



Diabetic Retinopathy Grade Classification using Vision Transformers

Workshop



Sara EL-ATEIF
ML GDE
@el_ateifSara

About Me

Google Ph.D. Fellow

Pursuing my Ph.D. @ENSIAS, Morocco

NVIDIA DLI Instructor/University
Ambassador

Co-Founder of AI Wonder Girls



Overview

Plan

- Vision Transformers
- CNN vs Vision Transformers
- Diabetic Retinopathy
- Workshop

Vision Transformers

Introduction

The Vision Transformer, or ViT, image classification model that employs a Transformer-like architecture over patches of the image.

(1) An image is split into fixed-size patches,

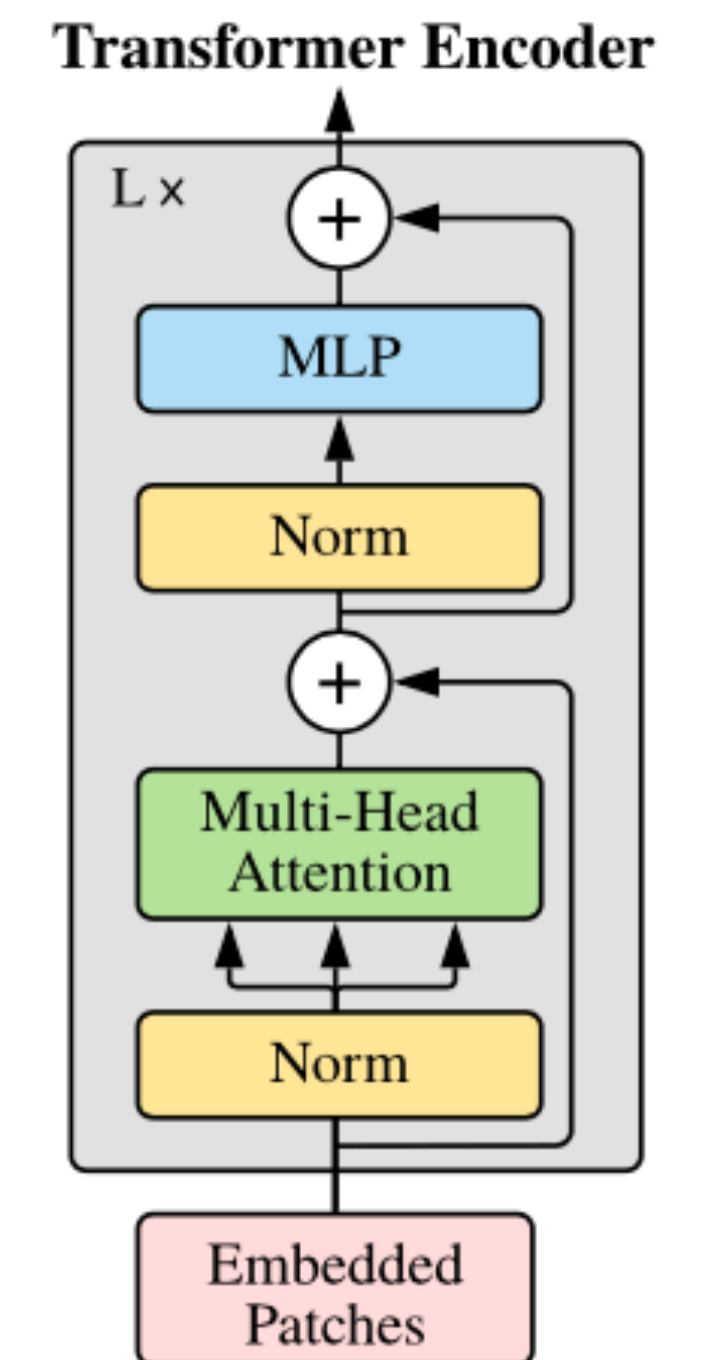
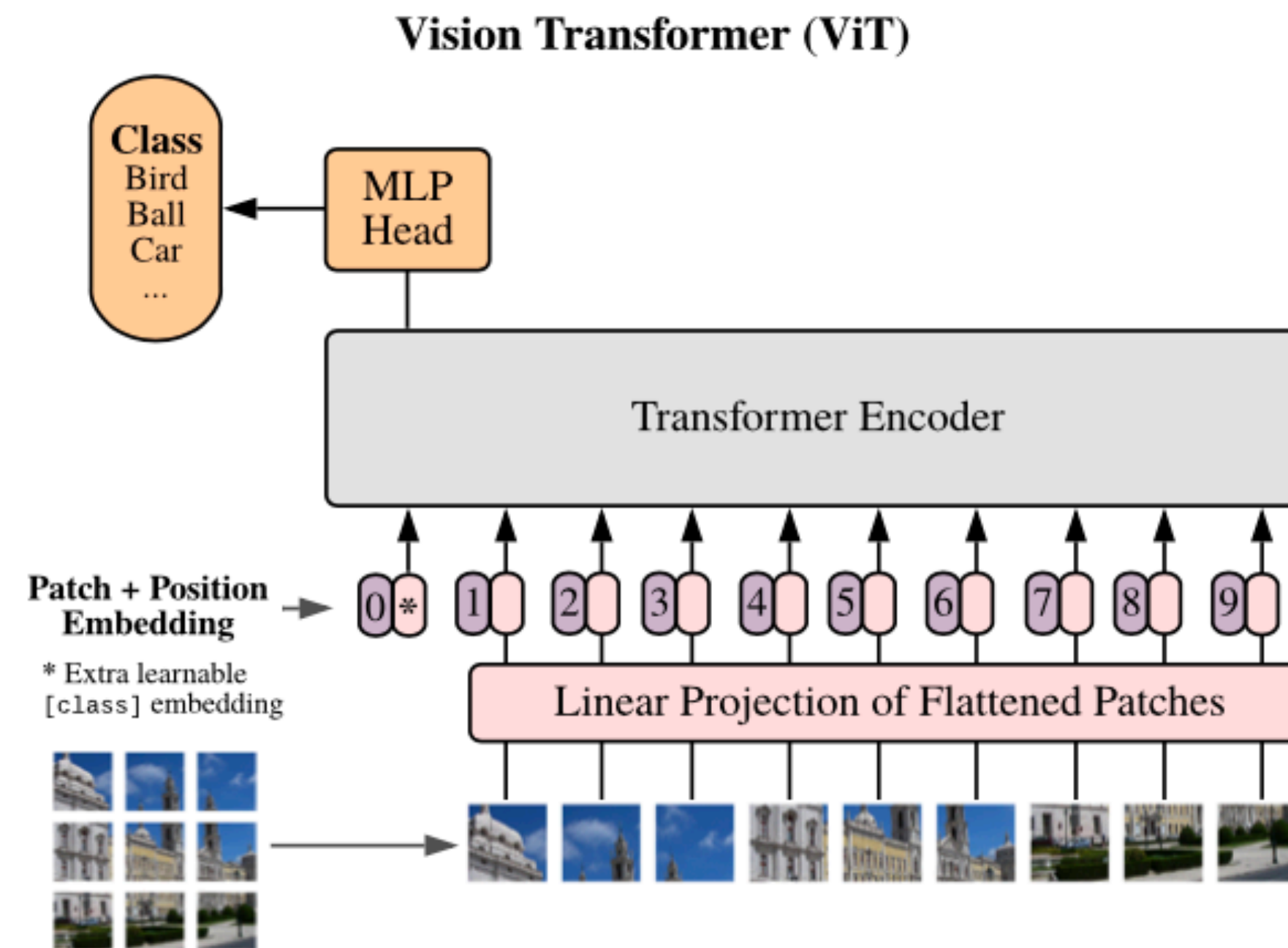
(2) Each image is then linearly embedded,

(3) Position embeddings are added,

And the resulting sequence of vectors is fed to a standard Transformer encoder.

To perform classification, an extra learnable “classification token” is added to the sequence.

Source: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (paper by Dosovitskiy et al.)

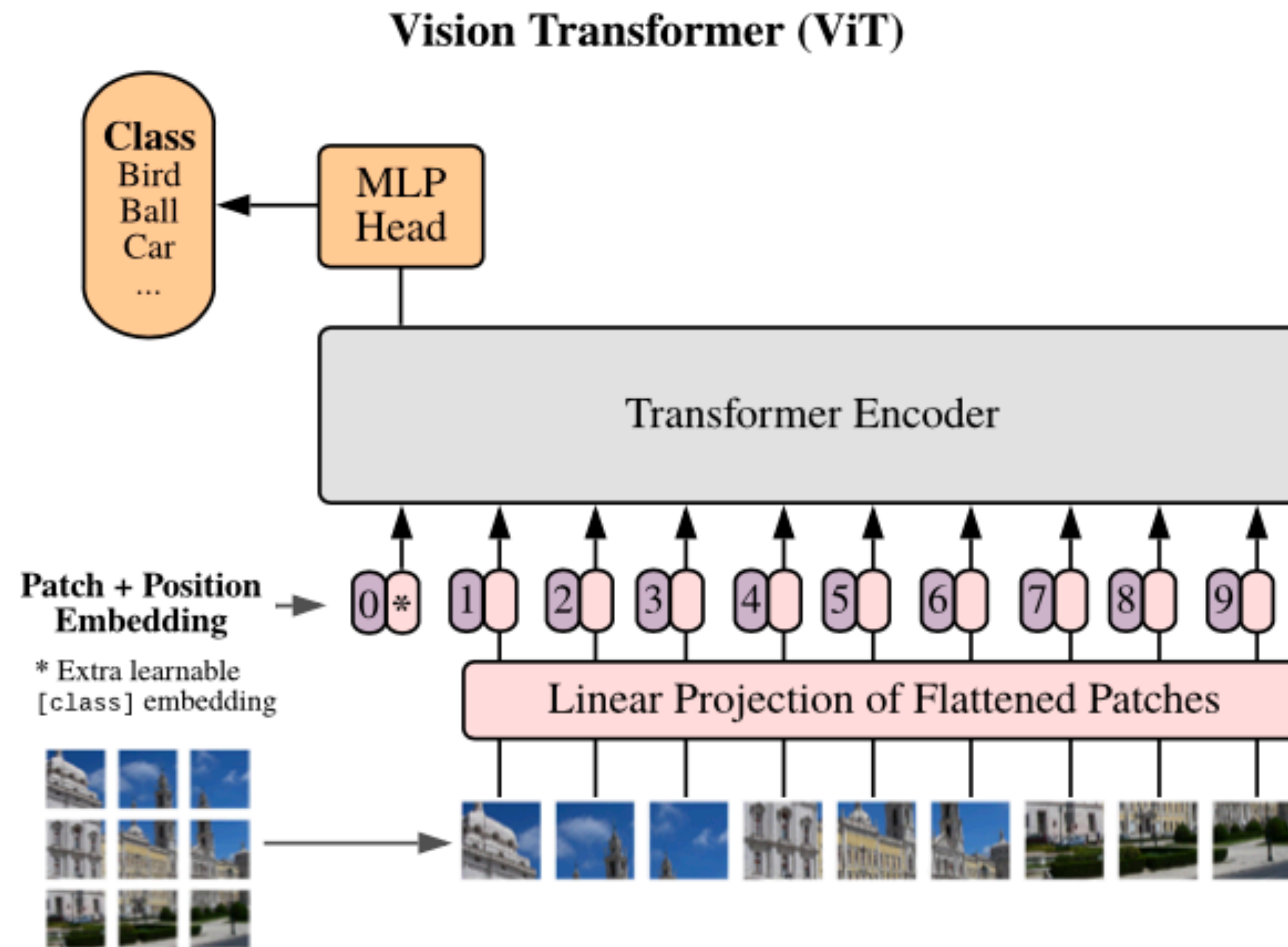


Source: [Lilian Weng](#)

Source: [Attention Is All You Need](#)

Introduction

Architecture



Source: [Lilian Weng](#)

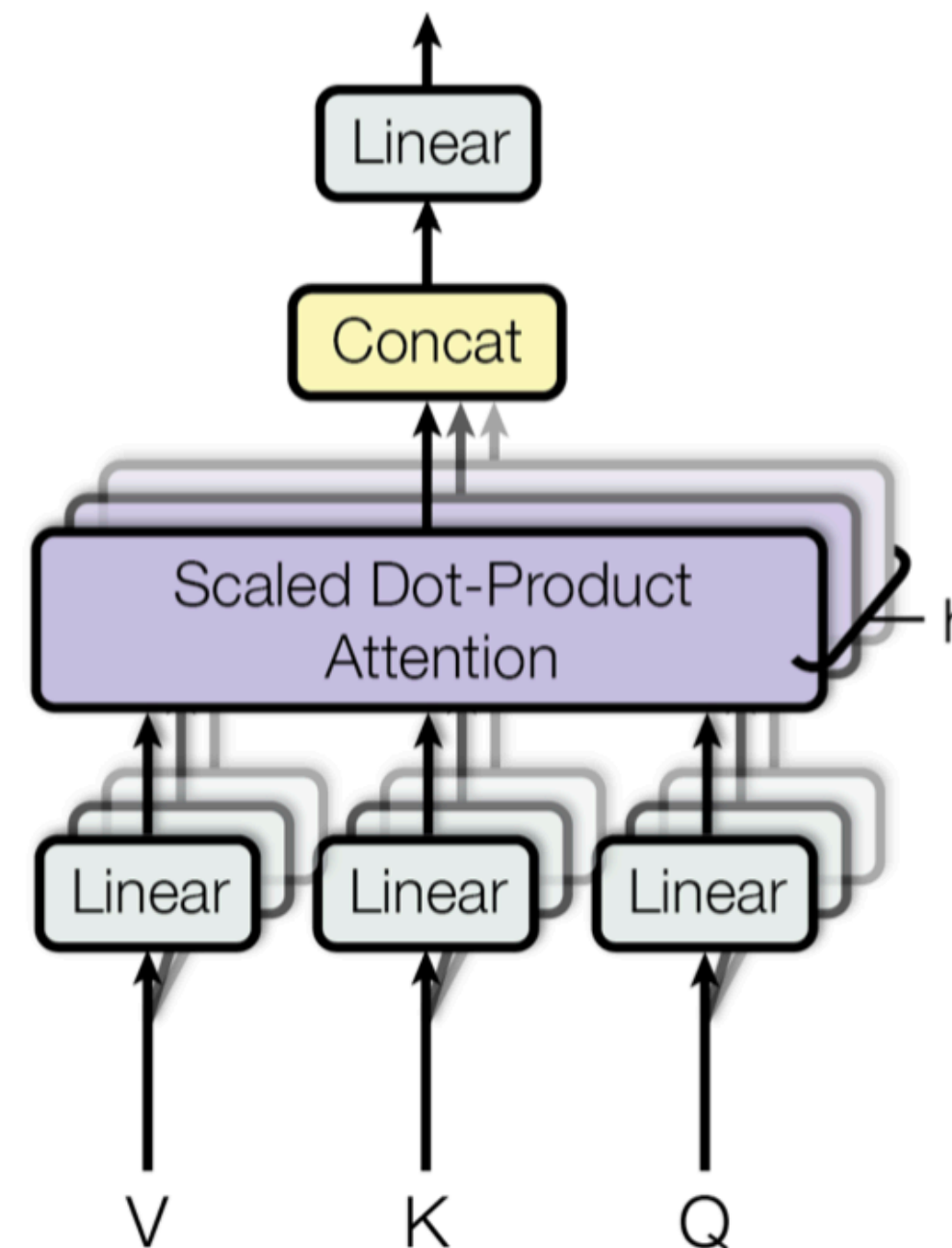
Source: [Attention Is All You Need](#)

Introduction

Transformer Encoder part

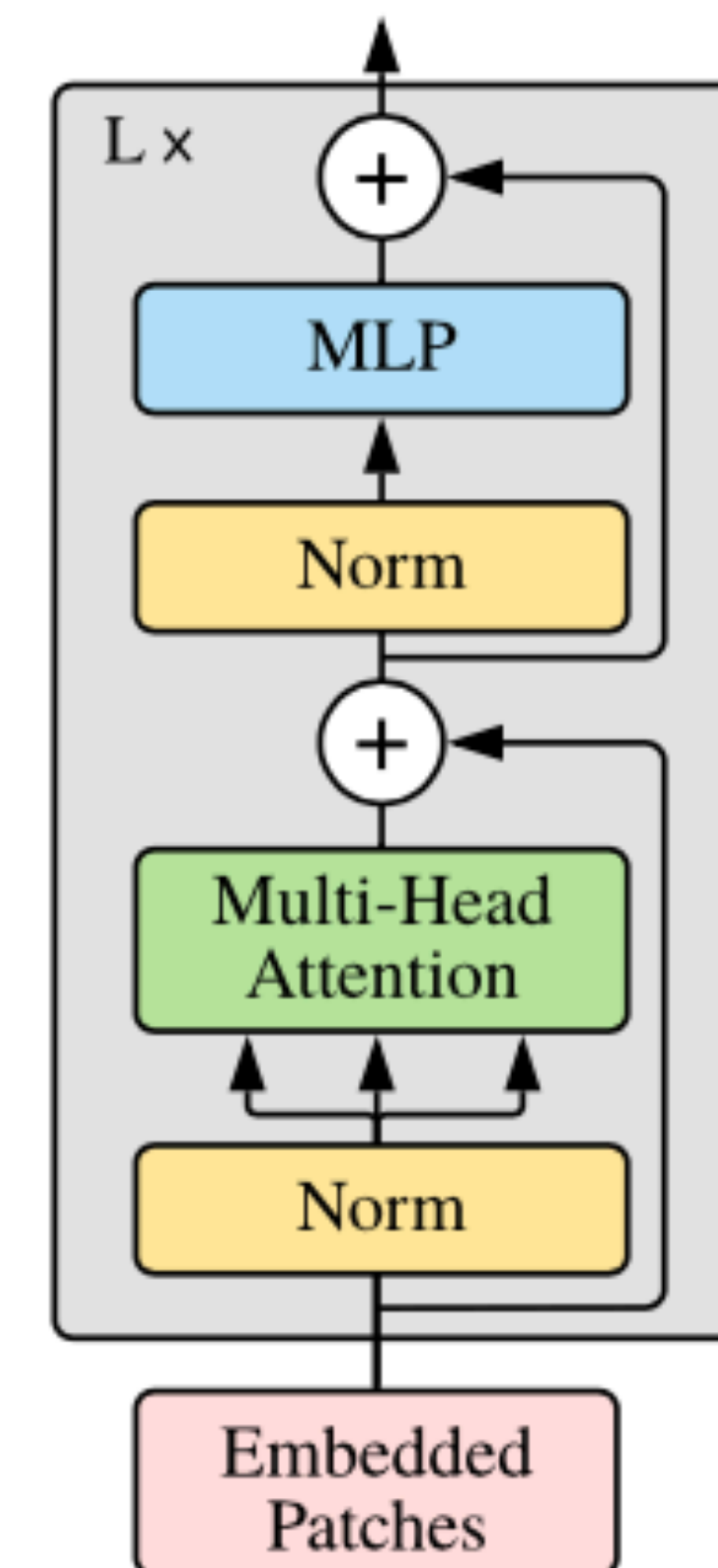
$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = [\text{head}_1, \dots, \text{head}_h] \mathbf{W}_0$$

where $\text{head}_i = \text{Attention}(\mathbf{Q} \mathbf{W}_i^Q, \mathbf{K} \mathbf{W}_i^K, \mathbf{V} \mathbf{W}_i^V)$



Multi-Head Attention

Transformer Encoder

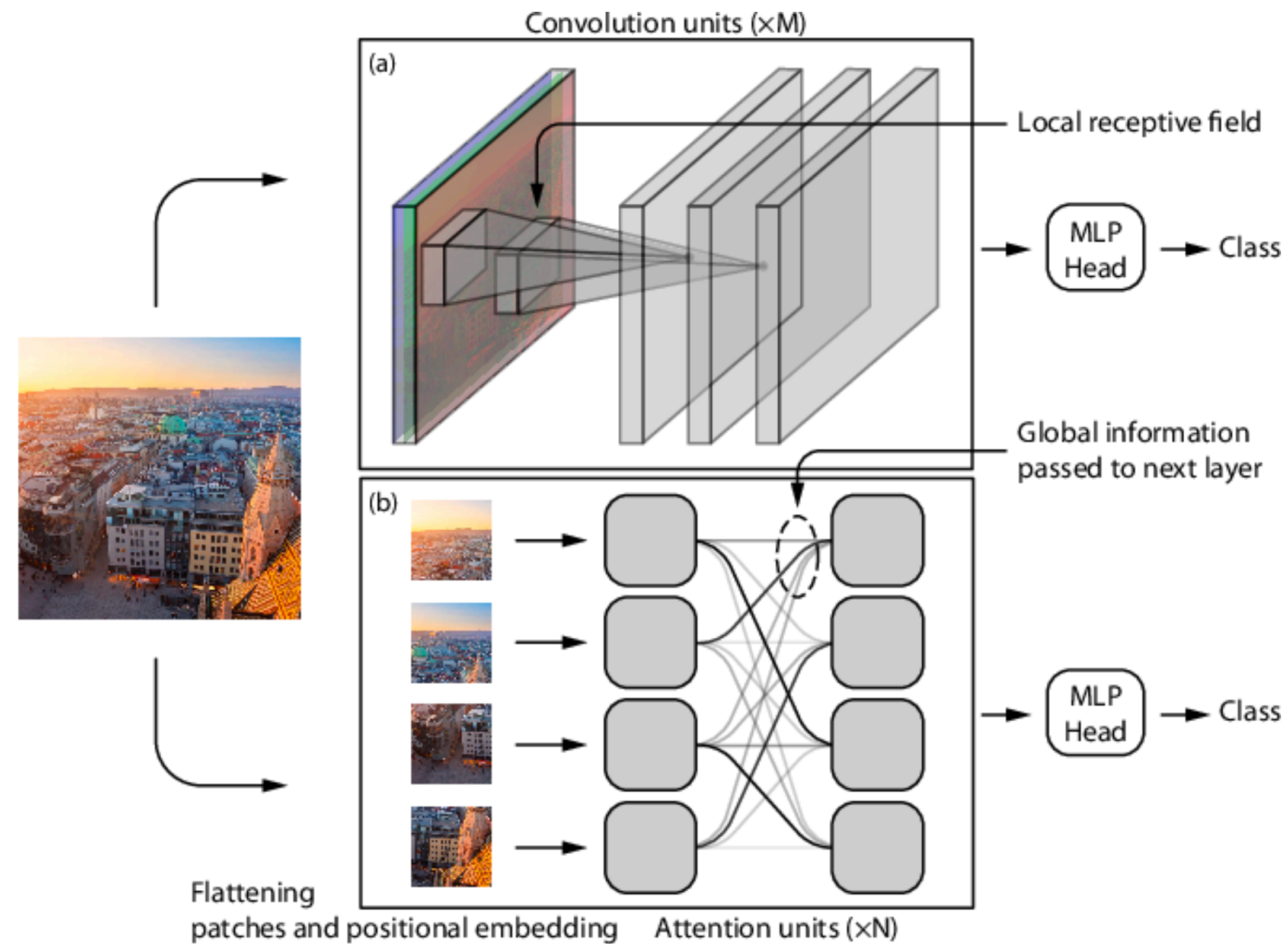


Source: [Lilian Weng](#)

Source: [Attention Is All You Need](#)

CNN vs ViT

Image representation



Source: [Are Convolutional Neural Networks or Transformers more like human vision?](#)

Shikhar Tuli et al.

CNNs vs ViTs

CNNs:

- Focus on local information
- Needs more training time
- Lots of resources
- Vulnerable to adversarial attacks or changes in the data

ViTs:

- Focus on global information
- Needs less training time
- Resources friendly
- Robust against adversarial attacks

Diabetic Retinopathy

Definition

Diabetic Retinopathy worldwide

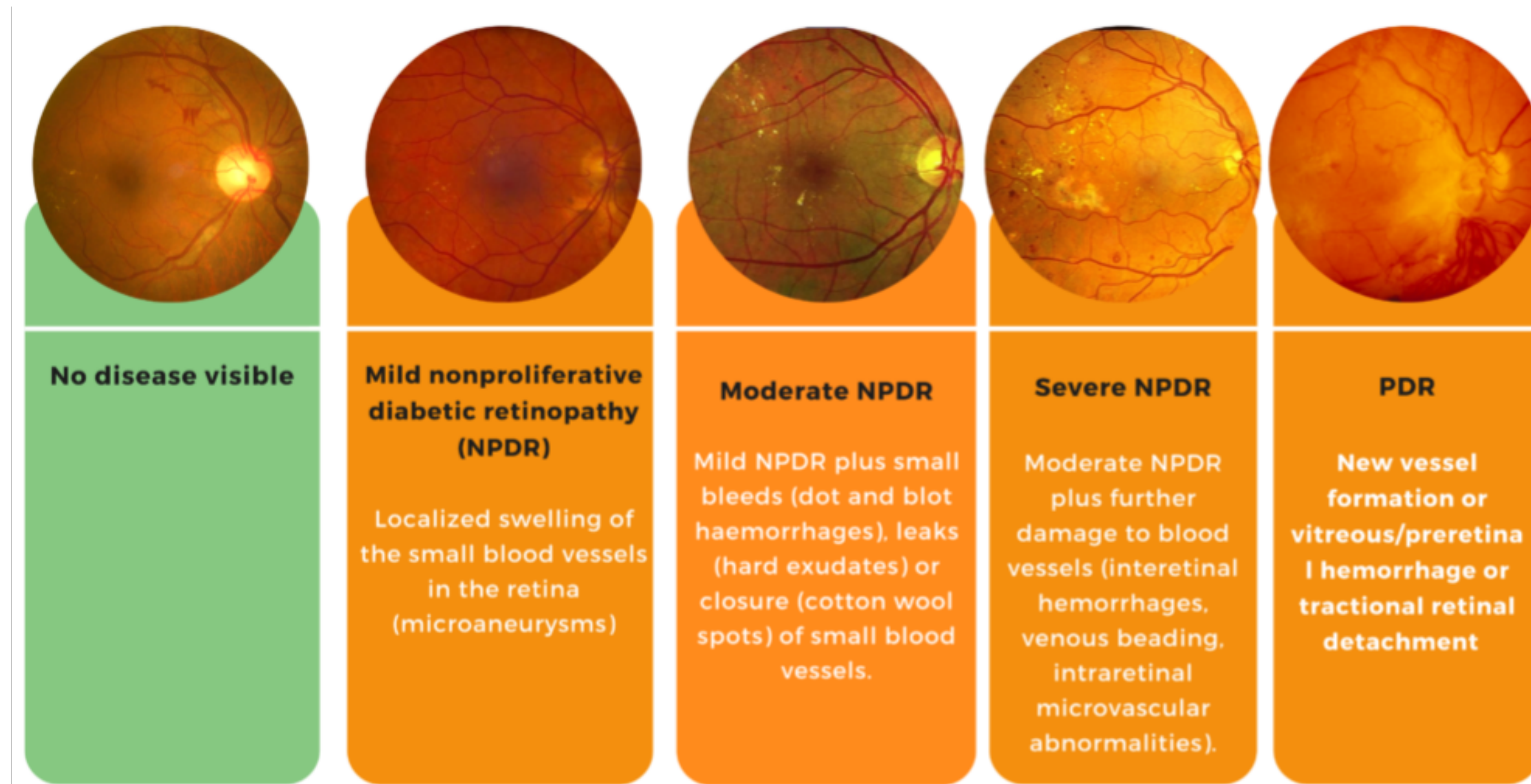
‘**Diabetic Retinopathy (DR)** is a complication of diabetes, caused by high blood sugar levels damaging the back of the eye (retina). It can cause blindness if left undiagnosed and untreated.’

Source: [NHS UK](#)

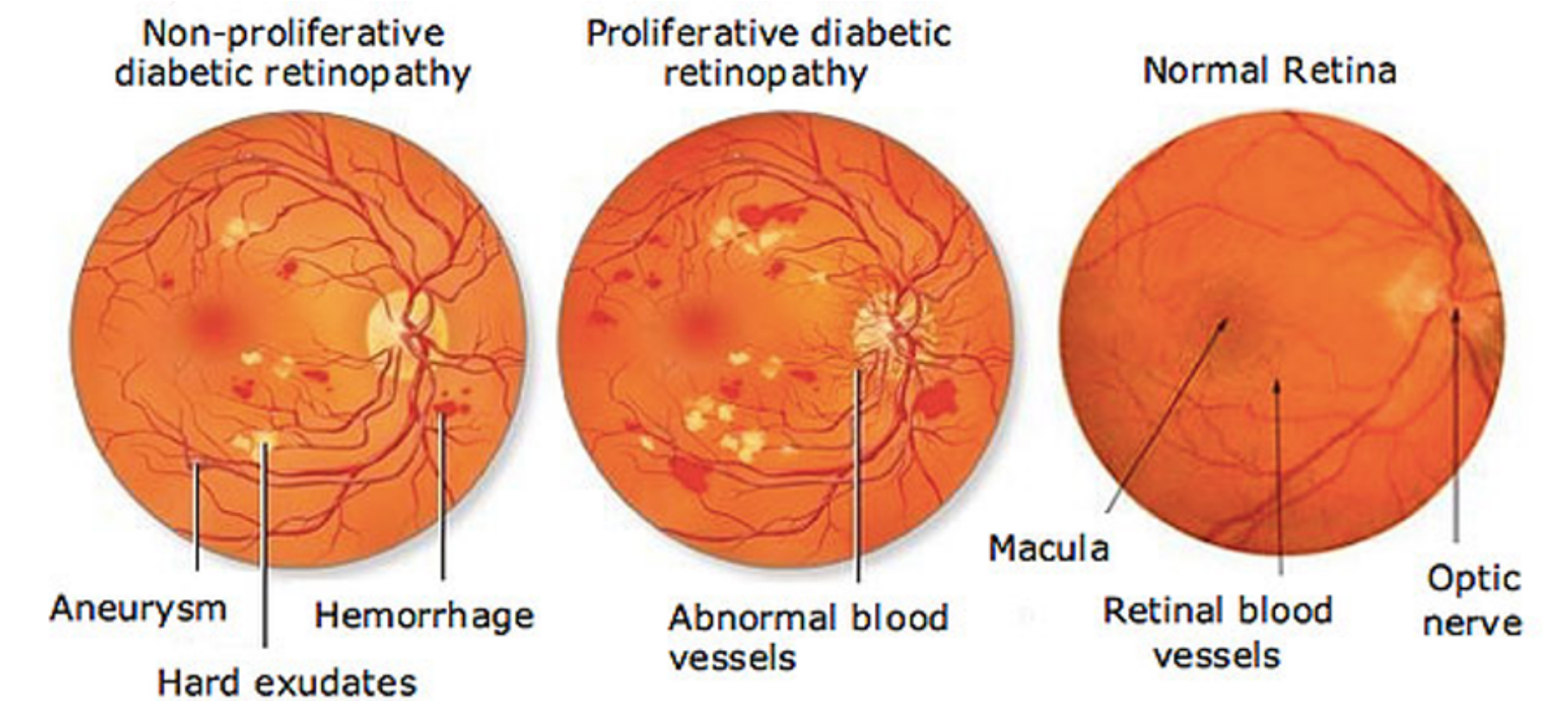
‘Globally, the number of people with DR will grow from **126.6 million in 2010** to 191.0 million by 2030.’

Source: [10.4103/0301-4738.100542](#)

Grades & Symptoms



Source: <https://www.ophtalytics.com/our-technology/diabetic-retinopathy/>



Source: <https://lowvisionaids.org/diabetic-retinopathy/>

Let's see the code

```
Epoch 1/7
183/183 [=====] - 1157s 6s/step - loss: 1.7917 - accuracy: 0.3765 - val_loss: 1.3220 - val_
accuracy: 0.6792

Epoch 00001: val_accuracy improved from -inf to 0.67917, saving model to ./model.hdf5
Epoch 2/7
183/183 [=====] - 1099s 6s/step - loss: 1.2129 - accuracy: 0.6669 - val_loss: 1.3682 - val_
accuracy: 0.6611

Epoch 00002: val_accuracy did not improve from 0.67917
Epoch 3/7
183/183 [=====] - 1068s 6s/step - loss: 1.1460 - accuracy: 0.7294 - val_loss: 1.1512 - val_
accuracy: 0.7431

Epoch 00003: val_accuracy improved from 0.67917 to 0.74306, saving model to ./model.hdf5
Epoch 4/7
183/183 [=====] - 1052s 6s/step - loss: 1.1401 - accuracy: 0.7225 - val_loss: 1.2206 - val_
accuracy: 0.7208

Epoch 00004: val_accuracy did not improve from 0.74306
Epoch 5/7
20/183 [==>.....] - ETA: 13:42 - loss: 1.0558 - accuracy: 0.7961
```

MacBook

Thank You!



Sara EL-ATEIF
ML GDE
@el_ateifSara

