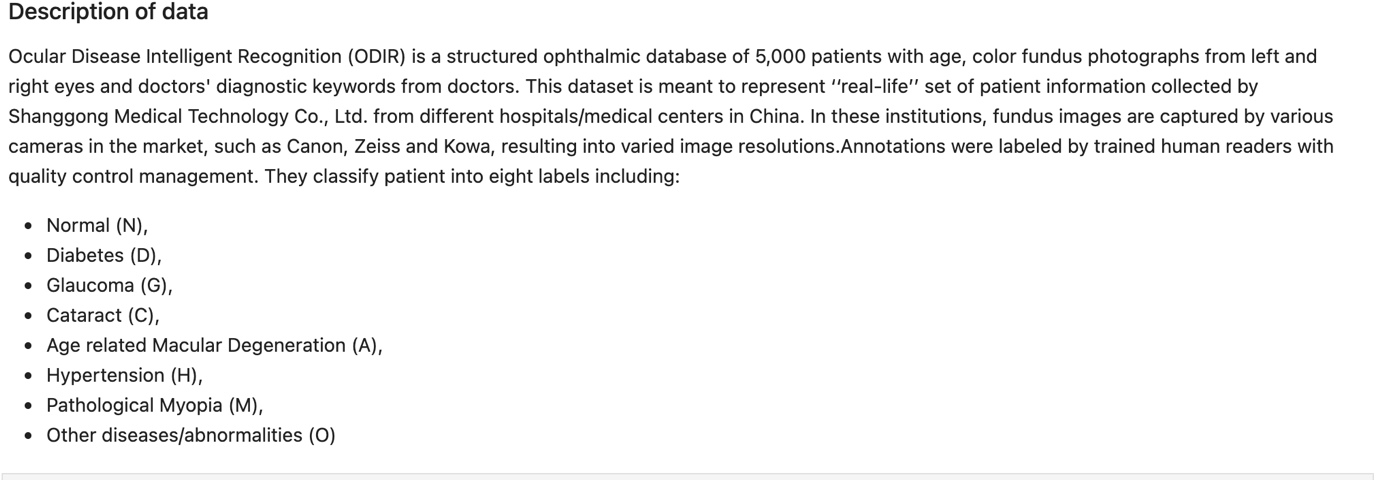
**Installation** :

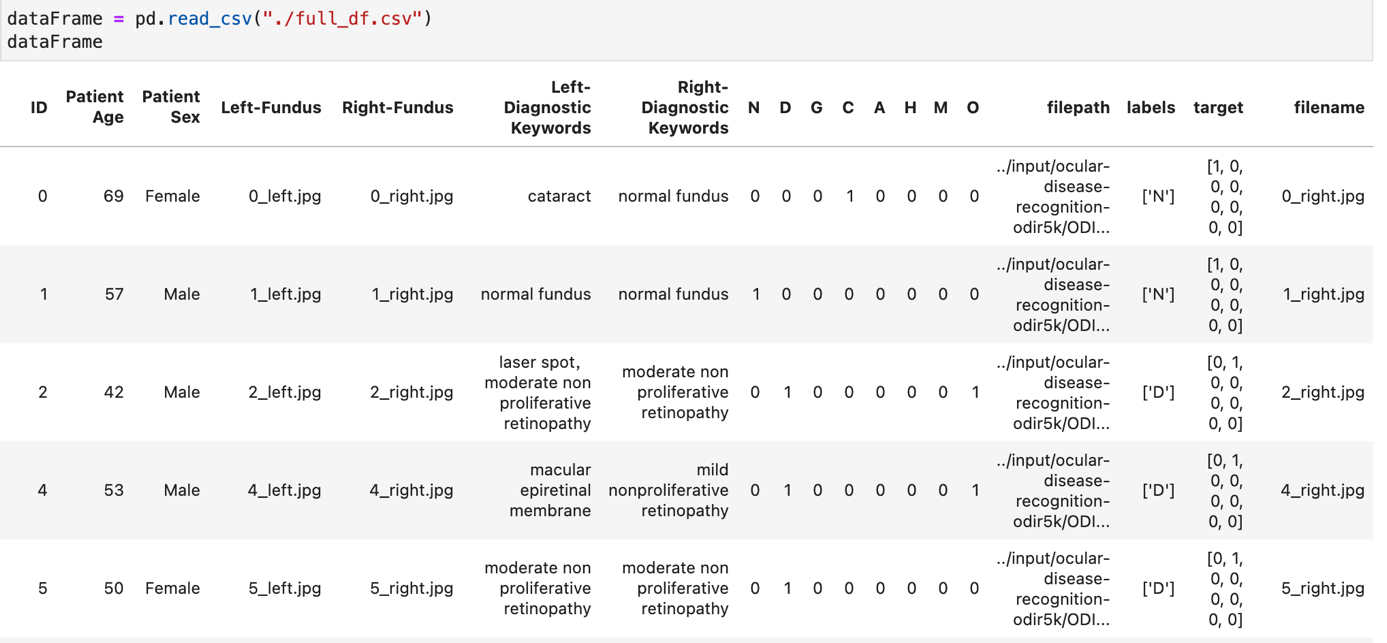
On a travaillé sur **jupyter lab** qui est installé en local avec un environnement python qui contient toutes les dépendances dont on a besoin(par exemple tensorflow, keras, pandas, numpy etc).

**Projet** :

Notre projet porte sur la reconnaissance de la maladie rétinographie diabétique qui est une maladie qui se manifeste au niveau des yeux. Donc pour pouvoir faire l’étude de notre projet il va nous falloir plusieurs données qui vont de permettre d’étudier comment la maladie se manifeste et ainsi creer un modèle de prédiction qui pourra nous aider à dire si la personne est atteinte de la maladie ou pas. Donc on est allé sur le site kaggle qui un répertoire de code et de données déjà disponible. On a puiser notre dataset sur ce site et pour chaque dataset on nous donne sa description et qui parle de description parle par exemple de nombre de variables, de la signification de chaque variable donc les informations concernant ses données. Les images **Figure 1** et **Figure 2** suivantes donnent une brève description des données collectées.

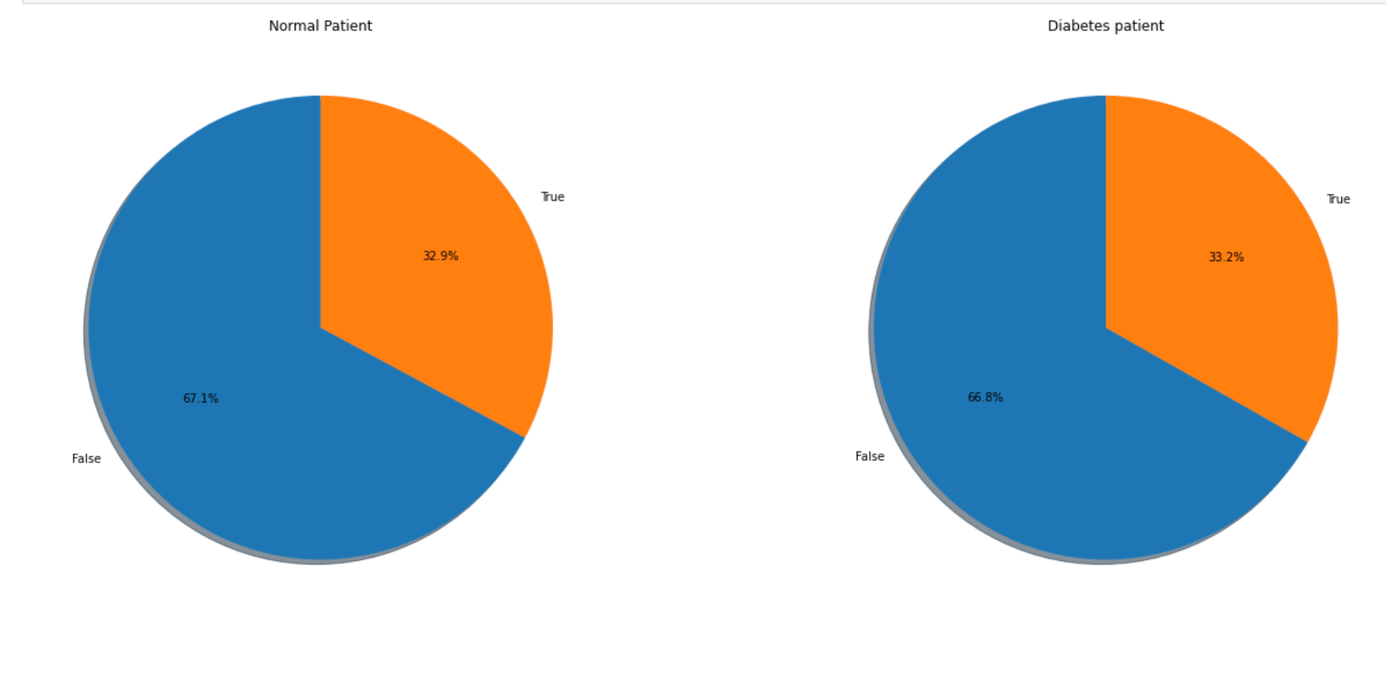
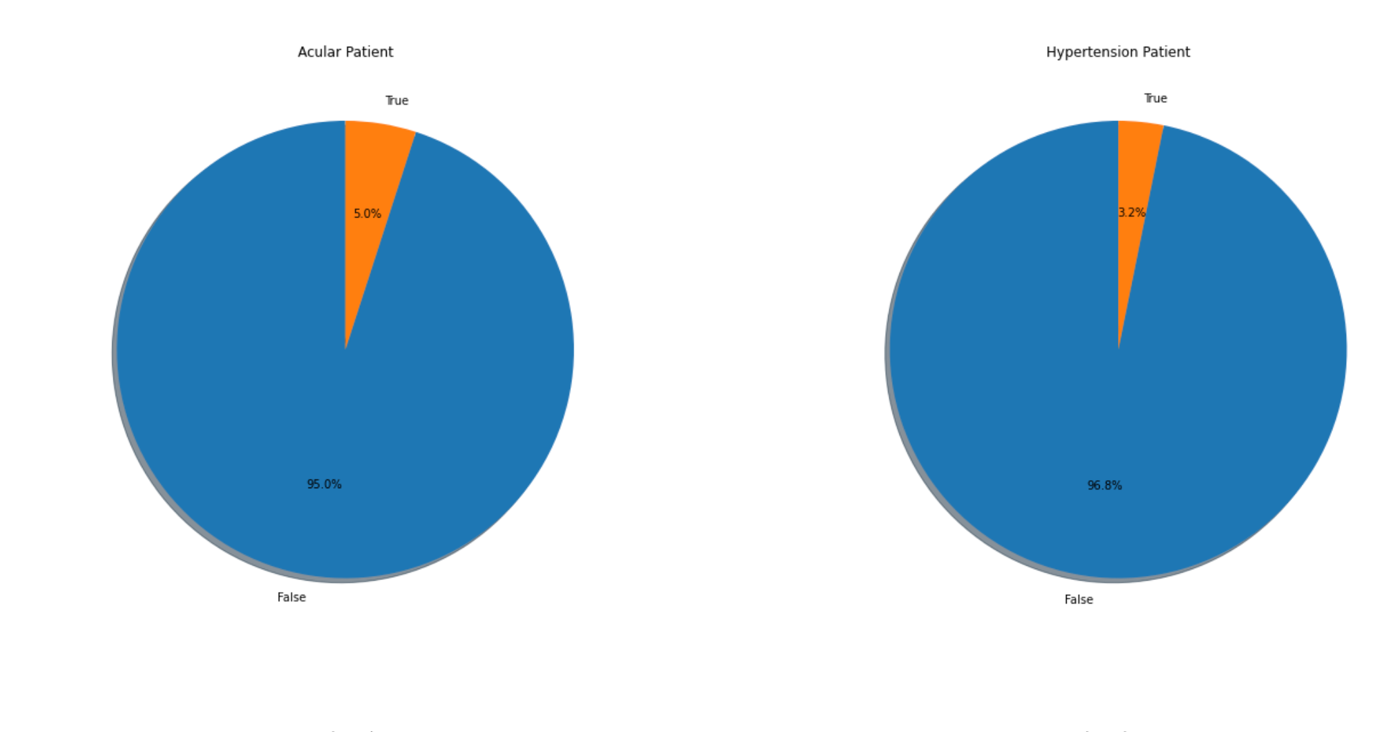


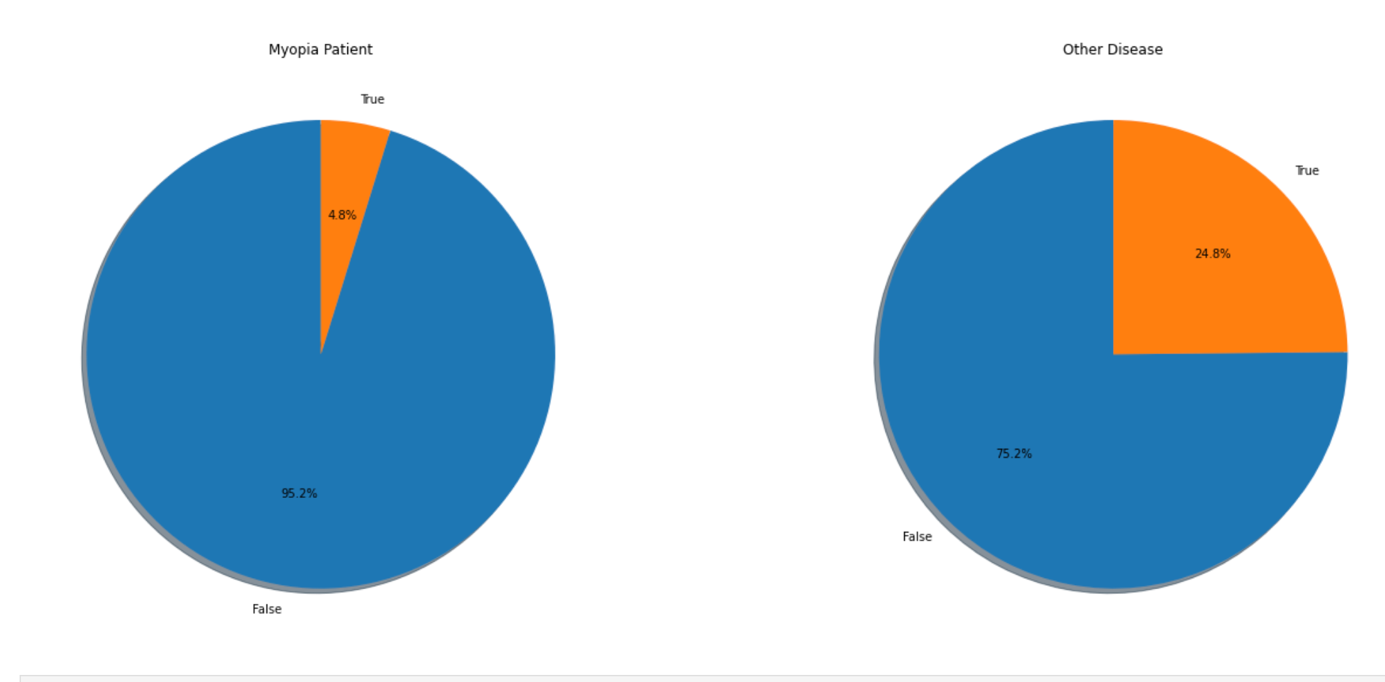
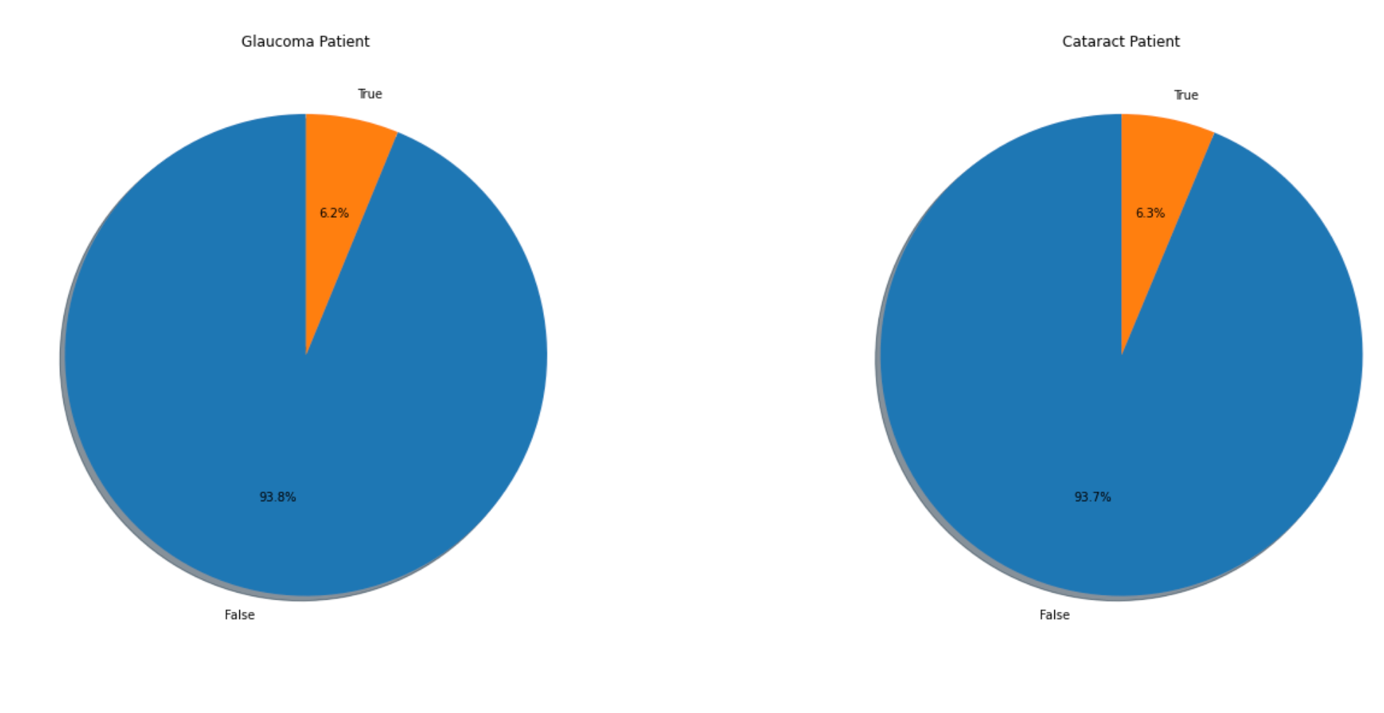
**Figure1**



**Figure 2**

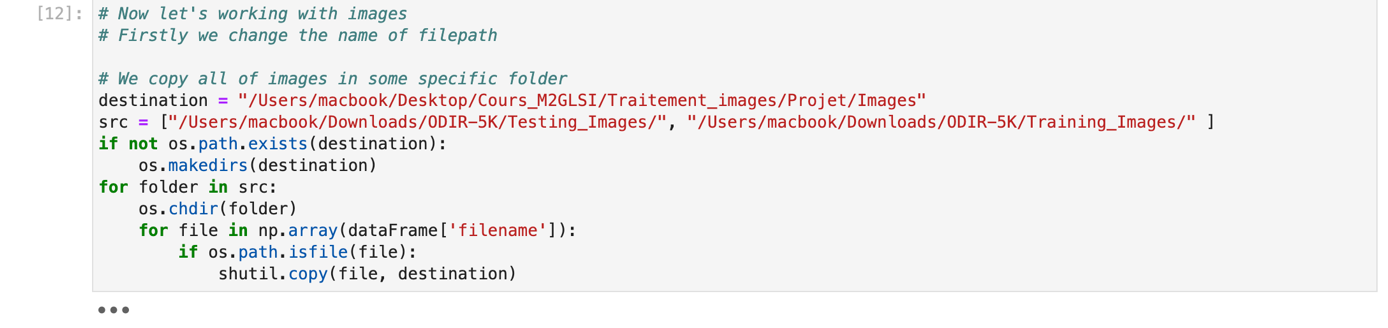
Ensuite là on étudie la composition des données, dans le dataset on a plusieurs type de maladies à prédire c’est-à-dire sur les données, il y a certaines qui appartiennent à des personnes atteintes d’hypertension, de l’acular, du cataracte etc. Donc sur l’image **Figure 3** on étudie le pourcentage de la présence de chacune de ces types de maladies. Les patients atteints par l’hypertension sont représentés au niveau de la colonne H par la variable 1, et les personnes saines sont représentés par 0. Et cela est valide pour toutes les types de maladies.





**Figure 3**

Lors du téléchargement des données, les données sont stockées sur le dossier téléchargement et le script suivant nous a permis de déplacé les fichiers téléchargés sur un autre documents

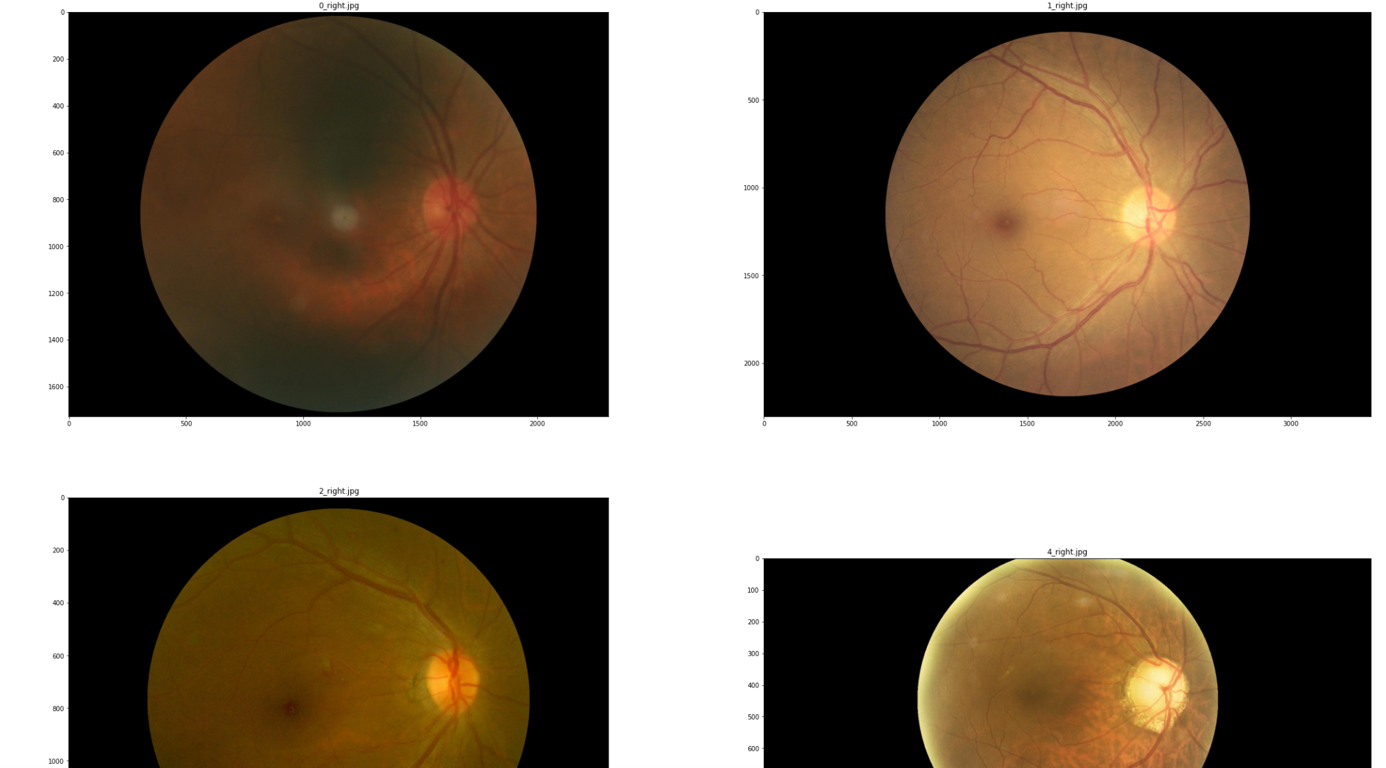


On supprime les colonnes qui ne sont pas nécessaire pour notre étude puisque cette étude dernière porte sur les personnes qui sont atteintes de diabète. Donc on supprime les autres types de maladies pour ne pas récupérer les informations(c’est à dire les photos) **Figure 4** .  
L’étude va porter que sur les personnes atteintes de diabète .

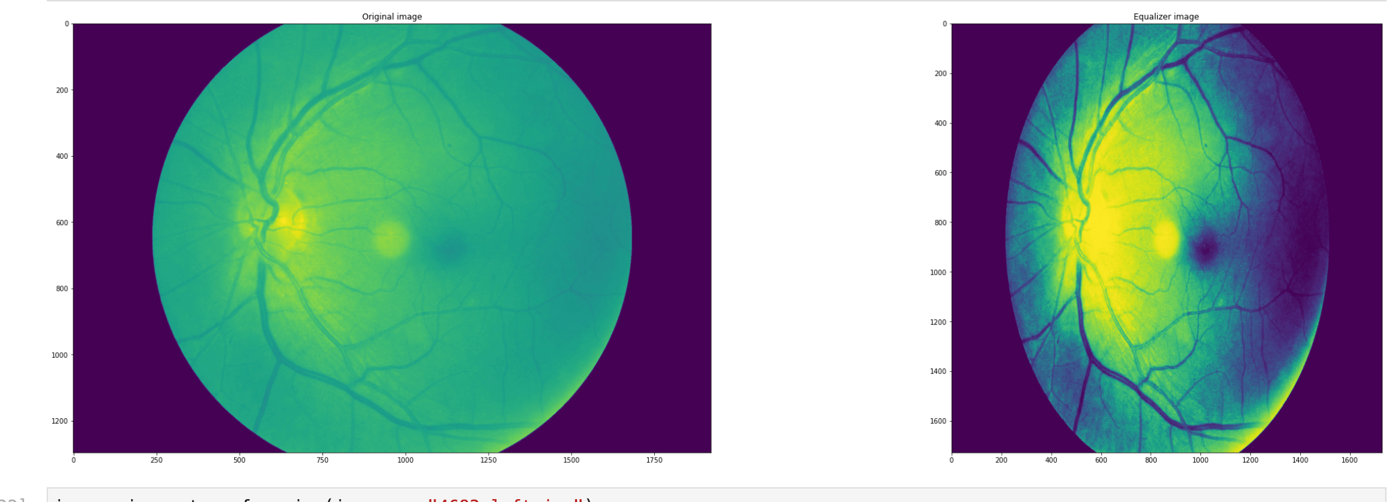


**Figure 4**

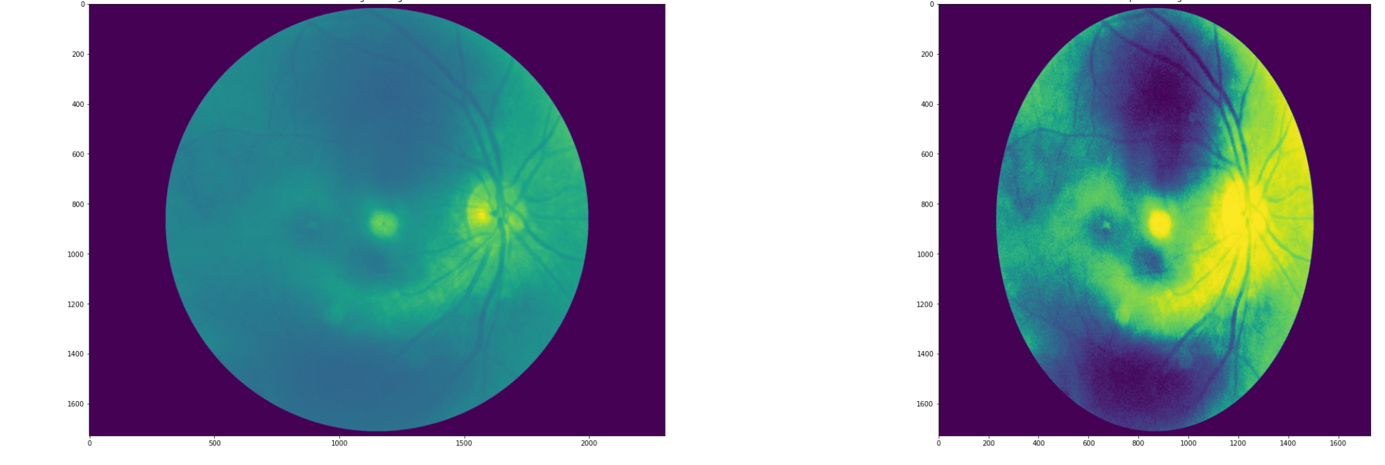
On fait la visualisation de quelques photos, à gauche ceux qui sont malades et a droite ceux qui ne sont pas malade.



En utilisant opencv on a égaliser les pixels, c’est ce qui nous permis d’obtenir les images suivantes. La **figure 5** présente les yeux d’un patient sain et la **figure 6** est l’image du patient malade



**Figure 5**



**Figure 6**

A partir de là on crée(**Figure 7**) une fonction qui va nous permettre de faire le traitement des images , et quand on parle de traitements dans notre cas c’est de récupérer les images sous forme de tableau de dimension et de les stockées sur une variable. La variable **channel** nous permet de récupérer une dimension spécifique de l’image. A noter que toutes les images ont une dimension de 128\*128 après l’application de la fonction pour uniformiser toute les images avec une même dimension.



On construit des modèles pour pouvoir faire des prédictions sur nos données, pour cela on a créé trois différents modèles et chacun d’eux avec des couches différentes. La dernière est faite en utilisant **Data Augmentation** qui est une librairie qui permet elle aussi de faire des predictions.



