Active Label Refinement for Robust Training of Imbalanced Medical Image Classification Tasks in the Presence of High Label Noise

Bidur Khanal[1(✉)], Tianhong Dai[3], Binod Bhattarai[3], and Cristian Linte[1,2]

[1] Center for Imaging Science, Rochester Institute of Technology, Rochester, NY, USA
bk9618@rit.edu
[2] Biomedical Engineering, Rochester Institute of Technology, Rochester, NY, USA
[3] University of Aberdeen, Aberdeen, UK

Abstract. The robustness of supervised deep learning-based medical image classification is significantly undermined by label noise in the training data. Although several methods have been proposed to enhance classification performance in the presence of noisy labels, they face some challenges: 1) a struggle with class-imbalanced datasets, leading to the frequent overlooking of minority classes as noisy samples; 2) a singular focus on maximizing performance using noisy datasets, without incorporating experts-in-the-loop for actively cleaning the noisy labels. To mitigate these challenges, we propose a two-phase approach that combines Learning with Noisy Labels (LNL) and active learning. This approach not only improves the robustness of medical image classification in the presence of noisy labels but also iteratively improves the quality of the dataset by relabeling the important incorrect labels, under a limited annotation budget. Furthermore, we introduce a novel Variance of Gradients approach in the LNL phase, which complements the loss-based sample selection by also sampling under-represented examples. Using two imbalanced noisy medical classification datasets, we demonstrate that our proposed technique is superior to its predecessors at handling class imbalance by not misidentifying clean samples from minority classes as mostly noisy samples. Code available at: https://github.com/Bidur-Khanal/imbalanced-medical-active-label-cleaning.git.

Keywords: Active label cleaning · Label noise · Learning with noisy labels (LNL) · Medical image classification · Imbalanced data · Active learning · Limited budget

B. Bhattarai and C. Linte—These authors share equal senior authorship.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-72120-5_4.

M. G. Linguraru et al. (Eds.): MICCAI 2024, LNCS 15011, pp. 37–47, 2024.
https://doi.org/10.1007/978-3-031-72120-5_4

1 Introduction

Label noise poses a significant hurdle in the robust training of classifiers for medical image datasets, as it can distort the supervised learning process and compromise generalizability [9,12]. In real-world scenarios, factors such as the lack of quality annotation [18], the use of NLP algorithms to extract labels from test reports [8], and the reliance on pseudo labels [13] lead to high label noise in datasets. The impact of label noise is particularly severe in imbalanced medical datasets, where the class distribution is skewed [14]. In recent years, numerous approaches, collectively referred to as Learning with Noisy Labels, have been proposed to train classifiers robustly in the presence of noisy labels [7,11,17]. These methods often employ a sample selection strategy based on the big-loss hypothesis, which suggests that samples with low incurred loss are likely to be clean, to distinguish clean samples from noisy ones. However, this simple hypothesis alone fails with highly imbalanced datasets, where minority or hard samples are mistakenly interpreted as noisy, necessitating a robust alternative.

Medical datasets are often highly imbalanced due to the varying prevalence of conditions or diseases, with some being rarer than others. For example, Dermatofibroma occurs less frequently than other skin conditions, so it is often under-represented in the dataset. Although there have been attempts to enhance robustness against noisy labels in imbalanced datasets [14,22], the performance still falls short of that achieved without noisy labels. Furthermore, a model trained solely on noisy labels is unlikely to be trusted for medical inference and has limited potential for improvement unless data is cleaned. Therefore, establishing a two-step mechanism to initially train optimally on a noisy, imbalanced dataset and then progressively correct labels over time to improve performance is crucial.

One such strategy involves incorporating the experts-in-the-loop to selectively relabel important noisy samples within a limited annotation budget over time. This approach is akin to active learning, which aims to label the most important examples from a pool of unlabeled samples to maximize task performance [3]. However, active learning assumes the existence of some initial accurately labeled data to train the first model, which is unfeasible with a noisy dataset, making it likely to fail without such a set from the start. A practical method should optimally train with the noisy dataset and then learn to identify the samples that need relabeling based on noise statistics. Several machine learning papers have proposed methods to actively clean labels [6,16,23]. Bernhardt *et al.* [2] proposed an active label-cleaning method for noisy Chest X-ray datasets with multiple annotators. However, this method, specifically designed for multi-annotator scenarios, struggles with highly imbalanced datasets.

In this work, we propose an approach to address two key challenges: robustly training a classifier on a noisy, imbalanced dataset and gradually cleaning important noisy samples by incorporating experts-in-the-loop to enhance classifier performance. For robust training, we modified the loss-based sample selection strategy used in LNL, which separates clean and noisy samples based on sample loss [7,15,17], by incorporating a Variance of Gradients-based selection. The loss-

based approach alone struggles with imbalanced datasets, as underrepresented samples often exhibit high loss values, leading to their mis-selection as noisy. The Variance of Gradients-based selection is robust against such bias and complements the loss-based method, compensating for the potential exclusion of underrepresented samples.

To summarize our contributions: (1) *we propose an active label cleaning pipeline to iteratively clean noisy labels for medical image classification by incorporating experts-in-the-loop under a limited annotation budget*; (2) *we introduce a novel Variance of Gradients-based example selection strategy to complement loss-based clean label selection, aiming to better handle highly underrepresented samples in highly imbalanced datasets with high label noise*; (3) *we demonstrate that our proposed method outperforms its predecessor baseline methods, while limiting the annotation budget, as shown using both the imbalanced ISIC-2019 and long-tailed NCT-CRC-HE-100K datasets.*

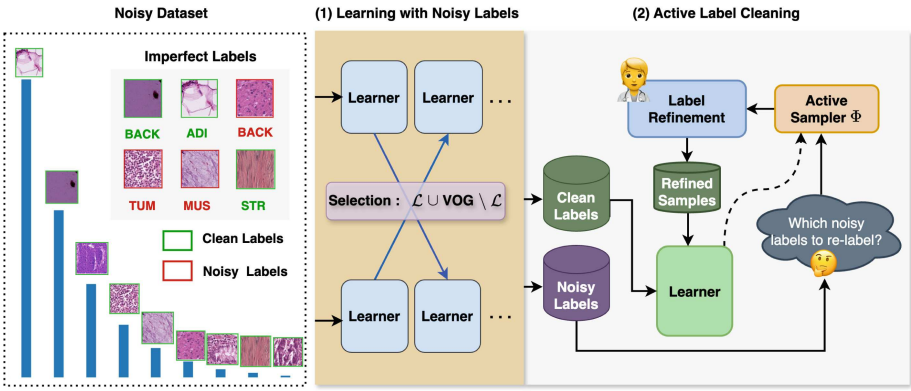

Fig. 1. Active Label Cleaning Pipeline: 1) Learning with Noisy Labels (LNL), where the clean-noisy selection process includes selections from both small Variance of Gradient (VOG) and small loss ($\mathcal{L}$) criteria; 2) Active Label Cleaning, wherein the noisy samples discarded by LNL are iteratively sampled using an active sampler ($\mathbf{\Phi}$) and relabeled.

2 Methods

2.1 Label Noise Injection

We denote an imbalanced classification dataset as $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, where N is the total number of instances, x_i is an input and $y_i \in \{1, 2, 3, ..., C\}$ is the corresponding true label. Noisy dataset is created by injecting label noise by randomly flipping the true label y_i with another class label $\hat{y}_i$ by a probability p (i.e., noise rate), such that $\hat{y}_i \overset{p}{\sim} \{1, 2, 3, ..., C\} \setminus \{y_i\}$.

2.2 Overall Pipeline

Overall, this is a two-stage pipeline as described in Fig. 1. In the first stage, we apply LNL to robustly train the model in the presence of noisy labels while concurrently identifying clean samples from the dataset. In the second stage, we use these clean samples to train a new model. Meanwhile, the remaining samples are ranked by an active sampler based on their importance using a ranking function. After ranking, the top a_l samples are cleaned in each annotation round and used to train the model again until the total annotation budget A_l is exhausted.

2.3 LNL Using Variance of Gradient

In this first phase, we robustly train our model using an LNL method. The typical loss-based sample selection used in LNL to separate noisy samples from clean samples does not handle underrepresented samples in imbalanced datasets well, often misidentifying them as noisy samples.

We used a novel approach to regularize sample selection by using the Variance of Gradients (VOG) [1], instead of relying solely on loss-based selection. Similar to loss, VOG can also separate samples with clean labels from noisy ones. Unlike loss-based selection, VOG estimates the change in gradients over epochs rather than making selections based on statistics from a single epoch, thereby avoiding potential bias. The original paper [1] computes the VOG of each sample at the image level and averages all the pixels to obtain a scalar value. This approach is unscalable as the dataset size and input-image resolution increase. Following [21], we compute the VOG at the feature level, which significantly reduces the memory footprint (example: from $256{\times}256{\times}N$ gradients to $512{\times}N$ gradients, where 512 is the dimension of the ResNet18 feature representation).

Mathematically, let's assume $S_{ij} \in \mathbb{R}^D$ is the gradient vector ($S_{ij} = \frac{\partial \mathcal{A}^l_{y_i}}{\partial x_i}$), for a sample x_i at an epoch j, where $i \in \{1, 2, ..N\}$, $j \in \{1, 2, ..E\}$, and $\mathcal{A}^l_{y_i}$ is the class activation w.r.t given label y_i. Here, N and E represent the number of data samples and the number of epochs, respectively, while D is the dimension of the gradient vector, equal to the feature dimension. Each sample x_i has a gradient vector computed at various epochs, i.e. $\{S_{i1}, S_{i2}, ...S_{iE}\}$. The Variance of Gradient (VOG) of x_i, at an epoch j, is given by:

$$VOG_{ij} = \frac{1}{D} \sum_{d=1}^{D} \sqrt{\frac{1}{t} \sum_{e=j-t}^{j} (S_{ie} - \mu_i)^2} \tag{1}$$

where $\mu_i = \frac{1}{t} \sum_{e=j-t}^{j} S_{ie}$ and t is the number of previous epochs used to compute the variance. If $t = 5$, VOG can be computed only after the 5^{th} epoch.

Co-teaching [7] is a loss-based selection approach that separates clean samples from noisy labels as $Cl = \{b \in B \mid \mathcal{L}(b) \text{ are the } R \text{ smallest values}\}$, where B is the mini-batch and $\mathcal{L}$ is the loss value, and R is the number of examples to be selected as clean given by $R = \lfloor (1 - \tau) * B \rfloor$. Usually, the forget rate τ is

chosen to match the noise rate p. In our approach, i.e., Co-teaching VOG, we select clean samples as: $Cl = \{b_1 \in B \mid \mathcal{L}(b_1)$ are the R_1 smallest values$\} \cup \{b_2 \in B \setminus b_1 \mid VOG(b_2)$ are the R_2 smallest values$\}$, where $R_1 = \lfloor (1-m) * R \rfloor$, $R_2 = \lfloor m * R \rfloor$, and m is a hyperparameter we refer to as mix ratio. When $m = 0$, no examples are selected using VOG. We only employ the VOG after a warm-up phase, because VOG is unstable in the early phase. The noisy subset is given by $\hat{Cl} = B \setminus Cl$. At the end of the training, we combine the samples selected at each mini-batch to obtain all the clean and noisy samples from the entire dataset containing N samples. Let $\hat{\mathbf{Cl}}$ represent the noisy samples set and $\mathbf{Cl}$ represent the clean samples set, from the whole dataset.

2.4 Active Label Cleaning

The first stage identifies $\mathbf{Cl}$ clean samples, while the remaining noisy samples $\hat{\mathbf{Cl}}$ undergo a label correction in this phase. We have a predefined annotation budget A_l that denotes the number of examples that we can afford to relabel. The noisy samples are annotated in batches up to M annotation rounds, at the rate of a_l samples per round. We apply an active learning sampler to select the most important samples which, after label correction, would improve test performance with fewer annotation rounds, given by: $L = \text{argmax}_{L \subseteq \hat{\mathbf{Cl}}, |L| = a_l} \mathbf{\Phi}(x | x \in \hat{\mathbf{Cl}})$, where $\mathbf{\Phi}$ is the scoring function. We then pass the selected samples to an expert annotator for relabeling: $L_{clean} = \mathcal{ORACLE}(L)$. After cleaning L samples, we update the noisy set and the clean set as $\hat{\mathbf{Cl}} : \hat{\mathbf{Cl}} \setminus L_{clean}$, and $\mathbf{Cl} : \mathbf{Cl} \cup L_{clean}$, respectively. After each annotation round, the original model is retrained with the updated clean set.

3 Experiments

3.1 Datasets

Long-tailed NCT-CRC-HE-100K: We created a long-tailed dataset from the original NCT-CRC-HE-100K [10] by modifying the class distribution. The original dataset has 100,000 histopathology images for training and 7,180 for testing, with nine classes. To create a long-tailed version, we randomly sampled examples from the training set using the Pareto distribution [5]: $N_c = N_0(r^{-(k-1)})^c$, where k is the total number of classes and c is the class being sampled. Here, we chose $N_0 = \min(N_0, N_1, \ldots, N_{k-1})$, representing the class with the minimum number of samples in the original dataset, and $r = 100$ as the imbalance factor. After creating the imbalanced dataset, we divided the training set into training and validation sets with split ratios of 0.8 and 0.2, respectively. Consequently, the final long-tailed training set contains 15,924 samples, and the validation set contains 3,982 samples, while the original test set remains unchanged. Label noise is injected only in the training set.

ISIC-2019: ISIC-2019[1] is an imbalanced dataset, which comprises 25,331 RGB images, each belonging to one of eight skin disease conditions. We divided the original dataset into training, validation, and test sets randomly, using split ratios of 0.7, 0.1, and 0.2, respectively. As a result, the training, validation, and test sets contain 17,731, 2,533, and 5,067 samples, respectively, where label noise is only injected into the training set.

3.2 Baselines

We compared our proposed method – Co-teaching VOG (CT_{VOG}) + Active Learning (Random, Entropy, Coreset [20]), where we robustly train the model using Co-teaching, regularized by VOG in sample selection – against the following baselines: 1) Active learning (Random and Entropy [4]), where we directly clean a few samples, exhausting annotation budget a_l, and train the initial base model. Then, we gradually clean additional samples, selected by active learning at each round, and fine-tune the model. This approach does not involve training with noisy labels in the initial phase. 2) Cross-Entropy (CE) + Active Learning (Random, Entropy), where we initially train the model using the noisy dataset, then gradually clean the samples selected by AL and fine-tune the model using only cleaned data. 3) ALC w/ Co-teaching (Bernhardt *et al.* [2]), where the model was fine-tuned using Co-teaching in each round using available data (both cleaned and still noisy). At each round, it uses a ranker function to select the samples to be cleaned. Since this method is primarily proposed for multi-annotator settings, we adapted it to our single-annotator setting and implemented it.

In our method, the samples selected in the initial phase as clean are retained and used to train the classifier using standard cross-entropy loss. The remaining noisy samples selected via active learning are gradually cleaned and added to this clean set at each round, while simultaneously tuning the model. *Evaluation: Since all our datasets are imbalanced, we evaluate the performance of our method and all the baseline performance using the macro-averaged F1-score, which captures both precision and recall, in the test set.* The test score is computed in the epoch where the validation set performed the best.

3.3 Implementation Details

We used ResNet18 pretrained on ImageNet as the feature extractor backbone for all our experiments. The batch size (b) was set to 256, and we trained the model with the SGD optimizer, an initial learning rate of 0.01, a momentum of 0.9, weight decay of 1×10^{-4}, and cosine scheduler. These parameters remained consistent across both datasets. Images from both datasets were resized to 244×224, and we applied basic data augmentations, including random crop, random flip, random Gaussian blur, and random color jittering. We selected two high noise rates, $p = \{0.4, 0.5\}$ and $p = \{0.7, 0.8\}$, for ISIC-2019 and Long-tailed NCT-CRC-HE-100K, respectively, the rates at which the classification performance degradation is high.

[1] https://challenge.isic-archive.com/landing/2019/.

For Co-teaching VOG, we set the warm-up epoch to 10. The number of instances selected as clean is determined by the forget rate, which depends on the maximum forget threshold τ and the decay rate c. Following [7], we set $\tau = p$ and $c = 1$, where p represents the label noise rate. The mix ratio (m) for Co-teaching VOG is a hyperparameter that depends on the dataset and noise rate. We found $m = 0.2$ and $m = 0$ to work best for ISIC-2019 at $p = 0.5$ and $p = 0.4$, respectively. Similarly, $m = 1$ was optimal for Long-tailed NCT-CRC-HE-100K for both $p = 0.8$ and $p = 0.7$.

In the active label cleaning phase, we set the annotation round (M) to 8. The per-round annotation budget (a_l) varied for both datasets and label noise rates. In ISIC-2019, $a_l = \{384, 492\}$ for $p = \{0.4, 0.5\}$, and in Long-tailed NCT-CRC-HE-100K, $a_l = \{70, 80\}$ for $p = \{0.7, 0.8\}$. We ran experiments across three seeds to obtain the average and standard deviation. All the training sessions were performed using the PyTorch 1.12.1 framework in Python 3.9 on a single A100 GPU.

4 Results

4.1 Overall Active Label Cleaning

In Fig. 2 and Fig. 3, we benchmark the performance of our approach against various baseline methods on the ISIC-2019 and the long-tailed NCT-CRC-HE-100K datasets, respectively. Our LNL strategy (CT_{VOG}) delivers a substantial performance uplift from the beginning, prior to any label cleaning efforts. Then, the active learning strategies, Entropy and Coreset, effectively select important noisy samples to re-label. These samples further enhanced the model's performance. Since strategies that rely solely on active learning (Random/Entropy) initially require clean samples to train the model, they consequently exhaust a round budget from the start. Initially, training with noisy labels using standard cross-entropy (CE) proves more advantageous than just solely relying on active learning from the beginning. However, the performances converge to solely using active learning in the later stages, with additional label-cleaning rounds. We observed no further improvement upon cleaning additional labels when using ALC with Co-teaching (Bernhardt *et al.* [2]). This method, proposed for multi-annotator settings and not intended for imbalanced datasets, still selected the same initial examples as clean, even after the noisy labels had been cleaned.

It is important to note that the macro-averaged test F1-score after training on the ISIC-2019 dataset with entirely clean labels is 0.767 ± 0.004. Our method achieves this performance by relabeling merely 3,152 samples at a noise rate of 0.4 and 3,936 samples at a noise rate of 0.5, out of 17,731 training examples. Similarly, the macro-averaged test F1-score for the long-tailed NCT-CRC-HE-100K dataset is 0.894 ± 0.12 with all clean labels. Our approach matches this score by relabeling just 300 samples at a noise rate of 0.7 and 400 samples at a noise rate of 0.8, from a total of 15,924 available training samples.

4.2 VOG as a Regularizer for LNL

In Table 1, we investigate the benefits of integrating VOG into Co-teaching (CT_{VOG}) for enhancing the identification of underrepresented samples. In ISIC-2019 at a noise rate of $p = 0.5$, we observed that Co-teaching alone tends to overlook minority classes while identifying clean samples (see class DF). By regularizing the sample selection with VOG, the accuracy of identifying samples from underrepresented classes improves, resulting in enhanced performance in LNL at the initial phase.

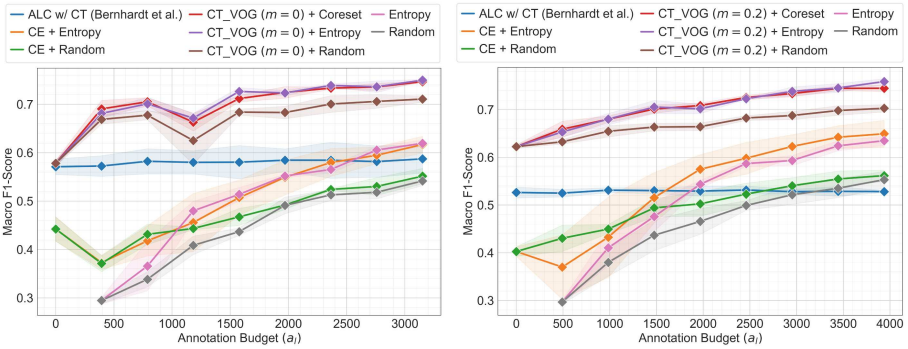

Fig. 2. Comparison of the macro-averaged test F1-score across various baselines in ISIC-2019 dataset at two noise rates: $p = 0.4$ (left) and $p = 0.5$ (right).

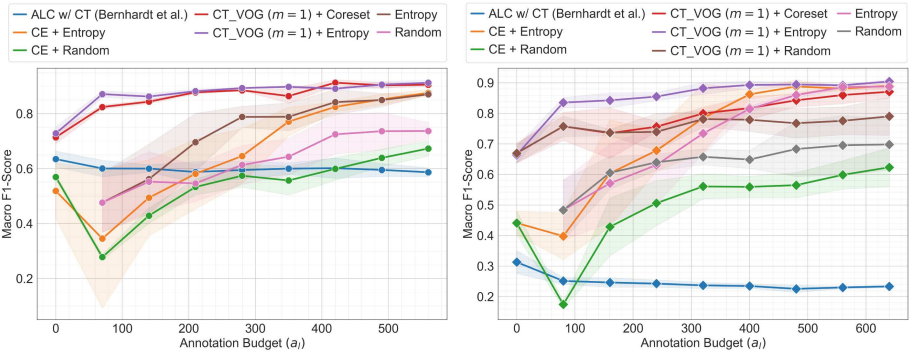

Fig. 3. Comparison of the macro-averaged test F1-score across various baselines in Long-tailed NCT-CRC-HE-100K dataset at two noise rates: $p = 0.7$ (left) and $p = 0.8$ (right).

Table 1. Comparing LNL performance of Co-teaching (CT) alone vs. Co-teaching with VOG as a regularizer (CT_{VOG}) on the ISIC-2019 dataset with a label noise rate ($p = 0.5$). We examine the Recall and Guess %, which indicate the proportion of samples identified as clean that belong to underrepresented classes.

	DF		VASC		SCC	
LNL	Recall	Guess (%)	Recall	Guess (%)	Recall	Guess (%)
CT	0.00 ± 0.00	0.00 ± 0.00	0.39 ± 0.34	50.73 ± 43.94	0.26 ± 0.22	36.11 ± 31.30
CT_{VOG}	0.10 ± 0.09	20.00 ± 17.33	0.57 ± 0.00	80.44 ± 3.92	0.37 ± 0.02	52.38 ± 2.06

5 Discussion and Conclusion

In this work, we present a strategy that combines learning with noisy labels and active learning to actively relabel noisy samples, thereby enhancing medical image classification performance in the presence of noisy labels. Our method of regularizing Co-teaching with VOG for sample selection has proven to handle imbalanced cases better. We show that by relabeling only a few samples, our method can match the performance achieved with clean labels in the ISIC-2019 and long-tailed NCT-CRC-HE-100K datasets.

While our method shows promising results compared to the baseline, some limitations in our work could be addressed in future research. First, we limited our study to a single CNN-based model. Exploring the behavior of larger models would be an interesting extension. Additionally, we adhered to common standard protocols by limiting our study to a uniform label noise distribution and specific noise rates. Investigating the performance of our method under different types of noise would provide further insights. Finally, it would also be valuable to examine how our method performs across various levels of skewness in imbalanced distributions.

Acknowledgments. Research reported in this publication was supported by the NIGMS Award No. R35GM128877 of the National Institutes of Health, and by OAC Award No. 1808530 and CBET Award No. 2245152, both of the National Science Foundation, and by the Aberdeen Startup Grant CF10834-10. We also acknowledge Research Computing at the Rochester Institute of Technology [19] for providing computing resources.

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

References

1. Agarwal, C., D'souza, D., Hooker, S.: Estimating example difficulty using variance of gradients. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022)

2. Bernhardt, M., Castro, D.C., Tanno, R., Schwaighofer, A., Tezcan, K.C., Monteiro, M., Bannur, S., Lungren, M.P., Nori, A., Glocker, B., et al.: Active label cleaning for improved dataset quality under resource constraints. Nature communications (2022)
3. Budd, S., Robinson, E.C., Kainz, B.: A survey on active learning and human-in-the-loop deep learning for medical image analysis. Medical Image Analysis (2021)
4. Cohn, D.A., Ghahramani, Z., Jordan, M.I.: Active learning with statistical models. Journal of artificial intelligence research (1996)
5. Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S.: Class-balanced loss based on effective number of samples. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (2019)
6. Goh, H.W., Mueller, J.: Activelab: Active learning with re-labeling by multiple annotators. In: ICLR Workshop on Trustworthy ML (2023)
7. Han, B., Yao, Q., Yu, X., Niu, G., Xu, M., Hu, W., Tsang, I., Sugiyama, M.: Co-teaching: Robust training of deep neural networks with extremely noisy labels. Advances in neural information processing systems (2018)
8. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., et al.: Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI conference on artificial intelligence (2019)
9. Karimi, D., Dou, H., Warfield, S.K., Gholipour, A.: Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. Medical image analysis (2020)
10. Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.A., Gaiser, T., Marx, A., Valous, N.A., Ferber, D., et al.: Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS medicine (2019)
11. Khanal, B., Bhattarai, B., Khanal, B., Linte, C.A.: Improving medical image classification in noisy labels using only self-supervised pretraining. In: MICCAI Workshop on Data Engineering in Medical Imaging. Springer (2023)
12. Khanal, B., Hasan, S.K., Khanal, B., Linte, C.A.: Investigating the impact of class-dependent label noise in medical image classification. In: Medical Imaging 2023: Image Processing. SPIE (2023)
13. Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., et al.: The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. International Journal of Computer Vision (2020)
14. Li, J., Cao, H., Wang, J., Liu, F., Dou, Q., Chen, G., Heng, P.A.: Learning robust classifier for imbalanced medical image dataset with noisy labels by minimizing invariant risk. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer (2023)
15. Li, J., Socher, R., Hoi, S.C.: Dividemix: Learning with noisy labels as semi-supervised learning. arXiv preprint arXiv:2002.07394 (2020)
16. Lin, C., Mausam, M., Weld, D.: Re-active learning: Active learning with relabeling. In: Proceedings of the AAAI Conference on Artificial Intelligence (2016)
17. Liu, J., Li, R., Sun, C.: Co-correcting: noise-tolerant medical image classification via mutual label correction. IEEE Transactions on Medical Imaging (2021)
18. Ørting, S.N., Doyle, A., van Hilten, A., Hirth, M., Inel, O., Madan, C.R., Mavridis, P., Spiers, H., Cheplygina, V.: A survey of crowdsourcing in medical image analysis. Human Computation (2020)

19. Rochester Institute of Technology: Research computing services (2022), https://www.rit.edu/researchcomputing/
20. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: International Conference on Learning Representations (2018)
21. Shin, S., Bae, H., Shin, D., Joo, W., Moon, I.C.: Loss-curvature matching for dataset selection and condensation. In: International Conference on Artificial Intelligence and Statistics. PMLR (2023)
22. Xue, C., Yu, L., Chen, P., Dou, Q., Heng, P.A.: Robust medical image classification from noisy labeled data with global and local representation guided co-training. IEEE Transactions on Medical Imaging (2022)
23. Zeni, M., Zhang, W., Bignotti, E., Passerini, A., Giunchiglia, F.: Fixing mislabeling by human annotators leveraging conflict resolution and prior knowledge. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (2019)