A logo with red and black text

AI-generated content may be incorrect.

**INSTITUTE OF ENGINEERING & MANAGEMENT**

Generative AI and Deep Learning

Project Report — Retinal OCT Classification using Vision Transformer (ViT**)**

DEPARTMENT OF COMPUTER SCIENCE AND BUSINESS SYSTEM

3rd Year, 6th Semester

Batch of 2022-26

Submitted by

|  |  |  |
| --- | --- | --- |
| **Name** | **Roll No.** | **Enrolment No.** |
| Anumoy Pathak | 11 | 12022002018015 |
| Sayan Saha | 26 | 12022002018040 |
| Padmanava Biswas | 35 | 12022002018050 |
| Ridhhima Chanda | 40 | 12022002018055 |
| Rohni Roy Banik | 45 | 12022002018062 |

**1. Title**

Retinal Disease Classification using Vision Transformer on OCT Images

**2. Introduction**

**CNV (Choroidal Neovascularization), DME (Diabetic Macular Edema), DRUSEN,** and **NORMAL** are some common eye diseases which may lead to visual impairment. Hence it is very important to detect these conditions early. We have a solution to improve treatment outcomes by using Optical Coherence Tomography (OCT). Our project applies the **Vision Transformer (ViT)** model to classify retinal OCT images using deep learning techniques into the four mentioned disease categories specified early.

### ****3. Objective****

We have developed an image classification model using **Vision Transformer (ViT)** for accurate diagnosis of retinal OCT scans by classifying them into:

* CNV (Choroidal Neovascularization)
* DME (Diabetic Macular Edema)
* DRUSEN
* NORMAL

**4. Dataset**

Kaggle Dataset: <https://www.kaggle.com/paultimothymooney/kermany2018>

**5. Data Preprocessing**

To adapt OCT grayscale images for **Vision Transformer** input:

* Resized all images to 224x224 pixels
* Converted grayscale images to RGB (ViT expects 3 channels)
* Normalized pixel values with mean and std = [0.5, 0.5, 0.5]
* Applied transformations using PyTorch torchvision.

**6. Model Architecture**

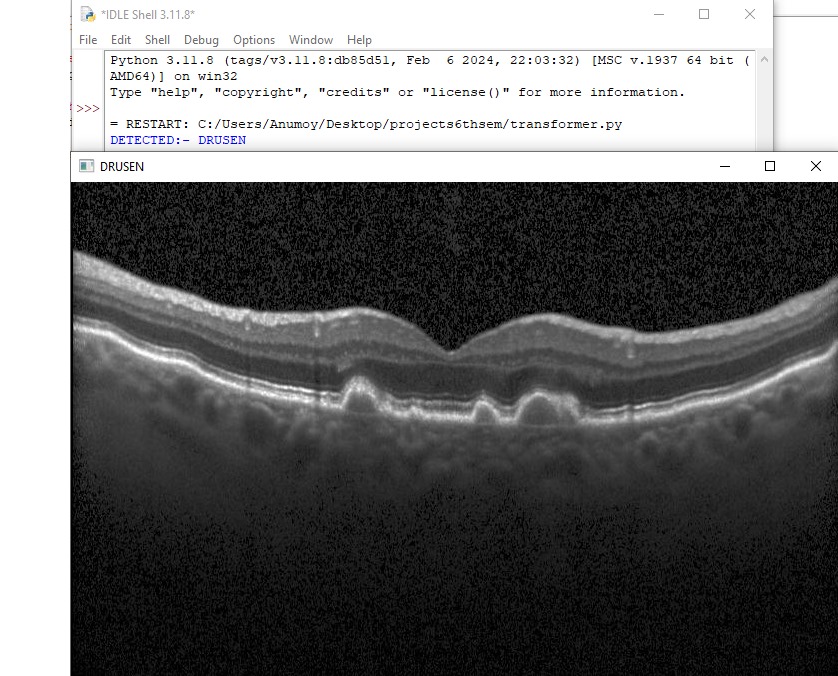
* **Backbone**: google/vit-base-patch16-224-in21k
* **Model**: Pretrained Vision Transformer (ViT) from Hugging Face Transformers
* **Modified Head**: Final classification layer adjusted to 4 output classes
* **Loss Function**: CrossEntropyLoss
* **Optimizer**: Adam (learning rate = 1e-4)

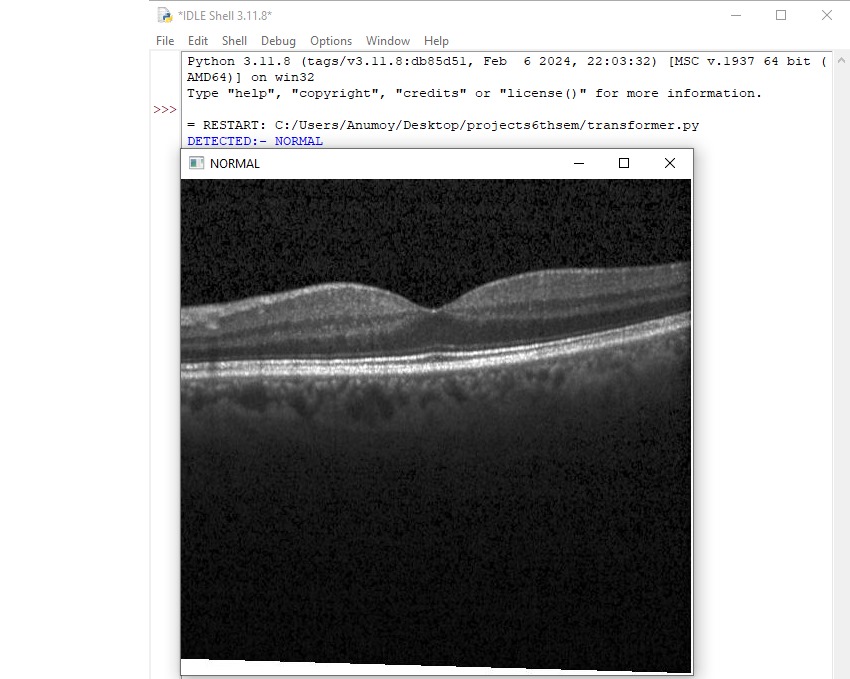
**7. Code**

The full source code in available in the Github link given below :-

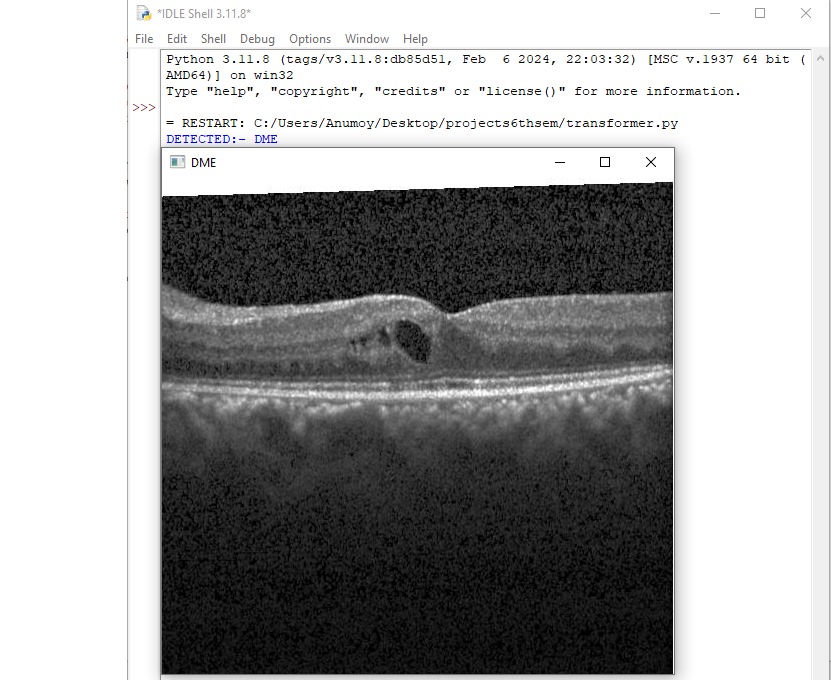
<https://github.com/SayanSaha11/DL-GenAI/blob/main/code.py>

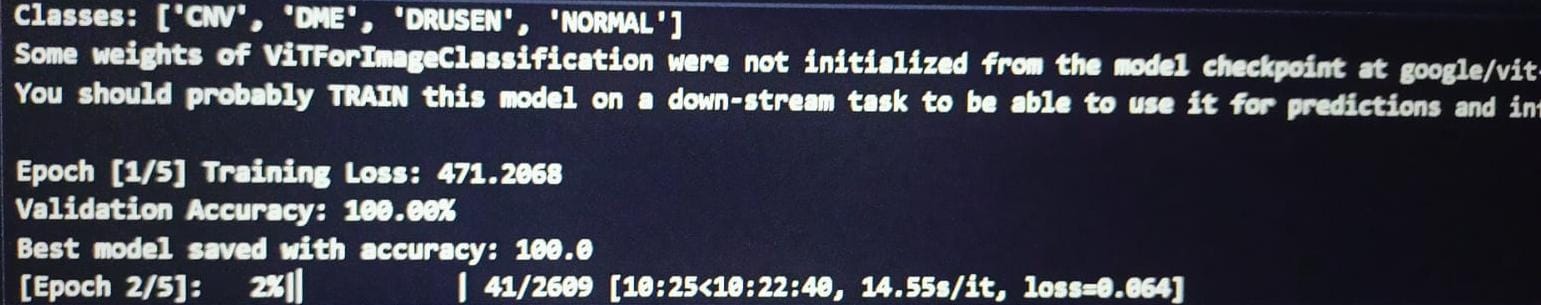
**8. Output**

****

****

****

****

****

**9. Observations**

* **Drusen -** Yellowish extracellular deposits under the retina.
* **Normal -** Uniform retinal layers with no structural abnormalities.
* **CNV -** Disruption in the RPE layer and irregular fluid-like regions beneath the retina
* **DME -** Retinal thickening with cystoid spaces

**10. Conclusion**

 Our deep learning model has demonstrated high accuracy and clinical relevance in classifying Retinal OCT images into the four major categories namely Drusen, Normal, CNV & DME.

 Our model has been able to accurately differentiate between normal and pathological retinas, with visual evidence aligning with known biomarkers for each condition.

 Our approach aids in early screening and potentially reducing diagnostic workload and speeding up treatment timelines.