmentation and classification with imbalanced annotations. *Medical Image Analysis*, Article 102867.

Park, T., Liu, M. Y., Wang, T. C., & Zhu, J. Y. (2019). Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2337–2346).

Qayyum, A., Sultani, W., Shamshad, F., Tufail, R., & Qadir, J. (2022). Single-shot retinal image enhancement using untrained and pretrained neural networks priors integrated with analytical image priors. *Computers in Biology and Medicine*, *148*, Article 105879.

Qiu, J., Chen, Q., Dong, Y., Zhang, J., Yang, H., Ding, M., et al. (2020). GCC: Graph contrastive coding for graph neural network pre-training. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 1150–1160). http://dx.doi.org/10.1145/3394486.3403168.

Rana, P., Sowmya, A., Meijering, E., & Song, Y. (2022a). Data augmentation with improved regularisation and sampling for imbalanced blood cell image classification. *Scientific Reports*, *12*(1), 18101.

Rana, P., Sowmya, A., Meijering, E., & Song, Y. (2022b). Imbalanced cell-cycle classification using WGAN-div and mixup. In *2022 IEEE 19th international symposium on biomedical imaging* (pp. 1–4). IEEE.

Rana, P., Sowmya, A., Meijering, E., & Song, Y. (2023). Imbalanced classification for protein subcellular localization with multilabel oversampling. *Bioinformatics*, *39*(1), btac841.

Ren, J., Yu, C., Ma, X., Zhao, H., Yi, S., et al. (2020). Balanced meta-softmax for long-tailed visual recognition. *Advances in Neural Information Processing Systems*, *33*, 4175–4186.

Reza, M. S., & Ma, J. (2018). Imbalanced histopathological breast cancer image classification with convolutional neural network. In *International conference on intelligent computing and signal processing* (pp. 619–624).

Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618–626).

Shen, Y., & Ke, J. (2020). A deformable crf model for histopathology whole-slide image classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 500–508). Springer.

Sivapuram, A. K., Ravi, V., Senthil, G., Gorthi, R. K., et al. (2023). VISAL—A novel learning strategy to address class imbalance. *Neural Networks*, *161*, 178–184.

Soltanzadeh, P., Feizi-Derakhshi, M. R., & Hashemzadeh, M. (2023). Addressing the class-imbalance and class-overlap problems by a metaheuristic-based under-sampling approach. *Pattern Recognition*, Article 109721.

Sun, Q., Li, J., Peng, H., Wu, J., Ning, Y., Yu, P. S., et al. (2021). SUGAR: Subgraph neural network with reinforcement pooling and self-supervised mutual information mechanism. In *Proceedings of the web conference 2021* (pp. 2081–2091). Association for Computing Machinery, http://dx.doi.org/10.1145/3442381.3449822.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., et al. (2017). Graph attention networks. In *International conference on learning representations*.

Wang, X., Lian, L., Miao, Z., Liu, Z., & Yu, S. X. (2020). Long-tailed recognition by routing diverse distribution-aware experts. arXiv preprint arXiv:2010.01809.

Wang, C., & Liu, Z. (2021). Learning graph representation by aggregating subgraphs via mutual information maximization. arXiv:2103.13125.

Wang, X., Liu, N., Han, H., & Shi, C. (2021). Self-supervised heterogeneous graph neural network with co-contrastive learning. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining* (pp. 1726–1736). Association for Computing Machinery, ISBN: 9781450383325.

Wang, J., Zhang, W., Zang, Y., Cao, Y., Pang, J., Gong, T., et al. (2021). Seesaw loss for long-tailed instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9695–9704).

Wei, M., Zhou, Y., Li, Z., & Xu, X. (2023). Class-imbalanced complementary-label learning via weighted loss. *Neural Networks*, *166*, 555–565.

Yang, Z., Pan, J., Yang, Y., Shi, X., Zhou, H. Y., Zhang, Z., et al. (2022). Proco: Prototype-aware contrastive learning for long-tailed medical image classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 173–182). Springer.

Yang, J., Shi, R., & Ni, B. (2021). Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. In *2021 IEEE 18th international symposium on biomedical imaging* (pp. 191–195). IEEE.

Yoon, C., Hamarneh, G., & Garbi, R. (2019). Generalizable feature learning in the presence of data bias and domain class imbalance with application to skin lesion classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 365–373). Springer.

Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). Mixup: Beyond Empirical Risk Minimization. In *International conference on learning representations*.

Zhang, Y., Hooi, B., Hong, L., & Feng, J. (2022). Self-supervised aggregation of diverse experts for test-agnostic long-tailed recognition. In *Advances in neural information processing systems*.

Zhang, S., Li, Z., Yan, S., He, X., & Sun, J. (2021). Distribution alignment: A unified framework for long-tail visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2361–2370).

Zhang, J., Ruan, Z., Li, Q., & Zhang, Z. Q. (2023). Towards robust and efficient musculoskeletal modelling using distributed physics-informed deep learning. *IEEE Transactions on Instrumentation and Measurement*.

Zhang, C., Tan, K. C., Li, H., & Hong, G. S. (2018). A cost-sensitive deep belief network for imbalanced classification. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(1), 109–122.

Zhang, X., Wu, Z., Weng, Z., Fu, H., Chen, J., Jiang, Y. G., et al. (2021). Videolt: large-scale long-tailed video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 7960–7969).

Zhang, J., Zhao, Y., Shone, F., Li, Z., Frangi, A. F., Xie, S. Q., et al. (2022). Physics-informed deep learning for musculoskeletal modeling: Predicting muscle forces and joint kinematics from surface EMG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 484–493.

Zhao, R., Chen, X., Chen, Z., & Li, S. (2022). Diagnosing glaucoma on imbalanced data with self-ensemble dual-curriculum learning. *Medical Image Analysis*, *75*, Article 102295.

Zhou, B., Cui, Q., Wei, X.-S., & Chen, Z.-M. (2020). Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9719–9728).

Zhu, B., Niu, Y., Hua, X.-S., & Zhang, H. (2022). Cross-domain empirical risk minimization for unbiased long-tailed classification. In *Proceedings of the AAAI conference on artificial intelligence, vol. 36* (pp. 3589–3597).

Zhu, Y., Xu, Y., Yu, F., Liu, Q., Wu, S., & Wang, L. (2020). Deep Graph Contrastive Representation Learning. In *ICML workshop on graph representation learning and beyond*. URL http://arxiv.org/abs/2006.04131.

Zhu, Q., Yang, C., Xu, Y., Wang, H., Zhang, C., & Han, J. (2020). Transfer learning of graph neural networks with ego-graph information maximization. arXiv preprint arXiv:2009.05204.

Zhuang, J. X., Cai, J., Zhang, J., Zheng, W. s., & Wang, R. (2023). Class attention to regions of lesion for imbalanced medical image recognition. *Neurocomputing*, *555*, Article 126577.