

# SSP-RACL: Classification of Noisy Fundus Images with Self-Supervised Pretraining and Robust Adaptive Credal Loss

1<sup>st</sup> Mengwen Ye

Academy for Engineering and Technology  
Fudan University  
Shanghai, China  
mwyee22@m.fudan.edu.cn

3<sup>rd</sup> You Li

Academy for Engineering and Technology  
Fudan University  
Shanghai, China  
youli22@m.fudan.edu.cn

2<sup>nd</sup> Yingzi Huangfu

Academy for Engineering and Technology  
Fudan University  
Shanghai, China  
fyzhuang23@m.fudan.edu.cn

4<sup>th</sup> Zekuan Yu\*

Academy for Engineering and Technology  
Fudan University  
Shanghai, China  
yzk@fudan.edu.cn

**Abstract**—Fundus image classification is crucial in the computer aided diagnosis tasks, but label noise significantly impairs the performance of deep neural networks. To address this challenge, we propose a robust framework, Self-Supervised Pre-training with Robust Adaptive Credal Loss (SSP-RACL), for handling label noise in fundus image datasets. First, we use Masked Autoencoders (MAE) for pre-training to extract features, unaffected by label noise. Subsequently, RACL employ a superset learning framework, setting confidence thresholds and adaptive label relaxation parameter to construct possibility distributions and provide more reliable ground-truth estimates, thus effectively suppressing the memorization effect. Additionally, we introduce clinical knowledge-based asymmetric noise generation to simulate real-world noisy fundus image datasets. Experimental results demonstrate that our proposed method outperforms existing approaches in handling label noise, showing superior performance.

**Index Terms**—Noisy labels, Fundus images, Self-supervised learning, Superset learning

## I. Introduction

Fundus image classification is pivotal in diagnosing various ocular diseases [1]. Researchers need a large amount of labeled fundus images to improve the classification performance of deep neural networks. Given the limited availability and high costs of annotated medical image data, self-supervised learning has demonstrated significant maturity in the field of fundus image classification by utilizing unlabeled data to enhance network generalization capabilities. Yet, fundus image annotation quality can be compromised by expertise, demanding years of specialized training and domain knowledge. Mislabeling is also likely due to the high feature similarity and image quality variations. Consequently, fundus image datasets often

inevitably contain label noise [2]. Fundus images with noisy labels can confuse supervised training, resulting in the model overfitting noisy labels and severely reducing classification and generalization performance [3]. Given the high demands for computer-aided diagnosis of ocular disease, addressing the issue of noisy labels is crucial.

In recent years, researchers have proposed various methods to address the problem of noisy labels in natural images datasets, including regularization methods [4]–[6], loss function design [7]–[9], sample selection strategies [3], [10], [11], label correction techniques [12]–[14], self-supervised learning [15] and integrated approaches [16]–[18]. Fundus datasets often contain hard samples and are imbalanced, so these methods for natural images can not achieve good performance on fundus datasets. Classification on highly noisy, imbalanced fundus image datasets is still a challenging task. Despite its clinical importance, research on addressing label noise in fundus images remains limited. Several methods have been proposed to tackle this issue. For example, [19] proposed an uncertainty estimation-based framework to handle disagreement label noise from inconsistent expert opinions and single-target label noise from biased aggregation of individual annotations. [20] addressed label noise by combining data cleaning, adaptive negative learning, and sharpness-aware minimization. [21] employed a method involving cost-sensitive classification constraints and atomic sub-task modeling, resulting in a 3-5% improvement in quadratic-weighted kappa scores at a negligible computational cost. However, these methods are often complex and require additional precisely annotated set.

To bridge this gap, we propose a novel method, Self-Supervised Pre-training and Robust Adaptive Credal Loss (SSP-RACL), for robust classification of noisy fun-

Key Laboratory of Medical Electronics and Digital Health of Zhejiang Province (MEDC202301) and National Student Entrepreneurship Practice Program of Fudan University (202310246001S)

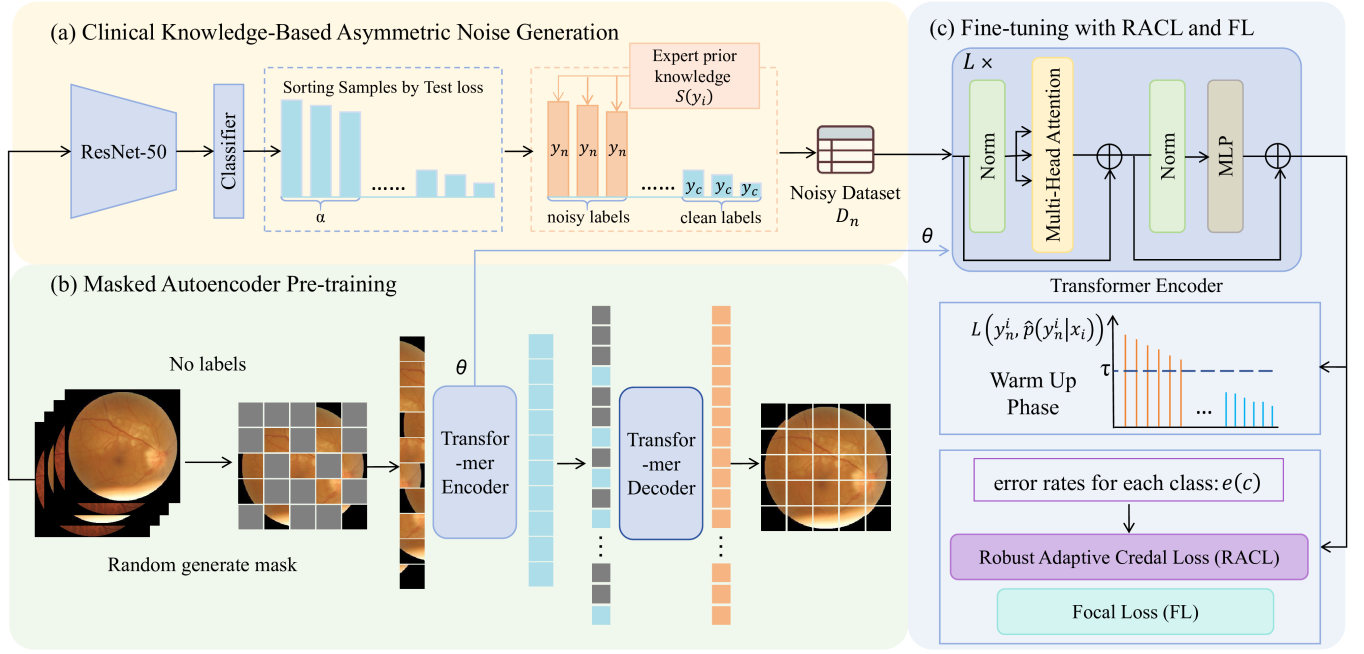


Fig. 1. The structure of the overall framework

dus images. Self-supervised pre-training (SSP) is robust against label noise and encompasses four strategies: innate relationship, generative, contrastive, and self-prediction [22]. We adopt for the Masked Autoencoder (MAE) approach, classified under the generative strategy, due to its notable scalability. Our approach is simple but effective, and does not rely on additional evaluation set. More importantly, it significantly enhances the ability to handle noisy data, ensuring the reliability of the model in clinical applications.

## II. Methodology

Initially, we employed clinical knowledge-based asymmetric noise generation techniques to create a simulated real-world noisy dataset. Subsequently, we employed MAE on this noisy dataset to ensure that the encoder captures fundamental features unaffected by label noise. Following this, we fine-tuned the model by integrating Robust Adaptive Credal Loss (RACL) with Focal Loss (FL). We innovatively propose an adaptive adjustment of the label relaxation parameter  $\alpha$ , offering increased flexibility in handling label noise. Additionally, the incorporation of FL further emphasizes focus on hard-to-classify samples. Fig.1 illustrates the structure of the overall framework.

### A. Clinical Knowledge-Based Asymmetric Noise Generation

In studies of label noise in natural images, datasets are typically configured to symmetric noise and pairflip noise. However, fundus images annotated by physicians often contain systematic errors, which are mostly asymmetric.

For example, in diabetic retinopathy grading tasks, the similar appearance of adjacent grades makes them more prone to misdiagnosis. To simulate these specific misdiagnosis patterns known in clinical practice, we propose clinical knowledge-based Asymmetric Noise Generation. Firstly, 30% of the samples in the dataset are randomly selected according to the original category proportions to form the training set. This subset is used to train a pre-trained ResNet-50 model. Next, the samples in test set are evaluated by the cross-entropy loss function. We select samples with the highest loss at a proportion  $r_n$  as candidates for noisy samples as high loss typically indicates low confidence in model predictions. These samples are likely misdiagnosed due to poor image quality, unclear features, or incorrect original labels. When generating noisy labels, instead of randomly assigning a different class to each selected sample, labels are adjusted based on expert knowledge. The noisy dataset  $D_n$  is defined as follows:

$$D_n = \{(x_i, y_i)\}_{i=1}^N \quad (1)$$

where  $x_i$  represents the sample, and  $y_i$  is the corresponding label, defined by:

$$y_i = \begin{cases} y_c & 1 - r_n \\ y_n \text{ randomly selected from } S(y_i) & r_n \end{cases} \quad (2)$$

where  $S(y_i)$  is the set of potential misdiagnosis labels determined based on expert knowledge,  $y_n$  represents the noisy label randomly selected from this set,  $y_c$  is the clean label, and  $r_n$  is the noise rate, indicating the probability that each sample is assigned a noisy label.

## B. Masked Autoencoder Pre-training

Recent studies have demonstrated the effectiveness of MAE in pre-training for natural image analysis [23]. By reconstructing complete images from partially masked inputs, the Vision Transformer encoder aggregates contextual information to infer the masked regions of the image. This capability to integrate contextual details is particularly crucial in fundus image analysis, as different anatomical structures are interdependent and closely connected. Therefore, we employ MAE to improve fundus image classification tasks with noisy labels. The fundus images are divided into non-overlapping patches, and around 80% of patches are randomly masked to reduce redundant information and encourage the model to learn global features. For each unmasked patch, a token is generated through linear projection and the addition of positional embeddings. Subsequently, the tokens undergo a shuffle operation to randomize their order, and then the top 20% are selected. These selected tokens serve as input for the Transformer encoder, as depicted in Fig. 1. These feature vectors are then concatenated with vectors from positional information of masked patches. The decoder receives these concatenated vectors as input, and the output of the decoder is transformed back to the pixel scale through a linear layer.

## C. Robust Adaptive Credal Loss

Algorithm 1 details the steps of the RACL algorithm which is used to handle noisy labels in downstream classification tasks [24], [25]. In traditional probabilistic learning settings, deterministic target labels  $y_i \in \mathcal{Y}$  are transformed into degenerate probability distributions  $p_{y_i} \in \mathbb{P}(\mathcal{Y})$ :

$$p_{y_i} = \begin{cases} 1 & \text{if } y = y_i \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

This representation implies full plausibility for the observed training label  $y_i$ , while regarding other labels as fully implausible. However, traditional methods completely rely on individual label information, making them overly sensitive to noisy labels and consequently degrading model performance. The credal set is an effect approach to address this issue. Representing the target distribution as a set of possible distributions, rather than a single point estimate, credal sets account for the uncertainty associated with each training instance. This approach enables the model to consider a range of potential true distributions, accommodating scenarios where the observed label may be incorrect. Rather than treating a mislabeled instance as a definitive representation of the true class, credal sets allow the model to consider alternative labels that may better align with the underlying data distribution, enhancing the model's robustness. Based on possibility theory [26], we define the possibility distribution  $\pi_i(y)$  for  $x_i$  via confidence threshold. This assigns a probability to each class in  $\mathcal{Y}$ , considering them as candidates for the true

outcomes associated with  $x_i$ .  $\pi_i(y)$  can be interpreted as an upper bound on the true probability  $p^*(y | x_i)$  given  $x_i$ . Specifically,  $\pi_i(y)$  is set to 1 when  $y$  equals the observed label  $y_i$  or the model's predicted probability  $\hat{p}(y | x_i)$  exceeds the threshold  $\beta$ , thereby considering this label as a fully credible candidate. For all other cases,  $\pi_i(y)$  is set to  $\alpha$ , indicating that even for labels with predicted probabilities below  $\beta$ , their possibility is not completely ruled out but instead assigned a lower possibility value. The mathematical representation is as follows:

$$\pi_i(y) = \begin{cases} 1 & \text{if } y = y_i \text{ or } \hat{p}(y|x) \geq \beta \\ \alpha(c) & \text{otherwise} \end{cases} \quad (4)$$

where  $\beta, \alpha(c) \in [0, 1]$ .  $\alpha(c)$  is a label relaxation parameter that is dynamically tuned based on the error rates  $e(c)$  of each class. We capitalize on the characteristic that clean samples have lower losses while noisy samples result in higher losses during the warm-up phase of model training. To calculate  $e(c)$  for each class, we use cross-entropy loss to assess the loss for each training sample during this warm-up phase. The error rate for a class is then defined as the proportion of samples from that class whose loss exceeds a threshold  $\tau$ :

$$e(c) = \frac{|\{x \in \mathcal{D}_c : L(y, \hat{p}(y|x)) > \tau\}|}{|\mathcal{D}_c|} \quad (5)$$

where  $\mathcal{D}_c$  is the set of samples belonging to class  $c$ . Then update  $\alpha(c)$  for each class based on its error rate:

$$\alpha(c) = \frac{k}{e(c) + \epsilon} \quad (6)$$

where  $k$  and  $\epsilon$  are constants chosen to scale and stabilize the adjustments, ensuring  $\alpha(c)$  remains within  $[0, 1]$ . We can derive the credal targets  $Q_\pi$  induced by  $\pi_i(y)$ , which define the set of all possible probability distributions  $p$  that satisfy the condition: for any subset  $Y \subseteq \mathcal{Y}$ , the sum of probabilities for all events within  $Y$  does not exceed the maximum possibility value corresponding to the events in  $\pi_i$ :

$$Q_{\pi_i} := \left\{ p \in \mathbb{P}(\mathcal{Y}) \mid \forall Y \subseteq \mathcal{Y} : \sum_{y \in Y} p(y) \leq \max_{y \in Y} \pi_i(y) \right\} \quad (7)$$

Probability distribution  $p^r$  that can be obtained by projecting the predicted probability  $\hat{p}$  onto the boundary of the credal set  $Q_{\pi_i}$ . This projection adjusts  $\hat{p}$  to ensure it complies with the constraints defined by the possibility distribution  $\pi_i$ , which specifies the upper bounds of probabilities for each class.  $p^r(c)$  is given as follows:

$$p^r(y) = \begin{cases} (1 - \alpha(c)) \cdot \frac{\hat{p}(y)}{\sum_{y' \in \mathcal{Y} : \pi_i(y')=1} \hat{p}(y')} & \text{if } \pi_i(y) = 1 \\ \alpha(c) \cdot \frac{\hat{p}(y)}{\sum_{y' \in \mathcal{Y} : \pi_i(y')=\alpha(c)} \hat{p}(y')} & \text{otherwise} \end{cases} \quad (8)$$

For the dataset  $\{(x_i, Q_{\pi_i})\}$  based on  $Q_{\pi_i}$ , we train the model by minimizing the Optimistic Superset Loss (OSL) [27]. The core of OSL lies in selecting the probability distribution  $p$  from the credal set  $Q_\pi$  that minimizes the

loss function  $\mathcal{L}(p, \hat{p})$ . Specifically, it involves calculating the loss associated with all potential distributions  $p$  within  $\mathcal{Q}_\pi$  and selecting the smallest of these losses for optimization. Furthermore, the computational complexity of the OSL is equivalent to common loss functions, and the function is convex [28], which means that the associated optimization problem can be effectively solved, guaranteeing the identification of a global optimum. The loss is given as follows:

$$\mathcal{L}^*(\mathcal{Q}_\pi, \hat{p}) = \begin{cases} 0 & \text{if } \hat{p} \in \mathcal{Q}_\pi \\ D_{KL}(p^r \parallel \hat{p}) & \text{otherwise} \end{cases}, \quad (9)$$

where Kullback-Leibler Divergence  $D_{KL}$  is used to measure the difference between two probability distributions. Through RACL, we can effectively reduce the model's reliance on incorrect training labels. Additionally, it enables us to be more cautious during prediction, avoiding erroneous judgments about potential mislabeled data. Therefore, this approach decreases the dependency on learning from noisy labels, while ensuring that the model's performance with clean samples remains unaffected.

---

#### Algorithm 1 Robust Adaptive Credal Loss (RACL)

---

Require: Training instance  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ , model prediction  $\hat{p}(x) \in \mathbb{P}(\mathcal{Y})$ , confidence threshold  $\beta \in [0, 1]$ , Error rates for each class  $e(c)$ , Adjustment parameter  $k, c$

- 1: Adjust  $\alpha(c)$  based on class-specific error rates  $e(c)$  as specified in (6)
- 2: Calculate  $\pi$  as specified in (4)
- 3: Calculate the credal set  $\mathcal{Q}_\pi$  derived from  $\pi$  as specified in (7)
- 4: return  $\mathcal{L}^*(\mathcal{Q}_\pi, \hat{p}(x))$  as specified in (9)

---

### III. Experiment

#### A. Datasets

We conducted experiments on two fundus image datasets namely the OIA-ODIR dataset [29] and the real-world dataset Kaggle Diabetic Retinopathy (DR) dataset [30]. The OIA-ODIR dataset consists of 10,000 fundus images, and contains 8 classes of labels: normal, hypertensive retinopathy, glaucoma, cataract, age-related macular degeneration, hypertension complications, pathologic myopia, and other diseases/abnormalities. Noise was generated using our proposed clinical knowledge-based Asymmetric Noise Generation. The Kaggle DR dataset consists of 34,423 fundus photographs: one photograph per eye corresponding to the DR grading labels, and it is estimated that the initially published dataset has approximately 30% noisy labels in the initially released dataset [19]. In the experiment, the distribution of data across the training, validation, and test sets was configured in a ratio of 6:1:3.

#### B. Implementation Details

All images were preprocessed to a size of 224x224. In the MAE pre-training phase, the mask ratio was set to 0.8, the batch size was set to 64, and the learning rate was set to  $1 \times 10^{-4}$ . Other hyperparameters of the MAE were set according to [23]. For the fine-tuning phase, the learning rate was set to  $1 \times 10^{-5}$ , with a total of 30 training epochs. During the warmup phase, the learning rate was set to  $1 \times 10^{-6}$ , and the number of warmup epochs was set to 5. The value of  $\beta$  in RACL was determined using the decay function described in [25]. All experiments were conducted using PyTorch on an NVIDIA GeForce RTX 3090 GPU.

#### C. Comparison with the State-of-the-Arts

We compared our SSP-RACL framework with several state-of-the-art methods on the OIA-ODIR and Kaggle DR datasets, including Generalized Cross-Entropy (GCE) [7], CORES [8], and Co-teaching [3]. The performance metrics used for comparison were Average Precision (AP), Area Under the Curve (AUC), Precision, F1 score, and Recall. Detailed performance comparisons are shown in Table I and Table II, demonstrating that our method achieves the best performance across most metrics.

TABLE I  
Performance comparison on the OIA-ODIR dataset

Noise rate $r_n$	0.1		0.2	
	AP	AUC	AP	AUC
GCE [7]	67.1	85.8	60.3	83.6
CORES [8]	63.2	83.1	56.7	81.2
Co-teaching [3]	71.5	85.5	69.6	83.1
Ours	72.4	86.4	71.9	84.3

TABLE II  
Performance comparison on the kaggle DR dataset

	Precision	F1 score	Recall
GCE [7]	72	65.9	60.7
CORES [8]	71.3	64.7	59.3
Co-teaching [3]	80.4	77.9	75.6
Ours	82.7	80.6	78.6

#### D. Ablation Study

To evaluate the contribution of each component in our SSP-RACL framework, we performed an ablation study by systematically removing different components and measuring the performance impact. The results, as shown in Table III, demonstrate that both RACL and SSP significantly contribute to the overall performance of our model.

#### Conclusion

In our paper, we propose an SSP-RACL robust framework to address label noise in fundus image datasets. The clinical knowledge-based Asymmetric Noise Generation method simulates label noise as observed in clinical

TABLE III  
Ablation study with  $r_n = 0.1$

SSP	RACL	AUC
×	×	81.7
✓	×	84.3
×	✓	85.2
✓	✓	86.4

practice. The MAE captures the intrinsic structure of the data, reducing dependence on precise annotations. The RACL adjusts label relaxation parameters and focuses on hard-to-classify samples, effectively handling label noise. Experiments on OIA-ODIR and Kaggle DR datasets demonstrated significant performance improvements. Future work will further optimize these techniques and explore their application to other medical imaging tasks.

#### Acknowledgment

This study was sponsored by Key Laboratory of Medical Electronics and Digital Health of Zhejiang Province (MEDC202301) and National Student Entrepreneurship Practice Program of Fudan University (202310246001S).

#### References

- [1] O. Ouda, E. Abdelmaksoud, A. Abd El-Aziz, and M. Elmogy, "Multiple ocular disease diagnosis using fundus images based on multi-label deep learning classification," *Electronics*, vol. 11, no. 13, p. 1966, 2022.
- [2] C. Zhu, W. Chen, T. Peng, Y. Wang, and M. Jin, "Hard sample aware noise robust learning for histopathology image classification," *IEEE transactions on medical imaging*, vol. 41, no. 4, pp. 881–894, 2021.
- [3] B. Han, Q. Yao, X. Yu, G. Niu, M. Xu, W. Hu, I. Tsang, and M. Sugiyama, "Co-teaching: Robust training of deep neural networks with extremely noisy labels," *Advances in neural information processing systems*, vol. 31, 2018.
- [4] X. Zhou, X. Liu, C. Wang, D. Zhai, J. Jiang, and X. Ji, "Learning with noisy labels via sparse regularization," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 72–81.
- [5] S. Liu, J. Niles-Weed, N. Razavian, and C. Fernandez-Granda, "Early-learning regularization prevents memorization of noisy labels," *Advances in neural information processing systems*, vol. 33, pp. 20331–20342, 2020.
- [6] Y. Lu, Y. Bo, and W. He, "Confidence adaptive regularization for deep learning with noisy labels," *arXiv preprint arXiv:2108.08212*, 2021.
- [7] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in neural information processing systems*, vol. 31, 2018.
- [8] H. Cheng, Z. Zhu, X. Li, Y. Gong, X. Sun, and Y. Liu, "Learning with instance-dependent label noise: A sample sieve approach," *arXiv preprint arXiv:2010.02347*, 2020.
- [9] L. Feng, S. Shu, Z. Lin, F. Lv, L. Li, and B. An, "Can cross entropy loss be robust to label noise?" in *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, 2021, pp. 2206–2212.
- [10] X. Xia, B. Han, Y. Zhan, J. Yu, M. Gong, C. Gong, and T. Liu, "Combating noisy labels with sample selection by mining high-discrepancy examples," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 1833–1843.
- [11] D. Patel and P. Sastry, "Adaptive sample selection for robust learning under label noise," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 3932–3942.
- [12] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa, "Joint optimization framework for learning with noisy labels," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5552–5560.
- [13] Y. Lu and W. He, "Selc: self-ensemble label correction improves learning with noisy labels," *arXiv preprint arXiv:2205.01156*, 2022.
- [14] Y. Wu, J. Shu, Q. Xie, Q. Zhao, and D. Meng, "Learning to purify noisy labels via meta soft label corrector," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, 2021, pp. 10388–10396.
- [15] H. Cheng, Z. Zhu, X. Sun, and Y. Liu, "Demystifying how self-supervised features improve training from noisy labels," 2021.
- [16] Y. Li, H. Han, S. Shan, and X. Chen, "Disc: Learning from noisy labels via dynamic instance-specific selection and correction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 24070–24079.
- [17] M. Sheng, Z. Sun, Z. Cai, T. Chen, Y. Zhou, and Y. Yao, "Adaptive integration of partial label learning and negative learning for enhanced noisy label learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 5, 2024, pp. 4820–4828.
- [18] P. Albert, E. Arazo, T. Krishna, N. E. O'Connor, and K. McGuinness, "Is your noise correction noisy? pls: Robustness to label noise with two stage detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 118–127.
- [19] L. Ju, X. Wang, L. Wang, D. Mahapatra, X. Zhao, Q. Zhou, T. Liu, and Z. Ge, "Improving medical images classification with label noise using dual-uncertainty estimation," *IEEE transactions on medical imaging*, vol. 41, no. 6, pp. 1533–1546, 2022.
- [20] T. Hu, B. Yang, J. Guo, W. Zhang, H. Liu, N. Wang, and H. Li, "A fundus image classification framework for learning with noisy labels," *Computerized Medical Imaging and Graphics*, vol. 108, p. 102278, 2023.
- [21] A. Galdran, J. Dolz, H. Chakor, H. Lombaert, and I. Ben Ayed, "Cost-sensitive regularization for diabetic retinopathy grading from eye fundus images," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23*. Springer, 2020, pp. 665–674.
- [22] S.-C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung, and A. S. Chaudhari, "Self-supervised learning for medical image classification: a systematic review and implementation guidelines," *NPJ Digital Medicine*, vol. 6, no. 1, p. 74, 2023.
- [23] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16000–16009.
- [24] J. Lienen, C. Demir, and E. Hüllermeier, "Conformal credal self-supervised learning," in *Conformal and Probabilistic Prediction with Applications*. PMLR, 2023, pp. 214–233.
- [25] J. Lienen and E. Hüllermeier, "Mitigating label noise through data ambiguity," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 12, 2024, pp. 13799–13807.
- [26] D. Dubois and H. Prade, "Possibility theory, probability theory and multiple-valued logics: A clarification," *Annals of mathematics and Artificial Intelligence*, vol. 32, pp. 35–66, 2001.
- [27] E. Hüllermeier, "Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization," *International Journal of Approximate Reasoning*, vol. 55, no. 7, pp. 1519–1534, 2014.
- [28] J. Lienen and E. Hüllermeier, "From label smoothing to label relaxation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 10, 2021, pp. 8583–8591.
- [29] N. Li, T. Li, C. Hu, K. Wang, and H. Kang, "A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection," in *Benchmarking, Measuring, and Optimizing: Third BenchCouncil International Symposium, Bench 2020, Virtual Event, November 15–16, 2020, Revised Selected Papers 3*. Springer, 2021, pp. 177–193.
- [30] Kaggle, "Diabetic retinopathy detection," 2015, accessed: 2024-08-19. [Online]. Available: <https://www.kaggle.com/c/diabetic-retinopathy-detection/data>