In order to obtain the features of the images, the region of interest must be first extracted. The region of interest is cropped from the original images. The regions of interest are depicted with red rectangle as shown below.

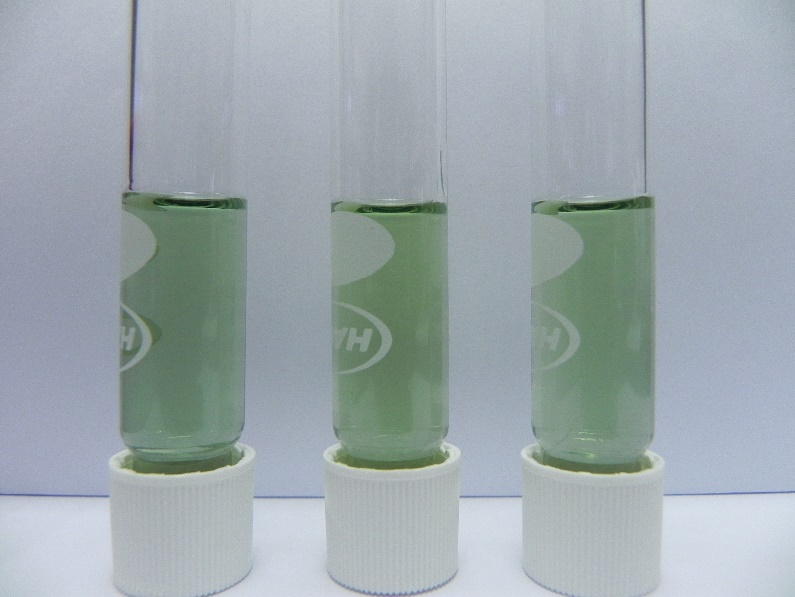


Figure X: Region of interest depicted with red rectangle

After the images are cropped, the crop images are then filtered using a median filter to remove the remove the designs on the test tubes. Next, the images are filtered using an average filter to remove the shading on the images. After the pre-processing is done, the mean and entropy of Red, Green, Blue, Hue, Saturation and Value are measured and saved into an excel file.

After obtaining the features of the images, the prediction of ammonia concentration is done using machine learning models which are Random Forest, Decision Tree and Naïve Bayes. The features measured are loaded and used to train the models. Each day have a total of 19 samples totaling 38 samples for both days. Due to the small sample size, K-Fold validation is used to measure the accuracy of the models. In the assignment 19-Fold validation is used. When the models are trained for the days separately, this means that for each iteration 18 samples are randomly selected for training the machine learning models while 1 sample is randomly selected for testing the machine learning models. When the models are trained for both days simultaneously, 36 samples are randomly selected for training and 2 are randomly selected for testing. The accuracy for the machine learning models is recorded.