Combating Medical Label Noise via Robust Semi-supervised Contrastive Learning

Bingzhi Chen[1,3], Zhanhao Ye[1], Yishu Liu[2(✉)], Zheng Zhang[2(✉)], Jiahui Pan[1], Biqing Zeng[1], and Guangming Lu[2,3]

[1] School of Software, South China Normal University, Guangzhou, China
[2] School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China
liuyishu@stu.hit.edu.cn, darrenzz219@gmail.com
[3] Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies, Shenzhen, China

Abstract. Deep learning-based AI diagnostic models rely heavily on high-quality exhaustive-annotated data for algorithm training but suffer from noisy label information. To enhance the model's robustness and prevent noisy label memorization, this paper proposes a robust Semi-supervised Contrastive Learning paradigm called SSCL, which can efficiently merge semi-supervised learning and contrastive learning for combating medical label noise. Specifically, the proposed SSCL framework consists of three well-designed components: the Mixup Feature Embedding (MFE) module, the Semi-supervised Learning (SSL) module, and the Similarity Contrastive Learning (SCL) module. By taking the hybrid augmented images as inputs, the MFE module with momentum update mechanism is designed to mine abstract distributed feature representations. Meanwhile, a flexible pseudo-labeling promotion strategy is introduced into the SSL module, which can refine the supervised information of the noisy data with pseudo-labels based on initial categorical predictions. Benefitting from the measure of similarity between classification distributions, the SCL module can effectively capture more reliable confident pairs, further reducing the effects of label noise on contrastive learning. Furthermore, a noise-robust loss function is also leveraged to ensure the samples with correct labels dominate the learning process. Extensive experiments on multiple benchmark datasets demonstrate the superiority of SSCL over state-of-the-art baselines. The code and pre-trained models are publicly available at https://github.com/Binz-Chen/MICCAI2023_SSCL.

Keywords: Medical Label Noise · Mixup · Semi-supervised Learning · Contrastive Learning

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43907-0_54.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14220, pp. 562–572, 2023.
https://doi.org/10.1007/978-3-031-43907-0_54

1 Introduction

The advancement of deep learning models heavily depends on the availability of large-scale datasets with high-quality annotated labels [4]. However, it is costly and time-consuming to obtain a sufficient number of accurate annotations from clinical systems, which inevitably introduces a certain level of noisy label [16]. This phenomenon seriously affects the stability and robustness of medical training and prediction procedures, leading to the production of corrupted representations and inaccurate classification boundaries [27]. Early approaches for combating the label noise mainly focused on the marginal improvements of model robustness, including designing robust loss functions [14,26], preprocessing the image [24,28], and estimating the noise transition matrix [8]. In recent years, some advanced methods [2,19] with semi-supervised learning [22] are designed to leverage the supervision signal provided by pseudo labels to re-correct the noise bias. With the development of contrastive learning technologies, some scholars [15,17] attempt to learn the contrastive representations between the clean and noisy labels by maximizing the similarity of positive pairs and minimizing the similarity of negative pairs. Despite the advances in conventional image processing, few studies have proposed to overcome the medical label noise.

To overcome this issue, this paper proposes a robust Semi-supervised Contrastive Learning (SSCL) paradigm, that simultaneously benefits from semi-supervised learning and contrastive learning for combating the medical noisy labels and promoting stability and robustness of the diagnostic model. Three important components, i.e., the Mixup Feature Embedding (MFE) module, the Semi-supervised Learning (SSL) module, and the Similarity Contrastive Learning (SCL) module, are proposed in the SSCL framework. The architecture of the SSCL framework is shown in Fig. 1. Specifically, the MFE module is built on a multi-branch architecture with the momentum update mechanism, which can effectively capture the abstract distributed feature representations from various levels of mixup augmented images. Based on the confidence scores provided by the initial classifier, a flexible pseudo-labeling promotion strategy is introduced into the SSL module to effectively select confident samples and generate the pseudo-labels, resulting in a more accurate supervision signal reconstruction. By calculating their similarity distribution of representation learning, a novel pair-wise selection strategy in the SCL module is designed to efficiently identify and select more reliable confident pairs out of noisy pairs for contrasting learning. In the training phase, we broaden the scope of penalization by incorporating loss functions, further reducing the impact of noise on statistical classification. The main contributions of our work are as follows:

- This paper presents a robust semi-supervised contrastive learning paradigm that effectively incorporates semi-supervised learning and contrastive learning to mitigate the effect of medical label noise. Our approach represents the first attempt to address this issue in the field of medical image analysis.
- The pseudo-labeling promotion strategy can re-correct the supervised information of noisy labels, while the pair-wise selection strategy can guide the confident pairs to dominate the contrasting learning process.

– The proposed SSCL framework is evaluated on multiple benchmark datasets, and extensive experiments demonstrate the generalization performance of our method in comparison with state-of-the-art baselines.

2 Related Work

2.1 Conventional Methods with Noisy Labels

To eliminate the memorization effect of noise labels in the training phase, recent works [14,26,29] were mainly devoted to exploring the effectiveness of robust loss function. Wang et al. [26] proposed a noise-robust loss function that combined with Cross-Entropy (CE) loss, to address the hard class learning problem and noisy label overfitting problem. Yi et al. [14] investigated the representational benefits of the contrastive regularization loss function to learn contrastive representations with noisy labels. With continuous in-depth research on image processing, data augmentation has been proven to have a significant role in combating noisy labels. Zhang et al. [28] presented a data-agnostic and straightforward data augmentation principle to mix up different images geometrically in the feature space, which has been widely used in the field of noise labeling. Moreover, researchers have attempted to leverage the label noise transfer matrix extracted from the data set to solve the noise label problem. Hendrycks et al. [8] directly used the matrix summarizing the probability of one class being flipped into another under noise. Ramaswamy et al. [18] proposed an efficient kernel mean embedding to overcome mixture proportion estimation.

2.2 Semi-supervised Learning

Compared with the existing learning methods, semi-supervised learning [3,12,23] has been recognized as an effective technique to solve the problem of noisy labels. As a semi-supervised learning technique, pseudo-labeling is frequently used in conjunction with confidence-based thresholding to retain unlabeled examples and increase the size of labeled training data. In recent years, some advanced works [2,19] demonstrate the capacity of pseudo-label based semi-supervised learning methods in combating medical label noise. Reed et al. [19] augmented the usual prediction objective with a notion of perceptual consistency, and the article referred to this approach as static hard guidance. Arazo et al. [2] proposed dynamic hard and soft bootstrapping losses by individual weight of each sample. Furthermore, the combination of semi-supervised learning and pseudo-labels can be used not only for single-label classification but also for negative learning and multi-label classification [20].

2.3 Contrastive Learning

With the advancement of deep learning, researchers have found the potential of contrastive-based similarity learning frameworks for representation learning [10,15,25]. Unsupervised contrastive learning [7,21] aims to maximize the similarity of positive pairs and minimize the similarity of negative pairs at the instance level. By incorporating clean label information, supervised contrastive learning [13] can obtain more supervised information and achieve better performance. Recently, many state-of-the-art works [15,17] have been proposed to make full use of contrastive learning for combating label noise. MOIT [17] adopted the method of interpolation contrastive learning, and used the supervised information obtained after semi-supervised learning for contrastive learning, so as to reduce the damage of noise labels on contrastive learning. Sel-CL [15] improved MOIT by introducing the concepts of confident use cases and confident pairs which improved the ability to filter noise labels with new detection strategies.

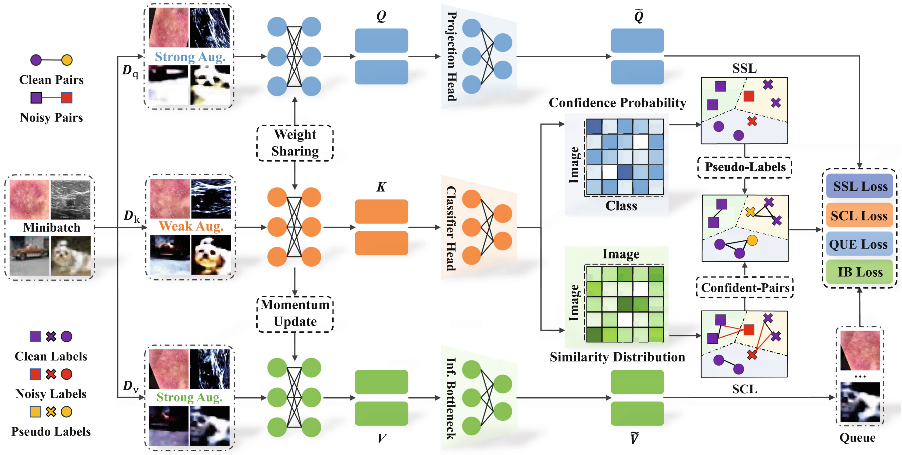


Fig. 1. Illustration of the proposed SSCL framework for combating medical label noise.

3 Semi-supervised Contrastive Learning

3.1 Mixup Feature Embedding

Mixup. Let $\boldsymbol{D} = \{(x_i, y_i)\}_i^N$ denotes the training minibatch of image-label pairs x_i and y_i, where N is the batch size. Initially, the operations of data augmentation are first conducted in the MFE module, to generate various levels of hybrid augmented images with Mixup [28]. As shown in Fig. 1, the obtained hybrid data set involves a group of weak augmentation images, i.e., $\boldsymbol{D}_k$, and two groups of strong augmentation images, i.e., $\boldsymbol{D}_q$ and $\boldsymbol{D}_v$. The principle of Mixup can be defined as:

$$x_i \leftarrow \lambda x_a + (1 - \lambda)\, x_b, \tag{1}$$

where $\lambda \in [0, 1] \sim Beta(\alpha_l, \alpha_r)$ is used to control the mixing strength of training samples, x_a and x_b are the training samples randomly drawn from each minibatch, and x_i is the enhanced image generated by the mixup preprocessing.

Feature Embedding. By taking the corresponding augmented images as inputs, the MFE module aims to capture abstract distributed feature representations. The MFE module consists of different branches, including a deep encoder with projection and classifier heads, and a momentum encoder with an information bottleneck (IB) [6]. Motivated from the physical perspective of optimization [7], the momentum update mechanism can be defined as:

$$\theta_v \leftarrow m\theta_v + (1-m)\theta_k, \tag{2}$$

where θ denotes the encoder parameters, and $m \in [0,1)$ is a momentum coefficient. Specifically, the feature representations $\boldsymbol{Q}$ would be mapped to the same-dimensional representations $\tilde{\boldsymbol{Q}}$ by the projection head, while $\boldsymbol{K}$ is used to predict the categorical outputs with classifier. The feature representation $\boldsymbol{V}$ is derived from the momentum encoder, possessing identical dimensions to $\boldsymbol{Q}$. Moreover, we also utilize Kullback-Leibler divergence [11] to implement the criterion of exploiting IB, resulting in a compact feature representation $\tilde{\boldsymbol{V}}$. By alternately learning robust representations from $\boldsymbol{D}_q$ and $\boldsymbol{D}_v$, a symmetry objective function is designed to gain the IB loss,

$$\mathcal{L}_{IB} = \sum_{x_i \in \boldsymbol{D}_q} \sum_{x_j \in \boldsymbol{D}_v} KL\left(\tilde{\boldsymbol{Q}}_i \parallel \boldsymbol{V}_j\right) + \sum_{x_i \in \boldsymbol{D}_v} \sum_{x_j \in \boldsymbol{D}_q} KL\left(\tilde{\boldsymbol{Q}}_i \parallel \boldsymbol{V}_j\right). \tag{3}$$

3.2 Semi-Supervised Learning

Selecting Confident Samples. The main core of the proposed SSL module is to recognize the confident samples with clean labels, and re-correct the supervision information of label noise. To this end, we first select confident samples based on their confidence score provided by the classifier. Denote the confident samples with clean label belonging to n-th class as $\boldsymbol{D}_c^n$,

$$\boldsymbol{D}_c^n = \{(x_i, y_i) \mid y_i \cdot p_i > \gamma_n\}, \tag{4}$$

where $p_i \in [0,1]$ is the classification probability of the enhanced image x_i, and γ_n is a dynamic confidence threshold for the n-th class to ensure a class-balanced set of identified confident examples.

Pseudo-Labeling. To generate accurate supervision signals, a flexible pseudo-labeling promotion strategy is introduced to replace noisy labels with pseudo-labels,

$$\tilde{y}_i = \begin{cases} y_i, & x_i \in \boldsymbol{D}_c, \\ p_i, & x_i \notin \boldsymbol{D}_c. \end{cases} \tag{5}$$

Benefiting from the unique semi-supervised learning structure, our SSL module can effectively reduce the impact of noise based on statistical classification. The objective function for semi-supervised learning is defined as:

$$\mathcal{L}_{SSL} = \sum_{x_i \in \boldsymbol{D}} \frac{1-(\tilde{y}_i \cdot p_i)^{\omega}}{\omega}, \tag{6}$$

where $\omega \in (0,1]$ is a tunable focusing parameter, which is utilized to exploit the benefits of both the noise-robustness and the implicit weighting scheme.

3.3 Similarity Contrastive Learning

Selecting Confident Pairs. To achieve a precise estimation of noisy pairs, a novel pair-wise selection strategy is proposed to identify the reliable confident pairs out of noisy pairs. By calculating their similarity distribution of representation learning, the SCL module can transform identified confident examples into a set of associated confident pairs $\mathcal{S}$ without knowing noise rates,

$$\mathcal{S} = \{(x_i, x_j) \mid p_i^\top p_j \geq \tau\}, \tag{7}$$

where τ is considered as a confidence threshold. Therefore, the objective loss on each sample pair for contrastive learning can be defined as:

$$\mathcal{L}(x_i, x_j) = \log\left(1 - \left\langle \tilde{\boldsymbol{Q}}_{\boldsymbol{i}}, \boldsymbol{V}_{\boldsymbol{j}} \right\rangle\right) \mathbb{1}\left[(x_i, x_j) \in \mathcal{S}\right]. \tag{8}$$

Consistent with Eq. 3, a symmetry loss function is applied for each minibatch,

$$\mathcal{L}_{SCL} = \sum_{x_q \in \boldsymbol{D}_q} \sum_{x_v \in \boldsymbol{D}_v} \mathcal{L}(x_q, x_v) + \sum_{x_v \in \boldsymbol{D}_v} \sum_{x_q \in \boldsymbol{D}_q} \mathcal{L}(x_v, x_q). \tag{9}$$

Queue. It is noted that blindly increasing the size of the minibatch will be limited by computing resources [15]. To overcome these issues, a queue with the length of L is also introduced into the SCL module, which can maintain a feature dictionary to store features and decouple the dictionary size from the mini-batch size [7,13]. As the new minibatch is added to the queue, the queue $\boldsymbol{M} = \{x_i, \boldsymbol{V}_i\}_{i=1}^{L}$ is gradually replaced and the oldest features in the queue are removed. The objective loss of the queue is defined as:

$$\mathcal{L}_{QUE} = \sum_{x_q \in \boldsymbol{D}_q} \sum_{x_l \in \boldsymbol{M}} \mathcal{L}(x_q, x_l) + \sum_{x_v \in \boldsymbol{D}_v} \sum_{x_l \in \boldsymbol{M}} \mathcal{L}(x_v, x_l). \tag{10}$$

Objective Loss of SSCL. Based on the analysis of the above modules, the total objective loss for the proposed SSCL method can be obtained by,

$$\mathcal{L}_{SSCL} = \mathcal{L}_{SSL} + \alpha \cdot \mathcal{L}_{QUE} + \beta \cdot \mathcal{L}_{IB}, \tag{11}$$

where α and β are loss weight. In our experiments, both α and β are set to 0.25.

4 Experiments

4.1 Implementation Details and Settings

By randomly replacing labels for a percentage of the training data with all possible labels, the proposed SSCL method is extensively validated on four benchmarks with symmetric noise. ISIC-19 [30] dataset boasts a training set of 20,400

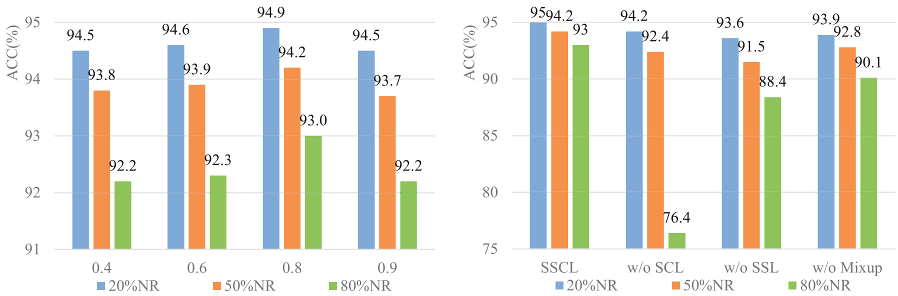
Table 1. Comparisons with the state-of-the-art baselines on benchmark datasets.

Methods		CE	Mixup	GCE	MOIT	Sel-CL	CTRR	SSCL	Improv.
Datasets	NR	–	–	–	CVPR'21	CVPR'22	CVPR'22	–	–
ISIC19	0.2	75.4	72.2	75.5	79.7	–	<u>79.9</u>	**81.1**	1.2 ↑
	0.3	73.1	70.2	72.6	<u>75.9</u>	–	75.8	**76.7**	0.9 ↑
	0.5	64.1	62.1	64.8	65.1	–	<u>65.4</u>	**66.8**	1.4 ↑
	Mean	70.9	68.2	71.0	73.6	–	73.7	**74.9**	1.2 ↑
BUSI	0.2	80.6	72.9	81.9	<u>84.5</u>	-	83.8	**86.9**	2.4 ↑
	0.3	76.7	67.8	77.4	77.2	–	<u>78.3</u>	**81.2**	2.9 ↑
	0.5	72.3	66.5	72.3	<u>73.5</u>	–	72.5	**74.9**	1.4 ↑
	Mean	76.5	69.1	77.2	78.4	–	78.2	**81.0**	2.6 ↑
CIFAR-10	0.2	82.7	92.3	86.6	94.1	**95.5**	93.9	<u>95.0</u>	0.5 ↓
	0.5	57.9	77.6	81.9	91.8	<u>93.9</u>	91.7	**94.2**	0.3 ↑
	0.8	26.1	46.7	54.6	81.1	<u>89.2</u>	88.1	**93.0**	3.8 ↑
	Mean	55.6	72.2	74.4	89.0	<u>92.9</u>	91.2	**94.1**	1.2 ↑
CIFAR-100	0.2	61.8	66.0	59.2	75.9	**76.5**	73.8	<u>75.0</u>	1.5 ↓
	0.5	37.3	46.6	47.8	70.6	<u>72.4</u>	72.2	**73.4**	1.0 ↑
	0.8	8.8	17.6	15.8	47.6	59.6	<u>63.9</u>	**66.5**	2.6 ↑
	Mean	36.0	43.4	40.9	64.7	69.5	<u>70.0</u>	**71.6**	1.6 ↑

dermoscopic images and a test set of 4,291 images, while BUSI [1] dataset encompasses 625 ultrasound images in the training set and 155 images in the test set. Both CIFAR-10 and CIFAR-100 contain 50,000 training images and 10,000 test images. The backbone of our SSCL framework is built on ResNet-18. In addition to Mixup, a series of augmentation techniques, such as random flipping, cropping, and Gaussian blur, are randomly applied to generate the hybrid augmented images during the training process. We compare the proposed SSCL method with several state-of-the-art baselines, including three conventional noise-robustness methods (i.e., CE [9], GCE [29], and Mixup [28]), three state-of-the-art baselines (i.e., MOIT [17], Sel-CL [15], and CTRR [14]). More implementation details are shown in the supplementary material.

4.2 Comparisons with the State-of-the-arts

In this part, we evaluate the performance of the proposed SSCL framework with classification accuracy under different label noise rates (NR). As shown in Table 1, the proposed SSCL significantly outperforms the state-of-the-art baselines on almost all evaluation metrics. For example, our SSCL achieves the highest classification accuracy on ISIC-19 and BUSI datasets, which can verify the effectiveness of our SSCL for medical image analysis with noisy supervision. Especially in the case of higher noise, SSCL has a more powerful capability to capture discriminative and robust features and minimize the effect of noisy



(a) Parameter analysis (b) Ablation studies

Fig. 2. Comparison of accuracy (%) in (a) parameter analysis and (b) ablation studies on CIFAR-10

labels. When the noise rate is 0.8, the proposed SSCL framework achieves a classification accuracy of 93.0% and 66.5% on CIFA-10 and CIFA-100, surpassing Sel-CL by 3.8% and 6.9%. The comparative results consistently demonstrate the superiority and generalizability of the proposed SSCL framework.

4.3 Parameter Analysis and Ablation Studies and Visualizations

As the key hyperparameter in Eq. 7, the threshold τ is designed to reduce the wrong sample pairs to achieve the best classification boundary construction. In this part, we empirically conduct the proposed SSCL framework with a range of different values τ. As shown in Fig. 2(a), our SSCL achieves the best performance when τ is increased to 0.8, which can avoid the adverse impact of noisy pairs. Moreover, we also conduct ablation studies by systematically removing each component within the SSCL. In Fig. 2(b), we can observe that all the modules are necessary for the function of the proposed SSCL framework. To further validate the discriminative power of the SSCL framework, we utilized t-SNE [5] to visualize the features extracted by vanilla ResNet-18 and SSCL. Compared with vanilla ResNet-18, the visualization results in Fig. 3 and Fig. 4 clearly demonstrate that SSCL has better characteristics for clustering and finer classification boundary, which can demonstrate the effectiveness of SSCL.

(a) ResNet:20% (b) SSCL:20% (c) ResNet:50% (d) SSCL:50%

Fig. 3. T-SNE visualization of the features learned by ResNet and SSCL on ISIC19.

(a) ResNet:20% (b) SSCL:20% (c) ResNet:80% (d) SSCL:80%

Fig. 4. T-SNE visualization of the features learned by ResNet and SSCL on CIFAR-10.

5 Conclusion

This paper presents a robust and reliable semi-supervised contrastive learning method that benefits greatly from the potential synergistic efficacy of semi-supervised learning and contrastive learning, which aims to tackle the challenge of learning with medical noisy labels. By explicitly selecting confident samples and pairs, our approach exhibits a powerful ability to learn discriminative feature representations, mitigating the impact of medical label noise. To demonstrate the effectiveness and versatility of our proposed approach in various practical scenarios, our future works would extend the proposed SSCL method to a broader range of real-world noisy datasets and tasks.

Acknowledgment. This work was supported in part by NSFC fund 62176077, in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515010057, in part by Shenzhen Science and Technology Program (Grant No. RCYX20221008092852077), in part by the Shenzhen Key Technical Project under Grant 2022N001, in part by the Shenzhen Fundamental Research Fund under Grant JCYJ20210324132210025, and in part by the Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies (2022B1212010005).

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data Brief **28**, 104863 (2020)
2. Arazo, E., Ortego, D., Albert, P., O'Connor, N., McGuinness, K.: Unsupervised label noise modeling and loss correction. In: Proceedings of the International Conference on Machine Learning, pp. 312–321 (2019)
3. Balaram, S., Nguyen, C.M., Kassim, A., Krishnaswamy, P.: Consistency-based semi-supervised evidential active learning for diagnostic radiograph classification. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention, pp. 675–685 (2022)
4. Chen, B., Zhang, Z., Li, Y., Lu, G., Zhang, D.: Multi-label chest x-ray image classification via semantic similarity graph embedding. IEEE Trans. Circ. Syst. Video Technol. **32**(4), 2455–2468 (2021)
5. Chen, B., Zhang, Z., Lu, Y., Chen, F., Lu, G., Zhang, D.: Semantic-interactive graph convolutional network for multilabel image recognition. IEEE Trans. Syst. Man Cybern. Syst. **52**(8), 4887–4899 (2021)

6. Harutyunyan, H., Reing, K., Ver Steeg, G., Galstyan, A.: Improving generalization by controlling label-noise information in neural network weights. In: Proceedings of the International Conference on Machine Learning, pp. 4071–4081 (2020)
7. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9729–9738 (2020)
8. Hendrycks, D., Mazeika, M., Wilson, D., Gimpel, K.: Using trusted data to train deep networks on labels corrupted by severe noise. Adv. Neural Inf. Process. Syst. **31**, 1–10 (2018)
9. Hinton, G., et al.: Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process. Maga. **29**(6), 82–97 (2012)
10. Imran, A.A.Z., Wang, S., Pal, D., Dutta, S., Zucker, E., Wang, A.: Multimodal contrastive learning for prospective personalized estimation of CT organ dose. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention, pp. 634–643 (2022)
11. Ji, S., Zhang, Z., Ying, S., Wang, L., Zhao, X., Gao, Y.: Kullback-leibler divergence metric learning. IEEE Trans. Cybern. **52**(4), 2047–2058 (2020)
12. Jiang, M., Yang, H., Li, X., Liu, Q., Heng, P.A., Dou, Q.: Dynamic bank learning for semi-supervised federated image diagnosis with class imbalance. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention, pp. 196–206 (2022)
13. Khosla, P., et al.: Supervised contrastive learning. Adv. Neural Inf. Process. Syst. **33**, 18661–18673 (2020)
14. Lee, S., Lee, H., Yoon, S.: Adversarial vertex mixup: toward better adversarially robust generalization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 272–281 (2020)
15. Li, S., Xia, X., Ge, S., Liu, T.: Selective-supervised contrastive learning with noisy labels. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 316–325 (2022)
16. Li, Y., Yang, J., Song, Y., Cao, L., Luo, J., Li, L.J.: Learning from noisy labels with distillation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1910–1918 (2017)
17. Ortego, D., Arazo, E., Albert, P., O'Connor, N.E., McGuinness, K.: Multi-objective interpolation training for robustness to label noise. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6606–6615 (2021)
18. Ramaswamy, H., Scott, C., Tewari, A.: Mixture proportion estimation via kernel embeddings of distributions. In: Proceedings of the International Conference on Machine Learning, pp. 2052–2060 (2016)
19. Reed, S., Lee, H., Anguelov, D., Szegedy, C., Erhan, D., Rabinovich, A.: Training deep neural networks on noisy labels with bootstrapping. arXiv preprint arXiv:1412.6596 (2014)
20. Rizve, M.N., Duarte, K., Rawat, Y.S., Shah, M.: In defense of pseudo-labeling: an uncertainty-aware pseudo-label selection framework for semi-supervised learning. In: Proceedings of the International Conference on Learning Representations (2021)
21. Seyfioğlu, M.S., et al.: Brain-aware replacements for supervised contrastive learning in detection of alzheimer's disease. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention, pp. 461–470 (2022)
22. Sohn, K., et al.: Fixmatch: simplifying semi-supervised learning with consistency and confidence. Adv. Neural Inf. Process. Syst. **33**, 596–608 (2020)

23. Tran, M., Wagner, S.J., Boxberg, M., Peng, T.: S5cl: unifying fully-supervised, self-supervised, and semi-supervised learning through hierarchical contrastive learning. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention, pp. 99–108 (2022)
24. Venkataramanan, S., Kijak, E., Amsaleg, L., Avrithis, Y.: Alignmixup: improving representations by interpolating aligned features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 19174–19183 (2022)
25. Wang, X., Yao, L., Rekik, I., Zhang, Y.: Contrastive functional connectivity graph learning for population-based fmri classification. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention, pp. 221–230 (2022)
26. Wang, Y., Ma, X., Chen, Z., Luo, Y., Yi, J., Bailey, J.: Symmetric cross entropy for robust learning with noisy labels. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 322–330 (2019)
27. Xia, X., et al.: Robust early-learning: hindering the memorization of noisy labels. In: Proceedings of the International Conference on Learning Representations (2021)
28. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: Mixup: beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
29. Zhang, Z., Sabuncu, M.: Generalized cross entropy loss for training deep neural networks with noisy labels, vol. 31 (2018)
30. Zhou, Y., et al.: Learning to bootstrap for combating label noise. arXiv preprint arXiv:2202.04291 (2022)