

Learning Self-Supervised Representations for Label Efficient Cross-Domain Knowledge Transfer on Diabetic Retinopathy Fundus Images

Ekta Gupta^γ, Varun Gupta^γ, Muskaan Chopra^γ, Prakash Chandra Chhipa^δ and Marcus Liwicki^δ

^δ Machine Learning Group, EISLAB, Luleå Tekniska Universitet, Luleå, Sweden
{prakash.chandra.chhipa, marcus.liwicki}@ltu.se

^γ Chandigarh College of Engineering and Technology, Punjab University, Chandigarh, India
ekta_cse@ccet.ac.in, varungupta@ccet.ac.in, chopramuskaan47@gmail.com

Abstract—This work presents a novel label-efficient self-supervised representation learning-based approach for classifying diabetic retinopathy (DR) images in cross-domain settings. Most of the existing DR image classification methods are based on supervised learning which requires a lot of time-consuming and expensive medical domain experts-annotated data for training. The proposed approach uses the prior learning from the source DR image dataset to classify images drawn from the target datasets. The image representations learned from the unlabeled source domain dataset through contrastive learning are used to classify DR images from the target domain dataset. Moreover, the proposed approach requires a few labeled images to perform successfully on DR image classification tasks in cross-domain settings. The proposed work experiments with four publicly available datasets: EyePACS, APTOS 2019, MESSIDOR-1, and Fundus Images for self-supervised representation learning-based DR image classification in cross-domain settings. The proposed method achieves state-of-the-art results on binary and multi-classification of DR images, even in cross-domain settings. The proposed method outperforms the existing DR image binary and multi-class classification methods proposed in the literature. The proposed method is also validated qualitatively using class activation maps, revealing that the method can learn explainable image representations. The source code and trained models are published on GitHub¹.

Index Terms—Self-supervised representation learning, domain adaptation.

I. INTRODUCTION

In the medical imaging area, artificial intelligence (AI), a topic characterized broadly by the building of computerized systems capable of doing tasks [1] & [2] that ordinarily require human intelligence, has significant potential. Automated radiology workflows have significantly benefited from machine learning and deep learning techniques. Although AI models have the potential to revolutionize clinical practice, they have been hampered by significant implementation and regulatory obstacles [3]. Almost all constraints may be traced back to a major issue: a dearth of medical image data to train and test AI algorithms [4]. The generalizability and accuracy of developed solutions are hampered because most research institutions and

enterprises only have limited access to annotated medical images. Large datasets, including high-quality images and annotations, are still necessary to train, validate, and test the AI systems [5]. In the absence of data that has been appropriately labeled, this procedure becomes prohibitively expensive, time-demanding, and inherently unstable. Labeled biomedical im-

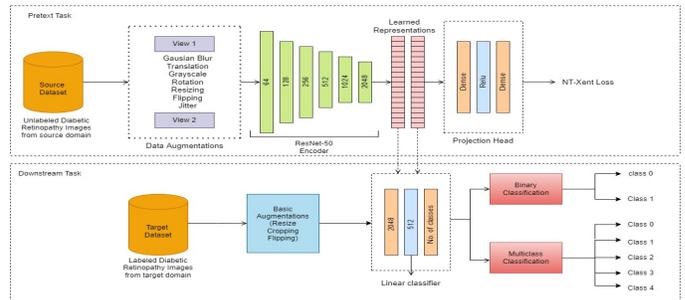


Fig. 1: Schematic presents the contrastive learning-based self-supervised cross-domain knowledge transfer. Pretraining is performed on the source dataset (EyePACS), and downstream tasks are performed on cross-domain targets (APTOS 2019, MESSIDOR, and Fundus Images).

ages are incredibly scarce, and multiple experts are required to annotate each image [6] manually. Massive amounts of health data are being generated and collected. These data range from in-hospital monitoring to wearable. Coding and annotating this data is impractical [6] [7]. In addition, the pretrained models obtained from natural images do not apply directly to medical images since their intensity distribution is different. Besides that, annotating natural images is simple; all that is required is basic human knowledge [8]. Nevertheless, in-depth knowledge is necessary for the annotation of medical images. The average medical image has over a billion pixels, significantly larger than others. The annotation process is highly error-prone and expensive, and experts cannot always identify a particular feature. A potential solution is to train models on unlabeled images using self-supervised learning [9] [10].

Most supervised learning methods require labeled data to train a machine. Unfortunately, obtaining good-quality labeled

¹<https://github.com/prakashchhipa/Learning-Self-Supervised-Representations-for-Label-Efficient-Cross-Domain-Knowledge-Transfer-on-DRF>

data can be cost-effective and time-consuming. Additionally, the data preparation lifecycle can be extremely long and complicated, including cleaning, filtering, annotating, reviewing, and restructuring according to a training framework [11]. Another approach has been used to deal with scarce biological data: domain adaptation-based self-supervised learning. Self-supervised learning (SSL), an alternative to supervised learning and transfer learning, has emerged as a viable possibility [12]. While self-supervised learning is distinct from transfer learning, both rely on acquiring representations from a secondary pretext activity and the subsequent transfer of those representations to the main focus task [13] [14]. The data utilized for the pretraining phase and the downstream task might be taken from one or more separate data sources in domain adaptation self-supervised learning, unlike transfer learning [15].

This work aims to prove that self-supervised learning can be used as a preliminary step in medical image classification. Contributions to this work are:

- This work proposes a domain adaptation-based self-supervised learning approach to learn image representations from diabetic retinopathy fundus images.
- Self-supervised learning of diabetic retinopathy image representations from unlabelled datasets has been validated in cross-domain settings, as shown in Figure 1.
- Results indicate that the proposed work outperforms the existing methods of DR classification.

II. RELATED WORK

In recent years, unsupervised learning has significantly progressed (SSL). Since it is useful for learning feature representations from image datasets without image labels, it has become a primary focus for academic investigation. Medical image classification tasks like detecting diabetic retinopathy, classifying brain age [35], recognizing cancer in histopathology [34], identifying pneumonia in X-ray [37], many others have shown progress using self-supervised learning methods, demonstrating state-of-the-art performance. This article focuses on self-supervised learning techniques pertaining to images of diabetic retinopathy.

In an approach, Truong et al. proposed the fusion of embeddings from multiple SSL models. Then the fused embeddings are combined with self-attention for feature learning. However, it did not use domain adaptation, as the datasets used in the pretext and downstream tasks are the same.

Taleb et al. [20] developed a series of five proxy tasks: 3D contrastive predicting coding, 3D rotation prediction, 3D jigsaw puzzles, 3D patch location, and 3D exemplar networks to learn the feature representations in the pretext task. Lin et al. [21] proposed a multilabel classification method using rotation SSL with graph CNN to learn fundus images' representations. Another work by Srinivasan et al. [22] trained a ResNet50 model using the MoCo-V2 approach in the pretext task to classify diabetic retinopathy images in the downstream task. The authors used a similar dataset for both the pretext and downstream tasks. Yap and Ng [23] proposed a contrastive

learning framework to create a patch-level dataset for pretext tasks by extracting the class activation maps from the labeled and unlabelled datasets.

A. SSL methods for DR segmentation

Segmentation of diabetic retinopathy using self-supervised learning has been explored partially. Tian et al. [18] proposed a multi-class Strong Augmentation via Contrastive Learning (MSACL) approach for detecting unsupervised anomalies. The author also proposes a contrastive loss that combines contrastive learning with multi-centered loss to cluster the samples of the same class. These unsupervised models need to be well-trained. Otherwise, they can learn ineffective image representations. Another author, Kukacka et al. [19], also proposed an approach for lesion segmentation by pretraining a U-net encoder in the pretext task.

B. Reconstruction-based SSL methods for DR classification

Many efforts have been made for the diabetic retinopathy classification task using reconstruction-based self-supervised learning methods. Holmberg et al. [16] proposed a cross-domain U-net-based system to generate the retinal thickness used for the classification during the downstream task. Other authors, Ngyyen et al., learned the features of the target dataset by using a self-supervised contrastive learning method on reconstructed retinal images. This work reconstructs images and features learning on the target dataset. In the proposed work, representations learning is performed on the source dataset of diabetic retinopathy and applied to those learned representations to a different dataset of diabetic retinopathy. In addition, a few authors also proposed a multi-modal-based reconstruction of images. Hervella et al. [3] performed multimodal reconstruction using U-Net for the segmentation task of the optic disk and cup in retinography and angiography images, and Li et al. [17] trained a CycleGAN model on the source dataset to learn the mapping function between the images and also learned both the modality-invariant and patient-similarity features in the pretext task. One more work by Cai et al. proposed a transformer-based framework in combination with a multitask decoder to learn the representations of the reconstructed images. Most works discussed above used adversarial learning methods to reconstruct the images. However, these methods provide inferior performance or have unstable training. Representation learning is pixel-based learning in the existing reconstruction-based methods, but the work focuses on learning representations at the visual concept level.

As was seen in the preceding review of the relevant literature, most existing SSL approaches employ the same dataset for both the pretext task in the source domain and the downstream task in the target domain. Progress has been seen in the knowledge transfer field, but no one has extensively explored the domain adaptation. The proposed work concretely focuses on cross-domain contrastive learning. To identify DR images from a different domain, this study presents an SSL

strategy to reuse the representations learnt on one unlabeled dataset from the source domain during the pretext job.

III. DIABETIC RETINOPATHY DATASET DESCRIPTION

Diabetic retinopathy is a major cause of blindness among people of working age in developed countries. It is a prevalent eye disease that affects more than 93 million people globally. Diabetic retinopathy detection is currently a time-consuming and laborious technique that requires a skilled person to analyze and interpret digital color fundus images of the retina. The public datasets for diabetic retinopathy are:

A. Subset of EyePACS

Eye disease Diabetic Retinopathy (DR)² is linked to long-term diabetes. If DR is caught early enough, visual loss can be halted. A comprehensive collection of high-resolution retina images captured using various imaging settings are accessible. Every subject has both a left and right field available to them. Images are identified not just with a subject id but also as being on the left or the right. A medical professional has determined diabetic retinopathy on a scale of 0 to 4.

B. APTOS 2019

Numerous people are affected by diabetic retinopathy, the most common reason for vision loss among adults in their 40s and 50s. Aravind Eye Hospital can help people in rural areas without easy access to medical screening in India's efforts to find and prevent this condition there. The answers will be available to other Ophthalmologists through the 4th Asia Pacific Tele-Ophthalmology Society (APTOS) Symposium³. A vast collection of retina images was collected using fundus photography in various situations made available. A clinical expert has determined that each image has been graded for its severity on a scale of 0 to 4.

C. MESSIDOR-I

Diabetic retinopathy detection is now a labor-intensive and time-consuming method that requires a qualified doctor to use digital color fundus images of the retina. It is known as MESSIDOR (Methods to Evaluate Segmentation and Indexing Techniques in the Field of Retinal Ophthalmology in French)⁴ [24]. The retinopathy grades are determined on a scale of 0 to 3.

D. Fundus Images

The Department of Ophthalmology provided the 757 color fundus images [37] included in this collection from the Hospital de Clinicas, Facultad de Ciencias Médicas, Universidad Nacional de Asunción, Paraguay. The Zeiss brand's Visucam 500 camera was utilized for the process of acquiring the retinographies. Fundus images have been classified into 7 distinct groups on a scale of 1 to 7.

Table I shows the dataset description of diabetic retinopathy.

<i>Dataset</i>	<i>Total Images</i>	<i>No. of Classes</i>
<i>Subset of EyePACS</i>	<i>31615</i>	<i>5</i>
<i>APTOS 2019</i>	<i>3660</i>	<i>5</i>
<i>MESSIDOR-I</i>	<i>1200</i>	<i>4</i>
<i>Fundus Images</i>	<i>747</i>	<i>7</i>

TABLE I: Diabetic retinopathy dataset description

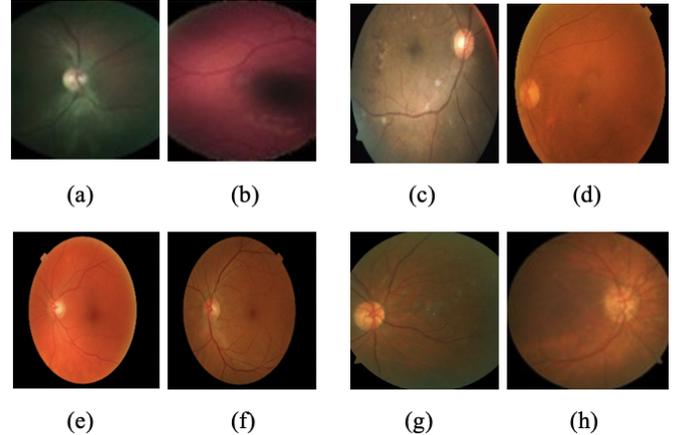


Fig. 2: Sample images- (a)(b) subset of EyePACS Dataset, (c)(d) Messidor-I dataset, (e)(f) APTOS 2019 dataset, (g)(h) Fundus images.

IV. SELF-SUPERVISED CROSS DOMAIN KNOWLEDGE TRANSFER FRAMEWORK

The proposed framework consists of two main tasks: (i) pretext task, i.e., representation learning of images from source domain DR dataset (a subset of EyePACS) (ii) downstream task, i.e., classification of DR(Diabetic Retinopathy) images from the target domain datasets (APTOS 2019, Messidor-I and Fundus Images). In the pretext task, the proposed approach applies various augmentations like flipping, affine transformations, jitter, grayscale, etc., to create different views from the images. The different views created from the same image act as positive pairs, and views from different images act as negative pairs. Then, image representations are learned through contrastive learning from positive and negative pairs of images. These learned representations of images act as input to the downstream task. This task does not require labeled images for representation learning as shown in Figure 1.

The downstream task involves binary as well as multi-class classification of DR images. The model pretrained for learning image representation during the pretext task act as initialization for performing the downstream task, i.e., classification of DR images. Now the downstream task requires fewer labeled images for performing DR classification. Figure 1 provides a detailed architecture of the proposed approach. The objective of the proposed approach is to obtain representations that are robust to domain shift and generalizable to the downstream task. The proposed approach uses an unlabeled source dataset

²<https://www.kaggle.com/c/diabetic-retinopathy-detection/data>

³<https://www.kaggle.com/competitions/aptos2019-blindness-detection/data>

⁴<https://www.adcis.net/en/third-party/messidor>

to learn the representations and a labeled target dataset to solve the classification task by reusing these features learned from the source dataset. The representations have been learned using the SimCLR (simple framework for contrastive learning) method [25]. As discussed, positive and negative pairs of DR images are created from unlabelled DR images using different augmentations like a Gaussian blur, flipping, translation, rotation, jitter, etc. These positive and negative pairs of images are fed to the encoder network. The encoder network consists of a ResNet-50 backbone and a projection head containing two fully connected layers of 2048 and 1024 neurons, respectively. This network is trained on positive and negative views of images using Normalized Temperature-scaled Cross-Entropy (NT-Xent) as the loss function, which tries to pull positive pairs close and push away the negative pairs. This loss function is defined as:

$$\ell(z_i, z_j) = -\log \frac{\exp(\text{sim}(Z_i Z_j)/T)}{\sum_{k=1}^{2n} \mathbb{1}_{k \neq i} \exp(Z_i Z_k/T)} \quad \dots(1)$$

Where z_i and z_j are representations of positive pairs, T is the temperature parameter, n is the number of images, and $\text{sim}()$ represents the similarity function. This loss function is the negative log-likelihood of similarity between positive pairs to the ratio of similarities between all possible positive and negative pairs. This loss function is a softmax function normalized using a temperature parameter. In the downstream task, the proposed approach performs binary and multi-classification of DR images separately from the target domain datasets (Messidor-I and APTOS 2019). During this phase, primary augmentations such as resizing, flipping, and cropping are applied to the target dataset of DR images to create different views. The encoder backbone ResNet-50 weights trained for the pretext task are used as the initialization for the network being trained for the downstream task. The projection head of the network used in the pretext task is replaced with two fully connected layers, i.e. (2048, 512) and (512, no. of classes). The proposed approach performs binary and multi-classification on APTOS 2019, Messidor-I, and Fundus Images datasets during this phase.

V. EXPERIMENTS AND RESULTS

To investigate the proposed domain adaptation framework, three datasets are explored - APTOS 2019, Messidor-I, and Fundus Image dataset. The investigation is performed in a manner that self-supervised pretrained the model on one source dataset, a Subset of EyePACS, and finetuned on the above-mentioned three target datasets. The classification task is performed on all the target datasets. Binary and multi-classification on all three target datasets are performed in the downstream task. This work also explored label efficiency by performing the experiments on 10, 30, 50, and 100 percent data from the datasets. The data augmentations are applied to generate two views of a single image which can be further tested for similarity. During the pretext task, flipping, cropping, translation, scaling, grayscale, rotation, blurring, and resizing augmentation techniques are applied to input medical images

to obtain better and more generalizable results. Due to the small dataset size, it is required to investigate various finetune scenarios. Only a few primary augmentations, like resizing and cropping, are used during the downstream task to make this approach more compelling. After performing numerous experiments with varied parameter settings, the hyperparameter values that gave promising results during the pretext task are given in Tables II & III. Table II shows the hyperparameters used for binary classification of diabetic retinopathy images for datasets APTOS 2019 and Messidor-I. It shows various probability values used for different augmentation techniques during the pretext task. The batch size used is 128, and the optimizer used for training is LARS (Layer-wise Adaptive Rate Scaling). The initial learning rate used is 0.79, and the weight decay is 10^{-6} . The performance metrics are defined below.

Augmentations	Parameters
<i>Resize</i>	224 X 224
<i>Horizontal Flip</i>	$P=0.5$
<i>Vertical Flip</i>	$P=0.5$
<i>Grayscale</i>	$P=0.2$
<i>Gaussian Blur</i>	$P = 0.5, \text{Kernel size} = [21, 21]$
<i>Batch size</i>	128
<i>Optimizer</i>	LARS
<i>Learning Rate</i>	0.79
<i>Weight-decay</i>	10^{-6}

TABLE II: Hyperparameters for binary classification

Table III shows the hyperparameters used for a multi-classification of diabetic retinopathy images for the dataset APTOS 2019. The augmentations in the multi-classification of DR images are – jitter, affine, and normalization, along with the augmentations used in the binary classification for better performance. The batch size used for multi-classification is 256, and the optimizer used for training is LARS.

The classification task is performed on three datasets – APTOS 2019, Messidor-I, and Fundus Images. Table IV shows the results obtained for the binary classification on APTOS 2019 dataset. For APTOS 2019, the proposed method obtains an accuracy of 99.59%, a precision of 100%, a recall of 99.54%, and an F1- score of 99.26% on only 10% images.

Augmentations	Parameters
<i>Resize</i>	224 X 224
<i>Horizontal Flip</i>	$P = 0.5$
<i>Normalization</i>	Mean=(0.425, 0.297, 0.212) Standard deviation = (0.276, 0.202, 0.168)
<i>Jitter</i>	Brightness: 0.4; Contrast: 0.4; Saturation: 0.4; Hue: 0.1
<i>Affine</i>	Degrees= (-180, 180), Translate= (0.2, 0.2)
<i>Grayscale</i>	$P = 0.2$
<i>Batch size</i>	256
<i>Optimizer</i>	LARS
<i>Learning Rate</i>	10^{-3}
<i>Weight-decay</i>	5×10^{-4}

TABLE III: Hyperparameters for multi-classification

Dataset used			Accuracy	Precision	Recall	F1-Score
Pretext	Downstream					
EyePACS	APTOS 2019	10%	96.88	100	94.23	96.44
		20%	99.48	100	98.54	99.01
		50%	99.56	99.01	99.62	99.18
		100%	99.59	100	99.54	99.26

TABLE IV: Results for binary classification using the proposed approach on APTOS 2019

Dataset used			Accuracy	Precision	Recall	F1-Score
Pretext	Downstream					
EyePACS	Messidor-I	10%	67.48	71.43	69.89	72.56
		20%	70.31	75.21	74.87	70.55
		50%	74.96	81.47	73.66	76.75
		100%	98.49	98.65	100.00	99.99

TABLE V: Results for binary classification using the proposed approach on Messidor-I

For 100% images, the accuracy improved by 2.71%, recall by 5%, and F1-Score by 3%.

For another dataset- Messidor-I, the highest accuracy obtained is 98.49%, precision is 98.65%, recall is 100%, and F1 score is 99.99% as shown in Table V.

For the third dataset- Fundus Images, the accuracy obtained on 100% images is 98.96%, precision is 96%, recall is 99.43%, and F1 score is 99.67% as shown in Table VI.

Dataset used			Accuracy	Precision	Recall	F1-Score
Pretext	Downstream					
EyePACS	Fundus Images	10%	92.15	93.44	91.56	93.32
		20%	93.75	95.66	91.54	95.98
		50%	96.43	98.32	98.22	96.44
		100%	98.96	100.00	99.43	99.67

TABLE VI: Results for binary classification on Fundus Images

Table VII shows the results of the multi-classification of DR images by using the same dataset for pretext and downstream tasks, i.e., the Subset of EyePACS. Table VIII shows the outcomes of the three datasets' multi-classification of DR images. The downstream dataset used for the multi-classification of diabetic retinopathy images is APTOS 2019, Messidor-I, and Fundus Images, where the proposed method obtained an accuracy of 83.43%, 66.39%, and 91.67%. The proposed self-supervised learning method outperforms prior state-of-the-art techniques on two datasets - Aptos 2019 and Fundus Images.

Dataset used			Accuracy	Precision	Recall	F1-Score
Pretext	Downstream					
Eyepacs	Eyepacs	10%	74.56	69.31	75.32	70.02
		20%	77.23	72.78	76.99	73.21
		50%	77.76	75.98	77.06	74.99
		100%	77.82	73.71	77.41	74.98

TABLE VII: Multi-classification results on eyepacs dataset

Figure 3 represents the class activation maps (CAMs) generated for the three downstream datasets.

A. Comparison with the existing work

Domain adaptation-based self-supervised learning on Messidor-I, Fundus Images, and APTOS 2019 datasets is

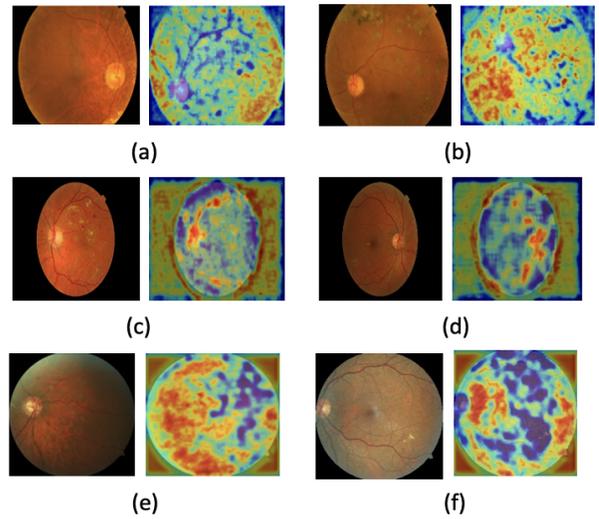


Fig. 3: CAMs representations for the datasets used. (a) & (b) APTOS 2019 (c) & (d) Messidor I (e) & (f) Fundus Images

Dataset used			Accuracy	Precision	Recall	F1-Score
Pretext	Downstream					
EyePACS	APTOS 2019	10%	76.04	78.45	89.23	88.45
		30%	78.15	62.32	83.34	78.76
		50%	83.12	81.12	82.76	84.45
		100%	83.43	81.09	85.54	77.86
	Messidor-I	10%	54.16	67.23	91.75	69.34
		30%	47.22	54.12	51.34	54.97
		50%	65.83	64.54	65.45	72.21
		100%	66.39	70.87	68.79	71.23
	Fundus Images	10%	75.0	67.98	67.65	80.40
		30%	81.25	85.90	91.43	88.49
		50%	82.00	84.12	80.32	91.23
		100%	91.67	86.01	92.43	94.54

TABLE VIII: Multi-classification results on three downstream datasets

unexplored. This work has considered various methods applied to these datasets for comparison purposes, including supervised learning-based methods. Table IX compares the results obtained from the proposed work with the existing methods for binary classification on Messidor-1, Fundud Images, and APTOS 2019 datasets. Chakraborty et al. [26] used ANN for binary classification and achieved an accuracy of 97.13% on the Messidor dataset. A CNN-based model proposed by Islam et al. [27] for the DR classification APTOS 2019 dataset achieved an accuracy of 98.36%. However, the proposed method reports improved results (accuracy of 99.59%, 98.49%, and 98.96%) on APTOS 2019, Messidor, and Fundus Images datasets.

Method	Accuracy	Precision	Recall
Dataset - Messidor			
Abramoff et al [28]	96.7	96.80	87.00
Chakraborty et al [26]	97.13	97.20	97.00
Dhanasekaran et al. [29] (SVM)	97.89	98.68	100.00
Dhanasekaran et al. [29] (PNN)	94.76	96.64	98.46
Proposed work - SSL Cross domain	98.49	98.00	100.00
Dataset - APTOS 2019			
Islam et al [27]	98.36	98.37	98.36
Proposed work - SSL Cross domain	99.59	100.00	99.00

TABLE IX: Comparative results for binary classification

Table X displays the results of a comparison between the proposed study and previous work in multi-class classification of DR images. Kassani et al. [32] reported the highest accuracy for multi-class classification was 83.09%. The remaining works reported an accuracy below 75% for classifying diabetic retinopathy images.

Authors	Accuracy	Precision	Recall
Kassani et al [32]	83.09	88.24	82.35
Gangwar & Ravi [33]	72.33	-	-
Proposed Work - SSL Cross domain	83.43	81.00	85.00

TABLE X: Comparative results for multi-classification on APTOS 2019 dataset. *No suitable previous work found for other datasets for multi-class classification*

The comparisons in Tables IX and X suggest that the performance achieved with the proposed work is improved over previous works for binary and multi-classification of diabetic retinopathy images.

B. Label efficiency in cross-domain knowledge transfer

The proposed self-supervised cross-domain knowledge method obtains concrete evidence for label efficiency. Result comparisons on both downstream tasks show that model achieves comparable performance when only 50% labels (half supervised) are used against fully supervised models with 100% labels. It is observed that downstream task performance differences between partially supervised and fully supervised are in close on at-least two datasets out of three target datasets for both classification tasks. Further, it is noticeable that the training portion of all the target datasets consists of only a few labeled examples in the range of 500 to 2500. Label efficiency is illustrated in the figures 4 & 5.

VI. CONCLUSION

This work proposes a label-efficient self-supervised representation learning-based method for diabetic retinopathy image classification in cross-domain settings. The proposed work has been evaluated qualitatively and quantitatively on the publicly available EyePACS, APTOS 2019, MESSIDOR-I, and Fundus Images datasets for binary and multi-classification of DR images. The qualitative evaluation shows that the proposed approach learns explainable image representations. Moreover, the proposed approach uses only a few training samples for training and outperforms the existing DR image

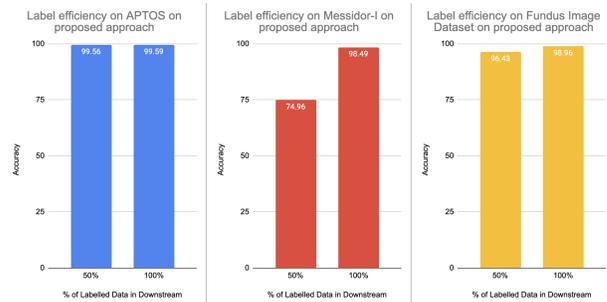


Fig. 4: Label efficiency on binary classification tasks for all three target datasets, which shows cross-domain knowledge transfer achieves comparable performance with only 50% label being used against the fully supervised model.

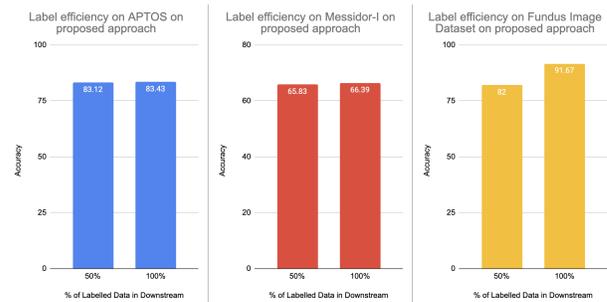


Fig. 5: Label efficiency on multi-class classification tasks for all three target datasets, which shows cross-domain knowledge transfer achieves comparable performance with only 50% label being used against the fully supervised model.

classification methods, even in cross-domain settings. In future work, the proposed approach can be used to investigate other downstream tasks, such as segmentation and localization. Further, non-contrastive methods for representation learning can be examined to perform downstream tasks on DR images in cross-domain settings.

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] Wang, Jian, et al. "A review of deep learning on medical image analysis." *Mobile Networks and Applications* 26 (2021): 351-380.
- [5] Fotedar, Gaurav, et al. "Extreme consistency: Overcoming annotation scarcity and domain shifts." *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I* 23. Springer International Publishing, 2020.
- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
- [7] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.

- [8] X. Zhang, J. Mu, X. Zhang, H. Liu, L. Zong, and Y. Li, "Deep anomaly detection with self-supervised learning and adversarial training," *Pattern Recognit.*, vol. 121, p. 108234, 2022.
- [9] A. Saeed, T. Ozcebe, and J. Lukkien, "Multi-task Self-Supervised Learning for Human Activity Detection," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 2, pp. 1–30, 2019.
- [10] T. Shen, C. Gou, F. Y. Wang, Z. He, and W. Chen, "Learning from adversarial medical images for X-ray breast mass segmentation," *Comput. Methods Programs Biomed.*, vol. 180, pp. 1–13, 2019.
- [11] O. Ciga, T. Xu, and A. L. Martel, "Self supervised contrastive learning for digital histopathology," *Mach. Learn. with Appl.*, vol. 7, no. May 2021, p. 100198, 2022.
- [12] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang, and T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification," *Proc. - 2017 10th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2017*, vol. 2018-Janua, no. 978, pp. 1–11, 2018.
- [13] Z. Li et al., "A novel multiple instance learning framework for COVID-19 severity assessment via data augmentation and self-supervised learning," *Med. Image Anal.*, vol. 69, p. 101978, 2021.
- [14] J. Lee and S. Y. Sohn, "Recommendation system for technology convergence opportunities based on self-supervised representation learning," *Scientometrics*, vol. 126, no. 1, pp. 1–25, 2021.
- [15] M. Hassan, S. Ali, H. Alqubayz, and K. Safdar, "Developing intelligent medical image modality classification system using deep transfer learning and LDA," *Sci. Rep.*, vol. 10, no. 1, pp. 1–14, 2020.
- [16] O. G. Holmberg et al., "Self-supervised retinal thickness prediction enables deep learning from unlabelled data to boost classification of diabetic retinopathy," *Nat. Mach. Intell.*, vol. 2, no. November, 2020.
- [17] X. Li, M. Jia, T. Islam, L. Yu, and L. Xing, "Self-supervised Feature Learning via Exploiting Multi-modal Data for Retinal Disease Diagnosis," vol. 0062, no. c, pp. 1–12, 2020.
- [18] Y. Tian et al., "Self-supervised multi-class pre-training for unsupervised anomaly detection and segmentation in medical images," *Institutional Knowl. Singapore Manag. Univ.*, pp. 1–10, 2021.
- [19] J. Kukačka, A. Zenz, M. Kollovieh, D. Jüstel, and V. Ntziachristos, "Self-Supervised Learning from Unlabeled Fundus Photographs Improves Segmentation of the Retina," 2021.
- [20] A. Taleb, W. Loetzsch, N. Danz, J. Severin, and T. Gaertner, "3D Self-Supervised Methods for Medical Imaging," no. *NeurIPS*, pp. 1–15, 2020.
- [21] J. Lin, Q. Cai, and M. Lin, "Multi-Label Classification of Fundus Images With Graph Convolutional Network and Self-Supervised Learning," vol. 28, pp. 454–458, 2021.
- [22] Srinivasan, V., Strodthoff, N., Ma, J., Binder, A., Müller, K. R., & Samek, W. (2021). On the robustness of pretraining and self-supervision for a deep learning-based analysis of diabetic retinopathy. *arXiv preprint arXiv:2106.13497*.
- [23] B. P. Yap and K. Ng, "Semi-weakly Supervised Contrastive Representation Learning for Retinal Fundus Images," pp. 1–9.
- [24] E. Decencière et al., "c," *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.
- [25] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," *37th Int. Conf. Mach. Learn. ICML 2020*, vol. PartF16814, no. Figure 1, pp. 1575–1585, 2020.
- [26] S. Chakraborty, G. C. Jana, D. Kumari, and A. Swetapadma, "An improved method using supervised learning technique for diabetic retinopathy detection," *Int. J. Inf. Technol.*, vol. 12, no. 2, pp. 473–477, 2020.
- [27] M. R. Islam et al., "Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images," *Comput. Biol. Med.*, vol. 146, no. May, p. 105602, 2022.
- [28] M. D. Abràmoff et al., "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investig. Ophthalmol. Vis. Sci.*, vol. 57, no. 13, pp. 5200–5206, 2016.
- [29] G. Mahendran and R. Dhanasekaran, "Investigation of the severity level of diabetic retinopathy using supervised classifier algorithms," *Comput. Electr. Eng.*, vol. 45, pp. 312–323, 2015.
- [30] G. Quéllec, H. Al Hajj, M. Lamard, P. H. Conze, P. Massin, and B. Cochener, "ExplAIIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis," *Med. Image Anal.*, vol. 72, no. 2016, 2021.
- [31] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA - J. Am. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [32] R. Kassani, Sara Hosseinzadeh; Kassani, Peyman Hosseinzadeh; Khazaeinezhad, Reza; Wesolowski, Michal J.; Schneider, Kevin A.; Deters, "Diabetic Retinopathy Classification Using a Modified Xception Architecture," *2019 IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, IEEE Access, vol. 4, pp. 1–16, 2019.
- [33] A. K. Gangwar and V. Ravi, *Diabetic Retinopathy Detection Using Transfer Learning and Deep Learning*, vol. 1176. Springer Singapore, 2021.
- [34] Chhipa, P. C., Upadhyay, R., Pihlgren, G. G., Saini, R., Uchida, S., & Liwicki, M. (2023). Magnification prior: a self-supervised method for learning representations on breast cancer histopathological images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 2717-2727).
- [35] Zhao, Q., Liu, Z., Adeli, E. and Pohl, K.M., 2021. Longitudinal self-supervised learning. *Medical image analysis*, 71, p.102051.
- [36] Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., ... & Norouzi, M. (2021). Big self-supervised models advance medical image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3478-3488).
- [37] V. E. Castillo Benítez, I. Castro Matto, J. C. Mello Román, J. L. Vázquez Noguera, M. García-Torres, J. Ayala, D. P. Pinto-Roa, P. E. Gardel-Sotomayor, J. Facon, and S. A. Grillo, *Dataset from fundus images for the study of diabetic retinopathy*, *Data in Brief*, vol. 36, p. 107068, Jun. 2021. doi: <https://doi.org/10.1016/j.dib.2021.107068>