Anusandhanvallari
Vol 2024, No.1
September 2024
ISSN 2229-3388

Role of Deep Learning in Early Detection and Prevention of Diabetic Retinopathy

Ahmad Talha Siddiqui*[1, 4], Harleen Kaur[2], Safdar Tanveer[3]
Research Scholar[1*], Department of Computer Science & Engineering
School of Engineering & Technology
Jamia Hamdard (Hamdard University)
Email-id: ahmadtalha2007@gmail.com
Professor[2], Department of Computer Science & Engineering
School of Engineering & Technology
Jamia Hamdard (Hamdard University)
Email-id: harleen@jamiahamdard.ac.in
Associate Professor[3], Department of Computer Science & Engineering
School of Engineering & Technology
Jamia Hamdard (Hamdard University)
Email-id:safdartanveer@jamiahamdard.ac.in
Assistant Professor[4], Department of CS & IT
Maulana Azad National Urdu University
Email-id: ahmadtalhasiddiqui@manuu.edu.in

ABSTRACT

Using unlabeled images, Self-Supervised Learning has evolved into a common method for learning image representations. Still, its use in medical picture analysis is not very well studied. In this study, we present a self-supervised Image Transformer that is led by saliency based on fundus pictures to grade Diabetic Retinopathy. Our method especially uses saliency maps in Self-Supervised Learning to direct the pre-training process using knowledge within a given domain. We particular suggest Two saliency-guided techniques for learning activities inside Self-Supervised Image Transformer: (1) Saliency-guided contrastive learning: To reduce. Unnecessary patches derived from momentum-updated key encoder input sequences, we utilize saliency maps of fundus pictures within conjunction with momentum contrast. As a result, the encoder for queries is directed to learn significant features from the prominent regions that the key encoder has focused on. (2) Saliency segmentation prediction: The query encoder is motivated to preserve detailed information in the acquired representations by being trained to predict saliency maps. Using four publicly accessible fundus imaging datasets, we do out comprehensive investigations. The efficiency of the representations learned through self-supervised Image Transformer is demonstrated by our results, which demonstrate that Self-Supervised Image Transformer performs noticeably better than a number of cutting-edge Self-Supervised Learning techniques across all datasets and evaluation circumstances.

KEYWORDS

Diabetic Retinopathy; Convolutional Neural Networks; Deep Learning; Self-Supervised Learning; Convolutional Neural Networks; Self-Supervised Image Transformer.

1. Introduction

The primary etiology of working-age blindness individuals in cultivated nations [1] is a serious complication of diabetes, diabetic retinopathy affects the eye's blood vessels and can ultimately cause irreversible vision damage loss if treatment is delayed. Fundus pictures are useful in identifying this disorder because they show particular biomarkers such as exudates, hemorrhages, microaneurysms, and retinal neovascularization [2]. But early

symptoms of Diabetic Retinopathy are sometimes subtle and hard to see, and even for seasoned professionals, screening is becoming more and more challenging as the number of diabetes patients rises. Automatic techniques to help with Diabetic Retinopathy detection are therefore desperately needed, particularly in places with little healthcare resources [3,45]. Convolutional Neural Networks which are extensively used in deep learning, have made major advancements in Diabetic Retinopathy detection over the last decades Diabetic Retinopathy grading is mechanized using this [4,41]. Diagnosing Diabetic Retinopathy can be challenging because the condition is silent and has no early warning indicators, making early detection challenging [44]. Historically, the diagnosis has been made by skilled medical professional's manually reviewing and assessing Digital Fundus Photography photographs [5, 43]. Depending on the amount of patients that require evaluation and the doctors' availability, this process may take several days. In addition, different doctors may get different results, and a doctor's accuracy greatly depends on their experience. Additionally, the technology and knowledge required may be inadequate in many areas with high Diabetic retinopathy is a widespread condition, affecting millions globally. Notably, convolutional neural networks (CNNs) have recently achieved breakthrough performance in medical imaging and other computer vision tasks

2. Literature Review

A. Deep Learning for DR Grading

The International Clinical Diabetic Retinopathy Scale divides the condition into four progressive stages: mild, moderate, and severe non-proliferative DR (NPDR), followed by the most advanced stage, proliferative DR (PDR). Clinicians use this classification system to assess disease progression and guide treatment decisions (0 is normal). [6]. Fundus imaging can identify biomarkers associated with DR, such as Hemorrhage, Key signs like exudates, microaneurysms, and retinal neovascularization can indicate diabetic retinopathy. To detect these markers, researchers are now using supervised deep learning approaches [8,42] have been more popular in recent years for Diabetic Retinopathy grading using fundus pictures. Because of their capacity to learn high-level features successfully, Convolutional Neural Networks are frequently used as the feature extraction module in these methods [9].

Recently, computer vision methods have become a go-to solution for medical image analysis, demonstrating remarkable accuracy in image identification applications [10,11] found that Vision Transformers are competitive, if not superior to, Convolutional Neural Networks in Diabetic Retinopathy grading, particularly for extensive datasets. Notwithstanding their potential, the application of Vision Transformers in medical imaging analysis remains restricted, owing to an absence of adequately annotated data and the fact that Vision Transformers have yet to be completely investigated. In our study, we suggest a framework for self-supervised learning for Vision Transformers to improve Diabetic Retinopathy grading by better using unannotated fundus pictures.

B. Self-supervised Learning in Natural Images

Self-supervised learning has attained significant achievement using computer vision. [12]. A popular in self-supervised learning, one common approach is to design a pretext task—an artificial challenge that helps the model learn meaningful representations from unlabeled data The key idea behind self-supervised learning? Create a pretext task that forces the model to learn useful features on its own that leverages the image's own information as a supervisory signal [13]. This challenge contributes to the training of neural network - an AI system modeled after the human brain - to automatically discover patterns Meaningful visual depictions. One of the most prevalent instances of prejudice used Techniques for self-supervised learning has revolutionized computer vision by allowing AI systems to learn from images without human annotations focuses on discriminating each Image by optimizing the congruence of representations from several enlarged perspectives of the identical image.

MoCo[14] and SimCLR[15] are at the forefront of instance discrimination techniques. MoCo, for example,

compares utilizing the method uses embedded features from a trained encoder to build a dynamic representation dictionary, which is continuously updated via a momentum encoder. In contrast, SimCLR compares images in batches. MoCo-v3 improves MoCo's performance for self-supervised vision transformers[16]. Another famous method is DINO[17], which employs self-distillation to train a Vision Transformers, where a student model learns to predict the feature representations generated by a momentum-based

While these algorithms excel at natural picture categorization, some studies have noted their reliance on global features, which may restrict their capacity to catch fine-grained details. To solve this, a new Self-supervised learning paradigm called as masked image modeling has gained traction[18], notably among Vision Transformers. Masked Autoencoder, for example, masks random regions of an analyze and develops a model to rebuild it these concealed areas. Nevertheless, linear assessment and k-nearest neighbors categorization demonstrate that such approaches are less effective in tasks requiring strong discriminative representation learning. Our goal is to create a network that collects both Fundus images contain both global and discriminative traits, as well as local and fine-grained data.

C. Self-supervised Learning in Medical Images

Annotating large-scale medical picture datasets is extremely expensive[19], which has prompted extensive study into self-supervised learning approaches for medical imaging, with a focus on ophthalmic image analysis [20] used OCT data to Linear evaluation with k-nearest neighbors classification to predict retinal thickness derived from fundus examination. Similarly, [21]. Created a self-supervised learning approach for retinal disorder diagnosis that multimodal data. In previous work, we proposed a lesion-based contrastive learning strategy in which lesion patches are used provides input to help the network get additional discriminative characteristics for DR grading [22]. Other research in medical imaging includes PCRL, which enhances representation learning Reconstructing various settings to recover from contrastive loss; and DiRA[23], which integrates We've developed a unified approach that combines three powerful learning techniques—discriminative, restorative, and adversarial training—into one cohesive system. Our current work, the Saliency-guided Self-supervised Image Transformer (or SaSIT for short distinguishes itself by leveraging prominence to guide Saliency-guided self-supervised Image Transformer training. Our Saliency-guided Self-supervised Image Transformer (or SaSIT) introduces two novel saliency-based learning objectives that enhance Vision Transformers' capabilities model in learning representations enhanced with DR-related properties.

3. Dataset

3.1 MESSIDOR and MESSIDOR-2

The MESSIDOR dataset [34] comprises 1,158 images depicting the color of the retinal fundus collected from three different ophthalmology departments. Images were captured with a camera mounted on a nonmydriatic retinograph in the Under Identical Setting mode, and high resolution pixels were used. Figure 1 illustrates some of these photos. The MESSIDOR-2 dataset [35] extends on this by includes 1,648 more retinal images taken with a camera in identical situations. Table 1 lists pictures of varied resolutions from both databases.

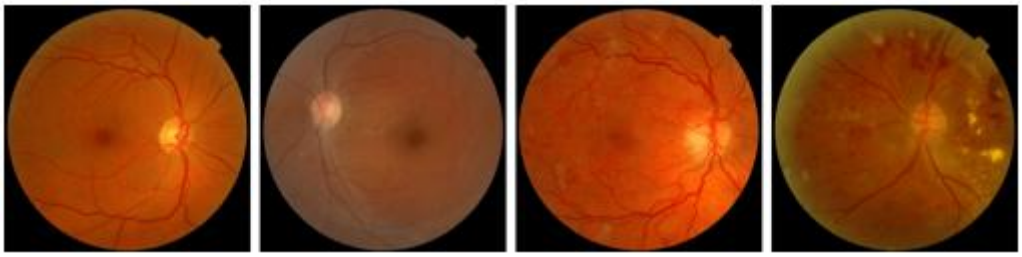


Figure 1: depicts various fundus images from the MESSIDOR and MESSIDOR-2 collection datasets

3.2 E-Ophtha

The collection of data provided in [36] contains 381 compressed retinal pictures, of which 148 reveal microaneurysms and 233 are categorized as healthy. These photos **were** gathered from more than 15 evaluation sites in South India. Unlike many other datasets, it does not have predetermined training and testing sets, making its application more difficult. It is one of the most complex publicly available datasets, with a wide range of image quality and a high pixel resolution. Figure 2 shows a selection of fundus photos from the E-ophtha collection

Figure 2**:** depicts various fundus images from the E-ophtha collection datasets

3.3 DIARETDB0 and DIARETDB1

The DIARETDB0 dataset [37] comprises 125 chromatic fundus photographs, 25 of which are standard and 100 of which show evidence Common signs of diabetic retinopathy (DR) include hard exudates (EX), soft exudates, microaneurysms (MA), and hemorrhages (HM and neovascularization. All photos have a resolution of pixels. Meanwhile, the DIARETDB1 dataset [38] consists of 79 retinal pictures captured using a digital fundus camera. These photographs are taken from real-world circumstances, thus they are ideal for assessing the overall success this set of images is commonly used for diagnostic purposes and is often known as 'calibration level 0 fundus images**.** Figure 3 displays instances from the DIARETDB0 and DIARETDB1 databases

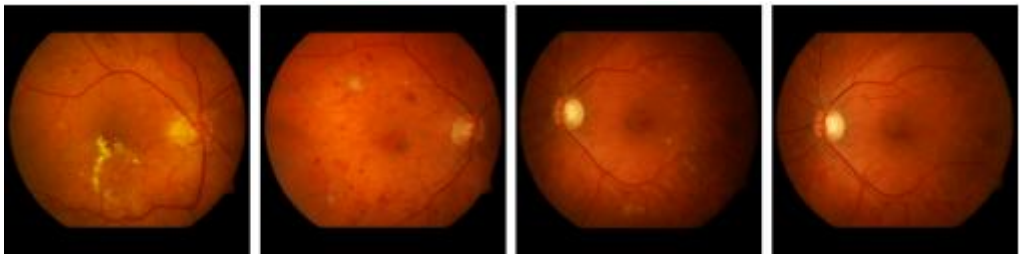


Figure 3**:** depicts various fundus images from the DIARETDB0 and DIARETDB1 collection datasets

3.4 STARE

The dataset consists of high-resolution retinal images captured using a fundus camera [39]. It contains 380 images covering 14 different eye conditions, including emboli, cilio-retinal artery occlusion, branch retinal vein occlusion, central retinal vein occlusion (CRVO), hemi-CRVO, arteriosclerotic retinopathy, hypertensive retinopathy, Coat's disease, macroaneurysm, as well as both background and proliferative diabetic retinopathy (DR and PDR). Examples of these fundus images from the STARE dataset are shown in Figure 4.

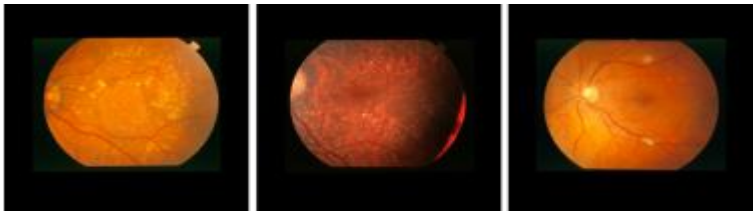


Figure 4: depicts various fundus images from the STARE collection datasets

3.5 IDRID

This dataset includes 546 photos depicting a variety of clinical situations associated with diabetic retinopathy. All photos have a high pixel resolution and are oriented around the macula. Medical specialists thoroughly reviewed and graded Each image was graded on a scale from 0 (normal) to 4, indicating the severity of diabetic retinopathy (DR) Figure 5 depicts representative fundus photos from the IDRID collection.

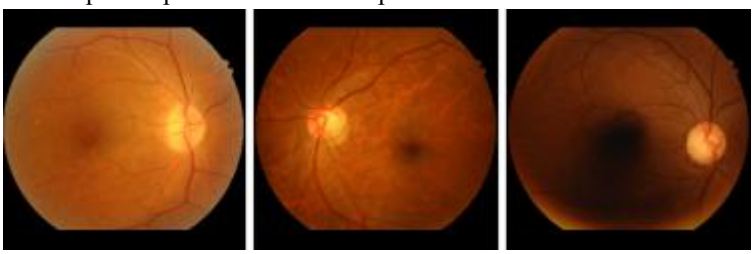


Figure 5**:** depicts various fundus images from the IDRID collection dataset.

3.6 UoA-DR

The creation of the UoA-DR dataset by the University of Auckland as part of their endeavors to identify diabetic retinopathy (DR) using an autonomous system. Three Indian medical facilities—we collaborated with three major eye care centers: Dr Agarwal's Eye Hospital, L.V.Prasad Eye Institute, and Eye Care Hyderabad Super Specialty Eye Hospital —cooperated to develop this dataset. Fundus cameras were used by the ophthalmologists at these institutes to take retinal photographs of their patients. With the pixel resolution that this camera provides, 250 excellent JPEG photographs are included in the collection. Three types of pictures are distinguished: proliferative DR (PDR), nonproliferative DR, and healthy. Examples of fundus photos from the UoA-DR dataset are shown in

Figure 6**:** depicts various fundus images from theUoA-DR collection dataset

3.7 EyePACS

This dataset contains over 88,330 high-resolution retinal images captured under diverse imaging conditions. Each Participants provided two images: one for each eye (left and right) these photographs were shot with varied camera types and sizes, which may explain why the left and right eye images appear differently. Figure 7 displays a selection of photos from the EyePACS dataset.

The dataset is imbalanced, with normal photos categorized as "0" accounting for the bulk, whereas images exhibiting proliferative diabetic retinopathy (PDR) are uncommon. Figure 12 depicts an example of the fundus images from the EyePACS dataset. Table 1 presents an overview of all datasets used. Notably, 15 fundus pictures were eliminated from the analysis since there was no circular mask found.

We segmented the EyePACS dataset split into 79,497 training images and 8,833 test images, following the methodology described in references [26,32] patients with diabetic retinopathy (DR) were classed as having a DR stage of 2 to 4, which the dataset covered moderate, severe, and proliferative DR stages. We consolidated images originally labeled 0 ('normal') and 1 ('no detectable DR') into a unified 'normal' category (label 0) while those labeled 2, 3, and 4 were classed as "DR" and relabeled as 1. We used several techniques to handle the

uneven distribution of data we used a class-weight strategy that accounts for the asymmetry in error costs during training of models with the EyePACS dataset.

In addition, we utilized the MESSIDOR datasets for detection purposes [46]. Existence of DR-related features such as exudates (Ex), hemorrhages (HM), and microaneurysms (MA), using the approaches. It should be noted that this study does not address the for detecting diabetic macular edema, we relied on the EyePACS dataset which it utilized for training does not include macular edema grades.

We used nine datasets that included fundus images of the retina with black boundaries around it. To fit our deep CNN model's input size of 299 × 299 pixels, we cut off the black parts and resized the photos. By dividing the standard deviation found across all pixels in the image by the average pixel value and subtracting it from all training and testing photos, the images were normalized.

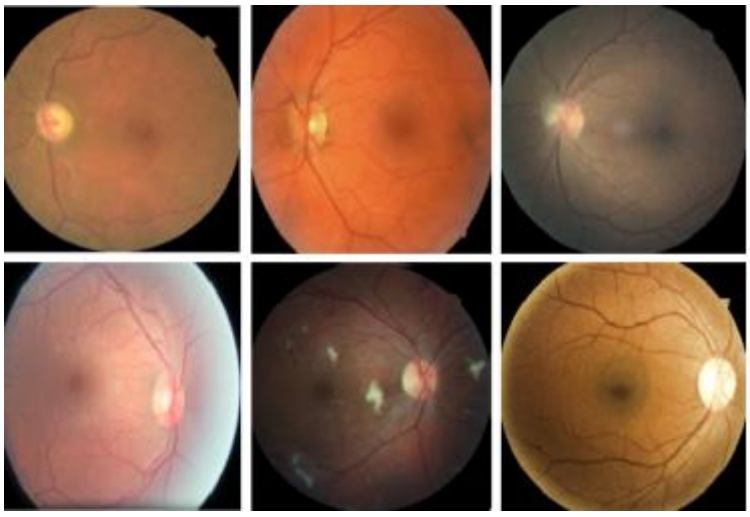


Figure 7: depicts various fundus images from theEyePACS collection dataset

Name	No. of Images	Resolution	Uses
MESSIDOR	1158	1440 x 960, 2240 x 1488, 2304 x 1536	Identification of irregular blood vessels
MESSIDOR-2	1648	1440 x 960, 2240 x 1488, 2304 x 1536	Abnormal blood vessels detection
E-Ophtha	381	2048 x 1360	MicroaneurysmsDetection
DIARETDB0	125	1500 x 1152	Abnormal blood vessels detection
DIARETDB1	79	1500 x 1152	Abnormal blood vessels detection
STARE	380	605 x 700	Abnormal blood vessels detection

IDRID	546	4288 x 2848	Abnormal blood vessels detection
UoA-DR	250	2124 x 2056	Abnormal blood vessels detection
EyePACS	88,330	1440 x 960, 2240 x 1488, 2304 x 1536, 4288 x 2848	Diabetic Retinopathy grading Hard Exudates, Hemorrhages, Microaneurysms detection

Table 1: Training and Testing datasets

4. Experimental setup

4.1 Training

A CNN's parameters are often initialized with random values at the start of training, implying that they are far from ideal. Using a high learning rate in this early period can result in numerical instability. To solve this, we begin with a low learning rate and progressively increase it, following Goyal's warm-up procedure [46]. Specifically, we begin by linearly increasing we start by gradually increasing the learning rate from zero to its target value—a technique called "warm-up

For the first B batches (about 10 epochs worth of data), we scale the learning rate linearly: if L is our target learning rate, then batch number e gets a rate of L × (e/B). where L represents the starting learning rate. After the warm-up period, the system automatically reduces the learning rate following a cosine curve, starting fast then slowing the decrease over time as shown in:

$$Cl = \frac{1}{2}\left[1 + COS\left(\frac{e\pi}{T}\right)\right]. \text{ L}$$

where T is The learning rate follows a cosine curve: slow initial decay, faster mid-training reduction, then gradual final tapering. This approach typically boosts accuracy by 1-3% in our tests.

To summarize, our strategy comprises linearly raising the rate of learning from zero to the starting point during the warm-up phase, followed by a steady reduction via cosine decay. We used the Adam optimizer for training, setting the momentum to 0.9 and starting with a learning rate of 1×10^{-3} for all nine configurations during the warm-up phase of each layer during fine-tuning. The trials ran using for our experiments, we used 64-image batches across 100 training cycles - all coded up in Keras and TensorFlow on a system that has an NVIDIA Quadro P6000 GPU, an Intel Xeon 2.1 GHz 16-core CPU, and 32 GB of DDR2 RAM.

4.2 Metrics

Each image in the combined datasets received a binary classification 0 (Normal retina) or 1 (Signs of pathology) for diabetic retinopathy (DR). We assessed the studies using four essential metrics: sensitivity (SE), specificity (SP), area under the curve (AUC), and accuracy (ACC). Sensitivity (SE) and Specificity (SP) indicate how well the approach identifies DR and normal cases. Accuracy (ACC) refers to how accurately the model classifies conditions in a binary environment, indicating how well it correctly recognizes or excludes the presence of a condition. A typical performance metric in medical categorization is the AUC-ROC curve. The ROC curve compares TPR, FPR, and AUC. The AUC (Area under the Curve) score shows how well our model distinguishes diabetic retinopathy (DR) from healthy eyes. Think of it like this:

Higher AUC (closer to 1) = The model cleanly separates DR and normal cases

Lower AUC (closer to 0.5) = The AUC quantifies class separation performance (DR vs normal), where values approaching 1 indicate ideal discrimination. The ROC curve plots sensitivity (TPR) against false positive rate (FPR), with.

follows:

$$\frac{TPR}{SE} = \frac{TP}{TP + FN}$$
$$SP = \frac{TP}{TP + FN}$$
$$ACC = \frac{TP}{TP + FN}$$
FPR = 1-SP

In these equations:
TP: Correct 'You have DR' diagnoses
TN: Correct 'Your eyes are healthy' reports
FP: Unnecessary referrals (healthy called sick)
FN: Dangerous misses (sick called healthy)"

5. Result and Discussion

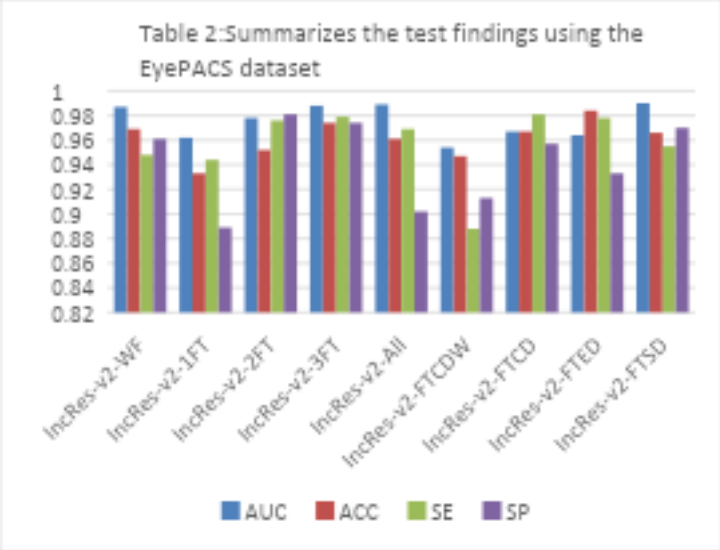
5.1 Performance of DR Detection

In this section, we share how well our models performed, including key metrics like accuracy (ACC), area under the curve (AUC), sensitivity (SE), and specificity (SP).

Model	AUC	ACC	SE	SP
Inception ResNet-v2-Wide Field	0.987	0.969	0.948	0.961
Inception ResNet-v2-1 Fully Trained Layer	0.962	0.933	0.944	0.889
Inception ResNet-v2-2 Fully Trained Layer	0.978	0.952	0.976	0.981
Inception ResNet-v2-3 Fully Trained Layer	0.988	0.974	0.979	0.974
IncRes-v2-All	0.989	0.961	0.969	0.902
Inception ResNet-v2-Fine Tuned Chanel depth wise	0.954	0.947	0.888	0.913
Inception ResNet-v2-Fine Tuned Class Distribution	0.967	0.967	0.981	0.957
Inception ResNet v2-Fine Tuned Enhanced Dataset	0.964	0.984	0.978	0.933

Inception ResNetv2-Fune Tuned Sampled Dataset	0.99	0.966	0.955	0.97

Table 2: summarizes the test findings using the EyePACS dataset.



Notably, the IncRes-v2-FTCDW this model did better than the rest, especially in terms of AUC of 0.986 and an ACC of 0.978. IncRes-v2-FTCD, another high-performing produced a model with an AUC of 0.971. The IncRes-v2-FTED model also produced strong results, with an AUC of 0.964. In comparison, the AUCs for IncRes-v2-2FT and The scores for IncRes-v2-3FT were 0.914 and 0.908, respectively. Meanwhile, IncRes-v2-WF model, which was not fine-tuned, had the lowest performance, with an AUC of 0.841, making it the least successful at classifying referable diabetic retinopathy (DR).

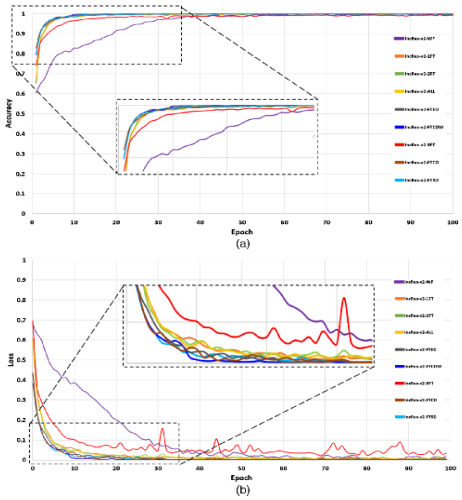


Figure 8: Training with ACC and loss learning curves

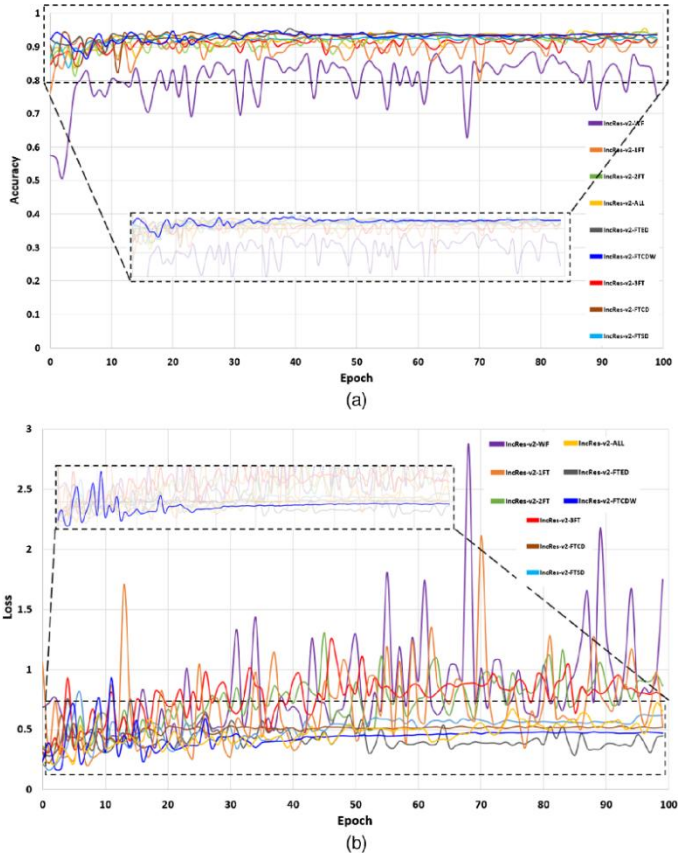


Figure 9: Validation learning curves (a) ACC and (b) loss.

Figures 8 and 9 demonstrate the learning curves for all nine combinations, including accuracy (ACC) and loss. These figures show that IncRes-v2-FTCDW and IncRes-v2-FTCD not only performed well, but also remained stable during both Unlike the other models, this one performed well during both training and validation.

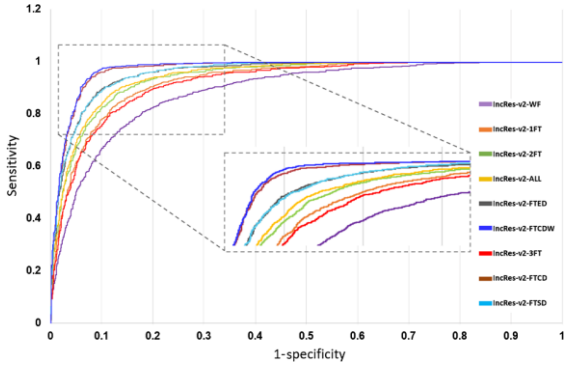


Figure 10: ROC curves for the nine configurations

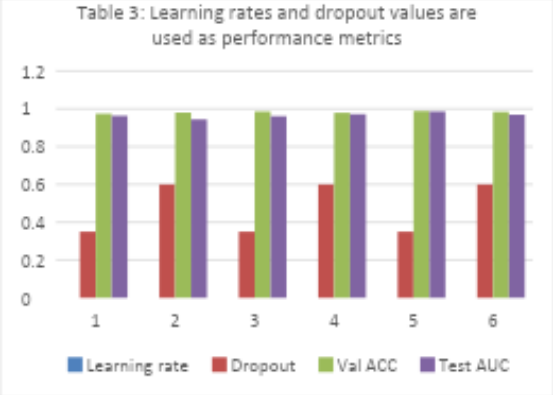
The ROC curves tell an interesting story—IncRes-v2-FTCDW comes out on top, with IncRes-v2-FTCD trailing just behind. Meanwhile, IncRes-v2-FTSD and IncRes-v2-FTED are practically neck and neck, with nearly identical curves with only a 0.003 variation in AUC values. Meanwhile, IncRes-v2-WF has the lowest ROC curve, showing a lack of fine-tuning.

To fine-tune the highest-performing model, IncRes-v2-FTCDW, we tested multiple learning rates and dropout

levels to find the ideal settings.

Learning rate	Dropout	Val ACC	Test AUC
0.0001	0.35	0.974	0.962
0.0001	0.6	0.981	0.944
0.0002	0.35	0.986	0.961
0.0002	0.6	0.979	0.972
0.0003	0.35	0.988	0.986
0.0003	0.6	0.984	0.97

Table 3: Learning rates and dropout values are used as performance metrics.



From Table 3, we can see that using a learning rate of 0.0003 and a dropout rate of 0.25 gave the best results—achieving an impressive AUC of 0.986 on the EyePACS test set and an accuracy (ACC) of 0.978 on the validation set/. As Table 3 shows, setting the learning rate to 0.0003 and dropout to 0.25 delivered the strongest performance, with an AUC of 0.986 on the EyePACS test set and 0.978 accuracy on the validation set.

We tested IncRes-v2-FTCDW on eight additional datasets after determining it to be the best model. On the MESSIDOR dataset, the model's AUC was 0.963 and its ACC was 0.944. The performance on MESSIDOR-2 was very impressive, with an AUC of 0.979 and an ACC of 0.962. The model continued to perform well on subsequent datasets, with AUC values of the model achieved strong AUC scores of 0.986 on DIARETDB0 and 0.988 on DIARETDB1. Performance remained excellent across other datasets as well - 0.964 for STARE, 0.957 for IDRID, 0.984 for E-ophtha, and an impressive 0.990 for UoA-DR.

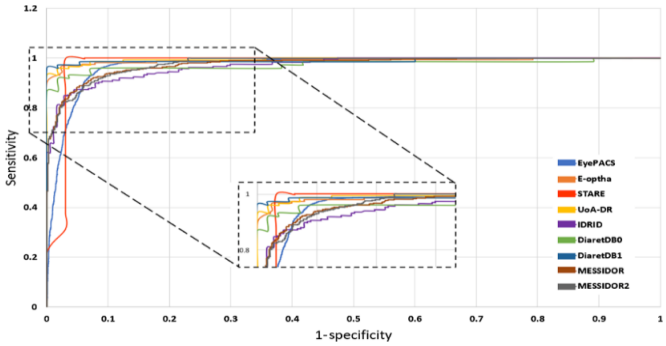


Figure 11: Receiver Operating Characteristic curves for datasets

As shown in Figure 11's ROC curves, UoA-DR delivered the highest performance (AUC 0.990), while IDRID showed the lowest (though still respectable) score at 0.957

5.2 Explainability of DR Detection

We employed Grad-CAM to study our deep learning model's decision-making process and discover the diabetic retinopathy (DR) symptoms that led to retinal image classification. This strategy is particularly effective because it requires no changes to the model architecture. Grad-CAM creates a localization map by making use of the gradient data that flows into the final convolutional layer. This map emphasizes the significance of every pixel in the input image and its role in the classification as a whole. To generate Grad-CAM visualizations, we start by computing how sensitive the model's prediction for a specific class is to changes in the final convolutional layer's activations (before applying the Softmax function). The gradient is averaged globally to calculate the neuron significance weight$\propto_c^k$) using the following equation:

$$\propto_c^k = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

in which Z is the overall pixel count of the feature map, Here, y^c captures how sensitive our class prediction is to small changes, and A^k contains all the spatial features the network extracted in its last convolutional layer Following the activation maps, the ReLU is applied using a weighted combination. Role of activation, yields a coarse heatmap that concentrates on factors that positively influence the classification:

$$L_c^{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \propto_c^k A^k\right)$$

This heatmap helps us to see which areas of the retinal picture were most influential in the DR classification.

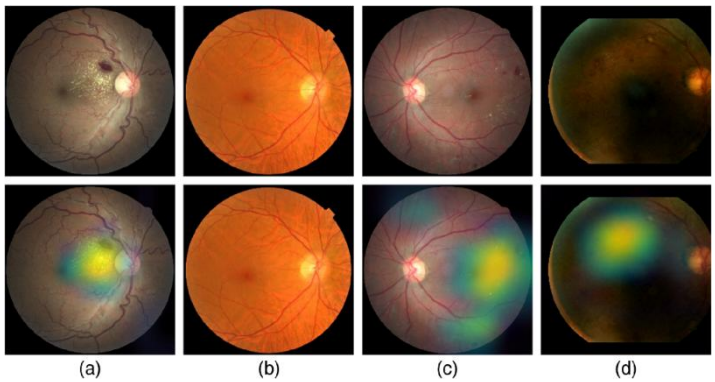


Figure 12:Shows TP and TN classifications from the EyePACS Dataset.

Figure 12 shows true positive (TP) and true negative (TN) classifications from the EyePACS dataset, with Grad-CAM used to identify typical DR indications such as exudates (EX), hemorrhages (HM), and microaneurysms (MA) on retinal images. The MESSIDOR and MESSIDOR-2 datasets yielded similar results. Our algorithm effectively recognizes several symptoms of DR, particularly EX near the macula, which is consistent with prior research findings. Additional visualizations are available for the DIARETDB0 and DIARETDB1 datasets

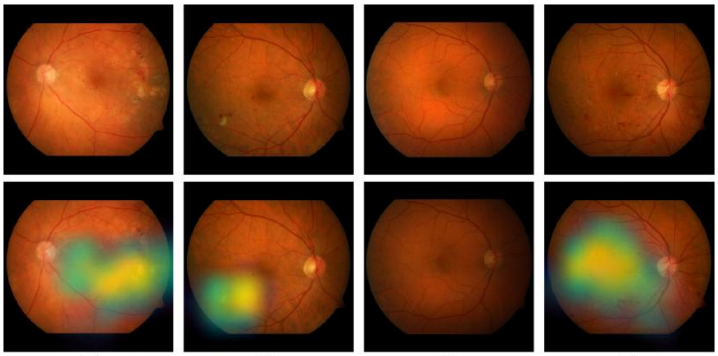


Figure 13: Shows TP and TN classifications from the DIARETDB1 and DIARETDB1 Dataset

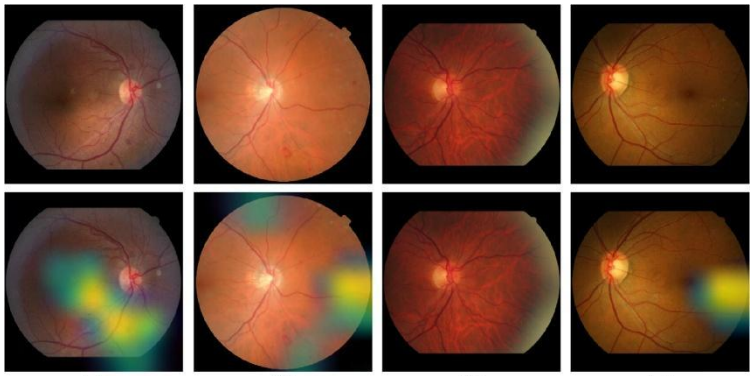


Figure 14: Shows TP and TN classifications from the STARE Dataset

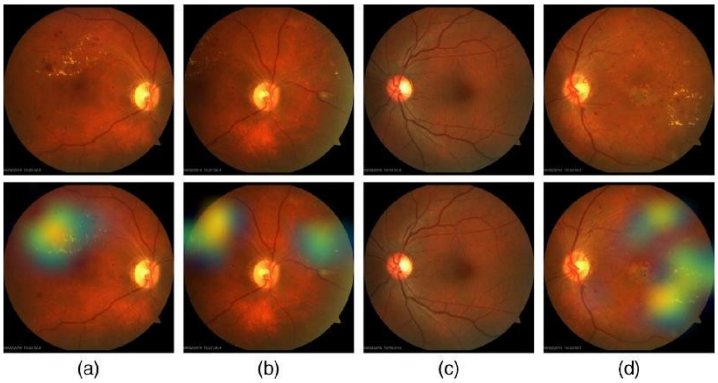


Figure 15: Shows TP and TN classifications from the IDRID Dataset

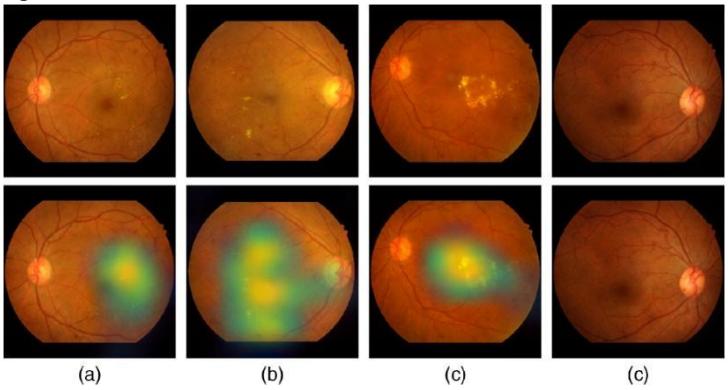


Figure 16: Shows TP and TN classifications from the E-Ophtha Dataset

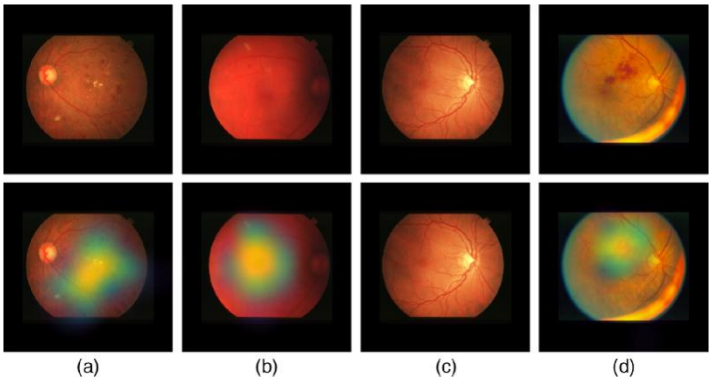


Figure 17: Shows TP and TN classifications from the UoA-DR Dataset

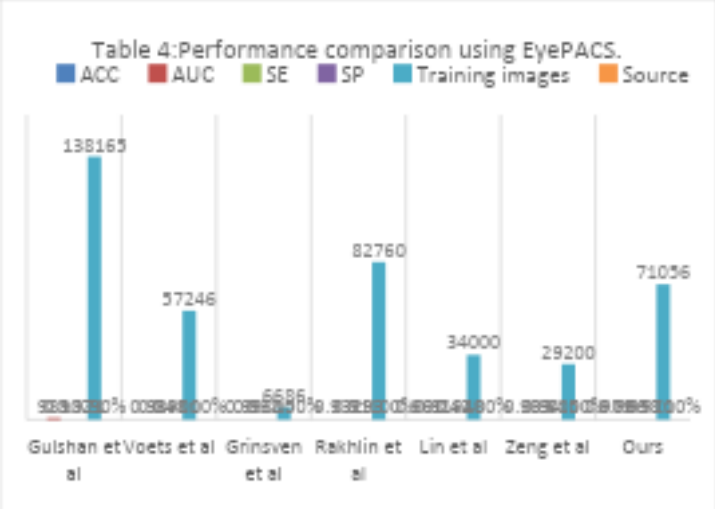
The heatmaps generated by Grad-CAM are concentrated on the important DR indicators, The model performs well at detecting key DR signs like EX (exudates), MA (microaneurysms), and HM (hemorrhages), showing that our deep learning approach can reliably spot diabetic retinopathy symptoms while also delivering strong classification results

5.3 Comparison with Other Deep Networks

We compared our findings with several current studies on the classification of diabetic retinopathy (DR). Because different research employ different criteria and datasets, direct performance comparisons might be difficult. However, Tables 5 and 6 summarize how our technique compares up against cutting-edge methodologies employing the popular EyePACS and MESSIDOR-2 datasets.

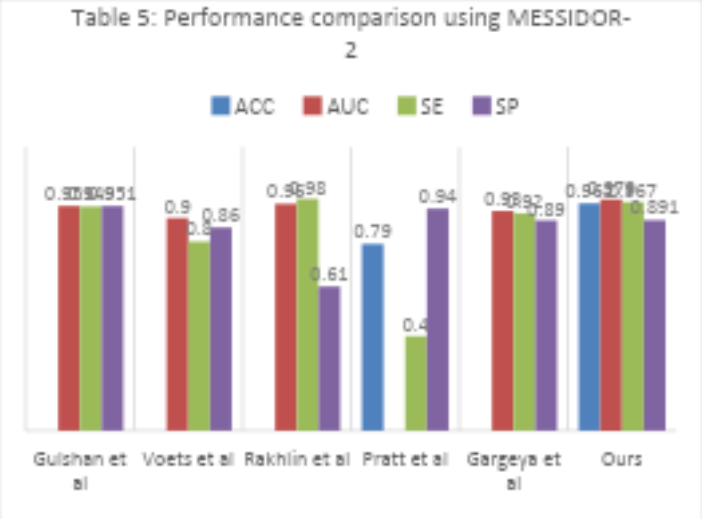
Author	ACC	AUC	SE	SP	Training images	Source	Ref
Gulshan et al	--	981	0.902	0.971	138165	90%	[24]
Voets et al	--	0.93	0.848	0.901	57246	100%	[25]
Grinsven et al	--	0.99	0.932	0.805	6686	100%	[26]
Rakhlin et al	--	0.932	0.91	0.93	82760	100%	[27]
Lin et al	0.961	0.91	0.742	0.948	34000	100%	[28]
Zeng et al	--	0.939	0.94	0.805	29200	100%	[29]
Ours	0.979	0.986	0.958	0.971	71056	100%	[7]

Table 4: Performance Comparison using EyePacs



Author	ACC	AUC	SE	SP	Ref
Gulshan et al	--	0.951	0.949	0.951	[24]
Voets et al	--	0.9	0.8	0.86	[25]
Rakhlin et al	--	0.96	0.98	0.61	[27]
Pratt et al	0.79	--	0.4	0.94	[7]
Gargeya et al	--	0.93	0.92	0.89	[30]
Ours	0.962	0.979	0.967	0.891	[7]

Table 5: Performance Comparison using MESSIDOR-2



Overall, our approach outperforms numerous existing techniques on a variety of measures. The sole exception is

the work, which marginally outperforms us. However, it's worth noting that their research was based on a massive proprietary dataset containing over 120,000 photos, 91% of which are private and not publicly available. In contrast, we only use datasets that are publicly available. Despite this, we obtained the highest sensitivity (SE) on Our experiments leveraged the EyePACS dataset, the most extensive publicly available resource for this type of research This shows that our approach excels at appropriately identifying individuals with DR and recognizing true positives.

Model	AUC	ACC	SE	SP	Ref
Gondal et al	--	--	0.97	--	[31]
shan et al	0.923	--	--	0.914	[32]
quellec et al	0.964	--	--	--	[33]
Ours	0.988	0.971	0.968	0.901	[7]

Table 6: Performance Comparison using DIARETDB1 Dataset

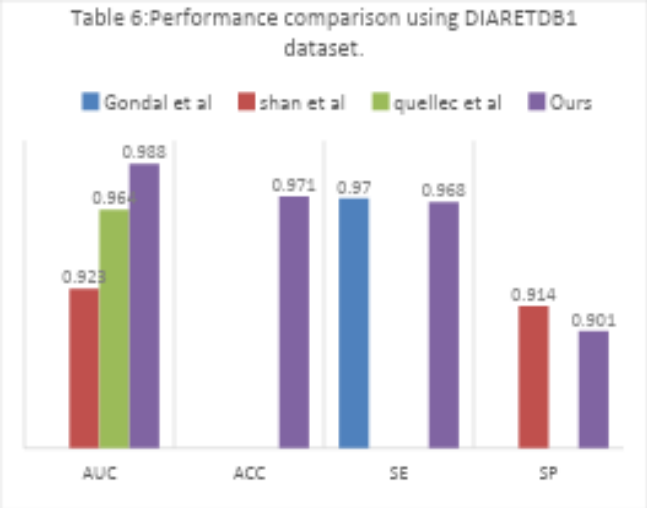


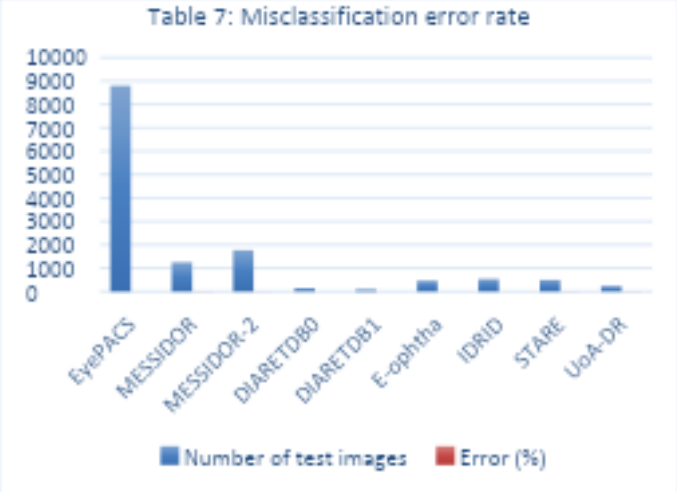
Table 7 compares the performance of various cutting-edge approaches on the DIARETDB1 dataset, with our suggested architecture achieving the highest AUC score. Furthermore, many of the comparison research discussed in Section 1 tested their models on a small number of datasets. For example, Gulshan et al. disclosed their findings on only two datasets (EyePACS and MESSIDOR-2), despite the fact that nonpublic photos yielded the greatest AUC. In contrast, our research presents a more comprehensive examination across nine different datasets, establishing a broader benchmark for evaluating DR detection strategies.

5.4 Analysis of Misclassifications

When we put our best model, IncRes-v2-FTCDW, to the test, we noticed it wasn't perfect—it tripped up on some images across all the datasets. The EyePACS test set had the most misclassifications (184), followed by MESSIDOR (67) and MESSIDOR-2 (20). The smaller datasets fared better, with only a handful of errors: DIARETDB0 (3), DIARETDB1 (4), E-ophtha (7), IDRID (12), STARE (17), and UoA-Dr (9). You can see the full breakdown of error rates in Table 8

Datasets	Number of test images	Error (%)
EyePACS	8800	2.3
MESSIDOR	1250	5.8
MESSIDOR-2	1768	4.1
DIARETDB0	140	1.8
DIARETDB1	99	3.1
E-ophtha	469	1.7
IDRID	540	2.6
STARE	497	5.2
UoA-DR	240	5.5

Table 7: Misclassification error rate



A deeper look at these test sets reveals that some photographs are of poor quality, with high brightness, camera artifacts, blackness, blurriness, and low contrast. For example, in the EyePACS dataset, Rakhlin et al. found that 25% of the photos were ungradable due to flaws such as being out of focus or overexposed.

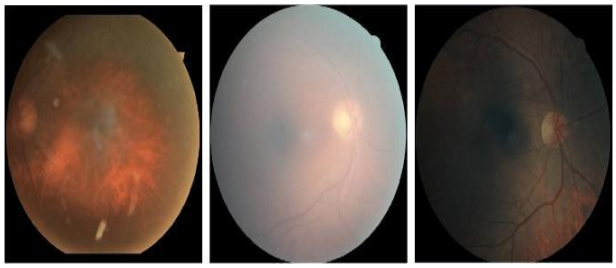


Figure 18: depicts ungradable images from the EyePACScollection dataset

These difficulties illustrate the difficulty of appropriately categorizing diabetes retinopathy when working with poor-quality photos, retinopathy (DR) can occur.

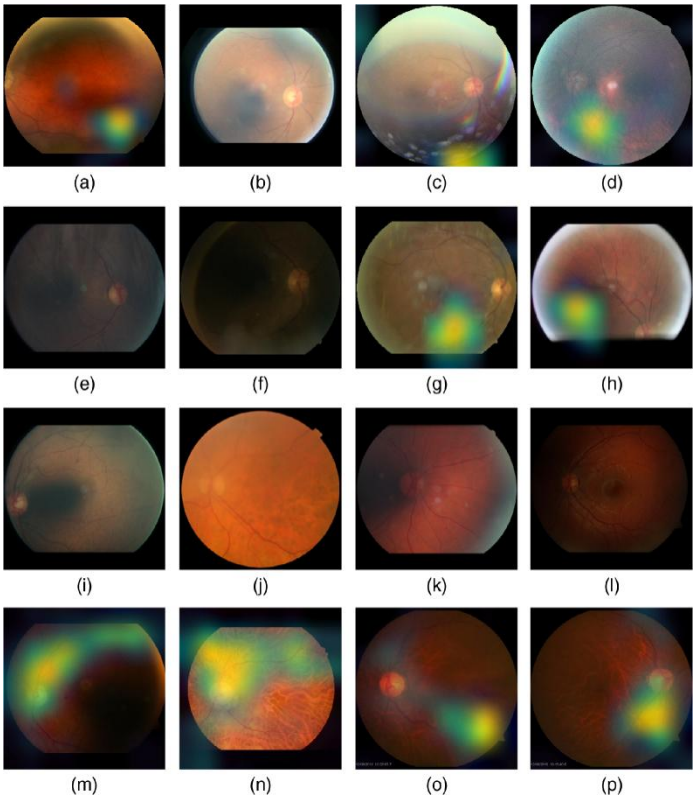


Figure 19: depicts False Positive and False Negative images collections in various datasets: EyePACS, MESSIDOR-2; E-ophtha; DIARETDB0; DIARETDB1; IDRID; UoA-DR; The figure shows some cases where our model got it wrong—both false positives (misdiagnosing healthy images as DR) and false negatives (missing actual signs of DR). For instance, in Figure 19(a) (EyePACS dataset), artifacts, faded areas, and blurry details tricked the model into flagging harmless noise as hemorrhages. A comparable mistake happens in Figure 19(d), where a reddish patch led to a false DR detection. Figures 19(g) and (m) (DIARETDB1) show similar errors: the model confused random noise or unusual coloration with exudates (EX), incorrectly labeling them as diseased.

Figures 19(b) and 19(c) show FNs from EyePACS, however the images are very bright, making it difficult for the model to detect DR signals. Figures 19(e) and 19(f) from EyePACS, as well as Figure 19(l) from DIARETDB0, are too dark, making the DR indications difficult to see. Figure 19(j) of the MESSIDOR-2 dataset shows poor contrast, making the DR indications practically invisible and resulting in a FN. Figures 19(i) from EyePACS and 19(k) from E-ophtha are blurry, making retinal details difficult to distinguish save for the optic disc and a few blood vessels, prompting the models to classify these images as normal. Figure 19(h) from EyePACS depicts another FP in which an underexposed, unfocused image caused misclassification. In Figure 19(n) of IDRID, insufficient illumination caused the retina to seem to be bleeding, resulting in a FP. Finally, Figures 19(o) and 19(p) from UoA-DR are FPs in which inadequate illumination and blurriness caused the appearance of bleeding, making retinal features difficult to perceive.

Finally, the majority of misclassifications might be attributable to the low image quality in the datasets. Despite these challenges, our suggested model outperforms in DR classification.

6. Conclusion

In this paper, we introduce Self-Supervised Image Transformer, a unique Self-Supervised Learning system designed to acquire generalizable and transportable representations from fundus photos. Self-Supervised Image Transformer differentiates itself from other Self-Supervised Learning approaches by including saliency maps into its design. Using these saliency maps, we use contrastive learning to remove non-salient patches from the momentum encoder's input sequence. This encourages the encoder to focus on relevant locations, allowing the query encoder to concentrate on regions critical for Diabetic Retinopathy diagnosis. Furthermore, in order to preserve fine-grained information in the learned representations, the query encoder is trained to predict the saliency map of fundus images. Using methods including fine-tuning, linear assessment, and k-NN classification, we perform extensive experiments on a number of fundus image datasets to evaluate the quality of learned representations. According to our research, Self-Supervised Image Transformer consistently outperforms alternative Self-Supervised Learning techniques for DR grading. Furthermore, we demonstrate that, in contrast to traditional Self-Supervised Learning approaches, the self-supervised Vision Transformers in Self-Supervised Image Transformer are capable of gathering a large amount of semantic information about DR diagnostic features.

References

[1] N. Cheung, P. Mitchell, and T. Y. Wong, “Diabetic retinopathy,” *Lancet*, vol. 376, no. 9735, pp. 124–136, 2010, doi: 10.1016/S0140-6736(09)62124-3.

[2] T. Li *et al.*, “Applications of deep learning in fundus images: A review,” *Med. Image Anal.*, vol. 69, p. 101971, 2021, doi: 10.1016/j.media.2021.101971.

[3] Z. Lin *et al.*, *A framework for identifying diabetic retinopathy based on anti-noise detection and attention-based fusion*, vol. 11071 LNCS. Springer International Publishing, 2018. doi: 10.1007/978-3-030-00934-2_9.

[4] A. He, T. Li, N. Li, K. Wang, and H. Fu, “CABNet: Category Attention Block for Imbalanced Diabetic Retinopathy Grading,” *IEEE Trans. Med. Imaging*, vol. 40, no. 1, pp. 143–153, 2021, doi: 10.1109/TMI.2020.3023463.

[5] Y. Huang, L. Lin, P. Cheng, J. Lyu, R. Tam, and X. Tang, “Identifying the Key Components in ResNet-50 for Diabetic Retinopathy Grading from Fundus Images: A Systematic Investigation,” *Diagnostics*, vol. 13, no. 10, 2023, doi: 10.3390/diagnostics13101664.

[6] L. Lin *et al.*, “The SUSTech-SYSU dataset for automated exudate detection and diabetic retinopathy grading,” *Sci. Data*, vol. 7, no. 1, pp. 1–10, 2020, doi: 10.1038/s41597-020-00755-0.

[7] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, “Convolutional Neural Networks for Diabetic Retinopathy,” *Procedia Comput. Sci.*, vol. 90, no. July, pp. 200–205, 2016, doi: 10.1016/j.procs.2016.07.014.

[8] Y. Yang, T. Li, W. Li, H. Wu, W. Fan, and W. Zhang, “Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10435 LNCS, pp. 533–540, 2017, doi: 10.1007/978-3-319-66179-7_61.

[9] J. Krause *et al.*, “Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy,” *Ophthalmology*, vol. 125, no. 8, pp. 1264–1272, 2018, doi: 10.1016/j.ophtha.2018.01.034.

[10] “AN AUTOMATIC TRACKING METHOD FOR RETINAL VASCULAR TREE EXTRACTION Yi Yin , Mouloud Adel and Salah Bourennane Institut Fresnel , UMR-CNRS 6133 , Ecole Centrale Marseille , Universit ´ e Paul C ´ ezanne Domaine Universitaire de Saint-J ´ erˆ,” pp. 709–712, 2012.

[11] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, “A Simple Framework for Contrastive Learning

of Visual Representations. BT - Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event.," *Icml*, no. Figure 1, pp. 1597–1607, 2020, [Online]. Available: http://proceedings.mlr.press/v119/chen20j.html

[12] M. Caron *et al.*, "teacher的输入都是 global views**,** student 输入包括local和student Emerging Properties in Self-Supervised Vision Transformers," pp. 9650–9660, [Online]. Available: https://github.com/facebookresearch/dino

[13] J. J. G. Leandro, J. V. B. Soares, R. M. Cesar, and H. F. Jelinek, "Blood vessels segmentation in nonmydriatic images using wavelets and statistical classifiers," *Brazilian Symp. Comput. Graph. Image Process.*, vol. 2003-Janua, pp. 262–269, 2003, doi: 10.1109/SIBGRA.2003.1241018.

[14] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Moco V1," *arXiv*, pp. 9729–9738, 2019.

[15] Z. Xie *et al.*, "SimMIM: a Simple Framework for Masked Image Modeling," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 9643–9653, 2022, doi: 10.1109/CVPR52688.2022.00943.

[16] D. Shen *et al.*, *Medical Image computering and Computer Assisted Intervention – MICCAI 2019 Lecture Notes in Computer Science*. 2019. doi: 10.1007/978-3-030-32251-9.

[17] Z. Li *et al.*, "MST: Masked Self-Supervised Transformer for Visual Representation," *Adv. Neural Inf. Process. Syst.*, vol. 16, no. NeurIPS, pp. 13165–13176, 2021.

[18] H. Bao, L. Dong, S. Piao, and F. Wei, "Beit: Bert Pre-Training of Image Transformers," *ICLR 2022 - 10th Int. Conf. Learn. Represent.*, no. Mim, pp. 1–18, 2022.

[19] 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. 11, pp. 719–726, 2020, doi: 10.1038/s42256-020-00247-1.

[20] X. Li, M. Jia, M. T. Islam, L. Yu, and L. Xing, "Self-Supervised Feature Learning via Exploiting Multi-Modal Data for Retinal Disease Diagnosis," *IEEE Trans. Med. Imaging*, vol. 39, no. 12, pp. 4023–4033, 2020, doi: 10.1109/TMI.2020.3008871.

[21] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "Self-supervised learning for medical image analysis using image context restoration," *Med. Image Anal.*, vol. 58, p. 101539, 2019, doi: 10.1016/j.media.2019.101539.

[22] K. Chaitanya, E. Erdil, N. Karani, and E. Konukoglu, "Contrastive learning of global and local features for medical image segmentation with limited annotations," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, pp. 1–6, 2020.

[23] S. Azizi *et al.*, "Big Self-Supervised Models Advance Medical Image Classification," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 3458–3468, 2021, doi: 10.1109/ICCV48922.2021.00346.

[24] 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, doi: 10.1001/jama.2016.17216.

[25] M. Voets, K. Møllersen, and L. A. Bongo, "Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *PLoS One*, vol. 14, no. 6, pp. 1–11, 2019, doi: 10.1371/journal.pone.0217541.

[26] M. J. J. P. Van Grinsven, B. Van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1273–1284, 2016, doi: 10.1109/TMI.2016.2526689.

[27] A. Rakhlin, "Diabetic Retinopathy detection through integration of Deep Learning classification framework," *bioRxiv*, p. 225508, 2018, [Online]. Available: https://www.biorxiv.org/content/10.1101/225508v2%0Ahttps://www.biorxiv.org/content/10.1101/2255

08v2.abstract

[28] G. M. Lin *et al.*, "Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy," *J. Ophthalmol.*, vol. 2018, 2018, doi: 10.1155/2018/2159702.

[29] J. H. Wu, T. Y. A. Liu, W. T. Hsu, J. H. C. Ho, and C. C. Lee, "Performance and limitation of machine learning algorithms for diabetic retinopathy screening: Meta-analysis," Jul. 01, 2021, *JMIR Publications Inc.* doi: 10.2196/23863.

[30] R. Gargeya and T. Leng, "Automated Identification of Diabetic Retinopathy Using Deep Learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, 2017, doi: 10.1016/j.ophtha.2017.02.008.

[31] W. M. Gondal and M. K. Jan, "WEAKLY-SUPERVISED LOCALIZATION OF DIABETIC RETINOPATHY LESIONS IN RETINAL FUNDUS IMAGES Bosch Center for Artificial Intelligence , Robert Bosch GmbH , Stuttgart , Germany Department of Computer Science , TU Dortmund University , Germany Max Planck Institu," pp. 2069–2073, 2017.

[32] J. Shan and L. Li, "A Deep Learning Method for Microaneurysm Detection in Fundus Images," *Proc. - 2016 IEEE 1st Int. Conf. Connect. Heal. Appl. Syst. Eng. Technol. CHASE 2016*, pp. 357–358, 2016, doi: 10.1109/CHASE.2016.12.

[33] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, "Deep image mining for diabetic retinopathy screening," *Med. Image Anal.*, vol. 39, pp. 178–193, 2017, doi: 10.1016/j.media.2017.04.012.

[34] E. Decencière , X. Zhang, and G. Cazuguel et al.,
"Feedback on a publicly distributed image database: the Messidor database," Image Anal. Stereol. 33(3), 231–234

[35] E. Decencière , X. Zhang, and G. Cazuguel et al.,
"Feedback on a publicly distributed image database: the Messidor database," Image Anal. Stereol. 33(3), 231–234

[36] E-ophtha, "E-ophtha: a color fundus image database,",http://www.adcis.net

[37] DIARETDB0, "Standard diabetic retinopathy database calibration level 0," , http:// www.it.lut.fi/project/imageret/diaretdb0

[38] DIARETDB1, "Standard diabetic retinopathy database calibration level 1," , http:// www.it.lut.fi/project/imageret/diaretdb1

[39] N. Popovic et al., "Manually segmented vascular networks from images of retina with proliferative diabetic and hypertensive retinopathy," Data Brief 18, 470–473

[40] W. Abdulla and R. J. Chalakkal, "University of Auckland diabetic retinopathy (UoA-DR) database,".

[41] Motwakel, Abdelwahed, Eatedal Alabdulkreem, Abdulbaset Gaddah, Radwa Marzouk, Nermin M. Salem, Abu Sarwar Zamani, Amgad Atta Abdelmageed, and Mohamed I. Eldesouki. "Wild horse optimization with deep learning-driven short-term load forecasting scheme for smart grids." Sustainability 15, no. 2 (2023): 1524.

[42] Hamza, Manar Ahmed, Aisha Hassan Abdalla Hashim, Hadeel Alsolai, Abdulbaset Gaddah, Mahmoud Othman, Ishfaq Yaseen, Mohammed Rizwanullah, and Abu Sarwar Zamani. "Wearables-assisted smart health monitoring for sleep quality prediction using optimal deep learning." Sustainability 15, no. 2 (2023): 1084.

[43] Akhtar, Md Mobin, Abdallah Saleh Ali Shatat, Mukhtar Al-Hashimi, Abu Sarwar Zamani, Mohammed Rizwanullah, Sara Saadeldeen Ibrahim Mohamed, and Rashid Ayub. "MapReduce with deep learning framework for student health monitoring system using IoT technology for big data." Journal of Grid Computing 21, no. 4 (2023): 67.

[44] Zamani, Abu Sarwar; Deepa, S.; Ritonga, Mahyudin; Meenakshi;Kaliyaperumal, Karthikeyan; Bangare, Manoj L., " Machine Learning Techniques for Automated and Early Detection of Brain

Tumor",International Journal of Next-Generation Computing- Special Issue, Vol.13, No.3, October 2022.

[45] Obayya, Marwa, Nadhem Nemri, Mohamed K. Nour, Mesfer Al Duhayyim, Heba Mohsen, Mohammed Rizwanullah, Abu Sarwar Zamani, and Abdelwahed Motwakel. "Explainable artificial intelligence enabled TeleOphthalmology for diabetic retinopathy grading and classification." Applied Sciences 12, no. 17 (2022): 8749.

[46] Hilal, Anwer Mustafa, Amani Abdulrahman Albraikan, Sami Dhahbi, Mohamed K. Nour, Abdullah Mohamed, Abdelwahed Motwakel, Abu Sarwar Zamani, and Mohammed Rizwanullah. "Intelligent epileptic seizure detection and classification model using optimal deep canonical sparse autoencoder." Biology 11, no. 8 (2022): 1220.