# coursework B

*Alexander Bieniek, David Nyrnberg, Wesley Quispel*

*March 15, 2019*

## Introduction

In this report, we investigate the effectiveness of transfer learning, a technique in which a model can be trained to solve one problem, and later refiend to solve another problem. In particular, a non-transfer model is trained on its given data alone, using cross validation. The transfer-based models are trained on other training data, and possibly after some refining, the model is tested on the same data that the non-transfer model is tested on. We have been given an abstract dataset which provides metadata about models performance in given experiments: * TeD - an identifier denoting which test data a model was tested on (TeD1... TeD7) * TrD1 - whether or not the model was trained with Training Data 1 * TrD2 - whether or not the model was trained with Training Data 2 * ... * TrD8 - whether or not the model was trained with Training Data 8 * model - an identifier describing which model was used in the experiment * score - the accuracy of the model on the given test data

Here is a set of summaries for the various features of the data on which these models wre trained and tested, as well as notes on the features of the models themselves, as described in the coursework prompt:

Models tested in the experiment: * B1 - Baseline model without transfer learning * B2 - Baseline model without transfer learning * B3 - Baseline model without transfer learning * S - Transfer learning model, using one transfer training set, refined for test set * MN - Multiple transfer training set model, without trtraining for test set * M1 - Multiple transfer training set model, with 1 part refined for test set * M2 - Multiple transfer training set model, with 2 parts refined for test set * M3 - Multiple transfer training set model, with 3 parts refined for test set * MF - Multiple transfer training set model, with all parts refined for test set

Testing sets: * TeD1 - Classification * TeD2 - Classification * TeD3 - Classification * TeD4 - Classification * TeD5 - Recommendation * TeD6 - Regression * TeD7 - Regression

## RQ1: Improvement in Transfer Learning

How much does transfer learning improve over typical non-transfer learning?

The following plots create histograms of performance for all of the transfer learning models on a given test set. Upon the histogram, we superimpose a line showing the performance of one of the three baseline models:

- red - B1
- green - B2
- blue - B3

```r
# dictionary mapping baseline model to its line color in the histograms
colorDict = Dict$new(c("B1" = "red", "B2" = "blue", "B3" = "green"))

# produce a new histogram of accuracies for each test set
for (ts in testSetNames) {
  blData = data[which(data$TeD==ts & data$model %in% baselineNames),]    # baseline models
  tlData = data[which(data$TeD==ts & !(data$model %in% baselineNames)),] # transfer models
  hist(tlData$score, xlim=c(0.0,1.0), breaks = 10,
       main=paste("Performance Histogram of Transfer Learning Models on", ts),
       xlab="accuracy")
  # superimpose a line for each baseline model
```
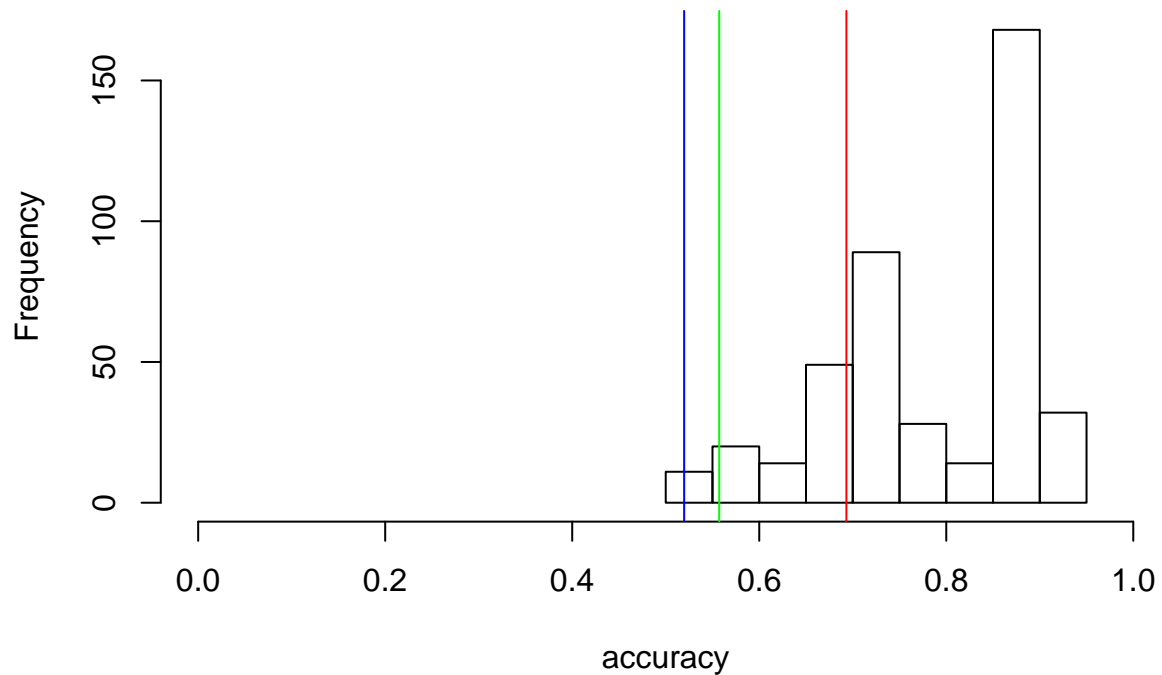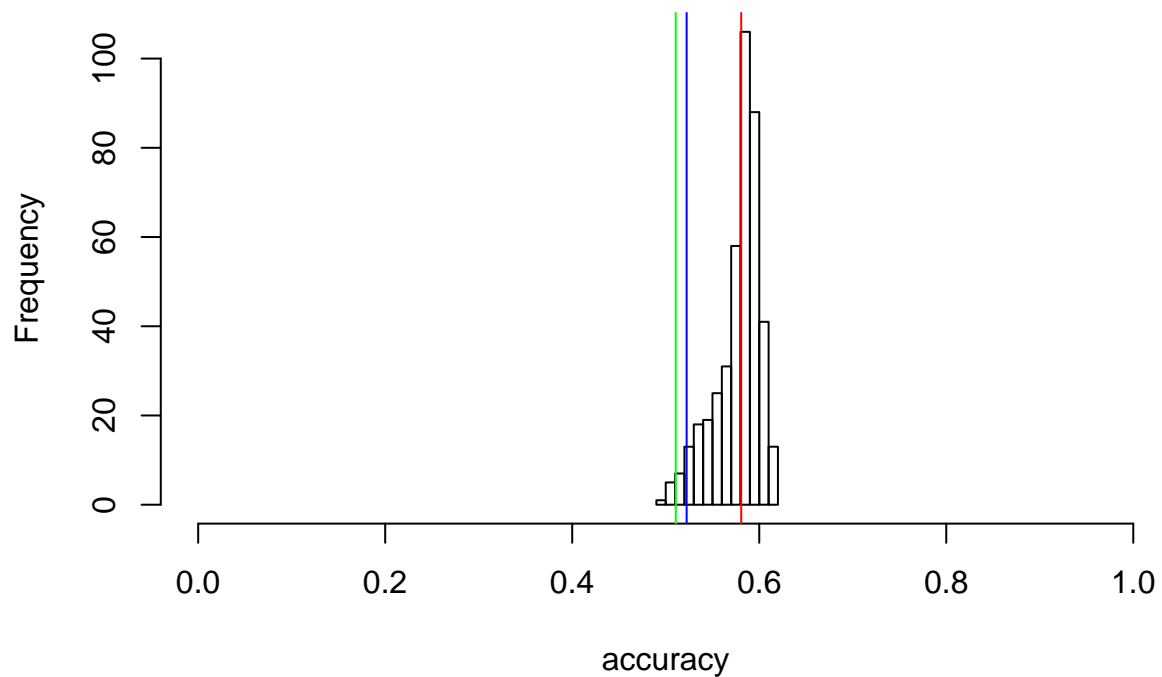
```
for (i in 1:nrow(blData)) {
  abline(v=blData[i,]$score, col=colorDict[blData[i,]$model])
}
}
```
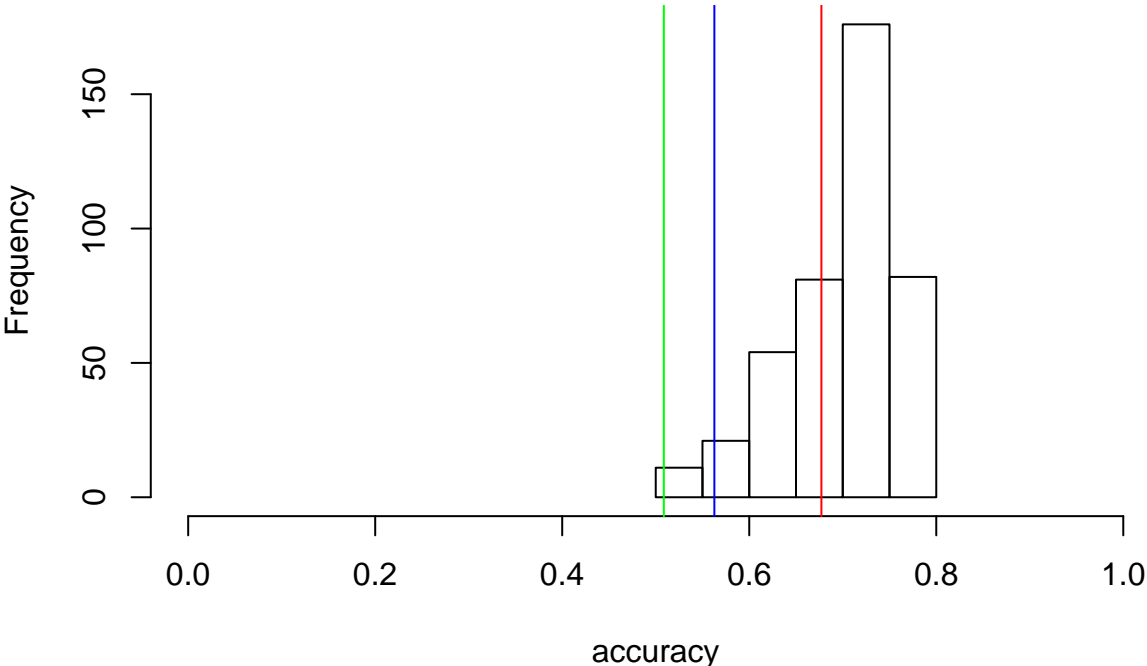
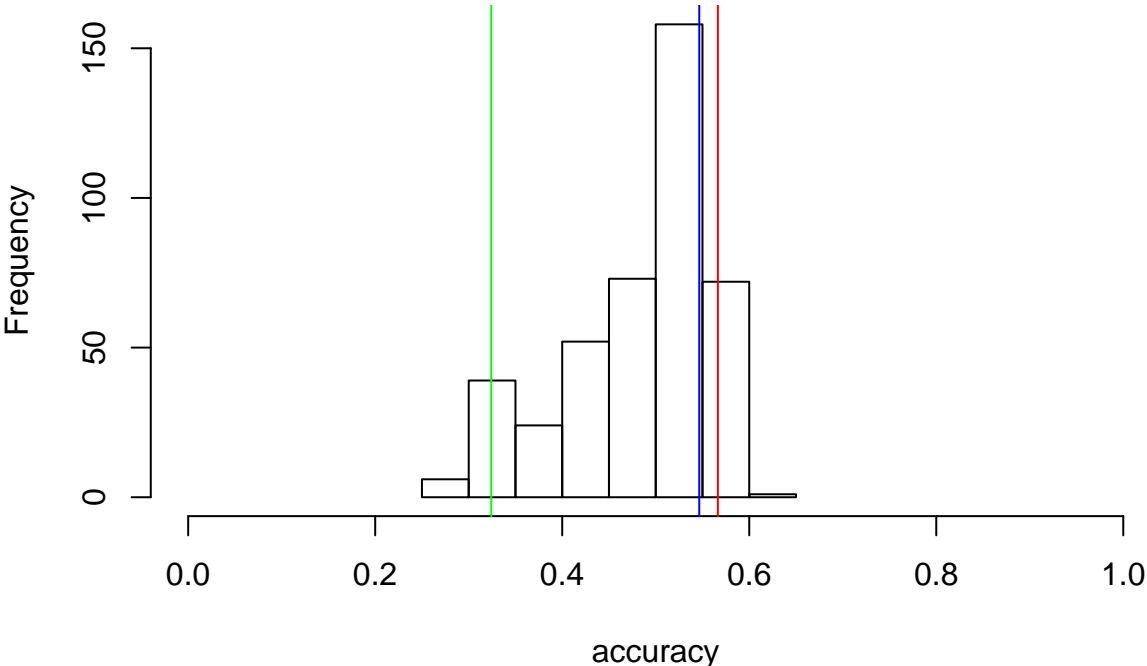## Performance Histogram of Transfer Learning Models on TeD1



## Performance Histogram of Transfer Learning Models on TeD2
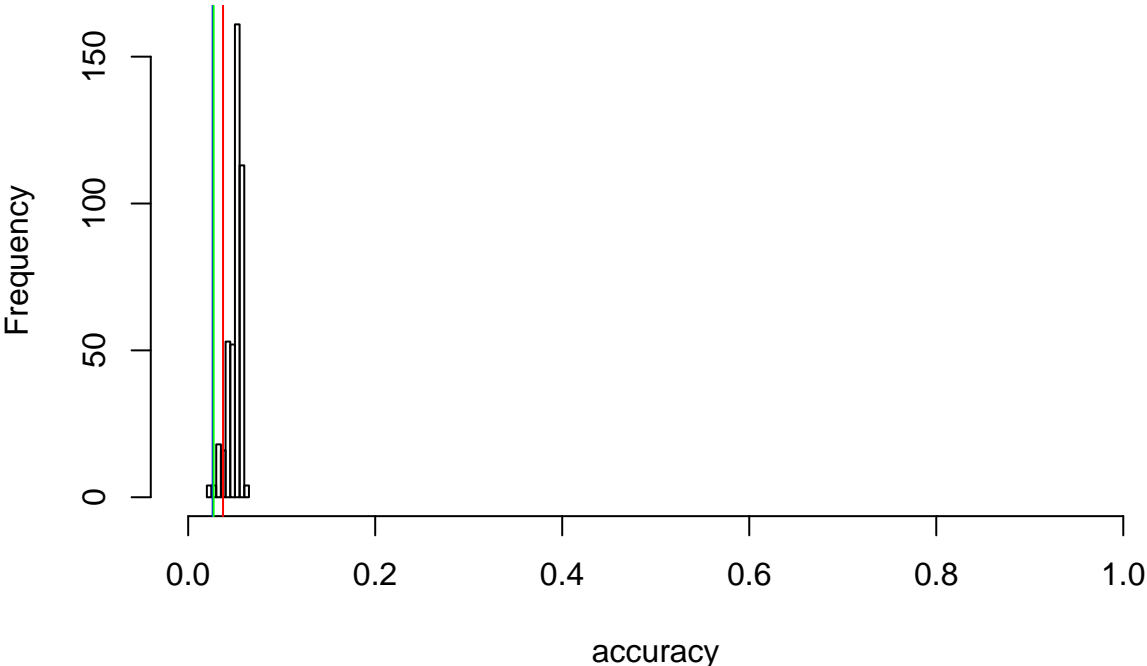
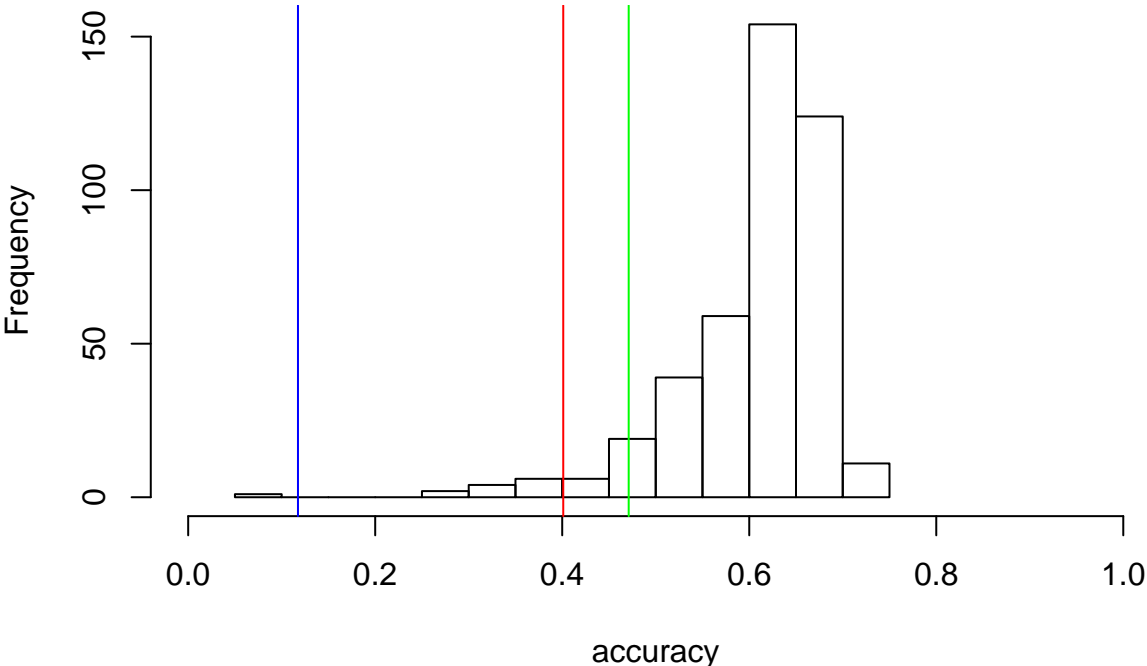# Performance Histogram of Transfer Learning Models on TeD3



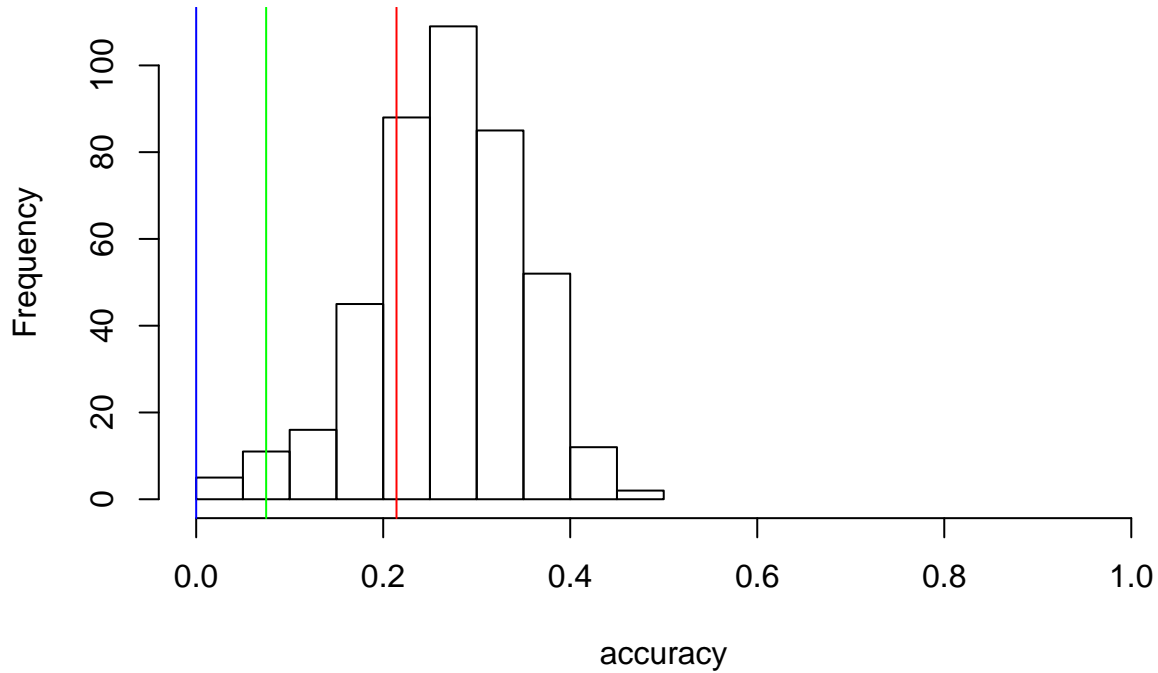# Performance Histogram of Transfer Learning Models on TeD4

# Performance Histogram of Transfer Learning Models on TeD5



# Performance Histogram of Transfer Learning Models on TeD6

## Performance Histogram of Transfer Learning Models on TeD7



accuracy

These visualizations tell quite different stories for each training set:

- No baseline model outperforms the other baseline models in every test set
- In most cases, the base models have performances that are lower than the most common performances among the transfer learning models
    - In TeD4, B1 and B3 appear to outperform the mode accuracy of the transfer learning models
- The models produced for the given tasks result in very different accuracy distributions
    - Most distributios appear to be unimodal and left skewed, although tests on TeD1 produce a bimodal accuracy distribution
    - Performance distributions have different medians, reaching as high as 80% and as low as ~5%

This preliminary visual analysis tells us that baselines tend to perform worse than the possibilities which can be achieved by the transfer learning models. However, the test sets appear to be very very different tasks, as their accuracy distributions are quite dissimilar. This visualization shows that, at the right most regions of our distributions, there are TL models which outperform the baseline models in all cases as well as the rest of the TL models, but we have yet to discover whether these high-performing models have any similarities.

### RQ2: Effect of Learning Strategy

What is the effect of different strategies to simultaneously learn one model from multiple TrD's?

### RQ3: Effect of Training Datasets

What is the effect of the TrD's on the final model performance?