

Article

Enhanced Semi-Supervised Medical Image Classification Based on Dynamic Sample Reweighting and Pseudo-Label Guided Contrastive Learning (DSRPGC)

Kun Liu ¹ , Ji Liu ^{1,*}  and Sidong Liu ² 

¹ School of Information Engineering, Shanghai Maritime University, Shanghai 200135, China; kunliu@shmtu.edu.cn

² Australia Institute of Health Innovation, Macquarie University, Sydney 2109, Australia; sidong.liu@mq.edu.au

* Correspondence: 202230310042@stu.shmtu.edu.cn; Tel.: +86-198-5561-2898

Abstract: In semi-supervised learning (SSL) for medical image classification, model performance is often hindered by the scarcity of labeled data and the complexity of unlabeled data. This paper proposes an enhanced SSL approach to address these challenges by effectively utilizing unlabeled data through a combination of pseudo-labeling and contrastive learning. The key contribution of our method is the introduction of a Dynamic Sample Reweighting strategy to select reliable unlabeled samples, thereby improving the model's utilization of unlabeled data. Additionally, we incorporate multiple data augmentation strategies based on the Mean Teacher (MT) model to ensure consistent outputs across different perturbations. To better capture and integrate multi-scale features, we propose a novel feature fusion network, the Medical Multi-scale Feature Fusion Network (MedFuseNet), which enhances the model's ability to classify complex medical images. Finally, we introduce a pseudo-label guided contrastive learning (PGC) loss function that improves intra-class compactness and inter-class separability of the model's feature representations. Extensive experiments on three public medical image datasets demonstrate that our method outperforms existing SSL approaches, achieving 93.16% accuracy on the ISIC2018 dataset using only 20% labeled data, highlighting the potential of our approach to advance medical image classification under limited supervision.



Citation: Liu, K.; Liu, J.; Liu, S. Enhanced Semi-Supervised Medical Image Classification Based on Dynamic Sample Reweighting and Pseudo-Label Guided Contrastive Learning (DSRPGC). *Mathematics* **2024**, *12*, 3572. <https://doi.org/10.3390/math12223572>

Academic Editor: Gintautas Dzemyda

Received: 22 October 2024

Revised: 10 November 2024

Accepted: 14 November 2024

Published: 15 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: semi-supervised learning; medical image classification; consistency regularization; feature fusion; contrastive learning

MSC: 68T45

1. Introduction

Deep learning has achieved significant advancements in medical image classification [1], primarily due to the availability of large-scale, high-quality, labeled datasets [2,3]. However, acquiring accurate and well-labeled medical image data is both time-consuming and costly in real-world applications. This challenge is particularly evident in clinical settings, where vast amounts of unlabeled medical images are available, but expert annotation remains expensive and labor-intensive. The difficulty is further compounded by the diversity of imaging modalities and pathological conditions. For example, the ISIC 2018 dataset, which focuses on dermoscopic images of skin lesions, and the NCT-CRC-HE dataset, containing histopathological images of colorectal cancer tissues, present distinct challenges in image classification. These images often suffer from missing labels, such as tumor grades, disease stages, or specific pathological features, which complicates the learning process. Additionally, rare diseases, such as early-stage melanoma or specific colorectal cancer subtypes, often comprise only a small fraction of the available data, making it difficult to obtain sufficient labeled samples for accurate diagnosis. Similarly, the Chest X-ray14

dataset, which contains over 112,120 frontal-view X-ray images of 14 different thoracic diseases, presents its own set of challenges. While it provides a large volume of labeled data, it suffers from class imbalance, with certain diseases being underrepresented in the dataset. Moreover, X-ray images may vary in quality due to differences in acquisition equipment or imaging protocols, further complicating classification tasks. The presence of noise and variations in image quality, combined with the challenge of obtaining high-quality labels for rare conditions like tuberculosis or interstitial lung diseases, makes this dataset particularly challenging for training robust machine learning models. Thus, the imbalance between the abundance of unlabeled data and the scarcity of labeled data presents a significant barrier to developing robust deep learning models. Consequently, semi-supervised learning (SSL) has emerged as a promising approach to leverage this abundant unlabeled data [4,5], aiming to bridge this gap and enhance model performance despite challenges associated with noisy and unreliable annotations, especially when labels are missing for critical features such as tumor subtypes or early disease markers.

The primary challenge in applying semi-supervised learning (SSL) to medical image classification lies in the noise present in unlabeled data and the potential for generating inaccurate pseudo-labels, particularly when data distributions are imbalanced or exhibit diverse pathology types [6,7]. Pseudo-labeling methods aim to utilize existing labeled data to train a model, which then makes predictions on unlabeled data to generate pseudo-labels. The effectiveness of this approach heavily depends on the reliability of these pseudo-labels—if they are inaccurate, the model can be adversely affected by propagating errors during training, leading to degraded performance [8,9]. Specifically, models are often not fully trained at the outset, making the reliability of the generated pseudo-labels difficult to guarantee, which exacerbates the issue of error propagation.

To mitigate the challenges posed by pseudo-labeling, advanced SSL methods have incorporated consistency regularization. This approach encourages the model to produce consistent outputs for the same input under different data augmentations, thereby improving robustness against noise and variations. For instance, the Mean Teacher model [10] utilizes a moving average of model weights to generate consistent predictions, significantly enhancing performance. Additionally, the Virtual Adversarial Training (VAT) method [11] creates adversarial examples to impose consistency constraints, further stabilizing the model's predictions. However, both methods exhibit notable limitations: pseudo-labeling is overly reliant on the initial quality of labeled data—if these initial labels are inaccurate, subsequent training results can be significantly compromised [12]. Meanwhile, consistency regularization may struggle to maintain performance in the presence of strong noise or outliers, particularly in scenarios where the data volume is limited, raising concerns about model stability [13]. Furthermore, existing methods often face difficulties in effectively capturing information across different spatial scales in unlabeled medical images, further limiting their practical applicability [14].

In this paper, we propose a novel Dynamic Sample Reweighting and Pseudo-label Guided Contrastive Learning Framework (DSRPGC), specifically designed to address the challenges faced by existing semi-supervised learning (SSL) methods. Our main contributions are summarized as follows:

1. **Dynamic Sample Reweighting Strategy (DSR):**
We are the first to combine meta-learning with gradient optimization to dynamically adjust the weights of unlabeled samples. In contrast to traditional methods with fixed weight allocation, the DSR module adaptively increases the weights of high-confidence samples, effectively reducing the impact of noisy samples on pseudo-label quality. This dynamic adjustment not only enhances the accuracy of pseudo-labels but also effectively reduces the risk of error propagation, thereby improving the model's robustness in uncertain environments.
2. **Medical Multi-scale Feature Fusion Network (MedFuseNet):**
To address the challenges of multi-scale medical image analysis, we propose MedFuseNet, which integrates feature information across different scales to improve

classification performance. Unlike traditional SSL methods that utilize single-scale features, MedFuseNet extracts both low-level and high-level features and fuses them through similarity matrices and element-wise addition. This comprehensive approach captures intricate details in medical images, leading to significantly improved classification accuracy, especially in complex medical imaging scenarios.

3. Pseudo-label Guided Contrastive Loss (PGC Loss):

We propose a PGC loss module that employs a momentum update strategy to optimize feature representations in contrastive learning. Unlike existing contrastive learning methods, our approach leverages pseudo-labels to guide the selection of positive and negative sample pairs, enabling high-quality feature representations even with a large proportion of unlabeled data. In contrast to methods that rely heavily on labeled data, our PGC loss improves robustness and classification accuracy, especially in noisy data environments.

The integration of DSR, MedFuseNet, and PGC Loss in the DSRPGC Framework significantly improves pseudo-label quality, leverages multi-scale features, and enhances model robustness. Our approach provides a more comprehensive and effective solution for medical image classification compared to conventional SSL methods, especially in challenging conditions with limited labeled data and noisy unlabeled data.

Finally, we evaluate our method on three different medical image datasets and conduct ablation studies to investigate the impact of each component. The experimental results demonstrate the effectiveness of our method on multi-class datasets. The paper is structured as follows: Section 2 discusses related work on semi-supervised learning and contrastive learning; Section 3 provides a detailed description of our proposed method; Section 4 presents and analyzes the experimental results; Section 5 offers a discussion on our method; and the final section concludes the paper.

2. Related Work

2.1. Consistency Regularization Methods

In recent years, various semi-supervised learning (SSL) methods [15,16] have increasingly found applications in medical image classification. Among these, consistency regularization methods stand out as they rely on the smoothness assumption to leverage unlabeled data by enhancing the model's prediction consistency under different perturbations [17,18]. For instance, Liu et al. [19] focus on ensuring semantic consistency across different samples, which helps the model discover additional semantic details from unlabeled data. A comprehensive review by Shakya [20] highlights various deep semi-supervised learning (DSSL) techniques and categorizes them into six main approaches, one of which is consistency regularization. The study emphasizes the role of consistency regularization in improving model generalization and addressing challenges such as limited labeled data and dataset heterogeneity in medical image classification. Our method, however, differentiates itself by addressing this problem through the elimination of Gaussian noise and the incorporation of diverse data augmentation strategies, thereby reinforcing consistency training. Despite the success of consistency regularization in semi-supervised learning, medical image datasets often encounter challenges such as class imbalance, noise, and misinformation. Consequently, to better utilize unlabeled data and improve model performance, consistency regularization methods frequently combine with other techniques to address these issues.

2.2. Pseudo-Labeling Methods

Pseudo-labeling techniques generate confidence-based prediction distributions through entropy minimization [21,22], which are then used as training targets alongside the standard cross-entropy loss function [23]. For example, Liu et al. [24] innovated by selecting unlabeled samples based on their information content, rather than using traditional threshold-based pseudo-label selection methods, and introduced Anti-Curriculum Pseudo-Labeling (ACPL), which combines deep learning with a KNN classifier to estimate pseudo-labels, thereby improving training balance and label accuracy. Zeng et al. [25] introduced a robust

technique to differentiate between high- and low-quality pseudo-labeled data by applying adversarial noise at the feature level, based on the principle that clean data tend to exhibit lower loss compared to noisy data. Mahmood et al. [26] utilized a weighted combination of class prototypes and classifiers to predict reliable pseudo-labels for unlabeled images and introduced alignment loss to reduce the model's bias towards the majority class. Despite these advancements, such methods face limitations: substantial amounts of data below the threshold are discarded, leading to underutilization of unlabeled data, and the model's initial training stages are constrained by a limited number of unlabeled samples, which impacts overall performance.

2.3. Contrastive Learning Methods

Contrastive learning enhances classifier performance on unlabeled data by maximizing similarity between samples of the same class and minimizing it between samples of different classes, helping the model understand and differentiate various medical image features. Techniques like SimCLR (Contrastive Learning of Visual Representations) [27] exemplify well-designed contrastive learning algorithms. In semi-supervised learning, contrastive learning can combine with pseudo-labels to define positive and negative sample pairs. For instance, Khosla et al. [28] proposed a method that enhances feature discrimination by utilizing a contrastive loss to pull together representations of similar labeled samples while pushing apart those of different classes, effectively leveraging both labeled and unlabeled data in semi-supervised medical image classification. Li et al. [29] improved pseudo-labels by applying a smoothness constraint on class probabilities through embeddings, while pseudo-labels regularized the structure of embeddings via graph-based contrastive learning [30]. This approach enhances the model's sensitivity to subtle features in medical images, especially those with low inter-class differentiation. Additionally, Biswas et al. [31] proposed pNNCLR, a refined nearest neighbor-based contrastive learning method, which improves the quality of the support set by introducing pseudo nearest neighbors, resulting in enhanced performance in both image and medical image recognition tasks. Wu et al. [32] further enhance semi-supervised segmentation with federated learning, combining prototype-based pseudo-labeling and contrastive learning to improve performance on COVID-19 X-ray and CT segmentation tasks. Liu et al. [33] proposed a semi-supervised learning framework, Contrastive Mutual Learning with Pseudo-Label Smoothing (CMLP), to address issues like pseudo-label noise and spectral variability in hyperspectral image classification. By combining mutual learning, pseudo-label smoothing, and contrastive learning, the framework enhances feature representation, with a dynamic threshold strategy (DTS) adjusting the use of unlabeled data during training.

3. Method

In this section, we provide a detailed explanation of the algorithm's implementation and functionality. The Enhanced Semi-Supervised Medical Image Classification Framework, as shown in Figure 1, outlines the core process of our method. The framework is built upon the Mean Teacher (MT) model, which utilizes an Exponential Moving Average (EMA) update strategy to update the student network, focusing on unlabeled medical image data. Both the teacher and student networks learn feature representations from the enhanced unlabeled data, which are then used for dynamic weight updates and improved consistency loss calculations.

To enhance clarity, the overall framework is divided into smaller sub-modules, each with its own specific role in the algorithm's functionality. Figure 2 illustrates the Dynamic Sample Reweighting Strategy (DSR), responsible for processing unlabeled data and adaptively adjusting sample weights. This helps to improve pseudo-label quality and reduce the impact of noisy samples on model training. Next, Figure 3 presents the Medical Multi-scale Feature Fusion Network (MedFuseNet), which integrates feature maps of unlabeled images at multiple scales. The fusion of low-level and high-level features enhances the classification accuracy, providing rich representations for further learning tasks. Finally, Figure 4

depicts the Pseudo-label Guided Contrastive Learning (PGC) module. Here, pseudo-labels are used to define positive and negative sample pairs for contrastive learning. The feature representations are updated using momentum, and dynamic weighting is applied to the contrastive learning loss. Additionally, the feature queue is updated, improving the learning effectiveness on the unlabeled data. Each of these modules contributes to the robust performance of the semi-supervised medical image classification task. By following these sub-modules in sequence, the entire framework operates cohesively, ensuring effective pseudo-label generation, high-quality feature fusion, and efficient contrastive learning. Collectively, these components play a critical role in improving the model’s performance on semi-supervised medical image classification tasks.

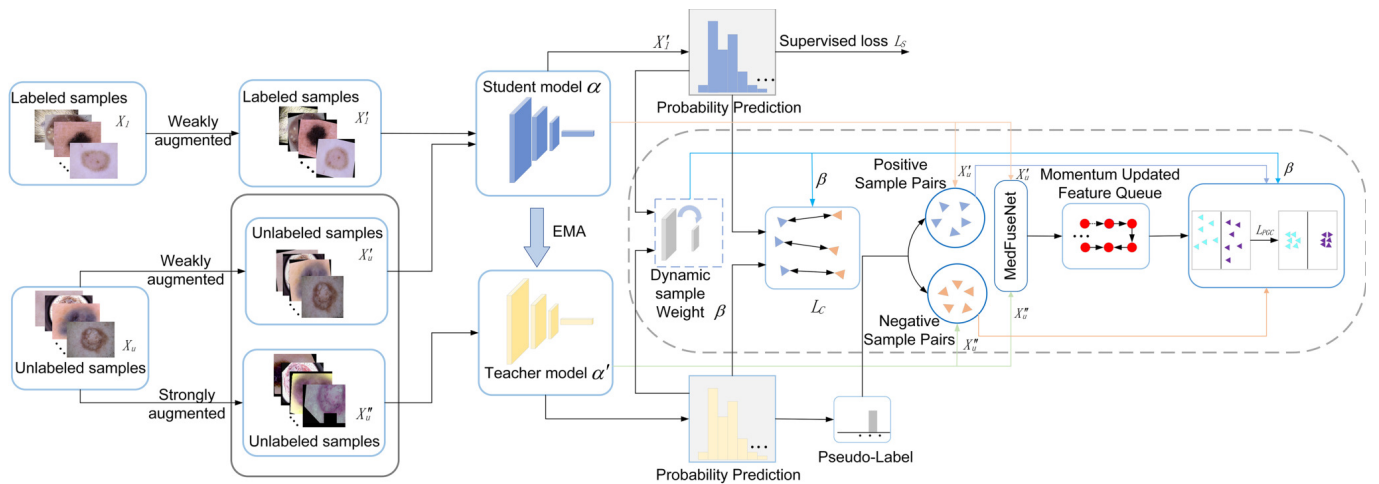


Figure 1. Enhanced Semi-Supervised Medical Image Classification Framework based on Dynamic Sample Reweighting and Pseudo-Label Guided Contrastive Learning.

Outer Loop

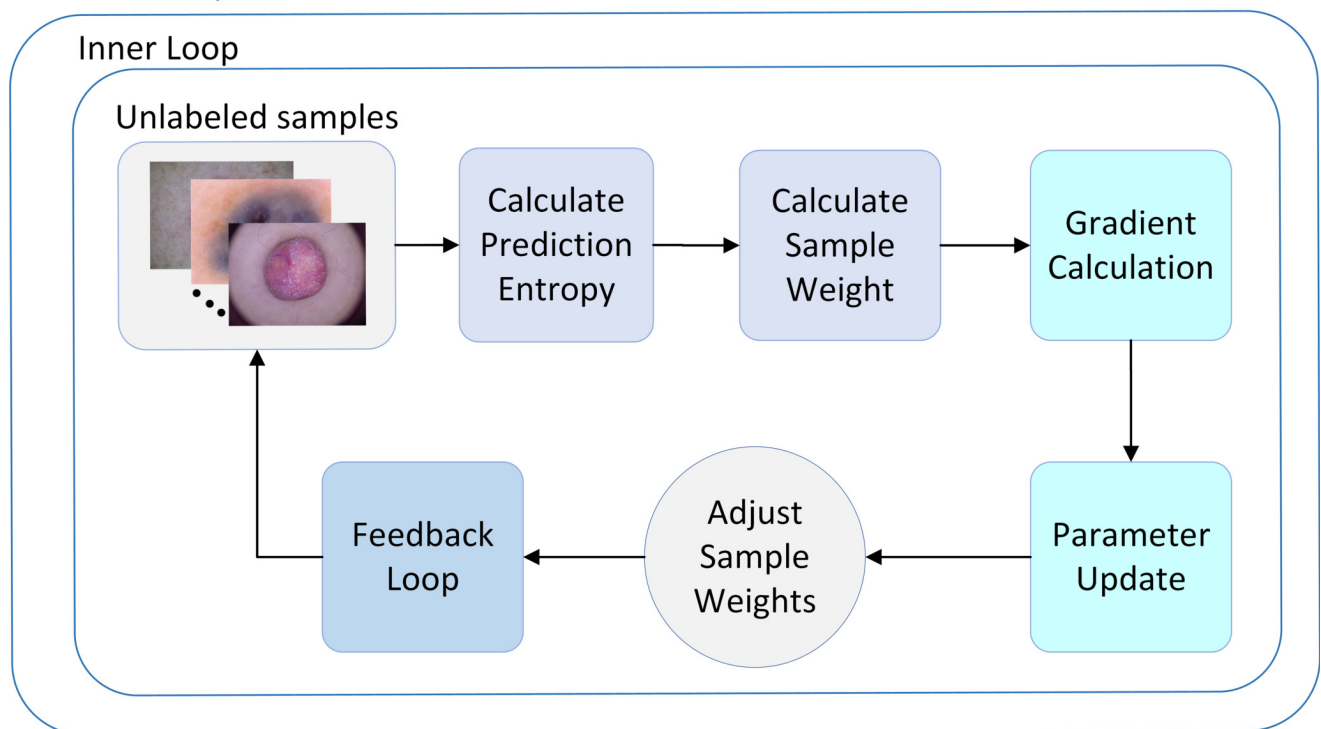


Figure 2. Dynamic Sample Reweighting implementation flow chart.

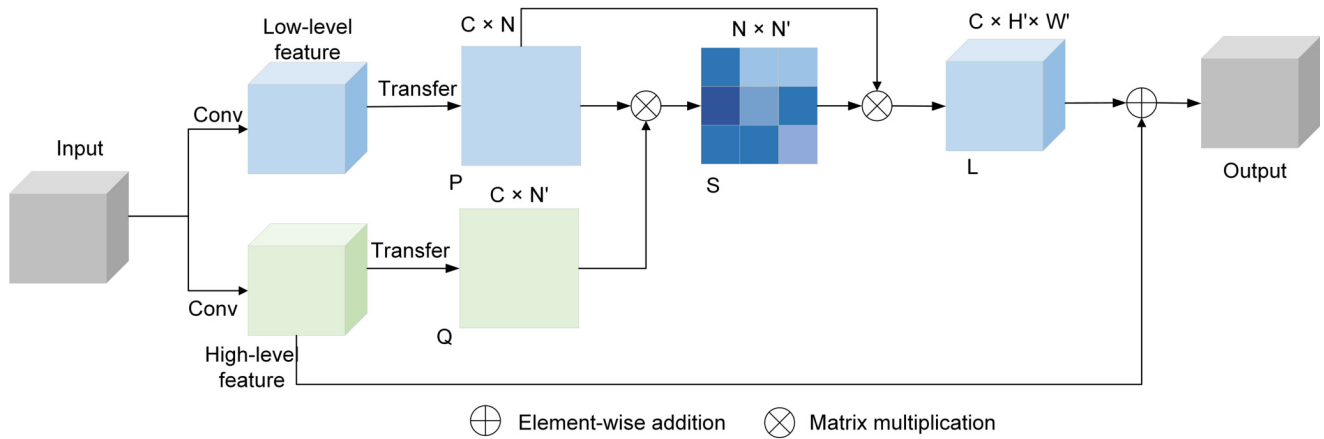


Figure 3. Medical Multi-scale Feature Fusion Network (MedFuseNet) diagram.

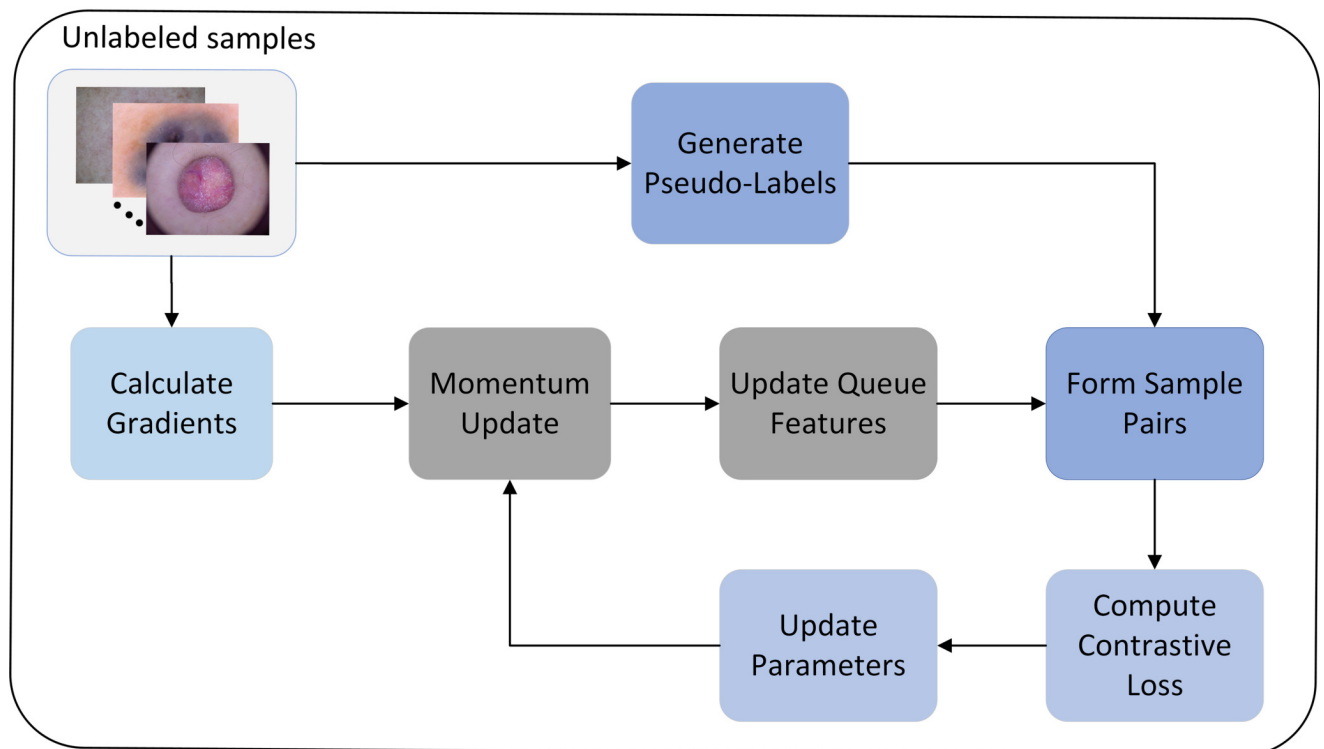


Figure 4. Pseudo-Label Guided Contrastive (PGC) Learning implementation flow chart.

3.1. Dynamic Sample Reweighting Strategy

In semi-supervised medical image classification, leveraging unlabeled data is essential for improving model performance. However, one of the key challenges lies in the potential for error propagation through noisy and imbalanced data, especially in the context of pseudo-labeling. Pseudo-labeling, while effective in utilizing unlabeled data, often suffers from the problem of noisy pseudo-labels, particularly during the early stages of training when the model’s predictions are less accurate. In such situations, the model may generate incorrect pseudo-labels that are subsequently used to train the model, amplifying errors as training progresses. This error propagation is further exacerbated when the data are imbalanced, as the model may rely too heavily on noisy pseudo-labels from underrepresented classes, leading to poor generalization.

To address this issue, we propose a Dynamic Sample Reweighting strategy that adjusts the weight of unlabeled samples based on their predicted reliability. As illustrated in Figure 2, by incorporating meta-learning concepts and gradient descent algorithms, our method dynamically reweights the unlabeled data throughout the training process. This approach helps reduce the influence of unreliable pseudo-labels by giving more weight to samples that the model is confident about. Consequently, this strategy mitigates the risk of error propagation, ensuring that the model's learning process remains stable and reliable, even when faced with noisy or imbalanced data.

Additionally, in our Dynamic Sample Reweighting strategy, the model adjusts the weight of unlabeled samples based on their predicted uncertainty, which is calculated using metrics like entropy. During the early training stages, when pseudo-labels are less reliable, we assign lower weights to uncertain samples, reducing their impact on training and mitigating error propagation. As training progresses and the model's predictions become more accurate, the weights of more reliable pseudo-labels are increased, allowing these samples to contribute more to the training process. This dynamic adjustment ensures that noisy pseudo-labels have minimal influence initially, while the model gradually learns from more confident pseudo-labels, leading to a stable and reliable learning process that can better generalize to unseen data.

Overall optimization loss function. In semi-supervised medical image classification, we aim to utilize both labeled and unlabeled data to improve model performance. To achieve this, we introduce an overall optimization loss function denoted as $L(\alpha, \beta)$. This loss function is designed to balance the learning process across different types of data, where α represents the model parameters and β represents the weights assigned to unlabeled samples based on their reliability. The optimization process incorporates the dynamic reweighting of unlabeled samples, which is crucial in the presence of noisy or imbalanced data. By adjusting the weights β_u based on the predicted entropy, the model focuses more on reliable unlabeled data while minimizing the impact of unreliable samples. This dynamic adjustment helps maintain the stability of the model's learning process, especially in the early training stages, and ensures that the performance gains are robust to noise and data imbalance. The overall optimization loss function can be expressed as

$$L(\alpha, \beta) = L_{out}(\alpha, \beta) + \lambda_{io}L_{in}(\alpha, \beta) \quad (1)$$

where $L_{out}(\alpha, \beta)$ represents the loss function of the outer loop, which focuses on optimizing the meta-parameters β . The outer loop adjusts the weights of the unlabeled samples based on their predicted reliability. Meanwhile, $L_{in}(\alpha, \beta)$ is the inner loop loss function, responsible for updating model parameters α to minimize the training loss on both labeled and weighted unlabeled data. The term λ_{io} is a hyperparameter that balances the contribution of the inner and outer loop objectives [34].

Gradient Calculation. Next, we calculate the gradient of the overall optimization loss function $L(\alpha, \beta)$ with respect to the parameters α and β . The gradients are essential for updating the model parameters and are computed using the following equation:

$$\nabla_{\alpha, \beta}L(\alpha, \beta) = \left(\frac{\partial L(\alpha, \beta)}{\partial \alpha}, \frac{\partial L(\alpha, \beta)}{\partial \beta} \right) \quad (2)$$

where $\nabla_{\alpha, \beta}L$ represents the gradient of the loss function with respect to model parameters α and meta-parameters β . The gradient $\frac{\partial L(\alpha, \beta)}{\partial \alpha}$ indicates how the loss function changes with respect to changes in the model parameters, while $\frac{\partial L(\alpha, \beta)}{\partial \beta}$ shows the sensitivity of the loss to changes in the weights assigned to the unlabeled samples. These gradients are computed using standard backpropagation techniques, allowing for iterative optimization of both labeled and unlabeled data during the training process.

By using this gradient-based approach, the model iteratively adjusts both the parameters α and the weights β of the samples. This two-loop optimization process not only updates the model to fit reliable data but also continuously refines the pseudo-label reliability assessment throughout training. As a result, it significantly reduces the risk of error amplification that often plagues pseudo-labeling approaches in noisy and imbalanced medical datasets.

Parameter update. We use the gradient descent algorithm to iteratively update the model parameters α and weights β as follows:

$$\alpha' = \alpha - l_\alpha \cdot \frac{\partial L(\alpha, \beta)}{\partial \alpha} \quad (3)$$

$$\beta' = \beta - l_\beta \cdot \frac{\partial L(\alpha', \beta)}{\partial \beta} \quad (4)$$

In these equations, l_α and l_β are the learning rates for updating the model parameters α and weights β , respectively. The learning rates control the size of the parameter update step, ensuring stable convergence during training.

In practice, this dynamic reweighting strategy helps the model focus more on unlabeled samples with higher predicted reliability, effectively reducing the influence of noisy pseudo-labels during the early stages of training. By adjusting the sample weights according to the predicted entropy (uncertainty), the model avoids relying heavily on samples with uncertain or erroneous pseudo-labels, thus reducing error propagation. In the early stages of training, when the model's predictions are less accurate and pseudo-labels are noisy, the dynamic weighting mechanism ensures that these uncertain samples are given lower weight, preventing them from negatively affecting the model's learning process. As the model improves and its confidence in the pseudo-labels increases, the weights of the unlabeled samples are dynamically updated, ensuring that only the most reliable samples contribute to the training, which reduces the risk of error propagation from incorrect pseudo-labels. This is particularly important in medical image classification tasks where noisy and imbalanced data are common, as it ensures that the model focuses on more reliable data, leading to better generalization and performance on unseen samples.

Unlabeled Data Weight. For each unlabeled sample x_u , we calculate the predicted entropy to assign a weight to the sample based on its predicted uncertainty. The entropy H of the prediction is given by

$$H(f(x_u, \alpha)) = - \sum_{j=1}^C f(x_u; \alpha)_j \log(f(x_u; \alpha)_j) \quad (5)$$

where $f(x_u; \alpha)_j$ is the predicted probability for the unlabeled sample x_u belonging to class j , and C is the total number of classes.

The weight β_u for each unlabeled sample is then defined as

$$\beta_u = 1 - \frac{H(f(x_u; \alpha))}{\log(C)} \quad (6)$$

This formulation allows us to convert the predicted entropy $H(f(x_u; \alpha))$ into a confidence score β_u . The weight for each unlabeled sample is computed based on its predicted entropy. The entropy reflects the model's uncertainty about the sample's class: higher entropy indicates greater uncertainty, resulting in a lower weight, while lower entropy suggests higher confidence, leading to a higher weight. This dynamic weighting mechanism ensures that the model prioritizes more reliable data and reduces the impact of noisy or incorrect pseudo-labels during training. As the model's confidence improves, the weights of unlabeled samples are updated to better reflect their reliability, minimizing the risk of error propagation from uncertain pseudo-labels.

In the early stages of training, when the model's predictions are less reliable, the entropy is higher for most of the unlabeled samples, meaning they will receive lower weights. This ensures that uncertain samples have minimal influence on the model's learning in the initial phases. As training progresses and the model's confidence increases, the entropy decreases for more confident samples, and the weights are gradually adjusted, allowing the model to learn more effectively from reliable pseudo-labels and reducing the influence of noisy ones.

The entire Dynamic Sample Reweighting strategy is a meta-learning-based parameter update process. The model inputs the weights β for the unlabeled samples and the parameters α of the neural network. The training process involves two nested loops. The inner loop updates the model parameters α to minimize the inner loop loss L_{in} given the current weights β , as shown in Equation (3). In contrast, the outer loop adjusts the weights β based on the updated model parameters α' , to maximize the outer loop loss L_{out} , as shown in Equation (4). This dynamic optimization process ensures that the weights of the unlabeled samples are continuously updated to reflect their reliability, improving the efficiency of unlabeled data utilization during training. Moreover, the use of a meta-learning framework allows the model to iteratively adjust these weights as the training progresses, reducing the reliance on noisy pseudo-labels, particularly in the early training stages. Unlike traditional pseudo-labeling approaches that apply fixed weights, our method continuously updates the weights based on the model's current confidence, thereby mitigating the risk of error propagation. This adaptive strategy ultimately enhances the performance of the semi-supervised medical image classification model, particularly in improving the consistency loss and pseudo-label-guided contrastive learning loss.

3.2. Enhancing Consistency Regularization

We apply consistency regularization to train the model with the goal of generating similar predictions from different perturbations of the same image. In our approach, we have removed the Gaussian noise from the original Mean Teacher (MT) model [10], as smaller Gaussian noise does not significantly enhance model performance. Instead, we use a diverse set of augmentation strategies. For each unlabeled sample, both weak and strong augmentation strategies are applied before the samples are processed by the student and teacher models. Additionally, Dynamic Sample Reweighting ensures the reliability of the data. This enhancement strategy better simulates real data diversity and improves the model's utilization efficiency of unlabeled data.

Weak and Strong Augmentation. For each unlabeled sample x_u , we apply weak augmentation t_w and strong augmentation t_s :

$$\tilde{x}_u^w = t_w(x_u), \quad \tilde{x}_u^s = t_s(x_u) \quad (7)$$

Here, \tilde{x}_u^w and \tilde{x}_u^s are the augmented versions of the sample x_u using weak and strong augmentations, respectively. The choice of augmentation strategies follows methods used in prior works for enhancing model robustness [35].

Weak Augmentation: Weak augmentation refers to mild perturbations applied to the original image, such as small geometric transformations or slight changes in color. These augmentations are designed to simulate slight variations in the input while preserving the core structure and information of the image. The goal of weak augmentation is to enforce consistency under minimal perturbation, ensuring that the model learns to generalize well under minor changes.

Strong Augmentation: In contrast, strong augmentation introduces more significant transformations to the image, such as larger crops, rotations, or applying more extreme color distortions. These augmentations are intended to challenge the model more and promote robustness to more substantial variations in the input data. The purpose of strong augmentation is to ensure that the model can maintain consistent predictions even under more substantial perturbations, simulating real-world variations in the data.

Model Outputs. The student model and the teacher model generate outputs based on the augmented samples:

$$p_u^w = h_\alpha(\tilde{x}_u^w), \quad p_u^s = h_{\alpha'}(\tilde{x}_u^s) \quad (8)$$

where h_α denotes the student model and $h_{\alpha'}$ denotes the teacher model. The parameters α and α' represent the model parameters for the student and teacher networks. The student model is updated based on the outputs of \tilde{x}_u^w , and the teacher model is updated based on \tilde{x}_u^s .

Enhanced Consistency Loss. The enhanced consistency loss is defined as

$$L_C = \frac{1}{|D_u|} \sum_{u=1}^{|D_u|} \beta_u \|p_u^s - p_u^w\|_2^2 \quad (9)$$

where D_u is the total number of unlabeled samples and β_u is the dynamic weight for each sample. The loss function L_C measures the consistency between the predictions of the student and teacher models on augmented samples. The use of dynamic weights β_u adjusts the importance of each sample based on its reliability, as determined through our Dynamic Sample Reweighting strategy [36].

The design of the enhanced consistency loss L_C aims to improve the model's robustness and generalization capabilities by leveraging different augmentation strategies. By ensuring that the outputs of the student and teacher models remain consistent under various perturbations, we can utilize unlabeled data more effectively, significantly enhancing classification performance in semi-supervised learning scenarios. This approach not only improves the model's efficiency in utilizing unlabeled data but also strengthens its adaptability to data diversity. Additionally, the model uses an Exponential Moving Average (EMA) strategy to smooth the parameter updates, which helps to reduce fluctuations during training and ensures the consistency of the outputs, which is crucial for effective semi-supervised learning.

3.3. Medical Multi-Scale Feature Fusion Network (MedFuseNet)

To better capture and integrate image features at different levels and optimize unlabeled data feature learning in semi-supervised medical image classification, this paper designs the Medical Multi-scale Feature Fusion Network (MedFuseNet). In MedFuseNet, feature extraction is conducted at multiple levels to capture a wide range of details from lesion images. The low-level features, extracted by the initial layers (conv1), focus on fine-grained details such as edges, textures, and local patterns, which are crucial for identifying subtle variations in the image. The high-level features, derived from the deeper layers (conv3), represent abstract, global information, including the overall shape and contextual relationships of lesions, which contribute to a holistic understanding of the image. This multi-level approach ensures that both detailed and abstract information are effectively utilized, leading to optimal classification results.

To address the challenges in semi-supervised learning (SSL) where labeled data are scarce and unlabeled data may be noisy or imbalanced, MedFuseNet incorporates multi-scale feature fusion in an SSL setting, optimizing the learning of both labeled and unlabeled data. By integrating low-level and high-level features in a manner that prioritizes the most informative parts of the image, we enable the model to better adapt to the complexities of semi-supervised tasks. Specifically, this fusion strategy boosts the model's ability to classify images with minimal labeled data while leveraging the abundant unlabeled data in the training process.

Observing Figure 3, the network extracts low- and high-level features using convolutional layers. These features are standardized to a common number of channels using 1×1 convolutions. A similarity matrix is then created to evaluate the correlation between features by flattening and multiplying them. The similarity matrix can be expressed as

$$S_{ij} = \frac{\exp(P_i \cdot Q_j)}{\sum_{k=1}^N \exp(P_i \cdot Q_k)} \quad (10)$$

where S_{ij} represents the similarity between the i -th feature in the low-level feature P of the image and the j -th feature in the high-level feature Q . N is the number of pixels, which is the product of the feature map height H and width W , where $P \in \mathbb{R}^{(H \times W \times C_{low})}$ and $Q \in \mathbb{R}^{(H' \times W' \times C_{high})}$.

Specifically, we first perform matrix multiplication on the similarity matrix $S_{ij} \in \mathbb{R}^{(N \times N')}$ obtained from the calculation above and the feature at the i -th position of the low-level feature p_i to obtain the weighted low-level features F_w . These features are then reshaped to match the shape of the high-level features $F_{reshaped}$. Finally, the adjusted low-level features are element-wise added to the high-level features to obtain the final fused features:

$$F_j = \sum_{i=1}^N (S_{ij} \cdot p_i) + B_j \quad (11)$$

where B_j represents the feature of the j -th channel of the high-level feature map. The fused features are further processed to map them to the number of channels required for the classification task, and the feature map size is adjusted through upsampling to match the original size of the input image, thereby obtaining the final classification result.

In contrast to traditional feature fusion techniques, MedFuseNet uses a similarity matrix to quantify the correlation between low-level and high-level features. Traditional methods typically fuse features from different scales through simple weighted sums or concatenation. In contrast, our design precisely adjusts the fusion of low-level and high-level features by calculating the similarity matrix (Equation (10)), allowing the fused features to better capture both the detailed and global information of the image. Additionally, the 1×1 convolution layers and matrix weighting mechanism in MedFuseNet provide a more effective way to adjust low-level features, effectively preventing the loss of detailed information in the context of high-level features, which is a common issue in traditional methods.

Regarding feature fusion, all extracted features from low and high levels are used. We apply 1×1 convolutions (convP, convQ) to standardize the channel dimensions of these features, ensuring they are compatible for the subsequent fusion process. The fusion process involves the use of a similarity matrix, which measures the correlation between low-level and high-level features. This matrix is used to weigh and combine the most relevant features from both levels. The weighted low-level features are then adjusted to match the shape of the high-level features before being fused via element-wise addition. This ensures that the fusion captures the most important information from each level, enhancing feature diversity without excluding any important details.

In MedFuseNet, the fusion of low-level and high-level features serves not only to enhance feature representation but also to closely interact with the pseudo-label mechanism in semi-supervised learning. By dynamically generating pseudo-labels for unlabeled data and weighting the low-level and high-level features based on these labels, MedFuseNet more effectively utilizes unlabeled data, thereby boosting the overall model's generalization ability. Unlike traditional feature fusion methods, MedFuseNet's pseudo-label-guided fusion significantly improves the model's adaptability to unlabeled data, making it more robust in real-world medical imaging scenarios with sparse labeled data.

3.4. Pseudo-Label Guided Contrastive Learning (PGC)

Momentum Update Strategy. In this paper, we adopt the momentum update method from MoCo V2 [37] to implement dynamic updates of the queue mechanism. This method introduces a momentum parameter to retain historical update information and combines it with current gradient information to achieve real-time updates of features and weights, optimizing the feature learning of unlabeled data.

Specifically, we apply momentum updates to both the features of unlabeled data stored in the queue and their corresponding weights. This ensures that the feature information in the queue is effectively updated and maintained as new data features are captured. Let f_t represent the features stored in the queue at time step t . The formula for momentum update can be expressed as follows:

$$m_t = \theta m_{t-1} + (1 - \theta)g_t \tag{12}$$

$$f_{t+1} = f_t + \eta m_t \tag{13}$$

In these equations, m_t represents the momentum vector at time t , which integrates both the historical momentum m_{t-1} and the current gradient g_t . The momentum coefficient θ governs the retention of historical information, while g_t , the gradient at time step t , signifies the instantaneous rate of change. Additionally, η denotes the learning rate, which determines the step size for the update. Together, these parameters play a crucial role in optimizing the learning process and enhancing the convergence of the model.

Pseudo-label Guided Contrastive (PGC) Loss Function. For unlabeled samples x_u , we make predictions using the trained model to obtain a class probability distribution. From this distribution, the category with the highest probability is selected as the pseudo-label:

$$\hat{y}_u = \arg \max \left\{ p \left(y = \frac{C}{x_u}; \alpha \right) \right\} \tag{14}$$

where C denotes the category index. For the current unlabeled sample, the generated pseudo-label is compared with the sample features stored in the queue. If the pseudo-labels match, they form a positive sample pair (samples of the same category) with the current sample; otherwise, they form a negative sample pair (samples of different categories).

To measure the similarity between samples, a contrastive loss function is constructed. The loss function must meet the following conditions:

1. For each sample x_u , when it is similar to the positive sample pair, its similarity is high and the contrast loss is small.
2. When the sample x_u is not similar to the positive sample pair or is similar to other negative sample pairs, the contrast loss should be high.

To address these requirements, this paper proposes a pseudo-label guided contrast loss function:

$$L_{PGC} = \frac{1}{D_u} \sum_{u=1}^{D_u} - \log \frac{\sum_{p \in P(u)} \exp \left(\frac{\beta_u z_u \cdot w_p z_p}{\tau} \right)}{\sum_{a \in A(u)} \exp \left(\frac{\beta_u z_u \cdot w_a z_a}{\tau} \right)} \tag{15}$$

where β_u denotes the weight of the unlabeled sample after dynamic reweighting, which ensures the reliability of the unlabeled data. The weight w_p signifies the importance of the positive sample pair, emphasizing the similarity between samples of the same category, while w_a represents the weight of all sample pairs, balancing the similarity differences across categories and minimizing the overlap in the feature space. Additionally, $P(u)$ and $A(u)$ correspond to the positive sample pairs and all sample pairs (including both positive and negative pairs), respectively. Finally, τ is the temperature parameter that controls the smoothness of the similarity distribution.

Observing Figure 4, The Pseudo-Label Guided Contrastive (PGC) loss function uses a momentum update strategy to continuously update the features in the queue mechanism, enhancing the model’s effective use of unlabeled data. Additionally, pseudo-labels are used to partition positive and negative sample pairs, which reduces the dependence on labeled data for contrastive learning. Importantly, PGC integrates contrastive learning with the features fused by MedFuseNet, optimizing and improving the classification decision boundary, thereby enhancing classification performance.

3.5. Total Loss Function

Minimizing the cross-entropy loss function ensures that the predicted probability distribution for labeled samples aligns with their corresponding one-hot encoded true labels. By optimizing this loss during training, the model adjusts its parameters α to accurately reflect the true category of each labeled sample, leading to improved classification performance.

The supervised loss function is defined as follows:

$$L_S = -\frac{1}{M} \sum_{l=1}^M \sum_{c=1}^C y_{l,c} \log f(x_l; \alpha)_c \quad (16)$$

where M denotes the number of labeled samples, $f(x_l; \alpha)_c$ represents the predicted probability distribution of the model for sample x_l in category c , and $y_{l,c}$ indicates the true label of sample x_l in category c .

The overall optimized loss function is expressed as follows:

$$L = L_S + \lambda_1 L_C + \lambda_2 L_{PGC} \quad (17)$$

where λ_1 and λ_2 are hyperparameters that control the balance between the consistency regularization loss L_C and the pseudo-label guided contrastive loss L_{PGC} , respectively.

3.6. Model Training

The following describes the training process of our semi-supervised learning model, which integrates both labeled and unlabeled medical images. In each iteration, the model processes labeled and unlabeled medical images, applying weak augmentation to labeled samples and both weak and strong augmentation to unlabeled samples. Feature representations are extracted using the baseline model DenseNet-121 and the Medical Multi-scale Feature Fusion Network (MedFuseNet), both of which are trainable throughout the training process. The model dynamically weights the samples based on the confidence scores assigned to unlabeled data, ensuring that higher confidence samples receive greater attention during training.

The specific training process is as follows:

1. Initialize model parameters and prepare the labeled and unlabeled datasets by applying normalization and data augmentation techniques.
2. Utilize DenseNet-121 to process both labeled and unlabeled samples. The resulting feature representations are then passed to MedFuseNet for multi-scale feature fusion, generating a comprehensive feature representation.
3. Calculate sample weights based on the model's confidence in its predictions for unlabeled samples. These weights are dynamically updated during training to prioritize learning from high-confidence samples.
4. Generate pseudo-labels for unlabeled data using the current model's predictions. Compute the Pseudo-label Guided Contrastive (PGC) Loss with these pseudo-labels to enhance the model's ability to distinguish similar samples.
5. Apply various augmentation operations to the input images and compute the consistency loss to ensure stable outputs under perturbations. This loss helps the model learn robust feature representations.
6. Extract features at different layers using MedFuseNet and combine them to create a comprehensive representation of the input images.
7. Update model parameters by minimizing the overall loss function, which includes both the contrastive loss and the consistency loss.
8. Repeat steps (2)–(7) for each training iteration until convergence.
9. Evaluate the performance of the trained model on the test dataset. The model's predictions are compared to the ground truth labels to compute accuracy and assess classification performance.

A more detailed explanation is illustrated in Algorithm 1.

Algorithm 1 Enhanced Semi-Supervised Medical Image Classification Based on Dynamic Sample Reweighting and Pseudo-Label Guided Contrastive (DSRPGC) Learning

Require: Labeled dataset $D_l = \{(x_i, y_i)\}_{i=1}^{N_l}$; Unlabeled dataset $D_u = \{x_j\}_{j=1}^{N_u}$; Initialize parameters (α, β) .

- 1: Initialize cache queue Q to be empty.
 - 2: Initialize the classification network F .
 - 3: **for** $t = 1$ to T **do**
 - 4: Select s labeled samples: $X_l = \{(x_l^{(i)}, y^{(i)})\}_{i=1}^s$.
 - 5: Select s unlabeled samples: $X_u = \{x_u^{(j)}\}_{j=1}^s$.
 - 6: Apply weak augmentation to unlabeled samples: $\tilde{x}_u^w = t_w(x_u)$.
 - 7: Apply strong augmentation to unlabeled samples: $\tilde{x}_u^s = t_s(x_u)$.
 - 8: Input x_l^w , \tilde{x}_u^w , and \tilde{x}_u^s into the model F .
 - 9: Calculate entropy for unlabeled samples and assign weights β_u based on uncertainty, as shown in Equations (5) and (6).
 - 10: Ensure consistency of predictions for the same image under different augmentations, as shown in Equation (9).
 - 11: Extract and fuse features at multiple levels using MedFuseNet, as shown in Equations (10) and (11).
 - 12: Use a momentum updating strategy to maintain dynamic updates of features and weights, as shown in Equations (12)–(15).
 - 13: Calculate the total training loss, as shown in Equation (17), enabling the model to learn the underlying structures and features within the data, thereby improving classification accuracy and robustness.
 - 14: Update parameters using Equations (3) and (4) to ensure that the model can quickly adapt and improve its classification performance when encountering new samples.
 - 15: **end for**
 - 16: Output the well-trained model F .
-

4. Experiments

4.1. Experimental Settings

Dataset setup. To validate the semi-supervised classification method proposed in this paper, we conducted experiments on three public medical image datasets: ISIC2018 [38], NCT-CRC-HE [39], and Chest X-ray14 [40]. Specifically, ISIC2018 contains 10,015 dermoscopic images categorized into seven classes. These images are RGB color images with three channels. The seven classes are as follows: 1113 melanomas (MEL), 6705 melanocytic nevi (NV), 514 basal cell carcinomas (BCC), 327 actinic keratoses (AKIEC), 1099 benign keratoses (BKL), 115 dermatofibromas (DF), and 142 vascular lesions (VASC). The training set consists of 7012 images, the test set contains 2003 images, and the validation set includes 1000 images. These seven types of images are shown in Figure 5. Each image in this dataset represents a single class, focusing on distinct skin lesions with specific characteristics.

NCT-CRC-HE consists of high-resolution pathology images from colorectal cancer patient tissue samples, containing 100,000 hematoxylin and eosin (HE) stained tissue slice images covering nine categories. These categories are adipose (ADI) with 10,407 samples, background (BACK) with 10,566 samples, debris (DEB) with 11,512 samples, lymphocytes (LYM) with 11,557 samples, mucus (MUC) with 8896 samples, smooth muscle (MUS) with 13,536 samples, normal colon mucosa (NORM) with 8763 samples, cancer-associated stroma (STR) with 10,446 samples, and colorectal adenocarcinoma epithelium (TUM) with 14,317 samples. The dataset is split into 70% for training, 20% for testing, and 10% for validation. All images are resized to 224×224 to ensure consistency in the experiments. These nine categories are shown in Figure 6. The images in this dataset may contain multiple tissue types, as each tissue slice may include various regions corresponding to different categories (e.g., tumor cells and surrounding stroma in the same image). Thus,

these images often present a multi-class tissue composition, making the classification task more challenging.

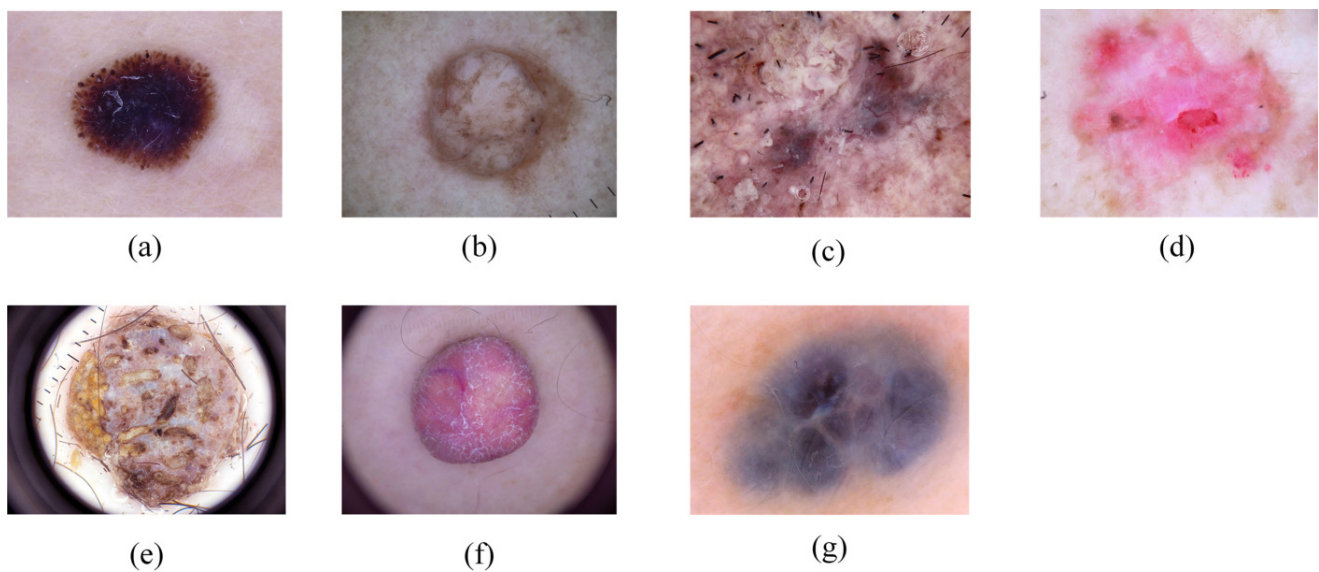


Figure 5. Samples from ISIC2018 dataset: (a) melanomas (MEL); (b) melanocytic nevi (NV); (c) basal cell carcinomas (BCC); (d) actinic keratoses (AKIEC); (e) benign keratoses (BKL); (f) dermatofibromas (DF); (g) vascular lesions (VASC).

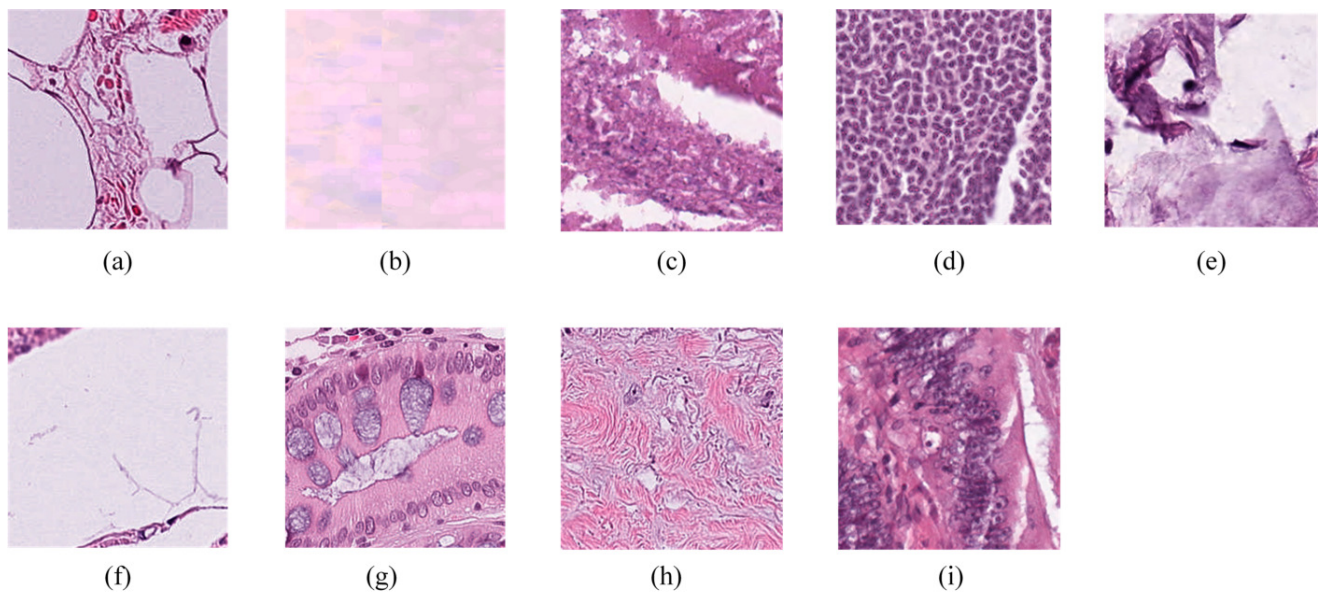


Figure 6. Samples from NCT-CRC-HE dataset: (a) adipose (ADI); (b) background (BACK); (c) debris (DEB); (d) lymphocytes (LYM); (e) mucus (MUC); (f) smooth muscle (MUS); (g) normal colon mucosa (NORM); (h) cancer-associated stroma (STR); (i) colorectal adenocarcinoma epithelium (TUM).

The Chest X-ray14 dataset contains a total of 112,120 X-ray images, representing 14 different types of lung diseases. For our experiments, we selected six common lung diseases: Atelectasis (2564 images), Effusion (2406 images), Infiltration (2296 images), Mass (1302 images), Nodule (1646 images), and Pneumothorax (1335 images), along with 10,711 normal lung images. The dataset is divided into a training set of 15,582 images, a validation set of 2226 images, and a test set of 4452 images. Examples of these six types of lung conditions are shown in Figure 7. This dataset's heterogeneity arises from the presence of multiple lung conditions in single images, which can complicate the classification task.

Implementation details. Our classification model framework is implemented in PyTorch and trained on a 24 GB RTX 3090 GPU using Python 3.8. The model employs two augmentation strategies with varying random parameters during training. The smoothing parameter for the EMA is set to 0.999. DenseNet-121 [41] is used as the backbone, and the Adam optimizer [42] with an initial learning rate of 1×10^{-4} is employed. The batch size is 64, consisting of 16 labeled samples and 48 unlabeled samples. The training process includes a warm-up phase followed by the formal training phase.

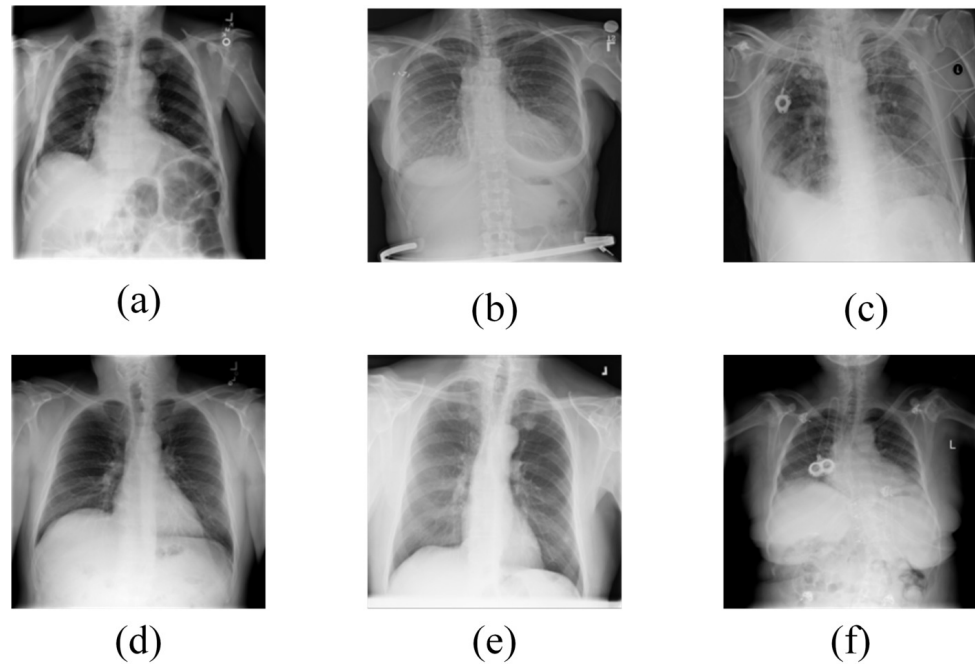


Figure 7. Samples from Chest X-ray14 dataset: (a) Atelectasis; (b) Effusion; (c) Infiltration; (d) Mass; (e) Nodule; (f) Pneumothorax.

Evaluation Metrics. To evaluate the performance of our proposed method, we use the following metrics:

1. **Accuracy:** The proportion of correctly classified samples out of all samples. It is defined as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives.

2. **Area Under the Curve (AUC):** The area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. It is computed as

$$AUC = \int_0^1 ROC(x) dx \quad (19)$$

where $ROC(x)$ represents the ROC curve.

3. **Sensitivity (Recall):** The ability of the model to correctly identify positive samples. It is defined as

$$Sensitivity = \frac{TP}{TP + FN} \quad (20)$$

4. **Specificity:** The ability of the model to correctly identify negative samples. It is defined as

$$Specificity = \frac{TN}{TN + FP} \quad (21)$$

5. **F1 Score:** The harmonic mean of precision and sensitivity, providing a balance between the two metrics. It is defined as

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (22)$$

6. **Precision:** The proportion of positive identifications that were actually correct. It is defined as

$$\text{Precision} = \frac{TP}{TP + FP} \quad (23)$$

These metrics provide a comprehensive evaluation of the model's performance in terms of classification accuracy, balance between precision and recall, and ability to discriminate between classes. Since evaluation metrics such as accuracy, sensitivity, and AUC rely on ground truth labels, we applied pseudo-labeling for unlabeled samples. In this process, the model assigned predicted labels to the unlabeled data, which were then treated as ground truth for evaluation purposes. These pseudo-labels allowed us to apply the standard evaluation metrics to the unlabeled samples.

Data Processing. In our study, we performed comprehensive image preprocessing and data augmentation to enhance the model's generalization and robustness. Initially, all images were converted to RGB format and resized to a fixed size of 224×224 pixels to ensure uniformity across the dataset. Figure 8 illustrates the entire preprocessing process. It includes the original image, the RGB image, the resized image at 224×224 pixels, the weakly augmented image resulting from random horizontal flipping, and the strongly augmented image created through random cropping and hue/saturation adjustments. We applied various data augmentation techniques to the pre-training and training datasets, including random cropping, color jittering, affine transformations, and horizontal flips. These augmentation methods allowed the model to learn more generalizable features by simulating various visual conditions. Each image was processed multiple times with different transformations to facilitate contrastive learning. For the validation and test datasets, we only applied resizing and normalization to maintain consistency in evaluation without augmentations. To address class imbalance, we employed a custom sampling strategy to ensure equal representation of both majority and minority classes in each batch. The class imbalance, particularly observed in the ISIC2018 and NCT-CRC-HE datasets, was mitigated by these strategies to prevent the model from being biased toward the more frequent classes. Additionally, this ensured that all classes were sufficiently represented in the training process, thus improving the model's ability to generalize across underrepresented classes. In addition to the ISIC2018 and NCT-CRC-HE datasets, we also included the Chest X-ray14 dataset, which comprised 112,120 chest X-ray images covering 14 lung diseases. For our experiments, we focused on six common lung diseases (Atelectasis, Effusion, Infiltration, Mass, Nodule, and Pneumothorax), along with normal cases. The class imbalance in this dataset was addressed similarly using our custom sampling strategy to ensure balanced training batches and improve the model's performance across all categories.

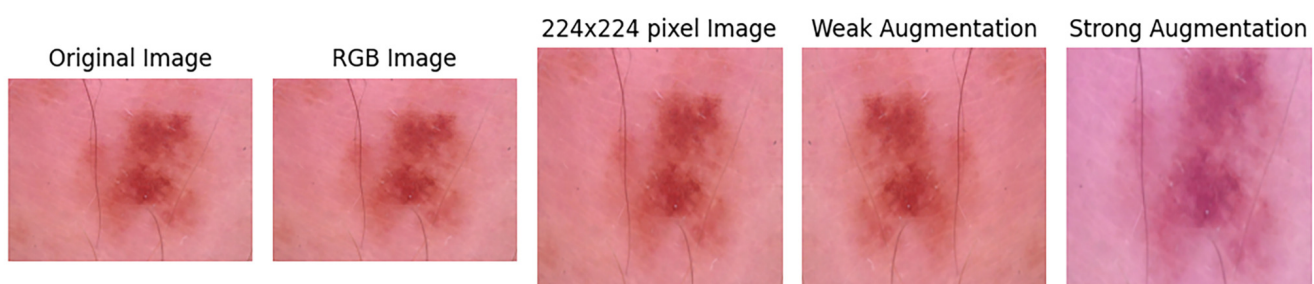


Figure 8. The entire preprocessing process of ISIC2018 dataset images.

4.2. Comparison with Other Semi-Supervised Classification Methods

Experimental results of our method on the ISIC2018 dataset. Observing Table 1, SelfMatch [43] utilized self-supervised learning and pseudo-label generation to effectively leverage unlabeled data. However, its limitations lie in its dependence on the quality of pseudo-labels and the effectiveness of the self-matching strategy, which may have led to the model's performance being affected by errors in pseudo-labels. In contrast, the MT [10] method addressed issues such as overfitting and generalization error in labeled data by leveraging consistency regularization and smooth updates of the teacher model. Compared to the MT method, SRC-MT [19] introduced a sample relationship consistency paradigm, further improving performance. Notably, our method incorporates dynamic weighting and random augmentation strategies into the MT framework, which not only filter out reliable unlabeled samples but also ensures consistency across different augmented samples in the feature space, achieving the best results across all evaluation metrics.

Table 1. Comparison with the latest semi-supervised methods on the ISIC2018 dataset with 20% labeled data.

Methods	Labeled	Unlabeled	AUC	Sensitivity	Specificity	Accuracy	F1
Baseline	20	0	91.12	68.69	91.82	91.85	58.21
SelfMatch [43]	20	80	92.92	72.10	92.45	91.96	59.67
MT [10]	20	80	93.34	70.15	92.46	92.38	60.36
SRC-MT [19]	20	80	93.59	70.45	92.52	92.59	60.68
Ours	20	80	94.49	76.92	92.68	93.16	65.82

Observing Figure 9, we compare the experimental results of our proposed method with the current state-of-the-art SRC-MT method on the confusion matrix. Each matrix shows the prediction accuracy for each category. Our method exhibits higher accuracy rates across most categories compared to SRC-MT. For instance, our method shows significantly better performance in categories like melanoma (MEL), basal cell carcinoma (BCC), melanocytic nevi (NV), actinic keratoses (AKIEC), benign keratoses (BKL), dermatofibromas (DF), and vascular lesions (VASC), where SRC-MT exhibits lower accuracy and higher misclassification rates. This indicates that SRC-MT struggles to correctly classify these challenging cases, leading to more false positives and false negatives. The superior accuracy of our method can be attributed to its dynamic weighting and random augmentation strategies. These techniques enhance the robustness of our model, allowing it to better handle variability and maintain high accuracy across different categories. This results in fewer misclassifications and more reliable classification performance overall. In summary, our method's confusion matrix highlights its ability to achieve higher accuracy and lower misclassification rates compared to SRC-MT, demonstrating its enhanced effectiveness in handling the complexities of the ISIC2018 dataset.

Experimental results of our method on the NCT-CRC-HE dataset. Observing Table 2, our method outperforms other SSL methods across multiple evaluation metrics. Compared to FixMatch [44], our approach introduces a dynamic weighting strategy, reducing the model's reliance on the accuracy of pseudo-labels, resulting in a 5.5% improvement in the accuracy metric. Even against the state-of-the-art SimMatch [45], our method remains superior, as SimMatch is still limited by the impact of noisy data, which leads to incorrect feature learning and pseudo-label generation. This further highlights the superiority of our approach, which not only ensures the quality of unlabeled data but also enhances the model's ability to differentiate colorectal cancer.

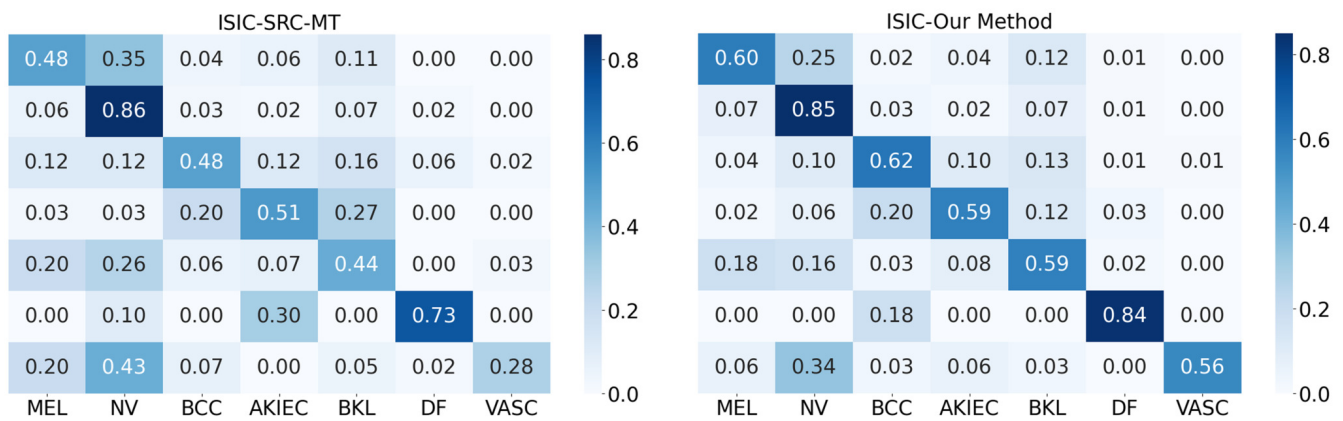


Figure 9. Confusion matrix of SRC-MT method and this method on ISIC 2018 dataset.

Table 2. Comparison with the latest semi-supervised methods on the NCT-CRC-HE dataset with 100 labeled samples.

Methods	NCT-CRC-HE (100 Labeled Data)				
	AUC	Sensitivity	Precision	Accuracy	F1
Baseline	96.32	74.01	75.98	73.65	73.29
MT [10]	97.15	77.51	78.81	77.97	77.07
FixMatch [44]	97.91	80.59	81.78	80.47	80.28
CoMatch [29]	97.99	84.70	84.58	83.93	84.12
SimMatch [45]	98.03	85.07	84.50	84.24	84.43
Ours	98.21	86.77	86.81	85.97	86.45

Additionally, Figure 10 illustrates the confusion matrices for our method and SimMatch on the NCT-CRC-HE dataset, revealing that our approach consistently surpasses SimMatch in classification accuracy across various categories, particularly in critical areas such as colorectal cancer. The categories in the NCT-CRC-HE dataset include adipose (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), and colorectal adenocarcinoma epithelium (TUM). This enhancement in performance is attributed to the advanced features of our method, notably, the integration of the Medical Multi-scale Feature Fusion Network (MedFuseNet) and the Pseudo-Label Guided Contrastive Learning (PGC) strategy. MedFuseNet improves feature extraction by effectively combining information from multiple scales, enabling the model to capture intricate details and subtle differences within the dataset. Simultaneously, the PGC method enhances learning stability through a momentum-updated feature queue, allowing for better alignment of features while systematically defining positive and negative sample pairs to refine the model’s discriminative power. In contrast, SimMatch’s reliance on potentially noisy pseudo-labels leads to lower accuracy and more misclassifications in key categories. Overall, our method effectively addresses the challenges posed by noisy and imbalanced data, underscoring its superior accuracy and robustness compared to SimMatch.

The results in Table 3 demonstrate that our method consistently outperforms all other models across different levels of labeled data in the Chest X-ray14 dataset. While Pseudo-Labeling shows the lowest performance, particularly with only 5% labeled data, FixMatch and MT achieve significant improvements through consistency regularization. However, our method achieves the highest accuracy at all annotated percentages, showcasing its superiority, likely due to Dynamic Sample Reweighting and medical multi-scale feature fusion, which enhance the model’s ability to effectively utilize both labeled and unlabeled data.

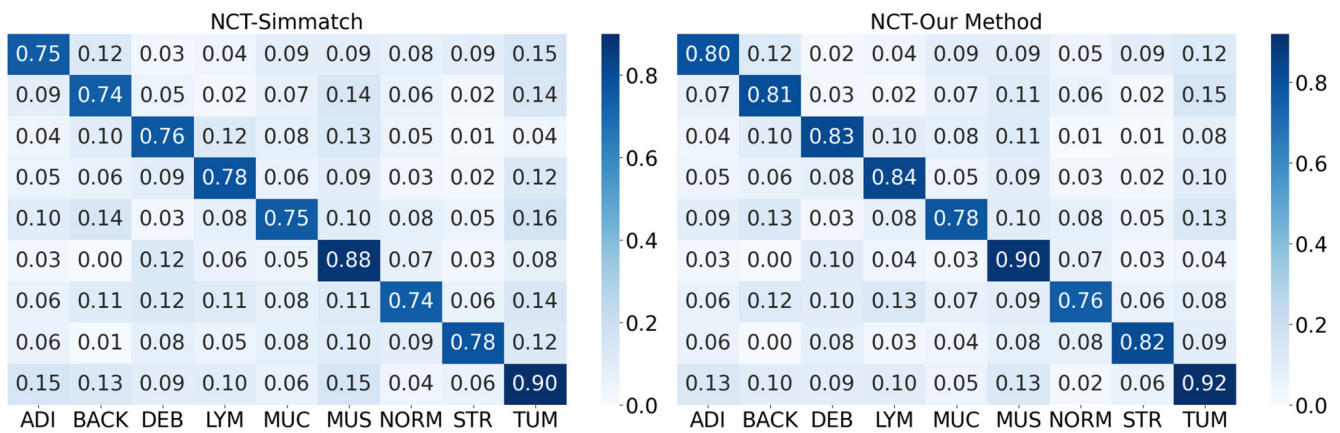


Figure 10. Confusion matrix of Simmatch method and this method on NCT-CRC-HE dataset.

Table 3. Classification accuracy with other models under different annotated percentages for the Chest X-ray14 dataset.

Method	5%	10%	20%
Pseudo-Labeling [17]	57.64%	65.27%	72.46%
FixMatch [44]	68.09%	73.01%	80.34%
MT [10]	67.96%	72.96%	79.07%
Ours	70.03%	75.20%	80.83%

The results in Table 4 demonstrate that our pseudo-labeling strategy effectively balances accuracy and coverage throughout training. By gradually increasing both metrics as the model stabilizes, our method selectively incorporates high-confidence unlabeled data, enhancing feature learning without introducing excessive noise. This adaptive approach ensures that the model leverages unlabeled data efficiently, leading to improved classification performance on the ISIC2018 dataset. Integration with our Pseudo-Label Guided Contrastive Learning (PGC) and Dynamic Sample Reweighting Strategy further refines the model’s robustness, contributing to more accurate and reliable predictions in semi-supervised medical image classification. It is worth noting that the model begins at epoch 80 because the first 80 epochs are dedicated to the warm-up phase.

Table 4. Pseudo-label accuracy and coverage at different training stages on the ISIC2018 dataset.

Epoch	Pseudo-label Accuracy (%)	Pseudo-Label Coverage (%)
80	60.3	28.7
100	68.9	45.2
120	75.6	57.8
140	81.4	65.3
160	85.2	72.5
180	88.5	78.9

4.3. Ablation Study

The proposed method is built upon three essential components: the Dynamic Sample Reweighting Strategy, the Medical Multi-scale Feature Fusion Network (MedFuseNet), and Pseudo-Label Guided Contrastive Learning (PGC). To assess the contribution of each component to the overall model performance, we conducted ablation experiments. In these experiments, we systematically remove or modify each component and evaluate how these changes affect the performance of the model. Ablation experiments are commonly used to isolate the individual impact of each part of a system, enabling us to understand the importance of each component in enhancing the model’s accuracy and robustness.

By conducting these experiments, we can identify which components are most crucial to the performance improvements observed in our proposed method.

To verify the impact of the Dynamic Sample Reweighting Strategy on model performance, we conducted an ablation study. In these experiments, we excluded this strategy (-/o Dynamic Weight) to evaluate its impact. Observing Figure 11, it is evident that when the weights of unlabeled data are evenly distributed, key metrics such as AUC and accuracy are significantly affected, particularly for edge cases or low-confidence samples. The introduction of the Dynamic Sample Reweighting Strategy results in significant improvements in model performance, with AUC and accuracy increasing by 4.28% and 4.02%, respectively. This strategy effectively filters out noisy data and addresses class imbalance by adjusting sample weights, thus enhancing the overall training process. When integrated with Pseudo-Label Guided Contrastive Learning (PGC) and the Medical Multi-scale Feature Fusion Network (MedFuseNet), the Dynamic Sample Reweighting Strategy synergistically improves feature representation and classification accuracy by ensuring that high-quality samples are prioritized during optimization and multi-scale feature integration. These mechanisms collectively contribute to the model's enhanced capability to manage diverse sample enhancement views, thereby significantly boosting performance in semi-supervised medical image classification tasks.

To assess the effectiveness of the Medical Multi-scale Feature Fusion Network (MedFuseNet) in our semi-supervised medical image classification model, we conducted an ablation study. The results, illustrated in Figure 11, show that removing MedFuseNet (-/o MedFuseNet) leads to a decline in AUC and accuracy, underscoring its significant contribution to the model's overall performance. MedFuseNet enhances the model's ability to consolidate features across multiple spatial scales, thereby improving intra-class compactness and inter-class separability. By integrating features from various scales, MedFuseNet enables a more cohesive representation of similar samples, simplifying classification and making different categories more distinguishable in the feature space. Figure 12 highlights how MedFuseNet contributes to feature integration, further enhancing these properties. This multi-scale feature fusion works in synergy with the Dynamic Sample Reweighting Strategy, ensuring that high-quality data inform feature extraction. Additionally, the Pseudo-Label Guided Contrastive Learning (PGC) loss function benefits from the diverse representations provided by MedFuseNet, enabling more effective sample pairing. The results from these ablation experiments confirm that MedFuseNet plays a crucial role in optimizing the utilization of unlabeled data, significantly enhancing the overall performance of the model in semi-supervised medical image classification tasks.

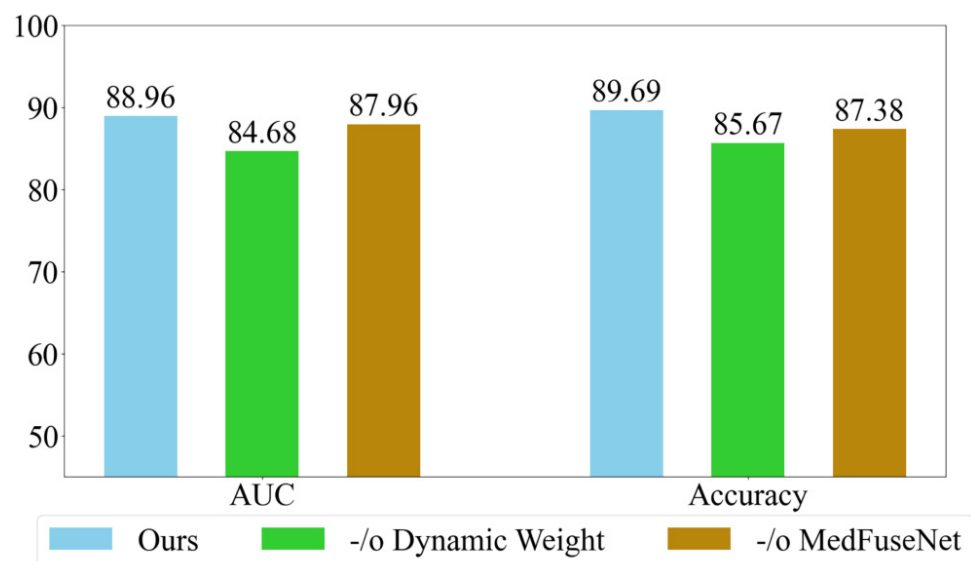


Figure 11. Ablation experiment results on the ISIC2018 dataset with 20% labeled data.

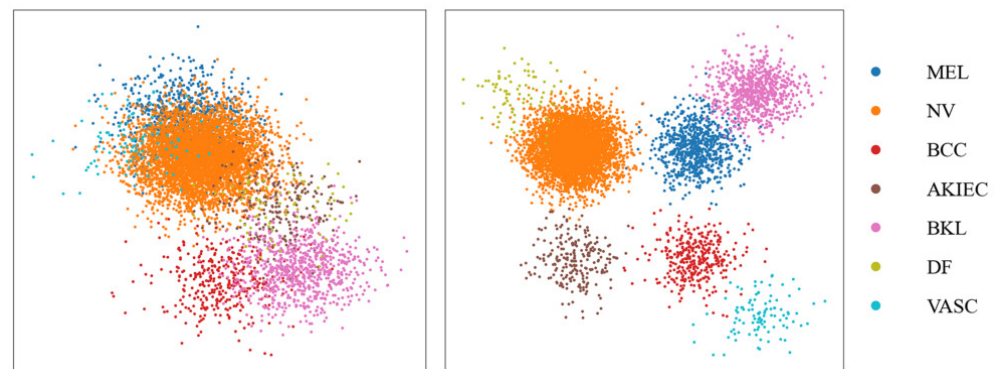


Figure 12. t-SNE visualization on the ISIC2018 dataset.

The results in Table 5 clearly highlight the superior performance of the PGC Loss function across all evaluation metrics. Compared to Contrastive Loss and Cross-Entropy Loss, the PGC Loss achieves the highest AUC (94.49%), sensitivity (76.85%), specificity (92.68%), accuracy (93.16%), and F1 score (65.92%). This demonstrates that the PGC Loss not only improves classification accuracy but also enhances the model's ability to distinguish between classes and correctly identify positive samples. The significant improvements in all metrics emphasize the effectiveness of the PGC Loss in enhancing semi-supervised medical image classification.

Table 5. Comparison of different loss functions on model performance.

Loss Function	AUC	Sensitivity	Specificity	Accuracy	F1 Score
Contrastive Loss	88.30	64.12	87.97	87.11	57.72
Cross-Entropy Loss	90.74	70.21	90.35	90.01	61.28
PGC Loss	94.49	76.85	92.68	93.16	65.92

To evaluate the effect of Pseudo-Label Guided Contrastive Learning (PGC) on model performance, we conducted ablation experiments on two datasets. The results, shown in Figure 13, reveal that excluding the PGC loss module and relying solely on the MT model led to fluctuating accuracy curves, indicating instability in the learning process. Additionally, using supervised contrastive learning (SupCon) for unlabeled data resulted in slow accuracy improvements due to its reliance on large labeled datasets, which presents a significant challenge in medical image classification. The proposed PGC loss strategy significantly enhances feature representation by leveraging a momentum-updated feature queue and optimizing pseudo-label-based sample pairs, improving both intra-class compactness and inter-class separability. Its integration with the Dynamic Sample Reweighting Strategy allows the model to prioritize reliable samples, reducing noise interference and ensuring that contrastive learning is guided by high-quality features. Moreover, the synergy between PGC and the Medical Multi-scale Feature Fusion Network (MedFuseNet) enriches feature representations across multiple spatial scales, further boosting classification performance. Overall, the PGC loss function plays a critical role in refining the model's learning objectives and enhancing its overall performance in semi-supervised medical image classification tasks.

As shown in Table 6, we conducted an ablation study on the ISIC2018 dataset to investigate the individual contributions of each component in our model. The baseline model, DenseNet-121, serves as the foundation for comparison. The inclusion of Dynamic Sample Reweighting (DSR) improves performance by handling uncertain pseudo-labels more effectively. Adding MedFuseNet, a medical multi-scale feature fusion network, further enhances the model's ability to capture and integrate diverse features, leading to higher accuracy. Finally, the incorporation of Pseudo-Label Guided Contrastive Learning (PGC) strengthens intra-class compactness and inter-class separability, resulting in even

better performance. The full model, incorporating all components, outperforms the baseline and individual components, demonstrating the effectiveness of each contribution.

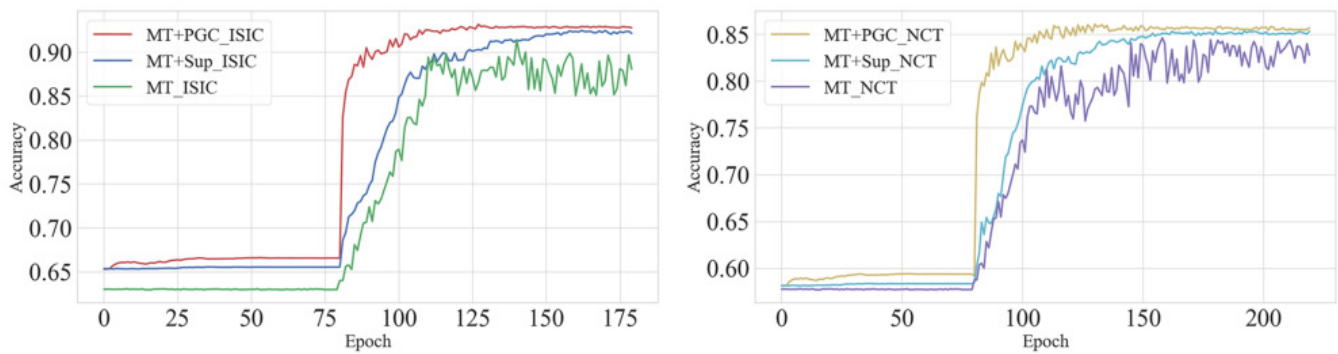


Figure 13. Classification accuracy of different contrastive learning modules under different datasets.

Table 6. Ablation Study: comparison of different model configurations on the ISIC2018 dataset with 20% labeled data.

Methods	AUC	Sensitivity	Specificity	Accuracy	F1
Baseline (DenseNet-121)	91.12	68.69	91.82	91.85	58.21
DSR (DenseNet-121 + DSR)	92.34	73.12	92.13	92.07	60.12
MedFuseNet (DenseNet-121 + MedFuseNet)	93.00	74.45	92.34	92.50	61.45
PGC (DenseNet-121 + PGC)	93.30	75.10	92.50	92.65	62.12
Ours (DenseNet-121 + DSR + MedFuseNet + PGC)	94.49	76.92	92.68	93.16	65.82

The results from Table 7 demonstrate the comparison between the performance of Pseudo-label Generation, Label-Guided Graph Contrastive Learning, and Pseudo-Label Guided Contrastive Learning (PGC). Among the methods, PGC stands out with the highest performance across all evaluation metrics. Specifically, PGC achieves an AUC of 94.49%, sensitivity of 76.85%, specificity of 92.68%, accuracy of 93.16%, and F1-score of 65.92%. This shows the significant advantages of incorporating the PGC module, which not only improves feature representation but also enhances both intra-class compactness and inter-class separability. By optimizing pseudo-label-based sample pairs and utilizing a momentum-updated feature queue, PGC outperforms the other methods, highlighting its robustness and effectiveness in semi-supervised medical image classification tasks.

Table 7. Comparison of Pseudo-label Generation, Label-Guided Graph Contrastive Learning, and PGC.

Methods	AUC	Sensitivity	Specificity	Accuracy	F1
Pseudo-label Generation	92.75	74.62	91.03	91.08	63.90
Label-guided Graph Contrastive Learning	91.23	72.50	90.12	89.45	60.74
PGC	94.49	76.85	92.68	93.16	65.92

To demonstrate the impact of varying proportions of labeled images on model performance, we analyzed the average accuracy at different levels of labeled data. As shown in Tables 8 and 9, the model’s accuracy increases as the proportion of labeled images rises. For example, on the ISIC2018 dataset, the model achieves an average accuracy of 89.69% when 5% of the data are labeled. When the labeled data proportion increases to 10%, the average accuracy improves significantly to 91.84%. However, further increasing the proportion to 20% only marginally improves the average accuracy to 93.16%, suggesting diminishing returns. This trend indicates that our method efficiently utilizes limited labeled data.

Table 8. Comparison of average accuracy on the ISIC2018 dataset with different labeled data ratios (%).

Methods	Label Percentage		
	5%	10%	20%
Baseline	85.96	89.62	91.85
SelfMatch	86.39	90.13	91.96
MT	87.59	90.65	92.38
SRC-MT	88.77	91.08	92.59
Ours	89.69	91.84	93.16

Table 9. Comparison of average accuracy on NCT-CRC-HE dataset with different labeled data (%).

Methods	The Number of Labeled Samples		
	50	100	200
MT	73.89	77.97	81.55
FixMatch	76.24	80.47	84.81
CoMatch	79.96	83.93	86.48
SimMatch	81.23	84.24	88.31
Ours	83.79	85.97	90.19

The five-fold cross-validation results on both the ISIC2018 (Table 10) dataset highlight the robustness and stability of our proposed approach across different medical image classification tasks. For the ISIC2018 dataset, using 20% labeled data, our method achieved an average AUC of 94.49% with a low standard deviation of ± 0.07 , along with consistently stable sensitivity and accuracy scores (standard deviations of ± 0.08 and ± 0.04 , respectively). This demonstrates the model's effectiveness in handling challenging and imbalanced skin lesion data. These consistent results confirm the model's strong generalization capability, making it suitable for real-world clinical applications involving noisy and limited labeled data.

Table 10. Five-fold cross-validation results on the ISIC2018 dataset with 20% labeled data.

Fold	AUC	Sensitivity	Specificity	Accuracy	F1
Fold 1	94.52	76.98	92.70	93.20	65.89
Fold 2	94.40	77.01	92.62	93.11	65.75
Fold 3	94.60	76.80	92.72	93.17	65.95
Fold 4	94.50	76.85	92.69	93.15	65.80
Fold 5	94.43	76.85	92.65	93.18	65.70
Mean \pm Std	94.49 \pm 0.07	76.90 \pm 0.08	92.68 \pm 0.04	93.16 \pm 0.04	65.82 \pm 0.09

In Figure 14, we present the relationship between training time and AUC for different model configurations. It can be observed that as the training time increases, the model performance (AUC) gradually improves. Specifically, the base model (with a training time of approximately 6 h) achieves an AUC of 90.12%. After incorporating MedFuseNet (Medical Multi-scale Feature Fusion Network), the AUC increases to 91.23%. This relatively small increase in AUC reflects the enhanced feature representation capability of MedFuseNet, demonstrating how even slight performance improvements can significantly enhance the model's ability to better distinguish between different medical image classes. In the context of medical image classification, even marginal gains in AUC can directly translate into more reliable and accurate results, offering substantial benefits for clinical applications. Although the training time slightly increases with the addition of MedFuseNet, the AUC improvement justifies this trade-off, as the performance boost is highly valuable in this domain.

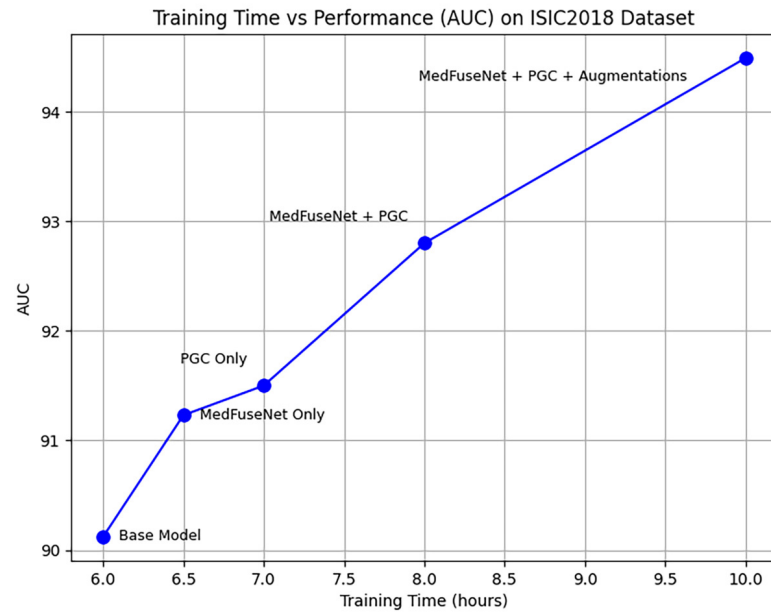


Figure 14. Training time vs. performance (AUC) on ISIC2018 dataset.

Furthermore, when combining PGC (Pseudo-Label Guided Contrastive Learning) and data augmentation (MedFuseNet + PGC + Augmentations), the training time increases to 10 h but the AUC rises to 94.49%. While this represents a more significant increase in training time, the AUC gain of over three percentage points is considerable, particularly in the context of medical image classification, where higher diagnostic accuracy can make a crucial difference in patient outcomes. Notably, this increase in training time is not excessive given the substantial improvement in performance. The results demonstrate that, despite the increased computational cost, the method achieves efficient and meaningful performance gains that are well worth the additional investment in computational resources. In medical image classification, where every percentage point in accuracy can have a major impact, such improvements justify the added complexity and time, confirming that the method is not only effective but also practical for real-world medical applications.

The analysis of the impact of λ_{DSR} and λ_{PGC} on model accuracy shows that both hyperparameters significantly affect the model’s performance. As shown in Figure 15, λ_{DSR} achieves the highest accuracy of 93.16% at 0.4, with accuracy decreasing as λ_{DSR} increases beyond this point. Similarly, λ_{PGC} reaches its peak accuracy at 0.6, and further increases lead to a slight decline in performance. These results suggest that the optimal values for λ_{DSR} and λ_{PGC} are 0.4 and 0.6, respectively, for maximizing model accuracy.

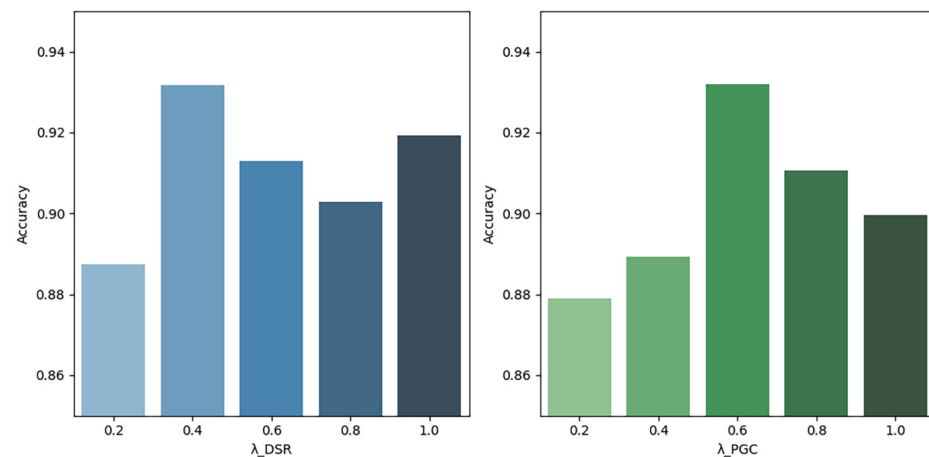


Figure 15. Impact of λ_{DSR} and λ_{PGC} on model accuracy.

5. Discussion

Medical image classification presents significant challenges due to the scarcity of well-labeled data, high annotation costs, and the variability in imaging techniques and pathological conditions. This imbalance between abundant unlabeled data and limited labeled samples, especially for rare diseases, hinders the development of accurate deep learning models. Additionally, the noise and imbalance in unlabeled data complicate the generation of reliable pseudo-labels, further challenging model training.

To address these challenges, we propose the Dynamic Sample Reweighting and Pseudo-Label Guided Contrastive Learning (DSRPGC) framework. This method utilizes DenseNet-121 to extract rich image features, capturing intricate patterns and representations from the input images. By dynamically adjusting the weights of unlabeled samples, we effectively filter out noisy data, enhancing pseudo-label reliability while significantly reducing the risk of error propagation. Furthermore, the Medical Multi-scale Feature Fusion Network (MedFuseNet) captures and integrates information across various scales, addressing the limitations of traditional methods in handling complex medical images. Finally, the Pseudo-Label Guided Contrastive Learning mechanism optimizes feature representations, reducing dependence on labeled data. Overall, our approach demonstrates notable improvements in pseudo-label quality, model robustness, and multi-scale feature processing, offering a more efficient solution for medical image classification.

Despite its strengths, the method has some limitations. While the Dynamic Sample Reweighting Strategy improves performance by prioritizing reliable samples, it may overemphasize certain samples, potentially leading to imbalanced class representation. Additionally, although effective in integrating multi-scale features, the Medical Multi-Scale Feature Fusion Network (MedFuseNet) may suffer from computational inefficiencies, particularly when applied to large-scale datasets. This inefficiency arises from the increased complexity and memory requirements associated with multi-scale feature extraction and fusion processes, which can significantly slow down the training and inference stages for large datasets, such as those commonly encountered in clinical settings. To address these computational challenges, future work could explore model optimization strategies, such as efficient feature fusion techniques, network pruning, or quantization, to reduce the computational burden without sacrificing performance.

The reliance on high-quality pseudo-labels also presents a challenge, as inaccurate labels can negatively impact the model's performance. Moreover, the complexity of hyperparameter tuning in the Pseudo-Label Guided Contrastive Learning (PGC) loss function complicates model optimization across diverse datasets. These factors highlight the need for more robust strategies to ensure the accuracy of pseudo-labels and simplify hyperparameter search processes, particularly in dynamic, real-world environments.

Looking ahead, in future work, we aim to test our method on additional medical imaging datasets to assess its adaptability across diverse image types and medical conditions. We believe that by refining the algorithm's scalability and robustness, it can be effectively applied to a broader range of medical imaging tasks. Furthermore, we aim to address the four key issues identified in the current method. First, adaptive weighting techniques or class balancing strategies will be explored to optimize the Dynamic Sample Reweighting process, alleviating class imbalance problems. Second, the computational efficiency of the Medical Multi-scale Feature Fusion Network (MedFuseNet) will be improved by investigating model compression and efficient fusion techniques, such as attention mechanisms and hierarchical feature aggregation. Third, the quality of pseudo-labels will be enhanced by incorporating self-training, co-training, or class-conditional optimization methods, particularly in cases of extreme class imbalance. Finally, automated hyperparameter optimization techniques will be employed to simplify the hyperparameter tuning process in the Pseudo-Label Guided Contrastive Learning (PGC) loss function, improving the model's performance and generalization across diverse datasets. These improvements will significantly enhance the model's robustness, computational efficiency, and adaptability, paving the way for its broader application in real-world clinical settings.

6. Conclusions

This paper introduces a novel approach to advancing semi-supervised medical image classification by employing Dynamic Sample Reweighting and Pseudo-Label Guided Contrastive Learning (DSRPGC). Our method effectively leverages extensive unlabeled data, addressing the issue of limited labeled data. The framework integrates a Mean Teacher (MT) model with a Dynamic Sample Reweighting strategy to enhance the reliability of unlabeled data. Additionally, we implemented diverse data augmentation strategies to maintain model consistency and stability across various perturbations. To improve the model's capacity to analyze intricate medical images, we present the Medical Multi-Level Feature Fusion Network (MedFuseNet), which combines multi-level feature information to improve classification accuracy. Furthermore, we propose a Pseudo-Label Guided Contrastive (PGC) loss function that incorporates a momentum-updated feature queue mechanism, pseudo-label-guided positive and negative sample pairs, and dynamically weighted contrastive learning. This combination not only strengthens the model's feature representation capabilities but also optimizes both intra-class compactness and inter-class separability. Overall, our experimental results on two public datasets validate the effectiveness of our method and provide valuable insights and directions for future research.

Author Contributions: Funding acquisition, K.L. and S.L.; methodology, J.L.; software, J.L.; supervision, K.L.; validation, J.L.; writing—original draft, K.L. and J.L.; writing—review and editing, K.L. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is sponsored by the National Natural Science Foundation of China Grant No. 62271302 and the Shanghai Municipal Natural Science Foundation Grant 20ZR1423500.

Data Availability Statement: The public datasets ISIC2018, NCT-CRC-HE, and ChestX-ray14 used in this study can be found at the following links: <https://challenge.isic-archive.com/data/#2018>, <https://zenodo.org/records/1214456>, and <https://nihcc.app.box.com/v/ChestXray-NIHCC> (accessed on 1 October 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Shen, D.; Wu, G.; Suk, H.-I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)] [[PubMed](#)]
- Suganyadevi, S.; Seethalakshmi, V.; Balasamy, K. A Review on Deep Learning in Medical Image Analysis. *Int. J. Multimed. Inf. Retr.* **2022**, *11*, 19–38. [[CrossRef](#)]
- Van Engelen, J.E.; Hoos, H.H. A Survey on Semi-Supervised Learning. *Mach. Learn.* **2020**, *109*, 373–440. [[CrossRef](#)]
- Ouali, Y.; Hudelot, C.; Tami, M. An Overview of Deep Semi-Supervised Learning. *arXiv* **2020**, arXiv:2006.05278. [[CrossRef](#)]
- Song, Z.; Yang, X.; Xu, Z.; King, I. Graph-Based Semi-Supervised Learning: A Comprehensive Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 8174–8194. [[CrossRef](#)]
- Feng, Z.; Zhou, Q.; Gu, Q.; Tan, X.; Cheng, G.; Lu, X.; Shi, J.; Ma, L. DMT: Dynamic Mutual Training for Semi-Supervised Learning. *Pattern Recognit.* **2022**, *130*, 108777. [[CrossRef](#)]
- Chang, J.-H.; Weng, H.-C. Fully Used Reliable Data and Attention Consistency for Semi-Supervised Learning. *Knowl.-Based Syst.* **2022**, *249*, 108837. [[CrossRef](#)]
- You, C.; Zhou, Y.; Zhao, R.; Staib, L.; Duncan, J.S. SimCVD: Simple Contrastive Voxel-Wise Representation Distillation for Semi-Supervised Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2022**, *41*, 2228–2237. [[CrossRef](#)]
- Qian, L.; Huang, H.; Xia, X.; Li, Y.; Zhou, X. Automatic Segmentation Method Using FCN with Multi-Scale Dilated Convolution for Medical Ultrasound Image. *Vis. Comput.* **2023**, *39*, 5953–5969. [[CrossRef](#)]
- Tarvainen, A.; Valpola, H. Mean Teachers Are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning Results. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
- Miyato, T.; Maeda, S.; Koyama, M.; Ishii, S. Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 1979–1993. [[CrossRef](#)] [[PubMed](#)]
- Zhang, Y.; Davison, B.D. Efficient Pre-Trained Features and Recurrent Pseudo-Labeling in Unsupervised Domain Adaptation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 2719–2728. [[CrossRef](#)]
- Laine, S.; Aila, T. Temporal Ensembling for Semi-Supervised Learning. *arXiv* **2016**, arXiv:1610.02242. [[CrossRef](#)]
- Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Correction: Corrigendum: Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* **2017**, *546*, 686. [[CrossRef](#)] [[PubMed](#)]

15. Guo, X.; Yuan, Y. Semi-Supervised WCE Image Classification with Adaptive Aggregated Attention. *Med. Image Anal.* **2020**, *64*, 101733. [[CrossRef](#)] [[PubMed](#)]
16. Hu, Z.; Yang, Z.; Hu, X.; Nevatia, R. SIMPLE: Similar Pseudo Label Exploitation for Semi-Supervised Classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 15099–15108. [[CrossRef](#)]
17. Lee, D.-H. Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks. In *Workshop on Challenges in Representation Learning*; ICML: Atlanta, GA, USA, 2013; Volume 3, p. 896.
18. Wang, J.; Qiao, L.; Zhou, S.; Zhou, J.; Wang, J.; Li, J.; Ying, S.; Chang, C.; Shi, J. Weakly Supervised Lesion Detection and Diagnosis for Breast Cancers with Partially Annotated Ultrasound Images. *IEEE Trans. Med. Imaging* **2024**, *43*, 2509–2521. [[CrossRef](#)]
19. Liu, Q.; Yu, L.; Luo, L.; Dou, Q.; Heng, P.A. Semi-Supervised Medical Image Classification with Relation-Driven Self-Ensembling Model. *IEEE Trans. Med. Imaging* **2020**, *39*, 3429–3440. [[CrossRef](#)]
20. Shakya, K.S.; Alavi, A.; Porteous, J.; K, P.; Laddi, A.; Jaiswal, M. A Critical Analysis of Deep Semi-Supervised Learning Approaches for Enhanced Medical Image Classification. *Information* **2024**, *15*, 246. [[CrossRef](#)]
21. Chen, X.; Yu, G.; Tan, Q.; Wang, J. Weighted Samples Based Semi-Supervised Classification. *Appl. Soft Comput.* **2019**, *79*, 46–58. [[CrossRef](#)]
22. Bagherzadeh, J.; Asil, H. A review of various semi-supervised learning models with a deep learning and memory approach. *Iran J. Comput. Sci.* **2019**, *2*, 65–80. [[CrossRef](#)]
23. Shi, W.; Gong, Y.; Ding, C.; Tao, Z.M.; Zheng, N. Transductive Semi-Supervised Deep Learning Using Min-Max Features. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 299–315.
24. Liu, F.; Tian, Y.; Chen, Y.; Liu, Y.; Belagiannis, V.; Carneiro, G. ACPL: Anti-Curriculum Pseudo-Labeling for Semi-Supervised Medical Image Classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 20697–20706. [[CrossRef](#)]
25. Zeng, Q.; Xie, Y.; Lu, Z.; Xia, Y. PEAT: Boosting Semi-Supervised Medical Image Classification via Pseudo-Loss Estimation and Feature Adversarial Training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 15671–15680.
26. Mahmood, M.J.; Raj, P.; Agarwal, D.; Kumari, S.; Singh, P. SPLAL: Similarity-Based Pseudo-Labeling with Alignment Loss for Semi-Supervised Medical Image Classification. *Biomed. Signal Process. Control* **2024**, *89*, 105665. [[CrossRef](#)]
27. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the International Conference on Machine Learning (ICML), Online, 13–18 July 2020; pp. 1597–1607.
28. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised Contrastive Learning. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2020**, *33*, 18661–18673.
29. Li, J.; Xiong, C.; Hoi, S.C. CoMatch: Semi-Supervised Learning with Contrastive Graph Regularization. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 9475–9484. [[CrossRef](#)]
30. Müller, R.; Kornblith, S.; Hinton, G.E. When Does Label Smoothing Help? *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2019**, *32*.
31. Biswas, M.; Buckchash, H.; Prasad, D.K. pNCLR: Stochastic pseudo neighborhoods for contrastive learning based unsupervised representation learning problems. *Neurocomputing* **2024**, *593*, 127810. [[CrossRef](#)]
32. Wu, H.; Zhang, B.; Chen, C.; Qin, J. Federated Semi-Supervised Medical Image Segmentation via Prototype-Based Pseudo-Labeling and Contrastive Learning. *IEEE Trans. Med. Imaging* **2023**, *42*, 1234–1245. [[CrossRef](#)]
33. Liu, L.; Zhang, H.; Wang, Y. Contrastive Mutual Learning with Pseudo-Label Smoothing for Hyperspectral Image Classification. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–11. [[CrossRef](#)]
34. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic Meta-learning for Fast Adaptation of Deep Networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1126–1135.
35. Xie, Q.; Dai, Z.; Hovy, E.; Luong, T.; Le, Q. Unsupervised Data Augmentation for Consistency Training. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6256–6268. [[CrossRef](#)]
36. Grandvalet, Y.; Bengio, Y. Semi-supervised Learning by Entropy Minimization. *Adv. Neural Inf. Process. Syst.* **2004**, *17*.
37. Chen, X.; Fan, H.; Girshick, R.; He, K. Improved Baselines with Momentum Contrastive Learning. *arXiv* **2020**, arXiv:2003.04297. [[CrossRef](#)]
38. Codella, N.; Rotemberg, V.; Tschandl, P.; Celebi, M.E.; Dusza, S.; Gutman, D.; Helba, B.; Kalloo, A.; Liopyris, K.; Marchetti, M.; et al. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). *arXiv* **2019**, arXiv:1902.03368. [[CrossRef](#)]
39. 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 Med.* **2019**, *16*, 1002730. [[CrossRef](#)]
40. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. ChestX-ray8: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 2097–2106. [[CrossRef](#)]

41. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708. [[CrossRef](#)]
42. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980. [[CrossRef](#)]
43. Kim, B.; Choo, J.; Kwon, Y.-D.; Joe, S.; Min, S.; Gwon, Y. SelfMatch: Combining Contrastive Self-Supervision and Consistency for Semi-Supervised Learning. *arXiv* **2021**, arXiv:2101.06480. [[CrossRef](#)]
44. Sohn, K.; Berthelot, D.; Carlini, N.; Zhang, Z.; Zhang, H.; Raffel, C.A.; Cubuk, E.D.; Kurakin, A.; Li, C.-L. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2020**, *33*, 596–608.
45. Zheng, M.; You, S.; Huang, L.; Wang, F.; Qian, C.; Xu, C. SimMatch: Semi-Supervised Learning with Similarity Matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 14471–14481. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.