Combat Long-tails in Medical Classification with Relation-aware Consistency and Virtual Features Compensation

Li Pan[1*], Yupei Zhang[2*], Qiushi Yang[3], Tan Li[4], Zhen Chen[5]

[1]The Chinese University of Hong Kong
[2]The Centre for Intelligent Multidimensional Data Analysis (CIMDA)
[3]City University of Hong Kong
[4]Department of Computer Science, The Hang Seng University of Hong Kong
[5]Centre for Artificial Intelligence and Robotics (CAIR), HKISI-CAS

*Equal contribution

Outline

- Background
- Long-tailed Challenge
- Our Method
  - Multi-view Relation-aware Consistency (MRC)
  - Virtual Feature Compensation (VFC)
- Experiments
- Conclusion

Background

- Class imbalance inherently exists in medical datasets due to the scarcity of target diseases, where normal samples are significantly more than diseased samples.
- The issue of class imbalance can be formulated as a long-tailed problem, where a few head classes contain numerous samples while the tail classes comprise only a few instances.

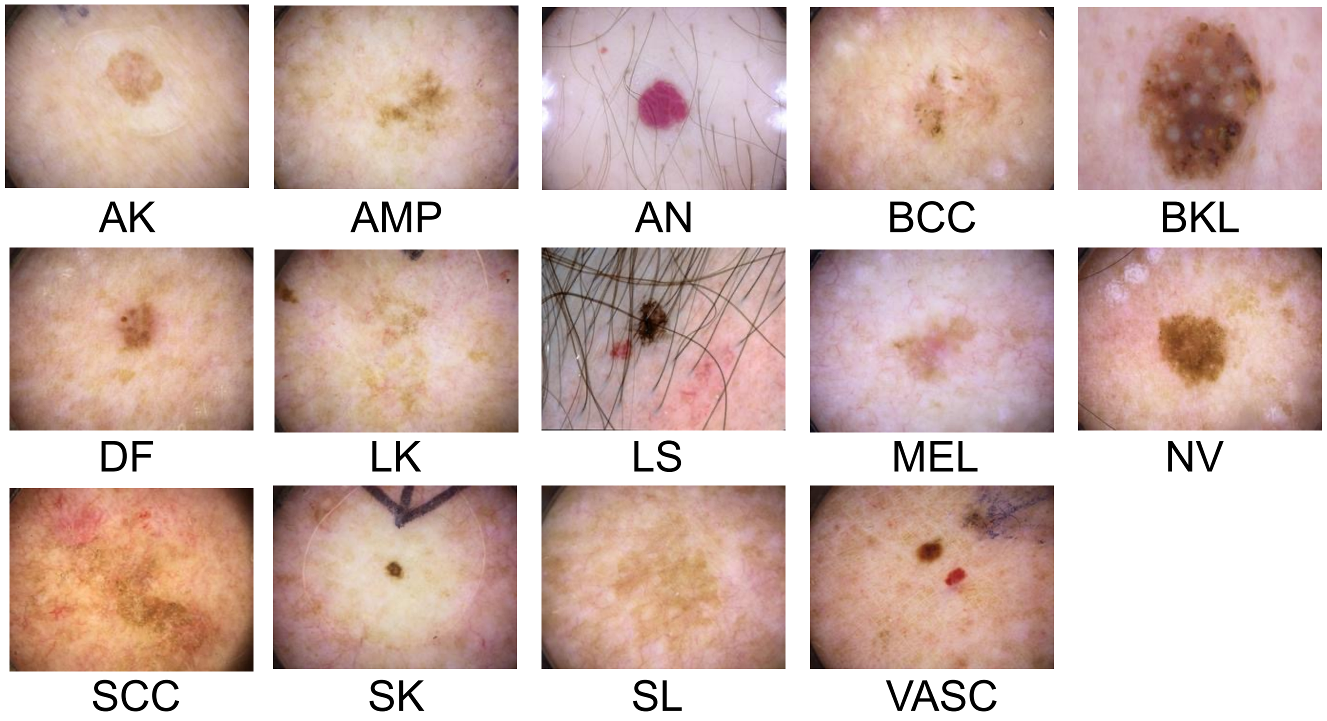

Fig.1 Examples from ISIC Archive LT dataset

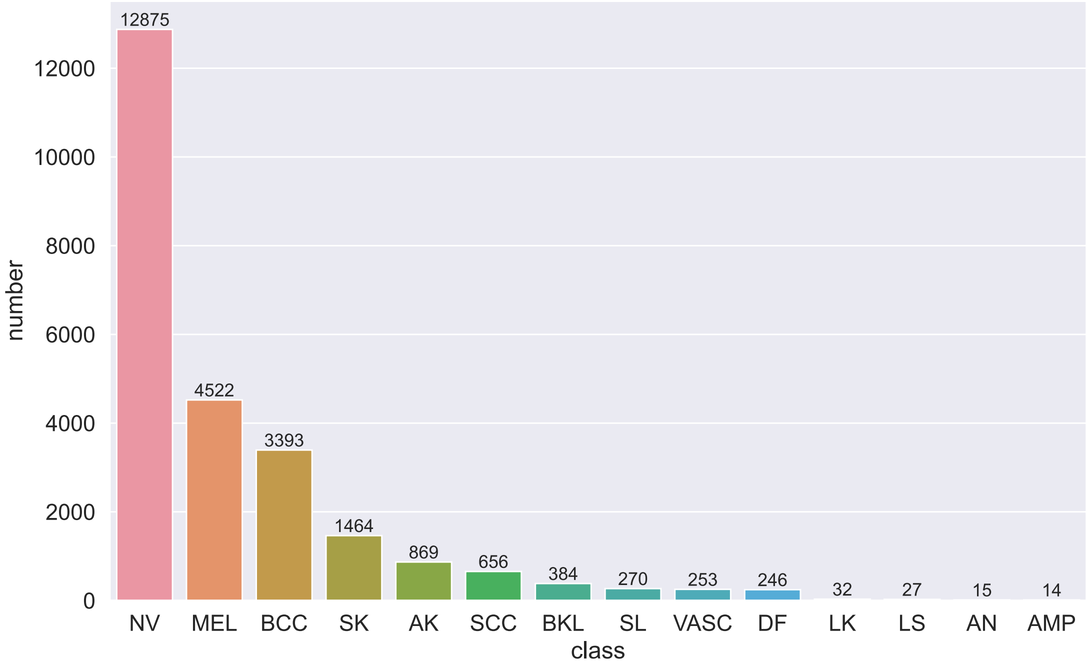

Fig.2 Class distribution of ISIC Archive LT dataset

Long-tailed Challenge

- This long-tailed distribution in medical datasets leads the model training biased to the majority categories and severely impairs the performance of diagnostic models in real-world scenarios.
- Current decoupling methods suffer from inefficient representation learning during the first stage and inadequate classifier recalibration in the second stage.

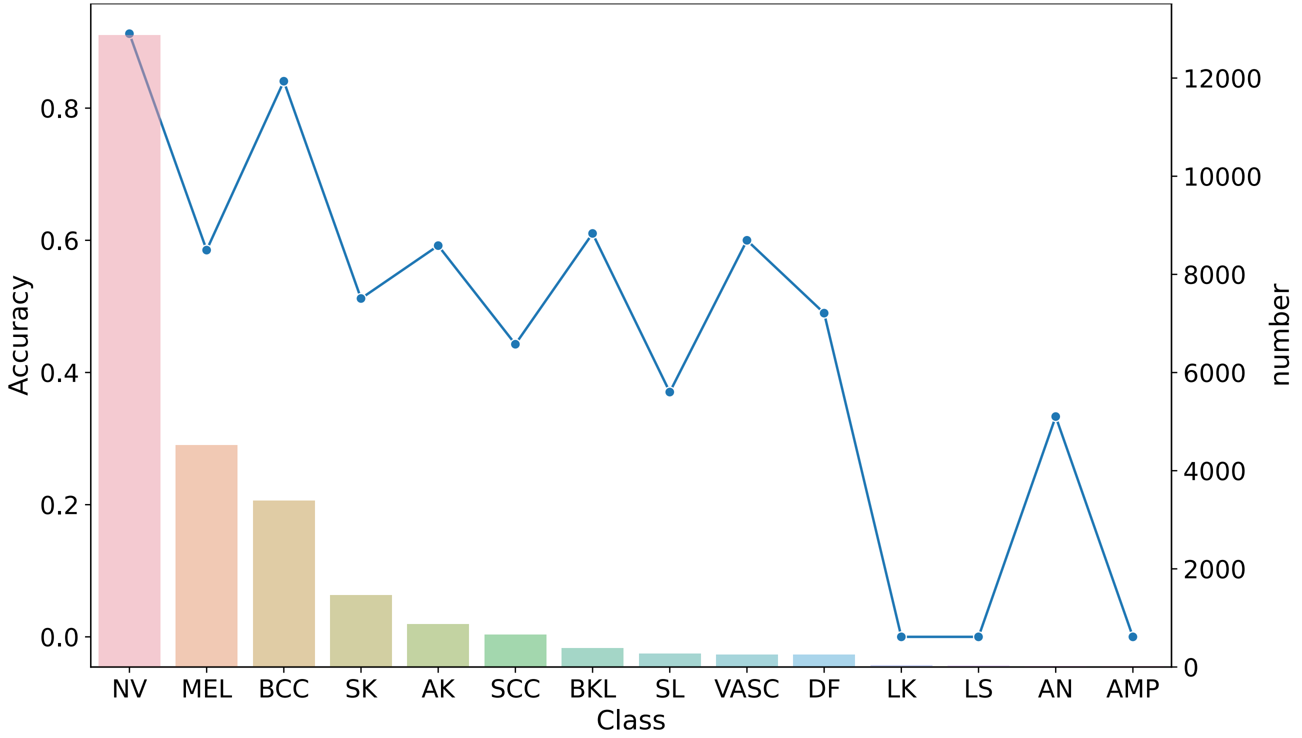

Fig.3 Accuracy on each Class of ISIC Archive LT dataset

Method: Overview

- We propose the MRC-VFC framework to combat long-tailed problems.
- **Multi-view Relation-aware Consistency (MRC)**, to enhance the encoder's representation ability, especially on the tail classes by constraining the encoder's consistency from multiple views.
- **Virtual Feature Compensation (VFC)**, to recalibrate the classifier by uniformly sampling instances for each class under multivariate Gaussian distribution in feature space, and an iterative optimization procedure.

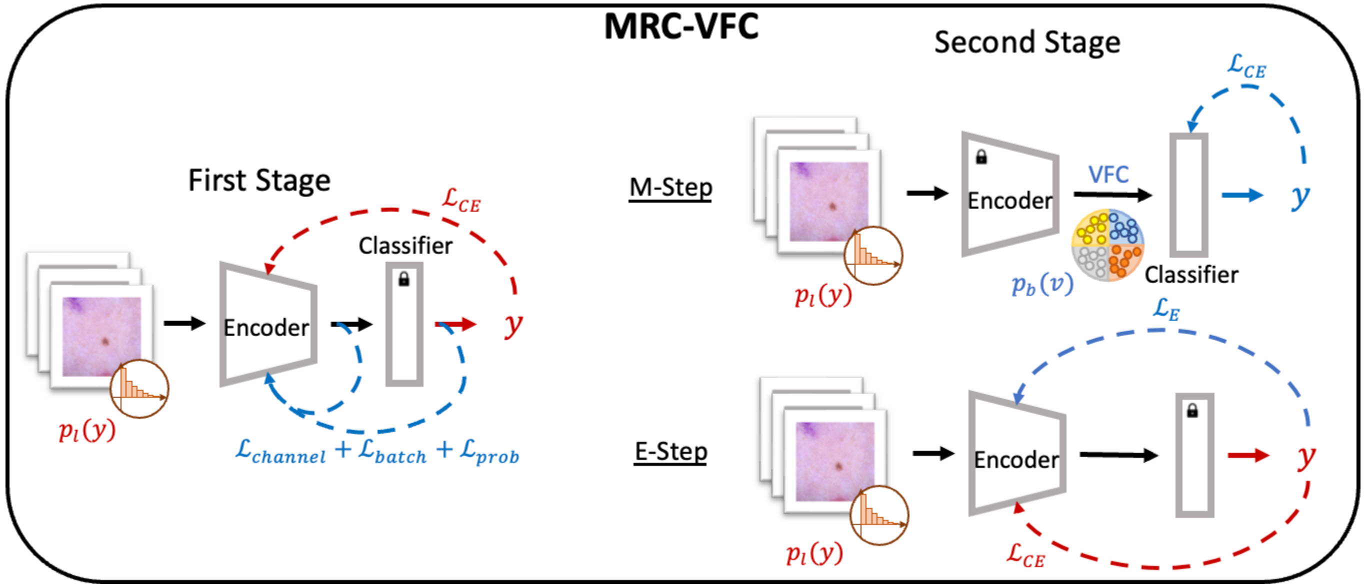

Fig.4 The overall framework of our MRC-VFC method.

Method: Multi-view Relation-aware Consistency (MRC)

- MRC is proposed to encourage the encoder to apprehend the inherent semantic features of the input images under different data augmentations as in Fig. 5.

1) Logits consistency: $\mathcal{L}_{\text{prob}} = \frac{1}{B}\text{KL}(f \cdot g(\boldsymbol{x}_s)\ ,\ f' \cdot g'(\boldsymbol{x}_w))$

2) Batch consistency: $\mathcal{L}_{\text{batch}} = \frac{1}{B}||\mathcal{S}_b(g(\boldsymbol{x}_s)) - \mathcal{S}_b(g'(\boldsymbol{x}_w))||_2$

3) Channel consistency: $\mathcal{L}_{\text{channel}} = \frac{1}{C}||\mathcal{S}_c(g(\boldsymbol{x}_s)) - \mathcal{S}_c(g'(\boldsymbol{x}_w))||_2$

4) Cross-entropy constraint: $\mathcal{L}_{\text{CE}} = \frac{1}{B}L(f \cdot g(\boldsymbol{x}_w),\ y)$

5) Overall loss function at Stage1:

$$\mathcal{L}_{\text{stage1}} = \mathcal{L}_{\text{CE}} + \lambda_1\mathcal{L}_{\text{batch}} + \lambda_2\mathcal{L}_{\text{channel}} + \lambda_3\mathcal{L}_{\text{prob}}$$

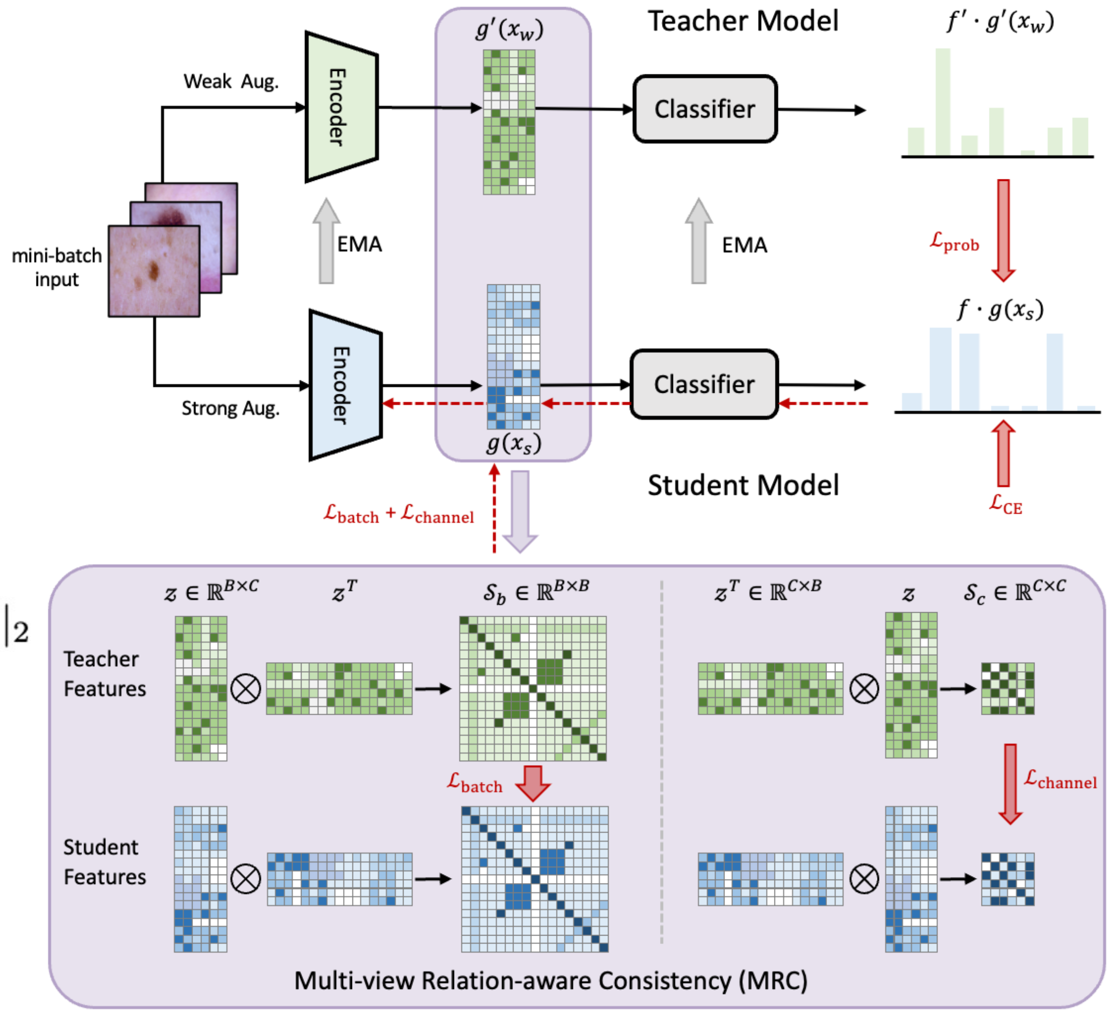

Fig.5 The Multi-view Relation-aware Consistency module.

Method: Virtual Feature Compensation (VFC)

- We propose Virtual Features Compensation (Fig.6), which generates virtual features for each class under multivariate Gaussian distribution [1] to combat the long-tailed problem.

1) The mean and covariance are:

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{\boldsymbol{x} \in X_k} g^I(\boldsymbol{x}), \ \boldsymbol{\Sigma}_k = \frac{1}{N_k - 1} \sum_{\boldsymbol{x} \in X_k} (\boldsymbol{x} - \boldsymbol{\mu}_k)^\intercal (\boldsymbol{x} - \boldsymbol{\mu}_k)$$

2) M step constraint:

$$\mathcal{L}^M_{\text{stage2}} = \frac{1}{RK} \sum_{k=1}^{K} \sum_{\boldsymbol{v}_i \in V_k} L_{\text{CE}}(f(\boldsymbol{v}_i), y)$$

3) E step constraint:

$$\mathcal{L}^E_{\text{stage2}} = \frac{1}{N} \sum_{\boldsymbol{x} \in X} \frac{(1 - (f \cdot g^I(\boldsymbol{x}) y)^q)}{q}$$

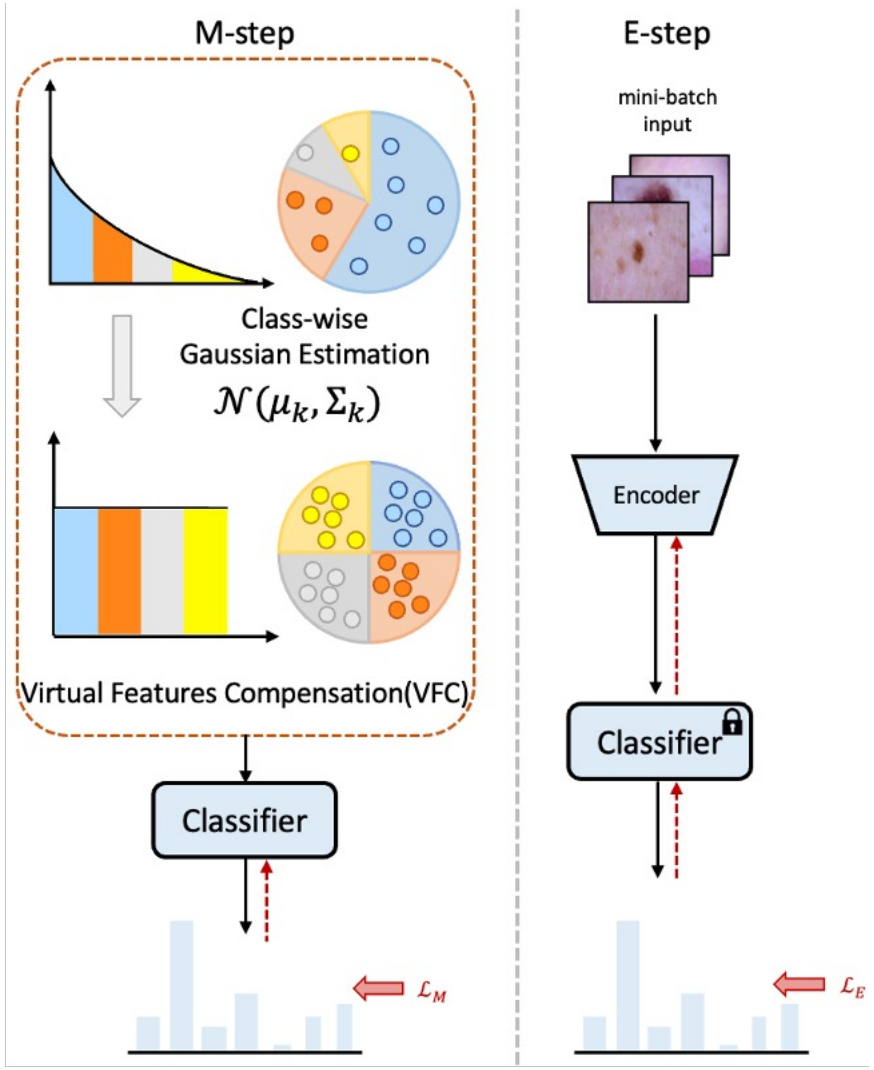

Fig.6 The Virtual Feature Compensation module.

[1] Ahrendt, Peter. "The multivariate gaussian probability distribution." Technical University of Denmark, Tech. Rep (2005): 203.

Experiment: Dataset

- To evaluate the performance on long-tailed medical image classification, we construct two dermatology datasets from ISIC1 [2] following [3].

Table 1. The class distribution of the *ISIC-Archive-LT* dataset.

ISIC-Archive-LT														
Group	Head				Middle				Tail					
Class	NV	MEL	BCC	SK	AK	SCC	BKL	SL	VASC	DF	LK	LS	AN	AMP
Training	9012	3165	2375	1024	608	459	268	189	177	172	22	18	10	9
Validation	1288	452	339	147	87	65	39	27	25	24	3	3	2	2
Testing	2575	905	679	293	174	132	77	54	51	50	7	6	3	3
Total	12875	4522	3393	1464	869	656	384	270	253	246	32	27	15	14

Table 2. The class distribution of the *ISIC-2019-LT* dataset with imbalance factor = {100, 200, 500}.

ISIC-2019-LT								
Class	NV	MEL	BCC	BKL	AK	SCC	VASC	DF
Original	12875	4522	3323	2624	867	628	253	239
Factor=100	12875	4140	2785	2013	609	404	149	129
Factor=200	12875	3750	2285	1496	410	247	83	65
Factor=500	12875	3290	1759	1010	243	128	38	26

[2] Tschandl, Philipp, Cliff Rosendahl, and Harald Kittler. "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions." Scientific data 5.1 (2018): 1-9.
[3] Ju, Lie, et al. "Flexible sampling for long-tailed skin lesion classification." International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022.

Experiment: Comparison with SOTA

- MRC-VFC outperforms state-of-the-art decoupling, reweighting, and resampling works.
- As illustrated in Table 4, we compare our MRC-VFC framework with the aforementioned methods on the ISIC-2019-LT dataset under different imbalance factors. Among these methods, our MRC-VFC framework achieves the best performance with an accuracy of 77.41%, 75.98%, and 74.62% under the imbalance factor of 100, 200, and 500, respectively.

Table 3. Comparison with state-of-the-art algorithms on the *ISIC-2019-LT* dataset.

ISIC-2019-LT			
Methods	Acc(%) @ Factor=100	Acc(%) @ Factor=200	Acc(%) @ Factor=500
CE	56.91	53.77	43.89
RS	61.41	55.12	47.76
MixUp	59.85	54.23	43.11
GCE+SR	64.57	58.28	54.36
Seesaw loss	68.82	65.84	62.92
Focal loss	67.54	65.93	61.66
CB loss	67.54	66.70	61.89
FCD	70.15	68.82	63.59
FS	71.97	69.30	65.22
Ours *w/o* MRC	75.04	73.13	70.13
Ours *w/o* VFC	72.91	71.07	67.48
Ours	**77.41**	**75.98**	**74.62**

Experiment: Comparison with SOTA

- We further perform the comparison with state-of-the-art algorithms on a more challenging ISIC-Archive-LT dataset for long-tailed diagnosis. As illustrated in Table 2, our MRC-VFC framework achieves the best overall performance with an accuracy of 67.84% among state-of-the-art algorithms, and results in a balanced performance over different classes, i.e., 69.71% for head classes and 70.34% for tail classes.

Table 4. Comparison with state-of-the-art algorithms on the *ISIC-Archive-LT* dataset.

ISIC-Archive-LT				
Methods	Head (Acc%)	Medium (Acc%)	Tail (Acc%)	All (Acc%)
CE	**71.31**	49.22	38.17	52.90
RS	70.17	55.29	34.29	53.25
GCE+SR	64.93	57.26	38.22	53.47
Seesaw loss	70.26	55.98	42.14	59.46
Focal loss	69.57	56.21	39.65	57.81
CB loss	64.98	57.01	61.61	61.20
FCD	66.39	61.17	60.54	62.70
FS	68.69	58.74	64.48	63.97
Ours *w/o* MRC	69.06	62.14	65.12	65.44
Ours *w/o* VFC	65.11	62.35	67.30	64.92
Ours	69.71	**63.47**	**70.34**	**67.84**

Conclusion

- To address the long-tails in computer-aided diagnosis, we propose the MRC-VFC framework to improve medical image classification with balanced performance in two stages.
- In the first stage, we design the MRC to facilitate the representation learning of the encoder by introducing multi-view relation-aware consistency.
- In the second stage, to recalibrate the classifier, we propose the VFC to train an unbiased classifier for the MRC-VFC framework by generating massive virtual features.
- Extensive experiments on the two long-tailed dermatology datasets demonstrate the effectiveness of the proposed MRC-VFC framework, which outperforms state-of-the-art algorithms remarkably.

Source Code

Research Collaboration

Thanks!