Assignment #2

Belina Jang
Student Number: V00924202
Dr. Xuekui Zhang
Stat 469 A01
October 16, 2022

First, YY is converted into matrix then using the provided cut-off values for each drug to convert the matrix YY into a matrix of 1 and 0 by comparing the resistance with their corresponding cut-off values where 1 indicates the sample is not resistant to the corresponding column of drug since the resistance is smaller than the cut-off and 0 indicates the sample is resistant to the drug since the resistance is greater than the cut-off. Then the format of the matrix is changed into data frame.

Then the n-fold is set to be 5. Then in order to do the stratified random sampling with 5 folds, K means clustering method was used. It randomly generates a list of 1 to 5 for each sample. Then the indices for 50 experiments are saved into a matrix to be used for each drug. So, all i-th experiment of different drug will use same sample data.

Then the format of the data is changed into dataframe so it's easier to work with. A matrix is created to store the confusion matrix values from different models for all drugs and experiments. Then a for loop is used for each experiment then another for loop inside for each drug. So for each i-th experiment (first loop), it will loop through all the drugs then move onto next (i+1)-th experiment. Then for each drug, it creates a matrix to store confusion matrix values from each experiment of each drug together. Then the predictor matrix XXdf is combined with a column of the drug from yBin. Then another loop for 5 folds is added. For each fold, it declares test and train data using the indices matrix created earlier to do stratified random sampling. I modelled and made predictions for LDA, classification tree and elastic net with the same train and test data. Then I evaluated the predictions and calculated values (tn, fp, fn, tp) for the confusion matrix for each model. Repeat it 5 times for 5 folds then add them together for each model, drug and experiment. Then those values are saved into the matrix (metrics) created earlier to save values of confusion matrix with corresponding model, drug name and experiment number. Then we repeat this process 50 times.

The resulting matrix will have 750 rows (50 experiments*5drugs*3models). Then it creates other columns namely mcr, acc, precision, recall, and F1 to save misclassification rate, accuracy, precision, recall, and F1 score respectively. To learn the uncertainties in my models, the experiment is repeated 50 times with different train and test data. Then I plotted different graphs to compare models in different drug.

I used wilcox.test to see if the differences between those models are significant to justify a certain model is the best model for each drug and which drug is the best model in average for all drugs.

From figure 1-4 below, we can observe which model was the best model in terms of different criteria for each drug.

Figure 1 shows misclassification rate of different models for 5 different drugs. The lower the misclassification rate, the higher the accuracy of the model. Elastic net has the best performance in terms of misclassification rate since the median misclassification rate of elastic net was the lowest in 4 out of 5 models and second lowest by little for DDI. We will check if elastic is in fact the best model in terms of the misclassification rate in average for all drugs later.

Figure 2 shows precision of different models for 5 different drugs. The higher precision, the better the model performance is. Classification tree has the best performance in terms of precision since the median precision of classification tree was the highest in 4 out of 5 models and second highest in 3TC. And Elastic net was the highest in 3TC and second highest in other drugs.

Figure 3 shows recall of different models for 5 different drugs. The higher the recall, the better the model is. It seems like Elastic net has the best performance in terms of recall since the median recall of elastic net was the highest in 2 out of 5 models and elastic net and LDA was a tie at the other 3 drugs. We will check if elastic is in fact the best model in average for all drugs later.

Figure 4 shows F1 score of different models for 5 different drugs. The higher the F1 score is, the better the model is. It seems like Elastic net has the best performance in terms of F1 score since the median F1 score of elastic net was the highest in 4 out of 5 models and elastic net and LDA was a tie in AZT. We will check if elastic is in fact the best model in terms of the F1 score in average for all drugs later and also if the difference between elastic net and LDA in AZT is significant later in the wilcox.test.

Figure 1: Misclassification Rate of all Models grouped by Drug name

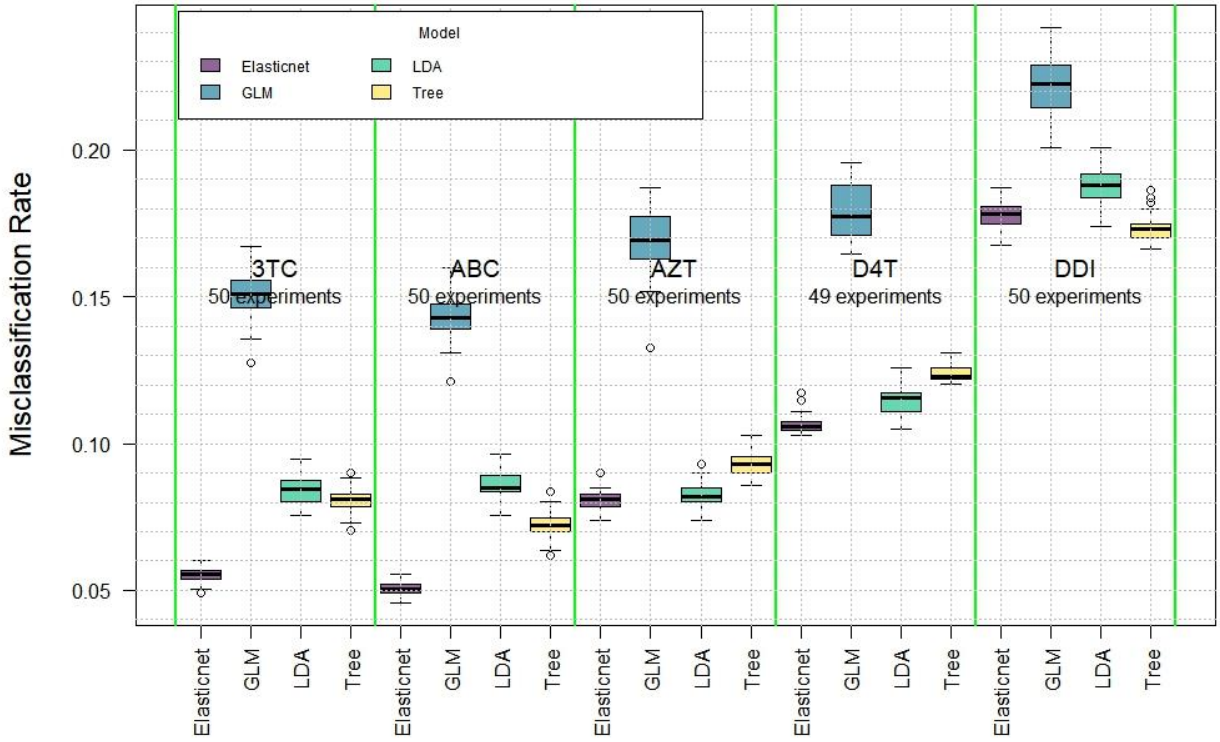


Figure 2: Precision of all Models grouped by Drug name

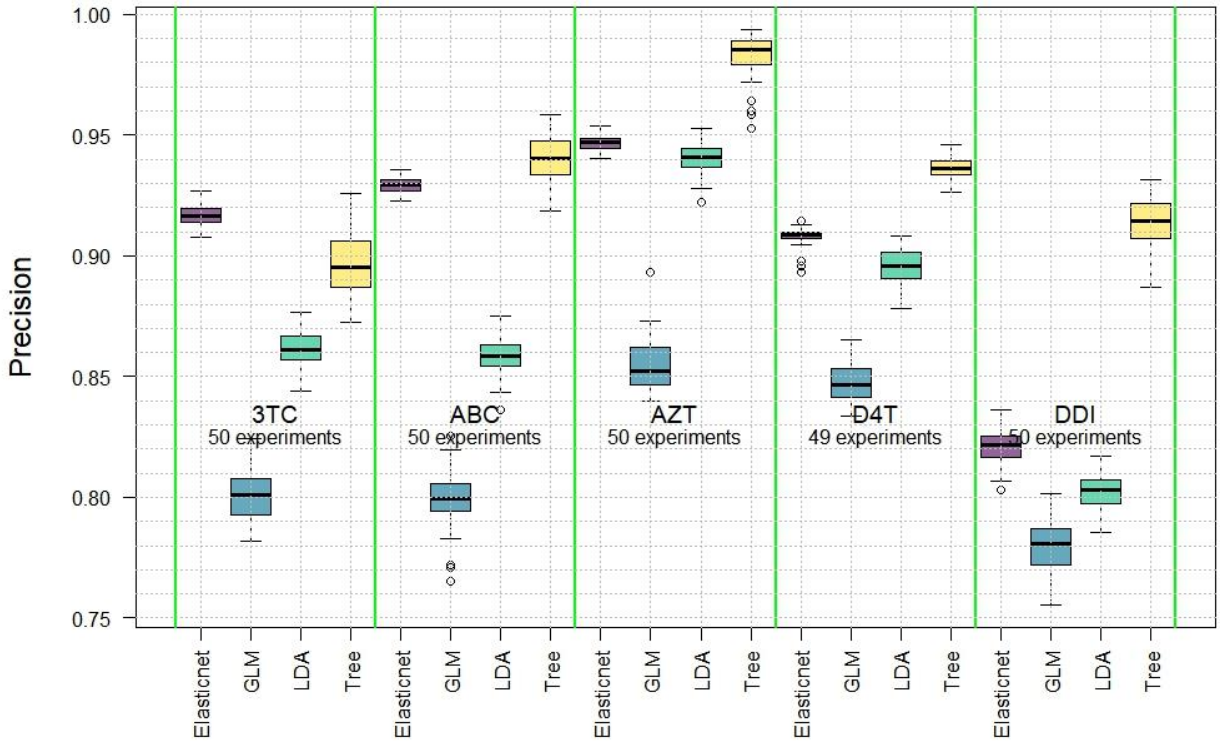


Figure 3: Recall of all Models grouped by Drug name

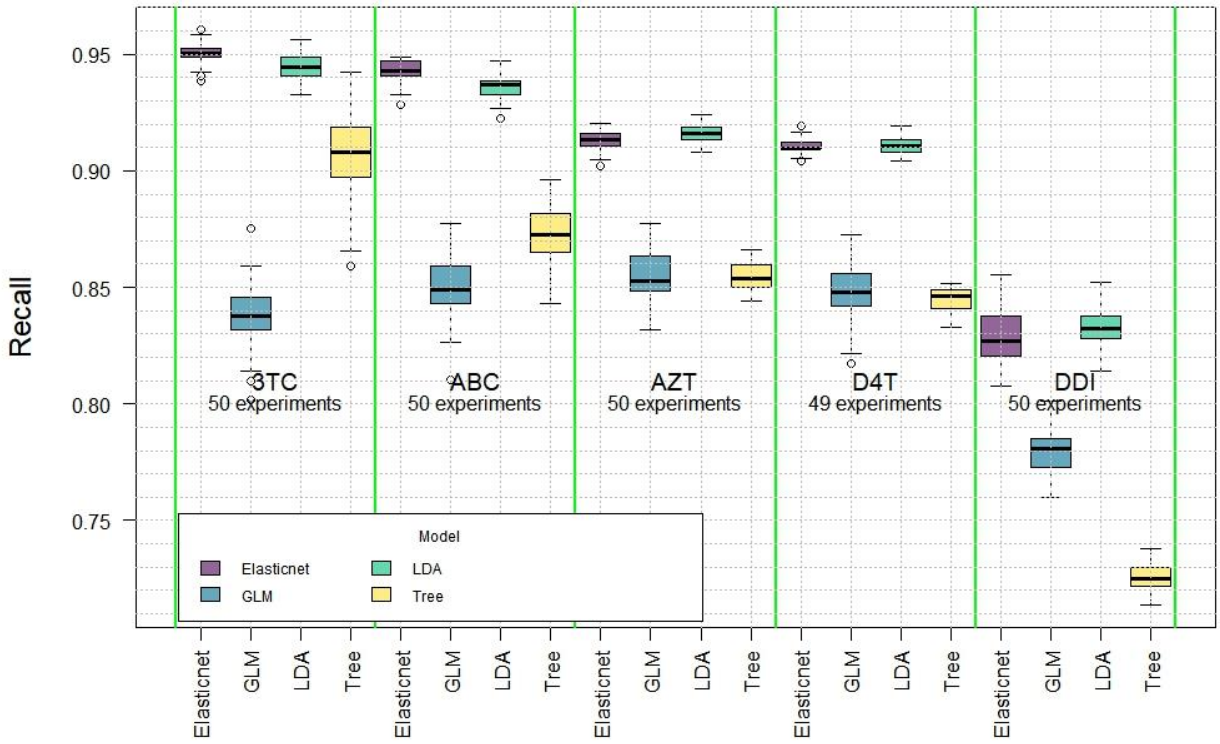
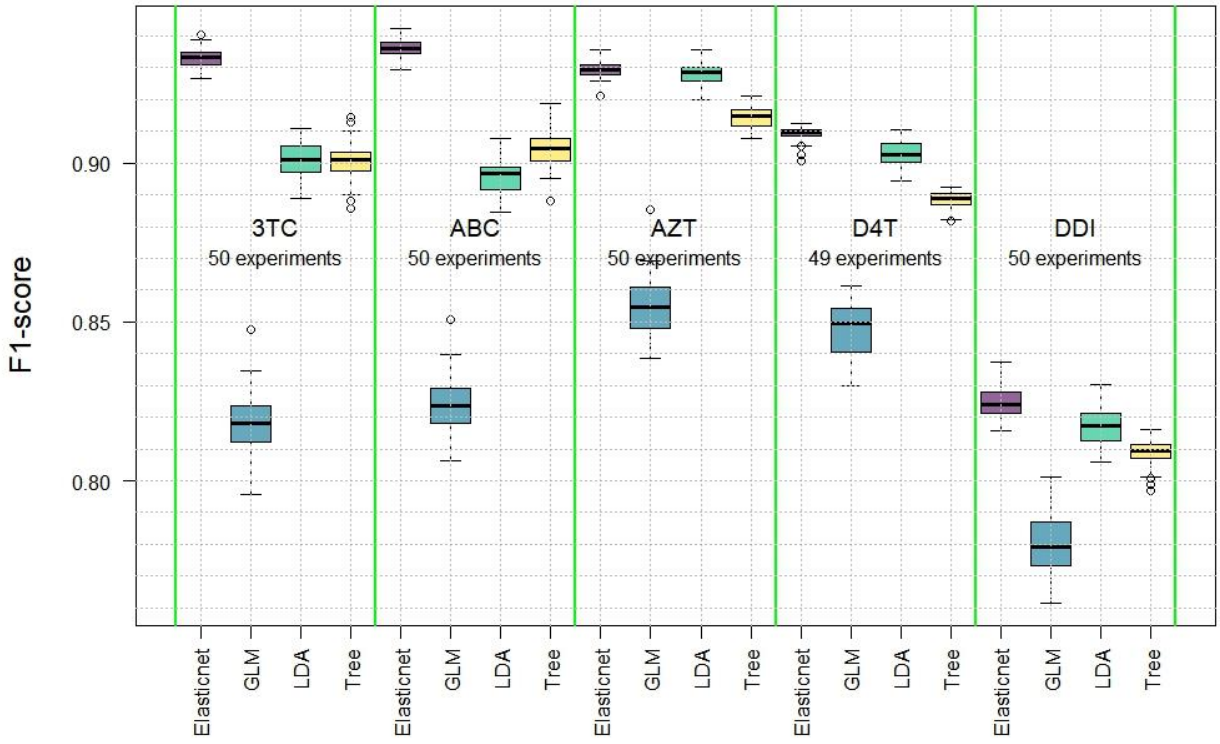


Figure 4: F1-score of all Models grouped by Drug name



From figure 5-8 below, we can observe which model was the best model in terms of different criteria regardless of the drug type.

From figure 5, we can observe that the elastic net was the best model in terms of misclassification rate. So, we can confirm the observation I made earlier in figure 1.

From figure 6, we can observe that classification tree was the best model in terms of precision and elastic net is the second best which also confirms that observation from figure 2.

From figure 7, we can observe that the median of the elastic net is similar to LDA and that overall box plot of elastic net locates higher than LDA. So, we can say either elastic net or LDA was the best model in terms of recall.

From figure 8, we can observe that the elastic net was the best model in terms of F1-score. So, we confirm the observation I made earlier in figure 4.

Figure 5: MCR of all Models all Drugs (249 exp)

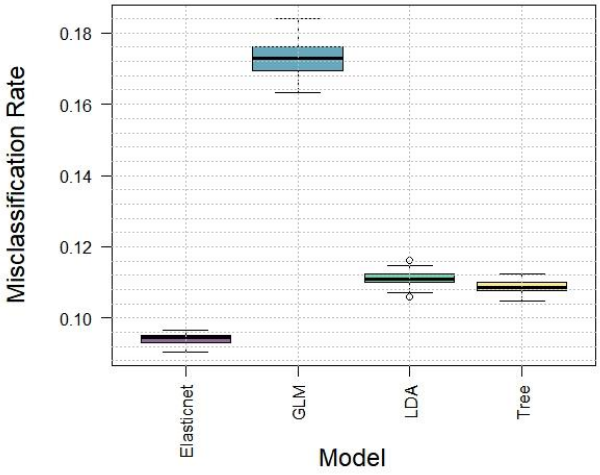


Figure 6: Precision of all Models all Drugs (249 exp)

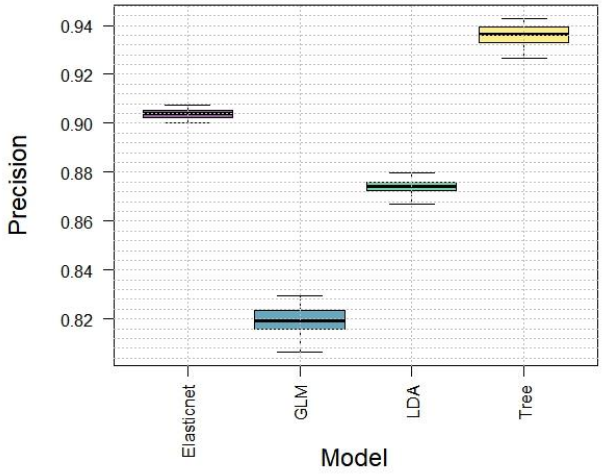


Figure 7: Recall of all Models all Drugs (249 exp)

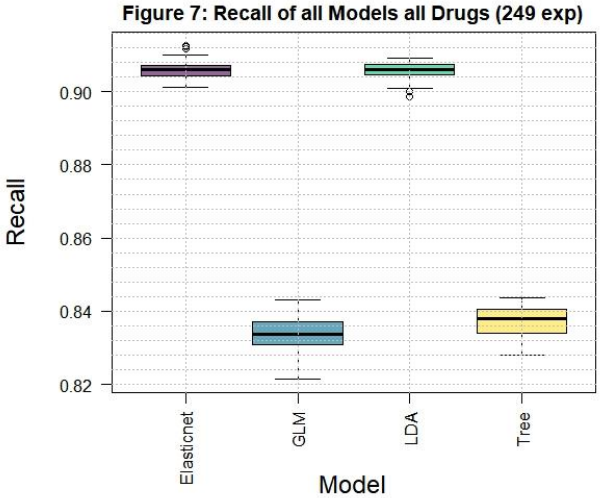


Figure 8: F1-Score of all Models all Drugs (249 exp)

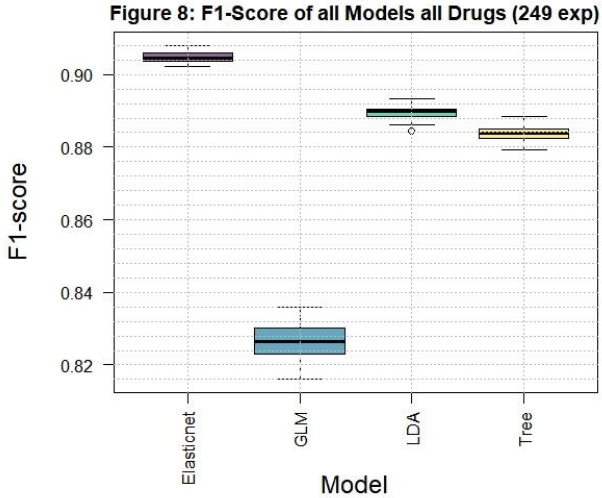


Table 3-7 shows p-value of wilcox pairs of F1-scores of different models. In most cases, p-values are very small except some cases.

In table 3, we can see the p-value for tree and LDA is big (> 0.1) for 3TC, which shows that the difference between tree and LDA was insignificant. And in figure 4, it does seem like LDA and tree have similar F1 score for 3TC. This does not affect the observation made about the best model. And the small p-values for between elastic net with other two models validate the observation that the elastic net is the best model for 3TC.

In table 4,6 and 7, we can see all the p-values are very small (< 0.001), so the differences between models for these drugs are very highly significant. So, it supports the observation made earlier about the drugs ABC, D4T and DDI.

In table 5, the p-value for LDA and elastic net for AZT is between 0.01 and 0.05, which indicates that there's moderate significance for the difference between LDA and elastic net for AZT. To check either Elastic net or LDA is better for AZT in terms of F1 score, I plotted the difference of F1 score with the p-value, see figure 9. Then we can conclude Elastic net is better for AZT. Therefore, Elastic net is better than other two models in all drugs.

Therefore, the p-values do support and explain the observations made earlier.

Table: 3: Wilcoxon Pairs of F1-Scores for Drug: 3TC (50 trials)

	Elasticnet	LDA	Tree
Elasticnet	NA	0.0000000	0.0000000
LDA	0	NA	0.9691988
Tree	0	0.9691988	NA

Table: 4: Wilcoxon Pairs of F1-Scores for Drug: ABC (50 trials)

	Elasticnet	LDA	Tree
Elasticnet	NA	0	0
LDA	0	NA	0
Tree	0	0	NA

Table: 7: Wilcoxon Pairs of F1-Scores for Drug: DDI (50 trials)

	Elasticnet	LDA	Tree
Elasticnet	NA	4e-07	0
LDA	4e-07	NA	0
Tree	0e+00	0e+00	NA

Table: 5: Wilcoxon Pairs of F1-Scores for Drug: AZT (50 trials)

	Elasticnet	LDA	Tree
Elasticnet	NA	0.0321109	0
LDA	0.0321109	NA	0
Tree	0.0000000	0.0000000	NA

Table: 6: Wilcoxon Pairs of F1-Scores for Drug: D4T (49 trials)

	Elasticnet	LDA	Tree
Elasticnet	NA	0	0
LDA	0	NA	0
Tree	0	0	NA

Figure 9: for AZT P value = 0.0321109

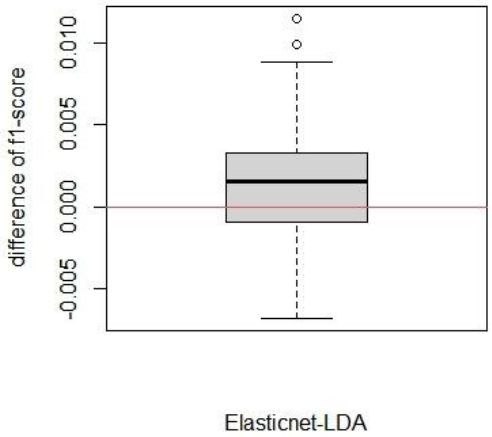


Table 8-10 shows the P-value of wilcoxon pairs of the F1-scores for each drugs for different models. Most p-values are very small (<0.001) except few cases. For elastic net, p-values are all very small so observations made about elastic net is valid. For LDA, pair D4T and 3TC had p-value of 0.0342454 which is between 0.01 and 0.05 so it's still moderately significant. For the classification tree, ABC and 3TC had a p-value of 0.0011107 which is between 0.001 and 0.01 so it's still strongly significant.

Table: 8: Wilcoxon Pairs of F1-Scores for Model: Elasticnet

	3TC	ABC	AZT	D4T	DDI
3TC	NA	8.2e-06	0	0	0
ABC	8.2e-06	NA	0	0	0
AZT	0.0e+00	0.0e+00	NA	0	0
D4T	0.0e+00	0.0e+00	0	NA	0
DDI	0.0e+00	0.0e+00	0	0	NA

Table: 9: Wilcoxon Pairs of F1-Scores for Model: LDA

	3TC	ABC	AZT	D4T	DDI
3TC	NA	5.2e-06	0	0.0342454	0
ABC	0.0000052	NA	0	0.0000000	0
AZT	0.0000000	0.0e+00	NA	0.0000000	0
D4T	0.0342454	0.0e+00	0	NA	0
DDI	0.0000000	0.0e+00	0	0.0000000	NA

Table: 10: Wilcoxon Pairs of F1-Scores for Model: Tree

	3TC	ABC	AZT	D4T	DDI
3TC	NA	0.0011107	0	0	0
ABC	0.0011107	NA	0	0	0
AZT	0.0000000	0.0000000	NA	0	0
D4T	0.0000000	0.0000000	0	NA	0
DDI	0.0000000	0.0000000	0	0	NA

In conclusion, for all the drugs, the best model was Elastic net since it has the largest F1-score and considering other performances. Therefore, the best model was consistent on all drugs.