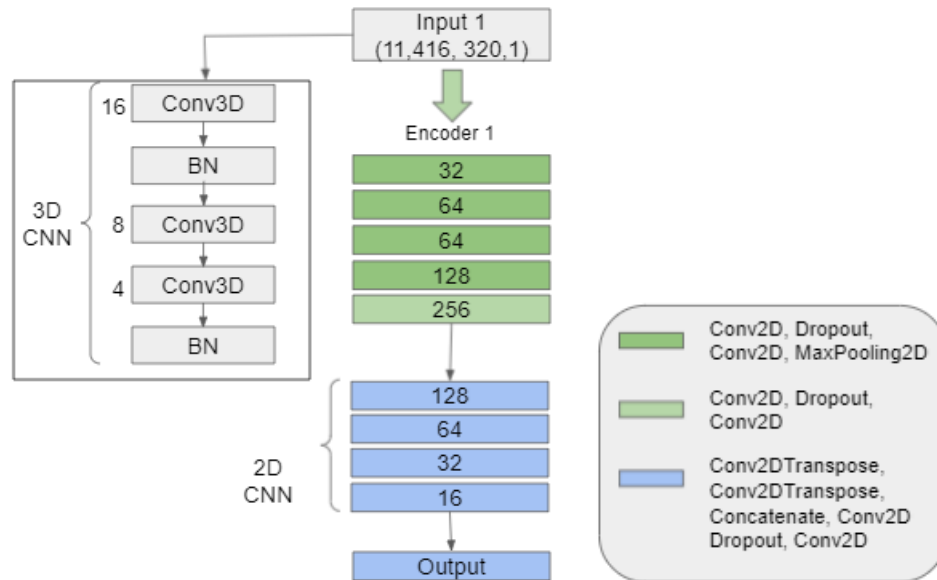


# Analytical Report

This report presents the results and analysis of the following deep learning model which is currently being implemented on LITS dataset:



## **Brief description of model**

The model takes preprocessed liver scans as input. These scans have been cropped to only contain liver and are padded to the size (11, 416, 320, 320). The model receives 11 slices for each slice containing tumor, i.e., 5 previous, 5 next and the slice with tumor. This makes sure the model receives enough information. The input is given to a 3D CNN, the output of which is fed to an encoder and then to a 2D CNN. Final output is the predicted mask of tumor.

Further details present [here](#).

## **Results**

These results are obtained on a model trained on 80 scans and validated on 20. Validation dice score is 0.5830.

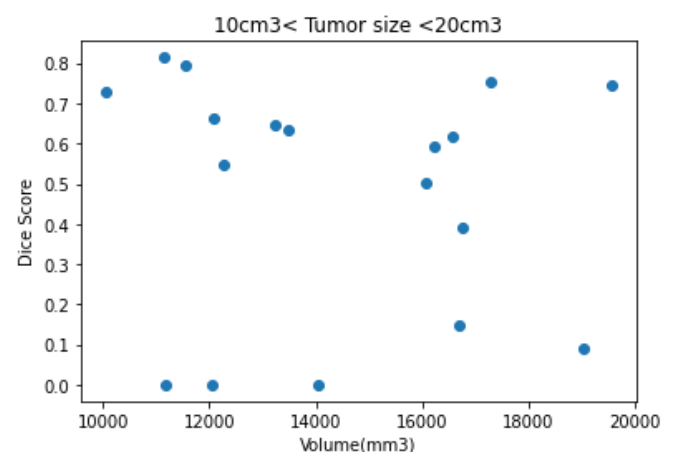
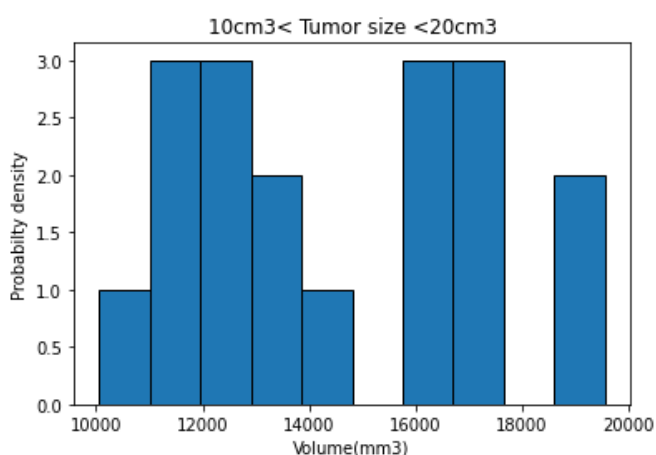
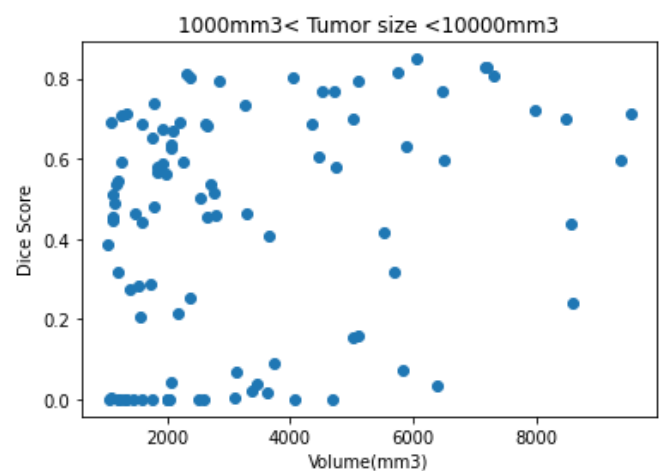
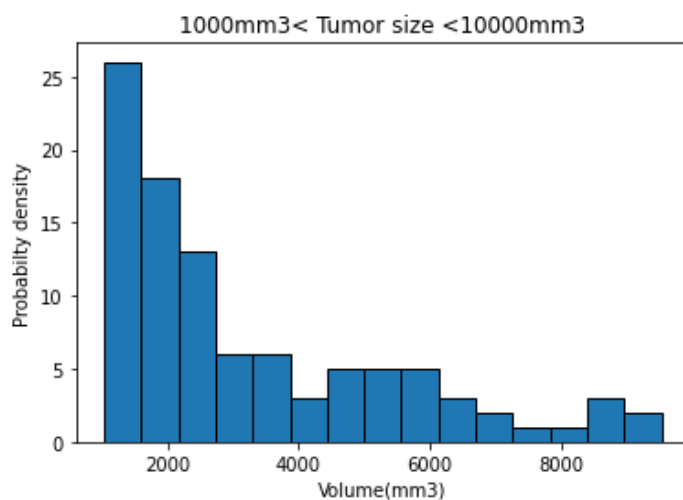
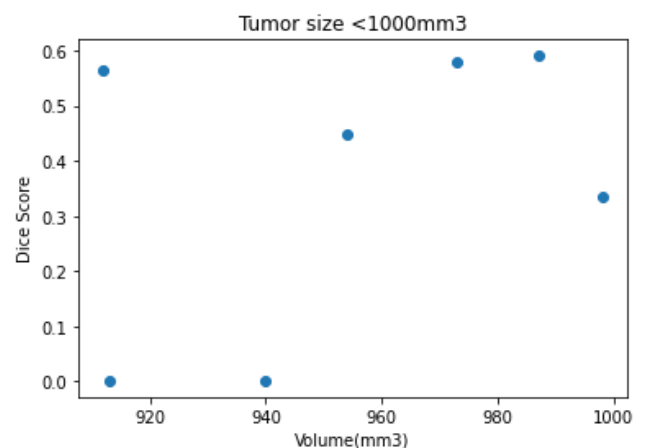
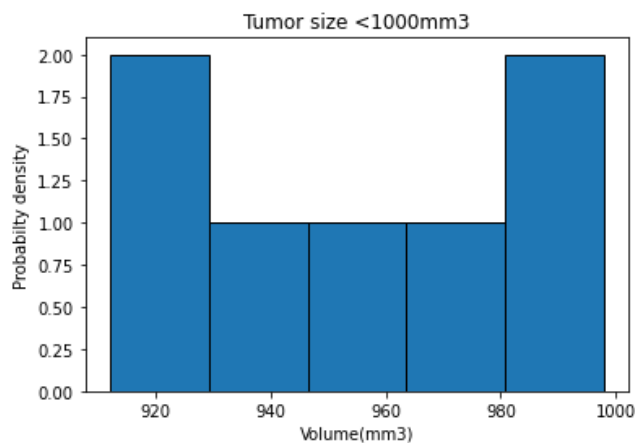
Similar to training, testing is done on scan of size (11, 416, 320, 320), where for every slice in the original scan, 5 previous, 5 next slices are given to the model. Output is the predicted tumor mask. This gives an Overall Dice Score of 0.3458.

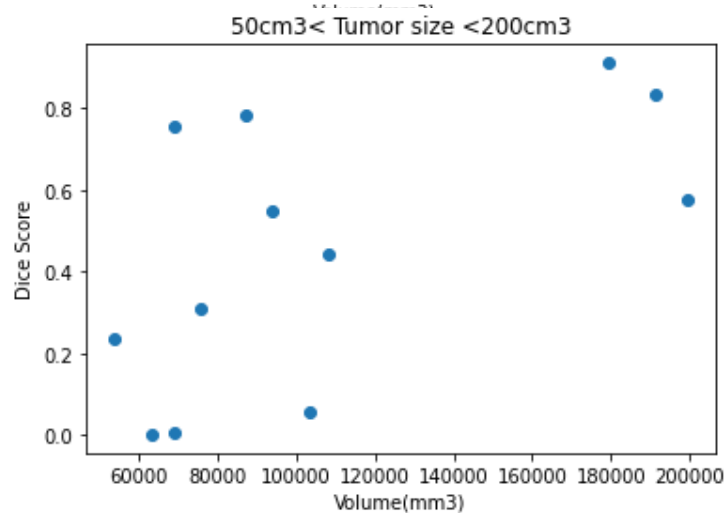
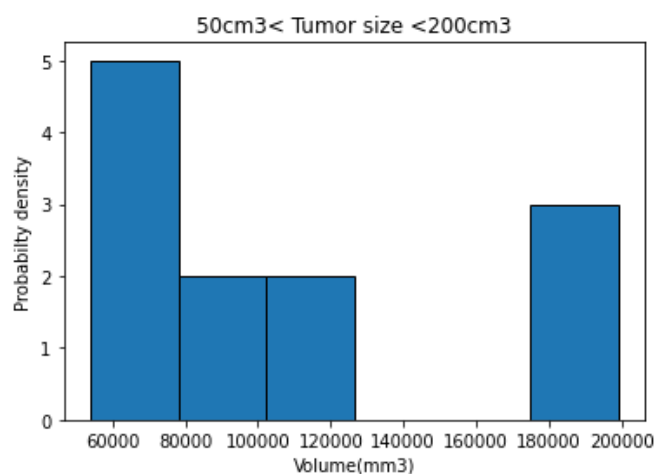
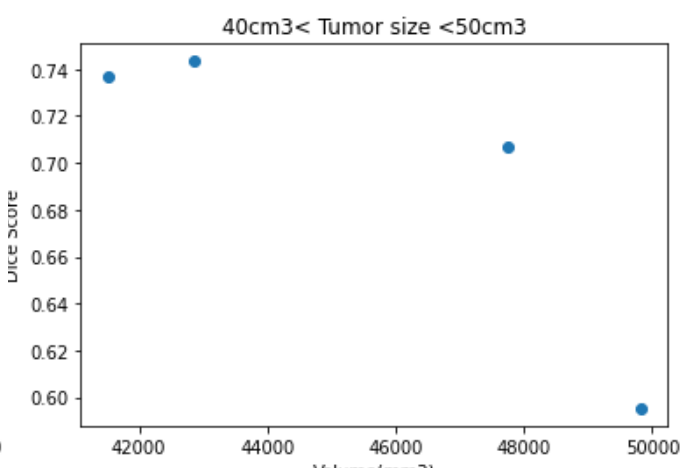
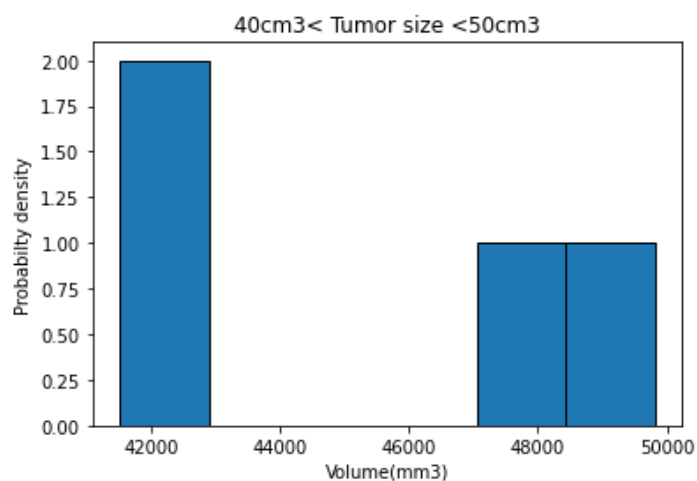
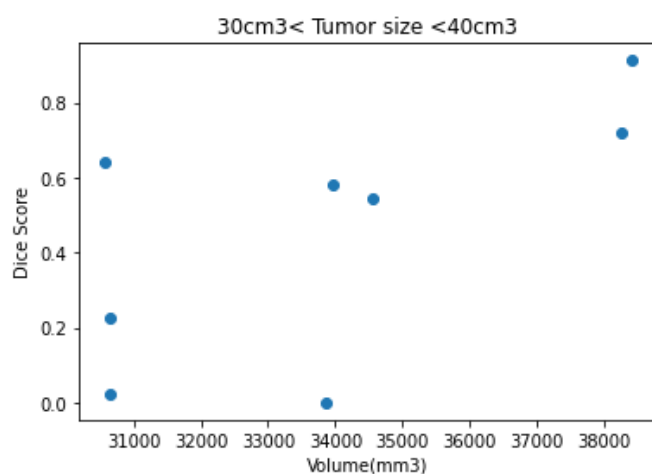
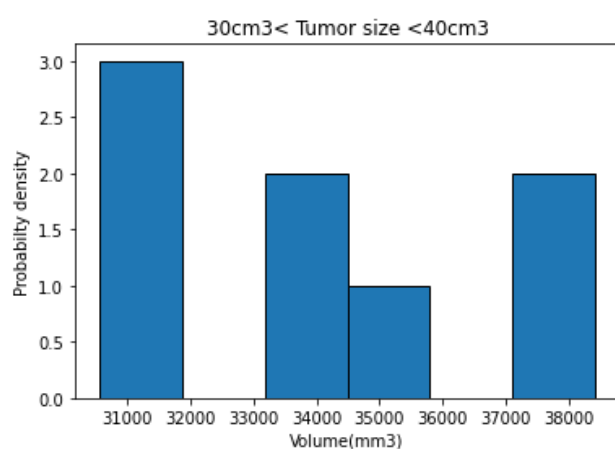
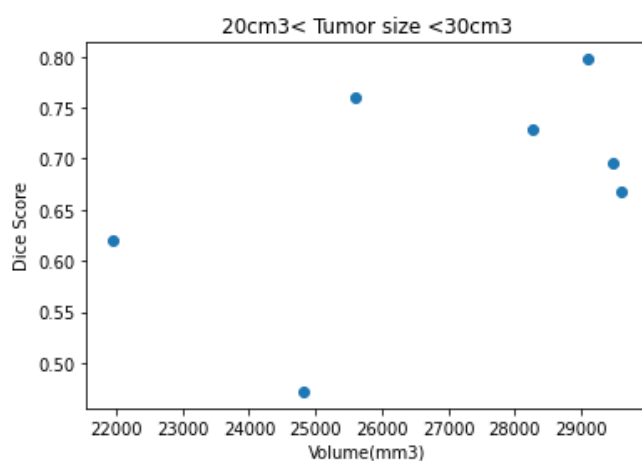
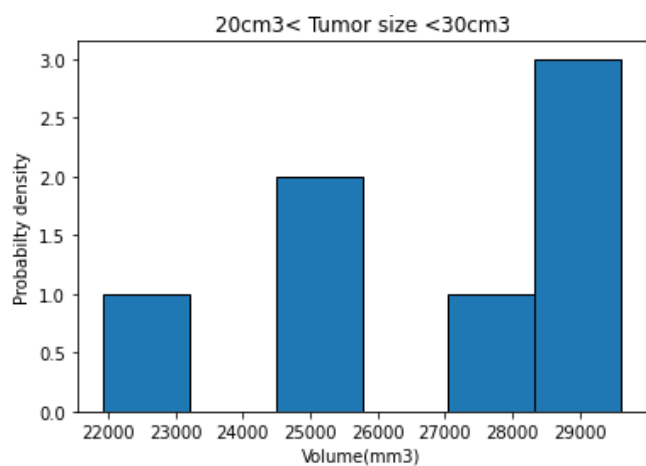
For tumor analysis, patches of individual tumors in ground truth were extracted and compared with patches in the same area in predicted scan. This was done to find what sort of tumors are detected and what sort are not. The conditions around a tumor play an important role in its detection. Mainly, these conditions include volume

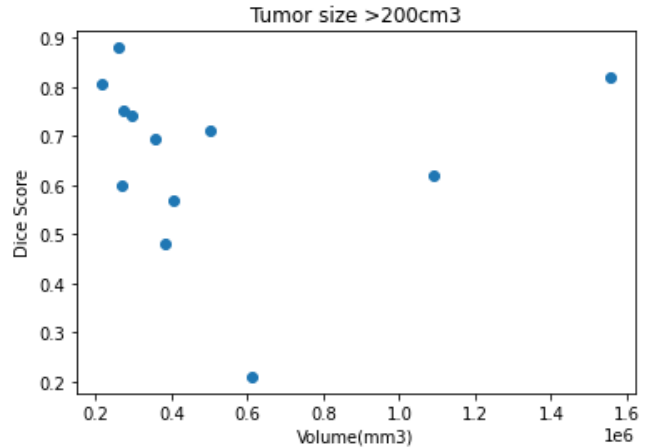
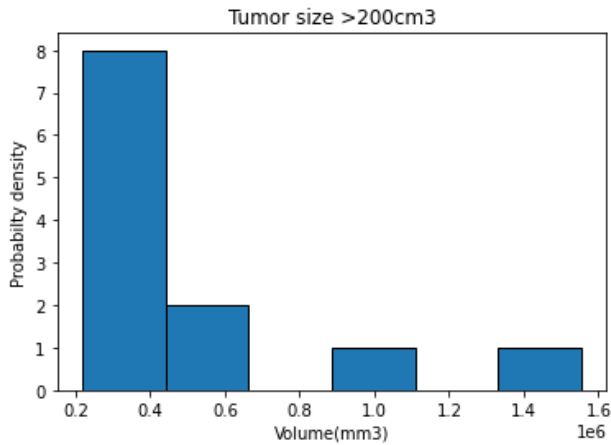
and intensity. There is a chance that bigger tumors have a higher probability of getting detected or that higher intensity ones could be classified accurately. To analyse these conditions, following results were studied.

### Volume

Histograms and scatter plots are plotted to find any meaningful relationship between ground truth tumor volumes and dice score. Volumes are divided into ranges to closely see if there is a trend. The range for which a graph is plotted is written at top.







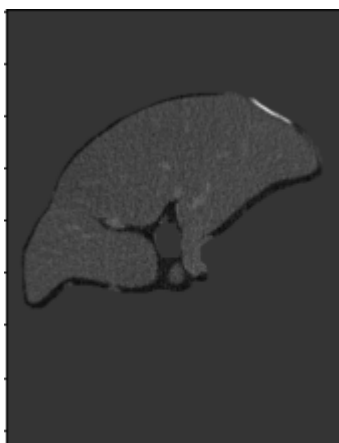
With these plots, it is concluded that most of the tumors are in the ranges 1cm<sup>3</sup> to 10cm<sup>3</sup> and 10cm<sup>3</sup> to 20cm<sup>3</sup>.

There are 50 tumors of a total 167 which had a dice score < 0.3. These misclassified tumors also lie in the same size ranges as the classified ones. Hence, it is possible that there is no correlation between tumor size and dice score. This means that volume has little effect on dice score.

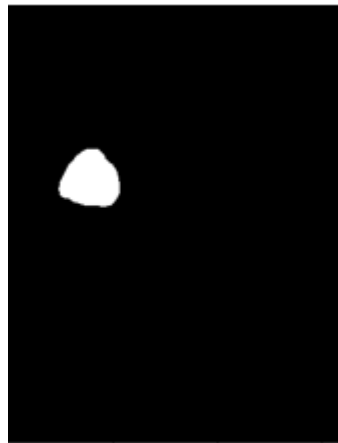
### Intensity

The next metric is pixel intensity of tumors and liver. To evaluate their relationship with dice scores, scans of different dice scores were plotted and visualized.

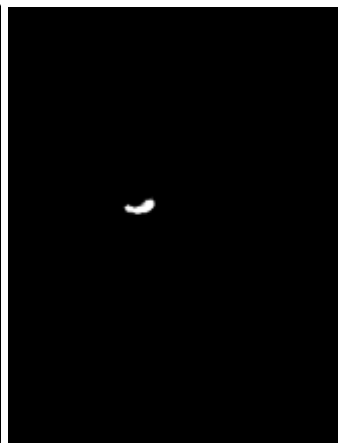
Overall Dice Score: 2.6484e-05



*Original Scan*

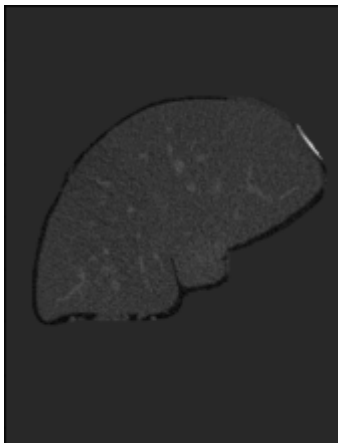


*Original Tumor Mask*



*Predicted Tumor Mask*

Overall Dice Score: 0.1873



*Original Scan*

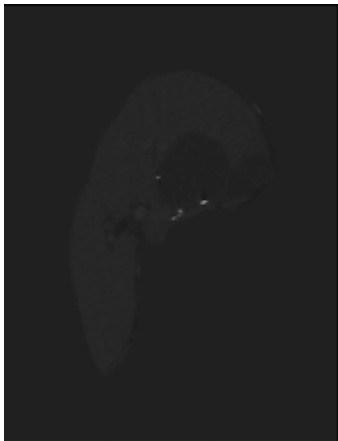


*Original Tumor Mask*



*Predicted Tumor Mask*

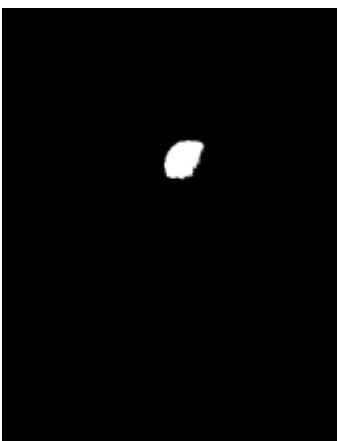
Dice Score: 0.5859



*Original Scan*



*Original Tumor Mask*



*Predicted Tumor Mask*

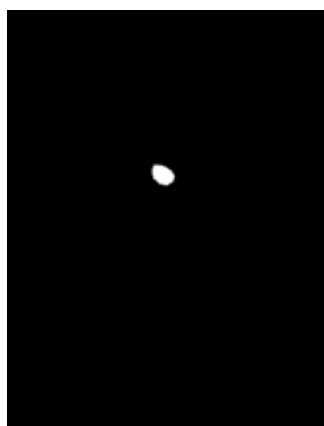
Here the model misses a huge part of tumor. From the same scan, the following slice was also visualized.



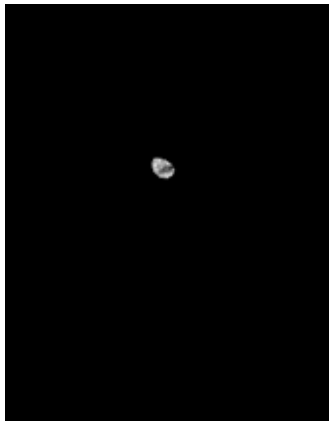
*Original Scan*



*Original Tumor Mask*



*Predicted Tumor Mask*

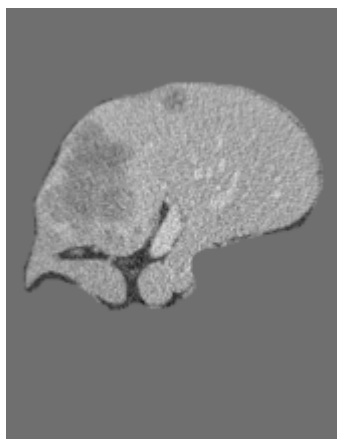


The above images are the beginning slices of a scan where the liver has just started to show, hence the liver size is small. Here, model predicts a tumor where there isn't one present, whereas it significantly misses the actual tumor from middle scans. This further proves that size is not a metric which affects accuracy of the model.

To see what the model was detecting, this predicted mask was mapped onto the original scan (shown on left). Slight discoloration is visible in the blot. On further evaluation, it

was found that the model predicted tumors on many starting and ending slices. This is because it is trained on middle slices and gets confused when given this input.

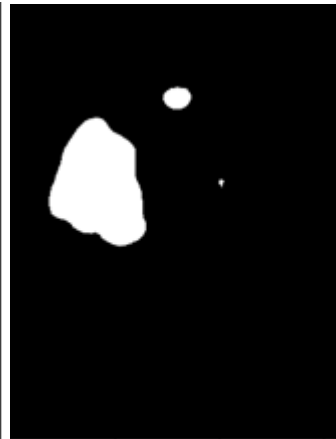
These scans have a high dice score of 0.8184



*Original Scan*



*Original Tumor Mask*



*Predicted Tumor Mask*

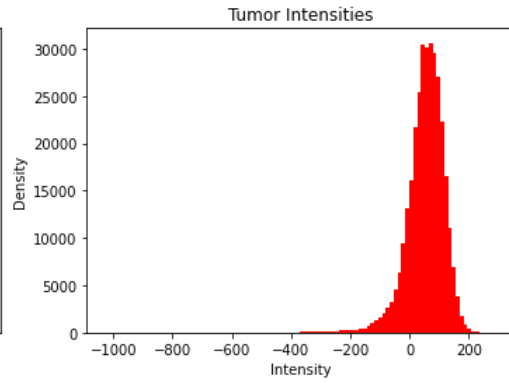
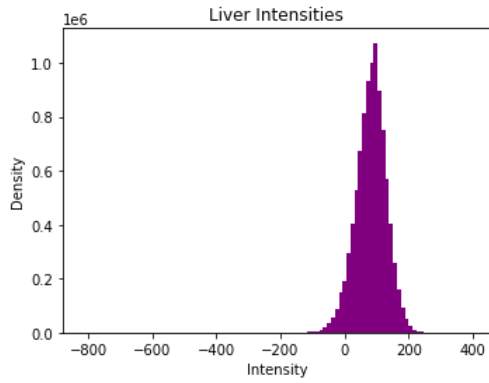
The prediction is close to ground truth. Some fine tuned edges can also be seen. An important detail to note is the contrast of these images. It is better than previous ones and the tumor is clearly visible as well. To further evaluate this, histograms of intensities of tumors and liver with different tumor dice scores were plotted.

Tumor dice score: 0.91072

Mean: 19.811237

Standard Deviation: 46.02159

Max: 259.0    Min: -676.0

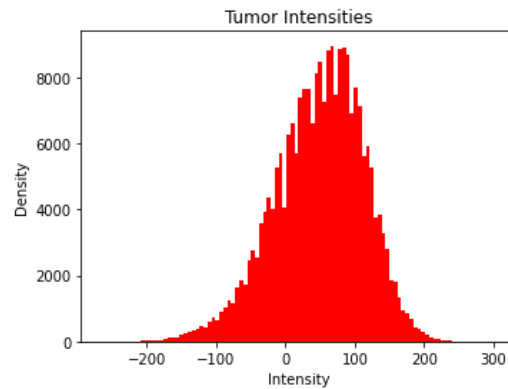
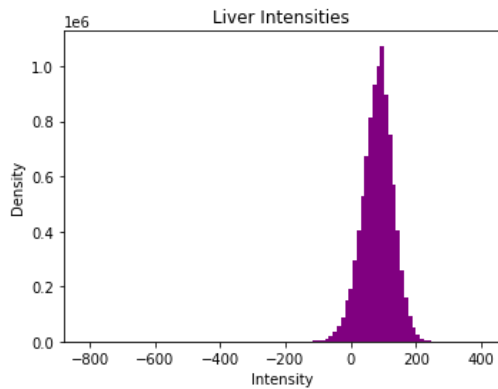


Tumor dice score: 0.44312

Mean: 37.61923

Standard Deviation: 59.75387

Max: 304.0    Min: -266.0

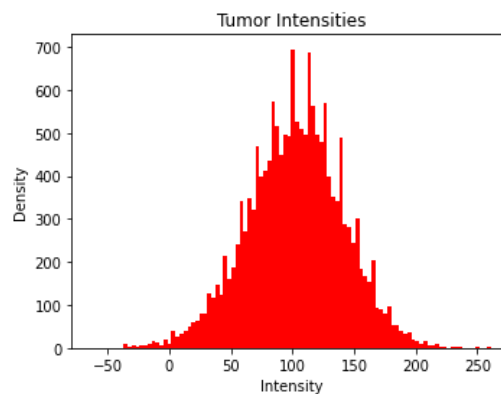
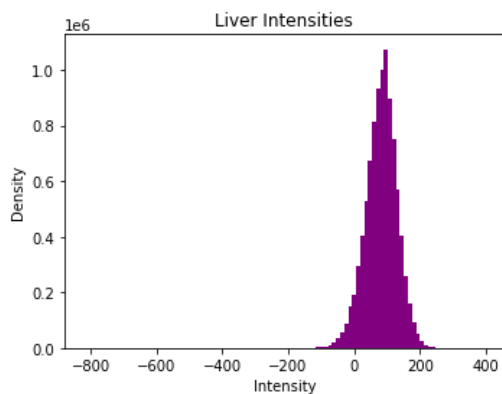


Tumor dice score: 0.15233

Mean: 100.83440

Standard Deviation: 40.39869

Max: 261.0    Min: -63.0

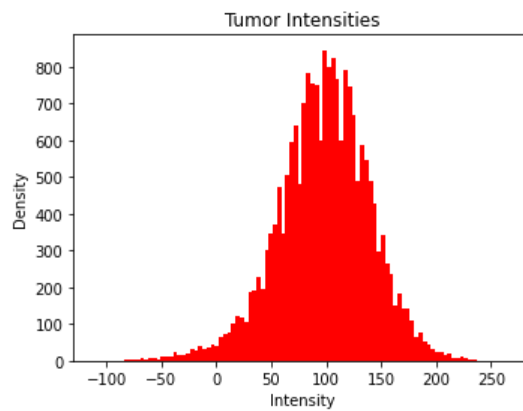
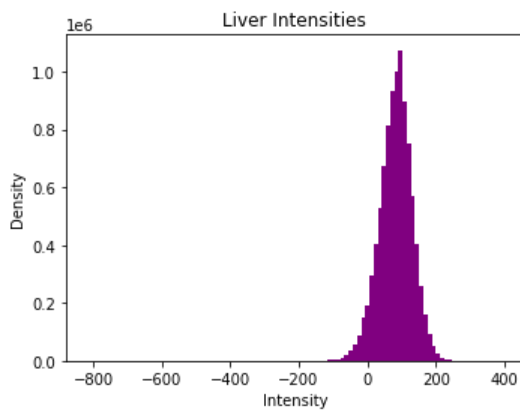


Tumor dice score: 0.07239

Mean: 97.93632

Standard Deviation: 42.10016

Max: 267.0 Min: -111.0

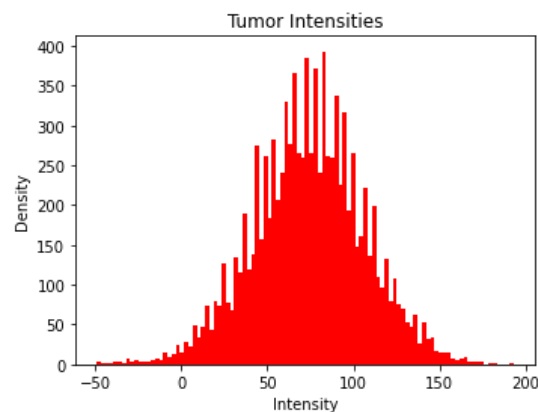
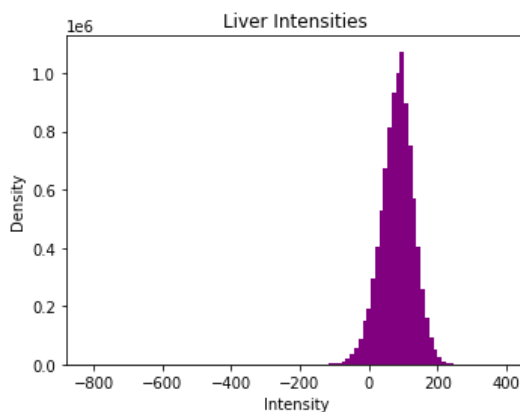


Tumor dice score: 0.000384

Mean: 73.79378

Standard Deviation: 32.61544

Max: 193.0 Min: -49.0



Smooth histograms means there is an even distribution of pixels in intensity bins. A low contrast image will result in a histogram with a large volume of pixels concentrated along a relatively narrow range of tones. A high contrast image will result in a histogram with a broad distribution along the tonal range. Thus, scans' histograms with better contrast are smooth while poor contrast scans' histograms will show spikes.

The above example scan has an overall dice score of 0.76144, with 9 tumors in the original scan and 9 tumors detected. As shown above on left graphs, the intensities of the liver are quite smoothly distributed, which means that the original scan has good contrast. As for tumor intensities shown on graphs on the right, the more the spikes in histogram, the poorer the dice scores. Spikes signify poor contrast. For higher dice score tumors, the graph is as smooth as the original liver scan. As the histograms get



more jumpy, the dice scores decrease. This concludes that contrast is an important metric in model performance and higher contrast images will give better dice scores.

### **Analysis**

Following statements have been concluded from the above results.

- Volume of tumor has little effect model accuracy, as volume of missed/misclassified tumors lie in the same range as detected ones.
- Liver scans with good contrast have better dice scores. This is because the tumor is better distinguishable and can be easily detected.
- Model gets confused with starting and ending slices of a scan. These are the beginning slices where the liver starts to show and vice versa. It classifies them as tumor or a large portion of it as tumor, hence giving a false positive.

### **Future Work**

To improve dice score, following procedures can be implemented.

- Apply dynamic windowing to scans which improves contrast
- Create a separate network which learns from liver scans with poor contrast. The output of which can be given to our model.
- Give some beginning and ending slices to model while training so that it does not get confused with them.