



# **iJRASET**

International Journal For Research in  
Applied Science and Engineering Technology



---

# **INTERNATIONAL JOURNAL FOR RESEARCH**

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 10    Issue: VI    Month of publication: June 2022**

**DOI: <https://doi.org/10.22214/ijraset.2022.44324>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Self-Supervised Learning for Medical Imaging

Dr. Bagam Laxmaiah<sup>1</sup>, Lekkala Harika Chowdary<sup>2</sup>, Koneru Pragna<sup>3</sup>, Nalla Bhavani<sup>4</sup>, A Kiran Kumar<sup>5</sup>

<sup>1</sup>Associate Professor, Dept of CSE, CMR Technical Campus, Hyderabad, Telangana, India

<sup>2, 3, 4</sup>Student, Dept of CSE, CMR Technical Campus, Hyderabad, Telangana, India

<sup>5</sup>Assistant Professor, Dept of CSE, CMR Technical Campus, Hyderabad, Telangana, India

**Abstract:** *The role of deep learning is growing quite effectively. As the created models are producing promising accuracy and the early detection and mitigation of diseases is becoming quite easy. As a result, the deep learning algorithm is receiving a variety of interest nowadays for fixing several problems within side the area of scientific imaging. In ophthalmology, one instance is detecting disorder or anomalies by using photos and classifying them into diverse disorder types or severity levels. This sort of project has been finished the use of quite a few machine learning algorithms which have been optimized, in addition to theoretical and empirical approaches. Diabetic Retinopathy is such a disease where in early detection plays a severe role as it could result in imaginative and prescient loss.*

*Diabetic Retinopathy disease recognition has been one of the active and challenging research areas in the field of image processing. Deep learning technique as well as hinders to work with disease recognition and find the accuracy of the model. To create a model in a supervised manner, we need a huge amount of dataset which is very costly. So, as to overcome this problem, we have implemented a self - supervised model for the detection of diabetic retinopathy which works with a very limited dataset. This model is implemented using one of the pretext/proxy task image rotations developed on Dense NET architecture. The model is fine-tuned with the various quantities of subsets of the original dataset and compared internally.*

**Keywords:** *Self-supervised model, pretext/proxy task, Dense NET architecture, fine-tuning*

## I. INTRODUCTION

The ophthalmology field has benefited from recent advances in deep learning, particularly in the case of deep convolutional neural networks (CNNs) when applied to large data sets, such as two-dimensional (2D) fundus photography, a low-key imaging technology that captures the back of the eye. Recently, self-supervised learning has achieved fantastic fulfillment within side the field of Computer Vision. Particularly, self-supervised learning can successfully serve the sector of medical imaging in which a big quantity of categorised facts is normally limited. The input data to the model is the diabetic retinopathy fundus images.

There are different types of proxy tasks. In this model we are using rotations as a proxy task. The model can predict the Diabetic Retinopathy[1] by self-supervised model and by using one pretext/proxy task image rotations developed on Dense NET architecture[2] and detect the different stages of it. To check various possibilities, we finetuned the model using different sets of data sizes i.e., 5%, 10%, 25%, 50% and 100% of the original dataset. With batch size of 32 with 5 repetitions and prediction architecture of simple multiclass. The model also finds the accuracy of the model and finds the value by using “Kappa-kaggle score”[3].

## II. LITERATURE SURVEY

Unsupervised learning in standard may be formulated as studying an embedding area, where the facts this is comparable semantically are nearer and vice versa. The self-supervised learning[4] does the identical through building such illustration area with the assist of proxy task[5] from the facts itself. The learning’s of version on the time of proxy task also can be utilized in numerous other downstream duties. Recently, numerous strategies within side the line of studies had been evolved and determined programs in several fields. Self-supervised learning includes predominant elements of processing first is the proxy task and 2nd is fine-tuning[6]. There are specific styles of self-supervised studying strategies that differ of their first constructing block i.e., proxy task. There are numerous styles of proxy duties evolved on this line of studies. Here we’re focusing in particular on photograph datasets.

Relative patch location[7]--It extracts random pairs of patches from each image and train a convolutional neural net to predict the position of the second patch relative to the first

Jigsaw Puzzles[8]-- The model solves the jigsaw puzzle as the proxy task which requires no manual labelling, and then later repurposed to solve object classification and detection.

Image color clustering[9]--The model is fed with grayscale image and asked to find the plausible colorization as the proxy task.

Image rotation prediction[10]-- It learns image features by training model to recognize the 2d rotation that is applied to the image that it gets as input.

Image Reconstruction[11]--A visual feature learning algorithm driven by context-based pixel prediction.

Object saliency[12]-- It learns background-agnostic representations by performing the salient object detection in a self-supervised manner.

Contrastive Prediction[13]-- The model learns representations by predicting the future in latent space by using powerful autoregressive models. It uses a probabilistic contrastive loss which induces the latent space to capture information that is maximally useful to predict future samples

Although all the techniques mentioned derives state of art results and getting better with time and research. This report deals with working of only one of the above techniques i.e., Image rotation prediction.

### III.PROPOSED SYSTEM

The model follows a self-supervised paradigm and proposes to learn image representations by training to recognize the geometric transformation that is applied to the image that it gets as input. The model is trained on the 4-way image classification task of recognizing one of the four image rotations i.e., 0, 90, 180, 270 degrees. More specifically, we first define a small set of discrete geometric transformations, then each of those geometric transformations are applied to each image on the dataset and the produced transformed images are fed to the model that is trained to recognize the transformation of each image.

To develop a self-supervised learning model for medical imaging, specifically for Diabetic Retinopathy detection. We employ the Rotation approach as a proxy task, and we use unlabeled data with various geometric progressions to train a Dense ConvNet model to predict geometric progression probability.

Then, as follows, fine-tune the model to determine the severity of the DR using some tagged data.

- 1) NO DR
- 2) MILD DR
- 3) MODERATE DR
- 4) SEVERE DR
- 5) PROLIFERATE DR

The model also finds the accuracy of the model and finds the value by using “Kappa-kaggle score”.

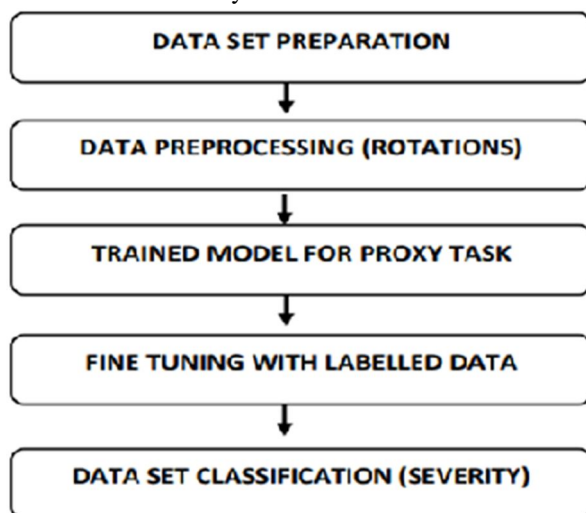


Figure 3.1: Architecture of the Model.

- *Dataset Preparation:* The data is obtained from Kaggle mentioned in Diabetic Retinopathy 2019 Kaggle challenge. X It contains images of retinal fundus resized into 224 x 224, categorized into five types, NO DR, mild, moderate, severe and proliferate. For the proxy task we combine all the types as it does not require any labelling. For finetuning, we use 5%, 10%, 25% and 50% of the original data set to check the efficiency at each specific size of data.

- **Data Pre-processing:** The data is resized into 244 x 244 and performed various geometric progressions i.e., rotations into multiples of 90 degrees, (0, 90, 180, 270 degrees). This pre-processed data is then fed to the ConvNET model with DenseNET121 encoder architecture. It was handled with learning rate of 1e-5, for 200 epochs with batch size of 32.
- **Fine-tuning with Labelled Data:** The ConvNET is trained to predict the geometric progressions of the image which is nothing short than learning characteristics of the image. It is utmost ready for classification but lacks knowledge of categorization, which can be achieved by finetuning the model with labelled data. To check various possibilities, we finetuned the model using different sets of data sizes i.e., 5%, 10%, 25%, 50% and 100% of the original dataset. With batch size of 32 with 5 repetitions and prediction architecture of simple multiclass.
- **Classification:** The final model is generated after the fine tuning which can classify or detect the Diabetic retinopathy stages. It is tested with “qw\_kappa\_kaggle” scores based on the accuracy and obtained very promising results in comparison to dataset size

#### IV.RESULT

Here is the sample code and result.

```

▶ ipython finetune.py self_supervised_3d_tasks/configs/finetune/rotation_2d.json
44/44 [-----] - 31s 71ms/step - loss: 0.3369 - accuracy: 0.8587 - val_loss: 0.5283
Epoch 17/17
44/44 [=====] - 31s 71ms/step - loss: 0.3369 - accuracy: 0.8587 - val_loss: 0.5283
qw_kappa_kaggle score: 0.7565178391296926
cat_acc_kaggle score: 0.638344262295082

-----
running test for: 25%
-----

Loading Test data
95.65%
['accuracy']
binary_crossentropy
Model: "model_1"

```

Fig 4.1: Code Execution

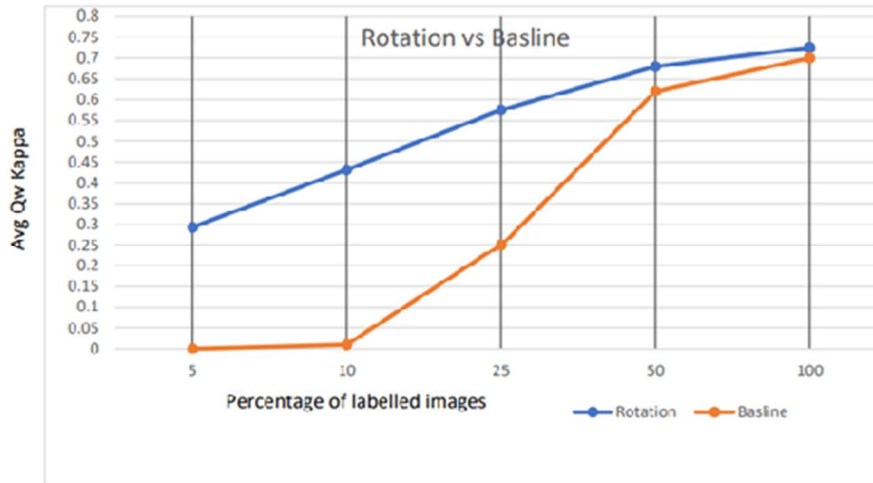
The final model recognizes and categorizes the retinal fundus data into one of the five kinds. The Kaggle dataset contains approximately 3600 photos, each of which has been scored on a scale of 0 to 4 by a clinician (NO DR, mild, moderate, severe, proliferate). To assess our performance on this benchmark, we pre-trained the model using all of the dataset's photos. Then they were fine-tuned on the same Kaggle data but with varied subset sizes, resulting in a data-efficient evaluation. When compared to other transfer learning methods that use a big corpus, the outcomes due to data efficient evaluation are not up to par. The dataset is being tested using 5-fold cross validation. The task's statistic is quadratic weighted kappa, which determines how well two ratings agree. Its values range from random (0) to total (1) agreement, and it can become negative if there is less agreement than chance.

| Train Split | Qw_kappa_kaggle MIN | Qw_kappa_kaggle AVG | Qw_kappa_kaggle MAX |
|-------------|---------------------|---------------------|---------------------|
| 10%         | 0.2888881102        | 0.4321430821        | 0.5084661884        |
| 5%          | 0.1751079345        | 0.2944362057        | 0.4669641719        |
| 50%         | 0.6116147969        | 0.6798179485        | 0.7365178391        |
| 25%         | 0.4738074393        | 0.5753686169        | 0.6937023326        |
| 100%        | 0.6955872731        | 0.7247486635        | 0.7555889924        |

Figure 4.4-Result of the final image



Avg Qw Kappa scores vs percentage of labeled images comparing rotation techniques and baseline values



## V. CONCLUSION

The final model we implemented by using Self-Supervised learning and Dense NET 121 convolution network is able to detect different stages of diabetic retinopathy by using the given fundus data. In this model we used proxy task as Rotations and we are also able to find the accuracy of the model by using Kappa-Kaggle score. Our findings, particularly in the low data regime, show that in the medical imaging sector, where data and annotation scarcity is a problem, it is possible to reduce the manual annotation labour necessary. We believe that utilising deep learning to diagnose Diabetic Retinopathy improves and mitigates the risk of vision loss for many patients, as well as making it cost-effective for regular check-ups.

## REFERENCES

- [1] The Four stages of Diabetic Retinopathy <https://modernod.com/articles/2019-june/the-four-stages-of-diabeticretinopathy?c4src=article:infinite-scroll>
- [2] Densenet - 121 Architecture <https://www.kaggle.com/datasets/pytorch/densenet121>
- [3] The five stages of Kappa – Kaggle score <https://www.kaggle.com/code/aroraaman/quadratic-kappa-metric-explained-in-5-simple-steps/notebook>
- [4] Self – Supervised Learning <https://neptune.ai/blog/self-supervised-learning>
- [5] Proxy task <https://medium.com/analytics-vidhya/what-is-self-supervised-learning-in-computer-vision-a-simple-introduction-def3302d883d>
- [6] Fine tuning with Keras And Deep Learning <https://pyimagesearch.com/2019/06/03/fine-tuning-with-keras-and-deep-learning/>
- [7] C. Doersch, A. Gupta and A. A. Efros, "Unsupervised Visual Representation Learning by Context Prediction," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1422-1430, doi: 10.1109/ICCV.2015.167. L.-C. Chen, G. Papandreou, I.Kokkinos, K.Murphy, and A. L. Yuille. Semantic Image Segmentationwith Deep Convolutional Nets and Fully Connected CRFs. *arXiv:1412.7062 [cs]*, Dec. 2014. arXiv: 1412.7062.
- [8] Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles [1603.09246v3] Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles (arxiv.org).
- [9] Zhang R., Isola P., Efros A.A. (2016) Colorful Image Colorization. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9907. Springer, Cham. [https://doi.org/10.1007/978-3-319-46487-9\\_40](https://doi.org/10.1007/978-3-319-46487-9_40)
- [10] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. CoRR, abs/1803.07728, 2018. URL <http://arxiv.org/abs/1803.07728>.
- [11] Deepak Pathak, Philipp Krähenbühl, Jeff Donahue, Trevor Darrell, and Alexei Efros. Context encoders: Feature learning by inpainting. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.
- [12] Jiawei Wang, Shuai Zhu, Jiao Xu, and Da Cao. The retrieval of the beautiful: Self supervised salient object detection for beauty product retrieval. In Proceedings of the 27th ACM International Conference on Multimedia, MM '19, page 2548–2552, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450368896. doi: 10.1145/3343031.3356059. URL: <https://doi.org/10.1145/3343031.3356059>.
- [13] Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. CoRR, abs/1807.03748, 2018. URL <http://arxiv.org/abs/1807.03748>.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)