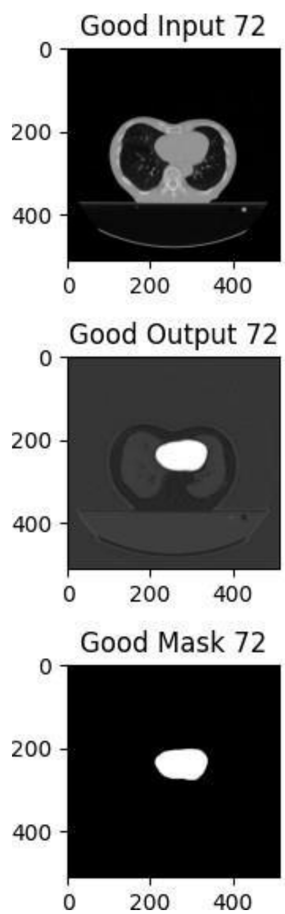
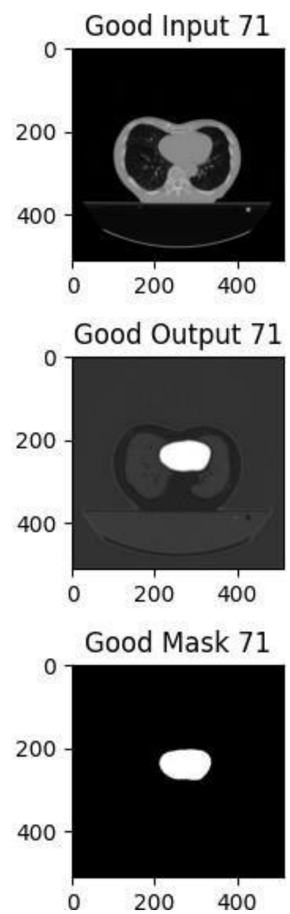
Kaan Dai 71651
Alp Akkanlar 76022

Medical Image Analysis HM3 Report

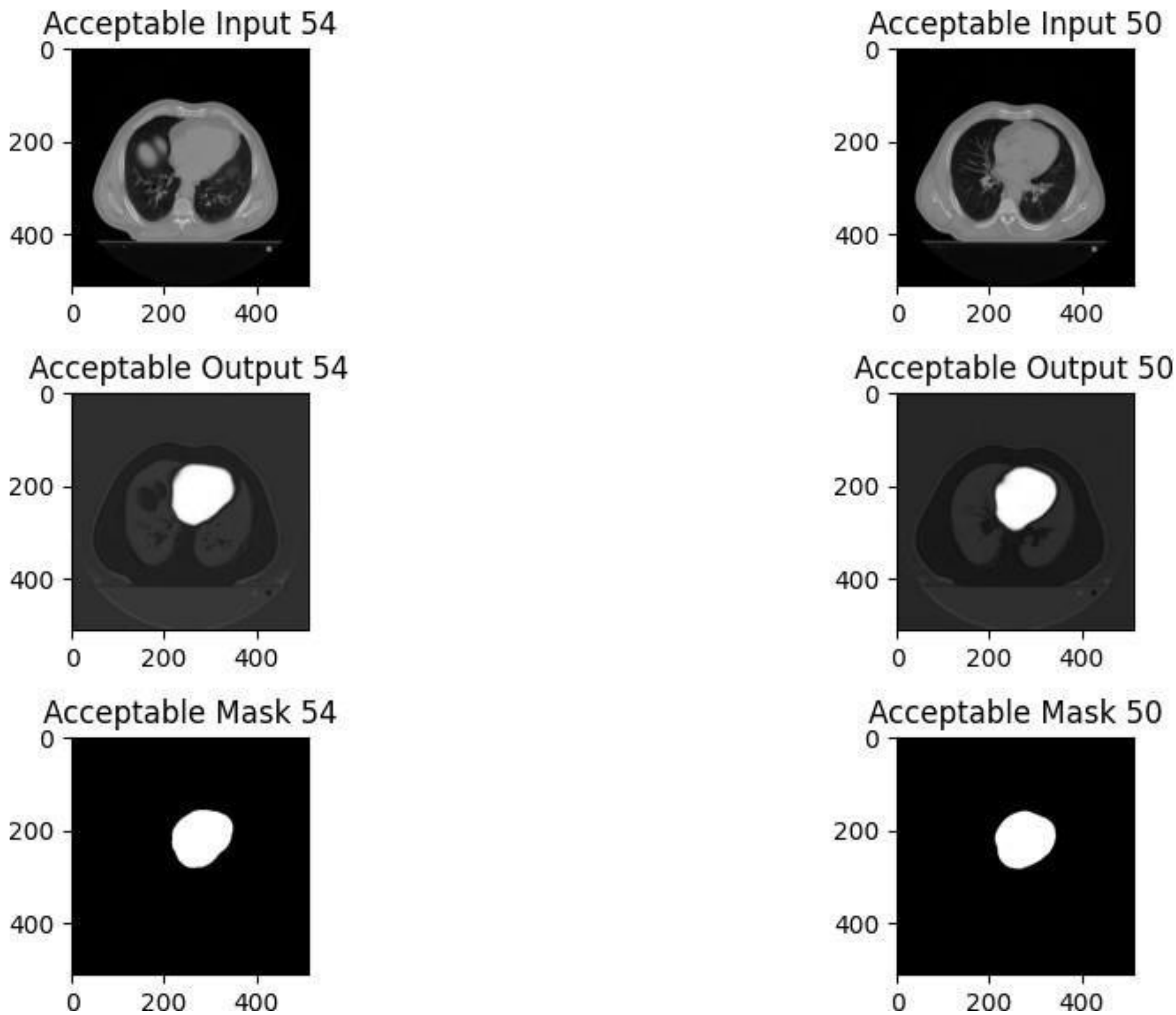
Part 1

We chose the optimizer as “Adam”. We chose it for its efficiency and ability to handle sparse gradients on noisy problems. The batch size used in the training is 32. We selected it based on a balance between memory capacity and the desire for quicker convergence. The training process used early stopping as a stopping condition. This monitors the validation loss and if it does not improve for a certain number of epochs, the training is stopped to prevent overfitting. The best model is saved for predicting the outputs for the images in the test set.

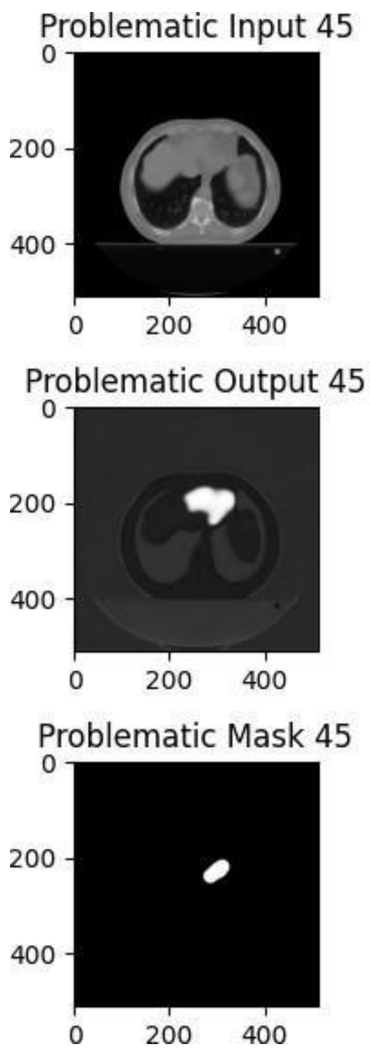


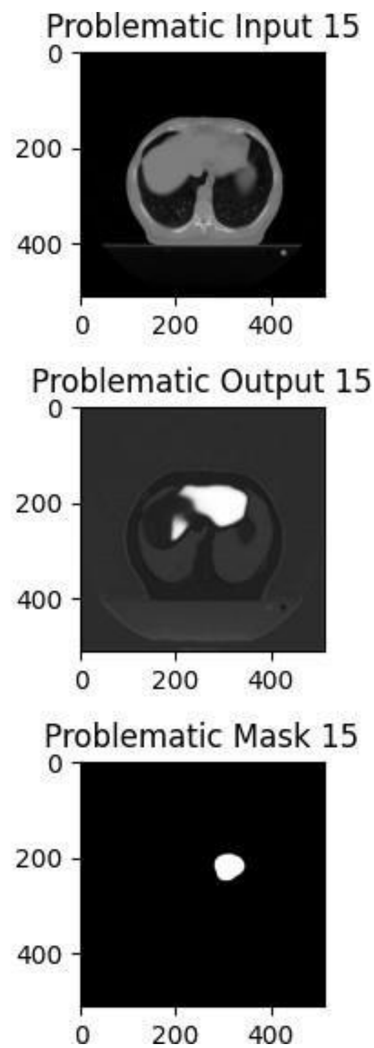


The model effectively captures the boundaries and details of the target objects, resulting in high precision and recall. In this image, the segmentation is clear and precise, with minimal noise or misclassification. The model demonstrates strong performance in identifying and segmenting the object of interest.



The segmentation result shows the model's ability to identify the primary regions of interest, but there are minor inaccuracies around the edges. This image illustrates an acceptable segmentation where the model correctly segments the main object but includes some extraneous regions. The recall is relatively high, but the precision is affected due to the inclusion of non-relevant areas.





The model struggles to accurately segment the object in this image, resulting in significant misclassification and noise. The boundaries are poorly defined, and the segmented area does not correspond well with the ground truth. In this example, the segmentation is highly inaccurate, with the model failing to capture the essential features of the target object.

Part 2

	Training Set			Validation Set			Test Set		
	Precison	Recall	F-score	Precison	Recall	F-score	Precison	Recall	F-score
Default network design	0.9742	0.9945	0.9842	0.8519	0.8981	0.8744	0.9091	0.9561	0.9320
Modification in the number of down and up sampling steps	0.9408	0.8931	0.9163	0.6083	0.8502	0.7091	0.8907	0.8427	0.8660
Modification in the number of feature channels	0.9721	0.9693	0.9707	0.8852	0.8829	0.8840	0.9147	0.9326	0.9236

UNet with 32 feature channels performs the best segmentation compared to other methods. The 32 feature channels provide a greater capacity for learning complex patterns and details in the images. The increased number of features helps in generalizing better. For segmentation, 8 Feature UNet has reduced performance and 32 Feature UNet demonstrates superior segmentation performance with high precision and recall. For training time, 8 Feature UNet has shorter training time due to reduced complexity and 32 Feature UNet has longer training time due to increased complexity and higher number of feature channels. For the gap between, 8 Feature UNet has a smaller gap due to underfitting but generally poor performance on both training and test sets and 32 Feature UNet has the smallest gap due to better generalization capabilities and sufficient capacity to learn from the data without overfitting. We prefer 32 Feature UNet among the three methods because it achieves the best segmentation performance, the architecture generalizes well to unseen data, showing minimal gap between training and test set performance and the increased number of feature channels allows the network to capture detailed and complex features.

Part 3

	Training Set			Validation Set			Test Set		
	Precison	Recall	F-score	Precison	Recall	F-score	Precison	Recall	F-score
Default network design	0.9742	0.9945	0.9842	0.8519	0.8981	0.8744	0.9091	0.9561	0.9320
Network w/dropout (p= 0.1)	0.9059	0.9812	0.9420	0.7412	0.9071	0.8158	0.8458	0.9703	0.9420
Network w/dropout (p= 0.3)	0.9920	0.6396	0.7777	0.9603	0.5401	0.6913	0.9715	0.6384	0.7705
Network w/dropout (p= 0.5)	0.9910	0.0035	0.0070	0.5190	0.0008	0.0012	0.9529	0.0203	0.0397

Yes, we observed the regularization effect after integrating the dropout layers. There is a noticeable decrease in validation loss compared to the model without dropout. Precision, Recall and F-score metrics improve on the validation set showing that the model can perform well on unseen data. The dropout rate of 0.3 performs the best segmentation compared to the other p-values because optimal regularization and balance between dropout and learning. Yes, we can say that increasing the p-value in the dropout layers decreases the difference between training and test set performance results. For 0.5 dropout rate, the results in more neurons being dropped during training which can lead to a significant reduction in overfitting. This reduces the performance gap between the training and test sets as the model becomes more robust to unseen data. For 0.3 dropout rate, it still provides effective regularization while maintaining a good learning capacity which is leading to the best observed performance. For 0.1 dropout rate, it provides minimal regularization leading to overfitting and a larger gap between training and test performance.