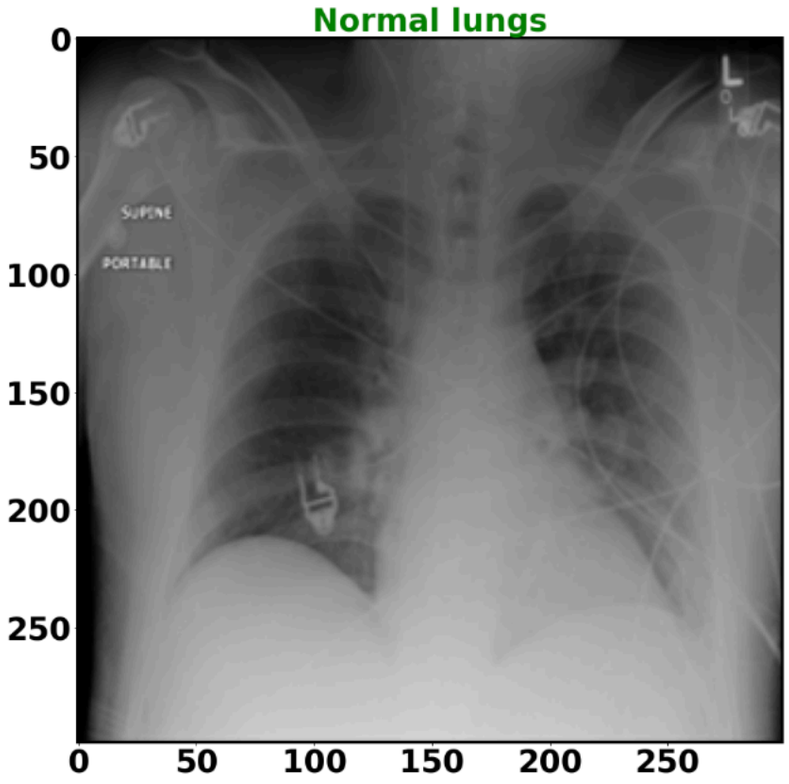
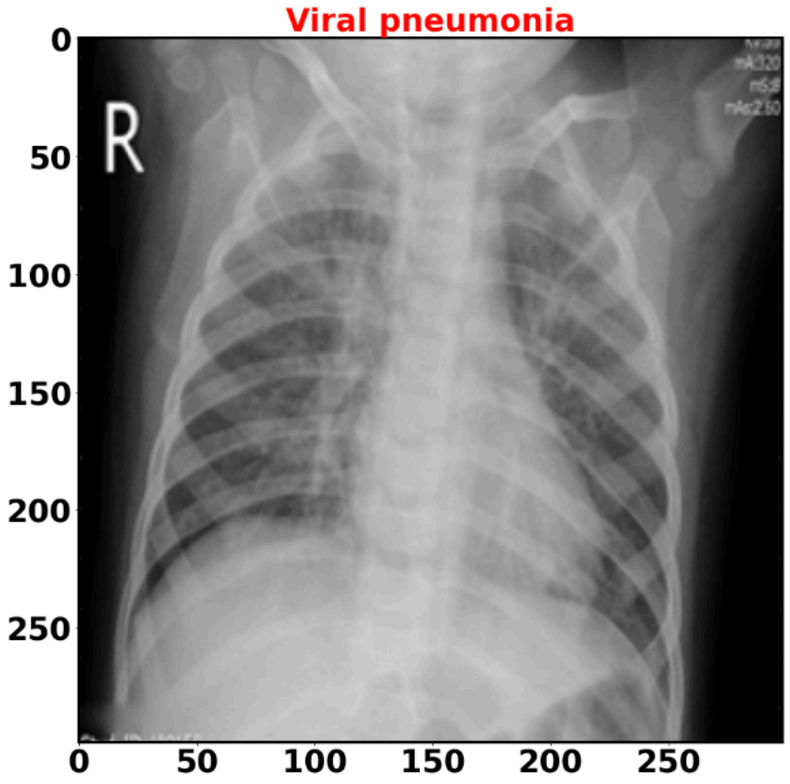
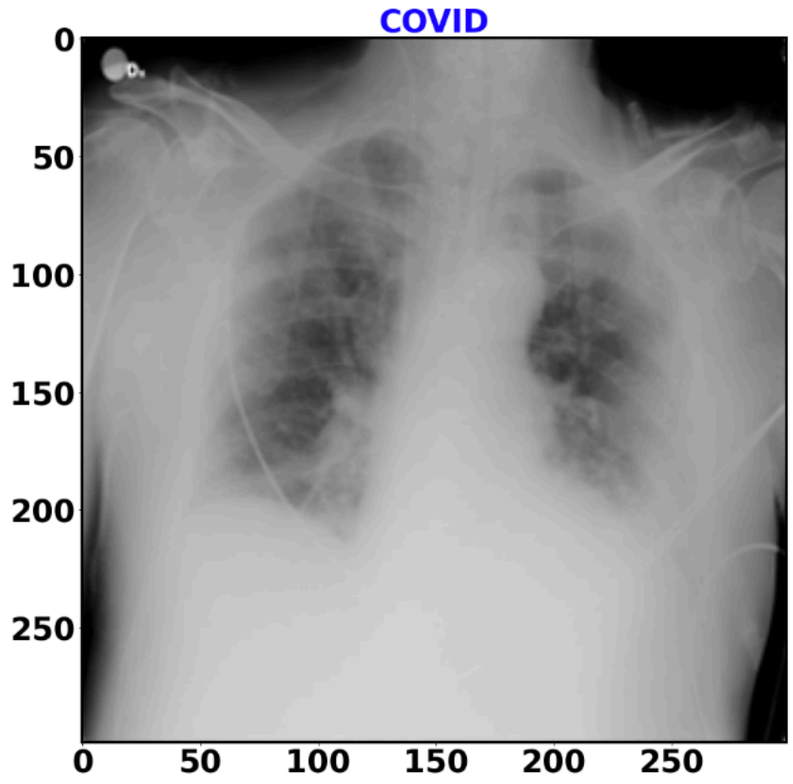
Classifying COVID, viral pneumonia and normal lungs from radiography dataset
Soutick Saha

Source dataset - M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M. T. Islam, “Can AI help in screening Viral and COVID-19 pneumonia?” IEEE Access, Vol. 8, 2020, pp. 132665 - 132676.
-Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S.B.A., Islam, M.T., Maadeed, S.A., Zughaier, S.M., Khan, M.S. and Chowdhury, M.E., 2020. Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-ray Images.
Data downloaded from - https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
Code - https://github.com/souticksaha21/Predicting-COVID-19-positivity-from-radiography-images/blob/main/Predicting_COVID_positivity_from_radiography_images.ipynb

Different lung types

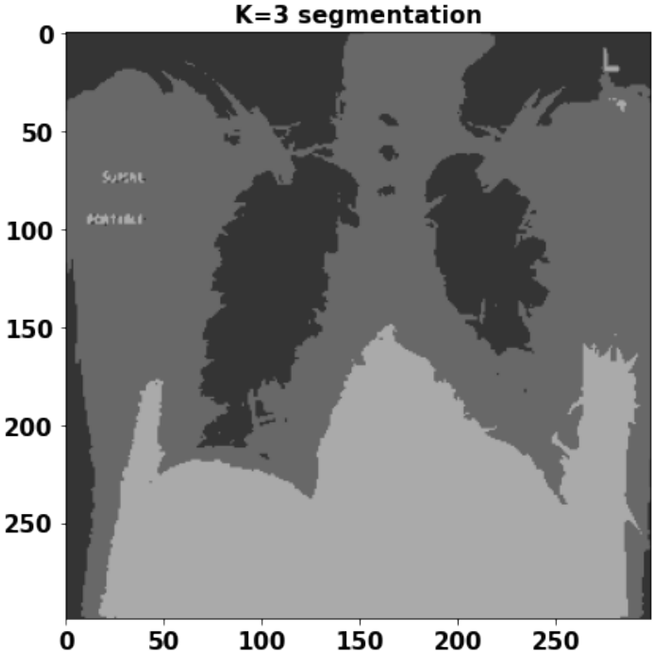


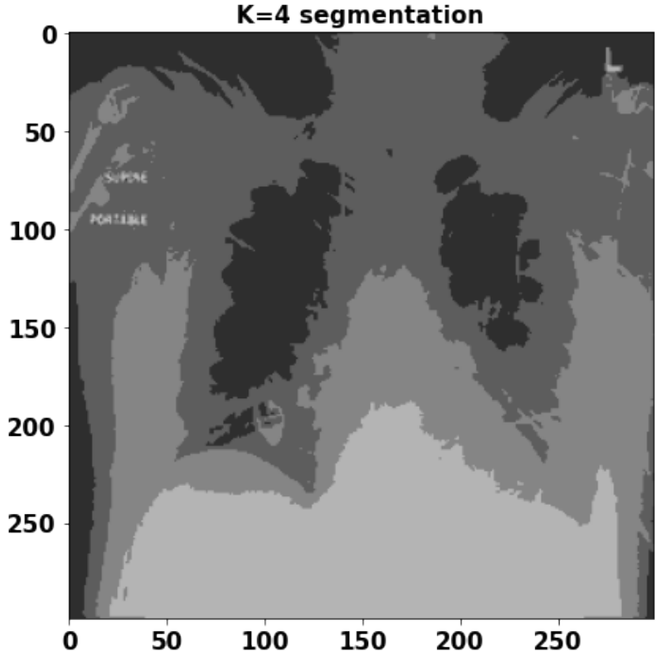




Effect of segmentation using K-means clustering

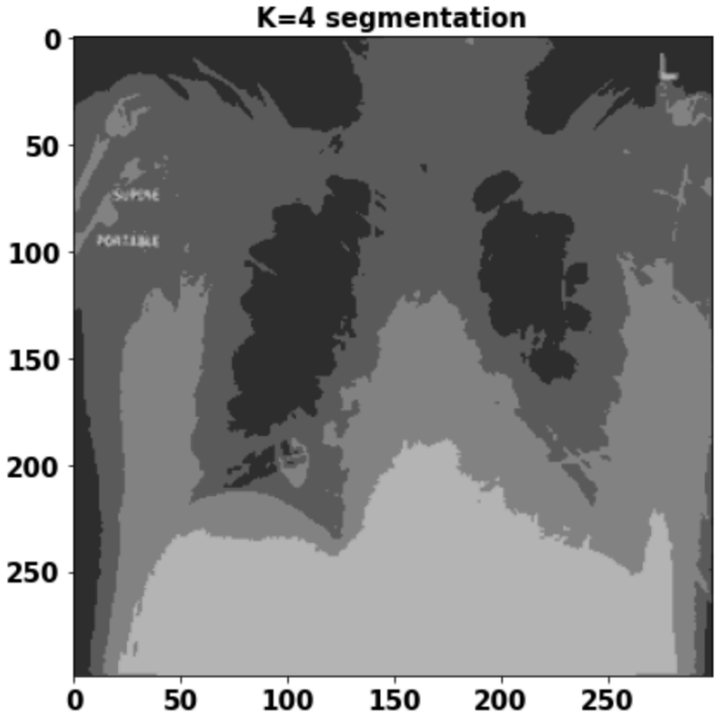
Normal lungs



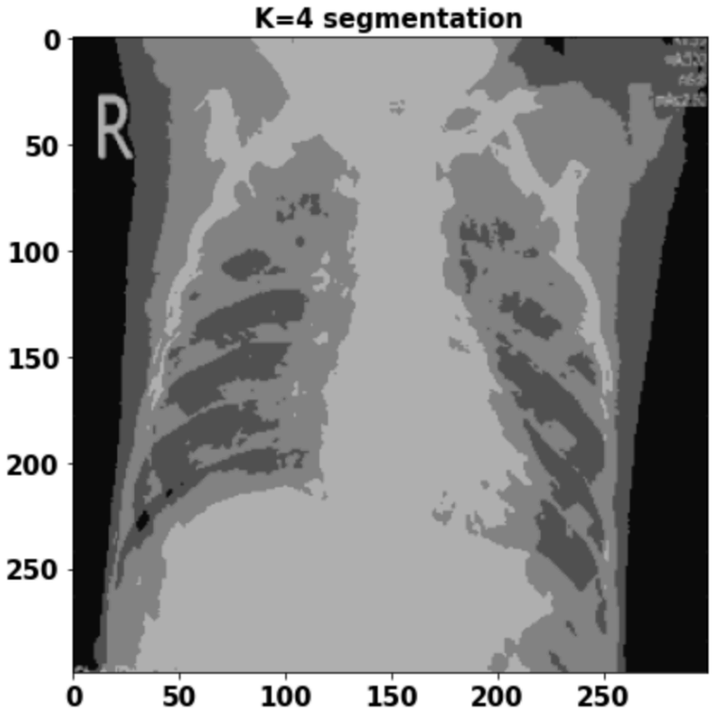


Effect of segmentation using K-means clustering

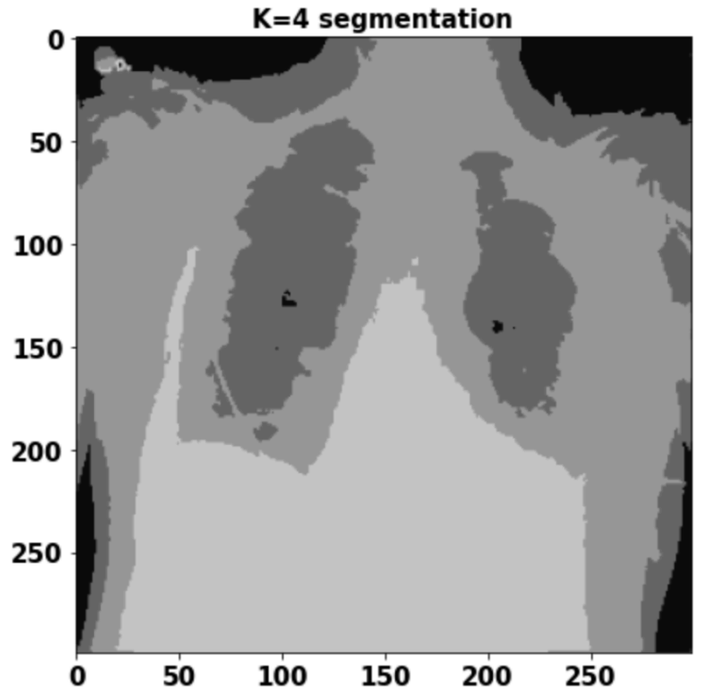
Normal Lungs



Viral pneumonia

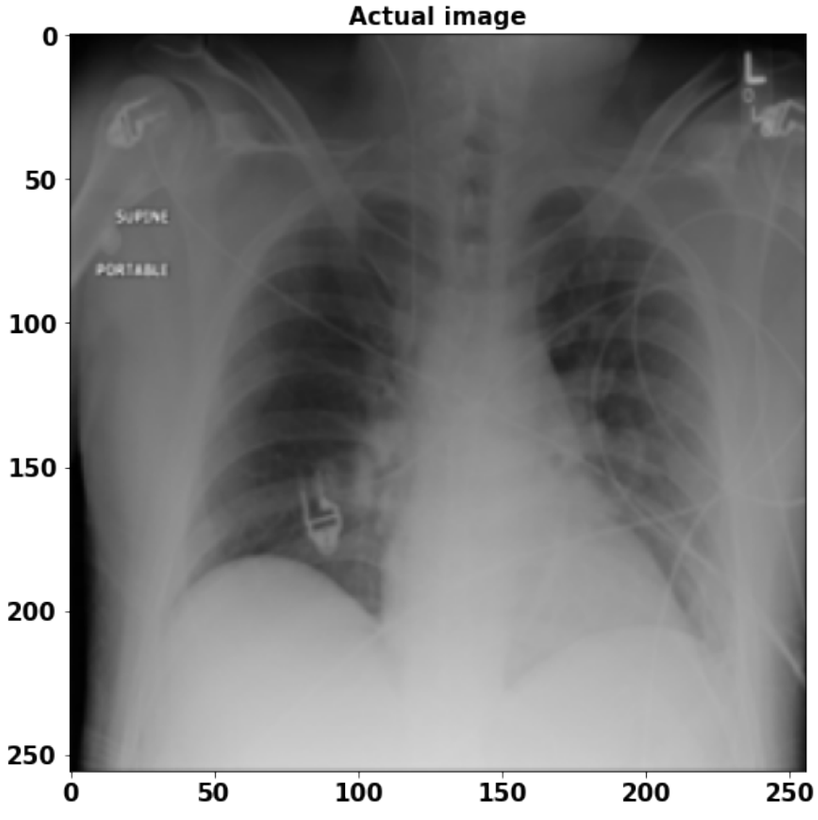


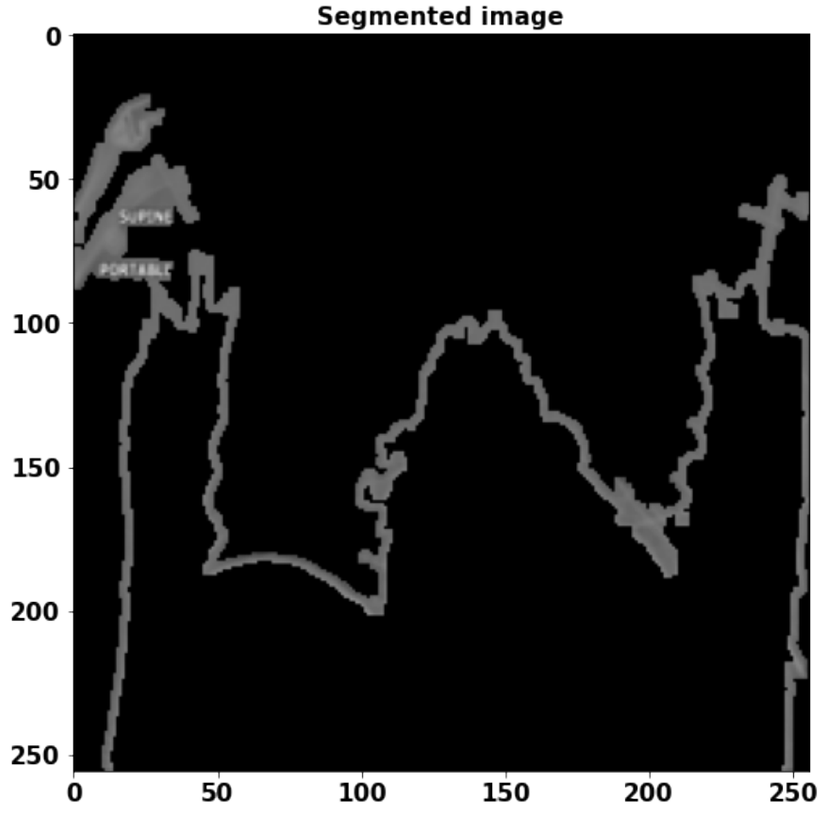
COVID



Effect of image segmentation using contour

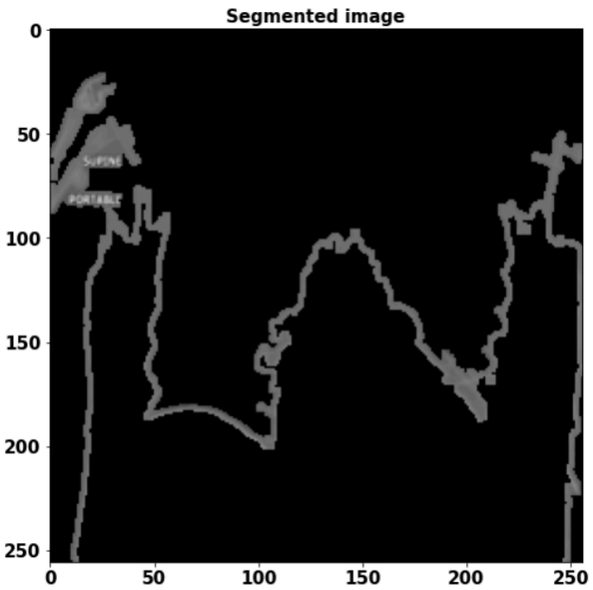
Normal lungs



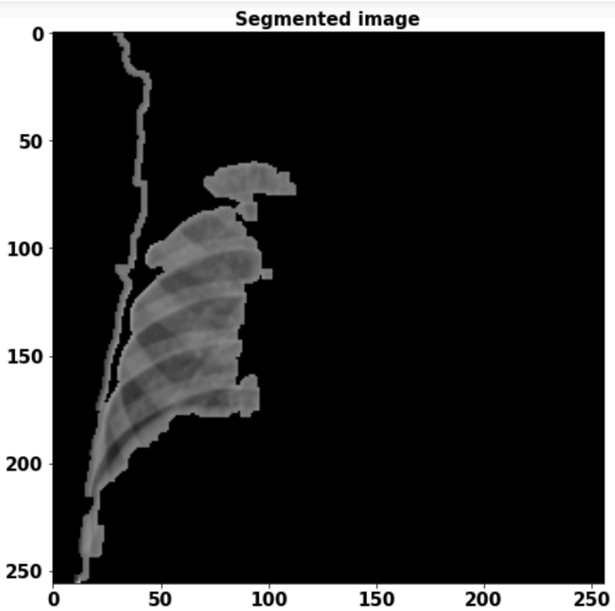


Effect of image segmentation using contour

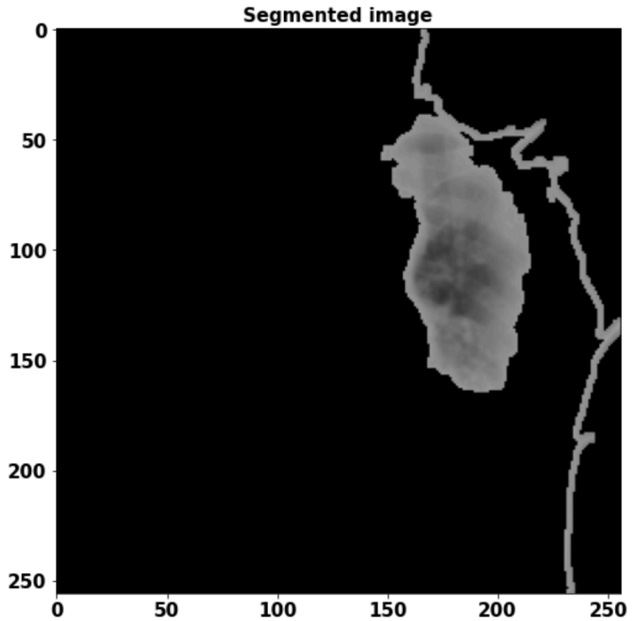
Normal Lungs



Viral pneumonia

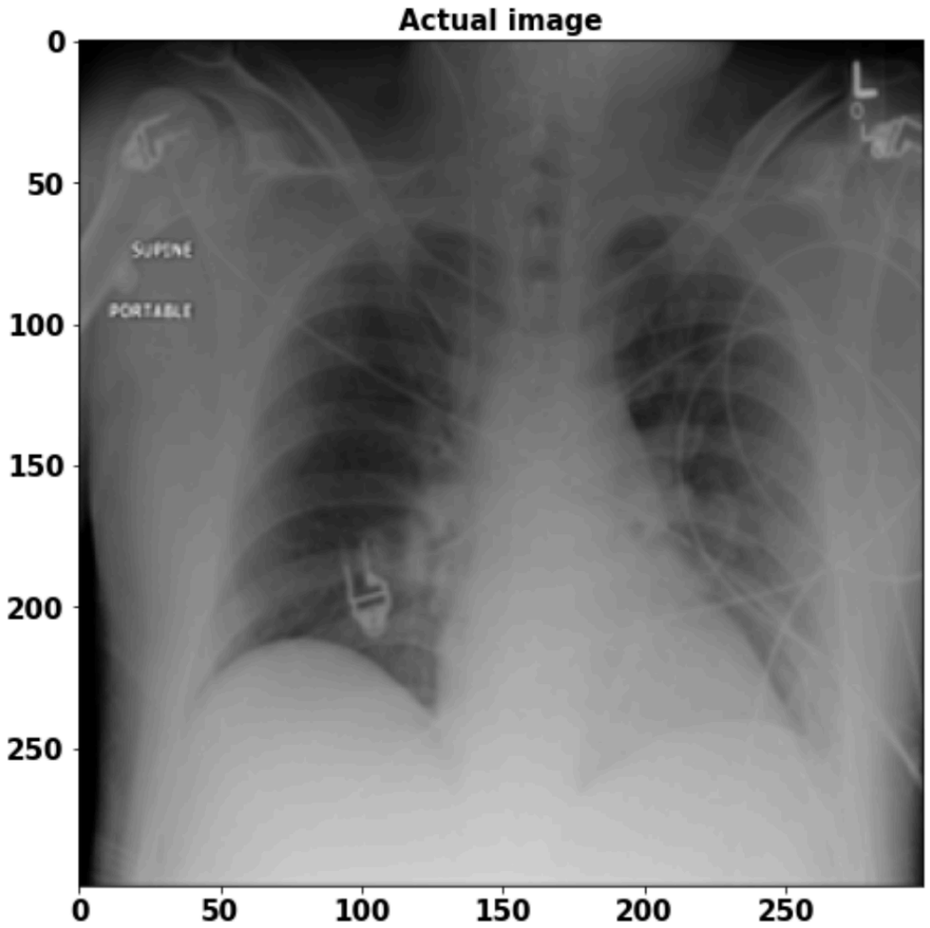


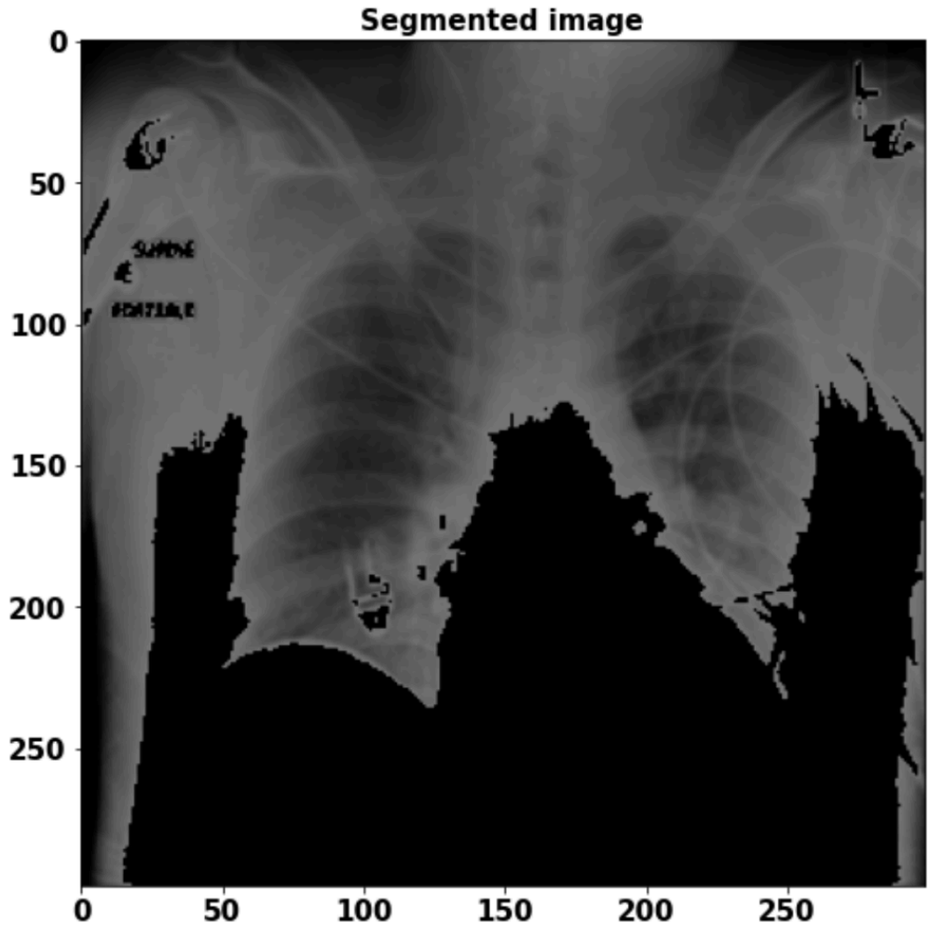
COVID



Effect of image segmentation using thresholding

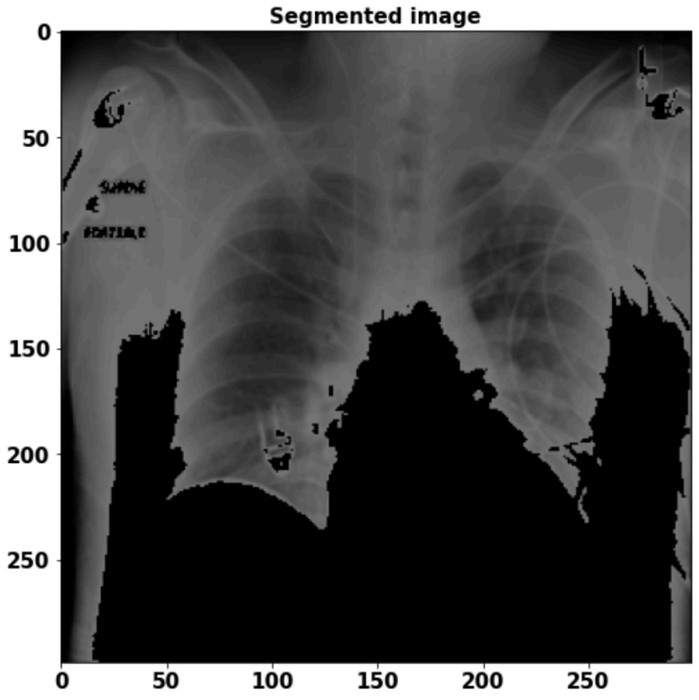
Normal Lungs



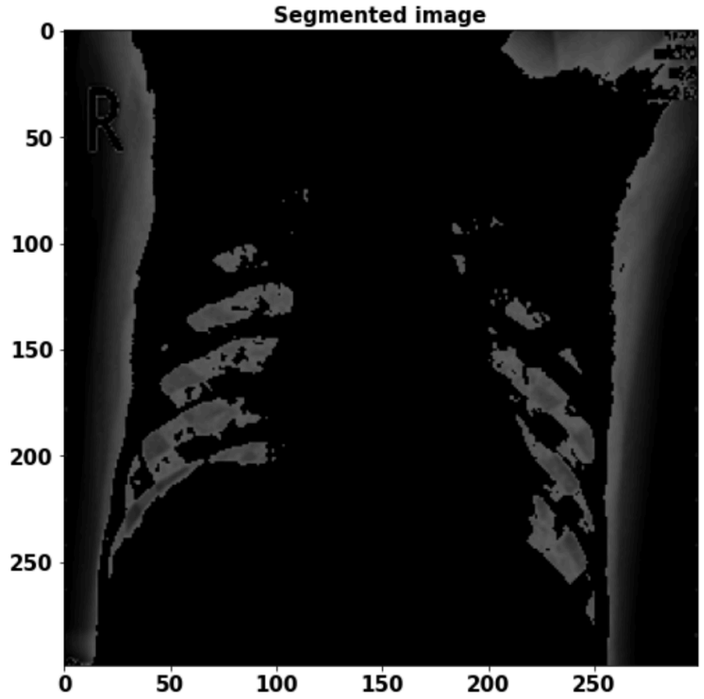


Effect of image segmentation using thresholding

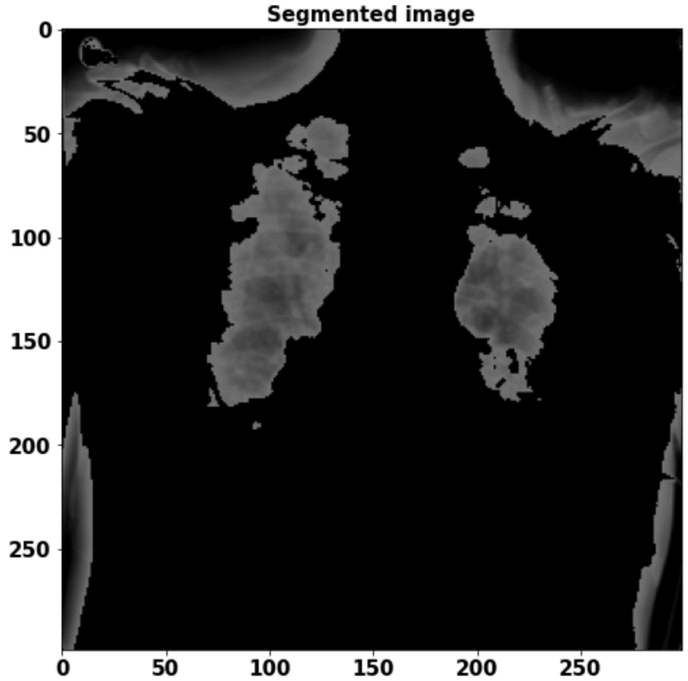
Normal Lungs



Viral pneumonia

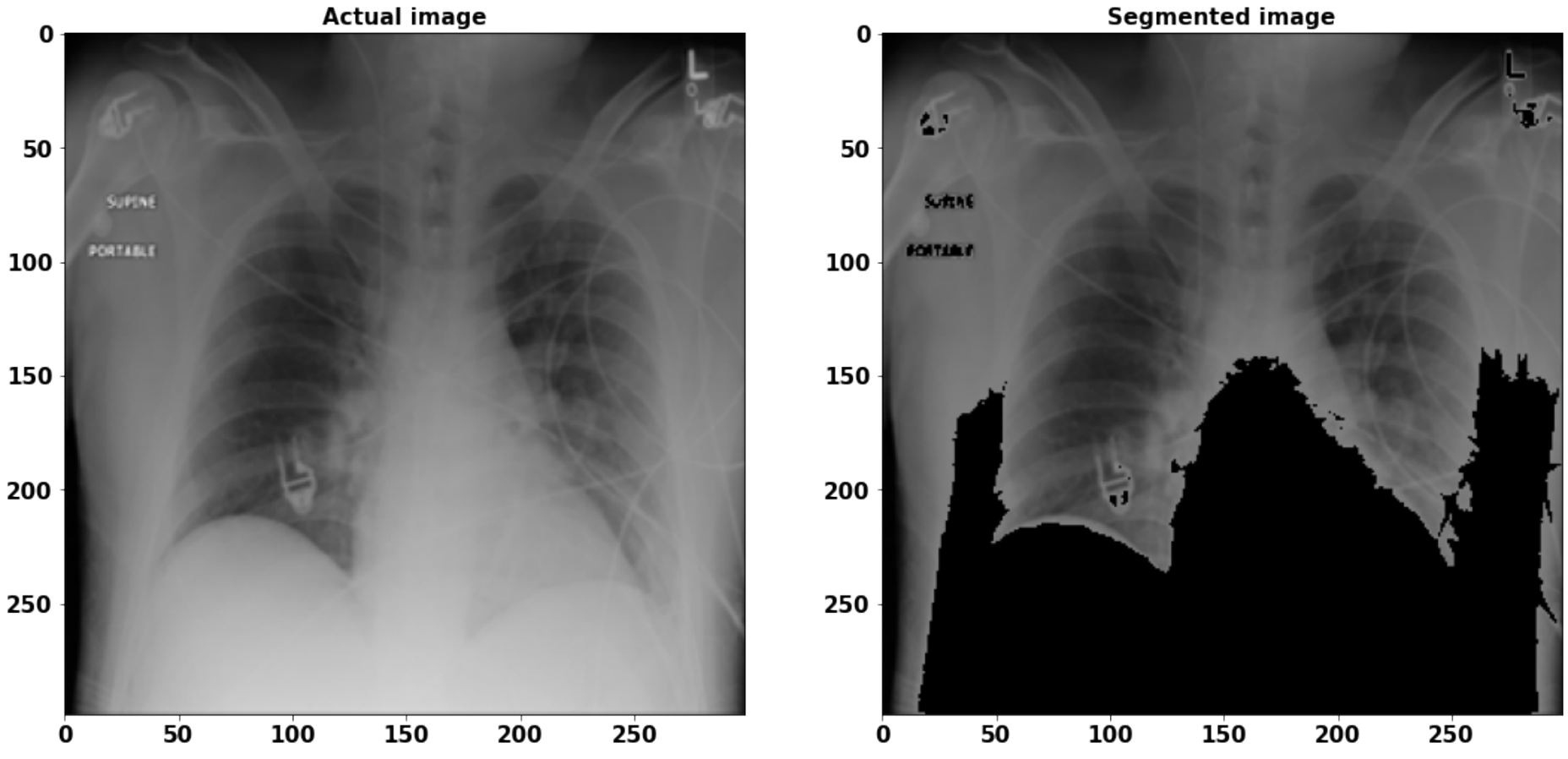


COVID



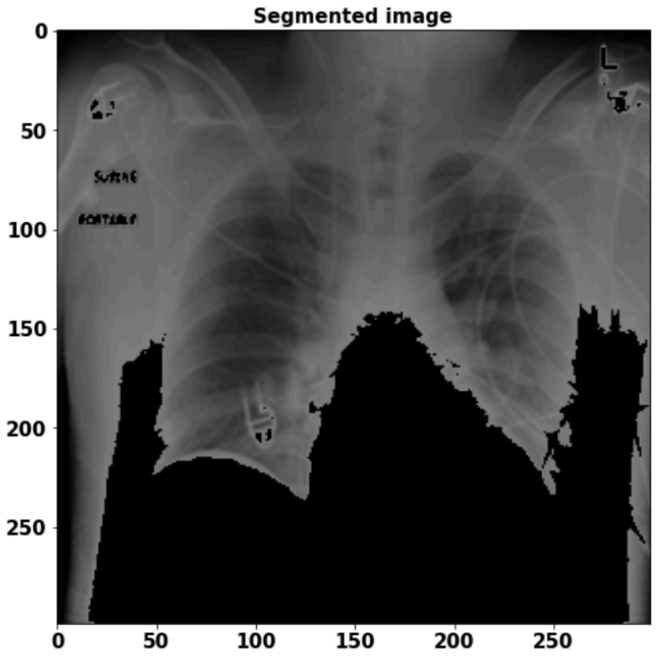
Effect of masking

Normal lungs

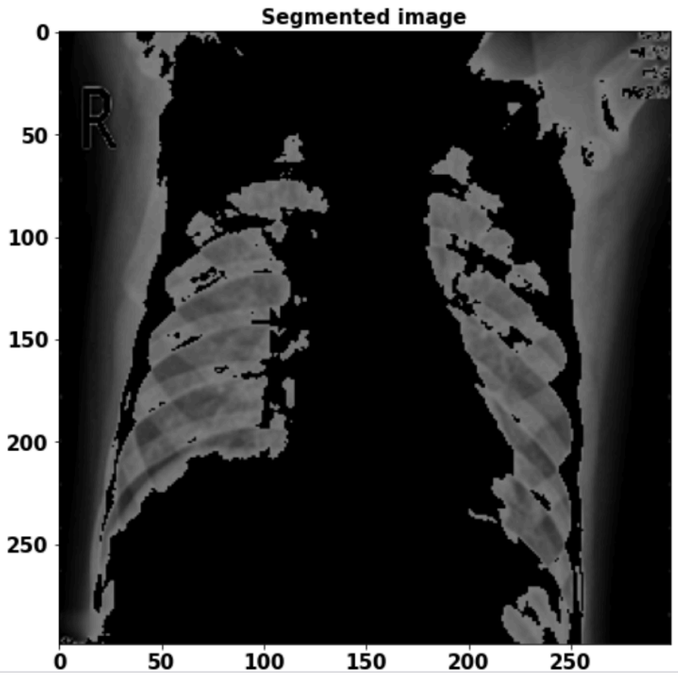


Effect of masking

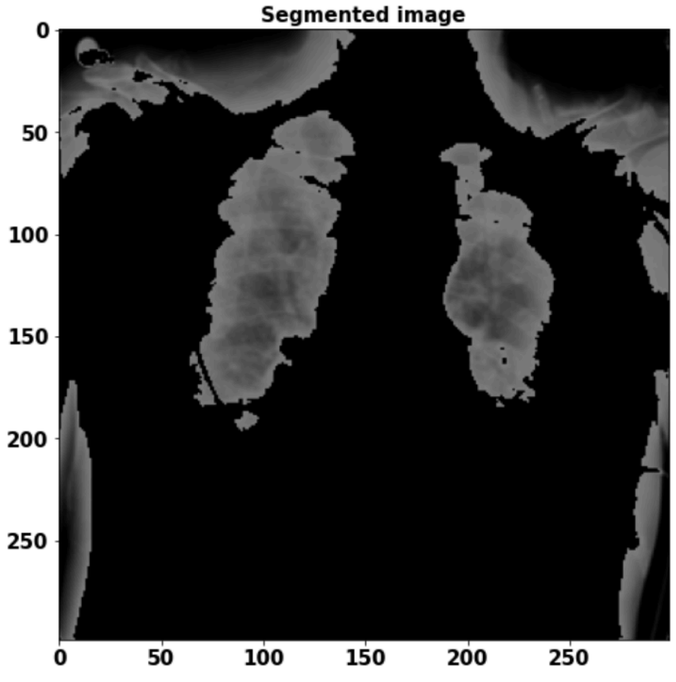
Normal Lungs



Viral pneumonia



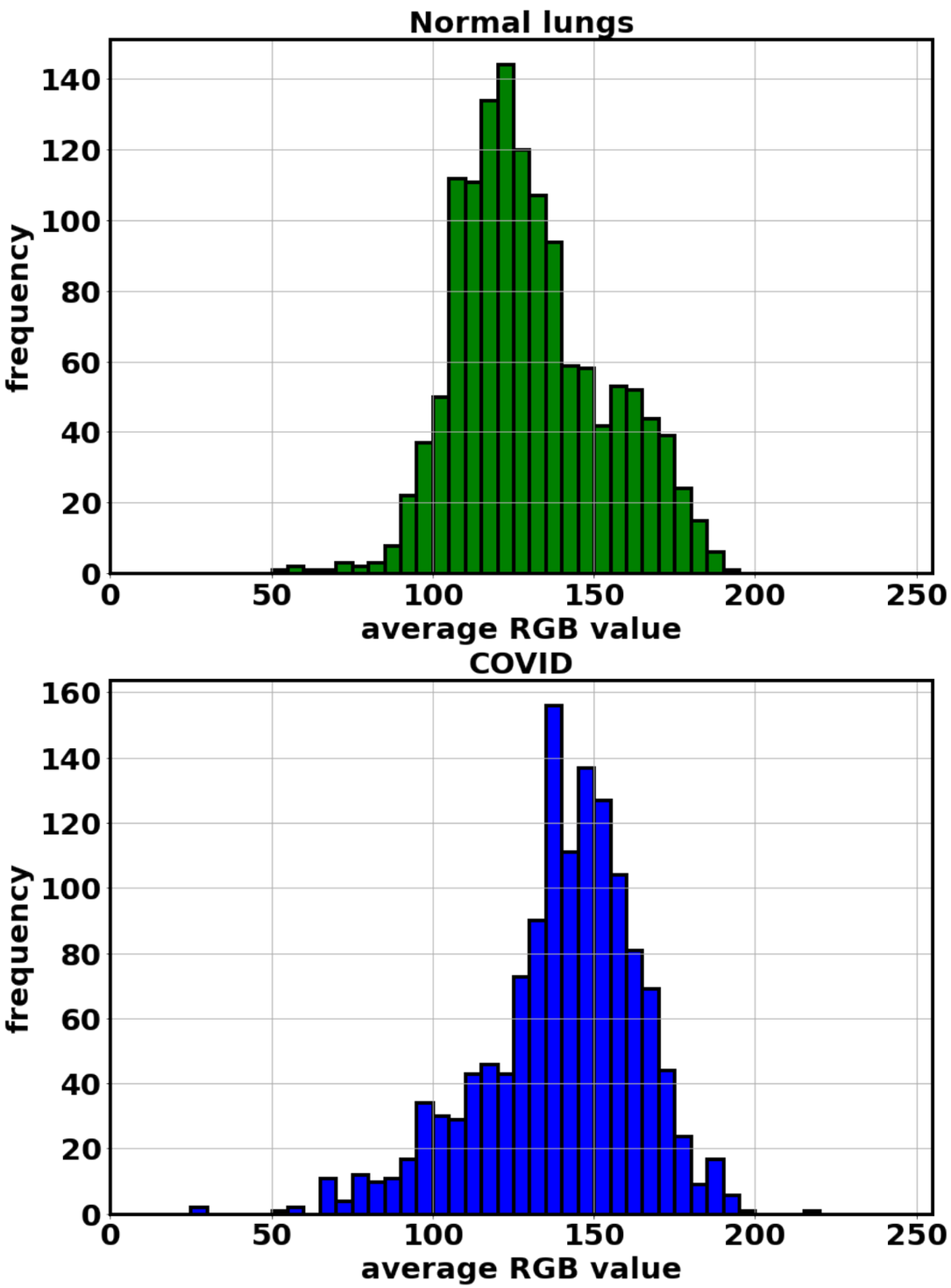
COVID



But these are one image from individual cases. What are the qualitative distinctive features if we analyze ~1300 images for both cases?

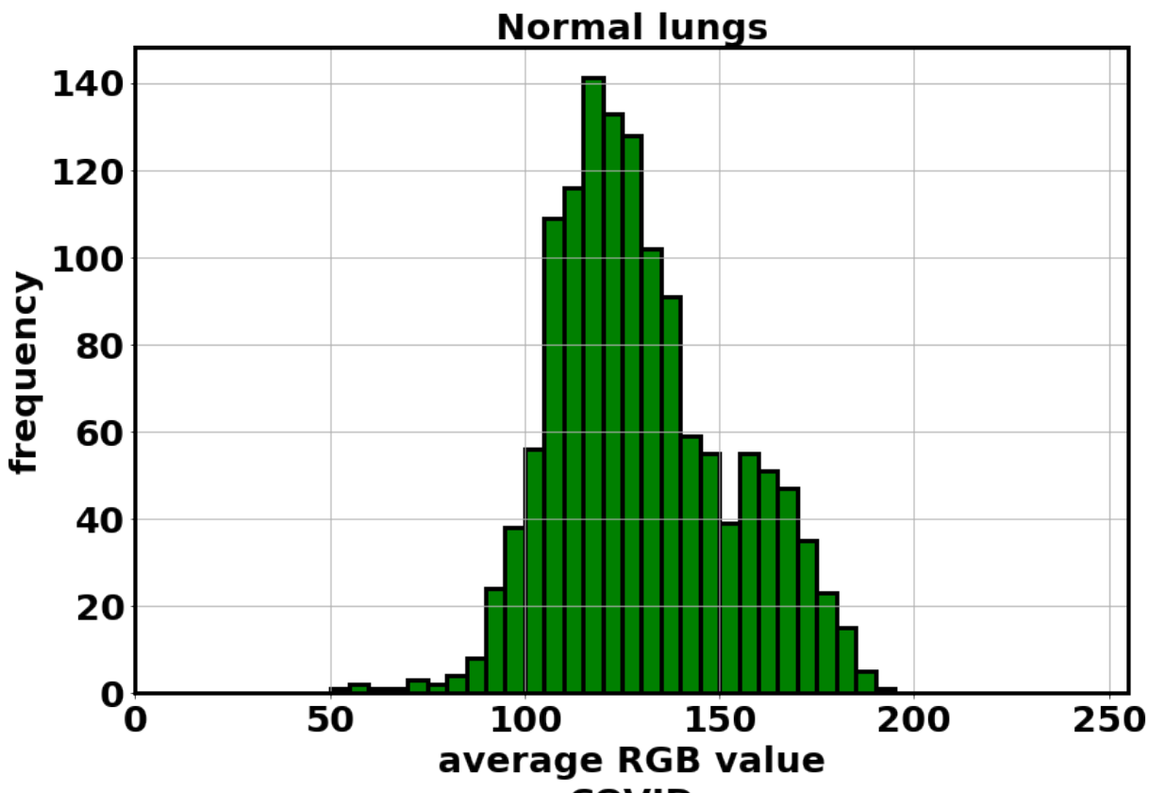
Features when considering average RGB levels for each image before any alteration

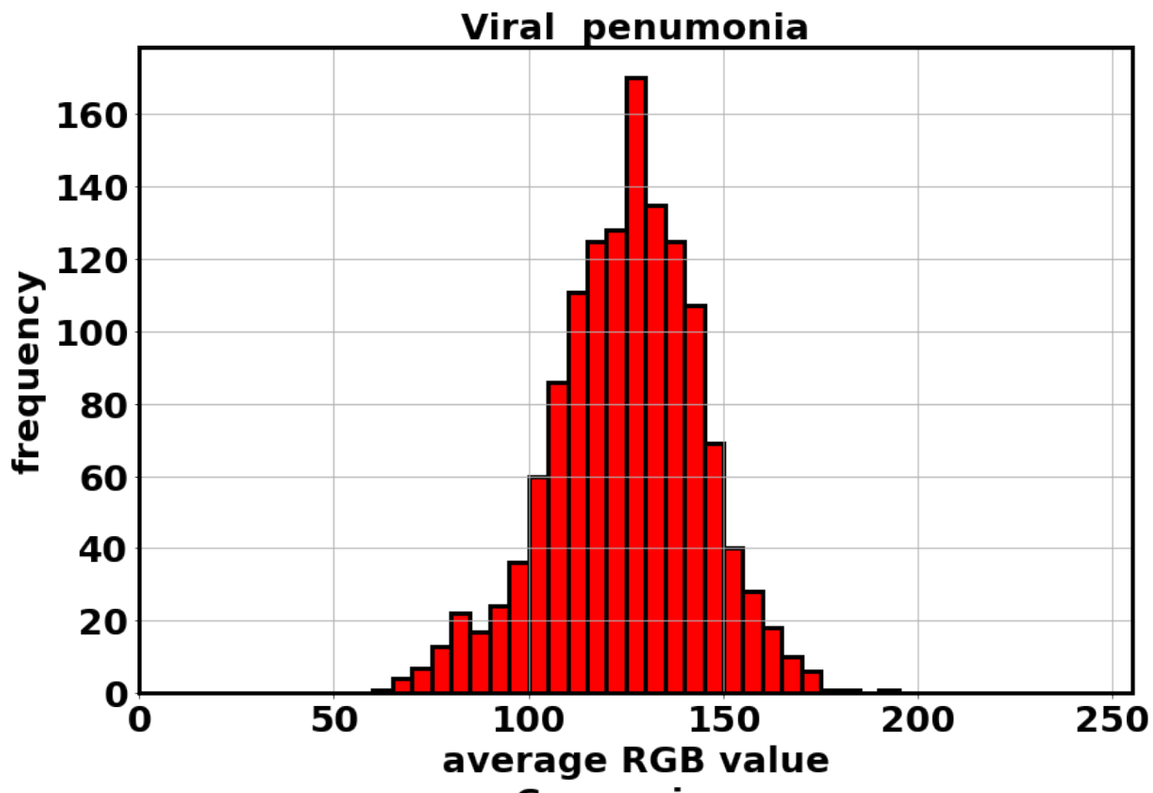
Image before segmentation

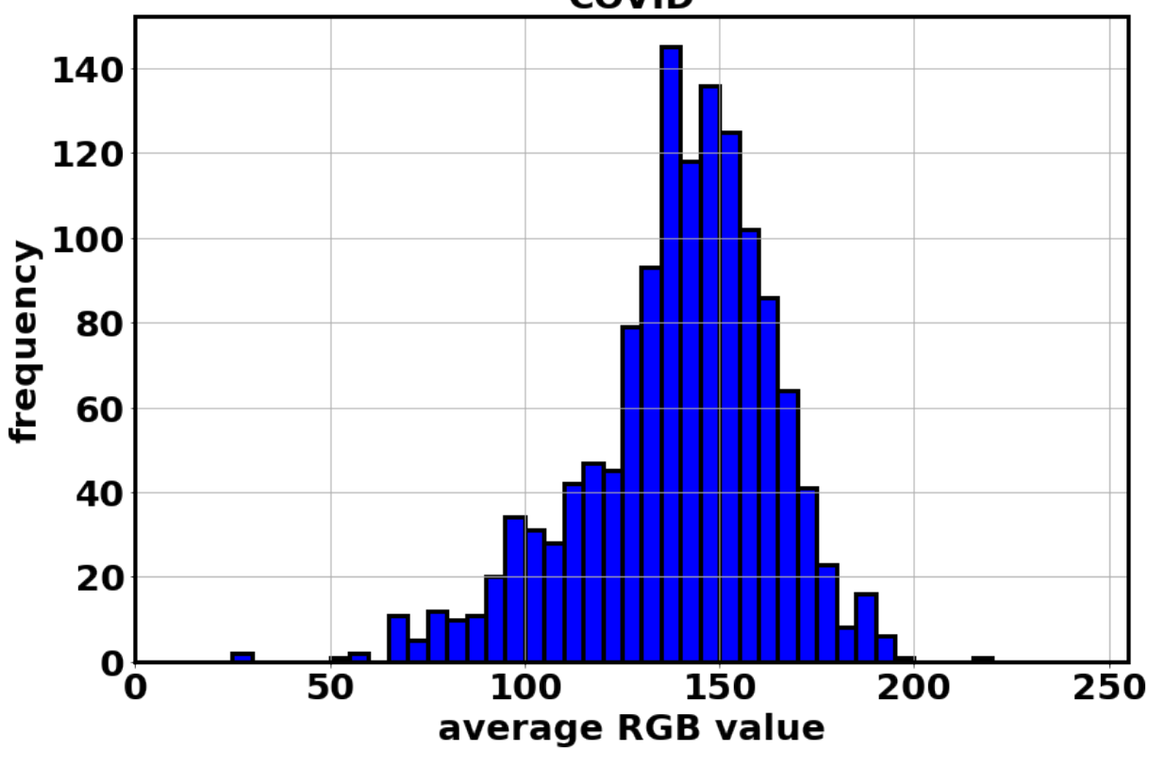


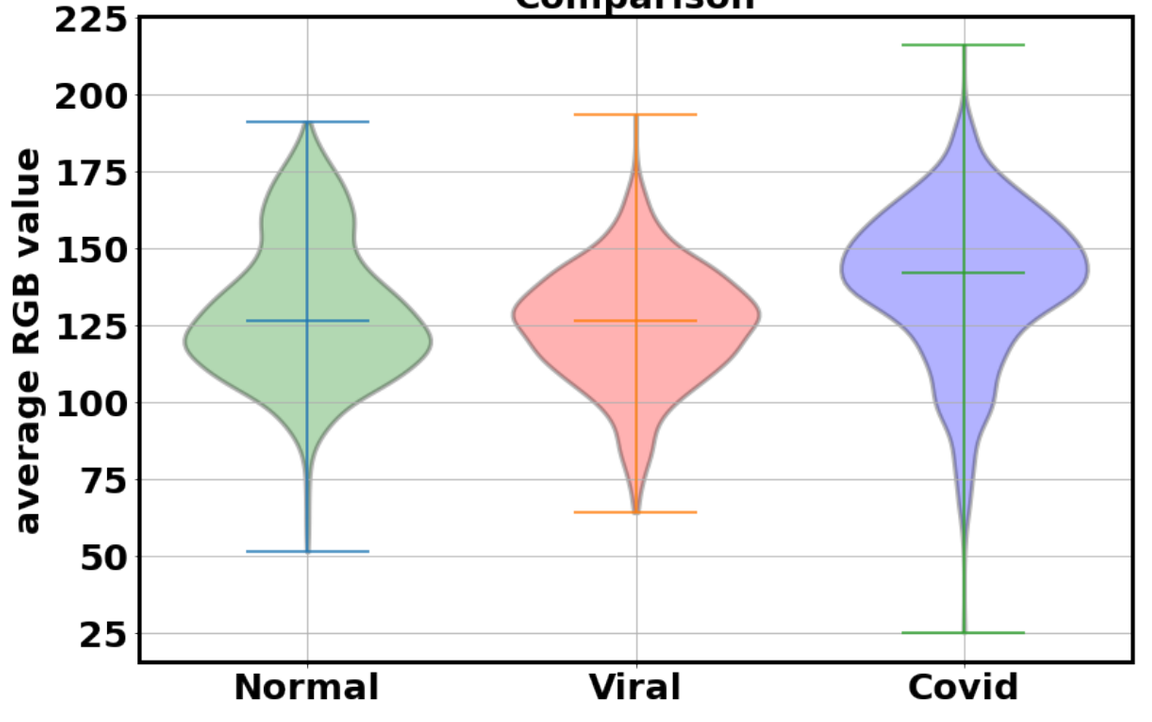
Features considering average RGB levels for each image with k=4 means clustering

k=4 mean clustering segmentation



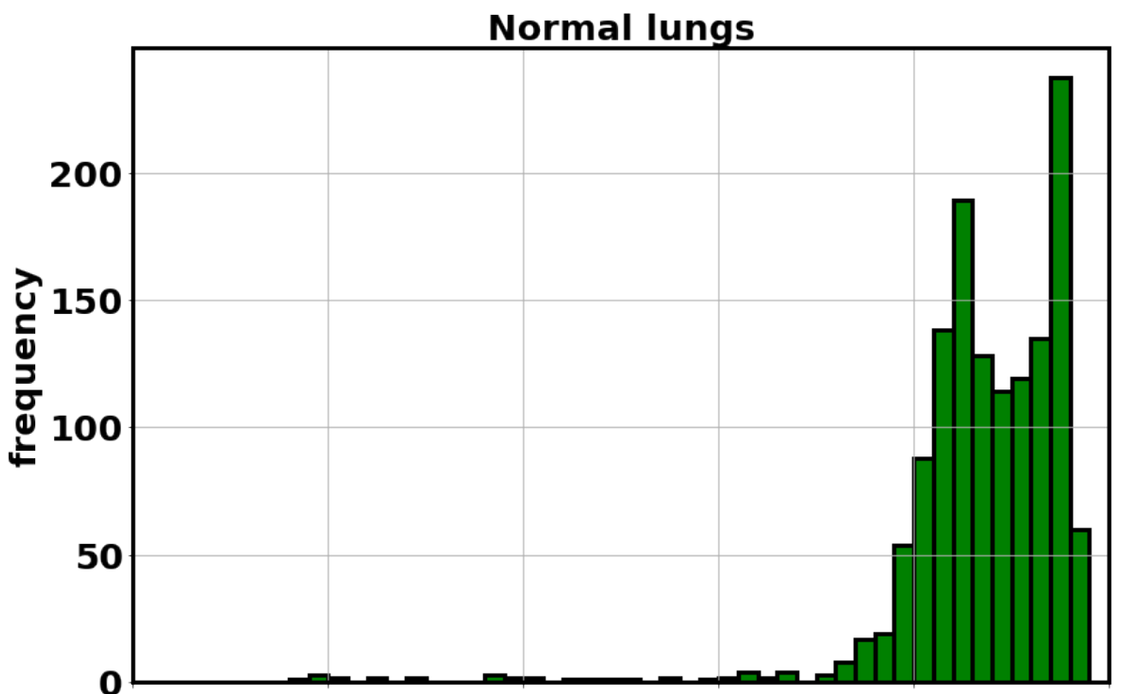


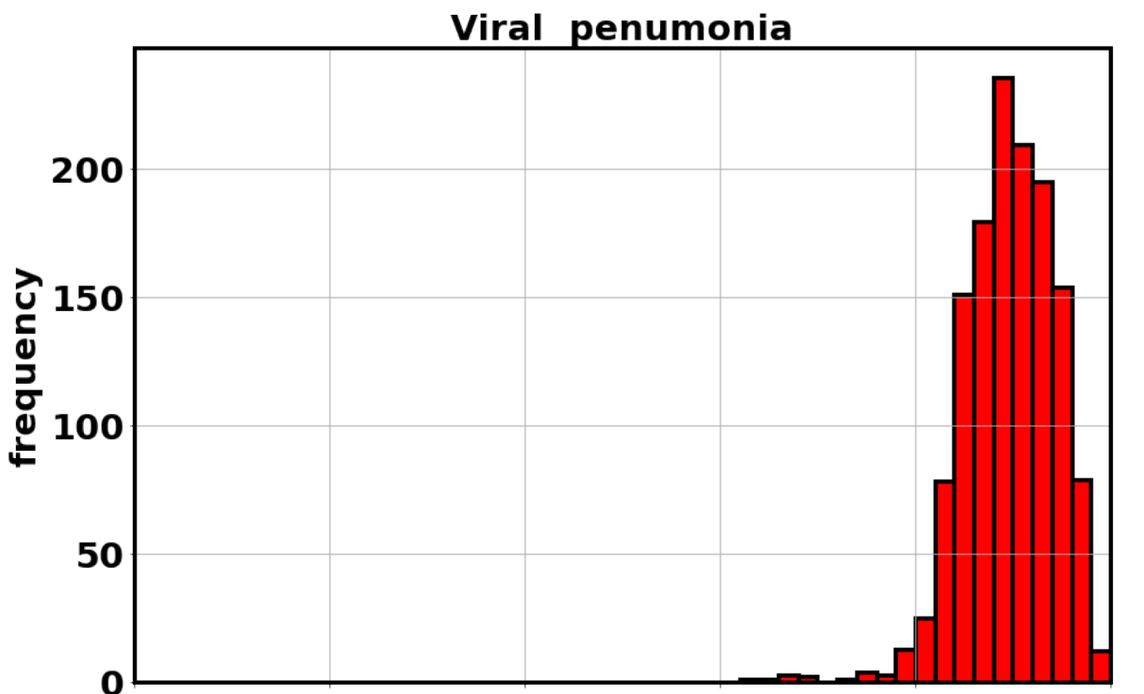


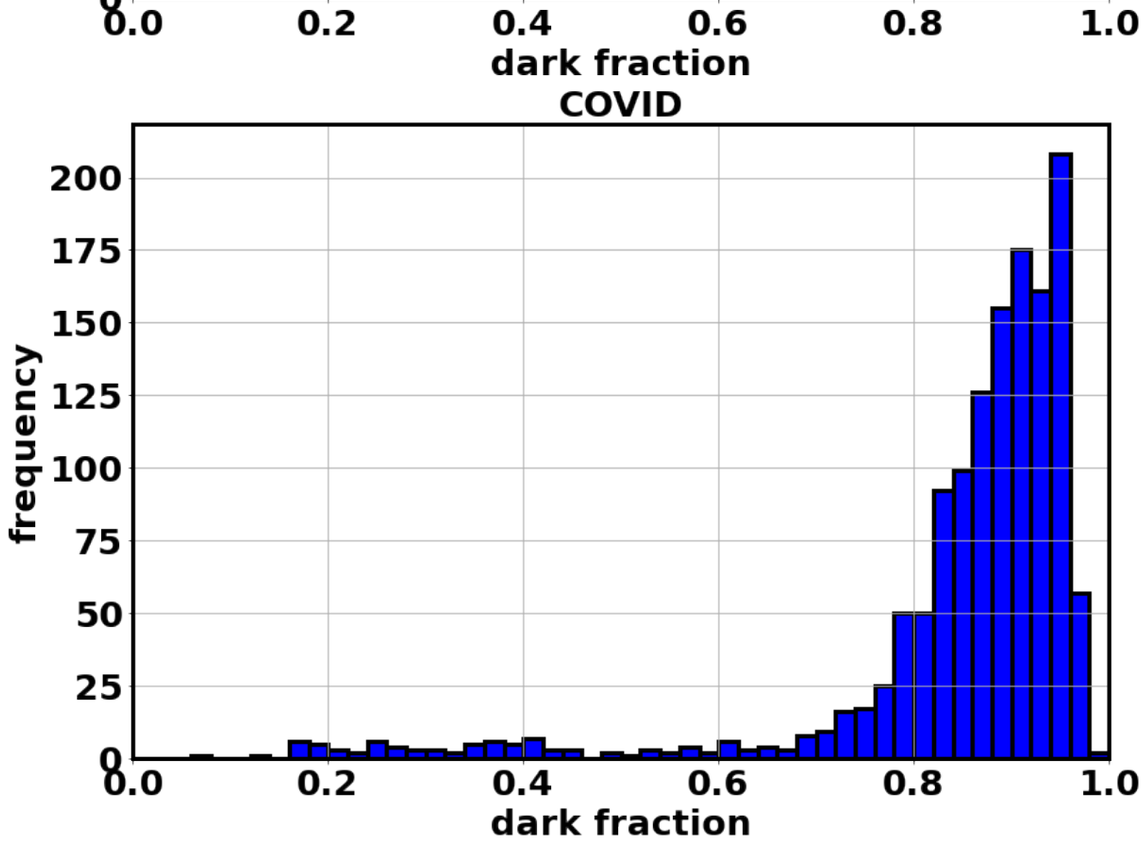


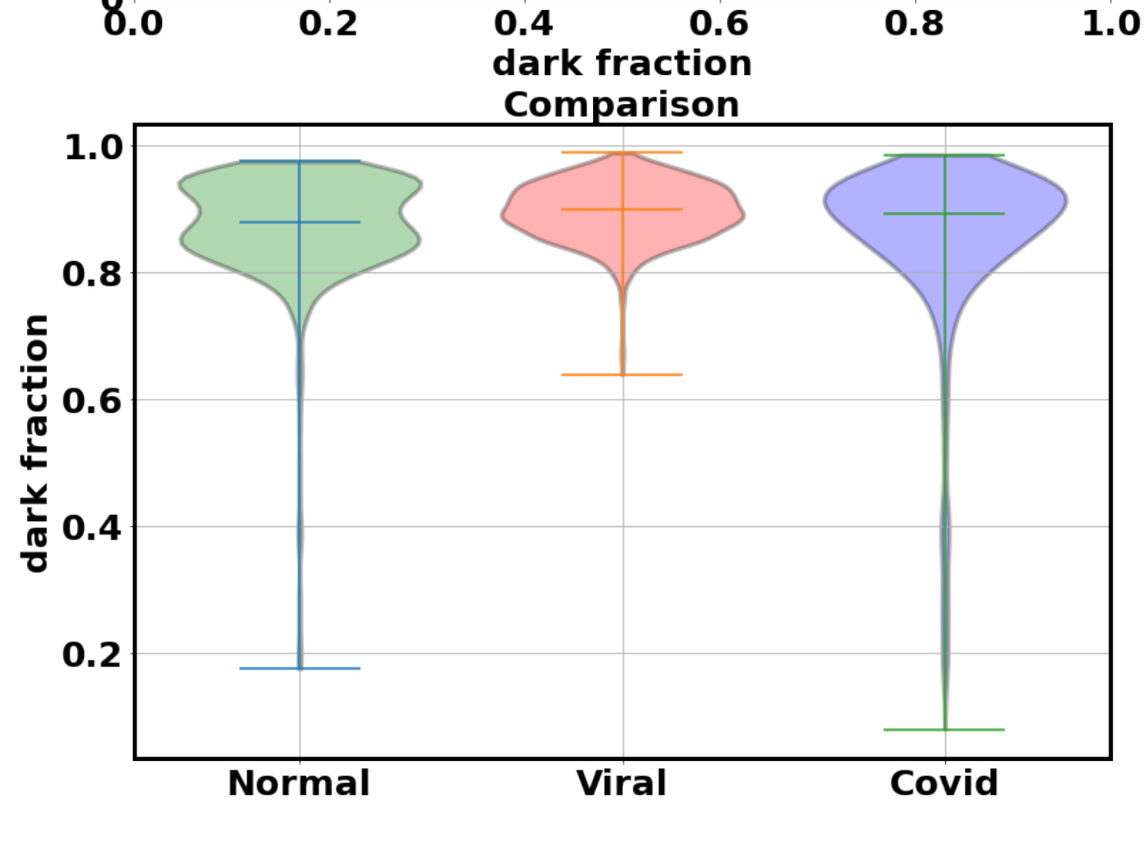
Fraction of black pixels for segmentation using contouring

Segmentation by contouring



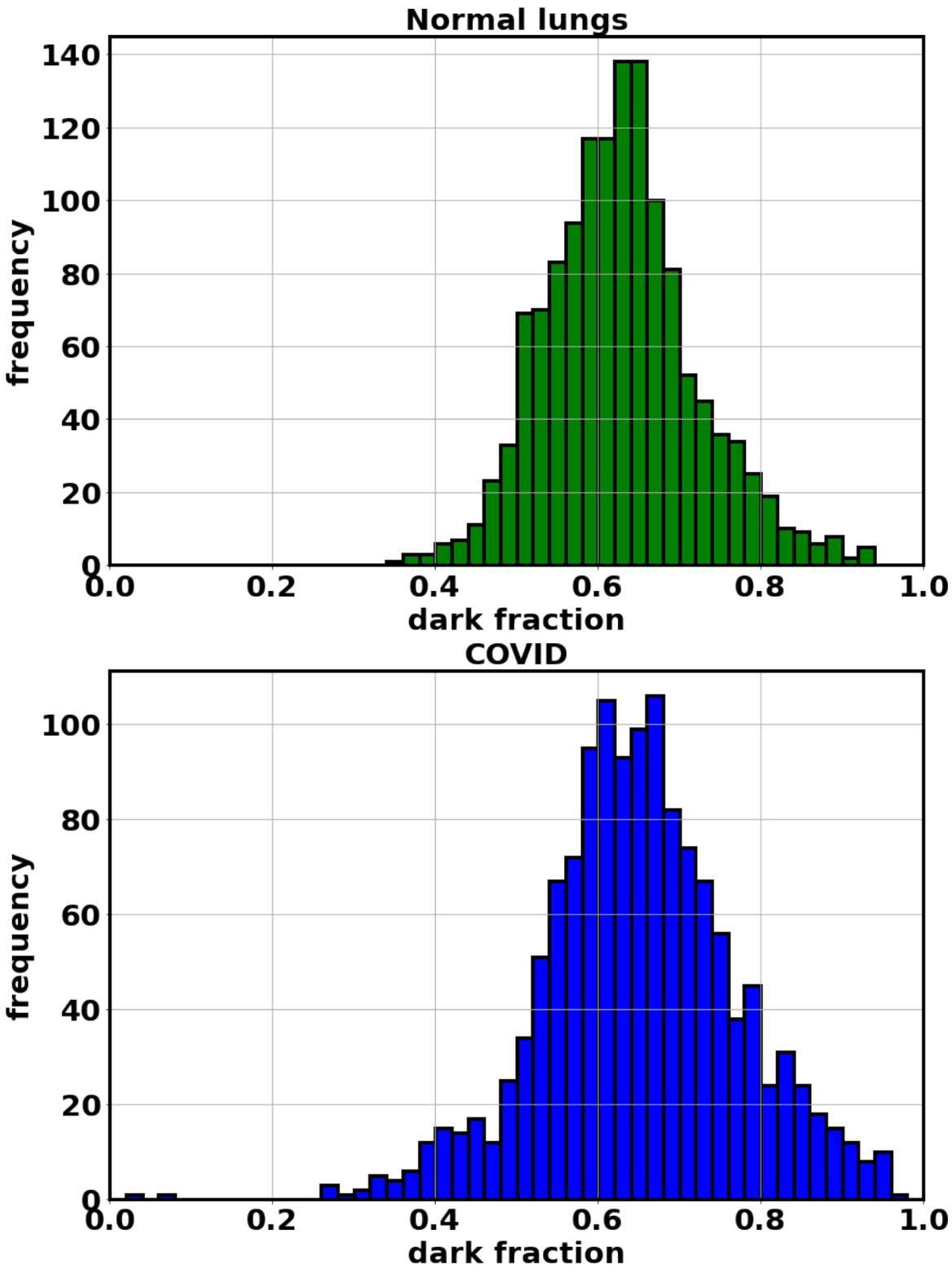


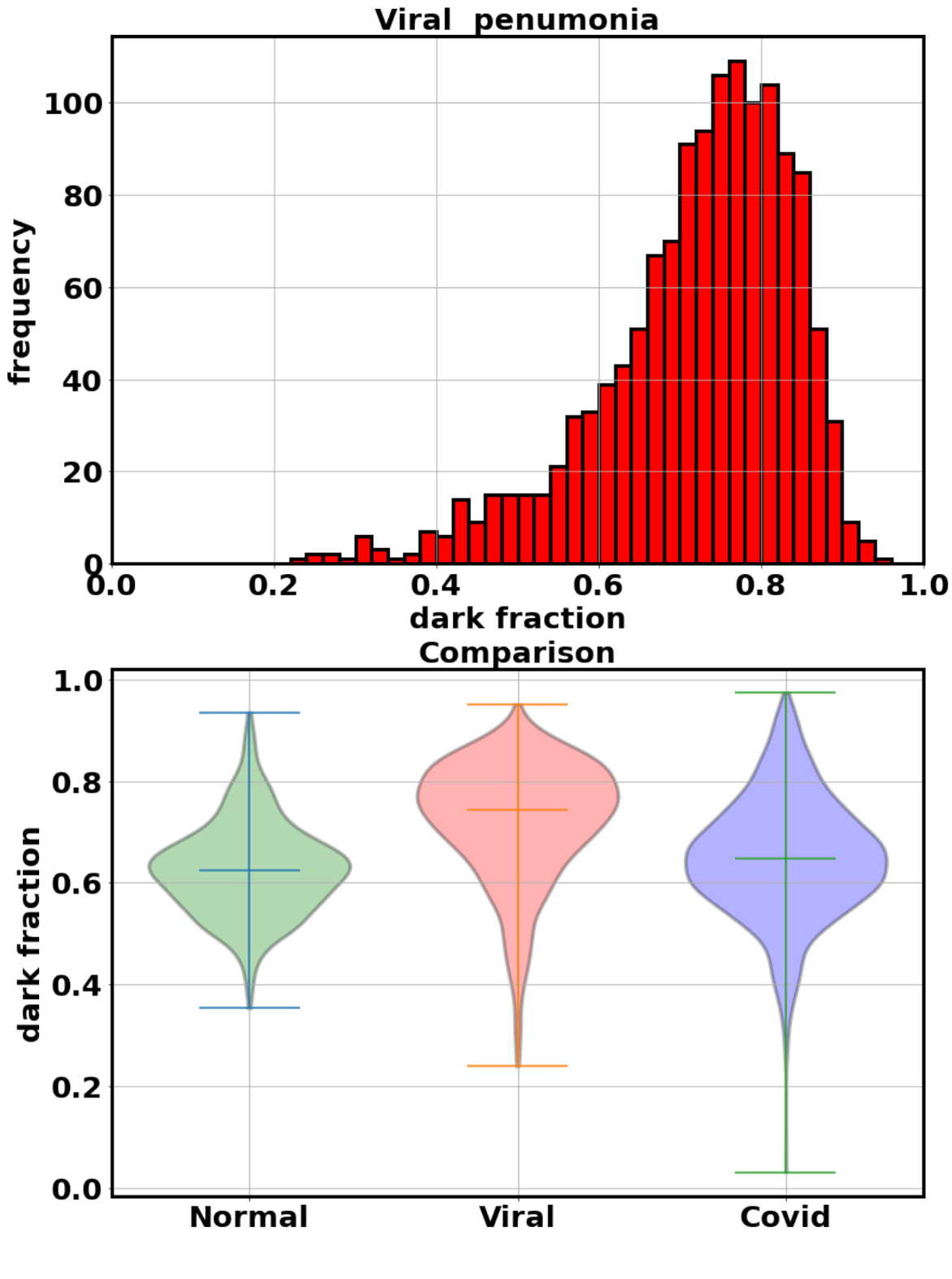




Fraction of black pixels for segmentation using thresholding

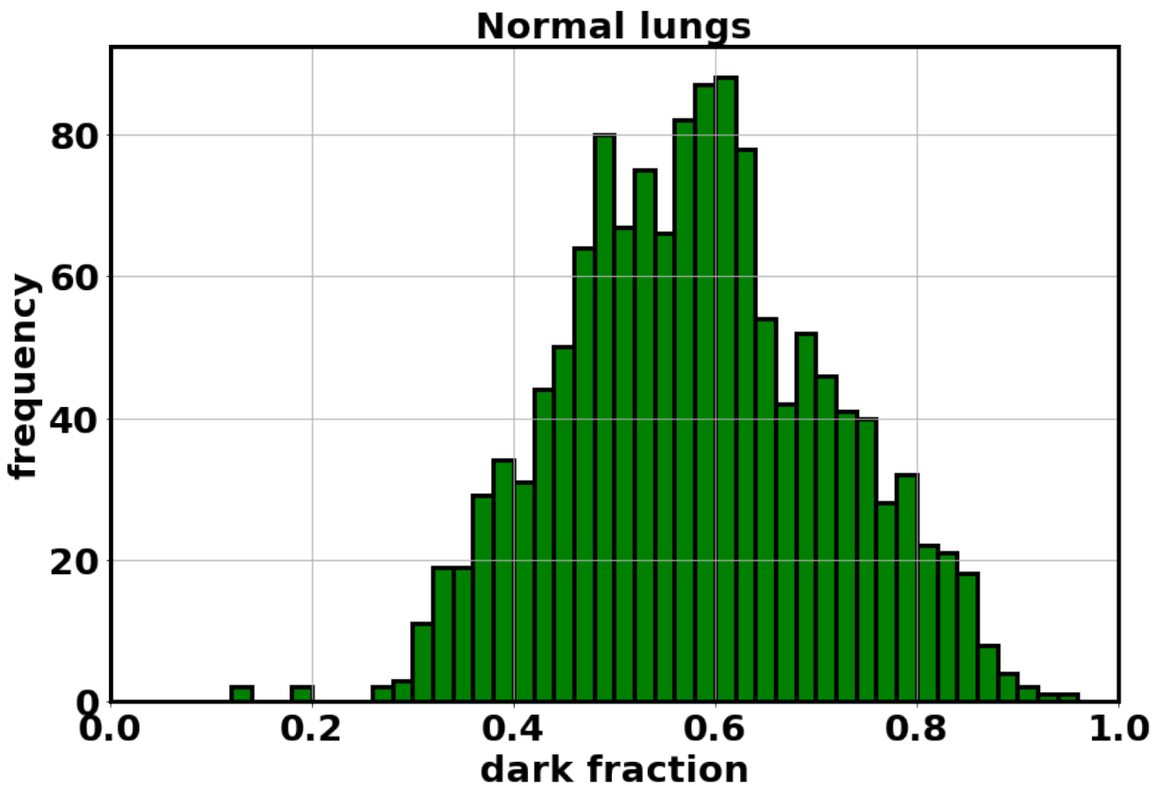
Segmentation by thresolding

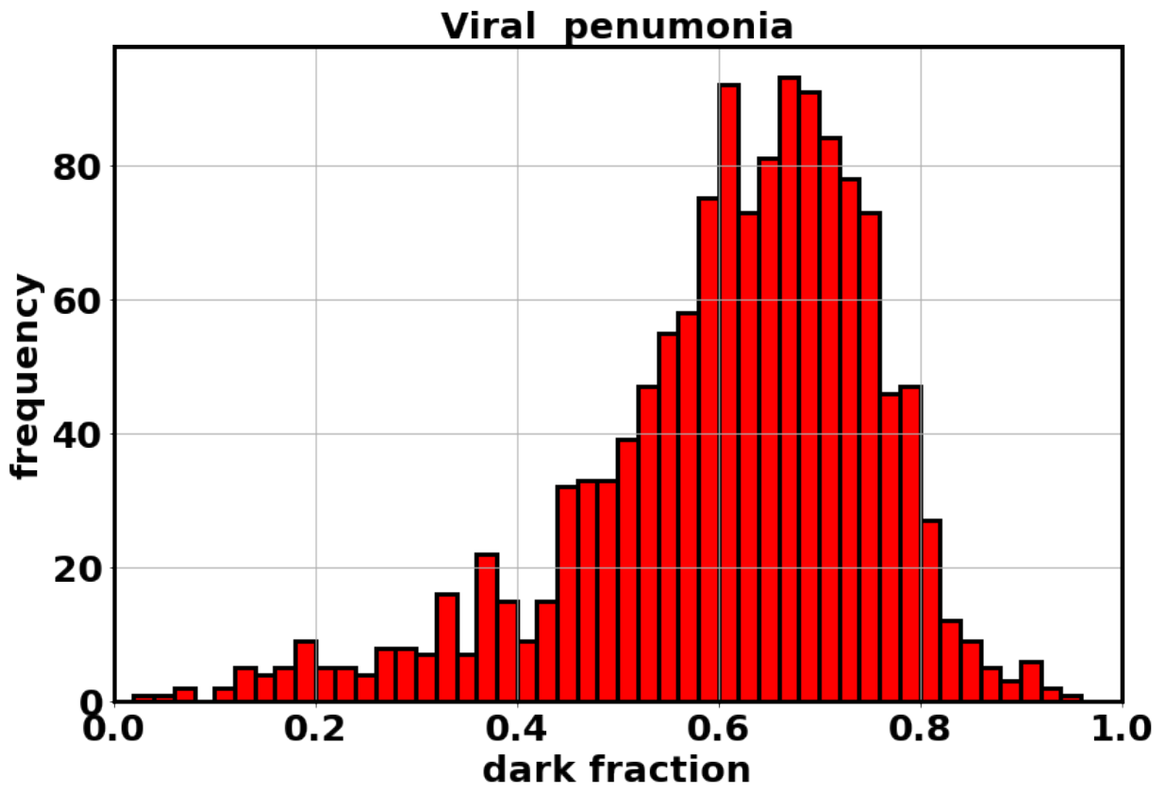


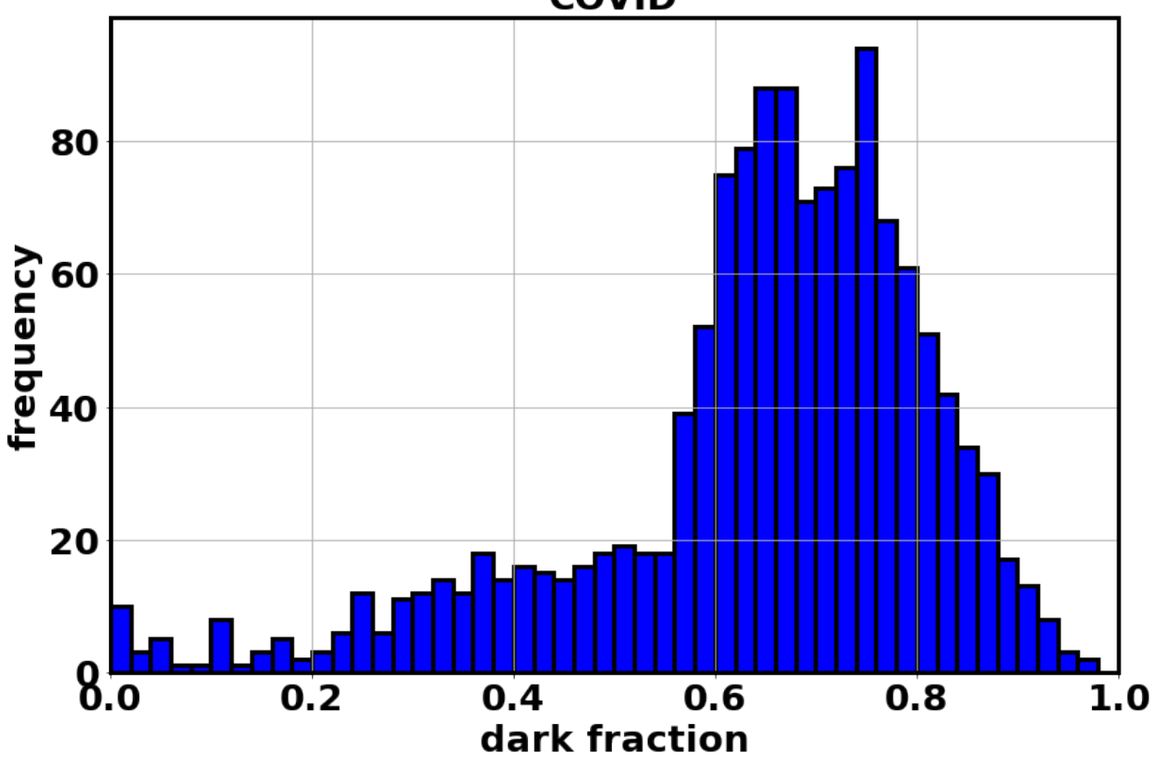


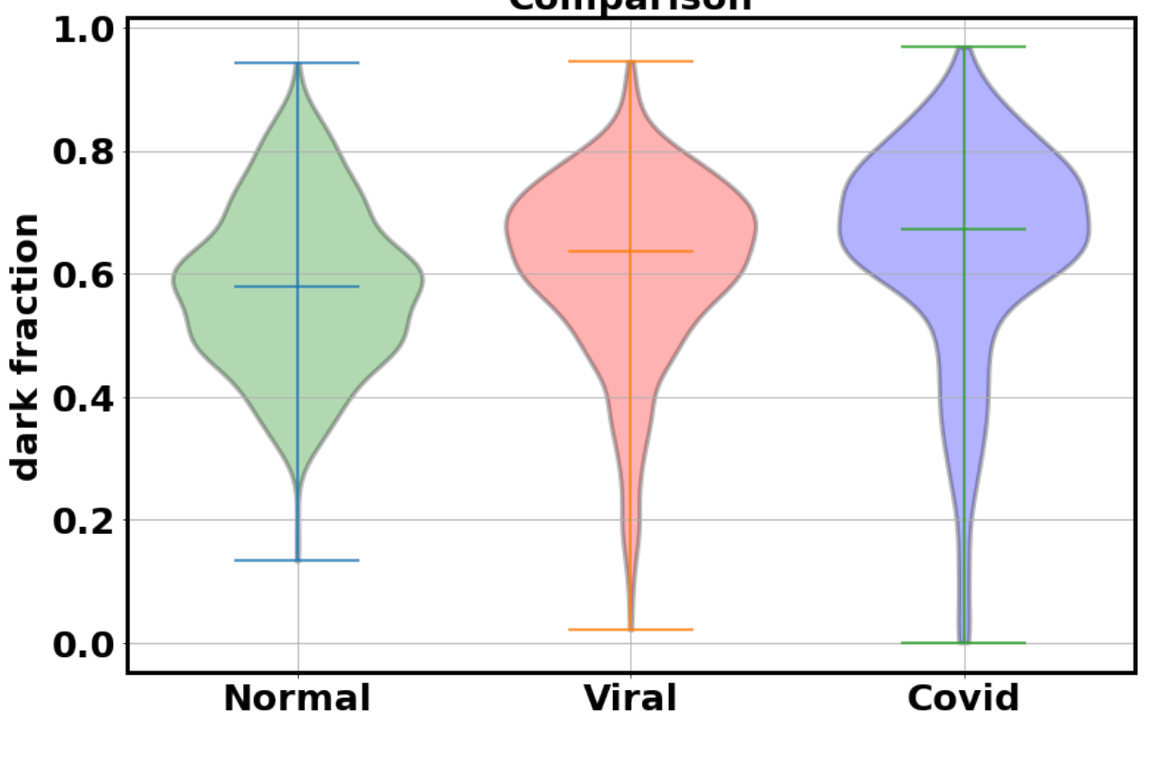
Fraction of black pixels for segmentation using masking

Segmentation by 50% masking



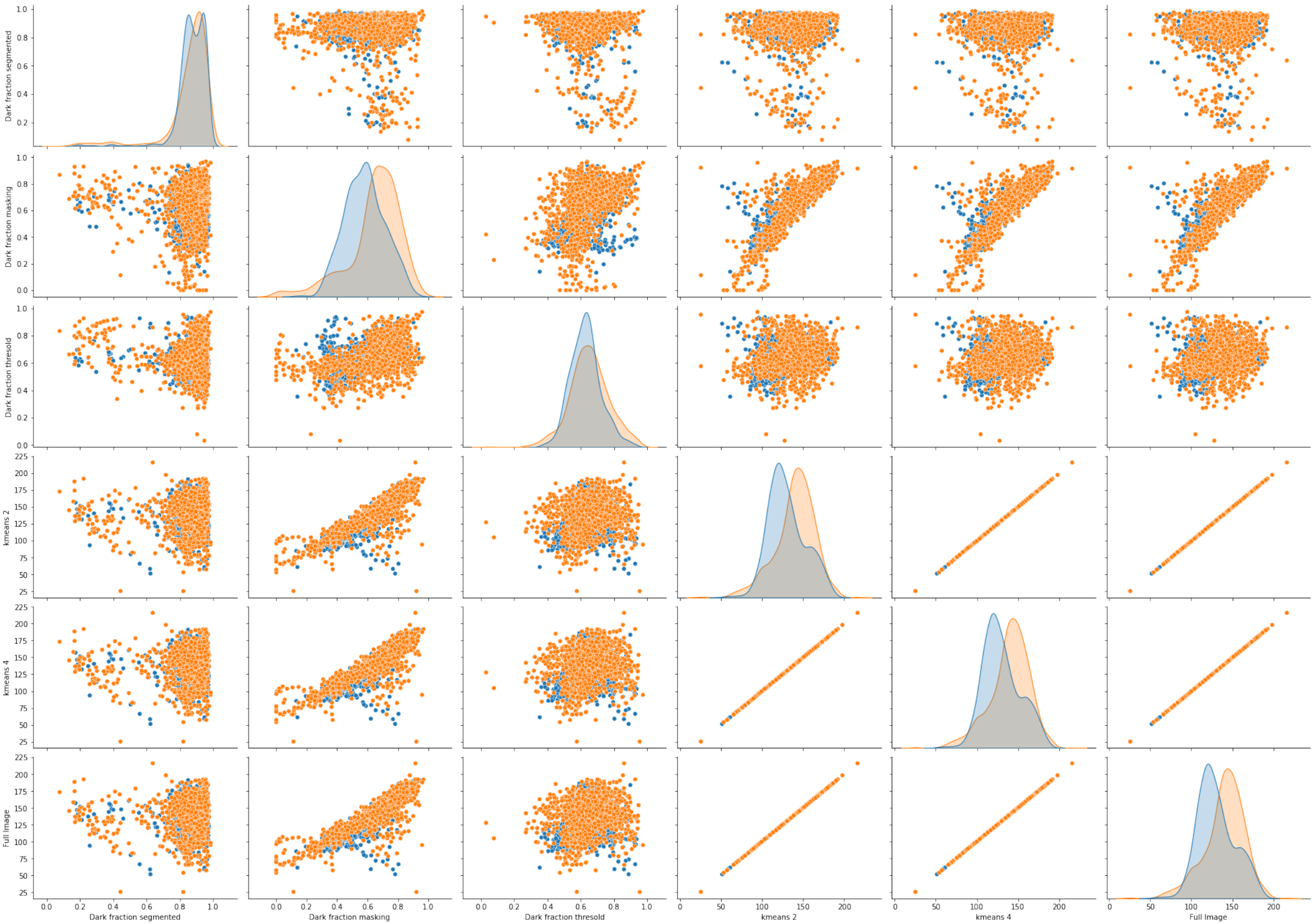






Now using these values from different segmentation techniques as inputs and the class of the images as output we train a machine learning model to differentiate between these three classes. We do pairwise analysis.

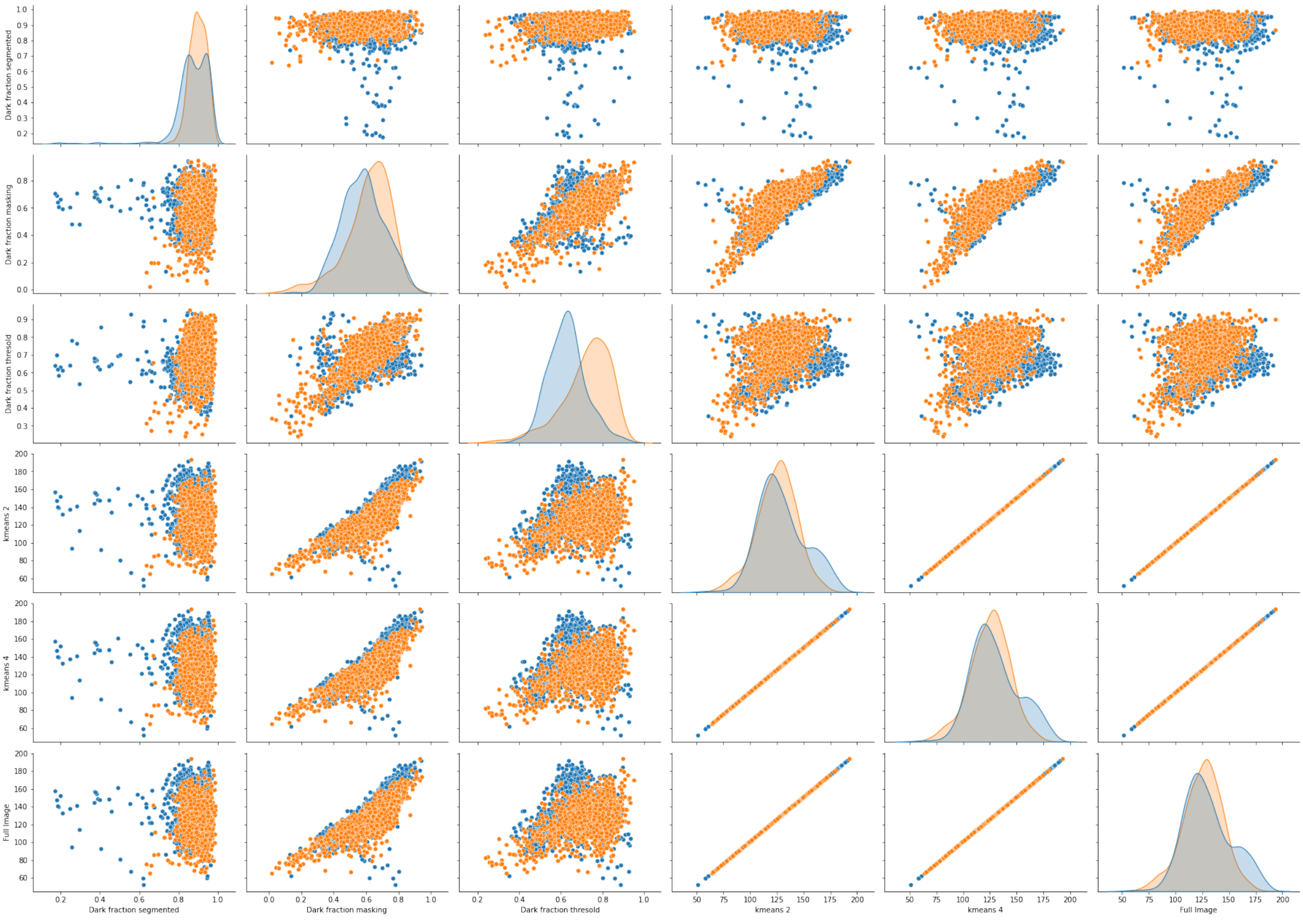
Normal vs COVID pairwise correlation plot



Normal

COVID

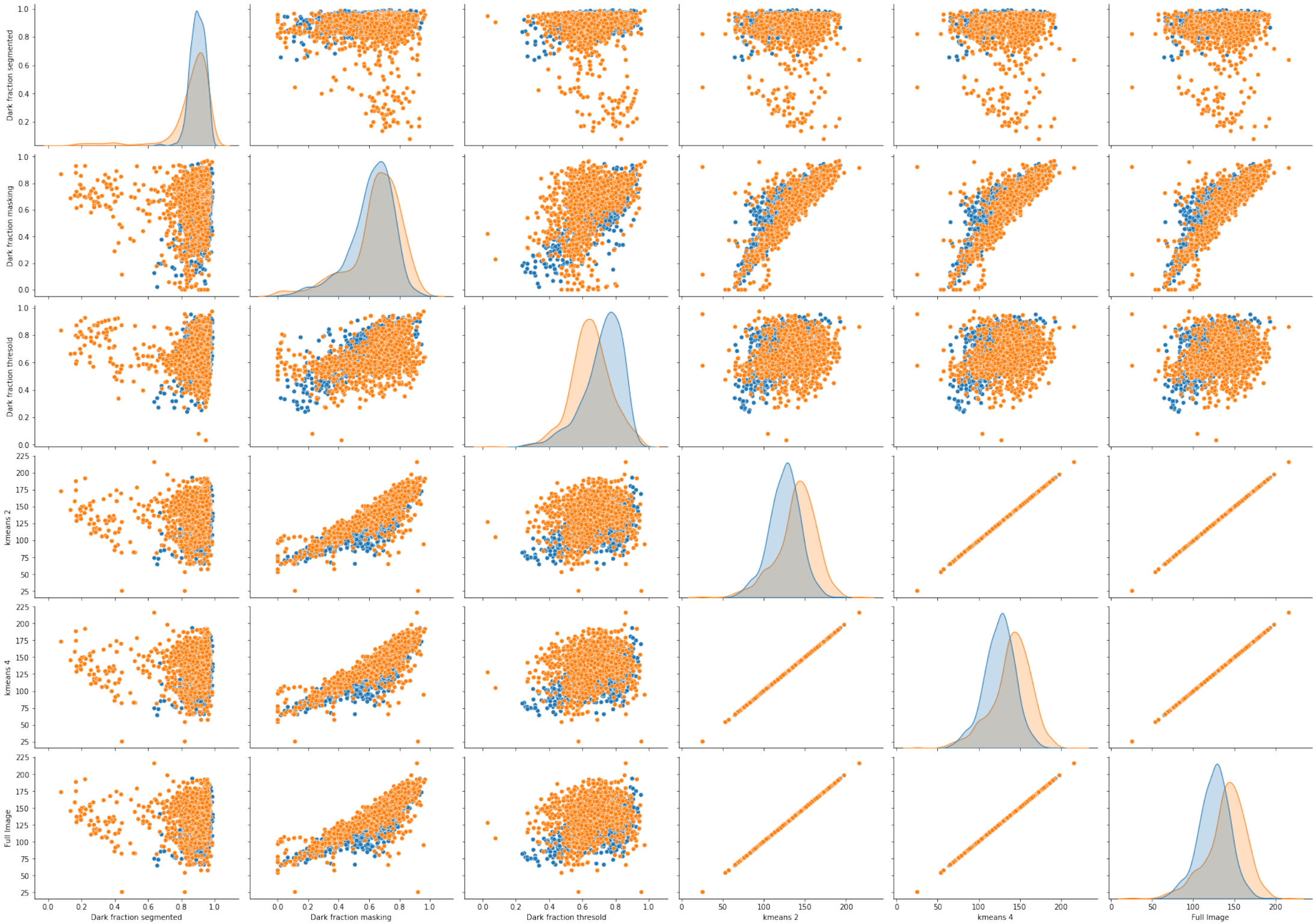
Normal vs Viral pneumonia pairwise correlation plot



Normal

Viral

Viral pneumonia vs COVID pairwise correlation plot

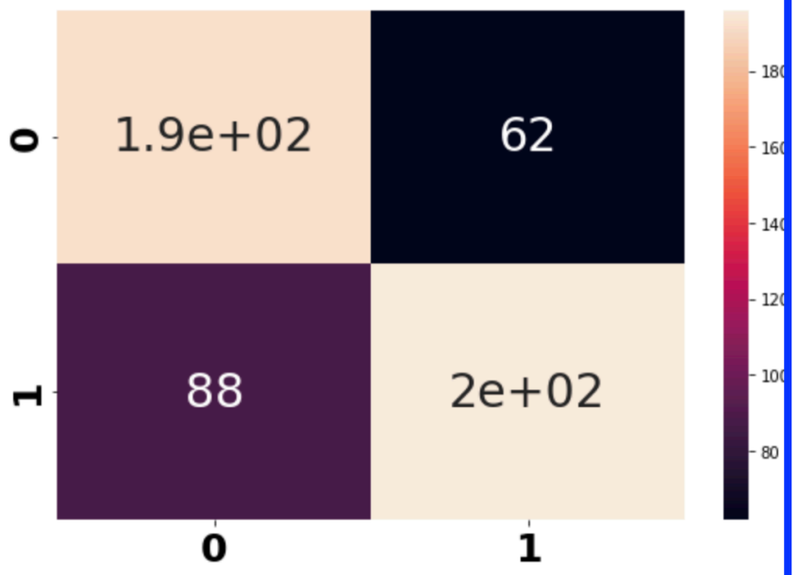


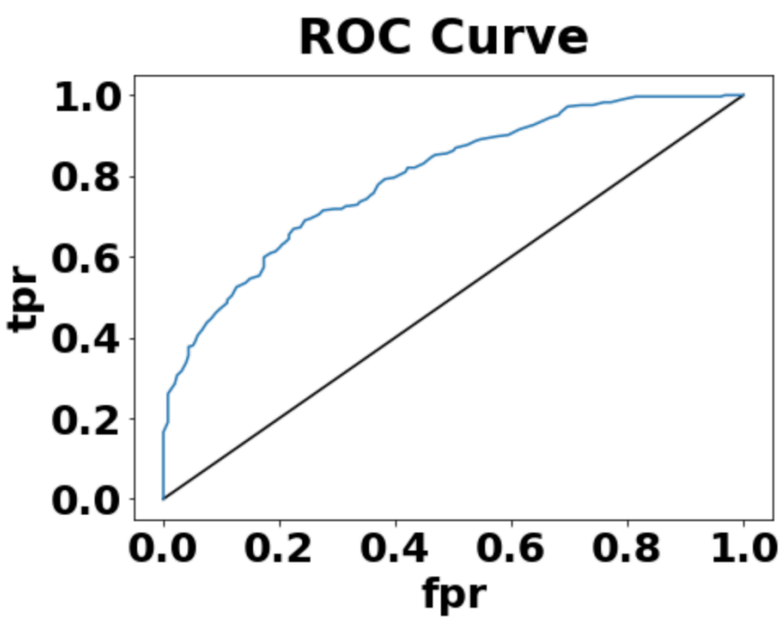
Viral

COVID

Confusion Matrix and ROC Curves

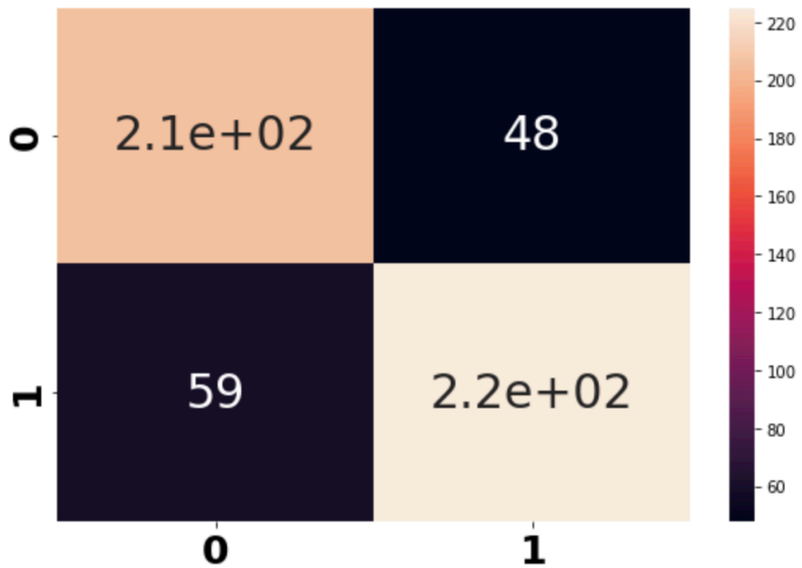
Normal=0 vs COVID=1

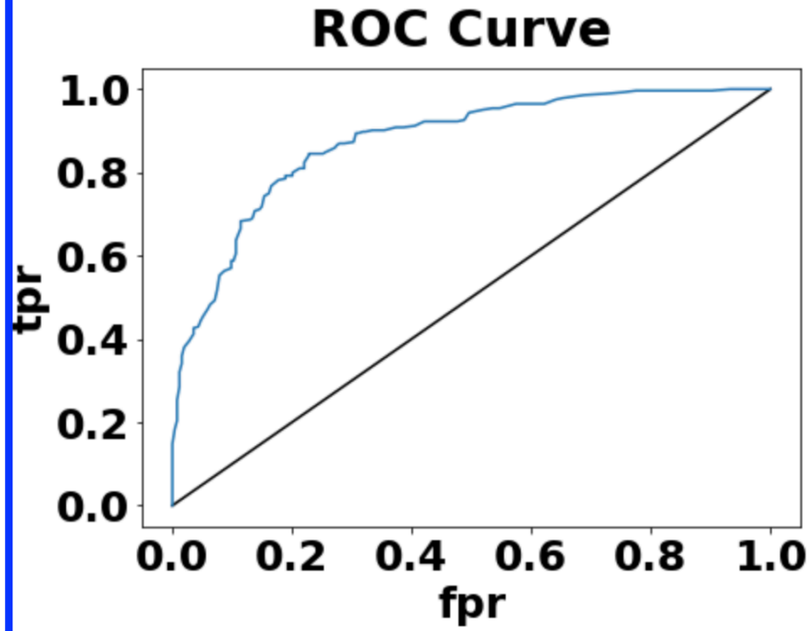




ROC AUC score = 0.798

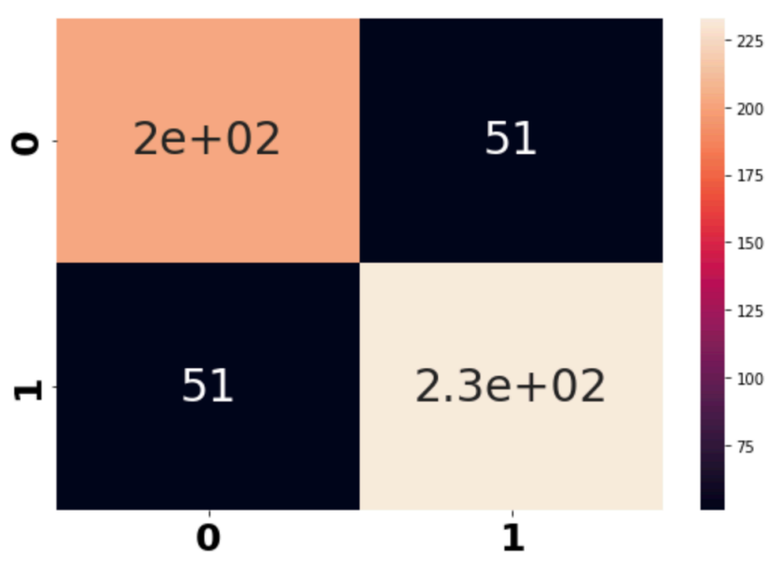
Normal=0 vs Viral=1

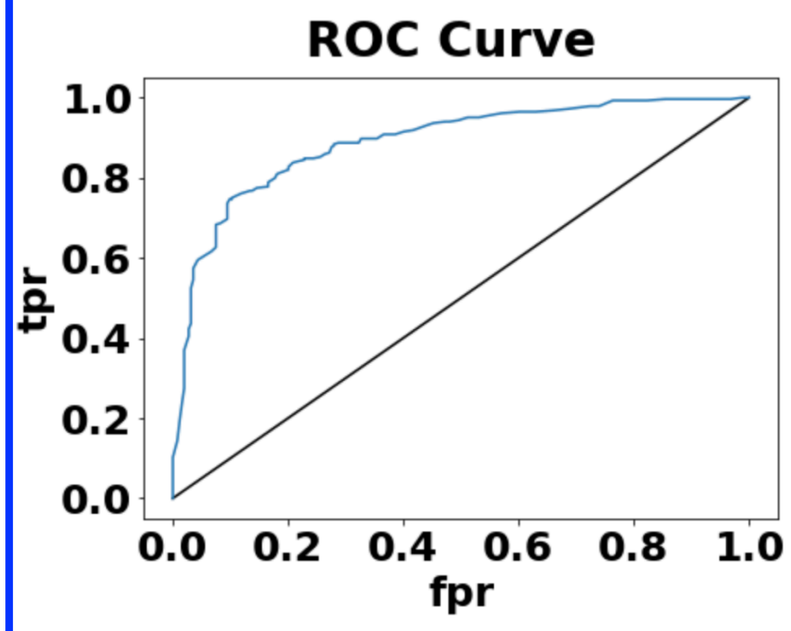




ROC AUC score = 0.875

Viral=0 vs COVID=1





ROC AUC score = 0.888

Future goals

- **Improve accuracy of classification by analyzing initial images for more scenarios and hence have greater training data for the classification model.**