**Test explanation**

A *p* value < .01 was chosen as significant.

## Differences between blood types

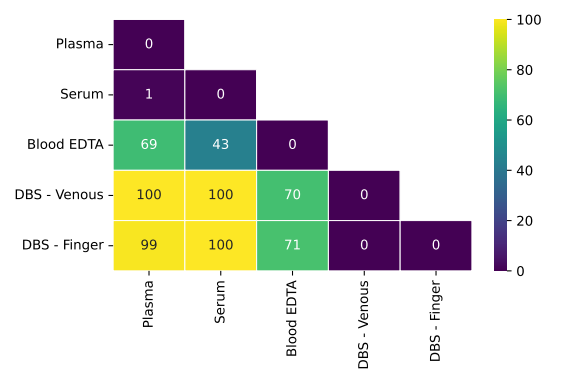
The next experiment shows if the concentration of a lipid differs in the blood types. For that, the date and person categories were omitted, and all test samples of a blood type were used as one data set. All blood types were compared to each other. Each lipid, which was measured in both blood types in every person at all time points, was part of the analysis. As the lipid concentration of one person of different blood types are dependant a Wilcoxon signed-rank test was performed, eg.g a high value in Plasma should also lead to a high value in Serum.

For comparison were EDTA was one of the blood types the second date was not used in the Wilcoxon test.

The comparison was computed for every lipid individually and the number of times H0 got rejected was counted. It is visible that Plasma and Serum are similar to each other in respect to lipid concentration. Additionally, DBS Finger and DBS Venous share similar concentrations. On the other hand Plasma and Serum are different to DBS Venous and DBS Finger. Blood EDTA has many lipids with a different distribution to each of the other blood types.

### Negative

127 lipids were tested.

Figure 1: Comparsion of mean of negative mode lipids between blood types. The number of times H\_0 got rejected was counted.

### Positive

Figure 2: Comparsion of mean of positive mode lipids between blood types. The number of times H\_0 got rejected was counted.

Positive = 197 lipids.

#### Differences between time points:

For the next experiment the person information was omitted and all samples at one time point were treated as one data set. The data sets are dependent, because the samples were taken from the same persons at different dates.

A Friedman test was used for all blood types, except Blood EDTA, because the measurements of three dates were compared. For Blood EDTA the Wilcoxon signed-rank test was used, because only two dates were available.

The statistical tests were performed per lipid and it was counted how often H0  got rejected. No highly significant difference arose. A table with only zeros is kind of useless, maybe a textual explanation of the experiment is enough.

**Caveat:** The compared samples in this experiment are very small and only three repeated measurements were compared.

Table 1: Number of lipids where the H\_0 hypothesis that all time points share the same mean could be rejected.

|  |  |  |
| --- | --- | --- |
| **Blood Type** | **Positive** | **Negative** |
| Plasma | 0 | 0 |
| Serum | 0 | 0 |
| Blood EDTA | 0 | 0 |
| DBS - Venous | 0 | 0 |
| DBS - Finger | 0 | 0 |

#### Differences between Persons:

To measure if the lipid concentration is significantly different between persons, we again omit the date information and treat all time points as one. This time we use the Kruskal-Wallis test, because we the samples of different persons are independent and we compare four different samples. Again, we count the number of times H0 got rejected. In no case the *p* value was small enough to reject the hypothesis that all persons share the same distribution of lipids.

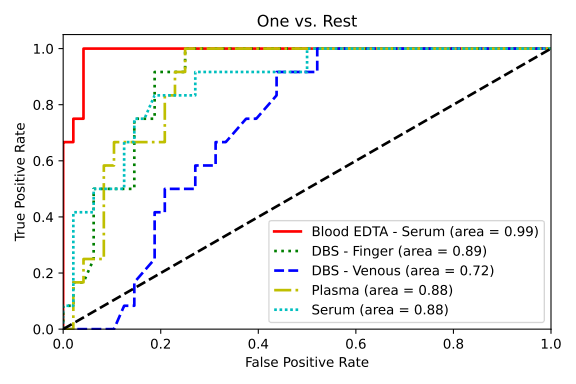


#### Classification:

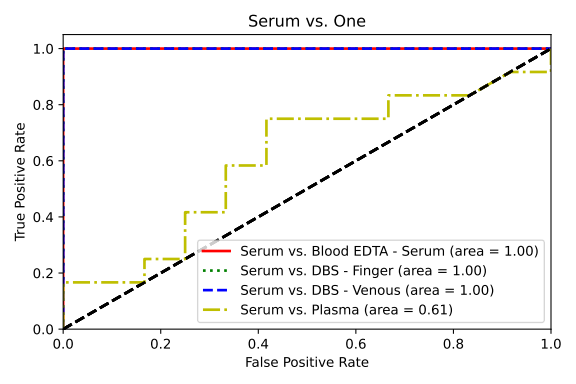
A classification random forest was trained in all experiments. The random forests were trained with a 3x4 cross-validation, where every fold contains the data points of one person. This split avoids information leakage from the training data to test data, which would occur if data samples of the same person would be in the training and in the test set.

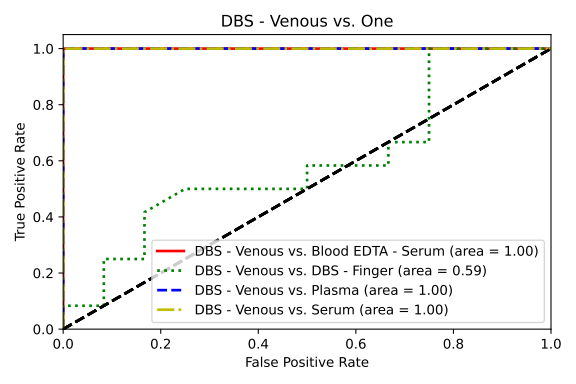
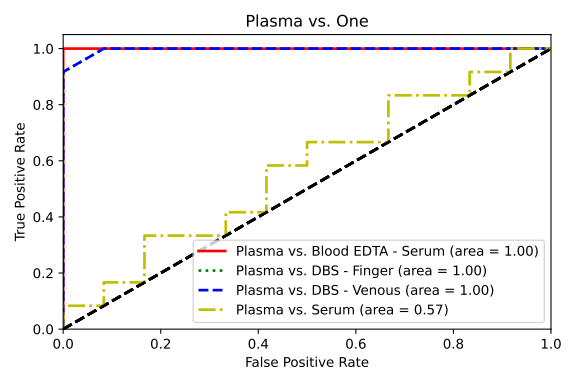
**Multiclass**: one random forest was trained to classify samples in to one of all possible blood types. Gives an overall overview. AUROC of 0.92.

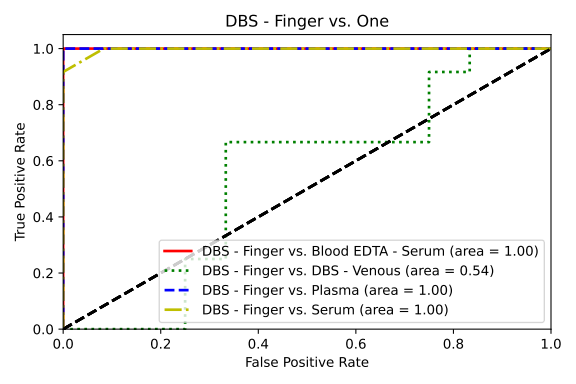
**One vs .rest:** For every blood type a random forest was trained. The samples of the corresponding blood type were treated as positive class and all remaining samples as negative class. This gives an overview of of the differences of one blood type to all other.

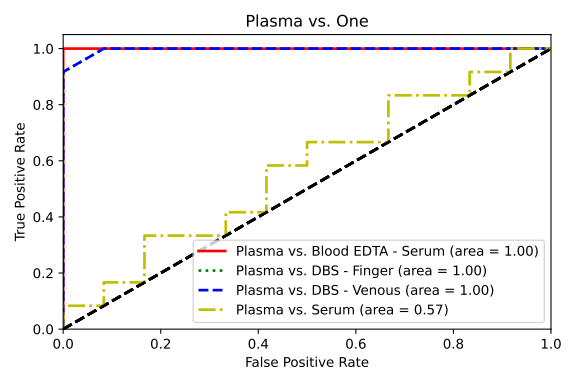
It is visible that the classifier reaches high AUROC values for all of the blood types. Furthermore, it can almost perfectly distinguish between blood EDTA and all other blood types.

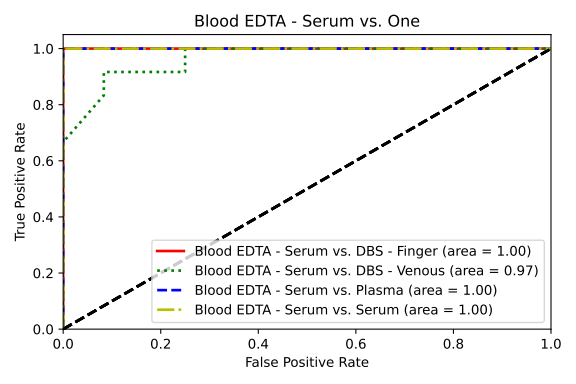
**One vs. one:** For every blood type for different random forests were trained. For every forest the samples of the corresponding blood type were used as positive class and the samples of one different classifier as negative class. This is suitable to analyse differences between individual blood types.









The classification one blood type vs. one blood type confirms our previous results. (AUROC = 0.5 = random, AUROC = 1 = perfect classification)

Considering all results together imply that there is significant difference in samples taken at different time points and between participants. As mentioned earlier the number of participants and repetitions are small and the results need to get verified in a large scale study.