**Problem statement:** 304 data points of heart disease containing 14 features were analyzed with the goal of predicting presence of heart disease.

Five models were tested and trained with this dataset. SVC, Logistic regression, Random Forest, and BaggingClassifier with an SVC base. The parameters were tuned using GridSearchCV, where a multitude of parameter combinations were compared.

In an article written by Beulah et al, methods to improve heart disease predictions were evaluated. They found that boosting added an extra 7.26% accuracy to their dataset (Beulah). Just like this project, the paper used the Cleveland dataset as well.

After trying the boosting method, testing accuracies rose to around 85.24 from the previous high of 81.967%.

This is a substantial boost in accuracy, and will be the main model that I will use in my final prediction model the web app I will create. Note, that there is a big difference between accuracy and precision. Although the model has a 89% accuracy, it does not imply it has a great precision score. For example, If a woman has breast cancer, a breast cancer test is positive with a probability of .95 P(Pos/Breastcancer)=0.95

-If a woman does not have breast cancer, test will be negative with probability .95.P(Neg/NoBreastcancer)=0.95

-If we supposed 5/1000 women have breast cancer in the population.

**-However,** if a woman tests positive for breast cancer, the probability she actually has breast cancer is only 0.087! P(BreastCancer/Pos)=0.087.

This presents an issue with the heart disease predictor, where accuracies of even 89% are not sufficient for determining prevalence of breast cancer in a reliable way.

The accuracies can be further improved upon with a larger sample size. With a sample size of n=304, there is simply not enough data to be able to have an accurate prediction for every population.

Next Steps: Implementation of bagging classifier to other models apart from svc will be tested. Further models will also be tested, as there are only 4 so far. My goal is to find a model with parameters that have an accuracy of at least 90%.

Bibliography:

1. <https://www.sciencedirect.com/science/article/pii/S235291481830217X>