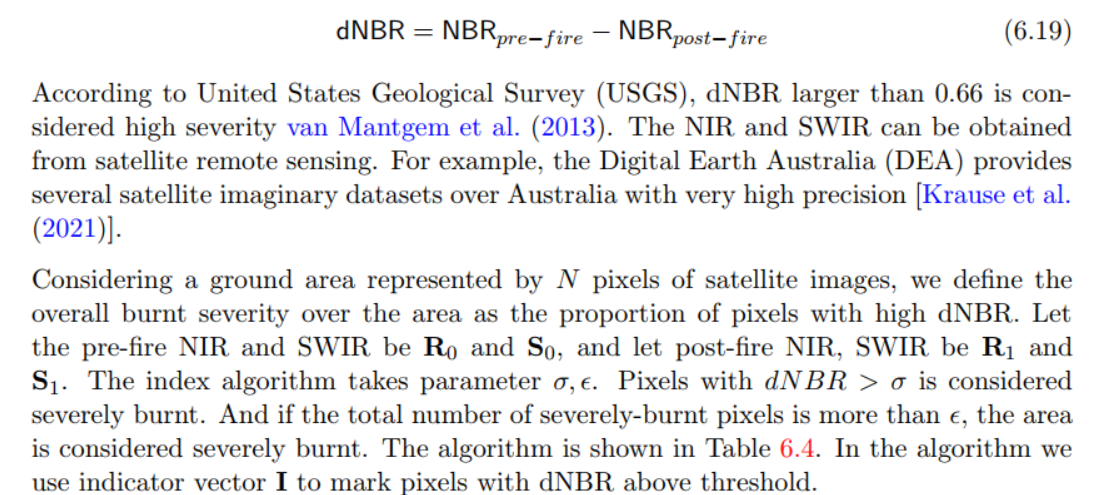
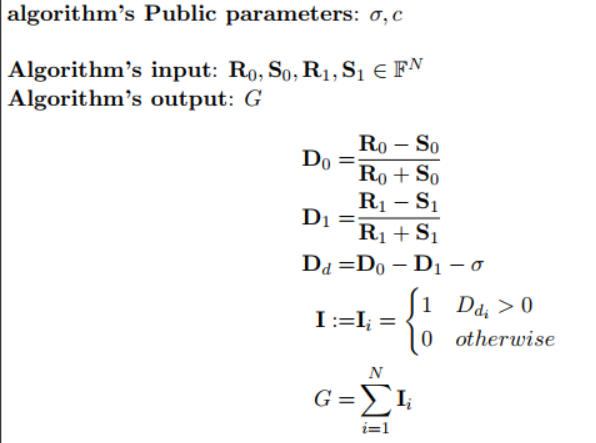
I researched on how many pixels to use in bushfire image. I want to compare the claiming result when using different number of pixels and the original satellite image.

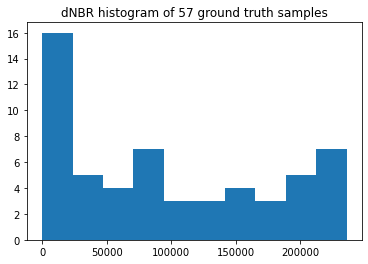
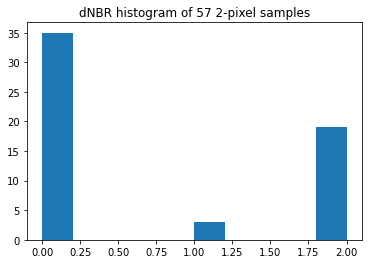
For data source, I manually selected 57 areas of forests, vary from around Brisbane to the south coast of Australia, where are very likely to be extremely suffered from the 2019-20 Australian bushfire season. For each area, I selected two timestamps, one around July 2019 and the other around February 2020, to show the change after the whole bushfire season. For each timestamp, I obtain a NIR image and a SWIR image from ga\_ls8c\_ard\_3 collection (introduction: <https://explorer.sandbox.dea.ga.gov.au/products/ga_ls8c_ard_3>), which are required in our model.

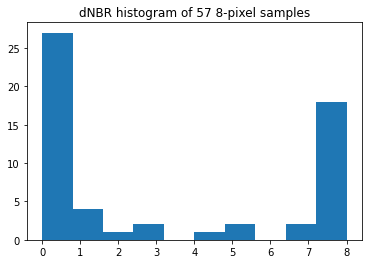
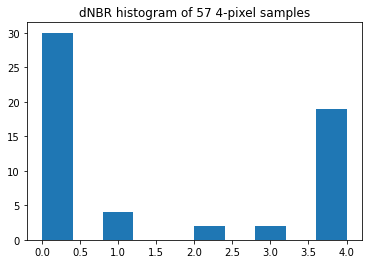
I first calculated dNBR of the original images (contains 250000 pixels each, represents about 22km \* 22km, used as the ground truth) and the compressed images of each area of forests. I tested 10 different sizes of compressed image: 2-pixel, 4-pixel, 8-pixel, 16-pixel, 32-pixel, 64-pixel, 128-pixel, 256-pixel, 512-pixel, 1024-pixel, 2048-pixel, 4096-pixel, 8192-pixel. Here I use dNBR > σ=0.3 to indicate high severity of bushfire for each pixel, as this works well in the ground truth samples, i.e. produce a meaningful distribution of G.

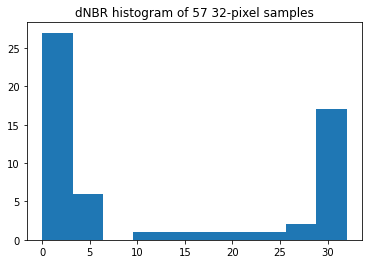
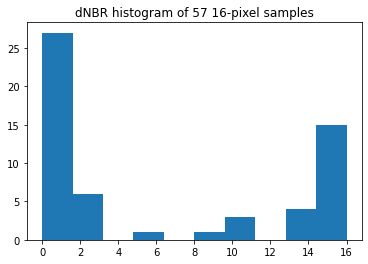
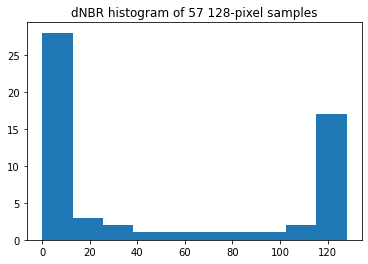
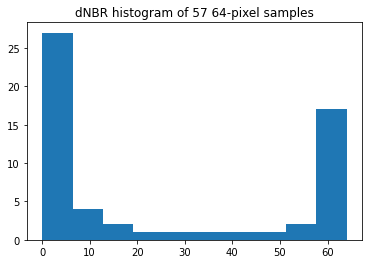
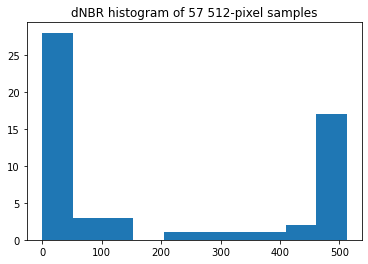
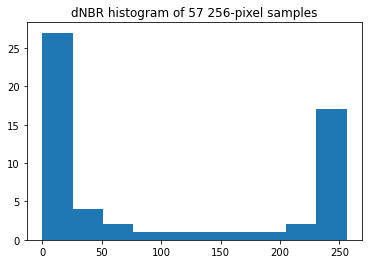
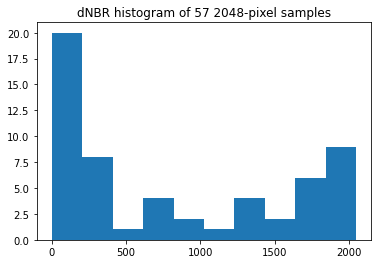
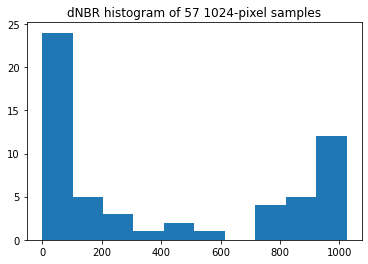
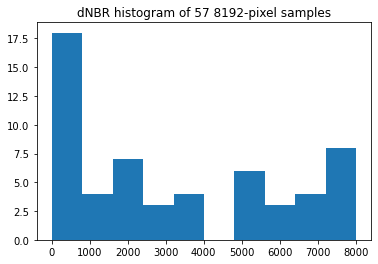
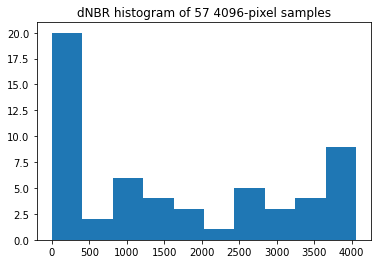




I think a rough but intuitional way of feeling their differences is to have a look at their dNBR histograms, as shown below.

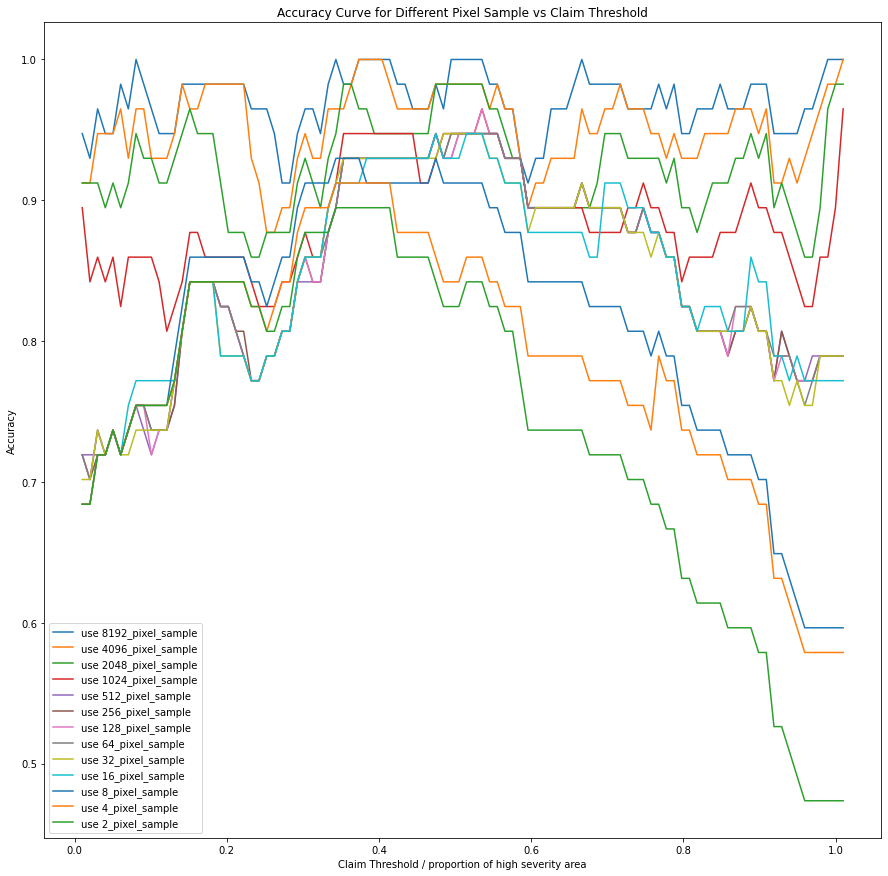
 

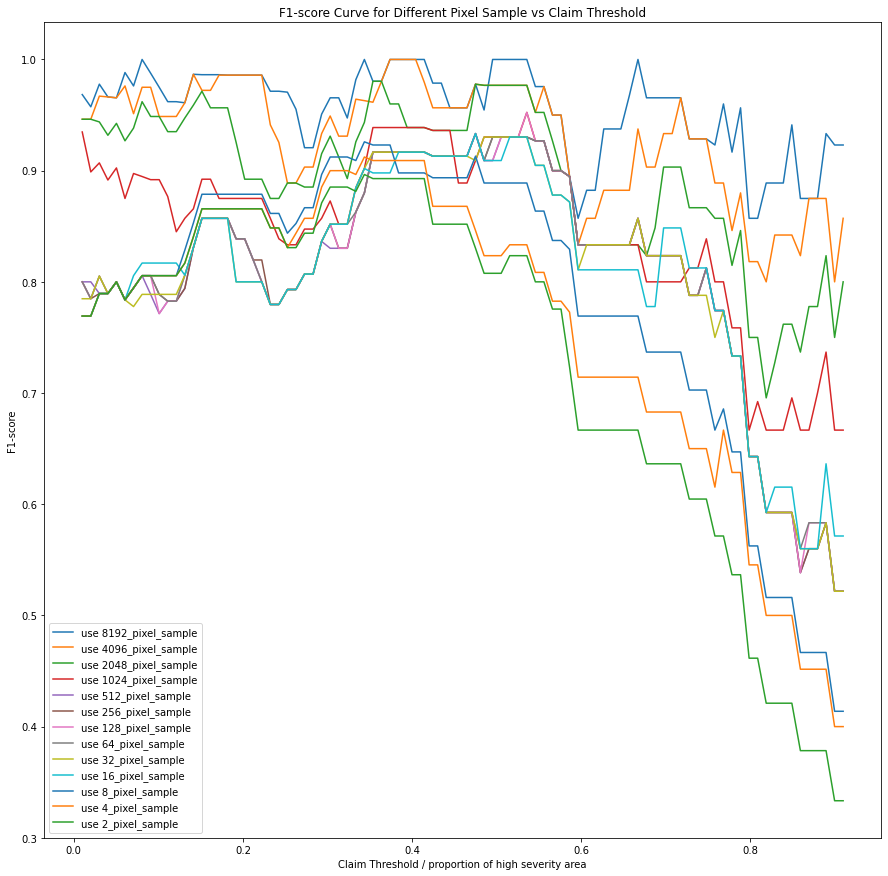


The above histograms confirm the fact that the more pixel we use, the better accuracy we may get. But it seems that the difference is not very large for similar sizes (like 32-pixel and 64-pixel).

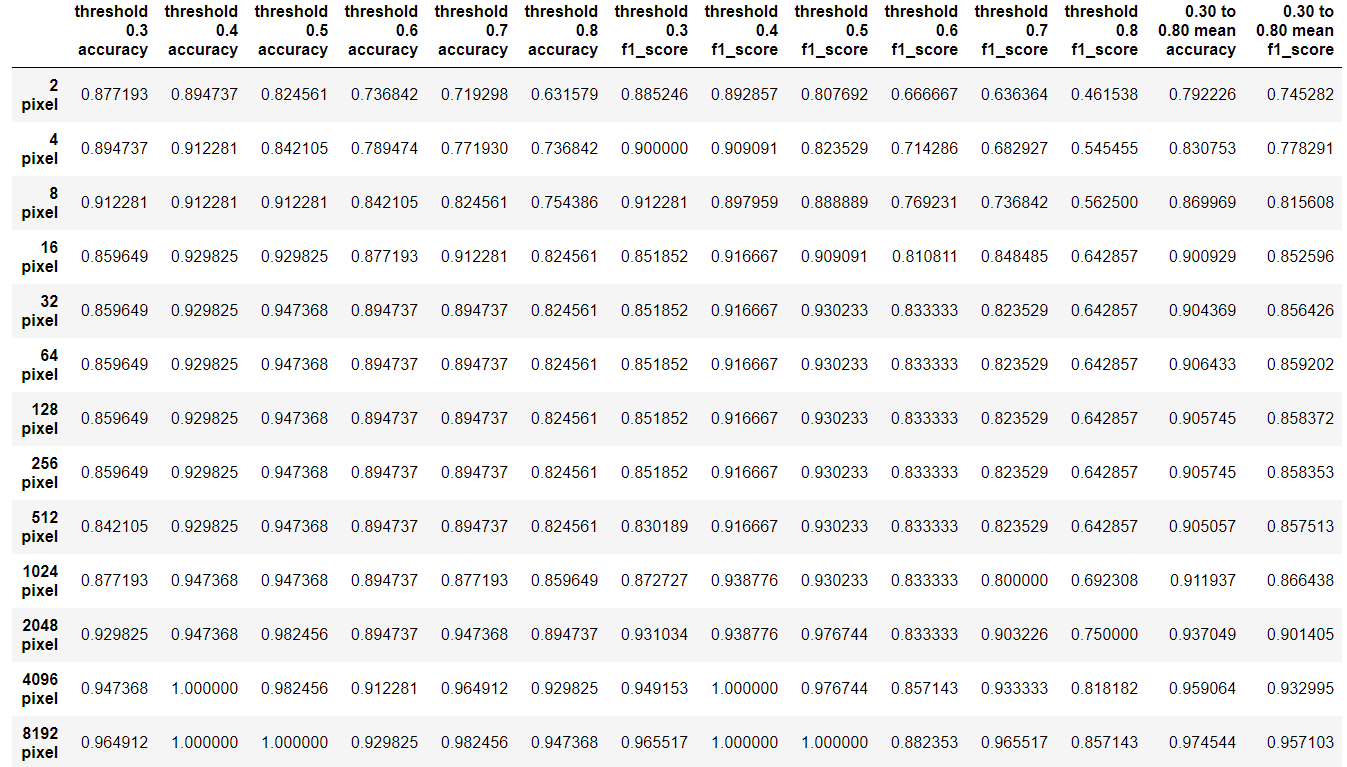
Then, to decide whether to accept the claim or not, our model requires a pre-defined threshold of proportion of high-severity area. The specific threshold (ε in the algorithm) needs to be decided by the insurance contract. So, let’s the threshold range from 1% to 100% and see the accuracy and f1-score of claim results would be for different image size, when I assume the claim result when using the original image is the ground truth.



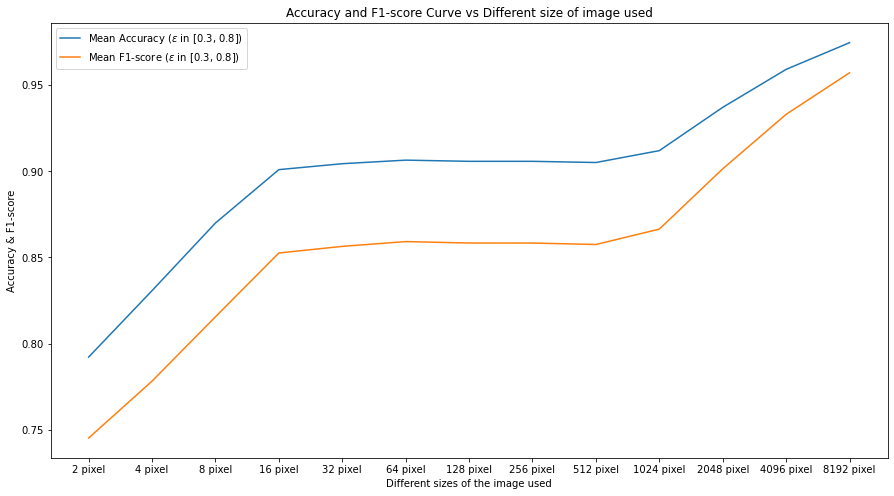


The larger the area under the curve is, the better the result is.

Specially, I think (30%, 40%, 50%, 60%, 70%, 80%, and numbers in between) are possible meaningful thresholds for **difference insurance policies**, because these thresholds would accept neither too many nor too few claims when using the ground truth data. So, I printed out a table showing the accuracy and f1-score for these special thresholds as well as the **average accuracy and f1-score** (last two columns which are the most important) when using different size of images.



The table is complex because of multiple possibilities of insurance policies. If we only consider the mean accuracy and f1-score:



The table is included in the attached output.xlsx file.

The answer to how many pixels to use depends on the balance of proving time and the accuracy & f1-score above. We can see that when we use 16 pixels or more for the claim image, the accuracy is greater than 90%.