Ian Sam
V00990790

STAT 469 Assignment #3

Introduction

This report evaluates the performance of four predictive models—LDA, Elastic Net, Random Forest, and MTPS (Residual Stacking)—on binary resistance outcomes for five drugs (ABC, 3TC, AZT, D4T, DDI). The objective is to compare the effectiveness of these models using F1-scores and determine whether the performance differences between them are statistically significant.

The evaluation is performed using 10 random splits with 5-fold stratified cross-validation, ensuring balanced representation of resistance and non-resistance cases. Pairwise Wilcoxon signed-rank tests are used to assess whether the differences in F1-scores between Random Forest, MTPS, LDA, and Elastic Net are statistically significant

Methods Used

The models used in this study include LDA, Elastic Net, Random Forest, and MTPS (Residual Stacking). LDA is performed using the lda function from the MASS package. Elastic Net is applied using the cv.glmnet function with alpha set to 0.5. Random Forest is implemented using the randomForest function with 100 trees. MTPS (Residual Stacking) is applied using the MTPS package, leveraging residual-based multi-task learning. Each model is trained and evaluated using stratified 5-fold cross-validation, repeated across 10 different random splits to ensure robust performance evaluation.

Results

After training and testing the models, I analyzed their performance across different drugs. The median F1-scores for each model and drug were computed to compare their effectiveness. To better visualize the results, I generated boxplots showing the distribution of F1-scores for each model.

F1 Scores for ABC

F1 Scores for 3TC

F1 Scores for AZT

F1 Scores for D4T
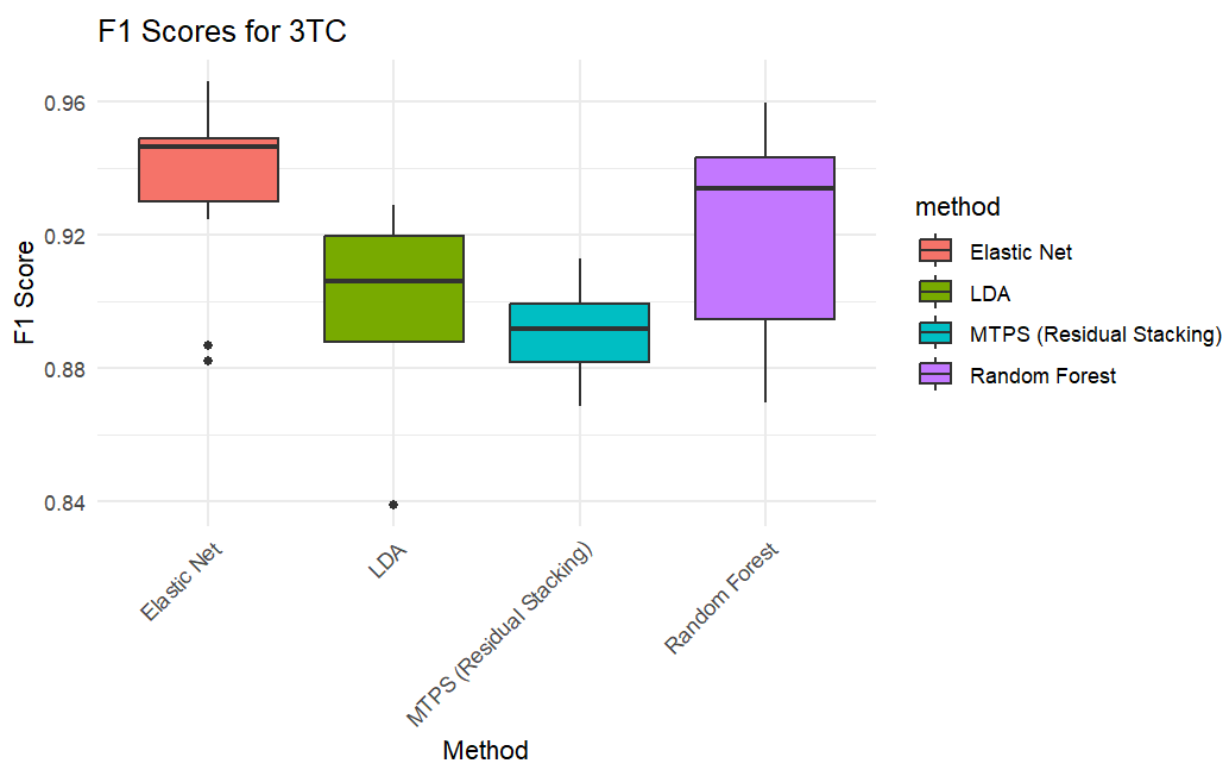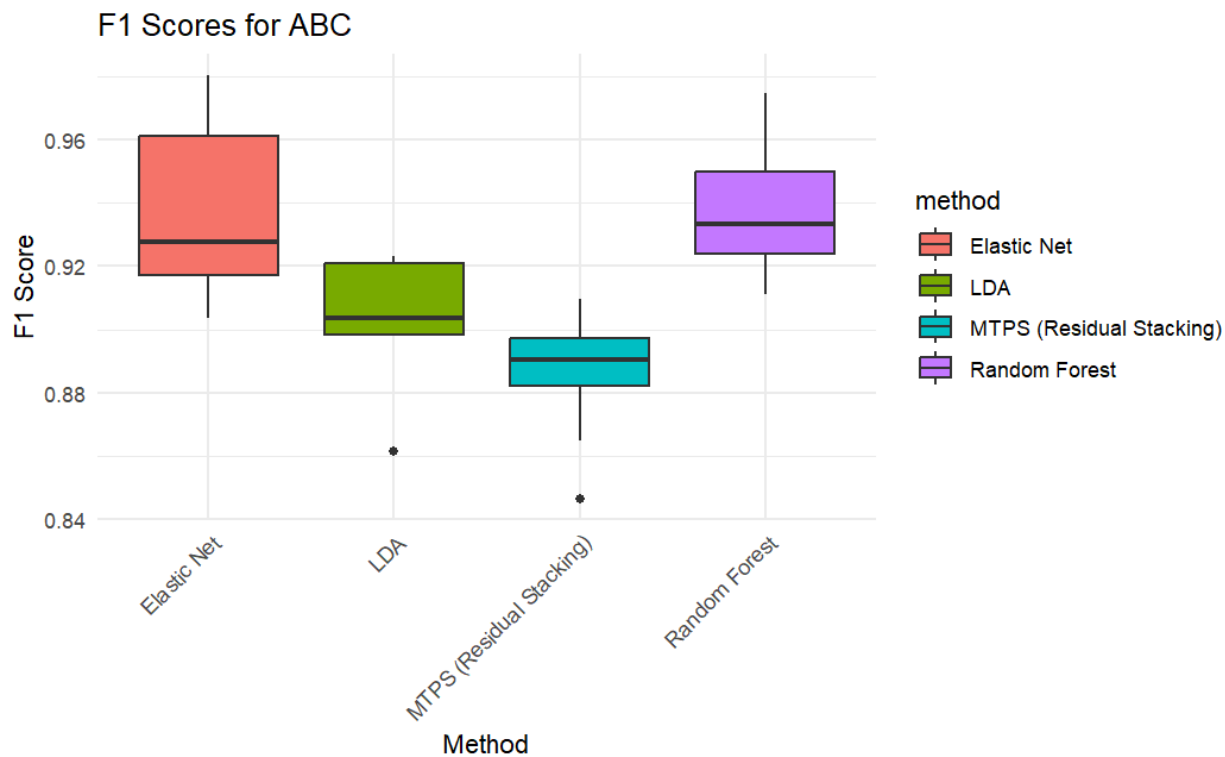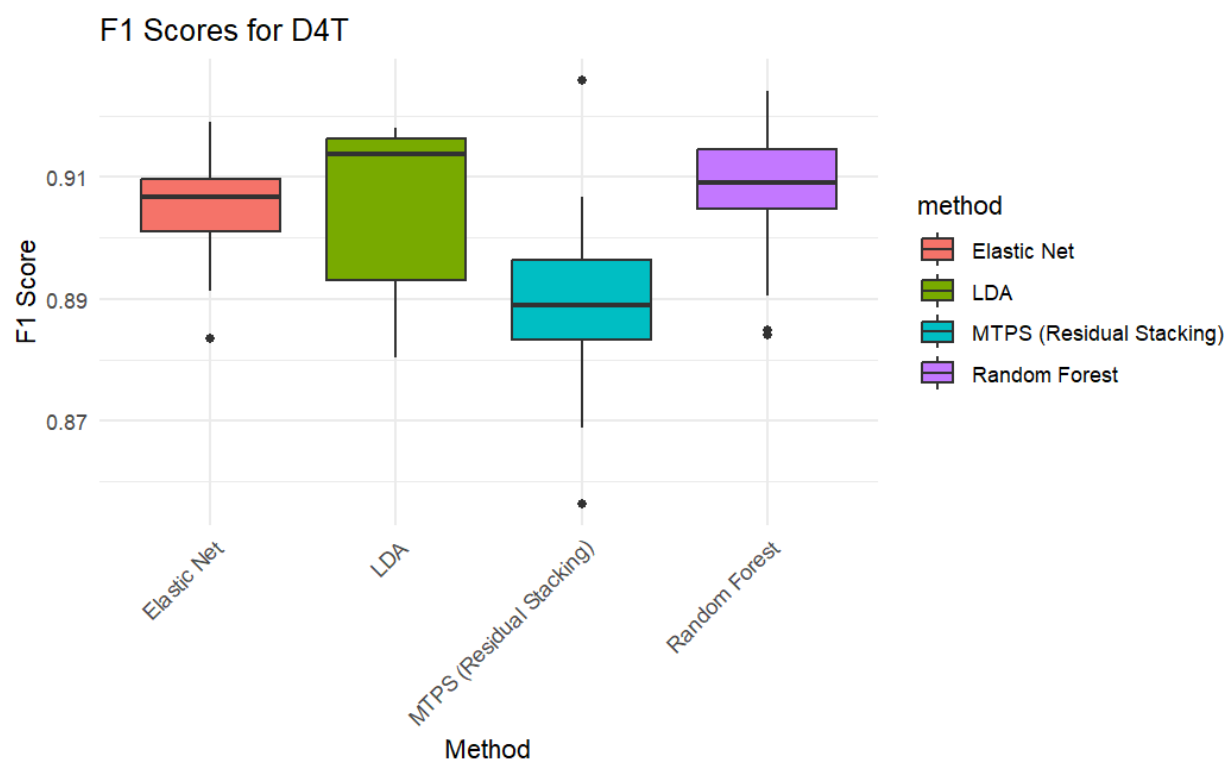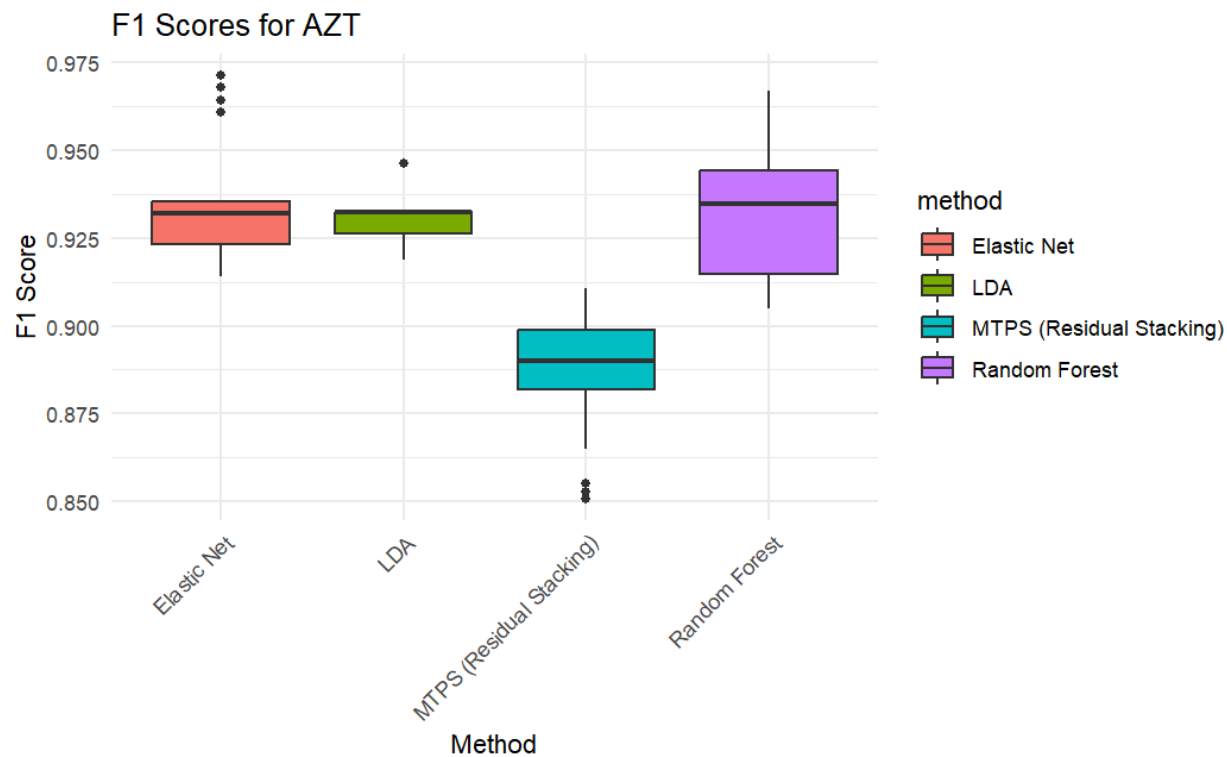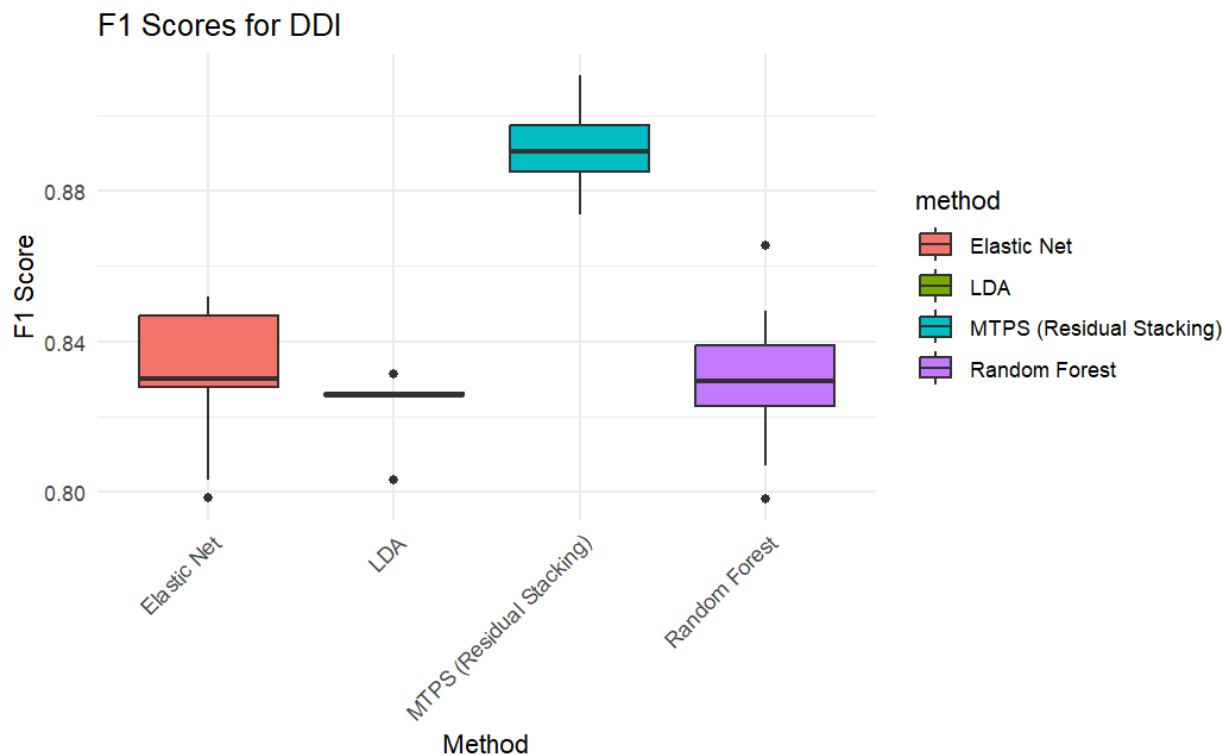
F1 Scores for DDI

Performance Comparison using F1 Scores

The boxplots below illustrate the distribution of F1-scores for LDA, Elastic Net, Random Forest, and MTPS across the five drugs.

From the boxplots, it is evident that Random Forest and Elastic Net consistently achieve the highest median F1-scores across most drugs. LDA shows moderate performance, generally lagging behind Elastic Net and Random Forest. MTPS, however, underperforms relative to the other models, with lower median F1-scores and higher variability, except in the last Drug DDI which it outperforms all methods

Statistical Significance of Model Comparisons

To assess whether the observed differences in performance between the models were statistically significant, I conducted pairwise Wilcoxon signed-rank tests. The p-values for each drug comparison are shown below:

```
### Drug: ABC
Random Forest vs LDA: p-value = 7.725658e-10
MTPS vs LDA: p-value = 0.00235964
Random Forest vs Elastic Net: p-value = 0.8544511
MTPS vs Elastic Net: p-value = 7.790492e-10
```

```
### Drug: 3TC
Random Forest vs LDA: p-value = 1.594613e-09
MTPS vs LDA: p-value = 0.1674577
Random Forest vs Elastic Net: p-value = 7.335608e-09
MTPS vs Elastic Net: p-value = 1.377778e-08

### Drug: AZT
Random Forest vs LDA: p-value = 0.953807
MTPS vs LDA: p-value = 7.790492e-10
Random Forest vs Elastic Net: p-value = 0.01240521
MTPS vs Elastic Net: p-value = 7.790492e-10

### Drug: D4T
Random Forest vs LDA: p-value = 0.08928622
MTPS vs LDA: p-value = 6.541656e-06
Random Forest vs Elastic Net: p-value = 0.005859894
MTPS vs Elastic Net: p-value = 1.260099e-06

### Drug: DDI
Random Forest vs LDA: p-value = 7.870617e-05
MTPS vs LDA: p-value = 7.790492e-10
Random Forest vs Elastic Net: p-value = 0.8242921
MTPS vs Elastic Net: p-value = 7.790492e-10
```

Results

Random Forest significantly outperforms LDA for most drugs, except for AZT and D4T, where the difference is not statistically significant.

MTPS significantly underperforms compared to LDA and Elastic Net for all drugs except 3TC, where the difference is not significant.

Elastic Net performs similarly to Random Forest in ABC and DDI, but Random Forest significantly outperforms it in 3TC, AZT, and D4T.

MTPS shows statistically significant lower performance compared to Elastic Net across all drugs, suggesting it may not be a suitable model for this classification task.

Drug ABC

method <chr>	misclassification_rate <dbl>	precision <dbl>	recall <dbl>	f1_score <dbl>
Elastic Net	0.05622490	0.9278351	0.9325843	0.9278351
LDA	0.07630522	0.8900000	0.9428571	0.9035533
MTPS (Residual Stacking)	0.10640000	0.9305244	0.8500738	0.8903616
Random Forest	0.05200000	0.9736543	0.8979592	0.9332702

Drug 3TC

method <chr>	misclassification_rate <dbl>	precision <dbl>	recall <dbl>	f1_score <dbl>
Elastic Net	0.0440	0.9339623	0.9417476	0.9463415
LDA	0.0840	0.8706897	0.9417476	0.9058296
MTPS (Residual Stacking)	0.1044	0.9264248	0.8593396	0.8916747
Random Forest	0.0520	0.9685855	0.8856728	0.9336717

Drug AZT

method <chr>	misclassification_rate <dbl>	precision <dbl>	recall <dbl>	f1_score <dbl>
Elastic Net	0.08064516	0.9461538	0.9179104	0.9318955
LDA	0.08064516	0.9517241	0.9019608	0.9324324
MTPS (Residual Stacking)	0.10560000	0.9283926	0.8550151	0.8899217
Random Forest	0.07444293	0.9633695	0.9104478	0.9345610

Drug D4T

method <chr>	misclassification_rate <dbl>	precision <dbl>	recall <dbl>	f1_score <dbl>
Elastic Net	0.1088710	0.8794326	0.9179104	0.9066667
LDA	0.1080000	0.8832117	0.9029851	0.9137380
MTPS (Residual Stacking)	0.1060000	0.9270453	0.8552563	0.8889875
Random Forest	0.1082169	0.9114379	0.8985507	0.9090909

Drug DDI

method <chr>	misclassification_rate <dbl>	precision <dbl>	recall <dbl>	f1_score <dbl>
Elastic Net	0.1686747	0.8270677	0.8333333	0.8301887
LDA	0.1800000	0.8153846	0.8225806	0.8257576
MTPS (Residual Stacking)	0.1056000	0.9328453	0.8584597	0.8903911
Random Forest	0.1640000	0.8816124	0.7878788	0.8294677

Best Model Based on F1-Scores

Random Forest is the best-performing model for most drugs based on median F1-scores, demonstrating consistently high predictive performance.

Best Model Based on Precision

Random Forest achieves the highest precision for most drugs, except for ABC and 3TC, where Elastic Net performs slightly better.

Best Model Based on Misclassification Rate

Random Forest and Elastic Net achieve the lowest misclassification rates across most drugs, with Random Forest performing slightly better in some cases.

Best Model Based on Recall

Elastic Net has the highest recall across most drugs, indicating strong sensitivity in identifying resistant cases.

Conclusion

Based on this comprehensive evaluation, Random Forest emerges as the best-performing model across multiple metrics, including F1-score, precision, and misclassification rate. Elastic Net also performs well, particularly in recall. LDA consistently lags behind these models, and MTPS (Residual Stacking) underperforms relative to all other models, making it the last choice for predicting HIV drug resistance outcomes. These findings suggest that Random Forest is the most reliable model for this classification task, with Elastic Net as a competitive alternative.