

Article

Semi-Supervised Medical Image Classification Combined with Unsupervised Deep Clustering

Bang Xiao and Chunyue Lu *

Digital Design and Intelligent Manufacturing Laboratory, North University of China, Taiyuan 030051, China; s202102015@st.nuc.edu.cn

* Correspondence: luchunyue@nuc.edu.cn; Tel.: +86-136-0351-4911

Abstract: An effective way to improve the performance of deep neural networks in most computer vision tasks is to improve the quantity of labeled data and the quality of labels. However, in the analysis and processing of medical images, high-quality annotation depends on the experience and professional knowledge of experts, which makes it very difficult to obtain a large number of high-quality annotations. Therefore, we propose a new semi-supervised framework for medical image classification. It combines semi-supervised classification with unsupervised deep clustering. Spreading label information to unlabeled data by alternately running two tasks helps the model to extract semantic information from unlabeled data, and prevents the model from overfitting to a small amount of labeled data. Compared with current methods, our framework enhances the robustness of the model and reduces the influence of outliers. We conducted a comparative experiment on the public benchmark medical image dataset to verify our method. On the ISIC 2018 Dataset, our method surpasses other methods by more than 0.85% on AUC and 1.08% on Sensitivity. On the ICIAR BACH 2018 dataset, our method achieved 94.12% AUC, 77.92% F1-score, 77.69% Recall, and 78.16% Precision. The error rate is at least 1.76% lower than that of other methods. The result shows the effectiveness of our method in medical image classification.

Keywords: medical image classification; semi-supervised learning; unsupervised learning; deep clustering; overclustering

**Citation:** Xiao, B.; Lu, C.Semi-Supervised Medical Image Classification Combined with Unsupervised Deep Clustering. *Appl. Sci.* **2023**, *13*, 5520. <https://doi.org/10.3390/app13095520>

Academic Editor: Yu-Dong Zhang

Received: 27 March 2023

Revised: 25 April 2023

Accepted: 27 April 2023

Published: 28 April 2023



Copyright: © 2023 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/>).

1. Introduction

With the development of deep learning technology, deep neural networks have become widely used in medical image analysis and processing. However, due to the need for professional knowledge, it is difficult to manually label medical images on a large scale. At the same time, there are constantly unlabeled data being generated in the clinical medical process, which results in a small number of labeled images and a large number of unlabeled images being generated. In this case, semi-supervised learning [1,2], which can use both labeled data and unlabeled data, has certain advantages. In this paper, we propose a semi-supervised learning framework combined with unsupervised learning [3–5], which can enhance the model's ability to extract information from unlabeled data, reduce the risk of overfitting, and effectively reduce the impact of outlier data.

Slight disturbances in an image do not affect a human's basic judgment of the image's semantics. Based on this feature, people apply the consistency regularization method [6–8] to semi-supervised learning, and generate different samples by adding different disturbances (such as Gaussian noise, flipping, cropping, etc.) to the input samples, and then encourage the model to generate similar predictions for the different disturbed versions of the same input. Theoretically, a good model can eliminate the influence of disturbance and make self-consistent predictions. For example, in the π -model [6], for any input x_i , the model generates two different disturbances and inputs them into the feature extraction network to obtain two predictions. The differences between these predictions are then

used as the objective function to optimize the model. The TE model [6] takes the EMA (exponential moving average) predictions in different epochs as the comparison target. The MT (mean teacher) model [7] has two parallel feature extraction networks, and different disturbance samples are input into different networks, in which the teacher network uses the EMA weight of the student network, and the output results of the two networks are regarded as the consistency target. However, the current semi-supervised method also has some limitations. For example, the unsupervised phase of the model depends on the label information of supervised learning, and it is difficult to avoid over-fitting when the scale of labeled data is small. For example, most semi-supervised methods assume that labeled data and unlabeled data obey the same distribution and have the same label space, which makes the occasional outlier in medical images (including instrument failure, improper operation, diseases that are not within the prediction range, etc.) often have negative effects on the quality of prediction.

In this paper, we introduce an unsupervised learning process that is completely independent of semi-supervised learning, which ensures that this part of the training is not affected by supervised learning, so we can produce pseudo-labels [9] that are not affected by label information. As shown in Figure 1, we represent the model as two phases: a semi-supervised phase and an unsupervised phase, which share the feature extraction network. Among them, the semi-supervised phase is the main task to perform the image classification task, and the unsupervised phase is the secondary task to cluster the data [10–13]. Our semi-supervised network architecture is based on the basic architecture of FixMatch [14]. Our work focuses on the unsupervised clustering phase.

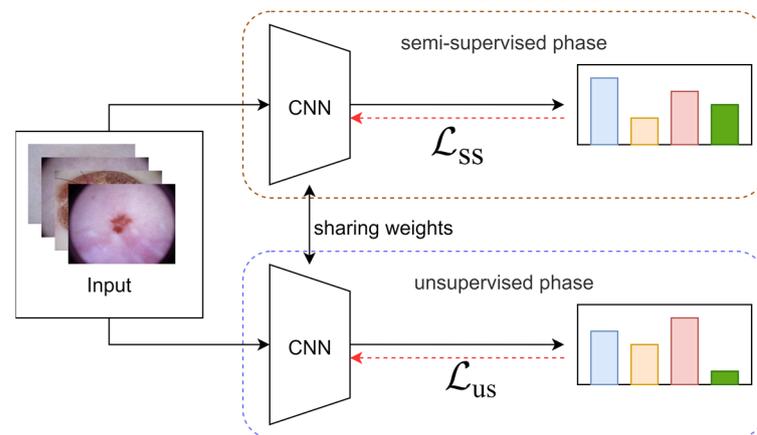


Figure 1. Overview of our two-phase framework for medical image classification. A semi-supervised phase and an unsupervised phase run alternately, and the feature extraction networks of the two phases share weights. In the semi-supervised phase, all data information is used to calculate the semi-supervised loss (\mathcal{L}_{SS}), and in the unsupervised phase, all data without label information is used to calculate the unsupervised loss (\mathcal{L}_{US}).

In order to prevent cluster degradation in the iterative process, we design a network structure similar to a Siamese Network and update the weights by momentum [15]. At the same time, the auxiliary overclustering method [16] is adopted to make the outliers less likely to affect the clustering results. Our main contributions are as follows:

1. We proposed a new semi-supervised learning framework to improve the classification performance of the semi-supervised model by introducing an unsupervised deep clustering algorithm.
2. Based on consistency regularization, a new objective function was designed to cross-verify different augmentations of the same image, and use an auxiliary overclustering method in the clustering process to reduce the influence of outliers on model training and ensure intra-class consistency in feature space.

3. We conducted experiments on two medical datasets, and the results show that our method is effective. Compared with the current semi-supervised medical image classification methods, our method has better performance.

2. Related Work

In this section, we review some basic methods related to our work and the recent developments in these fields. This includes semi-supervised learning in the field of medical image analysis, data augmentation methods, and deep clustering [17,18].

2.1. Semi-Supervised Learning in Medical Image Analysis

The characteristics of semi-supervised learning make it particularly advantageous for medical image analysis. In fact, the research on semi-supervised learning in this field [19–22] has been carried out for a long time. The tasks of semi-supervised learning in the field of medical image analysis can be divided into image classification [23,24] and semantic segmentation [25–28]. Methodologically, the field of medical image analysis mainly includes methods based on confrontation and methods based on consistency regularization.

In the field of medical image segmentation, Zhao et al. proposed a pseudo supervision method based on consistency regularization and uncertainty evaluation [29], which reduced the influence of the potential noise of pseudo-labels on segmentation results in semi-supervised learning. Dong et al. [30] proposed an unsupervised domain adaptive framework based on an antagonistic network, and applied it to semi-supervised learning on the JSRT (Japanese Society of Radiological Technology) datasets.

In the field of medical image classification, Liu et al. [31] put forward sample relation consistency on the basis of consistency regularization, so that the semi-supervised model can obtain more semantic information from unlabeled data. Pang et al. [32] proposed a data augmentation method based on a GAN (generative adversarial network) for semi-supervised learning.

2.2. Data Augmentation Methods

At present, there are many semi-supervised works based on consistency regularization, which improve the performance of the model by adding different disturbances to the same target and encouraging the model to generate consistent prediction. Although good results have been achieved, their methods of adding disturbances are still relatively primary, and most of them are based on Gaussian noise, dropout noise, or adversarial noise. Obviously, there are reasons for the complexity of data and training cost, but the effect of data augmentation is obviously not fully exerted. Some targeted data augmentation strategies have been proved to be effective in supervised learning. At present, data augmentation is also one of the keys to improving the model's performance in the field of semi-supervised learning. The data augmentation strategy of Mixup proposed by Berthelot et al. in [1] significantly enhances the robustness and generalization ability of the model and improves the model's performance. The AutoAugment strategy proposed by Cubuk et al. in [33] "Learning Augmentation Policies from Data," improves the data augmentation method through automatic search, and achieves better results in the same semi-supervised framework in multiple datasets. Subsequently, on the basis of AutoAugment, RandAugment [34] changed the search strategy and reduced the parameter space of data augmentation, and achieved better results while the complexity and training cost were significantly reduced.

2.3. Deep Clustering

Clustering is a widely used unsupervised algorithm, which divides a group of data into different clusters according to certain standards, so that the data in the same cluster are as similar as possible and the data in different clusters are as dissimilar as possible. Different clustering algorithms are based on different assumptions and data distribution, and the same group of data can also use different clustering methods, such as partition-based clustering methods such as k-means, density-based clustering methods such as DBSCAN

(Density-Based Spatial Clustering of Applications with Noise) [35], and distribution-based clustering methods such as Gaussian Mixture [36]. With the increase in data complexity, the effect of these traditional clustering algorithms is gradually declining. To solve this problem, Caron et al. put forward the concept of deep clustering in [17], that is, combining deep learning with clustering, and its basic framework is shown in Figure 2.

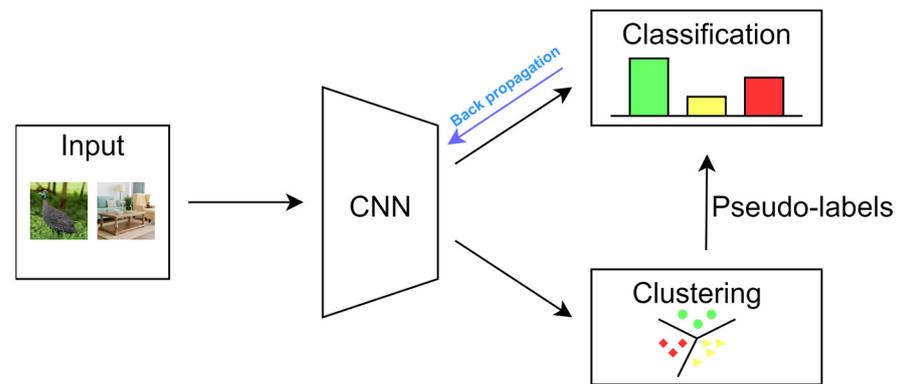


Figure 2. Overall process of deep clustering.

The deep clustering model firstly clusters the features, then generates pseudo-labels based on the clustering results, updating the parameters of the network, and making the network predict these pseudo-labels. These two processes are alternately carried out. Many excellent works have been improved on this basis. Joint Unsupervised Learning [37] proposes a cyclic framework of joint unsupervised learning for depth representation and image clustering, which updates network parameters in a hierarchical way. The learning presentation for clustering via prototype patterning and positive sampling [38] firstly maximizes the distance between different prototypes, and then closes the distance of samples in the same clustering center to improve the clustering effect. Its model takes into account both Contrastive Learning and non-Contrastive Learning, and trains in an end-to-end manner. Invariant information clustering (IIC) [16] maximizes the mutual information between two different augmentations of the same image and generates clustering of the data based on mutual information loss.

3. Method

In this paper, we propose a semi-supervised classification framework that combines unsupervised learning. As shown in Figure 1, our method includes two tasks: unsupervised clustering and semi-supervised classification. These tasks are carried out alternately in the training process. The semi-supervised learning phase uses all data containing label information, while the unsupervised phase uses all data without label information. When the number of labels is small, this method can effectively prevent the model from overfitting and improve its generalization ability.

This paper focuses on unsupervised clustering. We design an unsupervised clustering framework based on the basic framework of deep clustering and propose an objective function that is more conducive to clustering. This allows the unsupervised part to reduce the influence of outliers on the model and enhance its ability to extract semantic information from unlabeled data.

3.1. Semi-Supervised Phase

In the semi-supervised phase, we adopt the FixMatch semi-supervised framework, which is based on consistency regularization and pseudo-labeling. It combines and simplifies advanced ideas and methods used in models such as UDA (unsupervised data augmentation) [39] and ReMixMatch [40], and has achieved good results on massive datasets.

3.1.1. Data Augmentation Method

We use weak augmentation for labeled data and strong augmentation for unlabeled data. Weak augmentation is a standard flip-and-shift augmentation strategy that flips with a 50% probability and shifts with a 12.5% probability in both horizontal and vertical directions. Strong augmentation mainly applies two strategies: RandAugment and CTAugment, followed by Cutout [41].

3.1.2. Semi-Supervised Loss Function and Details

In the semi-supervised phase, we adopt different augmentation methods for labeled and unlabeled data (as shown in Figure 3).

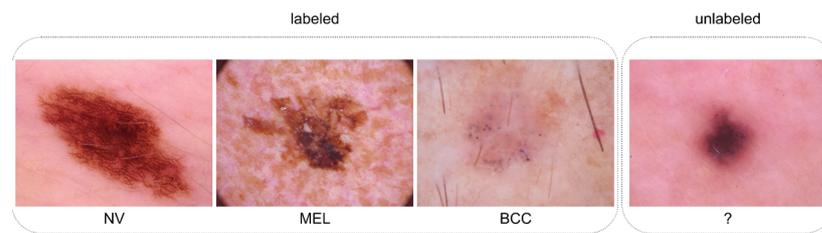


Figure 3. The type of input data divided into two categories according to whether they are labeled or not.

Specifically, labeled images are represented as $x = \{x_1, \dots, x_n\}$, and their labels are represented as $y = \{y_1, \dots, y_n\}$. The unlabeled images are represented as $u = \{u_1, \dots, u_{\mu n}\}$, and the network is represented as f_{θ} . $A(\cdot)$ represents a strong augmentation function and $a(\cdot)$ represents a weak augmentation function. The loss of annotation data can be expressed as:

$$\mathcal{L}_s = \frac{1}{n} \sum_{i=1}^n H(f_{\theta}(a(x_i)), y_i) \tag{1}$$

The Pseudo-labels of unlabeled data can be expressed as:

$$y'_i = \operatorname{argmax}(f_{\theta}(A(u_i))) \tag{2}$$

The loss of unlabeled data can be expressed as:

$$\mathcal{L}_u = \frac{1}{\mu n} \sum_{i=1}^{\mu n} 1(\max(f_{\theta}(A(u_i))) \geq \tau) H(f_{\theta}(A(u_i)), y'_i) \tag{3}$$

where μ represents the ratio of unlabeled data to labeled data; $H(\cdot, \cdot)$ represents cross entropy loss; τ is a scalar hyperparameter denouncing the threshold, and when the confidence exceeds the threshold, the pseudo-label is retained. Therefore, the final loss in the semi-supervised phase can be expressed as:

$$\mathcal{L}_{ss} = \mathcal{L}_s + \lambda_u \mathcal{L}_u \tag{4}$$

Among them, the hyperparameter λ_u is the relative weight of unsupervised loss, which is used to balance supervised loss and unsupervised loss (use the same hyperparameter settings as FixMatch did in most datasets).

3.2. Unsupervised Phase

In this paper, we use a clustering method for unsupervised learning. There are many kinds of clustering algorithms, and their results vary when used alone. However, considering that different clustering algorithms are similar within the framework of deep clustering, we adopt the standard k-means clustering algorithm to reduce computational complexity and

shorten iteration time. In this section, we describe the unsupervised algorithm framework used in this paper and introduce the total loss function in the unsupervised phase.

3.2.1. Unsupervised Clustering Framework

Figure 4 describes our basic process in the unsupervised learning phase. Different augmentations of the same image are input into the network, and then the output results are clustered separately. The clustering results are used as pseudo-labels to update the network parameters. Our framework regularizes the centroid distribution of different clusters to improve the clustering effect and make better use of unlabeled data.

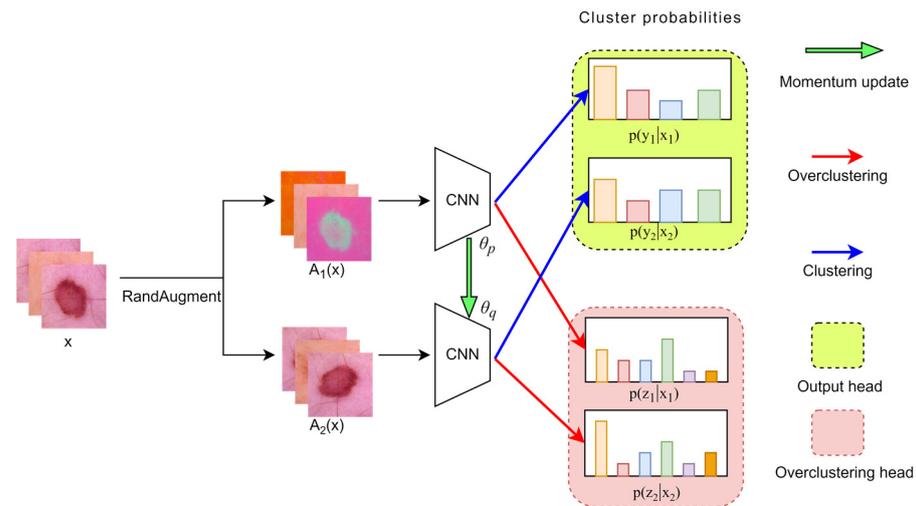


Figure 4. Overview of our unsupervised deep clustering framework. $A_1(x)$ and $A_2(x)$ are the outputs of x after augmentation. $p(y_1|x_1)$ and $p(y_2|x_2)$ are the results of clustering. $p(z_1|x_1)$ and $p(z_2|x_2)$ are the results of overclustering. The online network (above) and the target network (below) update the parameters in momentum. Image feature representation is clustered by the k-means method to get the pseudo-label. The over-clustering head is used to assist the training process, and only the output head is used in the test.

We denote unlabeled data as $x = \{x_1, \dots, x_N\}$, $y_i (i = 1, 2, \dots, N)$ represents the corresponding pseudo-label, f_θ represents the parameterized model, $A'(\cdot)$ represents the augmentation function, and η represents the clustering center. We adopt the RandAugment [34] augmentation method in the unsupervised phase (parameters settings are the same as the experiment on cifar10). The optimization goal can be expressed as:

$$\min_{\theta} \sum_i \mathcal{L}(f_\theta(A(x_i)), y_i, \eta) \tag{5}$$

$$\mathcal{L}(f_\theta(x_i), y_i, \eta) = -\log \left(\frac{e^{-d(f_\theta(x_i), \eta y_i)}}{\sum_j e^{-d(f_\theta(x_i), \eta y_j)}} \right)$$

Here $d(\cdot, \cdot)$ represents cosine distance.

We design a parallel network structure consisting of an online network and a target network. Two different augmentations of the same image are input into the online network and target network. The obtained vectors are then clustered, and a clustering center is assigned to each image x_i . This process does not depend on labels at all. Usually, unsupervised models eventually collapse into prone mode and output invalid solutions without any constraints. To prevent this situation, the usual practice is to add reconstruction loss [42,43], or explicitly constrain the uniform distribution of samples of all kinds [17]. Inspired by Contrastive Learning [15,44], we make the online network and target network not share weights but update their parameters by momentum. This keeps differences

between them at all times, providing motivation for continuous updating. Specifically, the parameters of the two networks are recorded as θ_p and θ_q , and we update θ_q by:

$$\theta_q \leftarrow m\theta_q + (1 - m)\theta_p \tag{6}$$

where m is a hyperparameter representing momentum. The greater the value of m , the slower the updating speed. In our experiment, $m = 0.99$.

3.2.2. Loss Function in Unsupervised Phase

Formally, for any picture x_i , the model generates two feature vectors $v_i^{(1)}$ and $v_i^{(2)}$:

$$\begin{aligned} v_i^{(1)} &= f_{\theta_p}(A'(x_i)) \\ v_i^{(2)} &= f_{\theta_q}(A'(x_i)) \end{aligned} \tag{7}$$

By clustering these two feature vectors, respectively, we can get two groups of pseudo-labels and centroids:

$$\begin{aligned} \mathbf{y}^{(1)}, \eta^{(1)} &= \underset{\mathbf{y}, \eta}{\operatorname{argmin}} \sum_i \|v_i^{(1)} - \eta_{y_i}\|^2 \\ \mathbf{y}^{(2)}, \eta^{(2)} &= \underset{\mathbf{y}, \eta}{\operatorname{argmin}} \sum_i \|v_i^{(2)} - \eta_{y_i}\|^2 \end{aligned} \tag{8}$$

We hope that the clustering label is consistent with its corresponding input, so we first calculate the internal loss of the same picture:

$$\mathcal{L}_{inward} = \sum_i \mathcal{L}(v_i^{(1)}, y_i^{(1)}, \eta^{(1)}) + \mathcal{L}(v_i^{(2)}, y_i^{(2)}, \eta^{(2)}) \tag{9}$$

At the same time, according to the principle of semantic consistency of the same picture, different augmentations will not change the semantic information, so we can also match the pseudo-label and centroid with different views:

$$\mathcal{L}_{cross} = \sum_i \mathcal{L}(v_i^{(1)}, y_i^{(2)}, \eta^{(2)}) + \mathcal{L}(v_i^{(2)}, y_i^{(1)}, \eta^{(1)}) \tag{10}$$

The final loss of clustering process \mathcal{L}_C can be simply expressed as:

$$\mathcal{L}_C = \frac{1}{N} (\mathcal{L}_{inward} + \mathcal{L}_{cross}) \tag{11}$$

To improve the generalization ability of the network and reduce the influence of outliers, we add an overclustering head to the network. We use k clustering centroids in the output head and $k + \lfloor k/2 \rfloor$ clustering centers in the overclustering head. Some data or outliers with low confidence will be divided into additional clustering centers during overclustering. The output head still maintains the same number of predictions as the actual category, improving the model's utilization efficiency for unlabeled data and increasing expressiveness in the learned feature representation. Similar to the previous clustering process, we obtain two groups of pseudo-labels and clustering centroids $(z^{(1)}, \rho^{(1)})$, $(z^{(2)}, \rho^{(2)})$ in the overclustering process. The loss can be calculated in the same way:

$$\mathcal{L}'_{inward} = \sum_i \mathcal{L}(v_i^{(1)}, z_i^{(1)}, \rho^{(1)}) + \mathcal{L}(v_i^{(2)}, z_i^{(2)}, \rho^{(2)}) \tag{12}$$

$$\mathcal{L}'_{cross} = \sum_i \mathcal{L}(v_i^{(1)}, z_i^{(2)}, \rho^{(2)}) + \mathcal{L}(v_i^{(2)}, z_i^{(1)}, \rho^{(1)}) \tag{13}$$

The final loss of the over-clustering process \mathcal{L}_{OC} is expressed as:

$$\mathcal{L}_{OC} = \frac{1}{N} (\mathcal{L}'_{inward} + \mathcal{L}'_{cross}) \tag{14}$$

Finally, we express the loss function of the unsupervised phase as:

$$\mathcal{L}_{US} = \mathcal{L}_C + \mathcal{L}_{OC} \quad (15)$$

3.3. Semi-Supervised Classification Method Combined with Unsupervised Learning

In both the unsupervised and semi-supervised phases, we use the same feature extraction network. The model alternates between semi-supervised classification and unsupervised clustering. In the semi-supervised learning phase, unsupervised clustering becomes easier by spreading label information. In the unsupervised phase, by ignoring label information, the model can obtain stronger generalization performance and avoid overfitting a small amount of labeled data. At the same time, using an overclustering method reduces the influence of low-confidence targets or outliers on network parameters. After trying different hyperparameters and based on final experimental results, we set the model to perform semi-supervised training four times first and then unsupervised training once.

4. Experiments

In this section, we evaluate our method on two large public datasets (the ISIC 2018 Dataset and the ICIAR BACH 2018 dataset) and compare it with state-of-the-art methods.

4.1. Datasets

4.1.1. ISIC 2018

The ISIC 2018 dataset [45] contains 10,015 skin lesion detection images marked as 7 common skin lesion types. The images in the ISIC 2018 dataset are dermoscopy images taken with a dermatoscope. Dermoscopy images provide more details and features of skin lesions than conventional photography, such as colors, patterns, and structures. The images in the dataset come from different sources and have varying resolutions, formats, and quality. We resize the image to 224×224 and normalize them using statistical data based on the ImageNet dataset. We divide the dataset into training, validation, and test sets in a 7:1:2 ratio. The number of clustering centers k is 7, and the number of overclustering centers is 10.

4.1.2. ICIAR BACH 2018

The ICIAR BACH 2018 [46] dataset contains hematoxylin and eosin-stained images of breast tissue commonly used for histopathological diagnosis of breast cancer. The images are divided into two types: microscopy images and whole-slide images. Microscopy images are 2048×1536 pixels in size and are extracted from different regions of interest within whole-slide images. Whole-slide images are very large (typically over $100,000 \times 100,000$ pixels) and contain multiple tissue samples. The images are annotated with four classes: normal, benign, in situ, and invasive. To eliminate overfitting, we augment the original dataset and adopt a nucleus-based patch extraction approach to extract 21,705 non-overlapping nucleus patches (a total of 22,105 non-overlapping nucleus patches are extracted), including 13,026 cancerous nucleus patches and 8679 non-cancerous nucleus patches. We divide the dataset into training, validation, and test sets in a 7:1:2 ratio. The number of clustering centers k is 4, and the number of overclustering centers is 6.

4.2. Metrics

For the ISIC 2018 dataset, our evaluation indicators include AUC (Area Under Curve), accuracy, sensitivity, and specificity.

For the ICIAR BACH 2018 dataset, our evaluation indicators include AUC, Error rate, Precision, Recall, and F1-score.

4.3. Implementation Details

We use Wide-ResNet-28-2 pre-trained on ImageNET as the backbone network of all experiments and use the same parameter settings as FixMatch in the semi-supervised

phase. In the unsupervised clustering phase, we use the SGD (stochastic gradient descent) optimizer with a batch size of 64. The learning rate and weight decay are set to 0.01 and 0.0002, respectively. All methods train for 100 epochs on the ISIC 2018 dataset (our method trains for 80 epochs in the semi-supervised phase and 20 epochs in the unsupervised phase) and 20 epochs on the ICIAR BACH 2018 dataset (our method trains for 16 epochs in the semi-supervised phase and 4 epochs in the unsupervised phase).

4.4. Results

We compare our method with current semi-supervised methods, including TCSE (transformation consistent self-ensembling) [47] (based on Pi-Model), Temporal Ensembling [6], MT [7], UDA [39], and the original FixMatch [14] framework to evaluate its effectiveness.

4.4.1. Comparison on ISIC 2018 Dataset

To clarify the actual effect of semi-supervised learning, we use a fully supervised model trained with 10% labeled data as a comparison baseline. We test the performance of these semi-supervised methods with 10% labeled data. As shown in Table 1, the performance of all models is significantly better than the reference baseline. The performance of the TE model is slightly better than that of the TCSE model because it integrates predictions from different periods and uses them as consistency targets, significantly reducing the model's calculation cost. Compared with the TE and TCSE models, the performance of the MT model is further improved thanks to its network structure and consistent target design. The good performance of the UDA model and FixMatch model proves that the advanced data augmentation method has obvious effects on improving the performance of the model. In addition, our method is better than FixMatch in all indicators, which shows that our proposed method can make better use of unlabeled data.

Table 1. Comparison with different semi-supervised learning methods on ISIC 2018 Dataset.

Method	Metrics			
	AUC	Accuracy	Sensitivity	Specificity
Baseline	85.17	89.42	63.11	87.25
TCSE	87.15	89.91	65.56	90.05
TE	87.23	90.05	65.69	90.14
MT	89.56	91.36	67.35	91.22
UDA	91.04	91.45	68.89	92.56
FixMatch	91.56	91.87	68.36	92.36
ours	92.41	92.16	69.44	92.45

At the same time, we also tested the actual performance of various methods with less labeled data. As shown in Table 2, under the condition of using 2% and 4% labeled data, our method shows more obvious advantages, and the indexes drop less than other methods, which further proves that our method can still perform well with less labeled data.

4.4.2. Comparison on ICIAR BACH 2018 Dataset

We trained a fully supervised model with all labeled data as the upper benchmark and a fully supervised model with 10% labeled data as the lower benchmark. Table 3 shows the comparison of various methods under different percentages of labeled data, and it can be observed that the model with more advanced data augmentation strategies (Auto Augment, CT Augment, Rand Augment, etc.) can perform better. At the same time, it can be noted that with the decrease in labeled data, the performance attenuation of our method is the least obvious, which reflects that our method has better generalization performance

than the pure semi-supervised method and is more robust in the face of the fluctuation of labeled data.

Table 2. Performance comparison of different methods under different scale labeled data.

Method	Labeled	Metrics			
		AUC	Accuracy	Sensitivity	Specificity
TCSE	2%	84.12	84.36	63.14	85.21
	4%	85.25	86.12	64.21	86.31
TE	2%	84.32	86.14	63.07	85.16
	4%	85.36	87.12	64.22	86.33
MT	2%	87.23	89.75	63.89	89.24
	4%	88.45	90.11	65.14	90.13
UDA	2%	89.36	91.07	66.19	91.14
	4%	90.71	91.34	67.33	91.79
FixMatch	2%	90.13	91.03	66.24	91.45
	4%	90.69	91.25	67.39	92.07
ours	2%	92.04	91.89	68.86	92.14
	4%	92.27	92.11	69.13	92.30

Table 3. Comparison of different semi-supervised methods under different scale labeled data.

Method	Labeled	Metrics			
		Error Rate	Precision	Recall	F1-Score
supervised	100%	21.23	81.25	81.69	81.46
	10%	36.24	59.41	56.36	57.84
TCSE	10%	32.51	64.17	63.21	63.69
	20%	30.09	68.45	66.31	67.36
	40%	28.72	71.84	68.34	70.05
TE	10%	31.21	68.19	67.23	67.71
	20%	28.74	70.13	68.14	69.12
	40%	27.55	72.54	70.04	71.27
MT	10%	30.11	70.88	68.91	69.88
	20%	27.52	72.18	70.21	71.18
	40%	26.25	74.89	72.14	73.49
UDA	10%	27.51	75.14	74.10	74.62
	20%	26.24	76.14	76.33	76.23
	40%	26.03	77.23	76.89	77.06
FixMatch	10%	28.72	73.16	72.89	73.02
	20%	26.27	75.21	74.01	74.61
	40%	25.06	77.14	77.07	77.10
ours	10%	26.11	76.13	76.22	76.17
	20%	25.08	77.89	77.24	77.56
	40%	23.39	78.16	77.69	77.92

We also compare the specific performance of different methods in terms of AUC. As shown in Table 4, with 40% labeled data, the average AUC of our method reaches 94.12%, ahead of all four categories. This proves the effectiveness of our method. At the same time, we notice that the performance of each model shows a gap in different categories. This may be because labeling noise in datasets reduces the effect of consistency regularization and our unsupervised clustering shows better classification ability in this case. The training method

of temporarily ignoring label information in the unsupervised training phase makes it less likely to overfit a small amount of training data.

Table 4. Evaluation of different methods on the ICIAR BACH 2018 dataset with AUC metric.

Method	Labeled	Normal Tissue	Benign Tissue	Ductal Carcinoma in Situ	Invasive Carcinoma	Average AUC
supervised	100%	95.61	94.69	93.16	94.36	94.46
TCSE	40%	89.36	88.12	88.07	88.71	88.57
TE	40%	89.44	87.91	88.26	88.64	88.56
MT	40%	91.12	90.08	89.68	90.11	90.25
UDA	40%	93.11	92.14	90.74	92.07	92.02
FixMatch	40%	94.69	93.25	91.45	93.25	93.16
ours	40%	95.44	94.36	92.89	93.77	94.12

5. Discussion

We designed a semi-supervised classification model with unsupervised learning and compared it with current semi-supervised learning methods on the ISIC 2018 and ICIAR BACH 2018 datasets. On the ISIC 2018 dataset, we achieved significant improvement in AUC and sensitivity. On the ICIAR BACH 2018 dataset, our method performed better on all metrics, especially in AUC and error rate. The results show that our method has stronger classification ability and can distinguish between positive and negative samples more clearly. We believe that this is the result of the combined action of semi-supervised and unsupervised classification mechanisms. When the proportion of labeled data decreases, the advantage of our method becomes more obvious. This shows that intermittently ignoring label information during neural network training can effectively avoid overfitting of the model. We improve the effect of unsupervised clustering by cross-comparison and auxiliary overclustering during unsupervised learning, ultimately improving the performance of the entire model. Considering the high initialization requirement of clustering algorithms, we use a network pre-trained on ImageNet as the feature extraction network. The results of random initialization in our experiments are not ideal, which is a problem that the deep clustering framework needs to address. In future studies, we will try to improve the performance of the model under random initialization and make it have stronger generalization ability.

Semi-supervised learning has great potential in medical image classification. Labeling medical images is often time-consuming and expensive, so semi-supervised learning can leverage large amounts of unlabeled data to improve model accuracy. In reality, a large number of unlabeled medical images are generated every day, making it practical to learn from these data in an unsupervised way. However, current semi-supervised learning methods still have some limitations, such as not adequately addressing the problem of class imbalance or being prone to overfitting when there is insufficient labeled data. We have shown that unsupervised clustering can be combined with semi-supervised learning to improve overall performance. Our work focuses on extracting meaningful semantic information from a large number of unlabeled data and dividing the same category into the same cluster to help the semi-supervised model classify better. In general, our method can be considered a supplementary means of semi-supervised learning, and we hope to reduce the model's dependence on manually labeled data.

6. Conclusions

In this paper, we study the problem of semi-supervised medical image classification with the aim of better utilizing medical images without labels. We propose a new learning framework that combines semi-supervised learning with unsupervised clustering.

The model alternates between semi-supervised and unsupervised learning. During the semi-supervised learning phase, the model propagates label information to accelerate the convergence speed of unsupervised clustering. In the unsupervised learning phase, the model ignores label information to reduce the risk of overfitting. Our experiments on two datasets have achieved better metrics, with a significant lead in AUC, indicating that our proposed method has stronger classification ability. When reducing the ratio of labeled data, our method has more obvious advantages in all metrics, indicating that it can still perform well with less labeled data. We have tried training the model without pre-training, but the overall effect was not satisfactory. Additionally, the unsupervised phase takes quite a long time, which lead to an increase in training costs. In future studies, we will continue to improve our methods so that they can still perform well under random initialization and lower training costs.

Author Contributions: Conceptualization, B.X.; methodology, B.X. and C.L.; software, B.X. and C.L.; validation, B.X.; writing—original draft preparation, B.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; Raffel, C.A. Mixmatch: A holistic approach to semi-supervised learning. *Adv. Neural Inf. Process. Syst.* **2019**, *32*. [[CrossRef](#)]
2. Van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109*, 373–440. [[CrossRef](#)]
3. Barlow, H.B. Unsupervised learning. *Neural Comput.* **1989**, *1*, 295–311. [[CrossRef](#)]
4. Ghahramani, Z. Unsupervised learning. In *Advanced Lectures on Machine Learning: ML Summer Schools 2003; Revised Lectures*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 72–112.
5. Hahne, F.; Huber, W.; Gentleman, R.; Falcon, S.; Gentleman, R.; Carey, V. Unsupervised machine learning. In *Bioconductor Case Studies*; Springer: New York, NY, USA, 2008; pp. 137–157.
6. Laine, S.; Aila, T. Temporal Ensembling for Semi-Supervised Learning. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
7. 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*. [[CrossRef](#)]
8. Luo, Y.; Zhu, J.; Li, M.; Ren, Y.; Zhang, B. Smooth neighbors on teacher graphs for semi-supervised learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8896–8905.
9. Lee, D.-H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Proceedings of the Workshop on Challenges in Representation Learning, ICML, Atlanta, GA, USA, 16–21 June 2013; p. 896.
10. Madhulatha, T.S. An overview on clustering methods. *IOSR J. Eng.* **2012**, *2*, 719–725. [[CrossRef](#)]
11. Sinaga, K.P.; Yang, M.-S. Unsupervised K-means clustering algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [[CrossRef](#)]
12. Murtagh, F.; Contreras, P. Algorithms for hierarchical clustering: An overview. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2012**, *2*, 86–97. [[CrossRef](#)]
13. Vesanto, J.; Alhoniemi, E. Clustering of the self-organizing map. *IEEE Trans. Neural Netw.* **2000**, *11*, 586–600. [[CrossRef](#)]
14. 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.* **2020**, *33*, 596–608.
15. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 9729–9738.
16. Ji, X.; Henriques, J.F.; Vedaldi, A. Invariant information clustering for unsupervised image classification and segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27–28 October 2019; pp. 9865–9874.
17. Caron, M.; Bojanowski, P.; Joulin, A.; Douze, M. Deep clustering for unsupervised learning of visual features. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 132–149.

18. Guo, X.; Liu, X.; Zhu, E.; Yin, J. Deep clustering with convolutional autoencoders. In Proceedings of the Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, 14–18 November 2017; Proceedings, Part II 24. Springer: Berlin/Heidelberg, Germany, 2017; pp. 373–382.
19. Cheplygina, V.; de Bruijne, M.; Pluim, J.P. Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med. Image Anal.* **2019**, *54*, 280–296. [[CrossRef](#)]
20. Bai, W.; Oktay, O.; Sinclair, M.; Suzuki, H.; Rajchl, M.; Tarroni, G.; Glocker, B.; King, A.; Matthews, P.M.; Rueckert, D. Semi-supervised learning for network-based cardiac MR image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, 11–13 September 2017; Proceedings, Part II 20. Springer: Berlin/Heidelberg, Germany, 2017; pp. 253–260.
21. Jin, Y.; Cheng, K.; Dou, Q.; Heng, P.-A. Incorporating temporal prior from motion flow for instrument segmentation in minimally invasive surgery video. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, 13–17 October 2019; Proceedings, Part V 22. Springer: Berlin/Heidelberg, Germany, 2019; pp. 440–448.
22. Zhang, Y.; Yang, L.; Chen, J.; Fredericksen, M.; Hughes, D.P.; Chen, D.Z. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, 11–13 September 2017; Proceedings, Part III 20. Springer: Berlin/Heidelberg, Germany, 2017; pp. 408–416.
23. Li, Q.; Cai, W.; Wang, X.; Zhou, Y.; Feng, D.D.; Chen, M. Medical image classification with convolutional neural network. In Proceedings of the 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), Singapore, 10–12 December 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 844–848.
24. Yadav, S.S.; Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* **2019**, *6*, 113. [[CrossRef](#)]
25. Pham, D.L.; Xu, C.; Prince, J.L. Current methods in medical image segmentation. *Annu. Rev. Biomed. Eng.* **2000**, *2*, 315–337. [[CrossRef](#)]
26. Norouzi, A.; Rahim, M.S.M.; Altameem, A.; Saba, T.; Rad, A.E.; Rehman, A.; Uddin, M. Medical image segmentation methods, algorithms, and applications. *IETE Tech. Rev.* **2014**, *31*, 199–213. [[CrossRef](#)]
27. Chartsias, A.; Joyce, T.; Papanastasiou, G.; Semple, S.; Williams, M.; Newby, D.; Dharmakumar, R.; Tsaftaris, S.A. Factorised spatial representation learning: Application in semi-supervised myocardial segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, 16–20 September 2018; Proceedings, Part II 11. Springer: Berlin/Heidelberg, Germany, 2018; pp. 490–498.
28. Nie, D.; Gao, Y.; Wang, L.; Shen, D. ASDNet: Attention based semi-supervised deep networks for medical image segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, 16–20 September 2018; Proceedings, Part IV 11. Springer: Berlin/Heidelberg, Germany, 2018; pp. 370–378.
29. Zhao, X.; Qi, Z.; Wang, S.; Wang, Q.; Wu, X.; Mao, Y.; Zhang, L. RCPS: Rectified Contrastive Pseudo Supervision for Semi-Supervised Medical Image Segmentation. *arXiv* **2023**, arXiv:2301.05500.
30. Dong, N.; Kampffmeyer, M.; Liang, X.; Wang, Z.; Dai, W.; Xing, E. Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, 16–20 September 2018; Proceedings, Part II 11. Springer: Berlin/Heidelberg, Germany, 2018; pp. 544–552.
31. 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](#)]
32. Pang, T.; Wong, J.H.D.; Ng, W.L.; Chan, C.S. Semi-supervised GAN-based radiomics model for data augmentation in breast ultrasound mass classification. *Comput. Methods Programs Biomed.* **2021**, *203*, 106018. [[CrossRef](#)] [[PubMed](#)]
33. Cubuk, E.D.; Zoph, B.; Mane, D.; Vasudevan, V.; Le, Q.V. AutoAugment: Learning Augmentation Policies from Data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2019; pp. 113–123.
34. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. Randaugment: Practical automated data augmentation with a reduced search space. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 702–703.
35. Ester, M.; Krieger, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *kdd*, 1996; pp. 226–231. Available online: <https://file.biolab.si/papers/1996-DBSCAN-KDD.pdf> (accessed on 2 August 1996).
36. Reynolds, D.A. Gaussian mixture models. *Encycl. Biom.* **2009**, *741*. [[CrossRef](#)]
37. Yang, J.; Parikh, D.; Batra, D. Joint unsupervised learning of deep representations and image clusters. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5147–5156.
38. Huang, Z.; Chen, J.; Zhang, J.; Shan, H. Learning Representation for Clustering Via Prototype Scattering and Positive Sampling. *IEEE Trans. Pattern Anal. Mach. Intell* **2022**, 1–16. [[CrossRef](#)]
39. 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.

40. Berthelot, D.; Carlini, N.; Cubuk, E.D.; Kurakin, A.; Sohn, K.; Zhang, H.; Raffel, C. ReMixMatch: Semi-Supervised Learning with Distribution Matching and Augmentation Anchoring. In Proceedings of the International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia, 26–30 April 2020.
41. DeVries, T.; Taylor, G.W. Improved regularization of convolutional neural networks with cutout. *arXiv* **2017**, arXiv:1708.04552.
42. Yang, B.; Fu, X.; Sidiropoulos, N.D.; Hong, M. Towards k-means-friendly spaces: Simultaneous deep learning and clustering. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 3861–3870.
43. Tian, K.; Zhou, S.; Guan, J. Deepcluster: A general clustering framework based on deep learning. In Proceedings of the Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, 18–22 September 2017; Proceedings, Part II 17. Springer: Berlin/Heidelberg, Germany, 2017; pp. 809–825.
44. 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, PMLR, Baltimore, MD, USA, 13–18 July 2020; pp. 1597–1607.
45. Codella, N.; Rotemberg, V.; Tschandl, P.; Celebi, M.E.; Dusza, S.; Gutman, D.; Helba, B.; Kalloo, A.; Liopyris, K.; Marchetti, M. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv* **2019**, arXiv:1902.03368.
46. Aresta, G.; Araújo, T.; Kwok, S.; Chennamsetty, S.S.; Safwan, M.; Alex, V.; Marami, B.; Prastawa, M.; Chan, M.; Donovan, M. Bach: Grand challenge on breast cancer histology images. *Med. Image Anal.* **2019**, *56*, 122–139. [[CrossRef](#)]
47. Li, X.; Yu, L.; Chen, H.; Fu, C.-W.; Heng, P.-A. Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model. In Proceedings of the British Machine Vision Conference 2018 (BMVC 2018), Newcastle upon Tyne, UK, 3–6 September 2018.

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.