2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

SISE-PC: Semi-supervised Image Subsampling for Explainable Pathology Classification

Sohini Roychowdhury
Director, Curriculum,
Fourthbrain.ai, roych@uw.edu

Kwok Sun Tang
University of Illinois
kwoksun2@illinois.edu

Mohith Ashok
AggDirect
mohithashok@gmail.com

Anoop Sanka
Fourthbrain.ai
anoopsanka@gmail.com

***Abstract*—Although automated pathology classification using deep learning (DL) has proved to be predictively efficient, DL methods are found to be data and compute cost intensive. In this work, we aim to reduce DL training costs by pre-training a ResNet feature extractor using SimCLR contrastive loss for latent encoding of OCT images. We propose a novel active learning framework that identifies a minimal sub-sampled dataset containing the most uncertain OCT image samples using label propagation on the SimCLR latent encodings. The pre-trained ResNet model is then fine-tuned with the labelled minimal sub-sampled data and the underlying pathological sites are visually explained. Our framework identifies upto 2% of OCT images to be most uncertain that need prioritized specialist attention and that can fine-tune a ResNet model to achieve upto 97% classification accuracy. The proposed method can be extended to other medical images to minimize prediction costs.**

***Index Terms*—SimCLR, contrastive loss, ResNet, semi-supervised, label spreading**

I. Introduction

Automated pathology classification has shown to significantly improve patient prioritization and resourcefulness of treatment procedures and patient care [1]. Although deep learning algorithms such as ResNet and InceptionV3 have been established as state-of-the-art [2] for several pathology classification tasks, training these models from scratch can be expensive from the labelled data acquisition and compute resource perspectives. In this work, we present a semi-supervised image sub-sampling method that identifies a minimal sub-sampled data set that represents the most sample uncertainty in a latent feature space that is encoded using a self-supervised contrastive model [3]. We demonstrate the classification performance of a pre-trained ResNet model that is fine tuned using only the minimal sub-sampled data for multi-class pathology classification.

Optical Coherence Tomography (OCT) is a commonly performed diagnostic test designed to assist doctors in identifying retinal diseases, such as choroidal neovascularization (CNV), Diabetic macular edema (DME), and Drusen, that are the most common pathologies resulting in acquired blindness [1]. With approximately 30 million invasive OCT scans being performed each year worldwide, there is a need to identify patients with OCT images that need prioritized attention. The existing work in [2] shows 92-99% classification accuracy for OCT image classification using large annotated OCT image data sets to train a CapsuleNet model. Contrarily, in this work we propose a novel active learning framework [4] to significantly reduce the training complexity by reducing the size of annotated training data set. The proposed system and steps are explained in the Fig. 1.

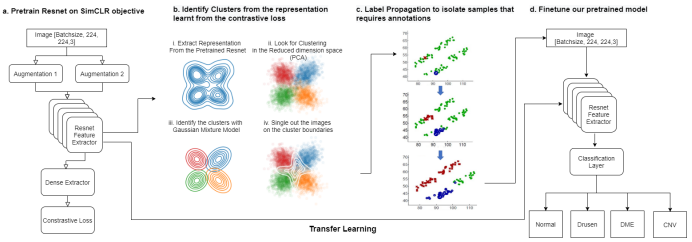


Fig. 1. Overview of the proposed workflow. a) A unsupervised model SimCLR is trained on the constrastive learning objective to learn useful representation of each image. b) Dimension reduction is applied followed by unsupervised clustering and identification of cluster peripheral samples. c) Label Propagation is applied using a pre-labelled annotated image set. Images with the most uncertain labels are selected as the optimal training sample subset. d) The ResNet pretrained with SimCLR is now fine-tuned with the annotated images selected in c) for classification.

This paper makes three key contributions. First, we present a novel semi-supervised framework to identify the most uncertain set of images that are not straightforward to classify or annotate. The proposed method can identify upto 2% of training images samples that are encoded using a self-supervised method [3] and that are found to lie along classification decision boundaries. The proposed active learning algorithm enables prioritization of patients with such uncertain images to be seen early for treatment. Second, we present a two-step classifer training method where feature extractor layers are first pre-trained using image augmentations without labels and SimCLR [3] framework. Next, the classifier is fine-tuning using the labelled uncertain image set for training. This process achieves 96.45% classification test accuracy for OCT-based pathologies by training on atmost 2000 images only. Third, we utilize the Gradcam library to analyze the regions of interest (ROIs) [1] that explain underlying pathology. We apply random image occlusion followed by feature analysis to identify primary and secondary ROIs for pathology to lie between the external limiting membrane and choroidal layers.

II. Data and Methods

The data set and methods used in this work are described below.

A. Data: OCT Image Dataset

In this work we apply the OCT data set in [1], where, training data set contains images annotated for 4 classes {CNV, DME, Drusen, Normal} with {26318, 8616, 11350, 37205},

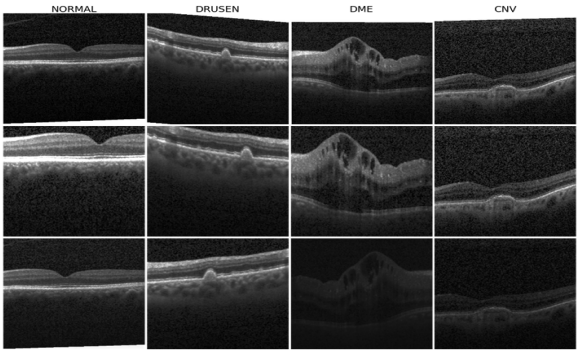


Fig. 2. Row 1 shows the original image with different label class. Rows 2, 3 show the same images in Row 1 with SimCLR default augmentations.

images per class, respectively. The test set has 968 images with the same 4 class labels.

B. Methods and Mathematical Frameworks

The first step to medical image classification is feature extraction while preserving the structural, contextual and textural features. The methods used to encode the OCT images to a latent vector space, and the proposed semi-supervised image sampling methods are described below [1].

1) Image Encoding by Self-supervised Learning: SimCLR [3] is a recent model proposed for unsupervised visual representation learning. A minibatch of n images is sampled and applied with pairs of augmentation to produce $2n$ images. These augmented pairs are fed into a feature encoder (ResNet model) followed by a projection layer (Multilinear Perceptron) to obtain an latent vector $\mathbf{z}$. The training objective is to maximize the agreement of the latent vectors from the same image with augmented views (positive pair), while repulsing from other different images ($2n-1$ negatives). In SimCLR, the similarity between two views is defined using cosine similarity, i.e. $\mathrm{sim}(\mathbf{u},\mathbf{v}) = \mathbf{u}^{\mathrm{T}}\mathbf{v}/||\mathbf{u}||\ ||\mathbf{v}||$. Thus, the loss function for each positive pair of sample (a, b) is defined to be

$$\mathcal{L}_{a,b} = -\log \frac{\exp(\mathrm{sim}(\mathbf{z_a}, \mathbf{z_b})/\tau)}{\sum_{\substack{c=1 \\ c\neq a}}^{2\mathrm{n}} \exp(\mathrm{sim}(\mathbf{z_a}, \mathbf{z_c})/\tau)}, \tag{1}$$

where τ refers to the temperature parameter.

The SimCLR model is trained with batch size of 128 for 120 epochs. A batch size of 128 provides 255 negative samples per positive pair from two augmented views. We apply the LAMB optimizer since training with a large batch size may lead to instability using SGD with momentum. Similar to the original SimCLR implementation in [3], we use a linear warmup for the first 10 epochs and decay the learning rate with the cosine decay schedule. Here, the default augmentation strategy from [3] is adopted. Sample augmentations are shown in Figure 2. The model training is stopped at epoch 75 with a contrastive accuracy of 99.77%.

Each OCT image I is resized to [224x224] followed by SimCLR self-supervised latent vector encoding process. The output of the ResNet encoder for each image of size [7x7x512] is subjected to global averaging to extract J, which is a [1x512] dimensional encoded vector representation for each image. We further reduce the dimensions of the latent vector representation by subjecting J to PCA-based decomposition, thereby retaining 8 top principal components per image that are found to be most representative. Thus, the data set used for subsequent selective sub-sampling is represented by $\{X \in \mathcal{R}^{[n\times d]}, Y \in \mathcal{R}^{[n\times 1]}\}$, where X depicts sample features with $d = 8$ dimensions for $n = 83,484$ samples, and Y represents the pathological class labels. Further, cluster-specific sub-sampling methods described below.

2) Semi-supervised Sub-sampling: With our intention to train a classification model with minimal training labels, we use k-means to cluster X into four categories/clusters with means and standard deviations, respectively, as $[\mu_k, \Sigma_k]\forall k = [1:4]$. Next, we identify uncertain samples as the ones that lie farther away from the cluster centers and are therefore more likely to lie along the classifier decision boundaries. In (2), we compute the probability for a sample x to belong to cluster k.

$$p_k(x) = \frac{exp(-\frac{1}{2}(x-\mu_k)\Sigma_k^{-1}(x-\mu_k)^T)}{\sqrt{(2\pi)^d|\Sigma|}} \tag{2}$$

Samples that have normalized probabilities of belonging to any cluster distribution in the uncertain range of [0.4, 0.6] can be considered to lie at cluster peripheries. These samples are collected as $(X_S \subset X), X_S \in \mathcal{R}^{[mxd]}$, where $m < n$. Other methods such as relative z-scores and Silhouette Scores per sample were also evaluated to identify uncertain samples, but the Gaussian distribution-based sample isolation proved most effective in identifying the corner cases per cluster. Next, we use a small set of labelled data ($L = \{X_l, Y_l\}$) with atmost $n_l = 80$ labelled samples as seed to apply label propagation [5] to all unlabelled samples in X_S. The goal here is identification of samples that demonstrate high variances for transductive labelling. Thus, uncertain samples that acquire different class labels at different sample runs can be indicative of the uncertain space around decision boundaries. Using the semi-supervised label propagation method shown in Algorithm 1, we identify the most uncertain minimum sub-sampled data set $(X_T \subset X_S), X_T \in \mathcal{R}^{[r\times d]}$, $r << n$, that needs to be labelled to adequately train a deep learning classifier [6].

In Algorithm 1, label-spreading using radial basis function (rbf)-kernel is applied with $\alpha = 0.01$ implying that once a label is acquired by a sample in X_S it is less likely to be modified again. Thus, we identify the samples that acquire highly variable labels across 10 label spread runs. The minimal sub-sampled data set X_T is returned for labelling and classifier training thereafter.

3) Pathology Classification and Explainability: Once the labels for the minimal sub-sampled data set are collected, we use this minimal training data to further fine-tune the SimCLR pre-trained ResNet classifier. Next we apply the Gradcam library and tf_explain libraries in Python [7] to identify primary and secondary features corresponding to ROIs that explain the underlying pathology as shown in Fig. 3. Here, we observe that the feature extraction layers of ResNet

[1]Code is available at https://github.com/anoopsanka/retinal_oct

Algorithm 1: Semi-supervised Sub-sampling

Output: Most uncertain sample set X_T
Input: X_S, $L = \{X_l, Y_l\}$
for $e = 0$ **to** 10 **do**
  Randomly select 5 samples per class from L;
  for $x_s \in X_S$ **do**
    Label-spreading with $\alpha = 0.01$, kernel= rbf;
  end
  return *labels* Y_S *for* X_S;
  U[:,e] $\longleftarrow$ Y_S;
end
Given: $U \in \mathcal{R}^{[m \times 10]}$, $X_S \in \mathcal{R}^{[m \times d]}$, $X_T = [.]$;
for $j = 0$ **to** m **do**
  l=unique(U[j,:]);
  if *length*$(l) >= 5$ **then**
    append $X_S(j,:)$ **to** X_T;
  end
end

focus on the regions between the inner segment layer and retinal pigment epithelium layer in the OCT images to classify pathological vs. normal images. In case this ROI is occluded, the secondary regions that are analyzed for each OCT image include the choroidal regions for macular OCT images.

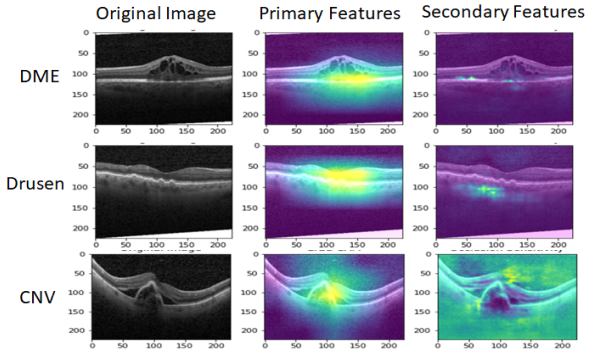


Fig. 3. Examples of primary and secondary ROIs identified on test data.

III. Experiments and Results

To enable minimal data sub-sampling followed by pathology classification, we perform three major experiments. First, we analyze the repeatability of the proposed semi-supervised sub-sampling method by analyzing across multiple runs and assessing the numbers of images detected per class. Second, we compare the classification performance of the proposed sub-sampling with random stratified sampling for ResNet classification. Third, we analyze the robustness of features explained by the sub-sampled uncertain images.

A. Analysis of Sub-sampling Repeatability

We repeat Algorithm 1 for 20 runs and record the numbers of images corresponding to each class category detected per run. The average and standard deviations in the number of samples per class are shown in Table I.

TABLE I
Analysis of sub-sampled images across 20 runs.

Metric	Normal	Drusen	DME	CNV	Total
Average	711	274	261	466	1707
Std. dev.	78	50	26	141	251

Here, we observe significant down sampling and relative consistency across the number of images sub-sampled across multiple runs when compared to the original sample size of 83,484 samples. Thus, the proposed method selects less than 2000 training samples, that represents about 2% of the total sample size, to be annotated for subsequent classification.

B. Classification of sub-sampled Images

To analyze the importance of the novel semi-supervised sub-sampling method proposed in this work, we further fine-tune the ResNet model that was previously trained using the contrastive loss for SimCLR. For this fine-tuning, we use several sets of sub-sampled labelled data to asses the impact of transfer learning and also to identify the minimum number of samples required to train an explainable pathology classifier.

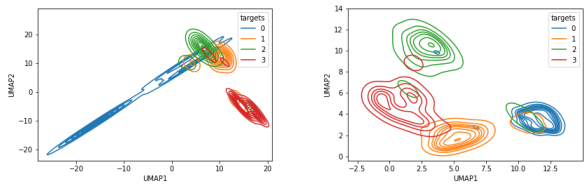
For this experiment, a set of labelled data (L) with n_l samples is subjected to stratified random 80/20 train/validation split. This data set is then used to train the ResNet model layers with categorical crossentropy loss, balanced class weights and Adam Optimizer with a learning rate 10^{-5} . Early stopping is applied when the validation loss has not improved for more than 3 epochs. To analyze the importance of data sub-sampling over using the complete training data set, we apply stratified random sampling to isolate 1, 5, 10% of samples per class to create labelled data L that is then used to fine-tune the ResNet model. The classification performances using stratified random sub-sampling in terms of average classification accuracy, precision, recall and F1 score [1] are shown in Table II. Here, we

TABLE II
Averaged classification performances from Stratified and Proposed Semi-supervised Sampling across several runs.

Stratified Sample Size (%)	Accuracy	Precision	Recall	F1
1%	0.9101	0.9221	0.9101	0.9083
5%	0.9834	0.9844	0.9834	0.9834
10%	0.9880	0.9883	0.9880	0.9880
100%	0.9989	0.9989	0.9989	0.9989
Proposed Semi-supervised Sampling				
2%	**0.9645**	**0.9675**	**0.9625**	**0.9625**

observe that training by atmost 5% random stratified samples, the ResNet model is significantly well trained to visually explain pathology classification. Improvements in classification performances thereafter are incremental and often not generalizable. Also, in Fig. 4 we observe the representations of classified samples with 1% and 98% of the data samples. We observe significant cluster separability as the training data set increases.

Our goal is to identify the minimal training data set to achieve significant cluster separability. Thus, we analyze the performance of the proposed sub-sampling method to train the classifier in Table II. We observe that the proposed sub-sampled data set that isolates 2% training images around

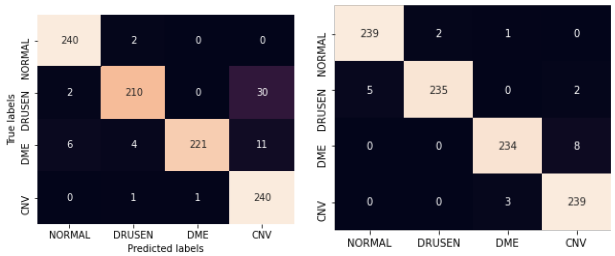


(a) UMAP with 1% data (b) UMAP with 98% data

Fig. 4. Uniform manifold approximation and projection (UMAP) representations of classified samples trained with L corresponding to 1% and 98% stratified random samples, respectively. Separability across clusters is significantly higher with 98% of training samples.

decision boundaries achieves about 96.45% classification accuracy. This performance is comparable to InceptionV3 based model in [2] that achieved 96.1% accuracy with complete training data set. Thus, the proposed sub-sampling method intuitively lies within the 5% sampling size as learned from the random stratified sampling. Additionally, the ResNet model fine-tuned on the proposed minimal sub-sampled data achieves consistent precision, and recall performances, which indicates consistencies in the numbers of classified false positives and false negatives.

The confusion matrix for ResNet classifier after being trained with 1% stratified random samples and with the 2% proposed sub-sampling method are shown in Fig. 5.



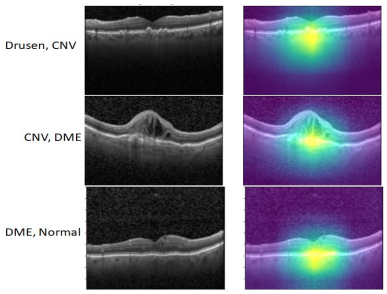
(a) 1% Stratified Random sampling. (b) 2% Proposed Sub-sampling.

Fig. 5. Confusion Matrix for Classification with varying training samples.

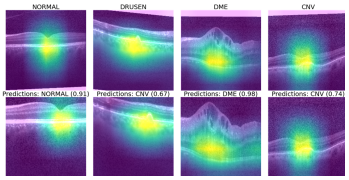
Here, we observe that using the proposed sub-sampled data, some DME images get mis-classified as CNV where small cystic regions develop close to the inner sub-retinal layers. Also some images with thin drusen membranes are mistaken for normal images and some images with CNV and poor contrast are classified as DME. Most of these classifications can be improved by pre-processed zooming in and contrast corrections applied on test images in future works.

C. Explainability of Pathology

Finally, we incorporate the Gradcam library to analyze the ROI that contributes to pathology classification. In Fig. 6 (a), we observe some sub-sampled images using Algorithm 1 that lie along the border of two distinct clusters, thereby enhancing classification performance for samples of those clusters. Further, Fig. 6 (b) shows the spatial robustness of the fine-tuned ResNet for pathology explainability.



(a) Examples of proposed sub-sampled images along cluster peripheries.



(b) ROI explanations for SimCLR augmentations shows the same ROI is highlighted per augmentation.

Fig. 6. Examples of pathology explanation on (a) sub-sampled training images, (b) test images, respectively.

IV. Conclusion

In this work we present a novel semi-supervised algorithm that uses encoded latent features per image to identify a minimal sub-sampled data set that can be useful to train and visually explain pathology. We incorporate a ResNet model for pathology classification that is first trained using contrastive loss to distinguish an original and augmented sample from other images. Next, we perform label spreading on the encoded latent space to identify a training set of upto 2% of samples that represent the most uncertainty in feature space. Once trained on this labelled minimal sab-sampled data, the final classifier can achieve upto 97% classification accuracy and is capable of visually explaining pathological sites. Future work can be directed towards extending the proposed image sub-sampling method and ResNet training to other pathologies.

References

[1] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[2] T. Tsuji, Y. Hirose, K. Fujimori, T. Hirose, A. Oyama, Y. Saikawa, T. Mimura, K. Shiraishi, T. Kobayashi, A. Mizota *et al.*, "Classification of optical coherence tomography images using a capsule network," *BMC ophthalmology*, vol. 20, no. 1, pp. 1–9, 2020.

[3] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020.

[4] S. Dasgupta and D. Hsu, "Hierarchical sampling for active learning," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 208–215.

[5] K. Sun, Z. Min, and J. Wang, "Pp-pll: Probability propagation for partial label learning," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2019, pp. 123–137.

[6] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[7] R. Meudec. (Accessed, Jan, 2021) tf_explain. [Online]. Available: https://github.com/sicara/tf-explain