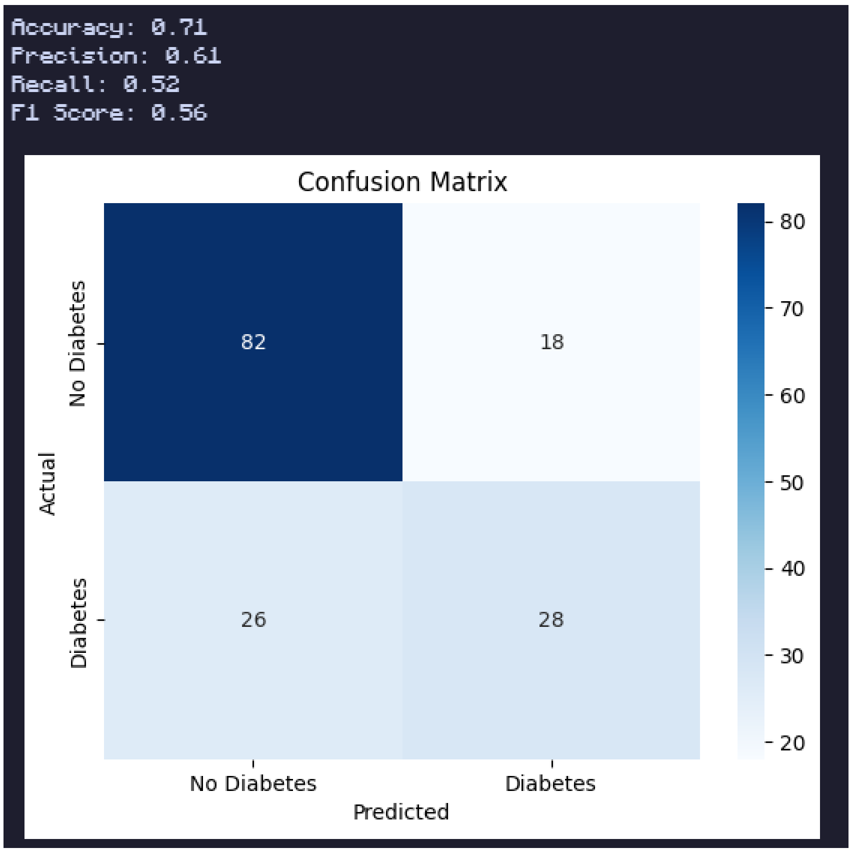
Homework 3

ECGR-4105-001

Joshua Foster, 801268118

Repo Link: https://github.com/jfoste81/ECGR4105_Homework/tree/main

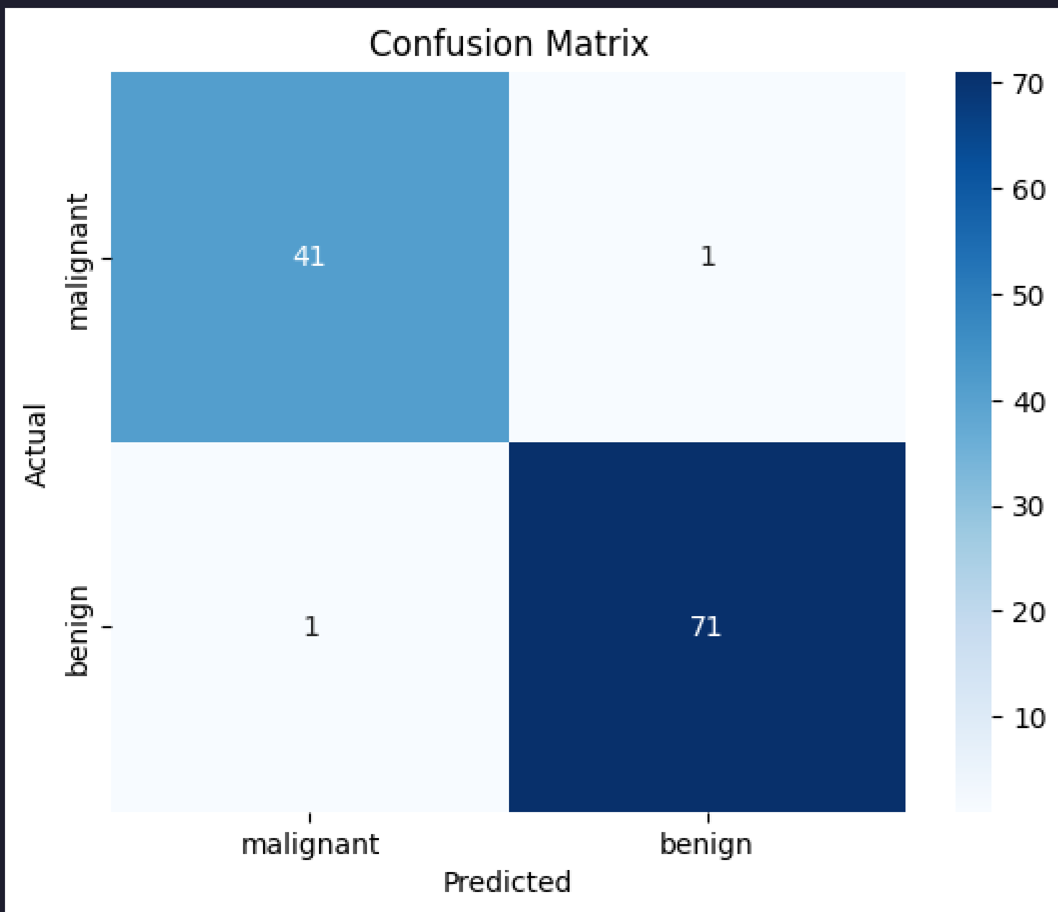
Problem 1:



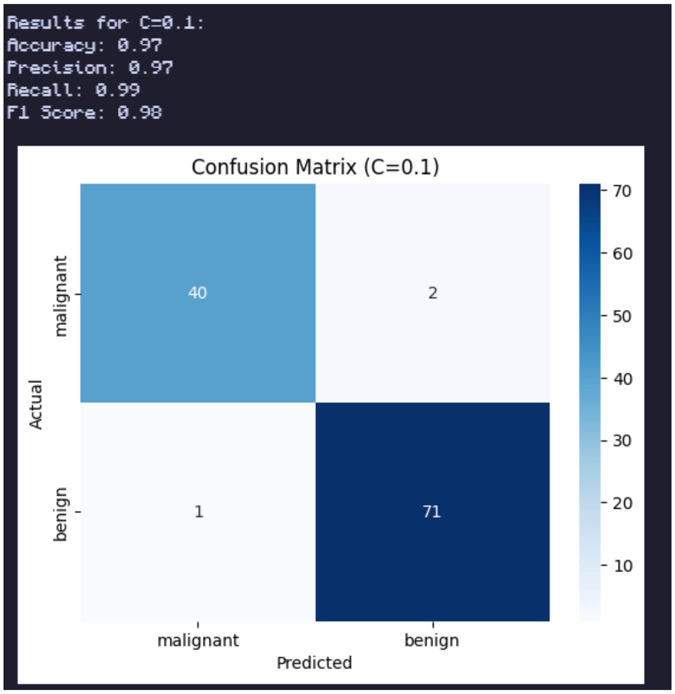
We can see from the precision and recall that the model struggled a bit with false positives (18/100 total samples) and false negatives (26/54 total samples) but definitely struggled with false negatives more. Which is reflected in the lower recall score. The model performed very well with identifying the absence of diabetes but struggled to properly identify diabetes positive states.

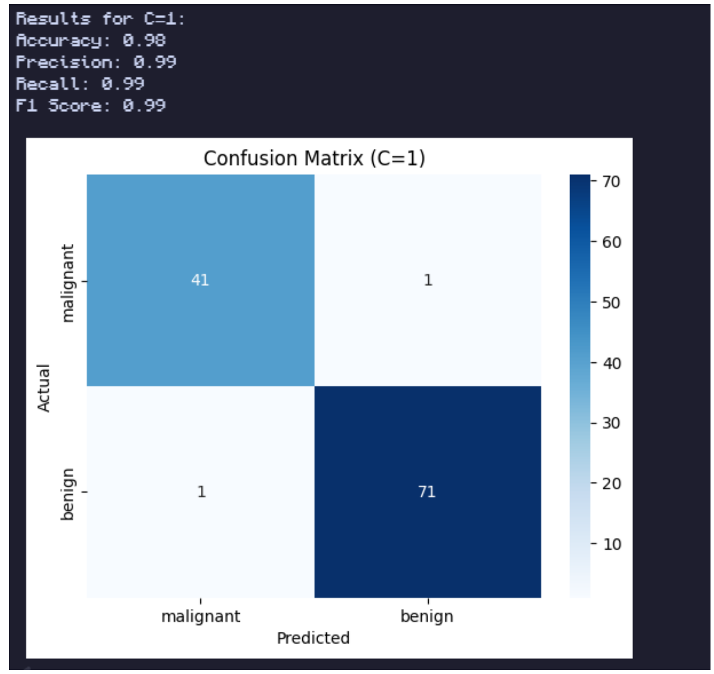
Problem 2:

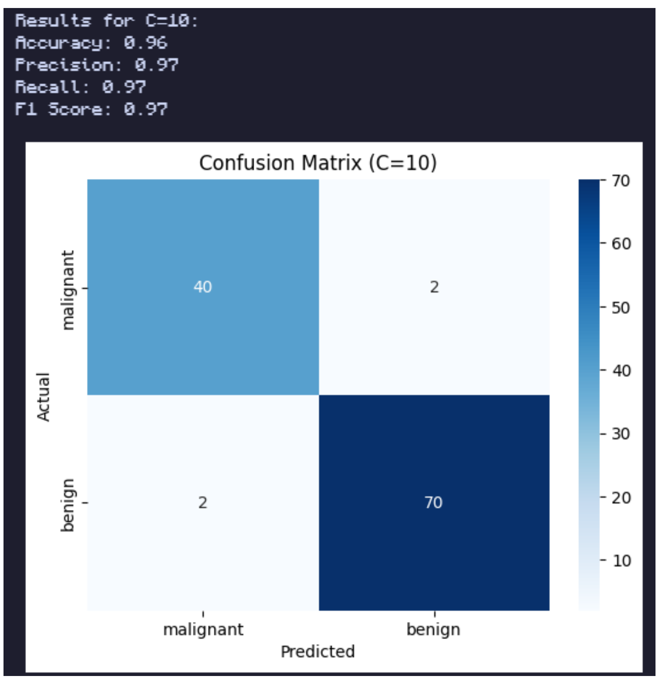
```
Accuracy: 0.98
Precision: 0.99
Recall: 0.99
F1 Score: 0.99
```



a) The model performed extremely well, almost perfectly, in this scenario. Despite using all 30 features, the model was able to learn very well and only had one false positive and false negative. Although, in a real world setting we would want to reduce the amount of false negatives as that would mean that we are missing malignant tumors.







b) I experimented with different weight penalty values (.1, 1, and 10) and found that the small weight penalty (.1) actually made performance drop by introducing an additional false negative, which is exactly what we are trying to avoid the most. Additionally, we find that the large penalty (10) performs the worst as it increases the amount of both false positives AND false negatives. Thus, we see that the middle weight penalty (1) performs the best, even though it doesn't necessarily increase the performance of the model overall.