Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 20 (2): 2439–2458.
DOI: 10.3934/mbe.2023114
Received: 06 June 2022
Revised: 25 October 2022
Accepted: 07 November 2022
Published: 21 November 2022

Research article

Predicting effectiveness of anti-VEGF injection through self-supervised learning in OCT images

Dehua Feng[1], Xi Chen[1,*], Xiaoyu Wang[1], Xuanqin Mou[1], Ling Bai[2], Shu Zhang[3] and Zhiguo Zhou[4]

[1] School of Information and Communications Engineering, Xi'an Jiaotong University, Shaanxi 710049, China
[2] Department of Ophthalmology, the Second Affiliated Hospital of Xi'an Jiaotong University, Shaanxi 710004, China
[3] Department of Geriatric Surgery, the Second Affiliated Hospital of Xi'an Jiaotong University, Shaanxi 710004, China
[4] Department of Biostatistics and Data Science, University of Kansas Medical Center, KS 66160, USA

* **Correspondence:** Email: xi_chen@mail.xjtu.edu.cn.

Abstract: Anti-vascular endothelial growth factor (Anti-VEGF) therapy has become a standard way for choroidal neovascularization (CNV) and cystoid macular edema (CME) treatment. However, anti-VEGF injection is a long-term therapy with expensive cost and may be not effective for some patients. Therefore, predicting the effectiveness of anti-VEGF injection before the therapy is necessary. In this study, a new optical coherence tomography (OCT) images based self-supervised learning (OCT-SSL) model for predicting the effectiveness of anti-VEGF injection is developed. In OCT-SSL, we pre-train a deep encoder-decoder network through self-supervised learning to learn the general features using a public OCT image dataset. Then, model fine-tuning is performed on our own OCT dataset to learn the discriminative features to predict the effectiveness of anti-VEGF. Finally, classifier trained by the features from fine-tuned encoder as a feature extractor is built to predict the response. Experimental results on our private OCT dataset demonstrated that the proposed OCT-SSL can achieve an average accuracy, area under the curve (AUC), sensitivity and specificity of 0.93, 0.98, 0.94 and 0.91, respectively. Meanwhile, it is found that not only the lesion region but also the normal region in OCT image is related to the effectiveness of anti-VEGF.

Keywords: anti-VEGF; effectiveness prediction; self-supervised learning; optical coherence tomography; choroidal neovascularization; cystoid macular edema

1. Introduction

The World Health Organization's survey reported that 8.7% of the worldwide population has age-related macular degeneration (AMD) in 2014, and the projected number of people with AMD will increase to 288 million in 2040 [1]. As the most common cause of loss vision for elderly people, AMD starts from the dry kind and always moves to the wet kind, i.e., choroidal neovascularization (CNV), then to the retinal layer [2,3]. And it is found that cystoid macular edema (CME), a main cause of loss vision related to vascular disease, is a common complication in patients with CNV associated with AMD [4,5]. Although the etiology of CME is not completely understood, pre-existing condition such as diabetes mellitus and uveitis as well as intraoperative complications can cause the risk of CME postoperatively [6].

As anti-vascular endothelial growth factor (anti-VEGF) therapy can slow down or stop damage from the abnormal blood vessels and even improve vision, it has become a standard way for CNV and CME treatment [7,8]. However, there are some issues on anti-VEGF therapy [9]. For example, anti-VEGF therapy is too expensive to be affordable for some patients, particularly in developing countries [10]. Meanwhile, it is a long-term therapy with three injections rounds and these injections are conducted at four-week interval [11]. Another challenge is that the anti-VEGF injection may not always be effective for patients [12]. Optical coherence tomography (OCT) imaging has been widely for ophthalmologists to analyze and diagnose retinal pathologies [13]. As such, it is necessary to develop an OCT based model for predicting anti-VEGF's effectiveness before injection.

Predicting effectiveness of anti-VEGF is essentially a treatment outcome prediction problem. According to the recent literature study, the strategies for treatment outcome prediction mainly consists of two categories: hand-crafted feature-based models and deep learning feature-based models. Hand-crafted features such as intensity features, texture features, and wavelet features have been widely used in treatment outcome prediction [14,15]. However, designing hand-crafted features requires domain knowledge and these features are low-level which may lead to inferior performance. Deep learning can obtain high-level features for specific task in an end-to-end way with superior performance in treatment outcome prediction tasks, such as survival time assessment [16–18], metastasis prediction [19,20], and treatment response prediction [21–23]. However, it is always difficult to collect amounts of images for the particular disease to feed the deep learning models. To handle this issue, transfer learning has been widely used and achieves great success [16, 21, 24–26]. For example, Paul et al. [16] predicted non-small cell adenocarcinoma lung cancer patients' short- or long-term survival time using CT scans' deep features as well as traditional features. Cha et al. [21] developed a predictive model to distinguish whether the patients of bladder cancer are in complete chemotherapy response via deep learning features through pre- and post-treatment CT images. However, these transfer learning-based pre-training are implemented based on large-scale natural image databases such as ImageNet. Our group also utilized transfer learning on ImageNet to predict effectiveness of anti-VEGF via OCT images in our previous study [27]. Although the results in [27] are acceptable, the large differences of imaging principle between natural image and OCT image may limit the performance of the fine-tuned model.

Since self-supervised learning (SSL) can learn features from raw images without explicit labels, it is an alternative solution to overcome the problem caused by small scale dataset and has attracted much attention [28–33]. SSL is implemented by designing a pretext task so that more semantic and intrinsic feature representation can be obtained through human-designed labels. As a novel unsupervised feature representation learning strategy, SSL has become a popular area in computer vision [29,30,34–40] and medical image analysis [31–33,41,42]. Several pretext tasks based on color transformation [34,35], geometric transformation [35], context [30,36,37], cross-modal [38], instance discrimination by contrastive learning [39,40] have been developed. In medical image analysis, the domain-specific pretext tasks have been developed to learn the visual feature representation. For example, changing multiple pairs of patches' positions and then recovering the deformed into original one are developed by Chen et al. [31]. This pretext task obtains promising results because the medical images for specific tissue or organ have similar and regular local structure information. This idea is also used in 3D self-supervised learning in [32] by creating three sub-tasks, they are cube ordering, cube orientation, and mask identification to learn structure and texture features. In another study [28], a unified SSL framework based on the task of recovering deformed volumes is built through multiple transformations to obtain multiple perspectives of medical images such as appearance, texture, context, etc.

To reduce this gap between source and target domain dataset, we proposed to obtain the pre-trained model on OCT image dataset using SSL based on unlabeled images. In this study, a public OCT dataset [43] (termed as UCSD dataset) is used for pre-training. Since the target in our study is different from UCSD dataset, labels of UCSD dataset are ignored and we only use the images. Inspired by the 3D medical pre-trained model on CT and MRI in [33], we employ the transformation-based pretext task to learn the general feature representation. And then the model is fine-tuned through our own OCT dataset with anti-VEGF effectiveness labels. Finally, classifier is trained through the features extracted from fine-tuned model to predict effectiveness of anti-VEGF. Experimental results demonstrate that the new OCT based SSL (OCT-SSL) model can obtain promising performance and the comparative studies show that OCT-SSL outperforms other available models.

2. Materials

Two datasets are used in this study, they are XJTU OCT dataset from our institution, and UCSD public dataset [43]. The details are described in the following subsections.

2.1. XJTU OCT dataset

The cohort consists of 228 patients collected from October 2017 to October 2019 at the Second Affiliated Hospital of Xi'an Jiaotong University (Xi'an, China). All 228 patients fulfilled the following criteria: (a) diagnosis of CNV, CME, or both; (b) availability of OCT images acquired by Heidelberg Retina Tomograph-IV (Heidelberg Engineering, Heidelberg, Germany) before anti-VEGF injection; (c) treatment of anti-VEGF injection and the effectiveness assessment performed on the 21st day the first anti-VEGF injection, which indicates the treatment's effectiveness. Anti-VEGF is effective for 171 patients, and the remaining 57 patients have no sign of recovery. Written informed consent was provided by each patient and approval of the study was obtained from the research ethics committee. The OCT image size is 766 × 596. Full retinal images were used in our proposed method for fine-tuning and final classification. Only one 2D image is used for one patient, which is the B-scan with the maximum lesion CME or CNV, whilst it was selected by an ophthalmologist with 15 years' experience and reviewed by an ophthalmologist with 25 years' experience. And the lesion regions were also contoured by these two ophthalmologists as shown in Figure 1. Therefore, we can investigate how the lesion region influence the prediction model. This investigation is discussed in Figure 7.

In this study, 20% cases (34 effective cases and 11 ineffective cases) were randomly selected as testing samples and the remaining cases were training samples. The cases in testing set never involved the training process. In fine-tuning stage, due to the limited scale of XJTU dataset, data augmentation was performed in the training images. Six data augmentation manners were utilized including horizontal flipping, vertical flipping, translation, rotation, Gaussian blurring and contrast changing. More detailed procedure for data augmentation is illustrated in our previous work [27].

2.2. UCSD OCT dataset

The UCSD OCT dataset was built by Kermany et al. [43] at the University of California San Diego. It is used for diagnosis of three kinds of retinal disorders, CNV, diabetic macular edema (DME) and drusen. The contributors also incorporate normal cases into the database. Examples of four classes are illustrated in Figure 2. The 2D image was horizontal foveal cut of OCT B-scans. Totally there are 26,315 normal images, 37,205 CNV images, 11,328 DME images and 8616 drusen images in training dataset, respectively. In testing dataset, each class containing 250 images are used to evaluate the model. The images were collected from the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, and Beijing Tongren Eye Center between July 2013 and March 2017. The scanning machine was from Heidelberg Engineering, Germany. In this study, we randomly sampled 250 images of each class from the original training dataset as training images and randomly sample 10 images of each class from the original testing dataset as the validation images in pre-training stage. The dataset in pre-training stage refers as UCSD_P dataset in this paper.

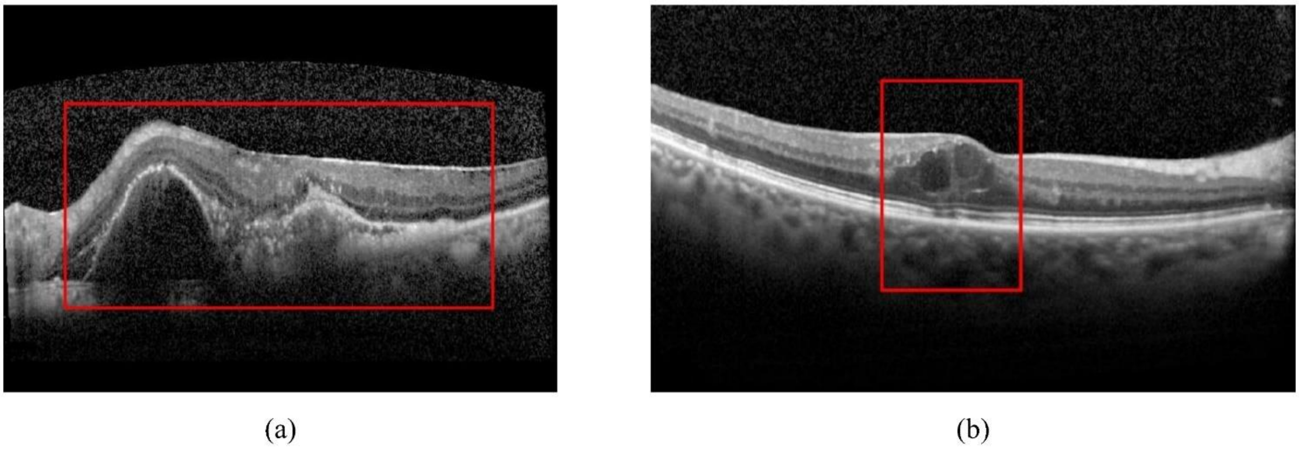

Figure 1. Full retinal images of two cases with contoured lesion region (red rectangular): (a) CNV and CME case, (b) CME case.

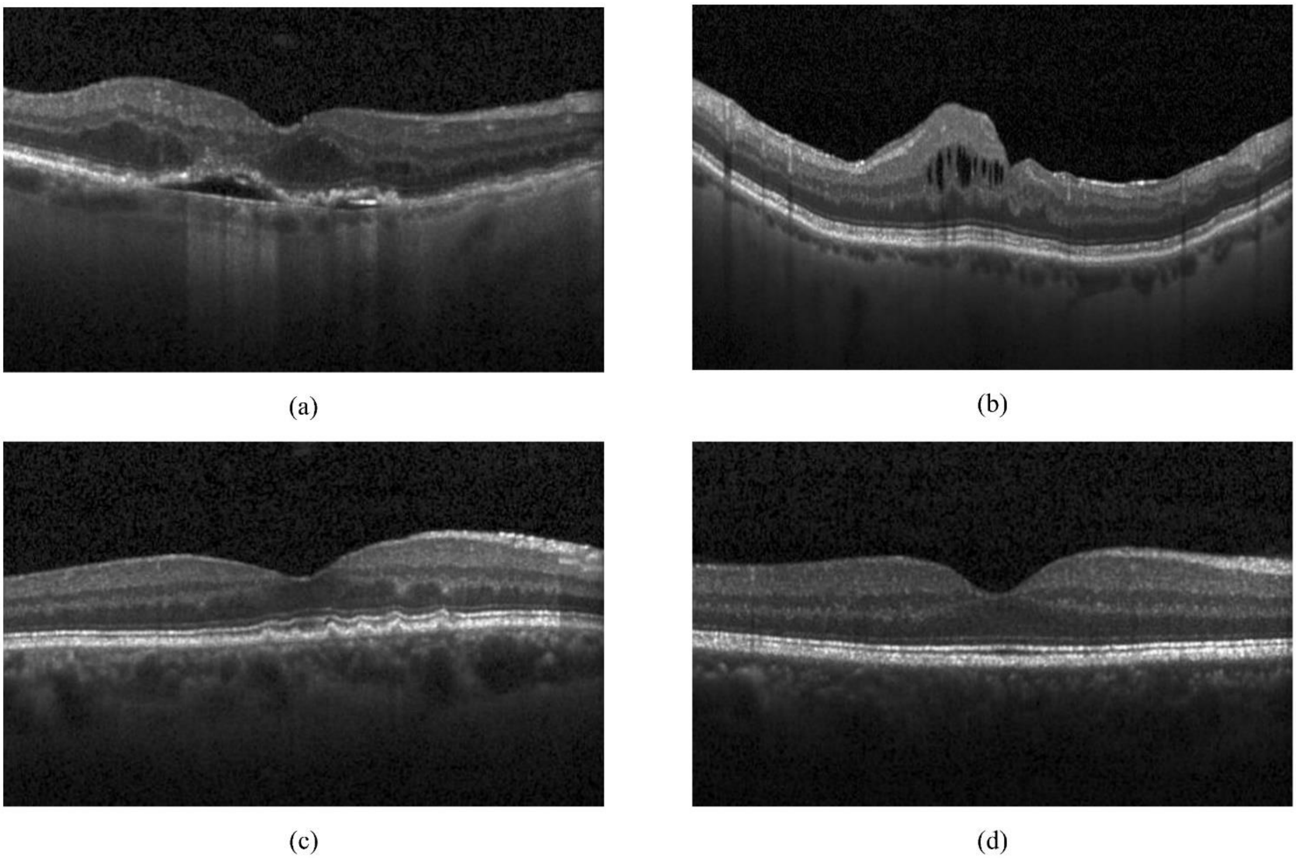

Figure 2. Four examples of CNV, DME, drusen and normal case from UCSD OCT dataset: (a) CNV case, (b) DME case, (c) Drusen case, (d) Normal case.

3. Methods

The OCT-SSL framework is shown in Figure 3, which consists of three steps: self-supervised pre-training, fine-tuning, and classification. First, UCSD_P dataset without label was used to build a U-Net to capture general features of OCT image. In this step, unlabeled OCT images are processed by combining multiple transformations to provide the deformed images. The second step is to learn task-specific feature representation through XJTU OCT dataset with label information. Finally, a classier fed with the features extracted by fine-tuned encoder in U-Net is built to predict the treatment response.

3.1. SSL pre-training

SSL pre-training includes two stages: image distortion and image restoration. In our study, the aim of pre-training based on SSL is to learn common OCT features that are transferable and generalizable. In image distortion stage, given original images, four image transformations proposed in [33] are consolidated to make model learn more robust feature representation. In image restoration stage, encoder-decoder is employed to reconstruct the image distortion stage's outputs as the original input images.

For each image I in UCSD_P dataset during image distortion stage, we apply image transformation:

$$\tilde{I} = f(I), \tag{1}$$

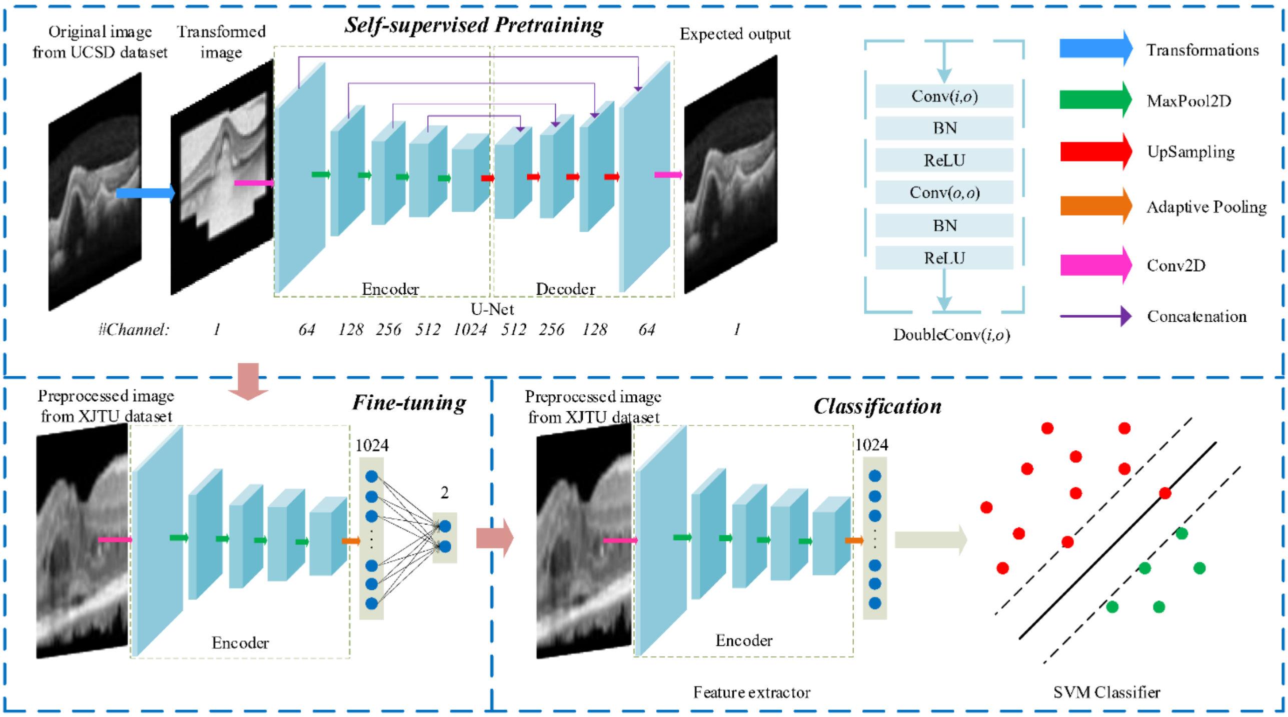

Figure 3. Pipeline of OCT-SSL. U-Net is to restore the transformed input OCT image from UCSD_P with the corresponding original image as label. The encoder of U-Net is the pre-trained model. Next, pre-processing and data augmentation are employed on training images of XJTU dataset. These data fine-tune the pre-trained model by label supervised learning. In classification part, the fine-tuned model without fully connected layer plays the role of feature extractor, and outputs features of XJTU OCT images only preprocessed without data augmentation. The 1,024 dimensions features are sent to SVM classifier to obtain the effectiveness of anti-VEGF. DoubleConv(i,o) and Conv(i,o) represent the DoubleConv module and Conv module with input of i channels and output of o channels.

where $f(\cdot)$ denotes transformation functions. Then in image restoration stage an encoder-decoder network is learned to approximate the function $g(\cdot)$ which maps the transformed image $\tilde{I}$ back to the original ones I:

$$g(\tilde{I}) = I = f^{-1}(\tilde{I}). \tag{2}$$

Four transformations are employed in this study, they are non-linear transformation, local pixel shuffling, image in-painting, and image out-painting. The operation procedures of the transformations are the same as those in [33]. For the integrity and readability of this paper, we also describe the four transformations and give the OCT illustrations.

(1) *Non-linear transformation*. Intensity value and its distribution represent appearance of organs and tissues. Restoring the image distorted with non-linear transformation can make the model learn organ appearance. In our study, we employ Bezier Curve [44] as the non-linear function, which is smooth, monotonous to ensure a one-to-one mapping. An example is illustrated in Figure 4(b).

(2) *Local pixel shuffling*. Texture information is of great importance to represent image. In OCT image, the local visual structure information such as retinal layer shows the severity of abnormality.

Local pixel shuffling transformation aims to learn local structure and texture feature in randomly selected patches. In an original patch, local pixel shuffling means shuffling the order of contained pixels resulting in a transformed patch. The patch size and position are determined randomly. In our study, the patch height and width are random integer values between 1 and 22. The patch position is determined by the left upper coordinate x and y which are random integer values between 0 and 202. The number of selected patches is 200 for each OCT image. To restore image from local pixel shuffling, the pre-trained model can learn local boundaries and texture. An example is shown in Figure 4(c).

(3) *Image in-painting.* Image in-painting is to paste constant gray-level windows randomly on the OCT image with random size, which means retaining the original intensities outside the window and replacing the intensity values of the inner pixels with a constant value. There are 5 windows for occlusion and the value of inner pixel is set to 0 in our study. The height and width are set as random integers between 28 and 56. The left upper position coordinate x and y are between 3 and 221-height. Compared with the local pixel shuffling, the image in-painting transformation increases the degree of difficulty which needs to assume the replaced region's pixel values according to the neighbor regions' intensity and structure information. Image in-painting requires the pre-trained model get local continuities of organs via interpolating. A typical example is shown in Figure 4(d).

(4) *Image out-painting*. As for image out-painting transformation, pixel values inside these windows are retained and the outside pixels are all replaced by background. In image out-painting, we also used 5 windows and the value of outside pixel is set to 0. The height and width of windows are between 96 and 128. And the left upper coordinate x and y are between 3 and 221-height. This transformation needs to assume the edge information from global perspective, which requires the pre-trained model to learn global geometry and spatial layout of organs via extrapolating. An example is shown in Figure 4(e) as well. Notice that image in-painting and out-painting transformations can't be conducted in one OCT image simultaneously.

In image distortion stage, each OCT image in UCSD_P is processed by the transformations ordered as local pixel shuffling, non-linear transformation, image in-painting or image out-painting, where image in-painting and image out-painting are exclusive. Figure 4(f) shows an example of the image processed by local pixel shuffling, non-linear transformation, image out-painting. After these transformations, OCT images have changes in appearance, texture and structure compared to the original ones. Then in image restoration stage, encoder-decoder network serves as a restoration model to transform the deformed images into original image. Since the decoder requires shape, texture, and structure information to restore the image, the encoder can provide richer semantic representation if the restoration model achieves great performance. In our study, we utilize U-Net as the backbone to accomplish model pre-training [45]. More details of U-Net are shown in Figure 3. During pre-training, U-Net served as the encoder-decoder to restore the deformed OCT image into original one where the encoder can learn the general representation of OCT images, such as intensity, texture, global structure, local structure, etc.

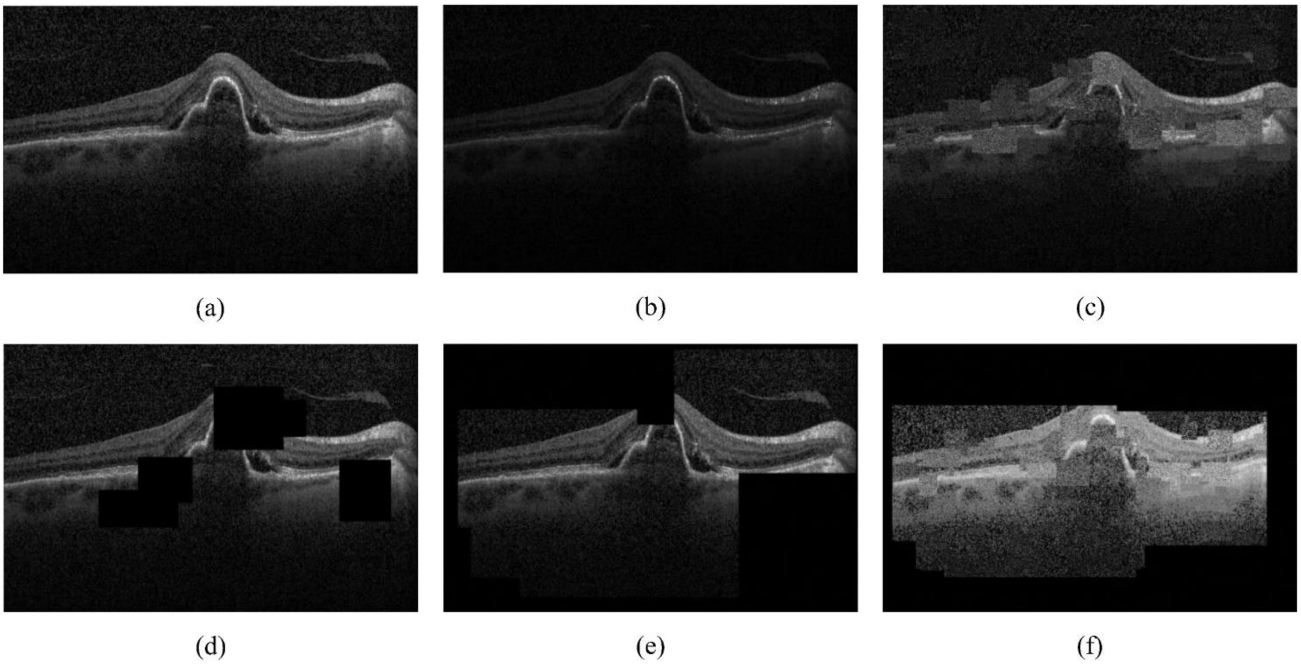

Figure 4. (a) Original OCT image of UCSD_P dataset. (b) Processed by Non-linear transformation. (c) Processed by Local pixel shuffling. (d) Processed by Image-in-painting. (e) Processed by Image-out-painting. (f) Processed by the combination of three transformations (Local pixel shuffling, Non-linear transformation, and Image-out-painting).

3.2. Model fine-tuning

In pre-trained encoder, we can obtain the general features of OCT images which can describe the OCT images comprehensively but are not specifically used to predict anti-VEGF effectiveness. In fine-tuning stage, these general features can be fine-tuned on the target dataset (XJTU dataset), where the anti-VEGF effectiveness is used as supervised label. In this stage, the encoder is used as a classifier. The feature maps of the last DoubleConv module in encoder are taken as the input of adaptive average pooling layer. Then feature maps of pooling layer are cropped by rectified linear unit (*ReLU*) activation mapping. In the next step, the output features are sent to a fully connected layer of size 1024×2, where two neurons output the predicted probabilities of each class. This procedure is based on training images of XJTU dataset with data augmentation.

3.3. Classification

In classification phase, features extracted from XJTU OCT images by fine-tuned encoder are taken into the SVM classifier [46] to predict the treatment response of anti-VEGF. And the predictive model is built based on XJTU OCT dataset (228 cases) without data augmentation. Since the number of effective cases and the number of non-effective cases are imbalanced, sample weighting strategy [47] is used in SVM training to mitigate the imbalance of classes and results in balance between sensitivity and specificity metrics.

4. Experiments

4.1. Set up

To avoid the predictive model being distracted by the background speckle, pre-processing is performed on XJTU dataset. The pixels in the top and bottom region of the retinal layers are set as 0 and pixels in the middle region, i.e., retinal layers are preserved without destroying the foreground information as illustrated in Figure 5. More detailed procedure for pre-processing is illustrated in our previous work [27].

During pre-training, all OCT images are resized to 224 × 224 before image deformation. And the probability thresholds of transformations, i.e., non-linear transformation, local pixel shuffling, image in-painting/image out-painting, are 0.9, 0.5, 0.9, respectively, as in [33]. We set the U-Net input channel and output channel as 1. The loss function for the encoder-decoder network is Minimum Square Error (MSE), and the network restores the image pixel by pixel. Adam algorithm is used as optimizer to update weights with a batch size of 16 and initial learning rate of 0.01. This network is trained through 100 epochs. We adopt the early stopping strategy so that if loss on validation set has not increased for 5 consecutive epochs, the training process will stop to avoid overfitting and improve training efficiency [48].

In fine-tuning, the input images are pre-processed full retinal images from XJTU dataset and cross entropy is used as the loss function. The weights are updated by the Adam optimizer on the mini-batch of 16 images and we empirically set the learning rate as 0.00001. The model is fine-tuned for 100 epochs and we employ testing set to evaluate model's performance by four metrics, accuracy, sensitivity, specificity, and AUC. Early stopping strategy is also used and the metric for early stopping is MSE.

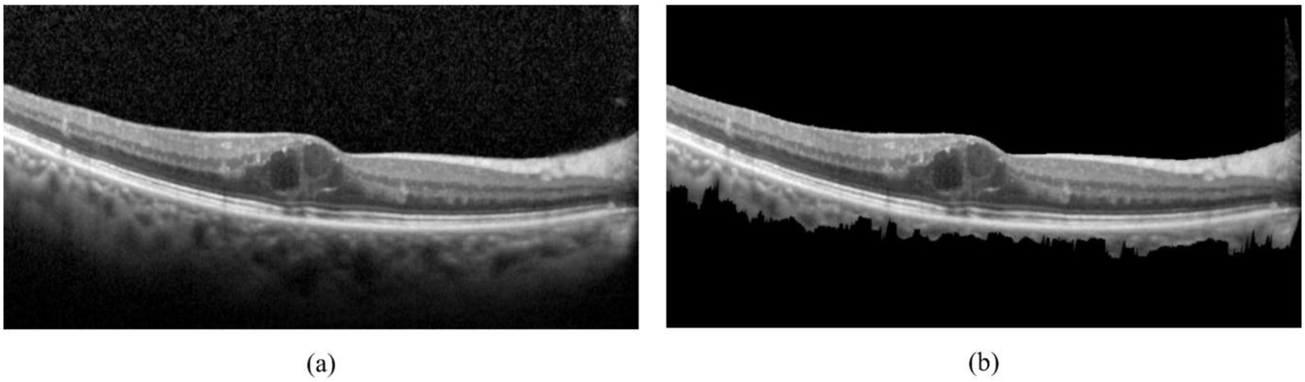

Figure 5. Original OCT image and its pre-processed image: (a) Original OCT image, (b) Pre-processed OCT image.

In classification model construction, we perform five-fold cross validation on training set of XJTU OCT images to select parameters of SVM's radial gaussian function (RBF) [46]. For RBF kernel, the parameter lists are [0.00002,0.0002, 0.002, 0.02, 0.2, 2, 20, 200] for *C* and [0.00002,0.0002, 0.002, 0.02, 0.2, 2, 20] for *gamma*. For each group of parameters, we run 10 times to get the average results. We find that SVM with RBF kernel of *C* = 2 and *gamma* = 0.0002 performs the best.

In addition, we conduct some comparative studies. First, the proposed OCT-SSL is compared with our previous method based on transfer learning on ImageNet. Second, since source domain dataset and pre-training manner are crucial to feature learning, we perform more experiments to investigate the influence of these factor on the results.

Meanwhile, we also investigate the performance of hand-crafted features for anti-VEGF effectiveness prediction. For OCT image-based pathological diagnosis classification tasks [43, 49–53], deep learning features or hand-crafted features can be used. Liu et al. [49] employ local binary pattern (LBP) and a multi-scale spatial pyramid to classify normal cases and three types of macular pathologies. And Srinivasa et al. [50] categorize normal cases, AMD, and DME utilizing multi-scale histogram of gradient (HoG) descriptors. More recently, Bogunović et al. [53] used quantitative spatial-temporal features computed from automated segmentation of retinal layers and fluid-filled regions in OCT images acquired after three initial monthly injections of anti-VEGF to predict further treatment requirements. In our proposed method, full retina images are used without any retinal layers or fluid-filled regions segmentation. To perform a reasonable comparison on the performance of hand-crafted features and deep features, we only use LBP and HoG features calculated from the full retina image. For LBP descriptor, 8 neighbors are deemed as a circle surrounding the central one with radius of 1 resulting in 64-dimesional feature. The HoG descriptor with 262,44 dimensions is computed on the resized OCT image of 224 × 224, and a cell size of 8 × 8 with 9 histogram bins. Principal component analysis (PCA) is used to obtain the top-200 dimensions with the explained variance ratio of 0.97. And the concatenation of LBP and HoG feature vectors is also studied and reduced by PCA. Intensity and texture features are also used. In our experiment, we extract 9 intensity features according to [54]. For the texture features, we extract 96 features based on gray-level co-occurrence matrix (GLCM) as in [54]. When calculating GLCM, we choose the distance or offset between the reference pixel and neighbor ones as 1 and 2 and pixel pairs with four directions including horizontal, vertical, diagonally up and diagonally down are employed. And then we combine the intensity and texture features as Intensity-Texture of 105 dimensions. The concatenation of deep features and LBP features is also investigated. The classifier is SVM with sample weighting to balance the effective cases and non-effective cases.

4.2. OCT-SSL results

In this section, model based on the proposed OCT-SSL is named $UNet_{SSL}^{UCSD}\_F\_SVM$ according to its configuration. The superscript of the name represents the source domain dataset used for pre-training, and the subscript represents the pre-training strategy. The prediction results of $UNet_{SSL}^{UCSD}\_F\_SVM$ are listed in Table 1. This model can achieve accuracy, AUC, sensitivity, and specificity of 0.93, 0.98, 0.94 and 0.91, respectively. Our previous study employed the classical transfer learning method on ImageNet database named $ResNet_{Sup}^{Img}$ where ResNet-50 [55] was the backbone and pre-trained model was also fine-tuned on XJTU OCT dataset, the prediction results achieved 0.72, 0.81, 0.78 and 0.71 for accuracy, AUC, sensitivity and specificity, respectively [27]. The better results demonstrate the superiority of the proposed method in predicting the effectiveness.

4.3. Comparison of different source domain datasets

In conventional transfer learning, the pre-trained models are created based on ImageNet dataset.

In this experiment, we compare the performance of ImageNet pre-trained model ($ResNet_{Sup}^{Img}$) with UCSD dataset pre-trained model ($ResNet_{Sup}^{UCSD}$). The pre-trained stage and fine-tuning procedure are both label-supervised. It can be observed that the results of $ResNet_{Sup}^{Img}$ performs better than $ResNet_{Sup}^{UCSD}$ (see Table 1). Although UCSD dataset has the same imaging modality with XJTU dataset, it gives inferior results, showing obvious imbalance between sensitivity and specificity. The reason may be that label-based pre-training manner cannot learn general feature representation of OCT images through UCSD dataset due to only four classes in UCSD dataset, but there are over 1,000 classes in ImageNet.

Table 1. Results of different source domain, pre-training strategy and network architecture.

Configuration	Accuracy	AUC	Sensitivity	Specificity
$\mathrm{ResNet}_{\mathrm{Sup}}^{\mathrm{Img}}$	0.72	0.81	0.78	0.71
$\mathrm{ResNet}_{\mathrm{Sup}}^{\mathrm{UCSD}}$	0.71	0.74	0.85	0.44
$\mathrm{UNet}_{\mathrm{SSL}}^{\mathrm{UCSD}}$	0.74	0.78	0.83	0.71
$\mathrm{UNet}_{\mathrm{SSL}}^{\mathrm{UCSD}}$_F	0.81	0.89	0.82	0.80
$\mathrm{ResNet}_{\mathrm{SSL}}^{\mathrm{UCSD}}$_F	0.75	0.84	0.76	0.75
$\mathrm{UNet}_{\mathrm{SSL}}^{\mathrm{UCSD}}$_F_SVM	0.93	0.98	0.94	0.91

4.4. Comparisons of using different pre-training strategies

Since the label-based pre-training may lead to the task-specific feature representation, the fine-tuning is hard to get a good performance in other target domain datasets. As such, we changed the pre-training strategy from label-supervised to self-supervised manner. In this study, we built a $UNet_{SSL}^{UCSD}$ model where fully connected layers (*FC3* in Table 2) as the classifier were applied directly on the on the frozen encoder from self-supervised pre-trained model. The results of $UNet_{SSL}^{UCSD}$ (shown in Table 1) are much better than $ResNet_{Sup}^{UCSD}$, especially in specificity, which indicates that the self-supervised strategy can learn more general feature representation and pre-training based on self-supervised strategy is more suitable than pre-training based on label-supervised strategy where source dataset and target dataset have different classification tasks.

4.5. Comparisons using different fine-tune modules

In our proposed method, the fine-tuned model without fully connected layer plays the role of feature extractor and the 1024 dimensions features are sent to SVM classifier to obtain the effectiveness of anti-VEGF. In this section, for fine-tune module, we design three types of fully connected layer *FC1*, *FC2* and *FC3* and use different strategies to freeze the encoder layers. The architectures of the fully connected layers and the number of parameters are listed in Table 2.

Table 2. Architecture and parameter number of three fully connected layers.

Fully connected layer	FC1	FC2	FC3
architecture	(1) Linear (1024, 512) (2) ReLU (3) Dropout (p=0.5) (4) Linear (512, 2)	(1) Linear (1024, 256) (2) ReLU (3) Dropout (p=0.5) (4) Linear (256, 2)	(1) Linear (1024, 2)
number of parameters	525,312	262,656	2048

Given the three fully connected classifiers, we fine-tune the model with different strategies to freeze the encoder layers. The U-Net's encoder has 5 DoubleConv modules and these convolutional modules are followed by max pooling layers. Therefore, we freeze the model's DoubleConv0 to DoubleConv1, DoubleConv0 to DoubleConv2 or DoubleConv0 to DoubleConv3 and we also freeze only DoubleConv0 if necessary. In other words, we always fine-tune parameters of the last DoubleConv module.

Figure 6 illustrates the results of different freezing manners of *FC1*, *FC2* and *FC3*. For *FC1*, we find that the model can achieve the best performance when freezing DoubleConv0 to DoubleConv2. And for *FC2*, the best performance is also achieved when freezing DoubleConv0 to DoubleConv2. However, for *FC3* without ReLU activation mapping, the best performance is achieved by freezing DoubleConv0 to DoubleConv1. This may be because the increase of parameters may overfit on the training set even though *FC1* and *FC2* use ReLU to enhance non-linear property of fully connected layer. *FC3*'s linear classifier needs tuning more layer to obtain the best results and the parameter with little update can gain the best result. Then we choose the best model of *FC1*, *FC2* and *FC3* for comparison. Table 3 shows the results of different architecture of fully connected layer on the testing set. *FC3* achieves the best result ($UNet_{SSL}^{UCSD}\_F$ in Table 1). But, if we use SVM following *FC3*, the improvement is about 0.1 of accuracy, AUC, sensitivity, and specificity ($UNet_{SSL}^{UCSD}\_F\_SVM$ in Table 1). The reason is that our dataset is small scale and SVM classifier with the maximum margin's optimizing objective obtains is a more robust classifier for small scale dataset. Using SVM following full connected layers can get better performance than only using full connected layers.

4.6. Comparisons using different encoder modules

In our proposed method, we use the original U-Net's encoder in pre-training, the results of accuracy, AUC, sensitivity, and specificity on $UNet_{SSL}^{UCSD}\_F$ are 0.81, 0.89, 0.82 and 0.80, respectively (Table 1.). We replace the U-Net's encoder with ResNet-50 ($ResNet_{SSL}^{UCSD}\_F$) for comparison, since we found that ResNet-50 is a promising model for predicting the effectiveness of anti-VEGF by OCT images using ImageNet for pre-training based on our previous study [27]. However, using the complicated ResNet-50 as the encoder obtain worse performance: accuracy, AUC, sensitivity, and specificity are 0.75, 0.84, 0.76 and 0.75, respectively. This may be because the complex structure of skip connection and deeper model learn more general feature representation of UCSD OCT images leading to harder fine-tuning on our XJTU dataset.

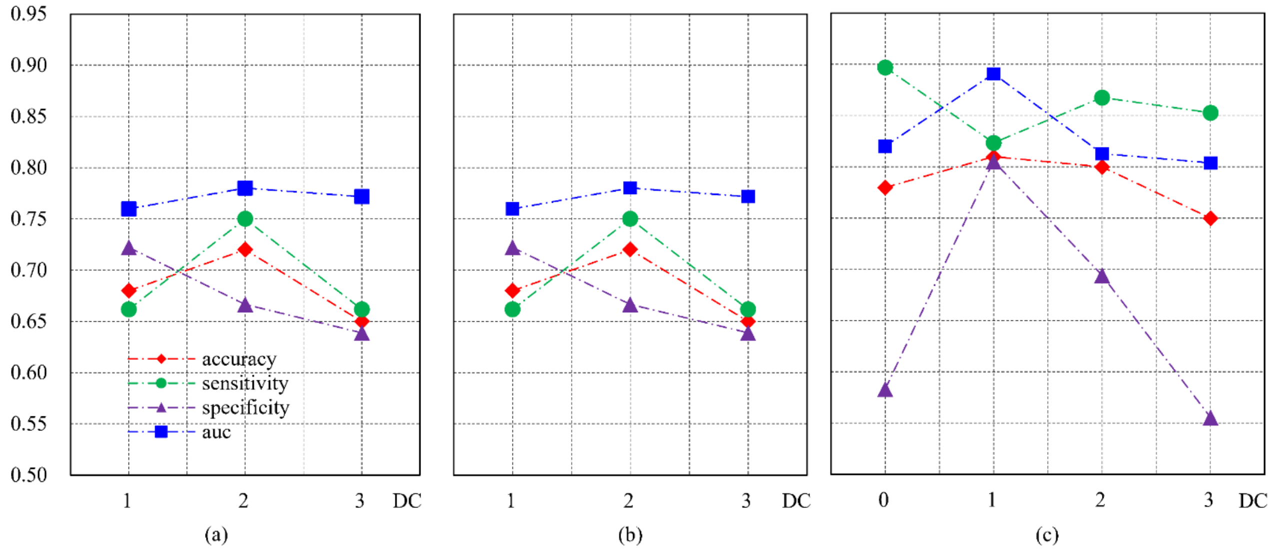

Figure 6. Results of different freezing manners in *FC1*, *FC2*, *FC3*. The horizontal axis DC (0,1,2,3) represents that modules of DoubleConv0 to DoubleConv DC are frozen and the remaining modules are updated during fine-tuning. (a) Performance of *FC1*, (b) Performance of *FC2*, (c) Performance of *FC3*.

Table 3. Best results of different fully connected layers.

Fully connected layers	Accuracy	AUC	Sensitivity	Specificity
FC1	0.72	0.78	0.75	0.67
FC2	0.73	0.84	0.72	0.75
FC3	0.81	0.89	0.82	0.80

4.7. Comparisons between hand-crafted features and deep learning features

Table 4 shows the prediction results by using different features. In hand-crafted feature experiments, we can observe that LBP features can make the best prediction although with a little imbalance between sensitivity and specificity. HoG-PCA, LBP-HoG-PCA and Intensity-Texture all show obvious imbalance between sensitivity and specificity. Table 4 shows that using deep features can make the best prediction. However, combing deep features and LBP features degrades the performance on accuracy and AUC. The hand-crafted features are extracted without the class labels' supervision, and the deep features are the results of model pre-trained on UCSD OCT images by self-supervised learning and fine-tuned on XJTU OCT images by supervised learning. Therefore, the deep learning feature-based model can convey more specific features for our task and achieve better classification results.

5. Discussion

During pre-training, U-Net served as the encoder-decoder to restore the deformed OCT image into original one where the encoder can learn the general feature representation of OCT images. Table 1 suggests that UCSD-U-Net can produce superior result than ImageNet-ResNet. In addition to no requirement of labeled OCT images for specific task in self-supervised learning manner, low amount of unlabeled source domain dataset can achieve satisfactory performance. In our study, only 1,000UCSD OCT images for pre-training can provide good feature representation. This is because the combined image transformations with random probabilities can lead to more complicated and enormous deformed images in different epochs.

Table 4. Results of hand-crafted features, deep features, and combinations for prediction.

Feature	Accuracy	AUC	Sensitivity	Specificity
LBP	0.66	0.68	0.66	0.60
HoG-PCA	0.67	0.47	0.79	0.31
LBP-HoG-PCA	0.66	0.48	0.78	0.30
Intensity-Texture	0.67	0.64	0.74	0.46
Deep	**0.93**	**0.98**	**0.94**	**0.91**
Deep-LBP	0.92	0.95	0.93	0.92

In image distortion stage of SSL, each OCT image is processed by a combination of transformations, local pixel shuffling, non-linear transformation, image in-painting or image out-painting, according to the probability threshold of each transformation. A variety of distortion operations ensure the generalization of the pre-trained model. For example, model learns tissue appearance via non-linear transformation, learns tissue texture via local pixel shuffling and learns context via image in-painting or image out-painting. In this study, we conducted ablation experiments where one type of image transformation was removed and other settings were unchanged. From the results in Table 5, even if one type of image transformation is removed, the performance of the model is unacceptable. AUC drops down to about 0.73 and there is an extreme imbalance between sensitivity and specificity, which means the three types of image distortion are all indispensable.

Table 5. Results of one type of image transformation removed in SSL based pre-training

Image transformation removed	Accuracy	AUC	Sensitivity	Specificity
Non-linear transformation	0.51	0.73	0.78	0.37
Local pixel shuffling	0.65	0.72	0.32	0.88
Image in-painting/out-painting	0.71	0.76	0.30	0.93

In our previous study, we concluded that the results of predicting effectiveness of anti-VEGF using full retinal images are better than that of using lesion region images [27]. In this study, we make more exploration by visualizing the heatmap of full retinal images based on the $UNet_{SSL}^{UCSD}\_F\_SVM$ method. Gradient-weighted class activation mapping (Grad-CAM) [56] is a technique of obtaining the visual explanations for decisions from CNN. Grad-CAM combines the feature maps from forward propagation phase and gradients of class score from backward propagations. The regions that contribute to decision of anti-VEGF effectiveness are identified by Grad-CAM. Figure 7 shows the Grad-CAM visualization result of one effective case. The color variations mean the contribution level and red represents highest contribution and blue represents the lowest. The illustration suggests that the decision of anti-VEGF's effectiveness not only depends on the lesion region but also is related to the non-lesion region. This conclusion is consistent with our previous study [27] as well.

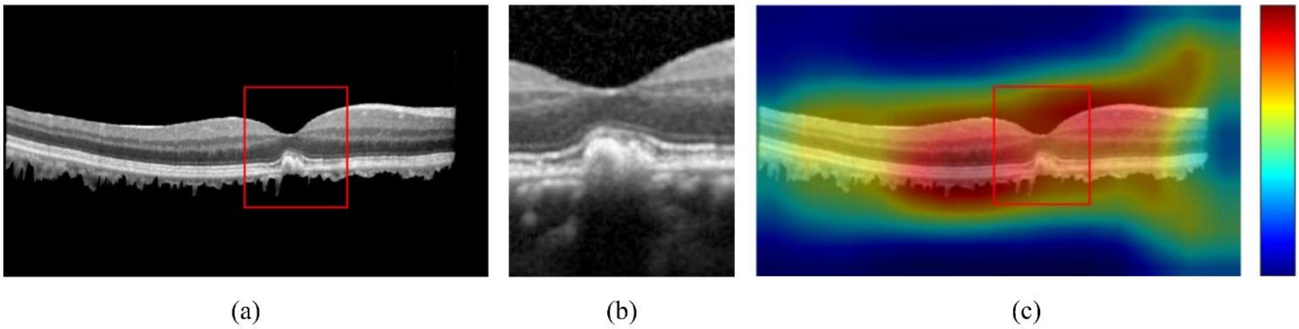

Figure 7. Visualization of contribution regions: (a) Full retinal image preprocessed with contoured lesion region, (b) Enlarged lesion region, (c) Grad-CAM visualization of contribution level to the effectiveness decision.

Though the proposed method can achieve good performance, there are still some limitations. The amount of target domain data set is insufficient and data augmentation strategy is used to increase the number for fine-tuning or training from scratch. Although data augmentation can alleviate the lack of data, more OCT images should be collected for better fine-tuning. And another limitation is that we conducted experiments without distinguishing CNV cases, CME cases and patients with both. These three kinds of patients can all be treated by anti-VEGF injection but they differ in pathology. For instance, the average CNV size for CNV patients with CME is 5.4 Macular Photocoagulation Study (MPS) disc areas and the average CNV size for CNV patients without CME is 5.6 MPS disc areas [4].

6. Conclusions

In this study, a new OCT-SSL model for predicting the effectiveness of anti-VEGF is developed. We introduce self-supervised learning to pre-train a U-Net on UCSD OCT dataset to obtain general feature representation. Then fine-tuning procedure is conducted on XJTU OCT dataset to learn more specific features. Finally, a classification model is built to obtain treatment outcome. The experimental results showed that OCT-SSL can achieve promising performance and can be used to assist ophthalmologists to make more optimal treatment plan. In future, we will collect more OCT images for predicting effectiveness of anti-VEGF with three groups: CNV patients, CME patients, and patients contracted with both CNV and CME.

Acknowledgments

This work is supported by the Science and Technology Program of Shaanxi Province (No. 2021KW-01) and the National Key Research and Development Program of China (No. 2016YFA0202003).

Conflict of interest

The authors do not have relevant conflict of interest to disclose.

References

1. W. L. Wong, X. Su, X. Li, C. M. G. Cheung, R. Klein, C. Y. Cheng, et al., Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: A systematic review and meta-analysis, *Lancet Glob. Health*, **2** (2014), e106–e116, https://doi.org/10.1016/S2214-109X(13)70145-1.
2. H. E. Grossniklaus, W. Green, Choroidal neovascularization, *Am. J. Ophthalmol.*, **137** (2004), 496–503. https://doi.org/10.1016/j.ajo.2003.09.042
3. M. R. Hee, C. R. Baumal, C. A. Puliafito, J. S. Duker, E. Reichel, J. R. Wilkins, et al., Optical coherence tomography of age-related macular degeneration and choroidal neovascularization, *Ophthalmology,* **103** (1996), 1260–1270. https://doi.org/10.1016/S0161-6420(96)30512-5
4. T. D. Ting, M. Oh, T. A. Cox, C. H. Meyer, C. A. Toth, Decreased visual acuity associated with cystoid macular mdema in meovascular age-related macular degeneration, *Arch. Ophthalmol.*, **120** (2002), 731–737. https://doi.org/10.1001/archopht.120.6.731
5. N. Shah, M. G. Maguire, D. F. Martin, J. Shaffer, G. S. Ying, J. E. Grunwald, et al., Angiographic cystoid macular edema and outcomes in the comparison of age-related macular degeneration treatments trials, *Ophthalmology*, **123** (2016), 858–864. https://doi.org/10.1016/j.ophtha.2015.11.030
6. A. Loewenstein, D. Zur, Postsurgical cystoid macular edema, in *Macular Edema*, (2010), 148–159. https://doi.org/10.1159/000320078
7. M. Rajappa, P. Saxena, J. Kaur, Chapter 6-ocular angiogenesis: Mechanisms and recent advances in therapy, *Adv. Clin. Chem.*, **50** (2010), 103–121. https://doi.org/10.1016/S0065-2423(10)50006-4
8. V. Tah, H. O. Orlans, J. Hyer, E. Casswell, N. Din, V. Sri Shanmuganathan, et al, Anti-VEGF therapy and the retina: An update, *J. Ophthalmol.*, **2015** (2015). https://doi.org/10.1155/2015/627674
9. H. Gerding, J. Mone´s, R. Tadayoni, F. Boscia, I. Pearce, S. Priglinger, Ranibizumab in retinal vein occlusion: Treatment recommendations by an expert panel, *Brit. J. Ophthalmol.*, **99** (2015), 297–304. https://doi.org/10.1136/bjophthalmol-2014-305041
10. D. Yorston, Anti-VEGF drugs in the prevention of blindness, *Community Eye Health,* **27** (2014), 44–46.
11. R. Marano, I. Toth, N. Wimmer, M. Brankov, P. Rakoczy, Dendrimer delivery of an anti-VEGF oligonucleotide into the eye: A long-term study into inhibition of laser-induced CNV, distribution, uptake and toxicity, *Gene Ther.*, **12** (2005), 1544–1550. https://doi.org/10.1038/sj.gt.3302579
12. J. H. Chang, N. K. Garg, E. Lunde, K. Y. Han, S. Jain, D. T. Azar, Corneal neovascularization: An anti-VEGF therapy review, *Surv. Ophthalmol.*, **57** (2012), 415–429. https://doi.org/10.1016/j.survophthal.2012.01.007
13. J. M. Schmitt, Optical coherence tomography (OCT): A review, *IEEE J. Sel. Top. Quant. Electron.*, **5** (1999), 1205–1215. https://doi.org/10.1109/2944.796348
14. K. Gnep, A. Fargeas, R. E. Gutie´rrez-Carvajal, F. Commandeur, R. Mathieu, J. D. Ospina, et al., Haralick textural features on T2-weighted MRI are associated with biochemical recurrence

following radiotherapy for peripheral zone prostate cancer, *J. Magn. Reson. Imaging*, **45** (2017), 103–117. https://doi.org/10.1002/jmri.25335

15. C. Shen, Z. Liu, Z. Wang, J. Guo, H. Zhang, Y. Wang, et al., Building CT radiomics based nomogram for preoperative esophageal cancer patients lymph node metastasis prediction, *Transl. Oncol.*, **11** (2018), 815–824. https://doi.org/10.1016/j.tranon.2018.04.005
16. R. Paul, S. H. Hawkins, Y. Balagurunathan, M. Schabath, R. J. Gillies, L. O. Hall, et al., Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma, *Tomography*, **2** (2016), 388–395. https://doi.org/10.18383/j.tom.2016.00211
17. W. Han, L. Qin, C. Bay, X. Chen, K. H. Yu, A. Li, et al., Integrating deep transfer learning and radiomics features in glioblastoma multiforme patient survival prediction, in *Medical Imaging 2020: Image Processing*, International Society for Optics and Photonics, (2020), 113132S. https://doi.org/10.1117/12.2549325
18. W. Han, L. Qin, C. Bay, X. Chen, K. H. Yu, N. Miskin, et al., Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas, *Am. J. Neuroradiology*, **41** (2020), 40–48. https://doi.org/10.3174/ajnr.A6365
19. N. Papandrianos, E. I. Papageorgiou, A. Anagnostis, Development of convolutional neural networks to identify bone metastasis for prostate cancer patients in bone scintigraphy, *Ann. Nucl. Med.*, **34** (2020), 824–832. https://doi.org/10.1007/s12149-020-01510-6
20. C. Fourcade, L. Ferrer, G. Santini, N. Moreau, C. Rousseau, M. Lacombe, et al., Combining superpixels and deep learning approaches to segment active organs in metastatic breast cancer PET images, in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, (2020), 1536–1539. https://doi.org/10.1109/EMBC44109.2020.9175683
21. K. H. Cha, L. Hadjiiski, H. P. Chan, A. Z. Weizer, A. Alva, R. H. Cohan, et al., Bladder cancer treatment response assessment in CT using radiomics with deep-learning, *Sci. Rep.*, **7** (2017), 1–12. https://doi.org/10.1038/s41598-017-09315-w
22. M. Byra, K. Dobruch-Sobczak, Z. Klimonda, H. Piotrzkowska-Wroblewska, J. Litniewski, Early prediction of response to neoadjuvant chemotherapy in breast cancer sonography using siamese convolutional neural networks, *IEEE J. Biomed. Health Inf.*, **25** (2020), 797–805. https://doi.org/10.1109/JBHI.2020.3008040
23. D. Romo-Bucheli, U. Erfurth, H. Bogunović, End-to-end deep learning model for predicting treatment requirements in neovascular AMD from longitudinal retinal OCT imaging, *IEEE J. Biomed. Health Inf.*, **24** (2020), 3456–3465. https://doi.org/10.1109/JBHI.2020.3000136
24. S. Starke, S. Leger, A. Zwanenburg, K. Leger, F. Lohaus, A. Linge, et al., 2D and 3D convolutional neural networks for outcome modelling of locally advanced head and neck squamous cell carcinoma, *Sci. Rep.*, **10** (2020), 1–13. https://doi.org/10.1038/s41598-020-70542-9
25. J. D. Cardosi, H. Shen, J. I. Groner, M. Armstrong, H. Xiang, Machine intelligence for outcome predictions of trauma patients during emergency department care, preprint, arXiv:2009.03873. https://doi.org/10.48550/arXiv.2009.03873
26. M. Pease, D. Arefan, J. Biligen, J. Sharpless, A. Puccio, K. Hochberger, et al., Deep neural network analysis of CT scans to predict outcomes in a prospective database of severe traumatic brain injury patients, *Neurosurgery*, **67** (2020), 417. https://doi.org/10.1093/neuros/nyaa447_417

27. D. Feng, X. Chen, Z. Zhou, H. Liu, Y. Wang, L. Bai, et al., A preliminary study of predicting effectiveness of anti-VEGF injection using OCT images based on deep learning, in 4*2nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (2020), 5428–5431. https://doi.org/10.1109/EMBC44109.2020.9176743
28. M. Noroozi, A. Vinjimoor, P. Favaro, H. Pirsiavash, Boosting self-supervised learning via knowledge transfer, in *2018 IEEE Conference on Computer Vision and Pattern Recognition*, (2018). https://doi.org/10.1109/CVPR.2018.00975
29. C. Doersch, A. Gupta, A. A. Efros, Unsupervised visual representation learning by context prediction, in *2015 IEEE International Conference on Computer Vision*, (2015), 1422–1430. https://doi.org/10.1109/ICCV.2015.167
30. M. Noroozi, P. Favaro, Unsupervised learning of visual representations by solving jigsaw puzzles, preprint, arXiv: 1603.09246. https://doi.org/10.48550/arXiv.1603.09246
31. L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, et al., Self-supervised learning for medical image analysis using image context restoration, *Med. Image Anal.*, **58** (2019), 101539. https://doi.org/10.1016/j.media.2019.101539
32. J. Zhu, Y. Li, Y. Hu, K. Ma, S. K. Zhou, Y. Zheng, Rubik's cube+: A self-supervised feature learning framework for 3D medical image analysis, *Med. Image Anal.*, **64** (2020), 101746. https://doi.org/10.1016/j.media.2020.101746
33. Z. Zhou, V. Sodha, M. M. R. Siddiquee, R. Feng, N. Tajbakhsh, M. B. Gotway, et al., Models genesis: Generic autodidactic models for 3D medical image analysis, in *International Conference on Medical Image Computing and Computer Assisted Intervention*, (2019), 384–393. https://doi.org/10.1007/978-3-030-32251-9_42
34. G. Larsson, M. Maire, G. Shakhnarovich, Colorization as a proxy task for visual understanding, in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 840–849. https://doi.org/10.1109/CVPR.2017.96
35. T. Chen, S. Kornblith, M. Norouzi, G. E. Hinton, A simple framework for contrastive learning of visual representations, in *Proceedings of the 37th International Conference on Machine Learning*, preprint, arXiv: 2002.05709.
36. A. Van den Oord, Y. Li, O. Vinyals, Representation learning with contrastive predictive coding, preprint, arXiv: 1807.03748.
37. G. Lorre, J. Rabarisoa, A. Orcesi, S. Ainouz, S. Canu, Temporal contrastive pretraining for video action recognition, in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, (2020), 662–670. https://doi.org/10.1109/WACV45572.2020.9093278
38. L. Tao, X. Wang, T. Yamasaki, Self-supervised video representation learning using inter- intra contrastive framework, in *MM'20: The 28th ACM International Conference on Multimedia*, (2020), 2193–2201. https://doi.org/10.1145/3394171.3413694
39. Z. Wu, Y. Xiong, S. X. Yu, D. Lin, Unsupervised feature learning via non-parametric instance discrimination, in *2018 IEEE Conference on Computer Vision and Pattern Recognition*, (2018), 3733–3742. https://doi.org/10.1109/CVPR.2018.00393
40. K. He, H. Fan, Y. Wu, S. Xie, R. B. Girshick, Momentum contrast for unsupervised visual representation learning, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2020), 9726–9735. https://doi.org/10.1109/CVPR42600.2020.00975

41. H. Spitzer, K. Kiwitz, K. Amunts, S. Harmeling, T. Dickscheid, Improving cytoarchitectonic segmentation of human brain areas with self-supervised siamese networks, in *Medical Image Computing and Computer Assisted Intervention-MICCAI 201-21st International Conference*, (2018), 663–671. https://doi.org/10.1007/978-3-030-00931-1
42. X. Li, M. Jia, M. T. Islam, L. Yu, L. Xing, Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis, *IEEE Trans. Med. Imaging*, **39** (2020), 4023–4033. https://doi.org/10.1109/TMI.2020.3008871
43. D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, *Cell*, **172** (2018), 1122–1131. https://doi.org/10.1016/j.cell.2018.02.010
44. M. E. Mortenson, *Mathematics For Computer Graphics Applications*, Industrial Press Inc., 1999.
45. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention-MIC-CAI 2015-18th International Conference Munich*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4
46. C. Chang, C. Lin, LIBSVM: A library for support vector machines, ACM Trans. *Intell. Syst. Technol.*, **2** (2011), 1–27. https://doi.org/10.1145/1961189.1961199
47. V. N. Vapnik, An overview of statistical learning theory, *IEEE Trans. Neural Networks,* **10** (1999), 988–999. https://doi.org/10.1109/72.788640
48. D. Lu, K. Popuri, G. W. Ding, R. Balachandar, M. F. Beg, Multiscale deep neural network ased analysis of FDG-PET images for the early diagnosis of Alzheimer's disease, *Med. Image Anal.*, **46** (2018), 26–34. https://doi.org/10.1016/j.media.2018.02.002
49. Y. Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, J. M. Rehg, Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding, *Med. Image Anal.*, **15** (2011), 748–59. https://doi.org/10.1016/j.media.2011.06.005
50. P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, et al., Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images, *Biomed. Optics Express*, **5** (2014), 3568–577. https://doi.org/10.1364/BOE.5.003568
51. R. Rasti, H. Rabbani, A. Mehridehnavi, F. Hajizadeh, Macular OCT classification using a multi-scale convolutional neural network ensemble, *IEEE Trans. Med. Imaging*, **37** (2017), 1024–034. https://doi.org/10.1109/TMI.2017.2780115
52. X. Wang, H. Chen, A. R. Ran, L. Luo, P. P. Chan, C. C. Tham, et al., Towards multi-center glaucoma OCT image screening with semi-supervised joint structure and function multi-task learning, *Med. Image Anal.*, **63** (2020), 101695. https://doi.org/10.1016/j.media.2020.101695
53. H. Bogunović, S. M. Waldstein, T. Schlegl, G. Langs, A. Sadeghipour, X. Liu, et al., Prediction of Anti-VEGF treatment requirements in neovascular amd using a machine learning approach, *Investi. Ophth. Vis. Sci.*, **58** (2017), 3240–3248. https://doi.org/10.1167/iovs.16-21053
54. Z. Zhou, M. Folkert, P. Iyengar, K. Westover, Y. Zhang, H. Choy, et al., Multi-objective radiomics model for predicting distant failure in lung SBRT, *Phys. Med. Biol.*, **62** (2017), 4460. https://doi.org/10.1088/1361-6560/aa6ae5

55. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770–78. https://doi.org/10.1109/CVPR.2016.90

56. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in *IEEE International Conference on Computer Vision*, (2017), 618–626. https://doi.org/10.1109/ICCV.2017.74