Résumé du Processus PCA

Ce résumé présente les résultats de l'analyse en composantes principales (PCA) effectuée sur un jeu de données d'images radiographiques associées à des masques. L'objectif de l'analyse était de réduire la dimensionnalité des données tout en conservant un maximum d'informations pertinentes, à savoir en expliquant une grande partie de la variance des données.

# Résultats

Voici les résultats de la PCA appliquée aux caractéristiques extraites des images :

1. Variance expliquée par chaque composante : [0.33753577, 0.17235869, 0.10687735, 0.06943919, 0.06053101, 0.04964212, 0.03888787, 0.02735815, 0.0207339, 0.01482933, 0.01200447]

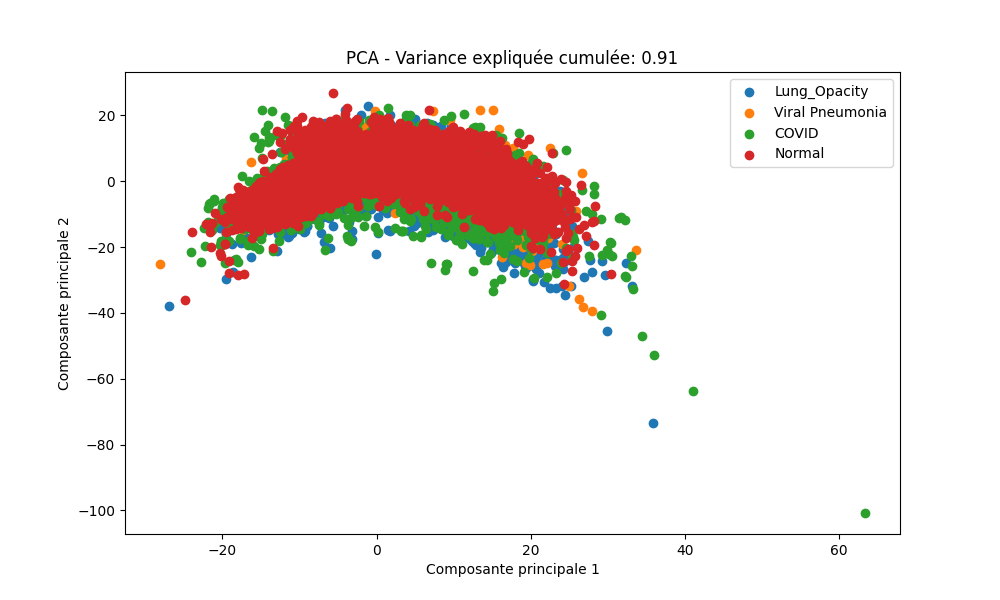
2. Variance totale expliquée : 0.9101978518735634 (soit environ 91% de la variance totale expliquée par 11 composantes principales).

3. Dimension initiale des données : 256 (nombre de caractéristiques extraites de chaque image).

4. Dimension après PCA (nombre de composantes choisies) : 11 composantes.

Grâce à la PCA, la dimension des données a été réduite **de 256 à 11 composantes principales,** **tout en expliquant environ 91% de la variance présente dans les données d'origine**. Cela permet de maintenir l'essentiel de l'information tout en simplifiant les données, ce qui est crucial pour les étapes suivantes d'analyse ou de modélisation.

# Graphiques

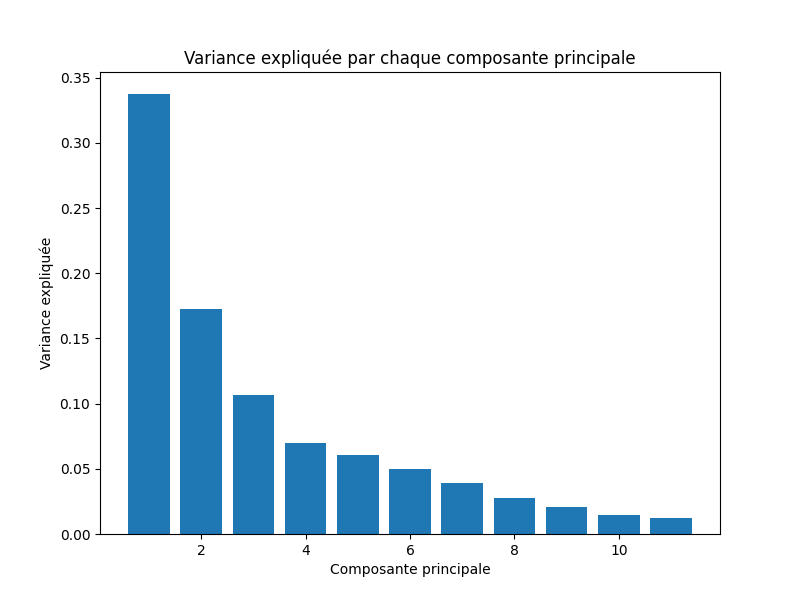
**Variance expliquée par chaque composante :**

Ce graphique en 2D montre la distribution des points projetés sur les deux premières composantes principales (PC1 et PC2). Les différentes classes (COVID, Normal, Lung\_Opacity, Viral Pneumonia) sont visibles et bien séparées. Cela indique que les deux premières composantes principales contiennent une grande partie de l'information discriminante nécessaire pour différencier les classes dans le jeu de données. Cette séparation est essentielle pour des algorithmes de classification ultérieurs.

**Une image contenant texte, capture d’écran, ligne, Tracé

Le contenu généré par l’IA peut être incorrect.Illustration de la réduction de dimension (avant vs après PCA) :**

Ce graphique montre la réduction de la dimensionnalité avant et après l'application de la PCA. La ligne bleue représente la variance cumulée expliquée par toutes les dimensions (jusqu'à 256 dimensions), tandis que la ligne rouge marque la réduction des dimensions après PCA, où 11 composantes sont choisies. Cela démontre clairement que la PCA réduit efficacement la complexité des données tout en conservant 91 % de la variance d'origine. La réduction de la dimension permet ainsi de simplifier les calculs tout en préservant l'essentiel de l'information.

**PCA - Variance expliquée cumulée :**

Le graphique montre la proportion de la variance expliquée par chaque composante principale (PC). On observe que la première composante (PC1) explique une part significative de la variance (environ 33,75 %), suivie de la deuxième composante (PC2) avec environ 17,24 %. Ces deux premières composantes couvrent à elles seules plus de 50 % de la variance. La contribution des autres composantes est de plus en plus faible. Cela montre que la réduction de dimension à 11 composantes principales est un choix judicieux, car on conserve 91 % de la variance totale.