DIABETIC RETINOPATHY GRADING USING VISION TRANSFORMERS

GUIDED BY
Dr. BIJOY A JOSE
ASSOCIATE PROFESSOR
DCS, CUSAT

SUBMITTED BY
ANUSREE M NAIR
MTECH CSE SE
DCS, CUSAT

INTRODUCTION

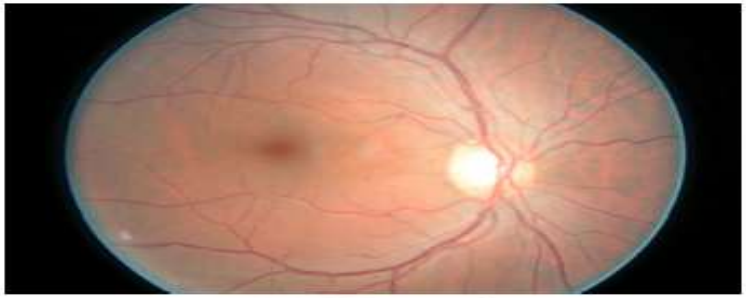
- Diabetic retinopathy(DR) is a medical condition in which damage occurs to retina of eye due to diabetic mellitus.
- It is caused by prolonged high blood glucose damaging the small blood vessels of the retina.
- It's the most recurrent cause of irreversible blindness and is highly likely when diabetes is poorly controlled.
- The primary cause of DR development and its consequences are avoiding the precautionary measure for blood sugar control and a healthy lifestyle.
- Its main symptom is that it produces mutilation of blood vessels of retina.

INTRODUCTION

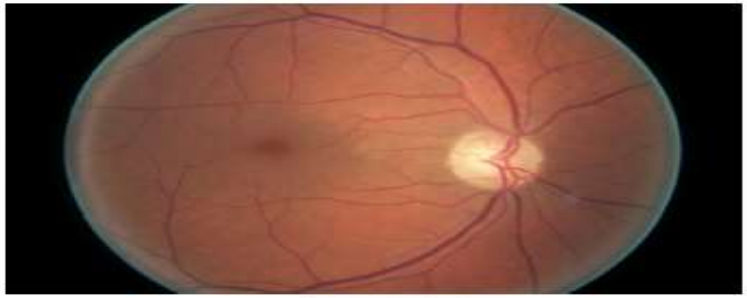
- The DR has two major types: the Non-Proliferative Diabetic Retinopathy (NPDR) and Proliferative Diabetic Retinopathy (PDR).
- The DR in the early stages is called NPDR which is further divided into Mild, Moderate, and Severe stages.
- The mild stage has one micro-aneurysm (MA), a small circular red dot at the end of blood vessels.
- In the Moderate stage the MAs rapture into deeper layers and form a flame-shaped hemorrhage in the retina.
- The severe stage contains more than 20 intra-retinal hemorrhages in each of the four quadrants, having definite venous bleeding with prominent intra-retinal microvascular abnormalities.
- PDR is the advanced stage of DR which leads to neovascularization, a natural formation of new blood vessels in the form of functional microvascular networks that grow on the inside surface of the retina

INTRODUCTION

DIFFERENT STAGES OF DIABETIC RETINOPATHY(DR)

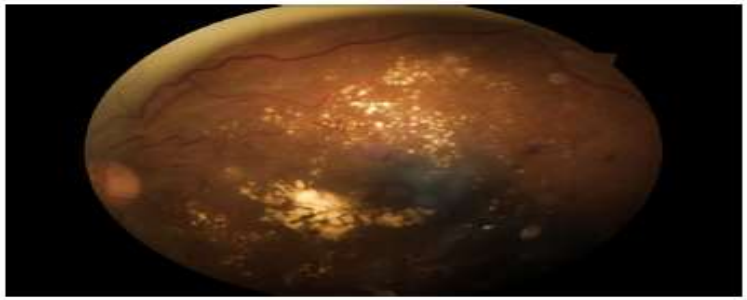


(a) Normal



(b) Mild

(c) Moderate



(d) Severe

(e) PDR

INTRODUCTION

- Detecting DR is a time consuming and manual process.
- Requires a trained clinician to examine and evaluate the digital color fundus photographs of retina.
- Delayed results leads to lost follow up ,miscommunication and delayed treatment.
- This project is an attempt towards finding an automatic way to classify a given set of fundus images using Vision transformers to classify the retinal disease and its stages based on its severity.
- It is an approach different from traditional DL techniques and is more accurate.

INTRODUCTION

- Diabetic retinopathy (DR) is a disease that damages retinal blood vessels and leads to blindness.
- Usually, colored fundus shots are used to diagnose this irreversible disease.
- Hence, various Machine learning based techniques were applied to predict the occurrences of the DR and its stages automatically.
- However, the performance of these ML based techniques depends upon features extracted. Also, these techniques are data hungry hence requires large data.

INTRODUCTION

- Here focus is on classifying the DR's different stages using vision transformers.
- The transformers employs an effective self attention mechanism that focus on specific portions of input fundus image to achieve more effective performance.
- It has the ability to replace the standard CNNs while achieving excellent results.

MOTIVATION

- Diabetic Retinopathy (DR) is an ophthalmic disease that damages retinal blood vessels
- DR causes impaired vision and may even lead to blindness if it is not diagnosed in early stages.
- Manual diagnosis is error proned and tedious.
- The number of patients are increasing day by day and hence its high time to find a quick solution.

PROBLEM STATEMENT

- To classy and detect different stages of diabetic retinopathy using vision transformers.

LITERATURE REVIEW

1. **Weakly supervised localisation of diabetic retinopathy lessions in retinal fundus images**

Authors : W. M. Gondal, J. M. Kohler, R. Grzeszick, G. A. Fink, and M. Hirsch

Publication : *IEEE Int. Conf. Image Process. (ICIP)*,Sep. 2017,

Summary:

- A CNN Referable Diabetic Retinopathy (RDR) approach is used.
- The network performance is evaluated on two different data sets for binary classification: kaggle dataset and DiaretDB1 dataset.
- Binary classification is performed by considering stage-0 and stage-1 as one group, and the rest of the 2,3 and 4 are grouped. Stage-1 features and lesions are challenging to detect because they may appear different or has fewer sample images.
- DR has five stages based on a disease that gradually increases the risk of eye-sight loss and hence it is important to detect any disease at its early stage to cure it on time and never led to the most dangerous and non-curable stage, stage 4 (PDR)

LITERATURE REVIEW

2. Lesion detection and Grading of Diabetic Retinopathy via Two stages Deep Convolutional Neural Networks

Authors : Y. Yang, T. Li, W. Li, H. Wu, W. Fan, and W. Zhang

Publication : Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Springer, 2017

Summary:

- Classification of two stages of DR (NDPR and normal) is done here.
- A DCNN with two networks, global and local is used.
- The local network highlights the lesions and sends them to the global network for further grading.
- The significant limitation of the study was that it has not considered the entire dataset of five stages.

LITERATURE REVIEW

3. A deep learning ensemble approach for diabetic retinopathy detection

Authors : S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. Ahmed Khan, and W. Jadoon,

Publication : *IEEE Access, Vol 7, 2019*

Summary:

- An Ensemble approach of classification is used.
- 5 deep ConvolutionNeural Network (CNN) models like Resnet50, Inceptionv3, Xception, Dense121, Dense169 are used to train a publicly available kaggle dataset.
- Stacking technique is used where stacking is created from a diverse group of strong classifiers.
- The limitation is that model gets too complex and loss transparency in reaching a model that adjusts for its errors.

LITERATURE REVIEW

4. A Deep Learning Approach for Diabetic Retinopathy detection using Transfer Learning

Authors : L. Andersen and P. Andersson,

Publication : Tech. Rep., 2020.

Summary:

- A transfer learning based approach for Diabetic Retinopathy categorization is used.
- Pretrained models called SEResNeXt32x4d and EfficientNetb3 are used.
- The pretraining of the aforementioned neural networks has been done on the ImageNet dataset.
- The Diabetic Retinopathy images are migrated to these models.
- Based on the dataset already available, the output is ultimately split into 5 levels according to the seriousness of the degree of DR.

PROPOSED SYSTEM

DATASET PREPROCESSING:

DATASET:

- Kaggle dataset is used.
- It contains high resolution retina images of different stages of diabetic retinopathy including normal, mild, moderate, severe and proliferative.
- The images are captured in different camera models and imaging techniques hence giving diverse range of datasets.

PROPOSED SYSTEM

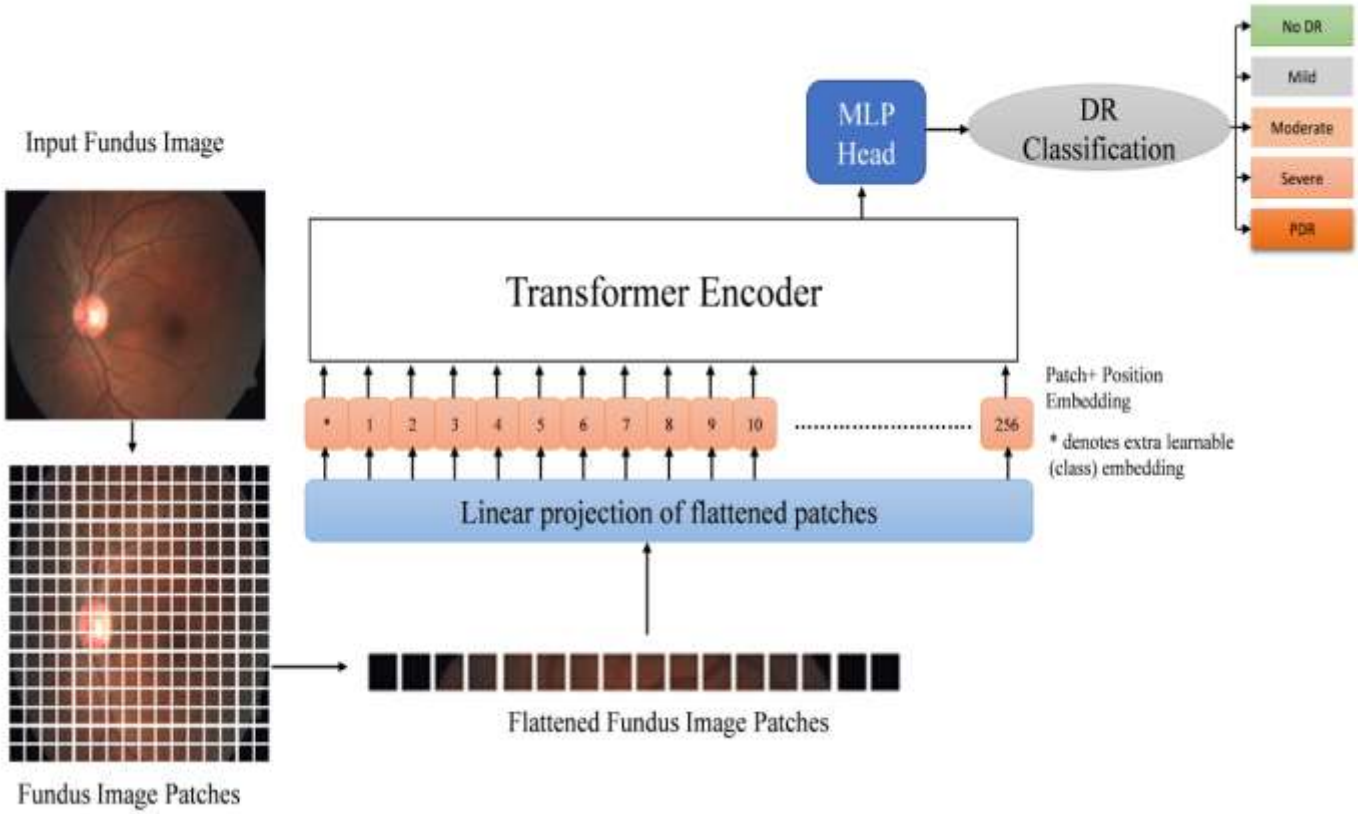
DATASET PREPROCESSING:

PREPROCESSING:

- Before feeding data into model some preprocessing steps are applied.
- First, we resize the images keeping in view the aspect ratio. It helps us to avoid features loss from images.
- Then, images are randomly cropped to a fixed size.
- Dataset is divided into training and test sets.

PROPOSED SYSTEM

ViT-DR ARCHITECTURE



PROPOSED SYSTEM

- A vision transformer based classification is used here.
- It is a pure Transformer based architecture used for image classification tasks.
- The sequence of steps followed in ViT is as follows:
  - Input image transformation
  - Class Tokens
  - Positional Embedding
  - Transformer Encoder
- The input image is first split into fixed-size non overlapping patches.
- Position embeddings are added and then fed into the Transformer Encoder.
- An extra learnable embedding/class token is added at the beginning for performing image classification.

PROPOSED SYSTEM

INPUT IMAGE TRANSFORMATION

- The first step is the transformation of a 2D image into a sequence of flattened patches.
- Input Image (I) = H x W x C, would be transformed into Patches (P) = N x (P^2 x C).
- N = (H x W) / (P x P)
  - Here,
  - H x W: refers to the height and width of the Input Image (I).
  - C: refers to the number of channels in the Input Image (I)
  - P^2: refers to the height and width of an individual patch.
  - N: refers to the number of patches.

- All the flattened patches are mapped into dimensions by passing through some learnable linear layers.

- The output features from these linear layers are called patch embeddings.

PROPOSED SYSTEM

CLASS TOKENS

- A learnable class token is added at the beginning of the sequence of the patch embeddings.
- The class token gathers information from the sequence of the patches over the different layers and learns the general image representation.
- Later when the ViT is used for image classification, an MLP (Multilayer Perceptron) is used.
- The MLP uses the information from the class token.
- In this way, the model is not biased towards any specific image patch.

PROPOSED SYSTEM

POSITION EMBEDDINGS

- Position embeddings are added to patch embeddings to retain the positional information.
- The positional encoding helps in ViT to determine the correct sequence of patches, such as the 1st patch referring to the top-left and so on.

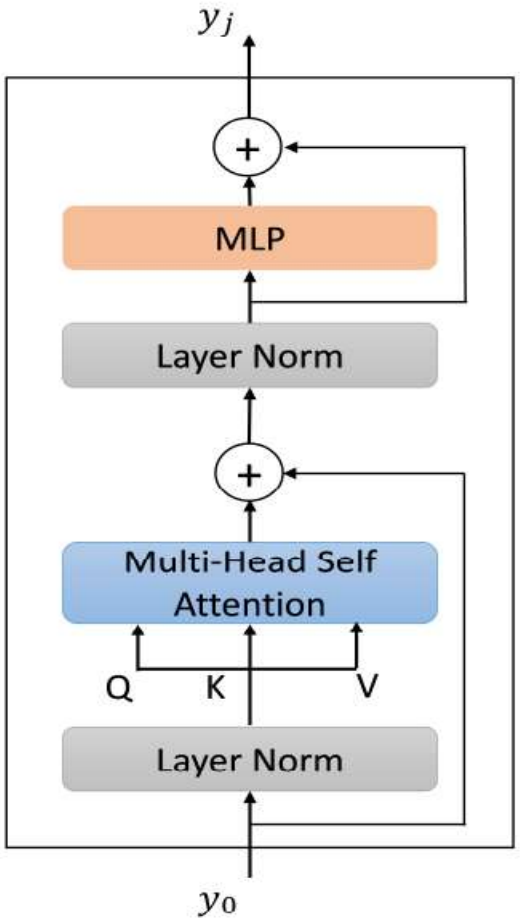
PROPOSED SYSTEM

TRANSFORMER ENCODER

- The embedded sequence of patches are passed on to transformer encoder.
- The transformer encoder begins with Normalization Layer (LayerNorm), Multi-Head Self Attention (MSA) and then is followed by a residual connection from the input.
- Next, again the Normalization Layer (LayerNorm), followed by a sequence of MLP (Multilayer Perceptron) and the residual connection.
- The MLP consists of two Linear (fully connected) layers and a GELU (Gaussian Error Linear Units) in between.

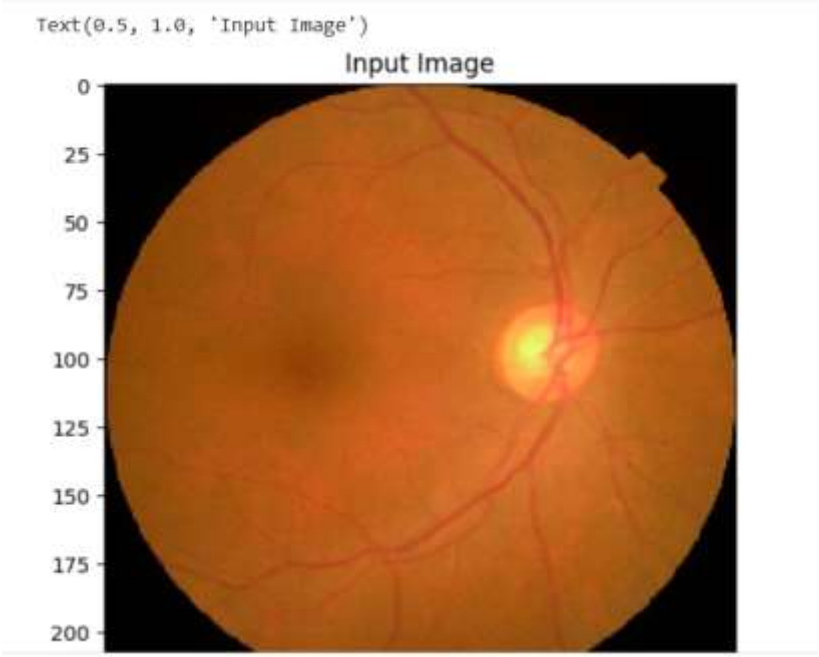
PROPOSED SYSTEM

- **TRANSFORMER ENCODER**



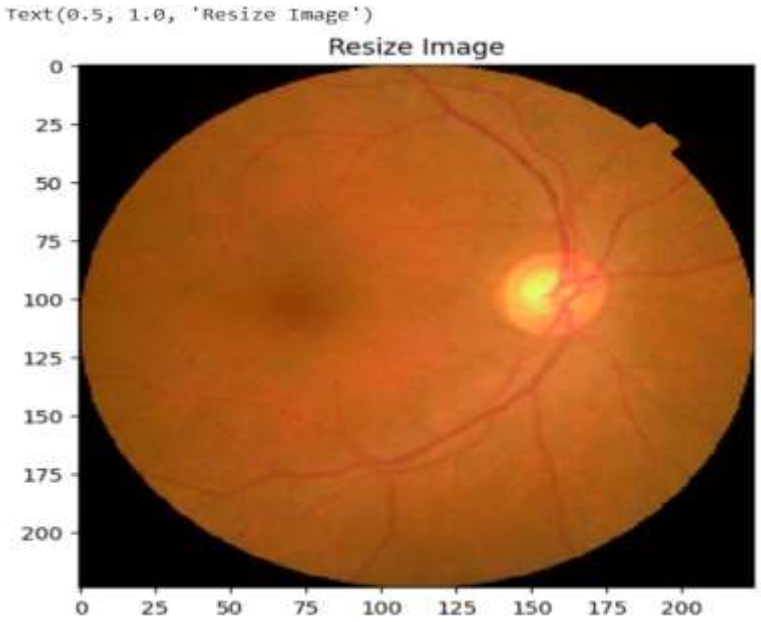
RESULTS

- **Input image:** The retinal fundus image is input for preprocessing.



RESULTS

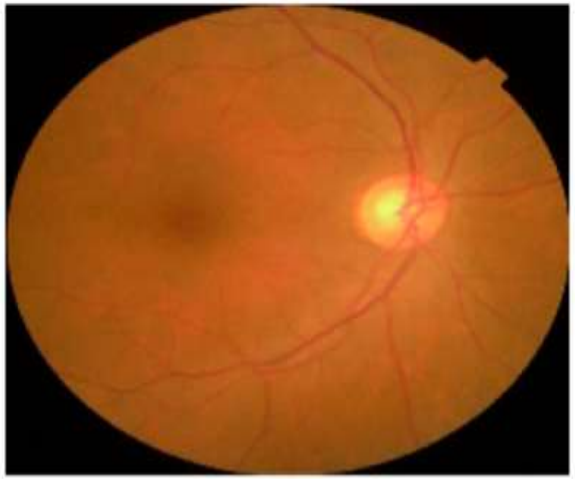
- **Resized Image:** The image is preprocessed and resized into a required fixed size.

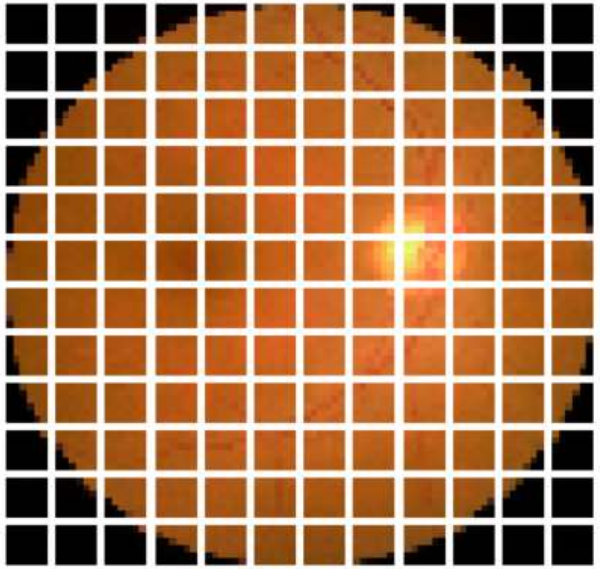


RESULTS

- **Dividing into patches:**

```
Image size: 72 X 72
Patch size: 6 X 6
Patches per image: 144
Elements per patch: 108
```





RESULTS

- **Model Summary:**

```
Model: "sequential_3"
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 conv2d_9 (Conv2D)            (None, 72, 72, 16)        208

 max_pooling2d_9 (MaxPooling  (None, 36, 36, 16)        0
 2D)

 conv2d_10 (Conv2D)           (None, 36, 36, 32)        2080

 max_pooling2d_10 (MaxPoolin  (None, 18, 18, 32)        0
 g2D)

 conv2d_11 (Conv2D)           (None, 18, 18, 64)        8256

 max_pooling2d_11 (MaxPoolin  (None, 9, 9, 64)          0
 g2D)

 dropout_6 (Dropout)          (None, 9, 9, 64)          0

 flatten_3 (Flatten)          (None, 5184)              0

 dense_6 (Dense)              (None, 500)               2592500

 dropout_7 (Dropout)          (None, 500)               0

 dense_7 (Dense)              (None, 6)                 3006

=================================================================
```

RESULTS

- **Model Prediction:**

```
=================================================================
Total params: 2,606,050
Trainable params: 2,606,050
Non-trainable params: 0
_________________________________________________________________
-------------------------------------
CONVOLUTIONAL NEURAL NETWORK (CNN)
-------------------------------------

842/842 [==============================] - 46s 53ms/step - loss: 0.6031
--------------------------------
ACCURACY FOR DEEP LEARNING
--------------------------------

1. Accuracy  = 99.39688324928284 %

2.Error Rate = 0.6031167507171631 %

 -------------
 PREDICTION
 -------------
 Identified = NORMAL
```

CONCLUSION

- Diabetic retinopathy causes irreversible visual loss if not detected and treated on time.
- This approach presents an improved transformer model for grading the severity levels of diabetic retinopathy.
- The performance of this model is improved by selecting the proper fundus image size, patch size, number of transformers, multi-layer perceptron head layers.
- It can be improved further by increasing the size of the database and patches at the cost of computation.
- The results achieved with the proposed model are better than the other approaches.
- In the future, the efficiency of the proposed model can be improved with better preprocessing techniques and can be tested with a real-time database.

REFERENCES

- W. M. Gondal, J. M. Kohler, R. Grzeszick, G. A. Fink, and M. Hirsch, ``Weakly supervised localization of diabetic retinopathy lesions in retinal fundus images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 20692073.

- 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," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Springer, 2017, pp. 533540.*

- S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. Ahmed Khan, and W. Jadoon, ``A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access, vol. 7, pp. 150530150539,* 2019.

- X. Zeng, H. Chen, Y. Luo, and W. Ye, ``Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network,“ *IEEE Access, vol. 7, pp. 3074430753, 2019.*

REFERENCES

- M. Lin, Q. Chen, and S. Yan, ``Network in network," 2013, *arXiv:1312.4400. [Online]. Available:* http://arxiv.org/abs/1312.4400

- E. Bhatti and P. Kaur, ``DRAODM: Diabetic retinopathy analysis through optimized deep learning with multi support vector machine for classification," in *Proc. Int. Conf. Recent Trends Image Process. Pattern Recognit.* Springer, 2018, pp. 174188.

- K. He, X. Zhang, S. Ren, and J. Sun, ``Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 9, pp. 19041916, Sep. 2015.*

THANK YOU