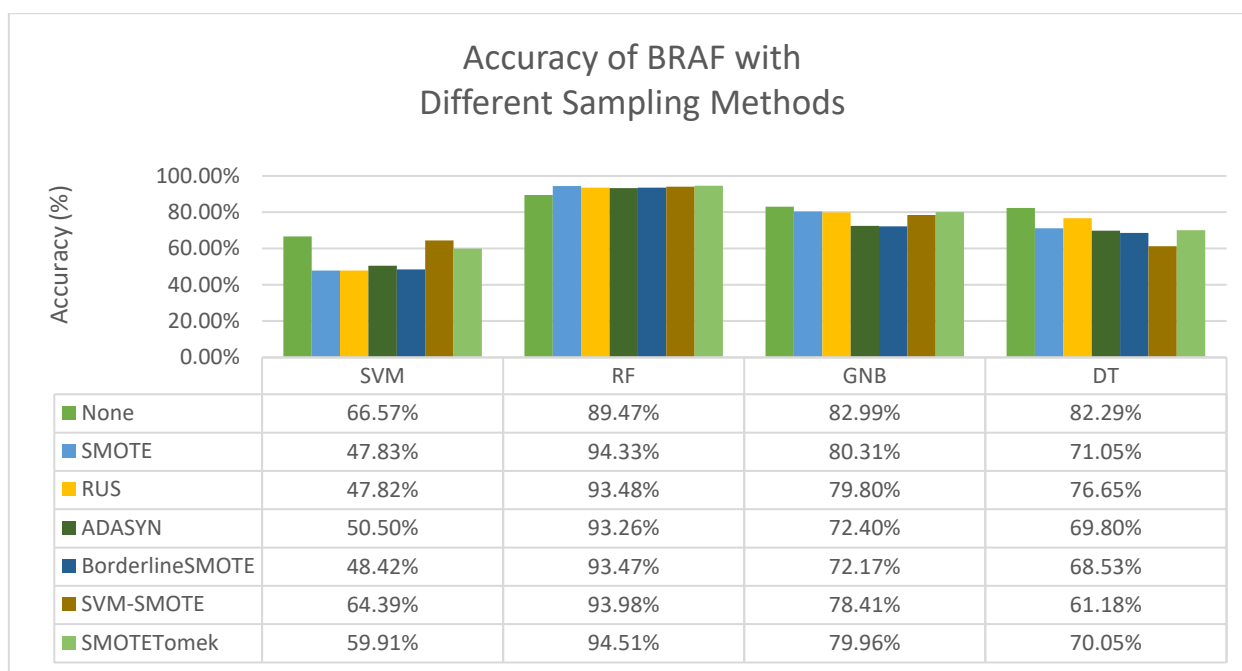


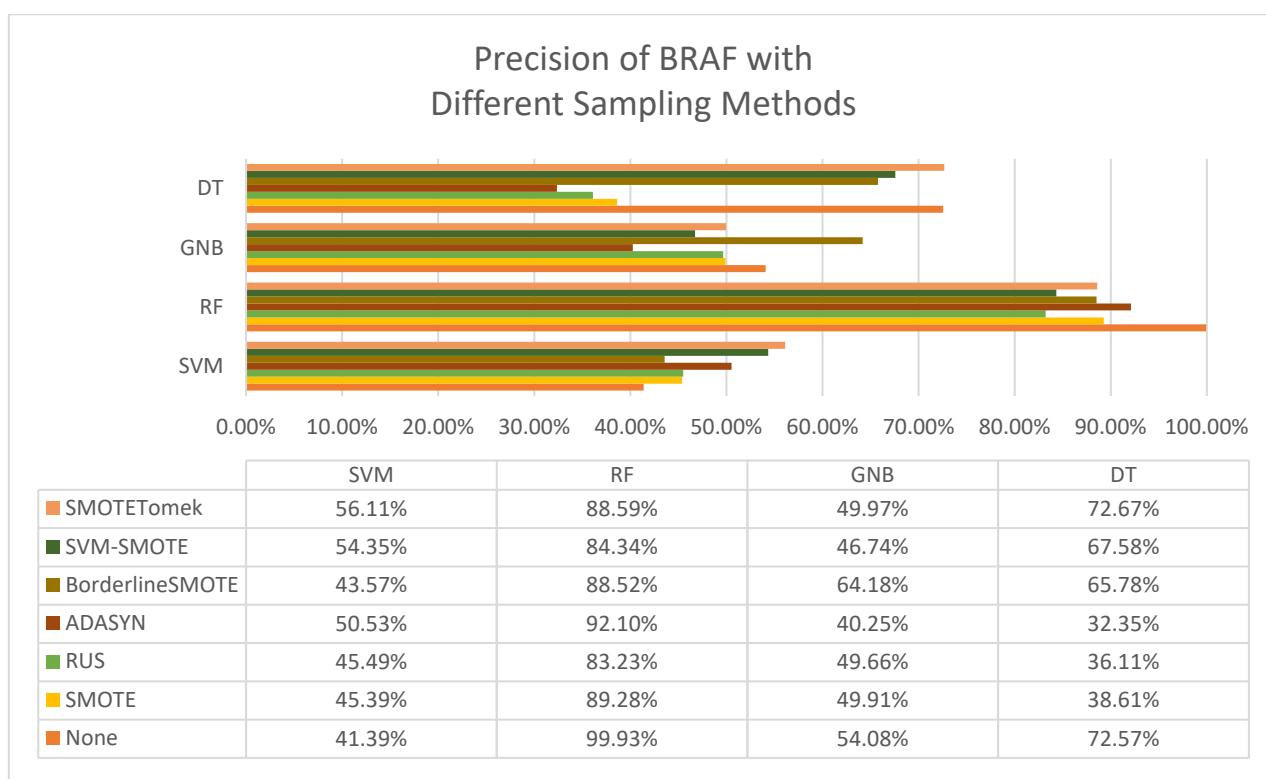
Comparative Studies on Resampling Techniques in Machine Learning and Deep Learning Models for Drug-Target Interaction Prediction

Azwaar Khan Azlim Khan and Nurul Hashimah Ahamed Hassain Malim

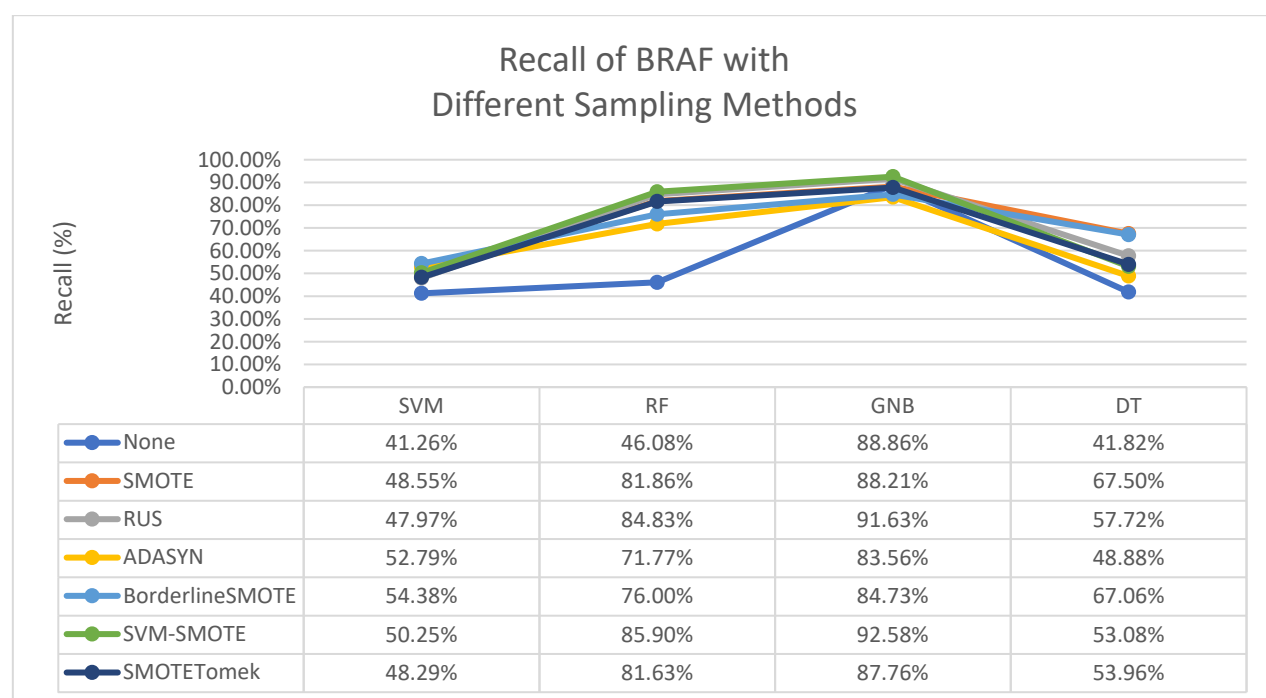
BRAF achieved the highest accuracy of 94.51% when SMOTETomek was applied on the RF classifier, where an increase of 5.04% is noticed from the original accuracy value of 89.47% when no resampling method was applied (Refer to Figure S1). Furthermore, when no resampling method was applied, the precision value is the highest (99.93%), and a drop of 7.83% is noticed when ADASYN was used, while the other resampling methods showed a drop in precision values even larger than 7.83% (Refer to Figure S1). In terms of recall, BRAF performs the best when SVM-SMOTE was used with the GNB classifier, yielding a total recall value of 92.58%, which is a 3.72% increase from when no resampling method is used in GNB, with a recall value of 88.86% (Refer to Figure S1). In general, RF is the best classifier, outperforming all the other classifiers when paired with resampling methods and SVM and DT turns out to be the weakest classifiers in terms of accuracy, precision and recall.



(a)



(b)



(c)

Figure S1. The accuracy, precision and recall values for BRAF. (a) Accuracy of BRAF when different sampling methods are applied; (b) Precision of BRAF when different sampling methods are applied; (c) Recall of BRAF when different sampling methods are applied.

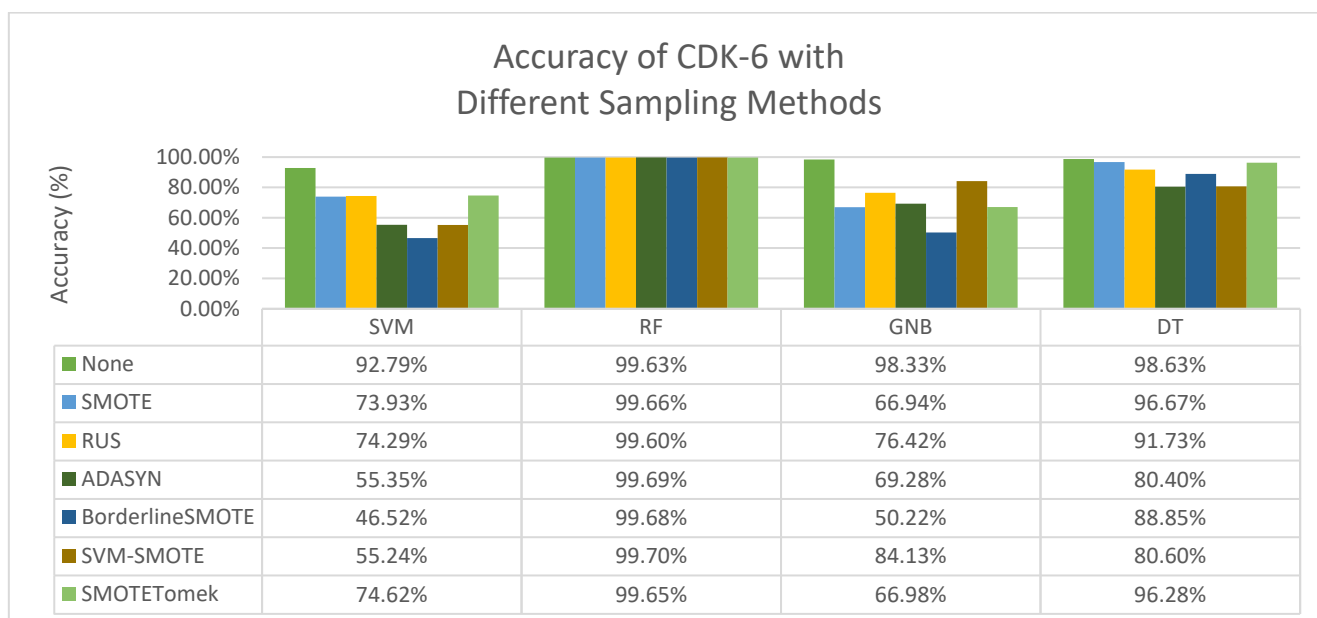
For CDK-6, the accuracy is the highest when SVM-SMOTE is used with the RF classifier (99.70%), which a slight increase from when RF is applied with no resampling method (Refer to Figure S2). In terms of precision, RF is the best classifier when no

19
20
21

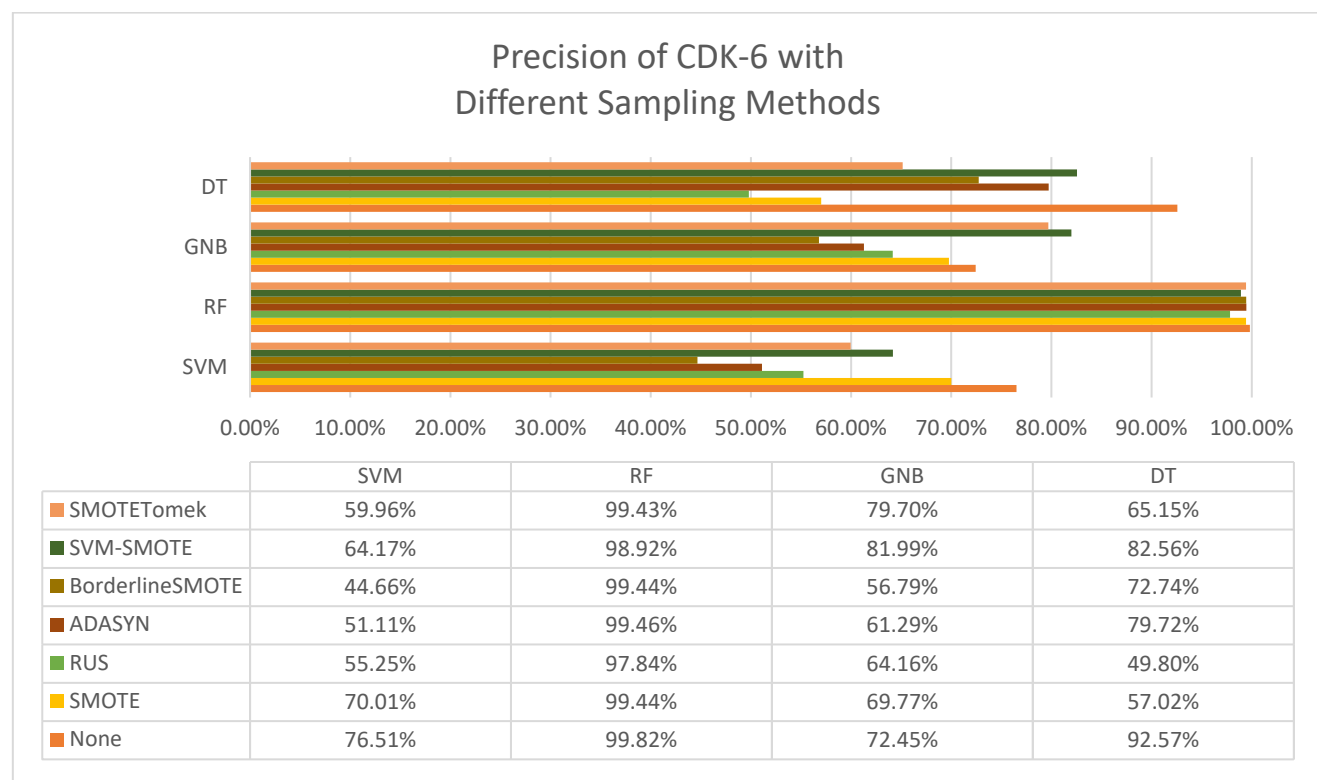
22
23
24

resampling method is applied, yielding a precision value of 99.82% and in terms of recall, it was found that GNB with SVM-SMOTE achieved the highest recall with a value of 96.56%, which is a 7.41% increase from the original value when GNB is applied with no resampling method (Refer to Figure S2). Similar to BRAF, SVM and DT are the weakest classifiers when paired with resampling methods in terms of accuracy, precision and recall while RF is the best classifier.

25
26
27
28
29
30
31



(a)



(b)

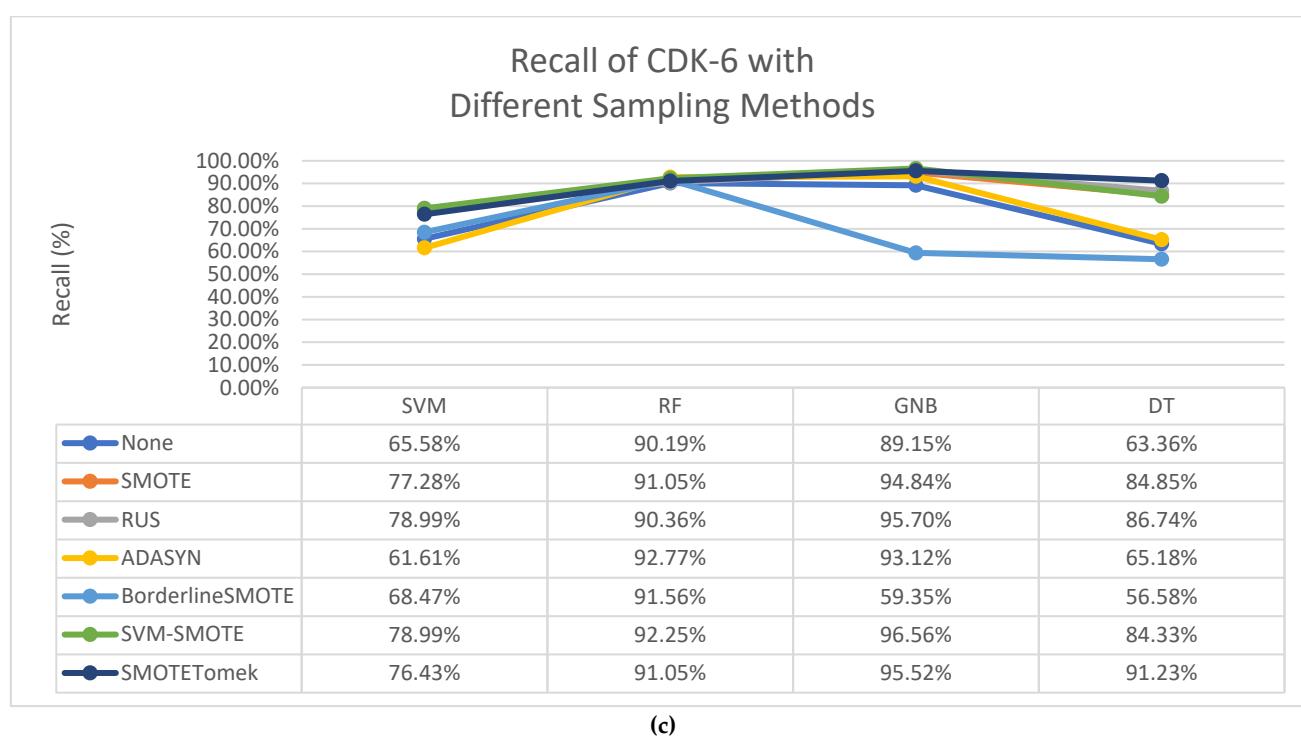
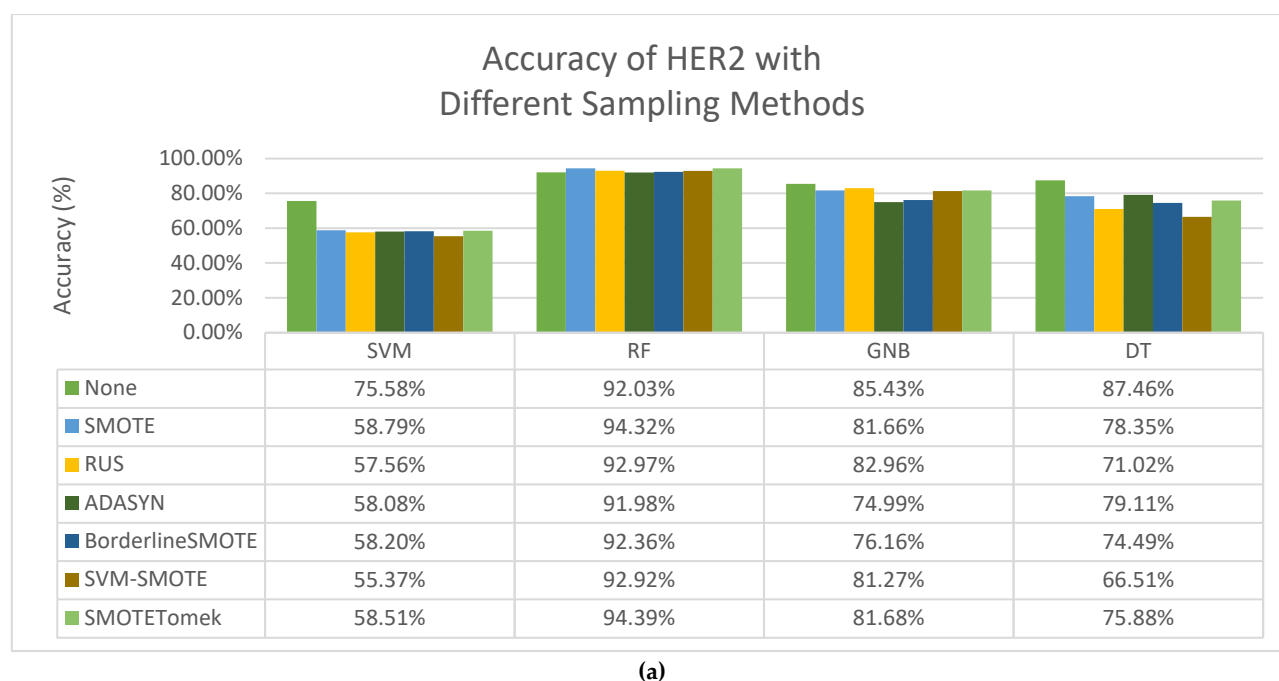
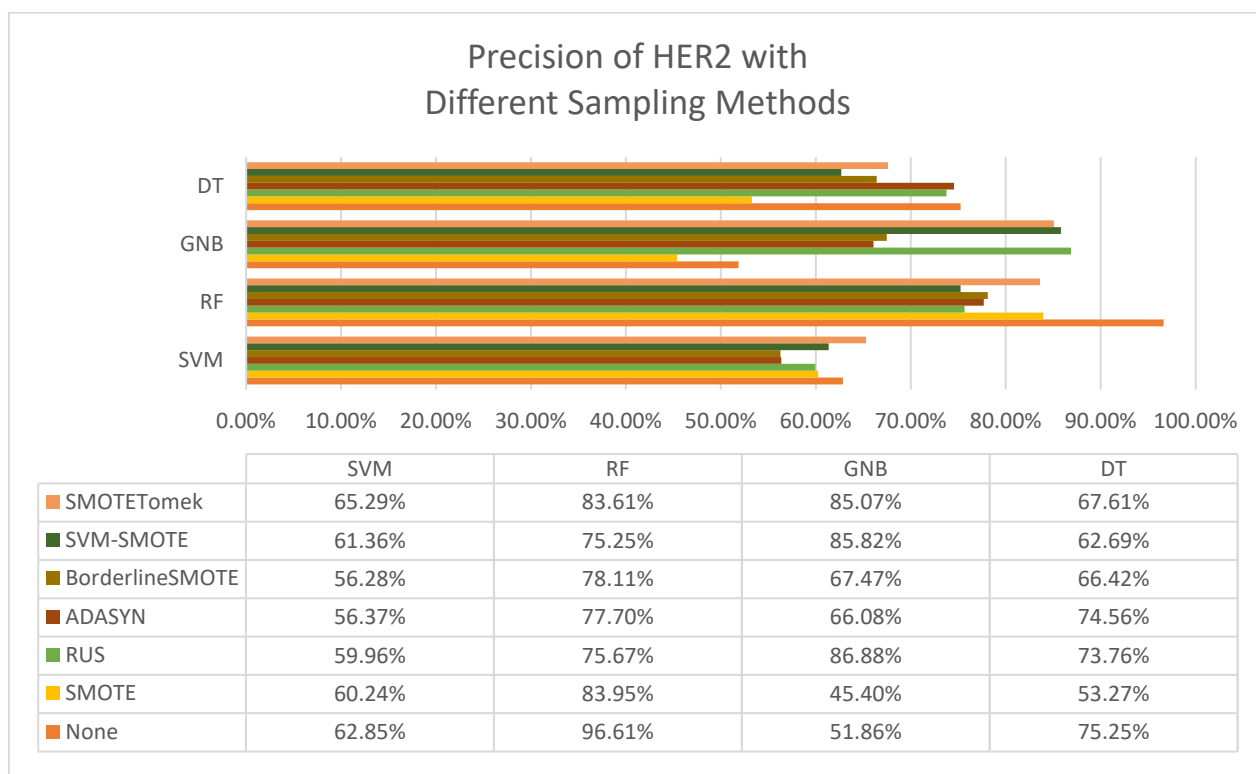


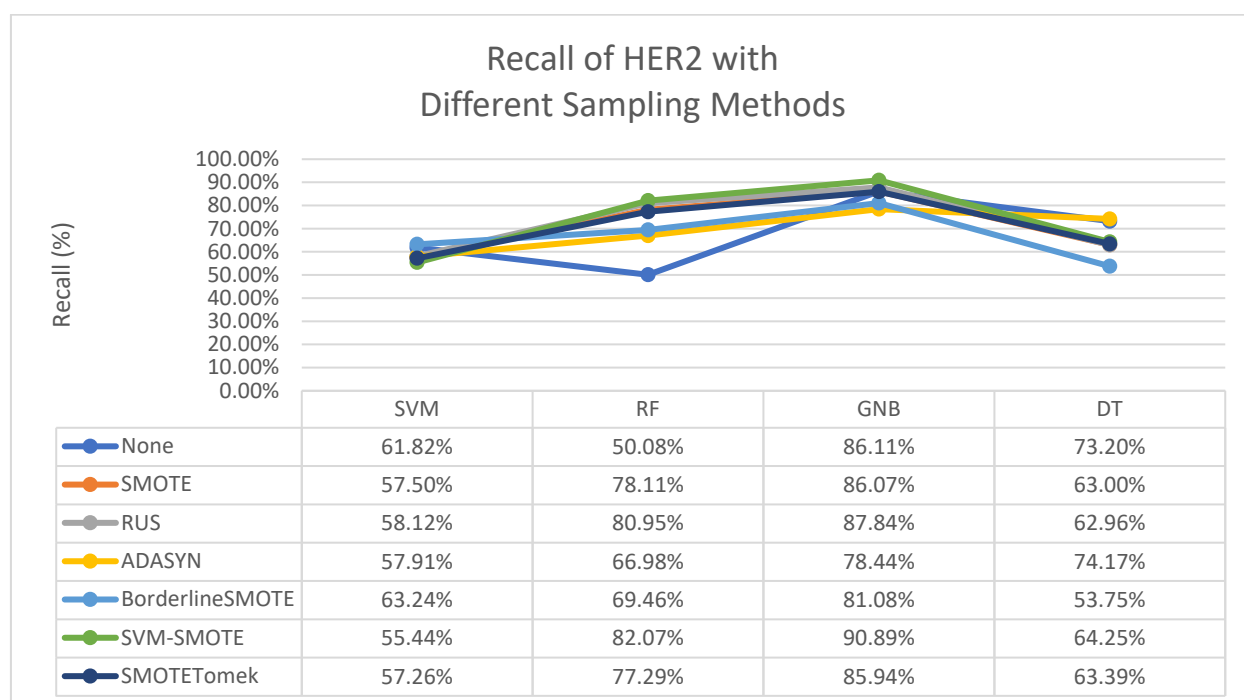
Figure S2. The accuracy, precision and recall values for CDK-6. (a) Accuracy of CDK-6 when different sampling methods are applied; (b) Precision of CDK-6 when different sampling methods are applied; (c) Recall of CDK-6 when different sampling methods are applied.

Next, in HER2, RF is also still the best classifier when paired with various resampling methods. The accuracy is the highest when SMOTETomek was applied with the RF classifier (94.39%), which is a 2.36% increase from when no resampling is applied to RF and in terms of precision, it is surprising that RF with no resampling method is the highest (96.61%) while a huge drop in the precision value is observed when resampling methods are applied (Refer to Figure S3). It was also found that when SVM-SMOTE is applied with the GNB classifier, the recall is the highest, which is a value of 90.89% (Refer to Figure S3).





(b)



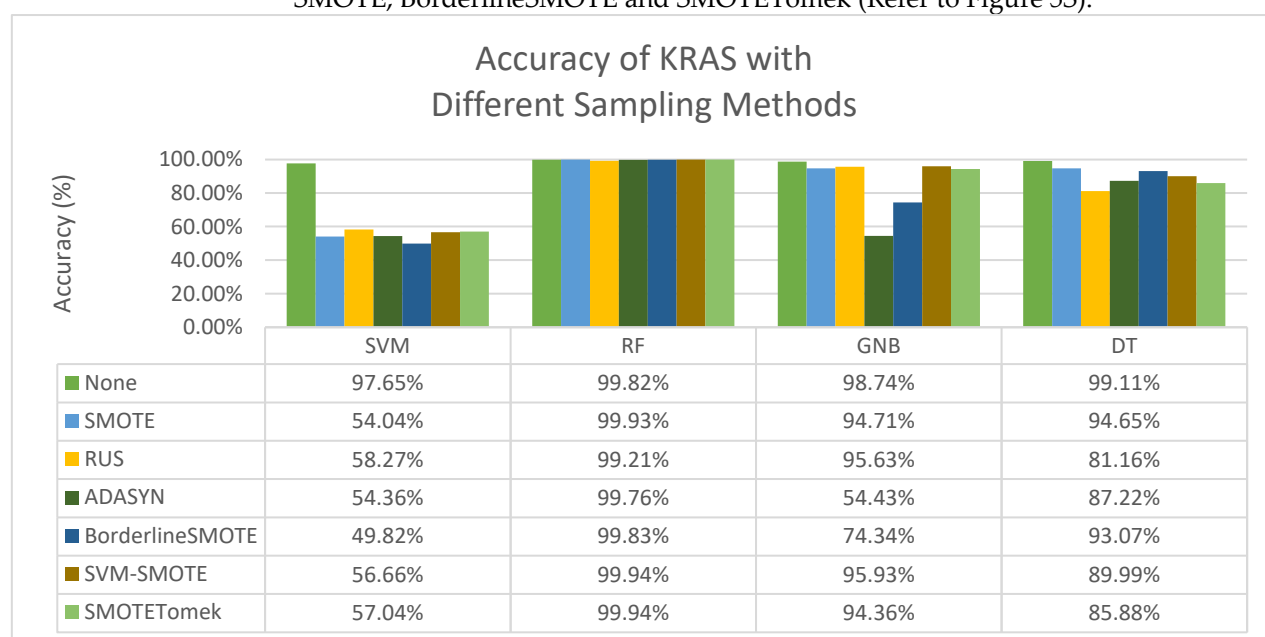
(c)

Figure S3. The accuracy, precision and recall values for HER2. (a) Accuracy of HER2 when different sampling methods are applied; (b) Precision of HER2 when different sampling methods are applied; (c) Recall of HER2 when different sampling methods are applied.

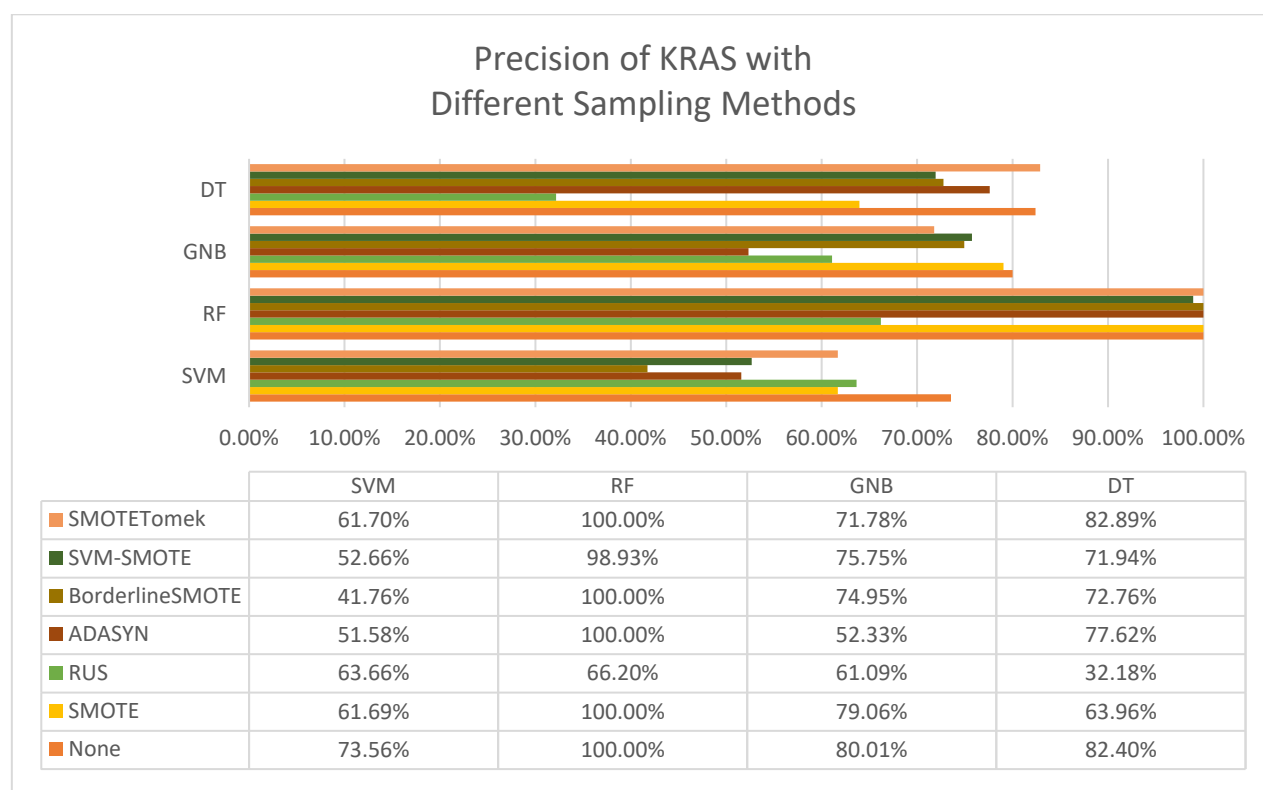
For KRAS, the accuracy is the highest when both SVM-SMOTE and SMOTETomek were applied on the RF classifier, with both achieving an accuracy of 99.94%, which is a

slight increase from when no resampling is applied on RF with a value of 99.82% (Refer to Figure S4). PD-1 also observed a high accuracy value of 99.96% when ADASYN, SVM-SMOTE and SMOTETomek were applied on the RF classifier (Refer to Figure S6). The precision of KRAS and PD-1 achieved an outstanding 100% when SMOTE, ADASYN, BorderlineSMOTE and SMOTETomek is applied (Refer to Figure S4 and Figure S6). The precision of PD-1 also reached 100% when SVM-SMOTE was used with the RF classifier. It is also very interesting to note that the precision for KRAS and PD-1 also reached 100% when no resampling method was applied on the RF classifier (Refer to Figures S4 and S6). Furthermore, the recall for both KRAS and PD-1 was also the highest when GNB was paired with SMOTE for KRAS, achieving a recall of 99.50%, and 99.36% for PD-1 when GNB was paired with SVM-SMOTE (Refer to Figures S4 and S6). Both of these recall values showed an increase from the original value when no sampling was applied, whereby a 7.92% increase was seen for KRAS and a 1.91% increase for PD-1 (Refer to Figures S4 and S6).

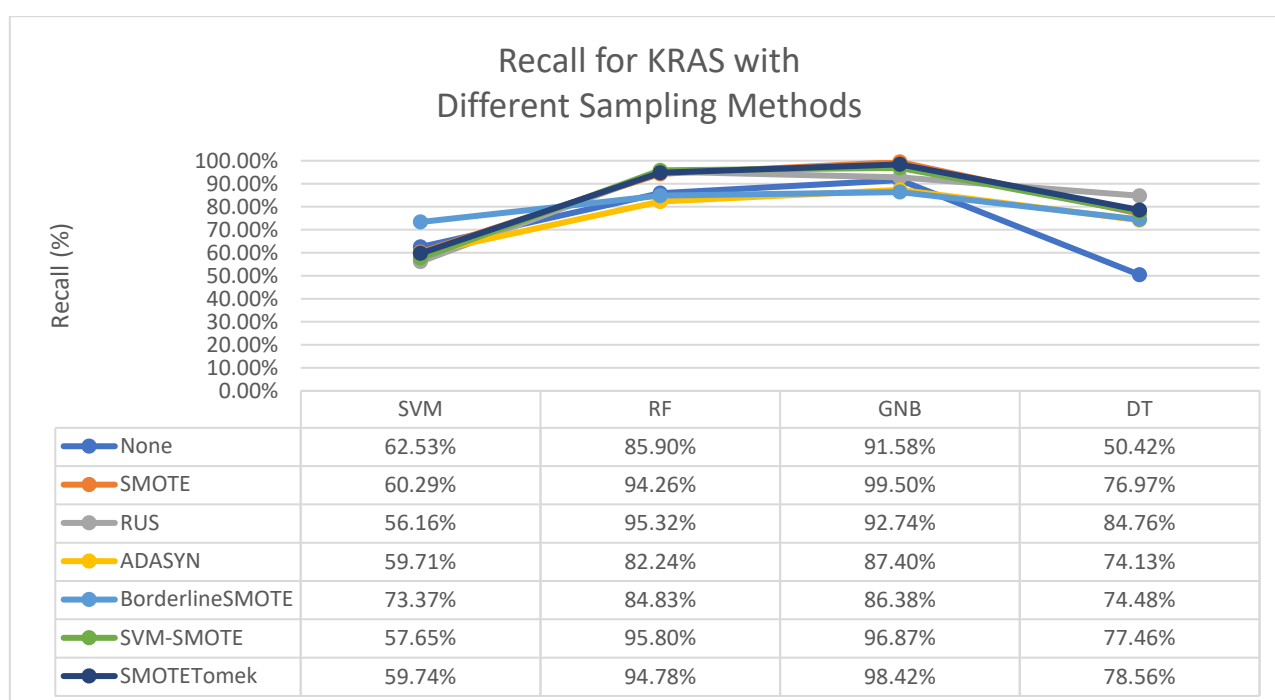
Then, for the MEK1 activity class, the highest accuracy is 99.95% when ADASYN and SMOTETomek was paired with the RF classifier (Refer to Figure S5). In terms of precision, the highest value is achieved when ADASYN was paired with the RF classifier, which is a value of 97.50% (Refer to Figure S5). The same precision value was observed when no resampling method was applied on the RF classifier as well. Finally, the highest recall value is also 97.50% and this value is achieved when the GNB classifier is applied with SMOTE, BorderlineSMOTE and SMOTETomek (Refer to Figure S5).



(a)

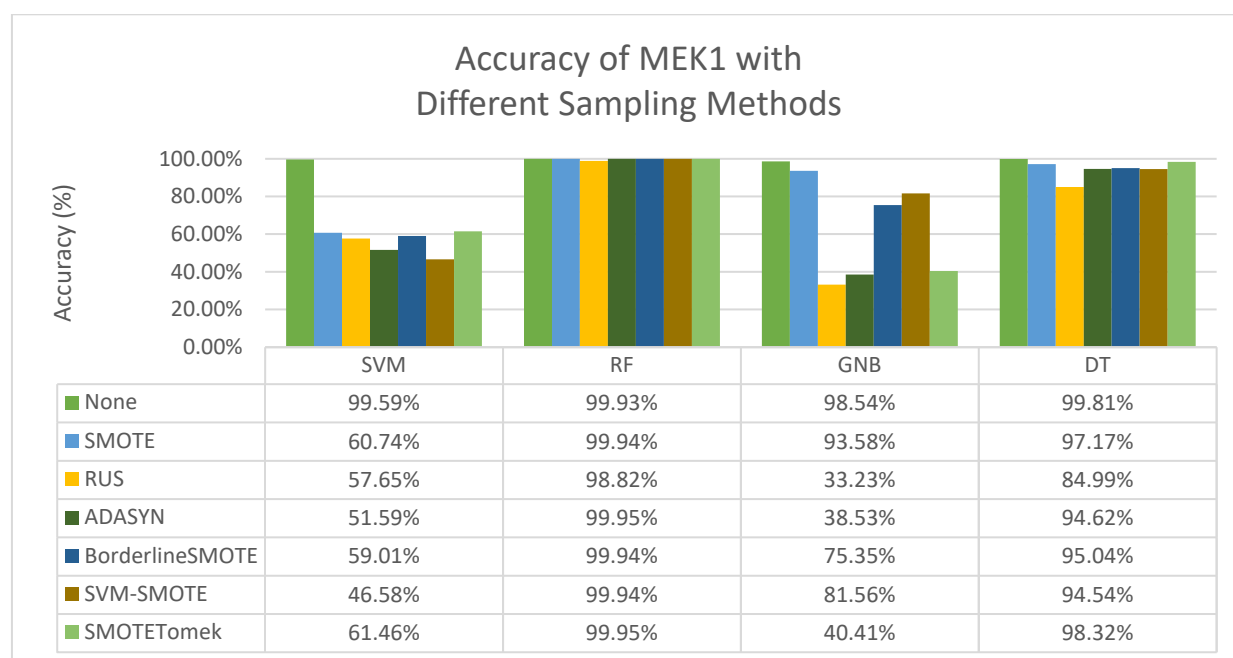


(b)

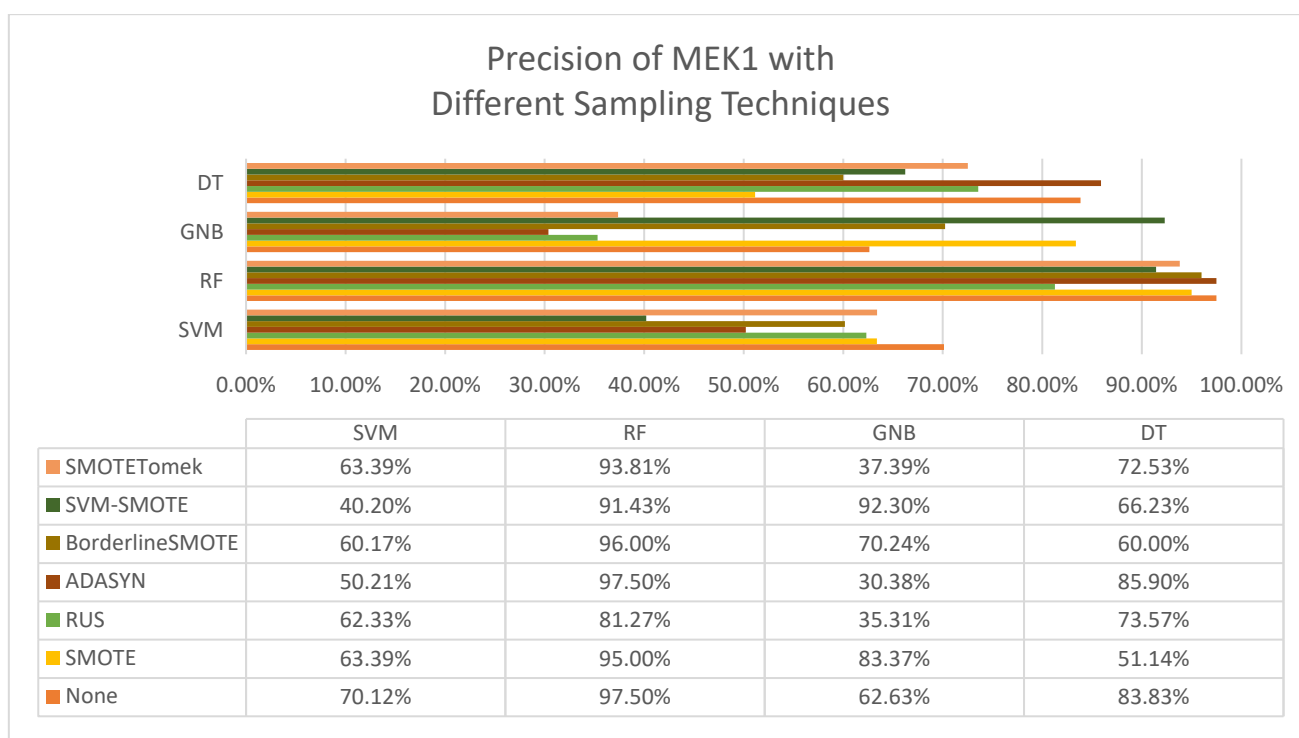


(c)

Figure S4. The accuracy, precision and recall values for KRAS. (a) Accuracy of KRAS when different sampling methods are applied; (b) Precision of KRAS when different sampling methods are applied; (c) Recall of KRAS when different sampling methods are applied.



(a)



(b)

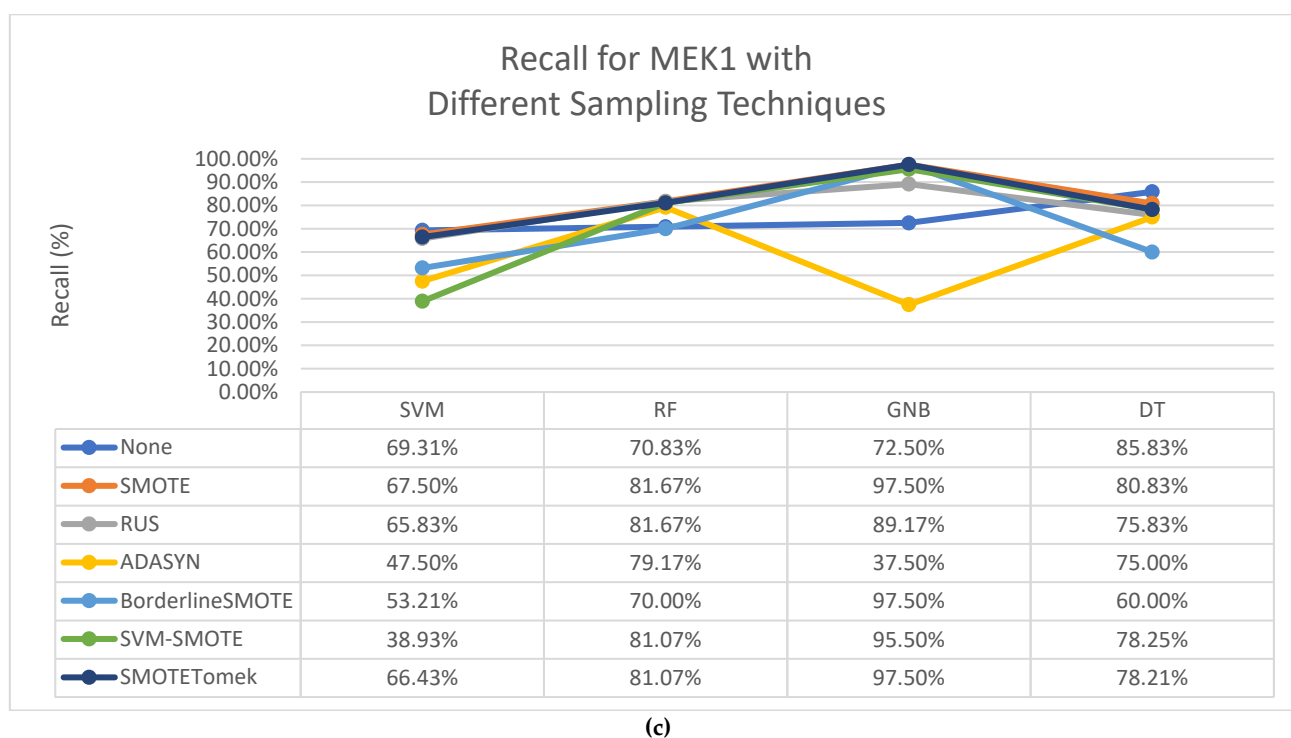
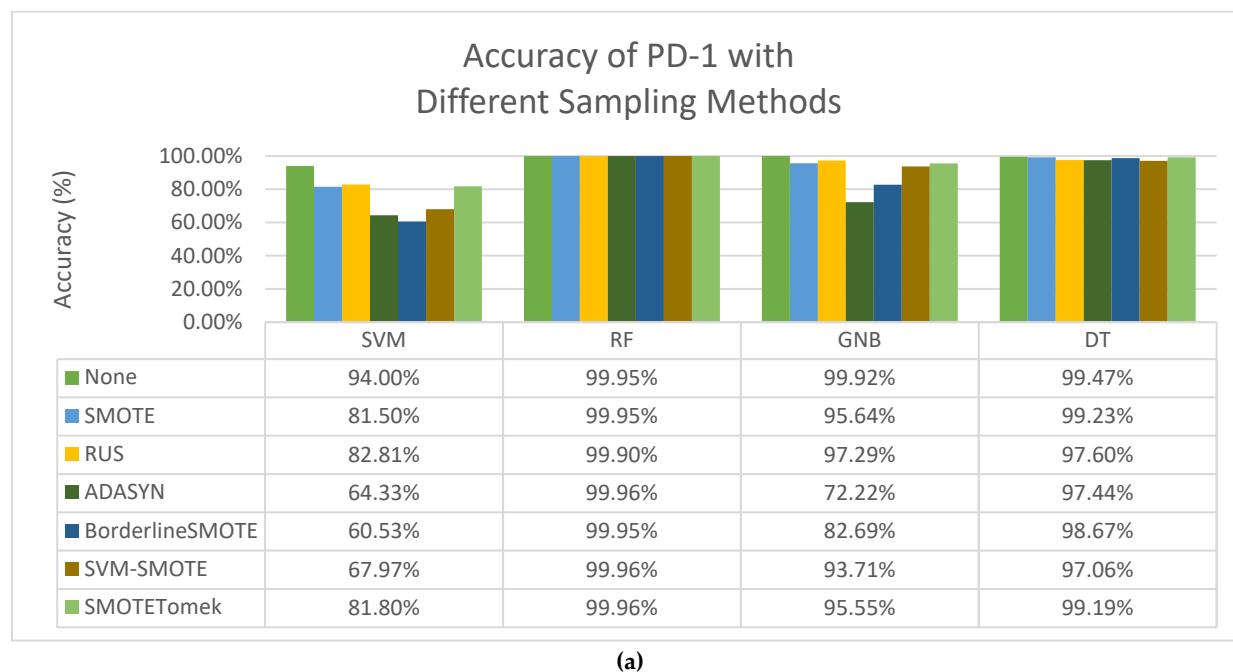
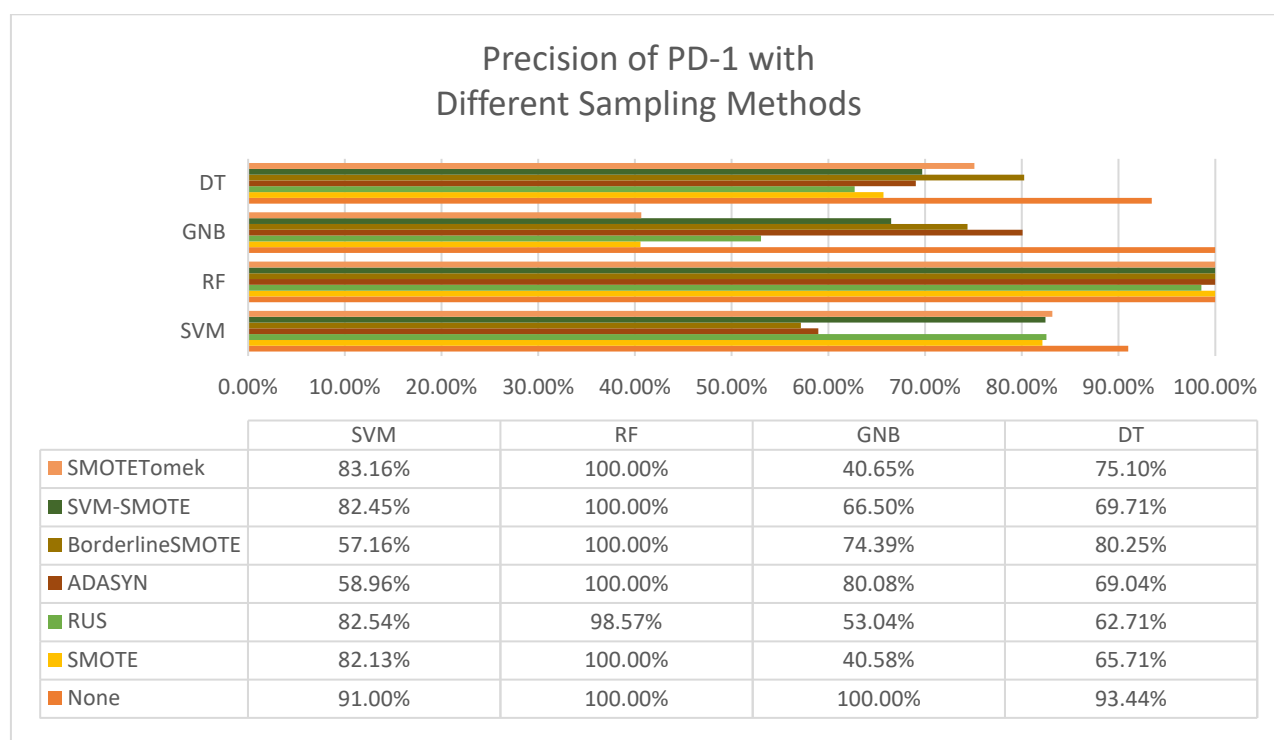


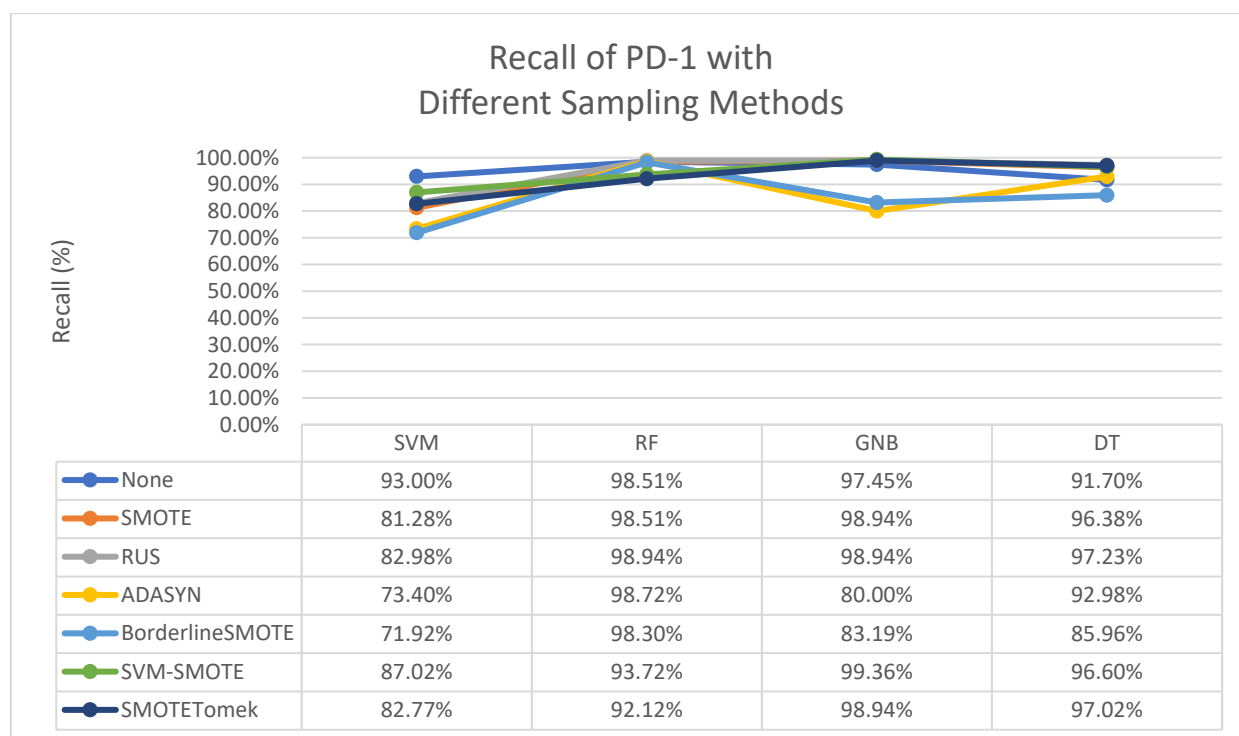
Figure S5. The accuracy, precision and recall values for MEK1. **(a)** Accuracy of MEK1 when different sampling methods are applied; **(b)** Precision of MEK1 when different sampling methods are applied; **(c)** Recall of MEK1 when different sampling methods are applied.

71
72
73





(b)

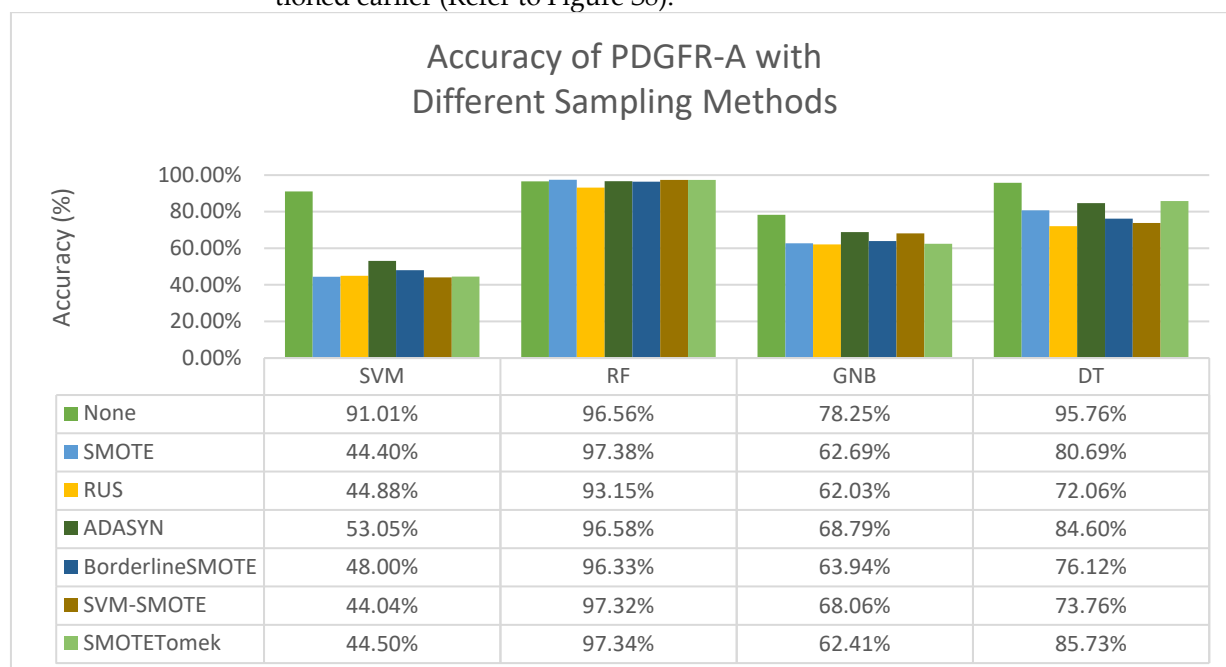


(c)

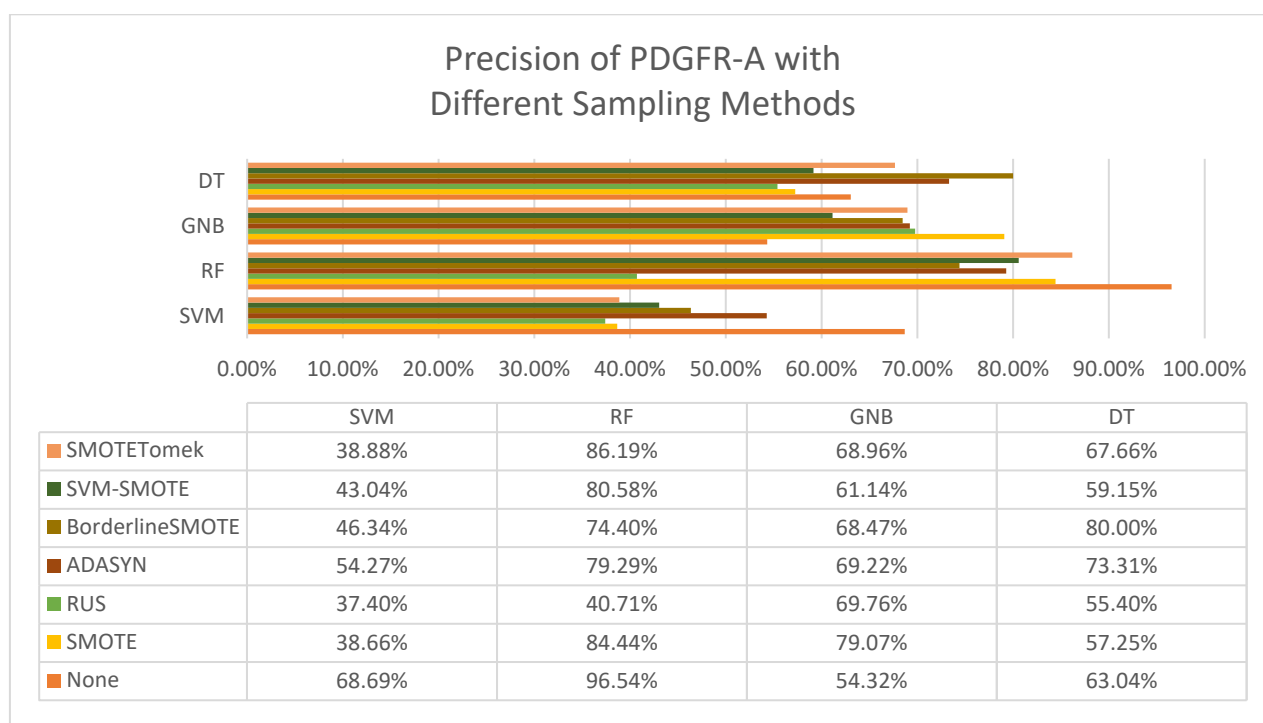
Figure S6. The accuracy, precision and recall values for PD-1. (a) Accuracy of PD-1 when different sampling methods are applied; (b) Precision of PD-1 when different sampling methods are applied; (c) Recall of PD-1 when different sampling methods are applied.

Next, for PDGFR-A, it was found that when SMOTE is applied on the RF classifier, the accuracy is the highest, yielding an accuracy value of 97.38% (Refer to Figure S7). However, in terms of precision, PDGFR-A seems to be performing well when no

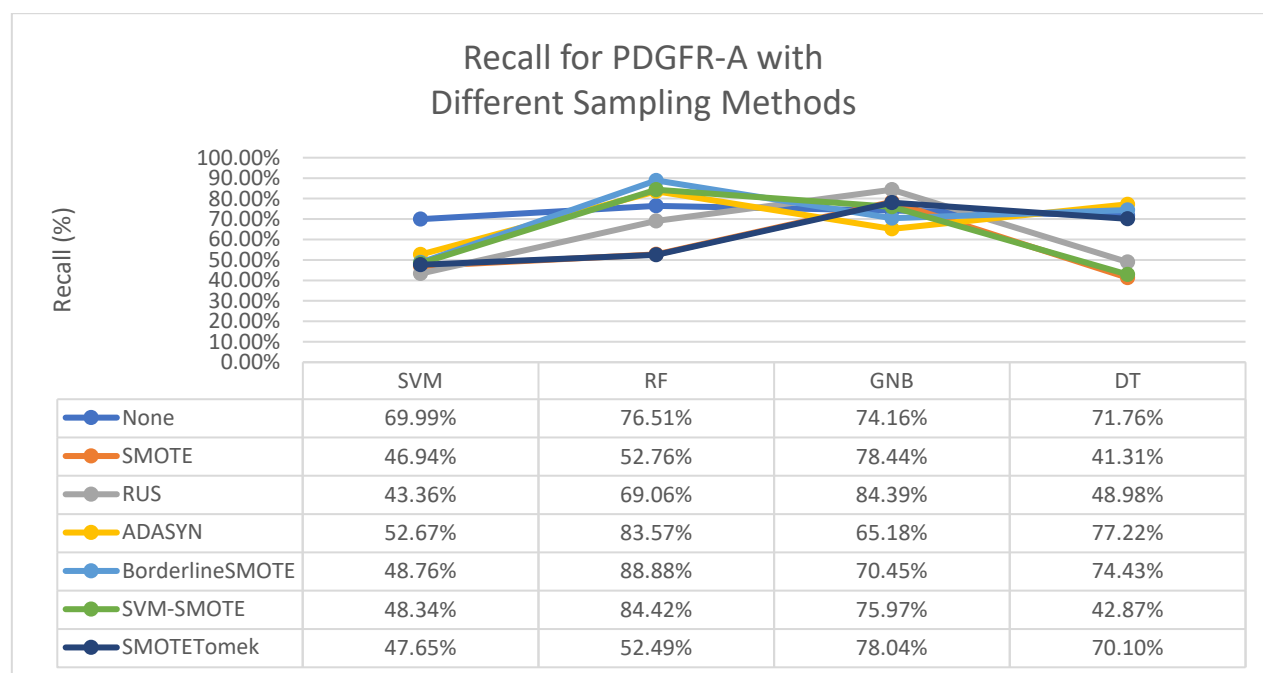
resampling method is used on the RF classifier, achieving a value of 96.54% (Refer to Figure S7). Nevertheless, a drop in the precision value can be seen across most of the classifiers when different resampling methods are applied, except for GNB in which when SMOTE is applied on the GNB classifier, a sharp increase of 24.75% in the precision value can be seen (Refer to Figure S7). For the recall of PDGFR-A, the value is the highest when BorderlineSMOTE is applied with the RF classifier with a value of 88.88% (Refer to Figure S7). Now for PDGFR-B, the accuracy value is the highest when the RF classifier is applied with SMOTE and BorderlineSMOTE, achieving a value of 99.87% respectively (Refer to Figure S8). The highest precision value recorded was 92.46%, and this was when RUS was paired with the RF classifier. However, when RUS is applied to the other classifiers, a sharp decrease is observed in the precision value (Refer to Figure S8). In terms of recall, the highest value achieved is when RUS was paired the RF classifier (99.50%) while a drop in recall values can also be seen in both the SVM and GNB classifiers when different resampling methods are applied, especially with ADASYN, BorderlineSMOTE and SVM-SMOTE, which is a similar case to the nature of the precision values of PDGFR-B mentioned earlier (Refer to Figure S8).



(a)



(b)



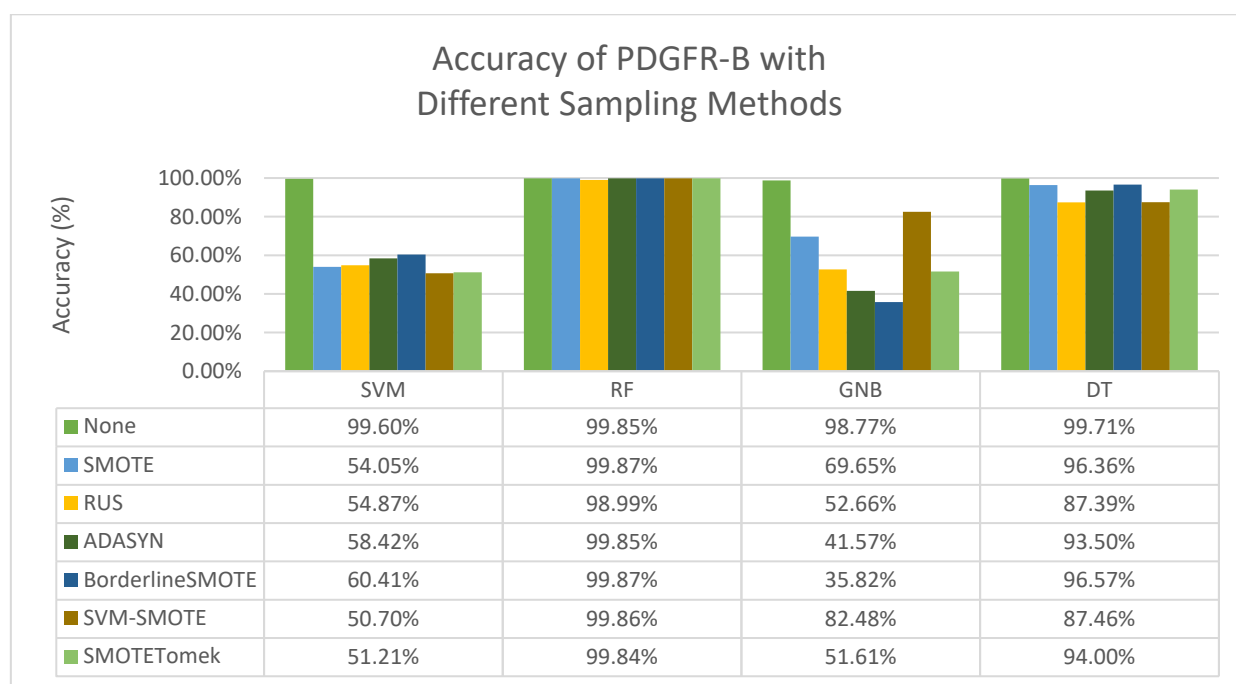
(c)

Figure S7. The accuracy, precision and recall values for PDGFR-A. (a) Accuracy of PDGFR-A when different sampling methods are applied; (b) Precision of PDGFR-A when different sampling methods are applied; (c) Recall of PDGFR-A when different sampling methods are applied.

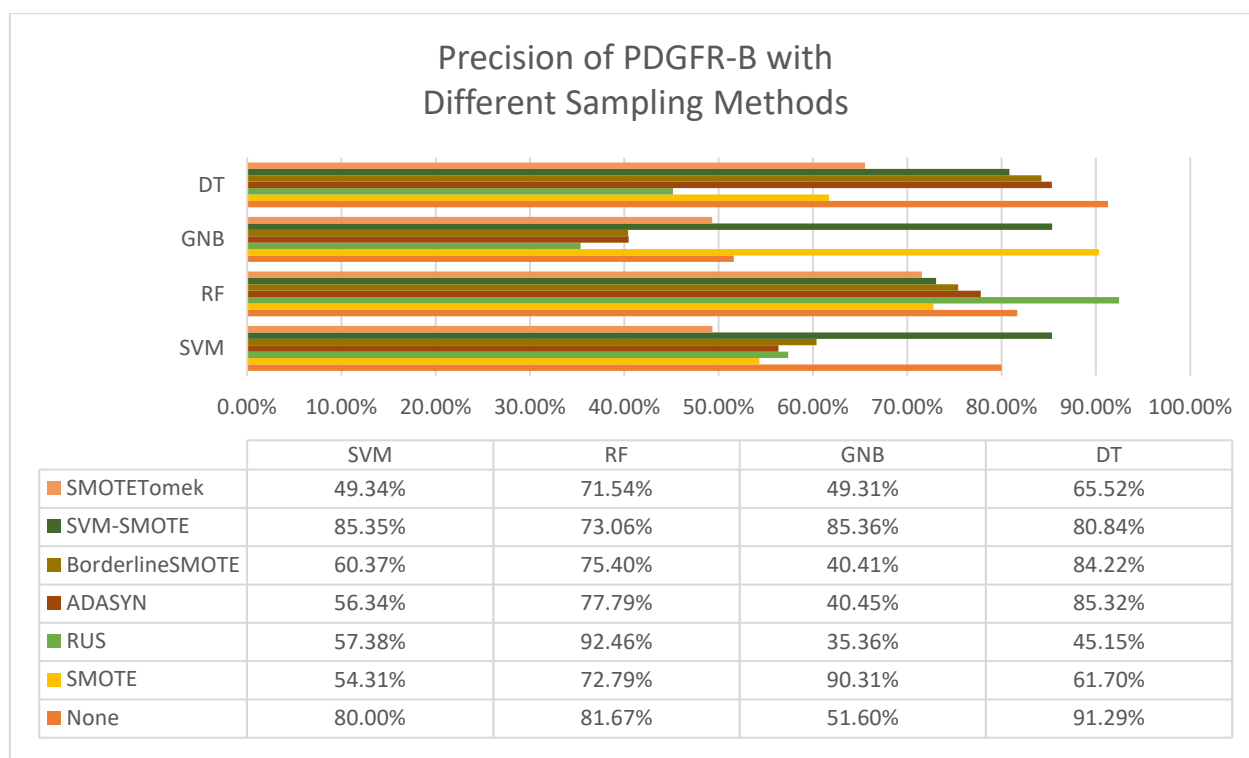
96

97

98



(a)



(b)

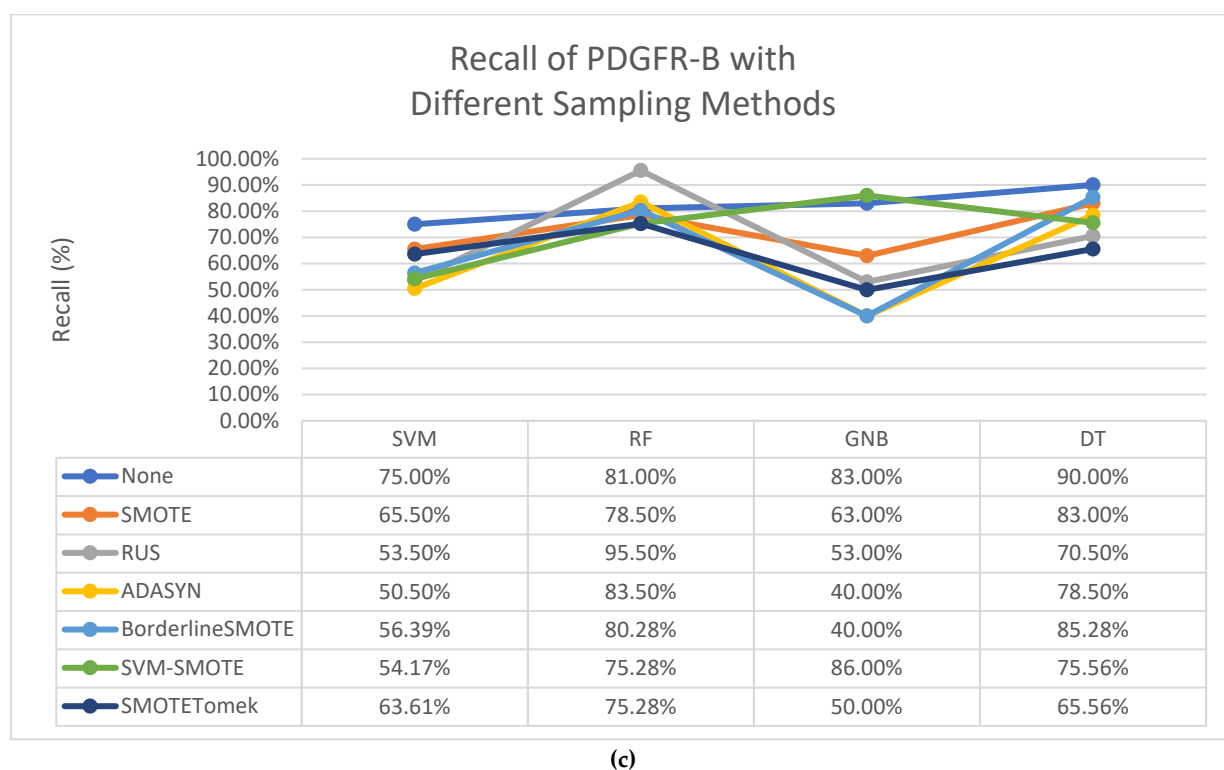
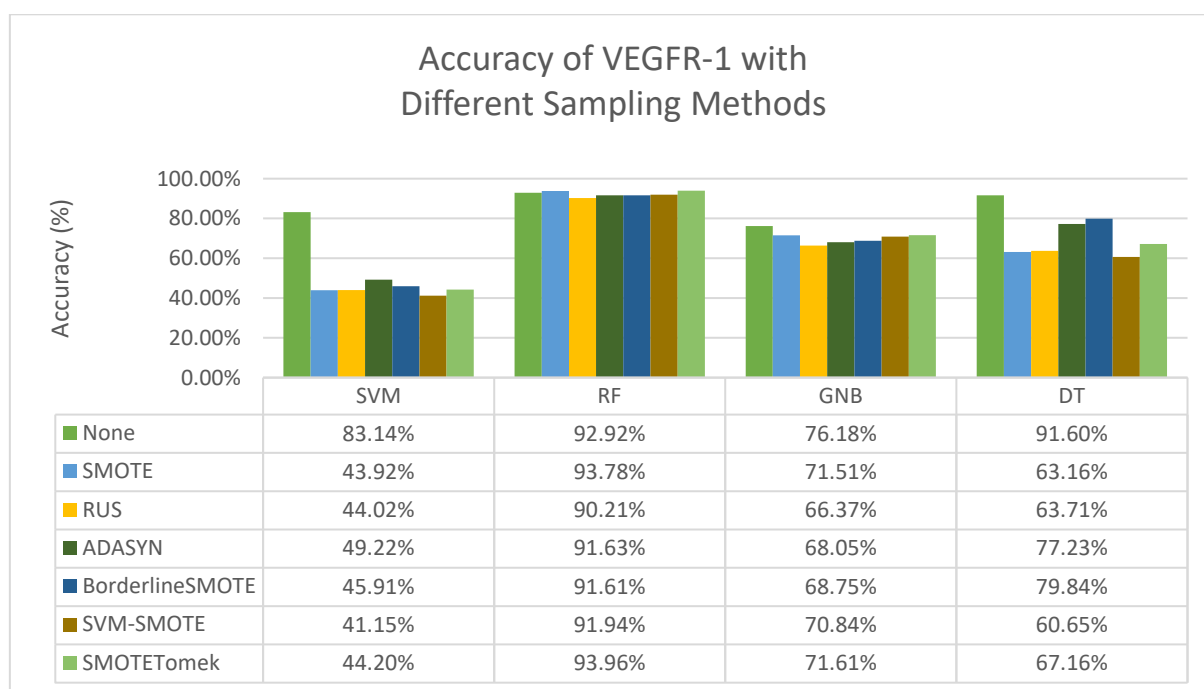
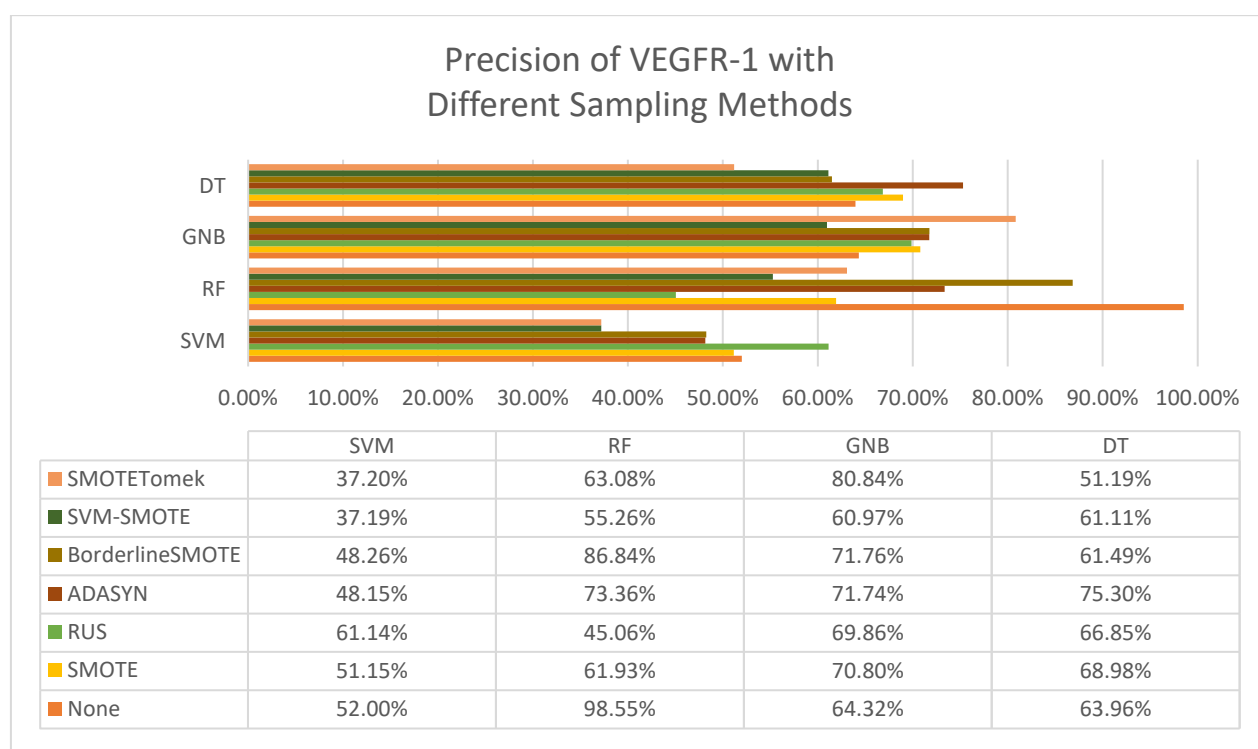


Figure S8. The accuracy, precision and recall values for PDGFR-B. (a) Accuracy of PDGFR-B when different sampling methods are applied; (b) Precision of PDGFR-B when different sampling methods are applied; (c) Recall of PDGFR-B when different sampling methods are applied.

Additionally, the accuracy for VEGFR-1 is the highest when SMOTETomek is applied on the RF classifier, achieving a value of 93.96%, which is a 1.04% increase compared to its original accuracy value when no resampling method is used with the RF classifier, 92.92% (Refer to Figure S9). In terms of precision, it is interesting to note that when no resampling method is applied on the RF classifier, the precision is the highest with a value of 98.55%, similar to activity classes KRAS and PD-1 (Refer to Figures S4, S6 and S9). Similarly, the same case is also happening to the recall value of VEGFR-1, whereby the value is the highest when no resampling method is used (similar to activity classes KRAS and PD-1), achieving a value of 88.94% (Refer to Figures S4, S6 and S9). Finally, the accuracy of VEGFR-2 is the highest when SVM-SMOTE is paired with the RF classifier, yielding an accuracy value of 91.94%, which is a 10.99% increase from the original value of 80.95% when no resampling was used (Refer to Figure S10). In addition, when RF is paired with BorderlineSMOTE, the best precision value recorded was 87.08% and in terms of recall, the value is surprisingly the highest when no resampling method is used on the RF classifier, yielding a value of 99.48% which is similar to the case of the recall values of the activity classes KRAS, PD-1 and VEGFR-1 (Refer to Figures S4, S6, S9 and S10).



(a)



(b)

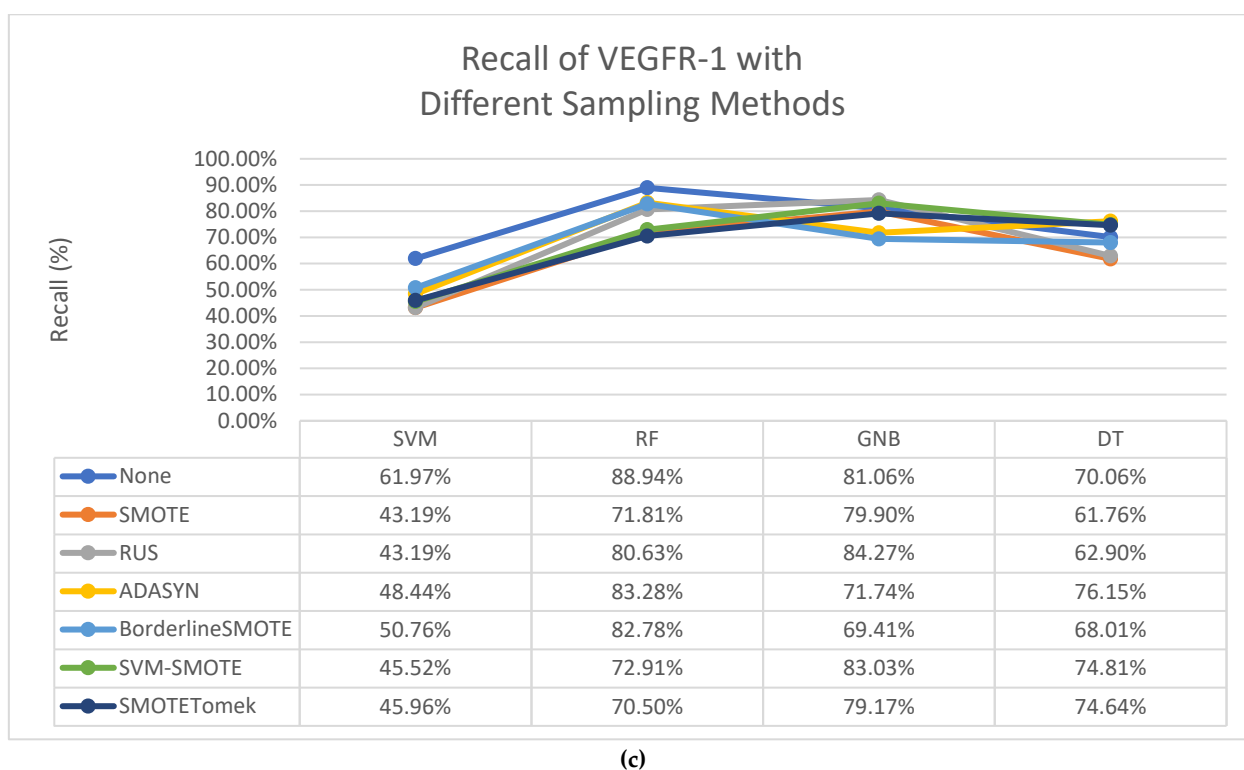
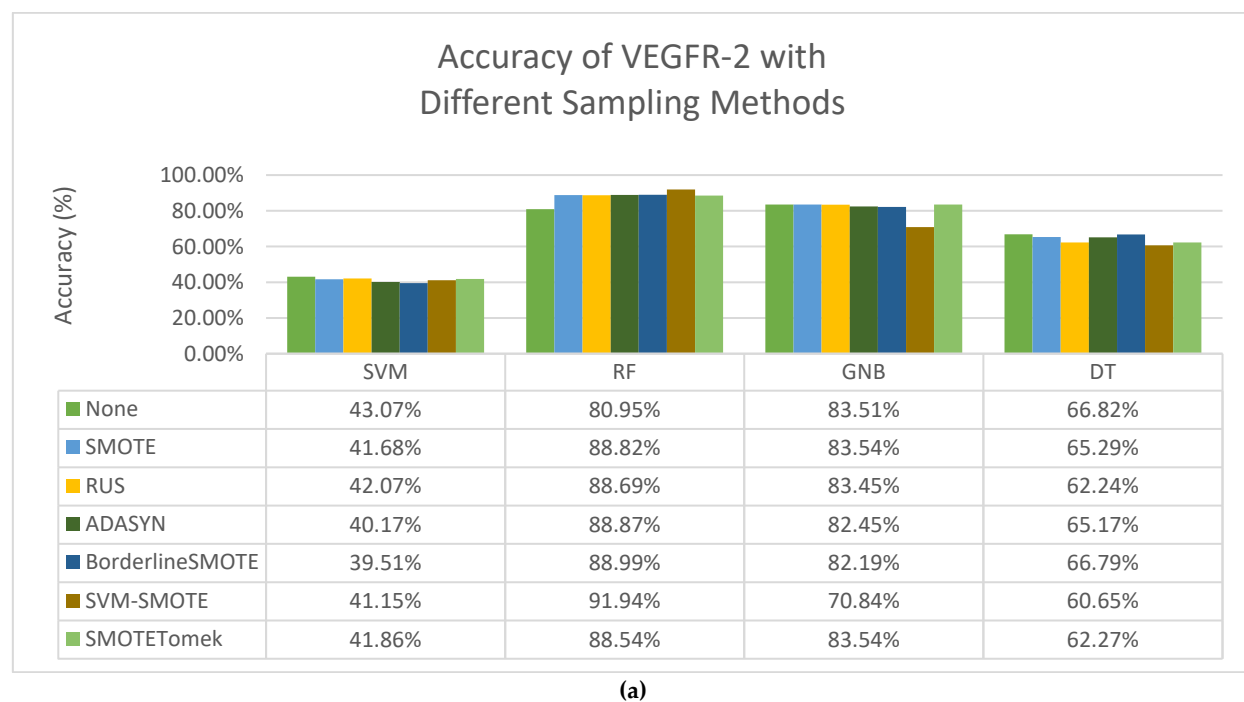
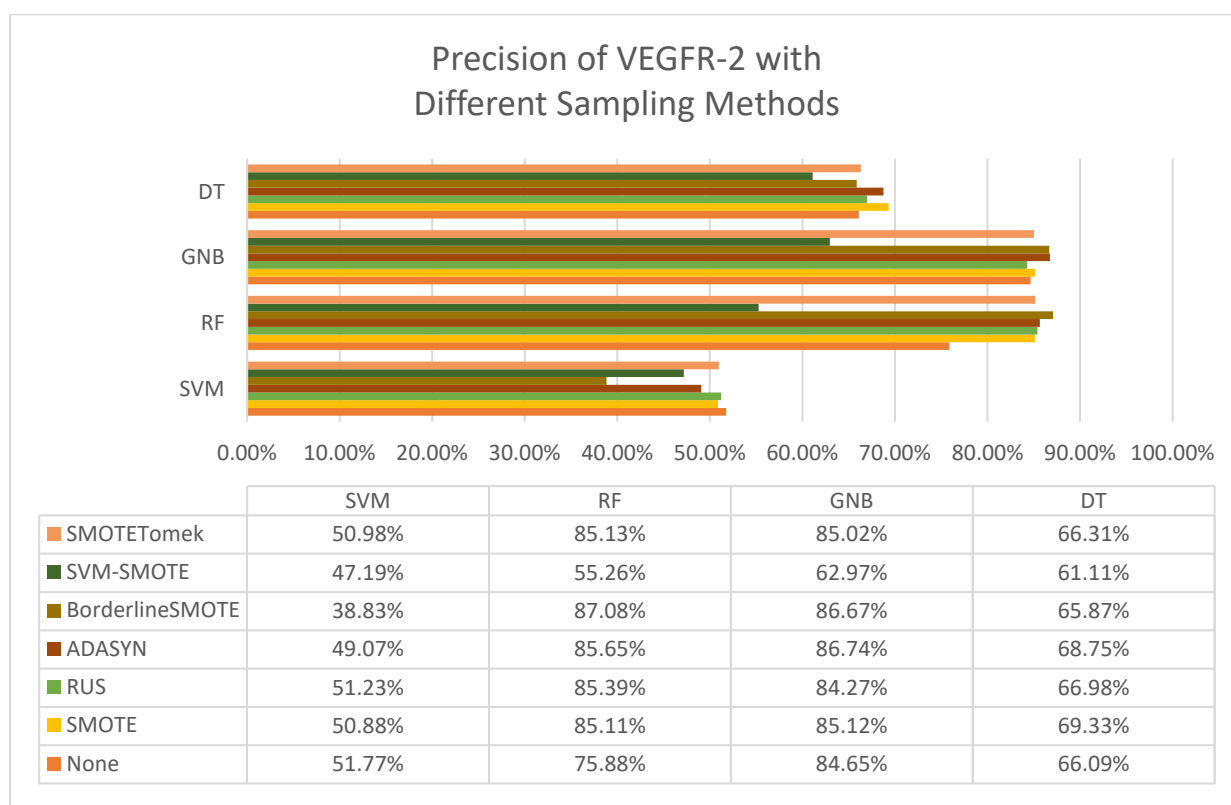
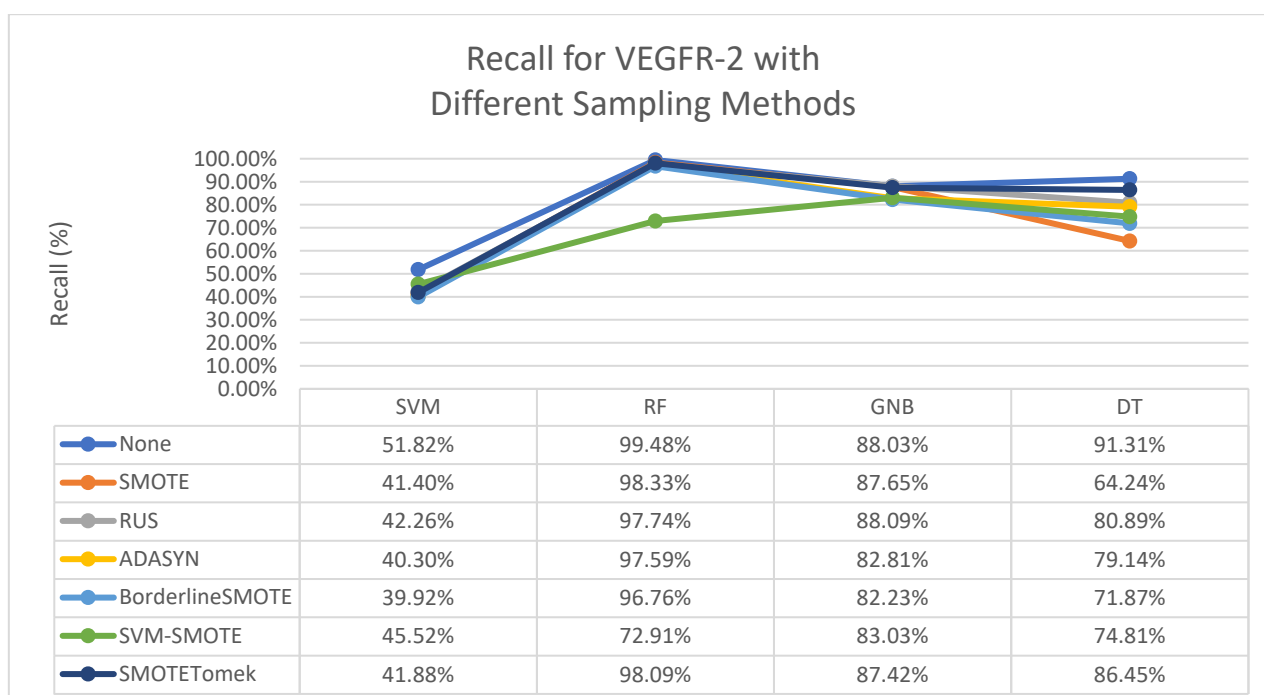


Figure S9. The accuracy, precision and recall values for VEGFR-1. (a) Accuracy of VEGFR-1 when different sampling methods are applied; (b) Precision of VEGFR-1 when different sampling methods are applied; (c) Recall of VEGFR-1 when different sampling methods are applied.





(b)



(c)

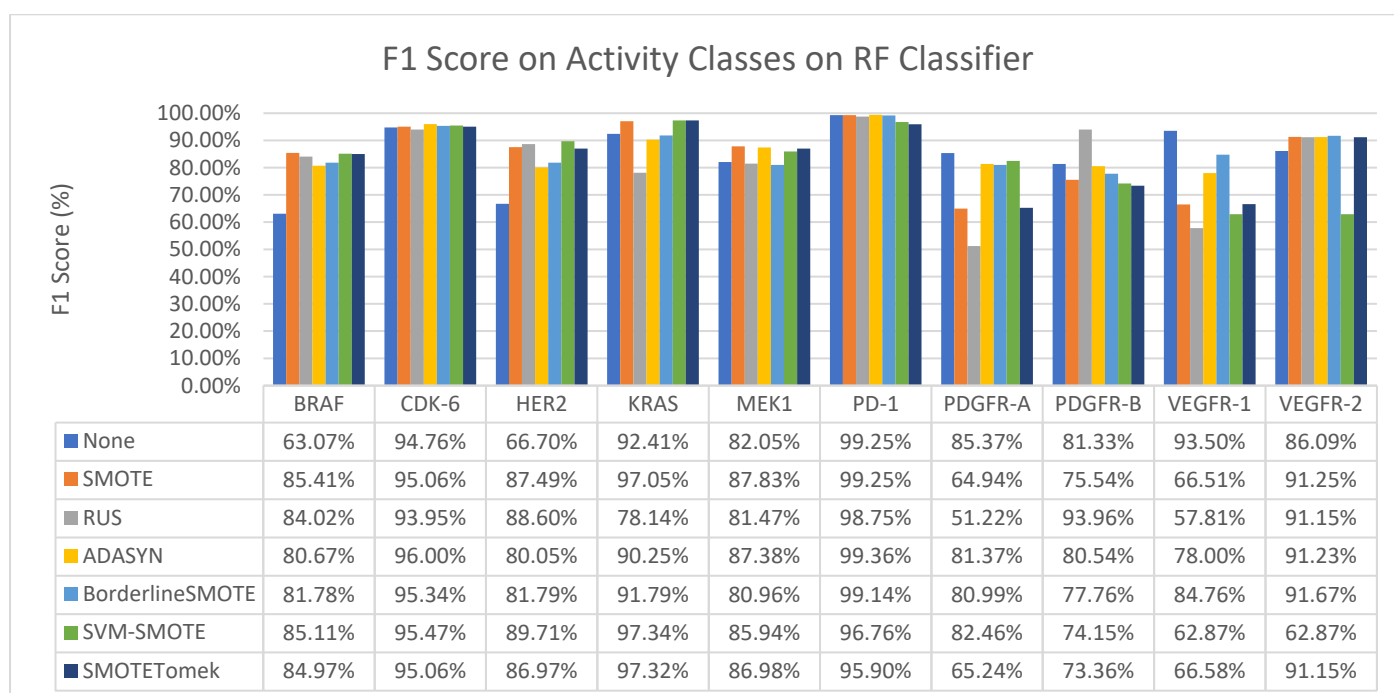
Figure S10. The accuracy, precision and recall values for VEGFR-2. (a) Accuracy of VEGFR-2 when different sampling methods are applied; (b) Precision of VEGFR-2 when different sampling methods are applied; (c) Recall of VEGFR-2 when different sampling methods are applied.

122

123

124

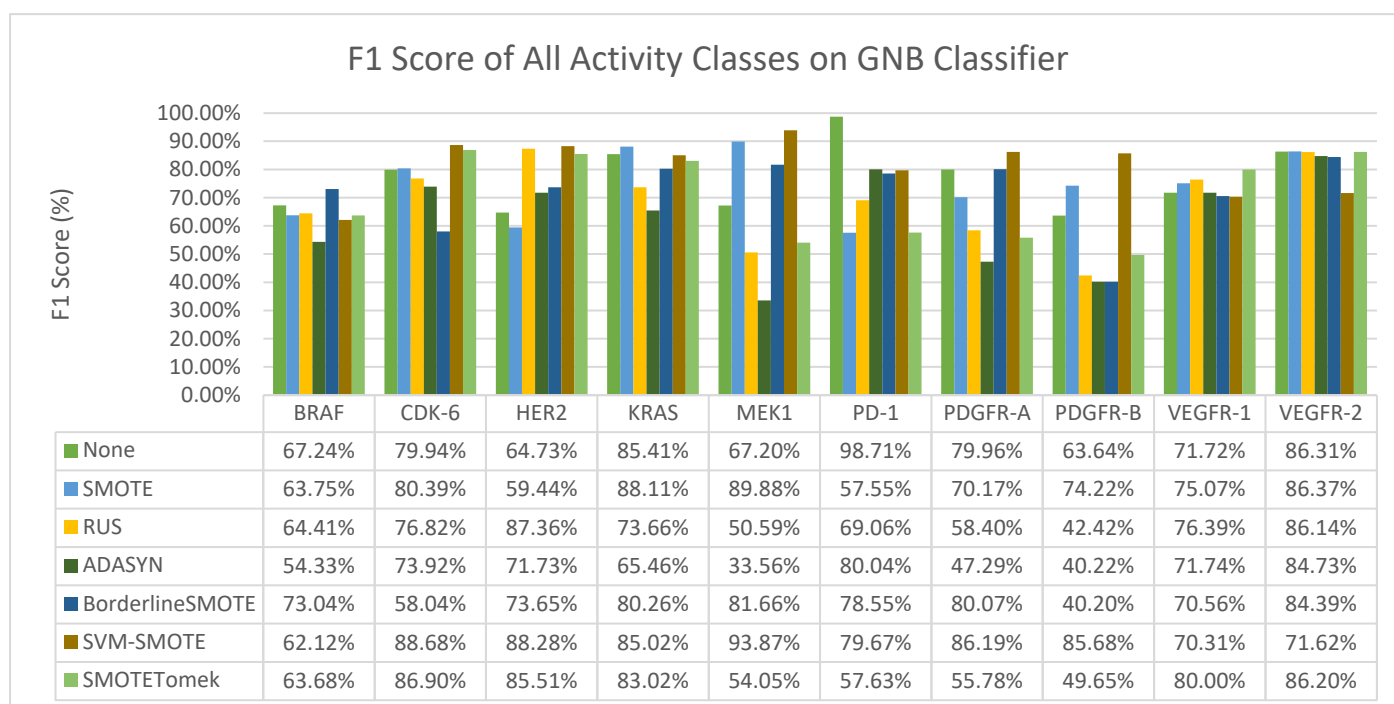
125



126

Figure S11. F1 scores for all activity classes with the RF classifier

127



128

Figure S12. F1 score for all activity classes with the GNB classifier

129