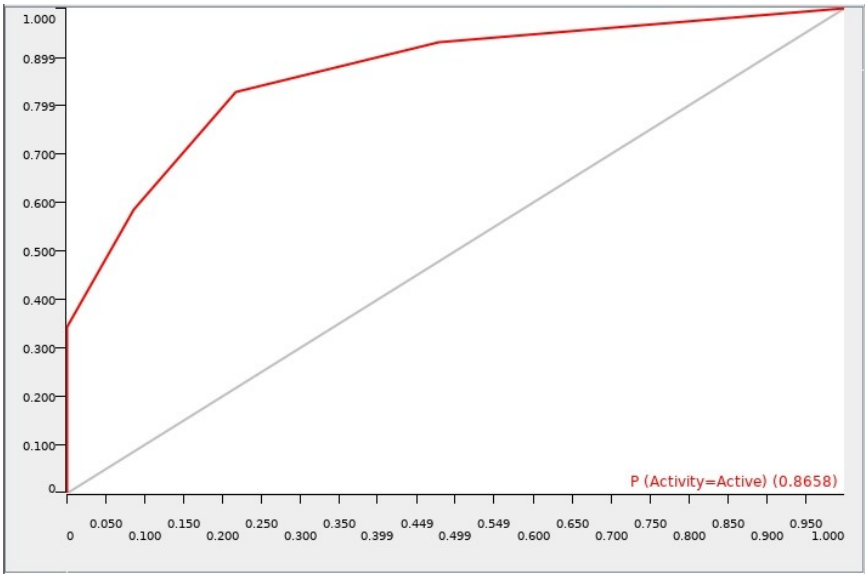
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

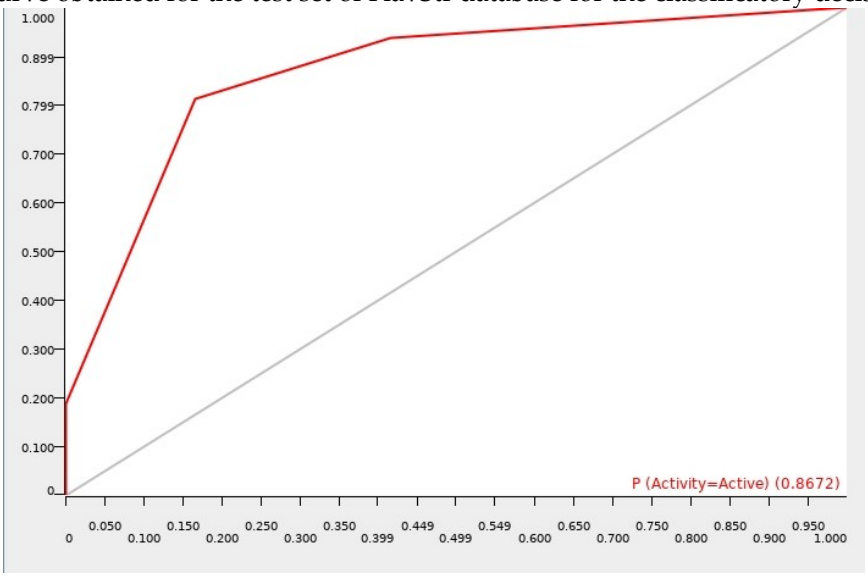
Adriana de Oliveira Fernandes, Valéria Vieira Moura Paixão, Yria Jaine Andrade Santos, Eduardo Borba Alves, Ricardo Pereira Rodrigues, Daniela Aparecida Chagas-Paula, Aurélia Santos Faraoni, Rosana Casoti, Marcus Vinicius de Aragão Batista, Marcel Bermudez, Silvio Santana Dolabella and Tiago Branquinho Oliveira

Figure S1. ROC curve obtained for the training set of FlavStr database for the classificatory decision tree (DTc).



The y axis represents the true positive rates, while the x axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

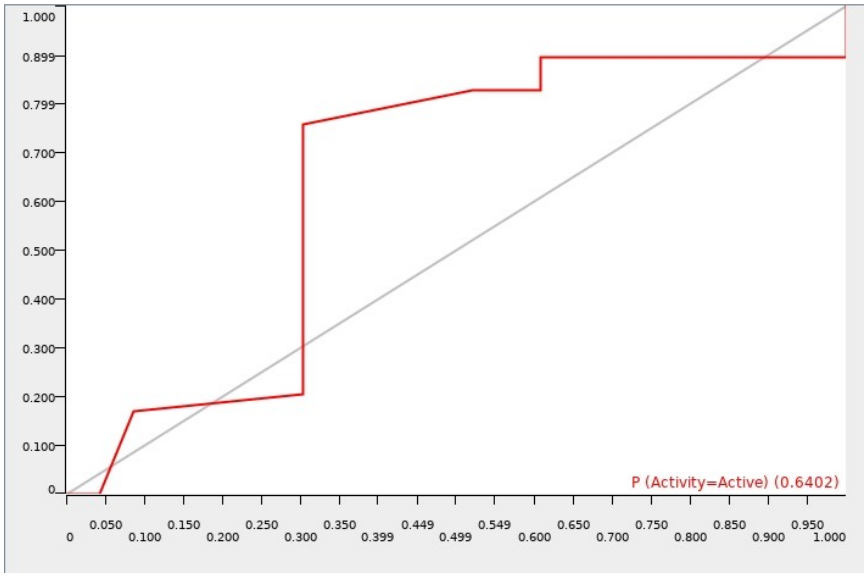
Figure S2. ROC curve obtained for the test set of FlavStr database for the classificatory decision tree (DTc).



The y axis represents the true positive rates, while the x axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

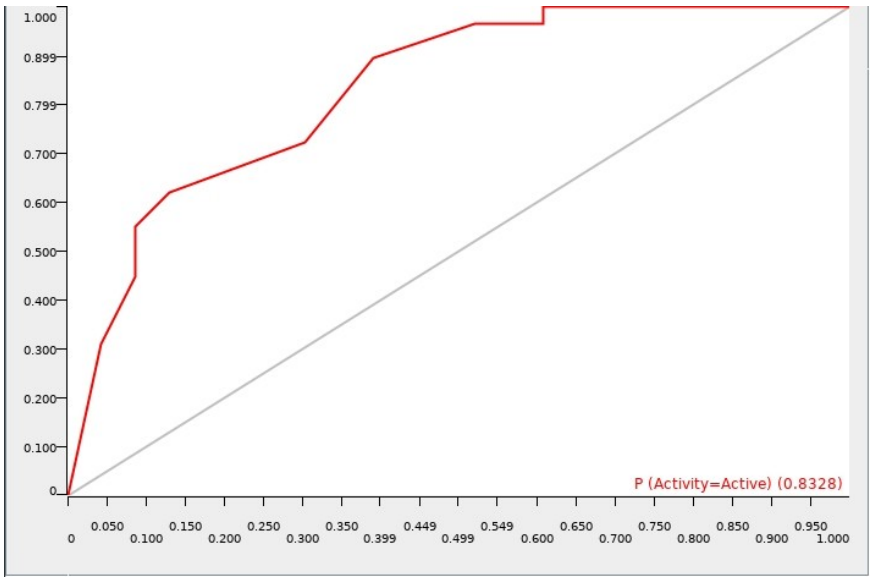
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S3. ROC curve obtained for the cross-validation set of FlavStr database for the classificatory decision tree (DTc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

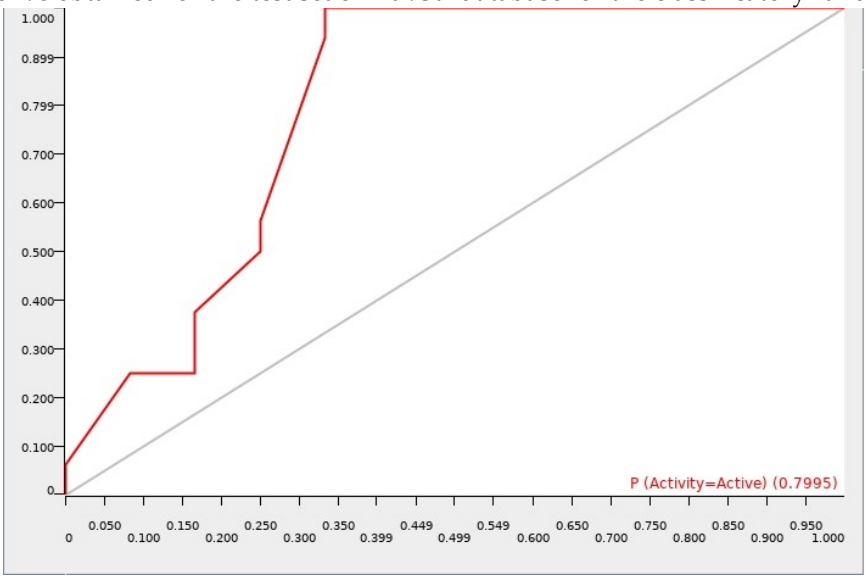
Figure S4. ROC curve obtained for the training set of FlavStr database for the classificatory random forest (RFc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

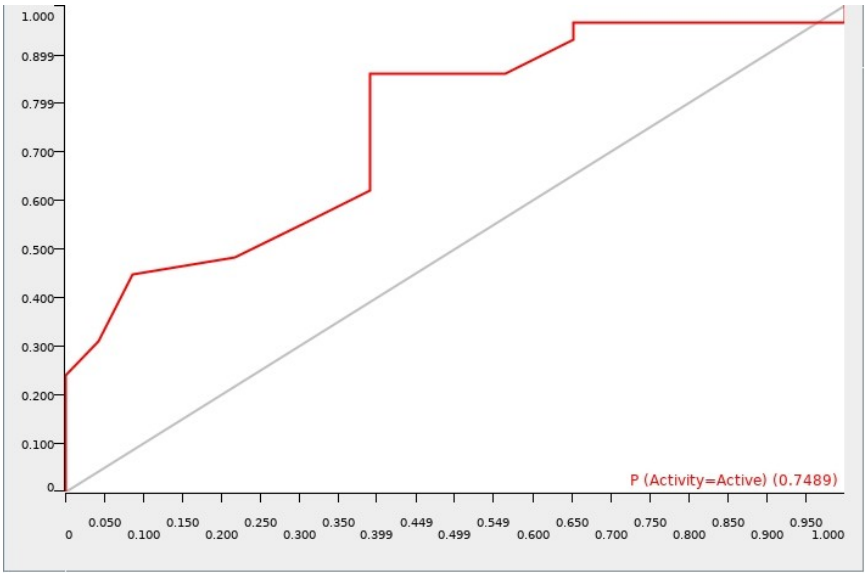
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S5. ROC curve obtained for the test set of FlavStr database for the classificatory random forest (RFc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

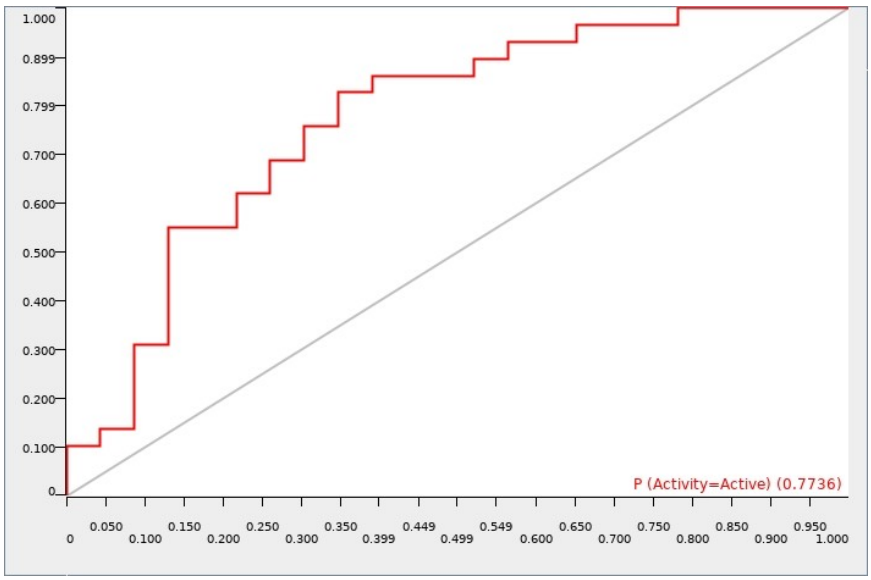
Figure S6. ROC curve obtained for the cross-validation set of FlavStr database for the classificatory random forest (RFc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

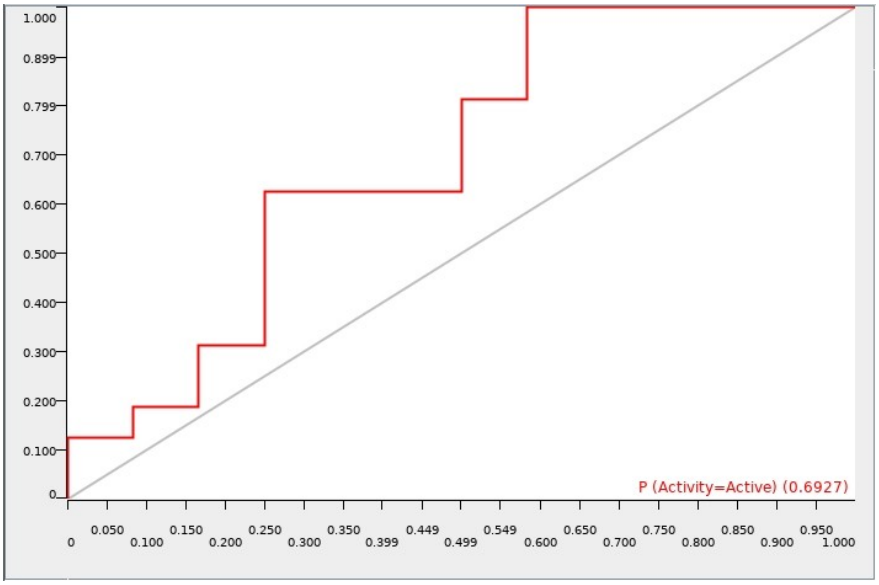
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S7. ROC curve obtained for the training set of FlavStr database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

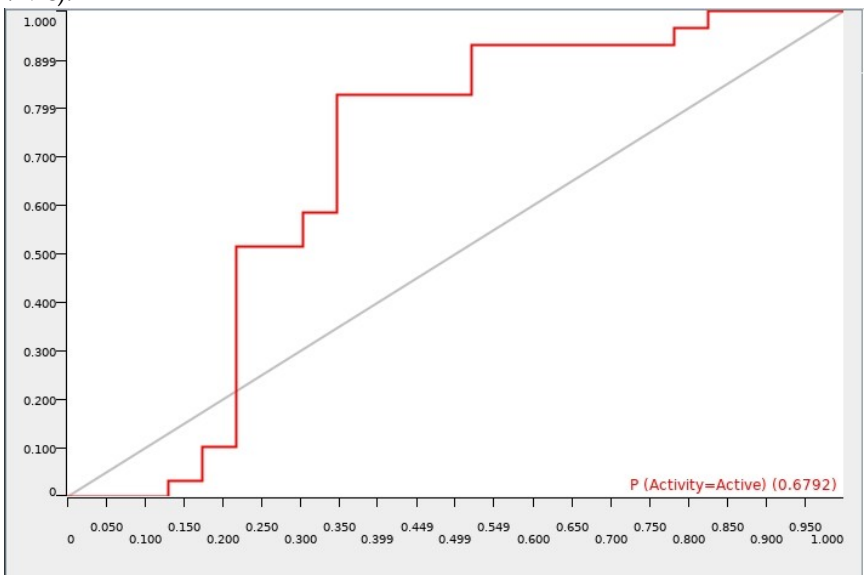
Figure S8. ROC curve obtained for the test set of FlavStr database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

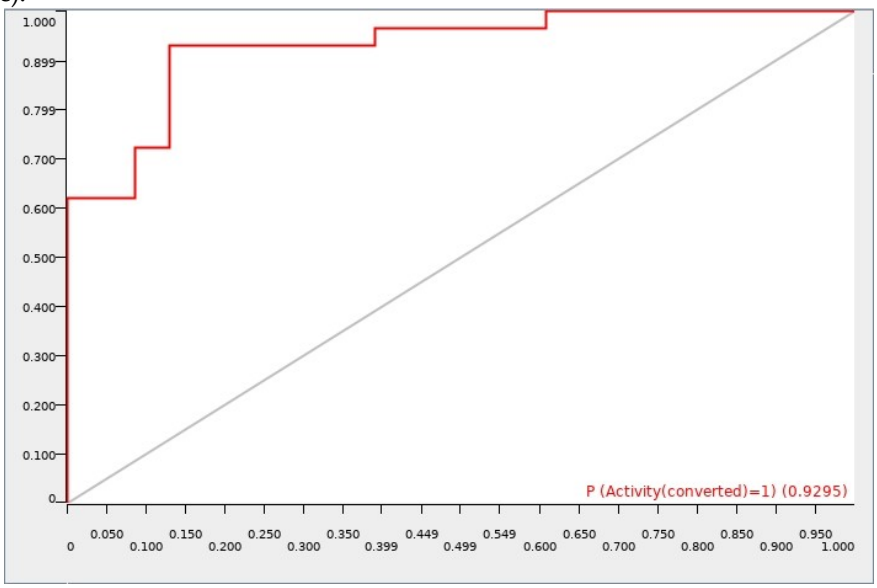
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S9. ROC curve obtained for the cross-validation set of FlavStr database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

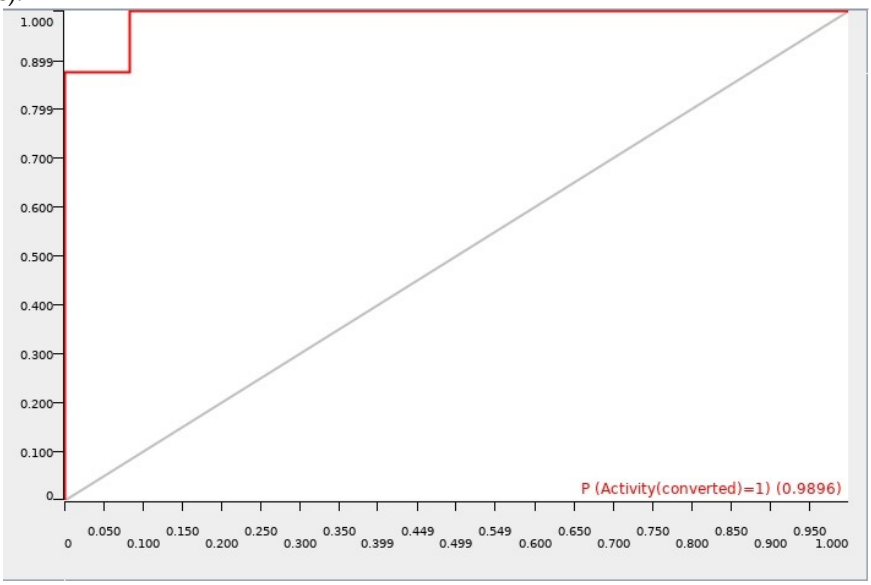
Figure S10. ROC curve obtained for the training set of FlavStr database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

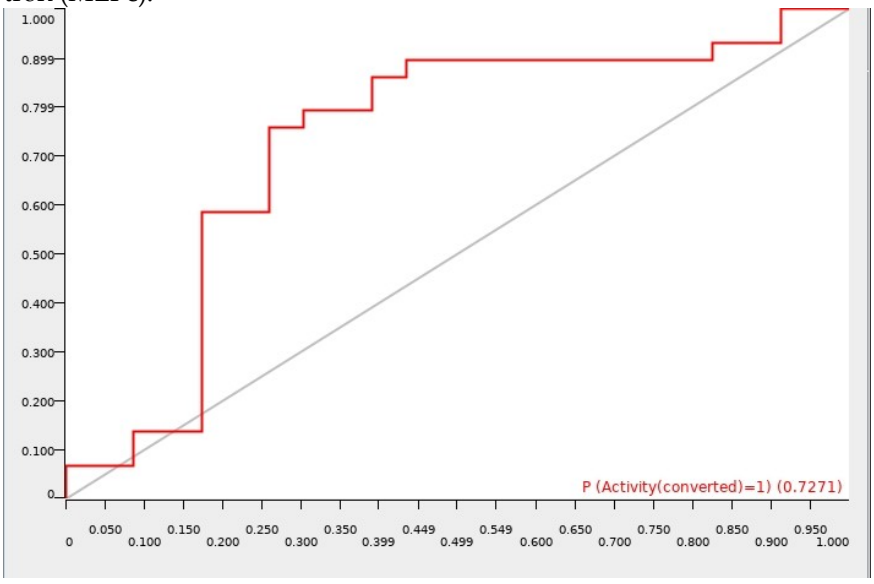
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S11. ROC curve obtained for the test set of FlavStr database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

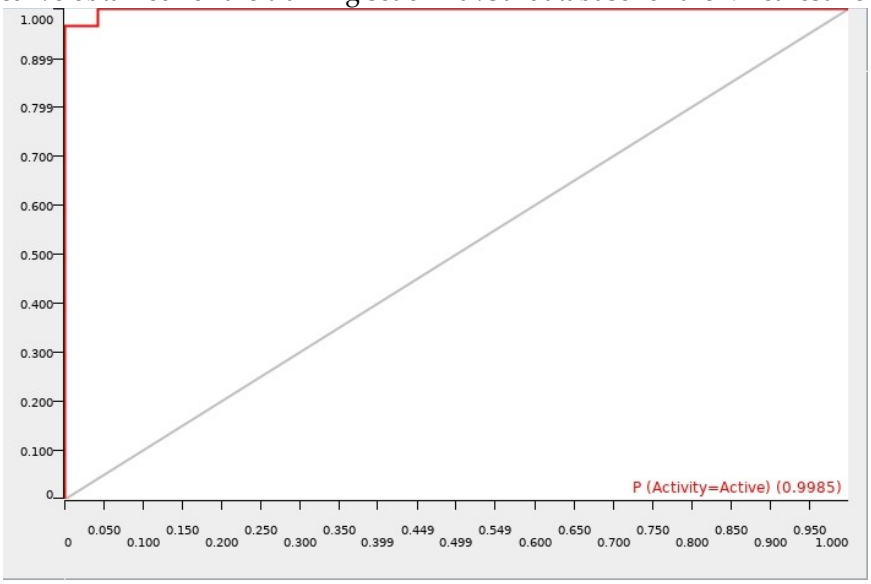
Figure S12. ROC curve obtained for the cross-validation set of FlavStr database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

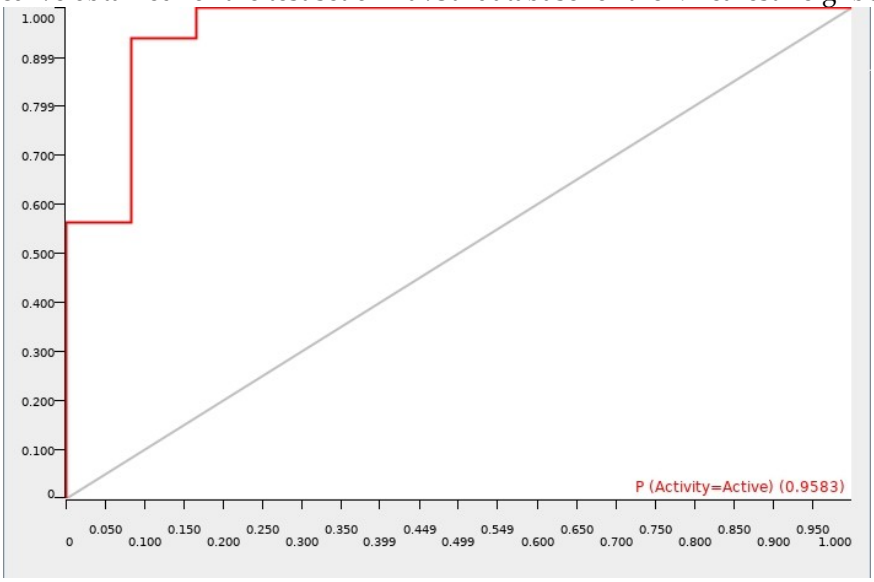
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S13. ROC curve obtained for the training set of FlavStr database for the k-nearest neighbour (KNN).



The y axis represents the true positive rates, while the x axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

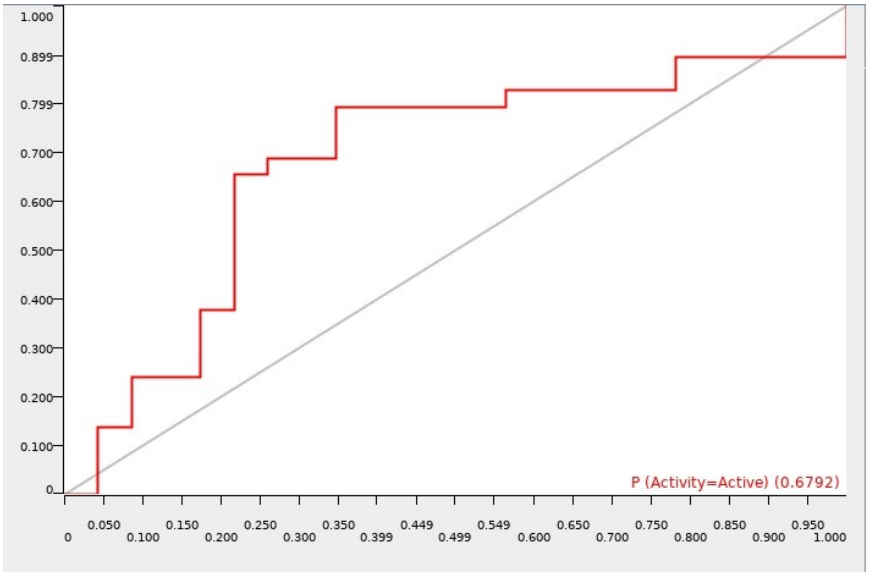
Figure S14. ROC curve obtained for the test set of FlavStr database for the k-nearest neighbour (KNN).



The y axis represents the true positive rates, while the x axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

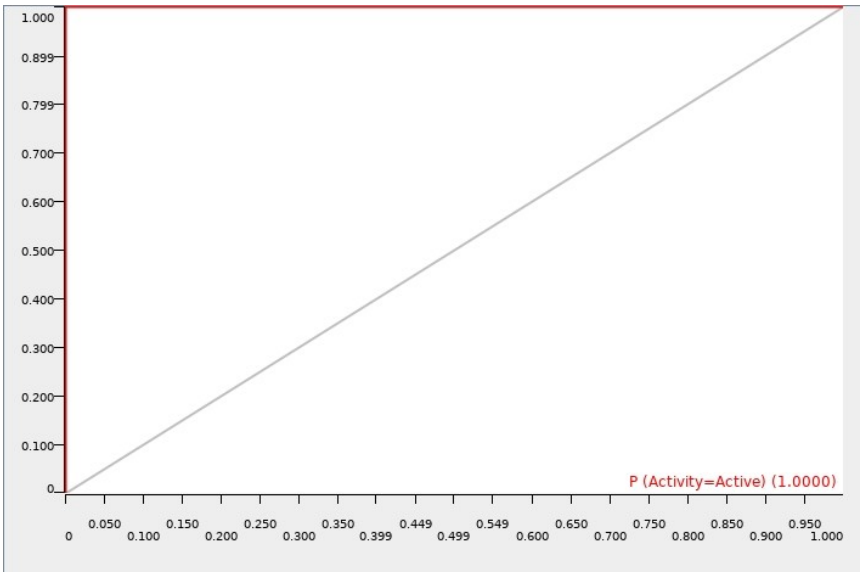
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S15. ROC curve obtained for the cross-validation set of FlavStr database for the k-nearest neighbour (KNN).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

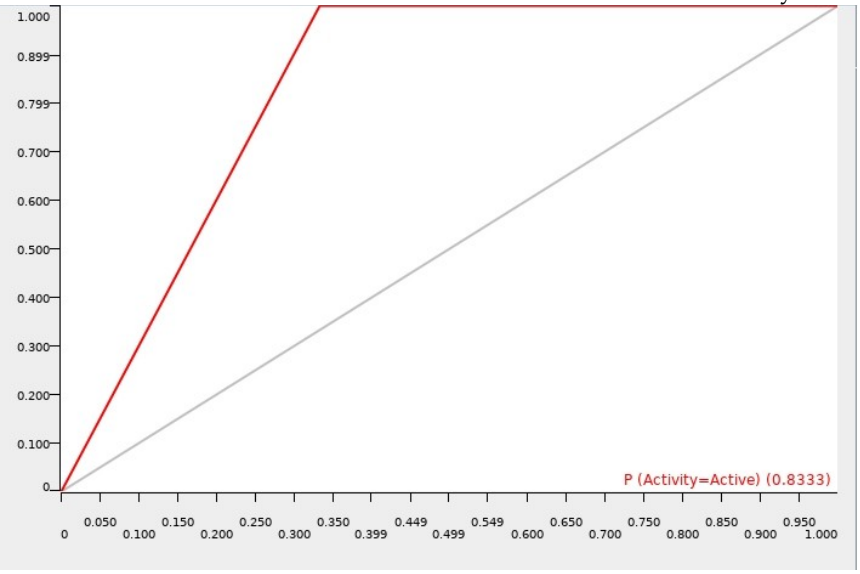
Figure S16. ROC curve obtained for the training set of FabG database for the classificatory decision tree (DTc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

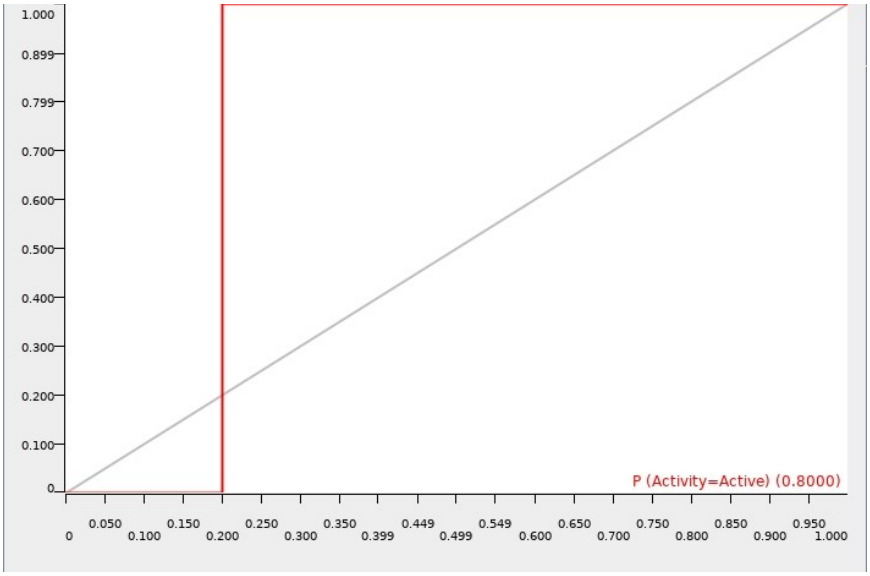
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S17. ROC curve obtained for the test set of FabG database for the classificatory decision tree (DTc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

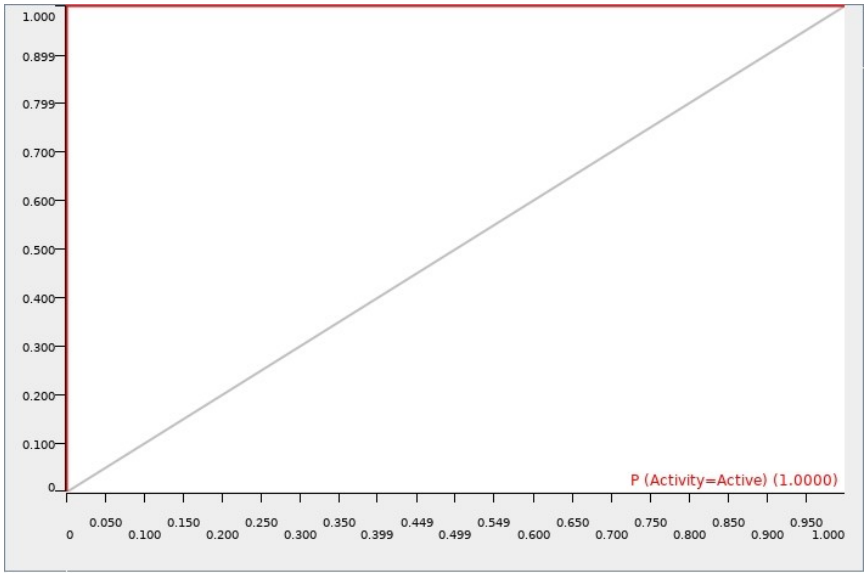
Figure S18. ROC curve obtained for the cross-validation set of FabG database for the classificatory decision tree (DTc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

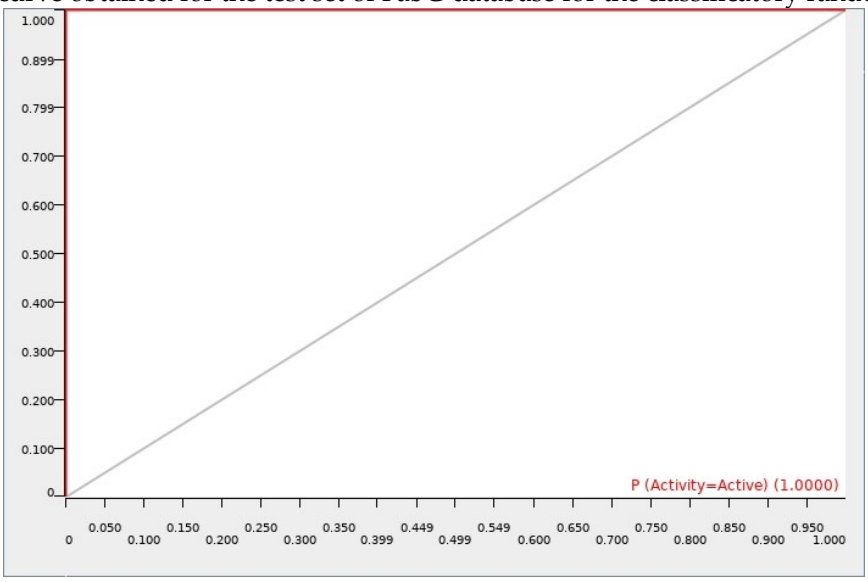
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S19. ROC curve obtained for the training set of FabG database for the classificatory random forest (RFc).



The y axis represents the true positive rates, while the x axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

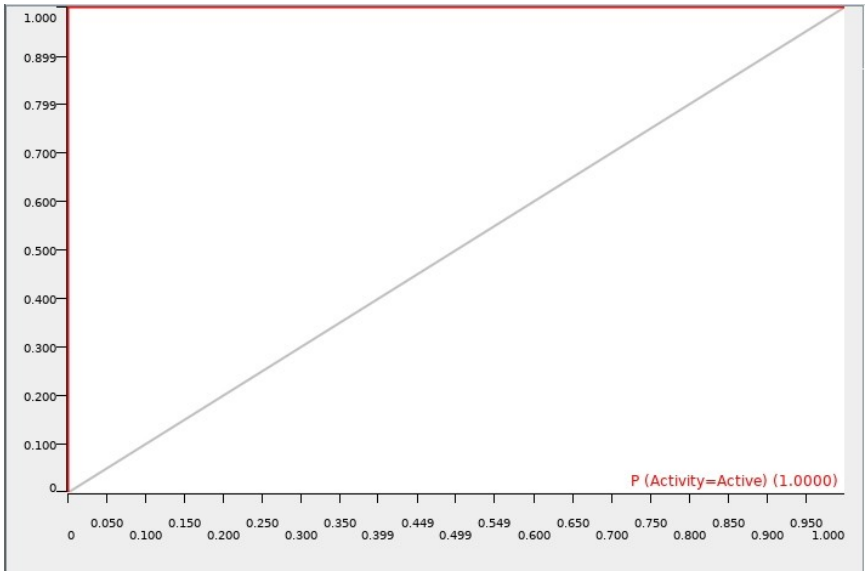
Figure S20. ROC curve obtained for the test set of FabG database for the classificatory random forest (RFc).



The y axis represents the true positive rates, while the x axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

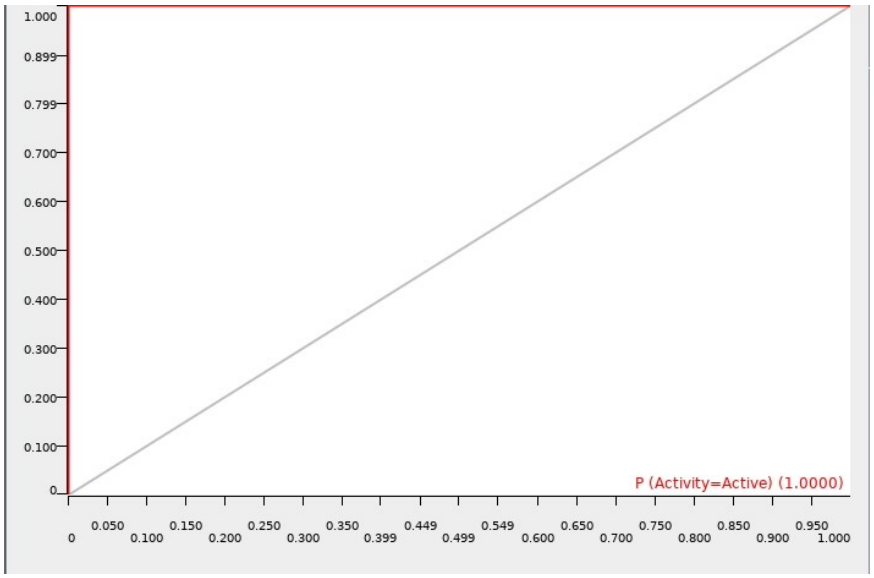
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S21. ROC curve obtained for the cross-validation set of FabG database for the classificatory random forest (RFc).



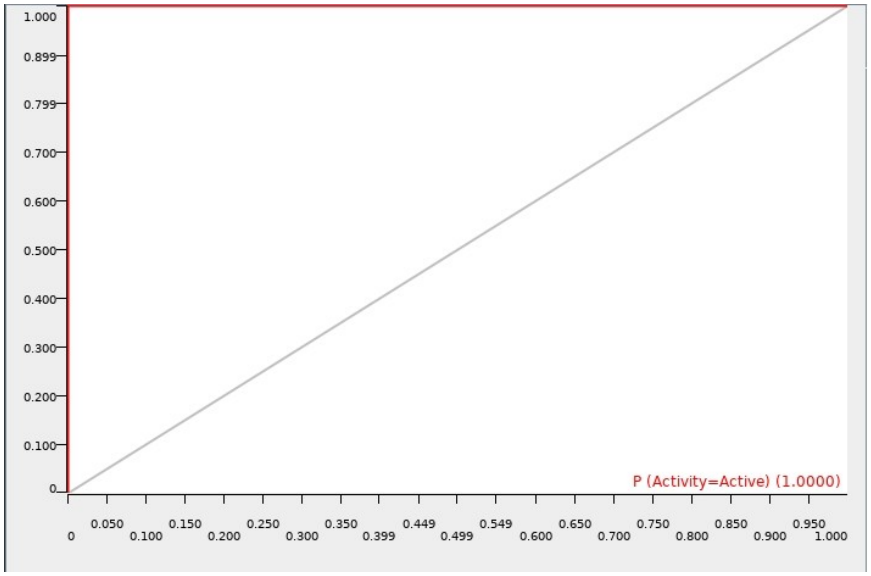
The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

Figure S22. ROC curve obtained for the training set of FabG database for the classificatory support vector machine (SVMc).



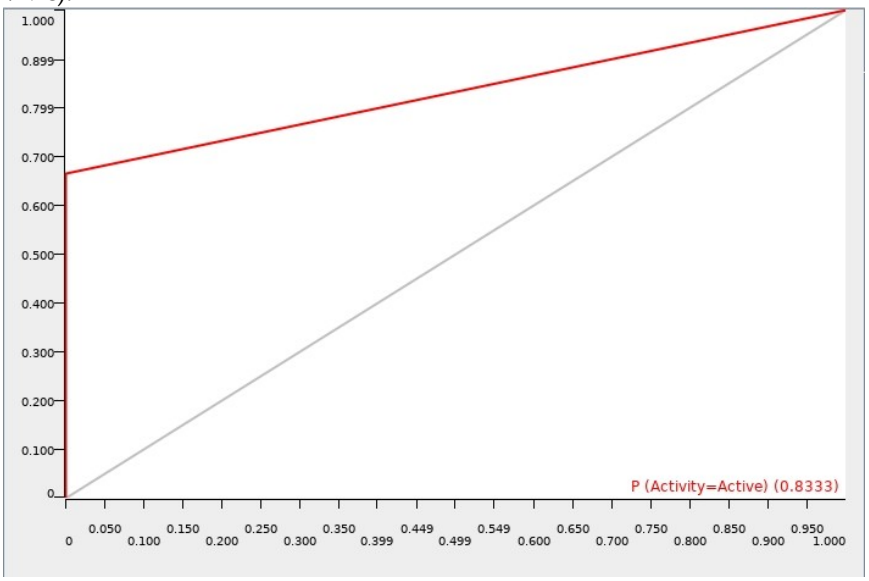
The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

Figure S23. ROC curve obtained for the test set of FabG database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

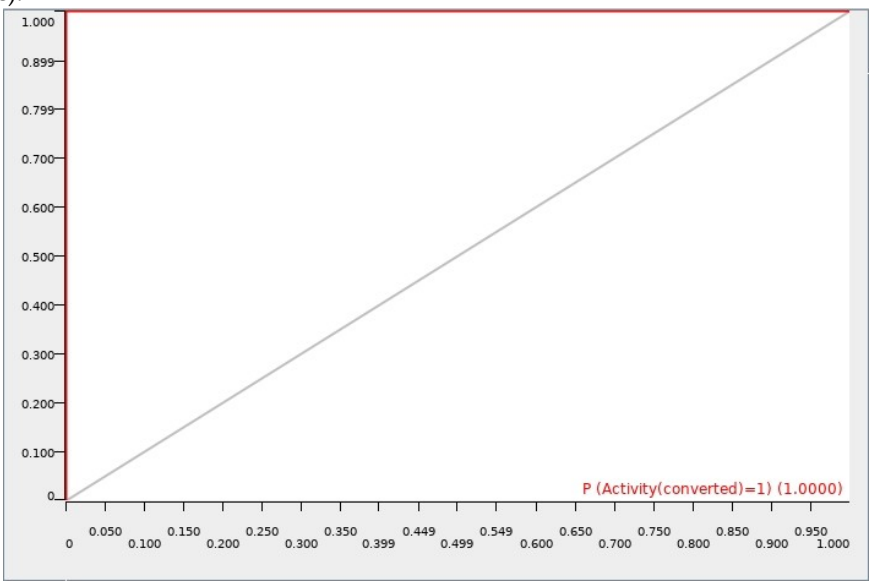
Figure S24. ROC curve obtained for the cross-validation set of FabG database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

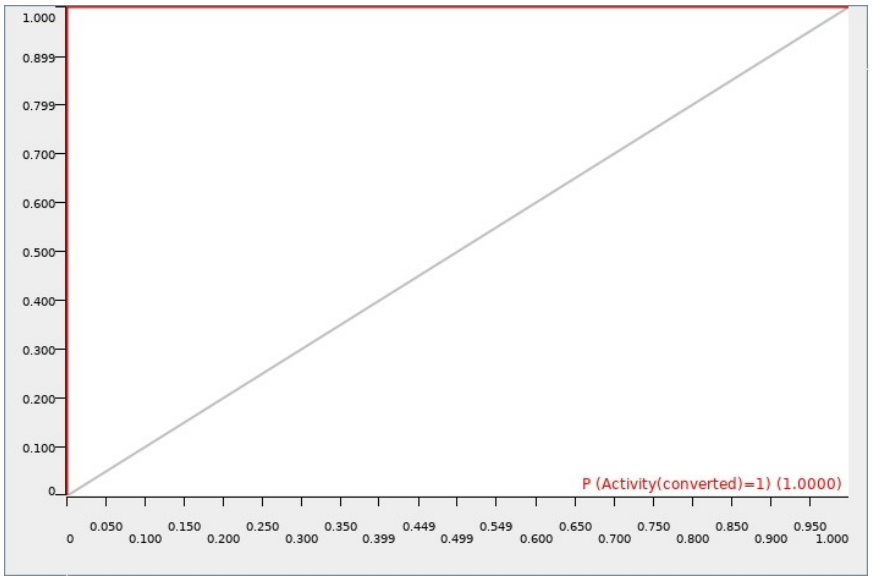
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S25. ROC curve obtained for the training set of FabG database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

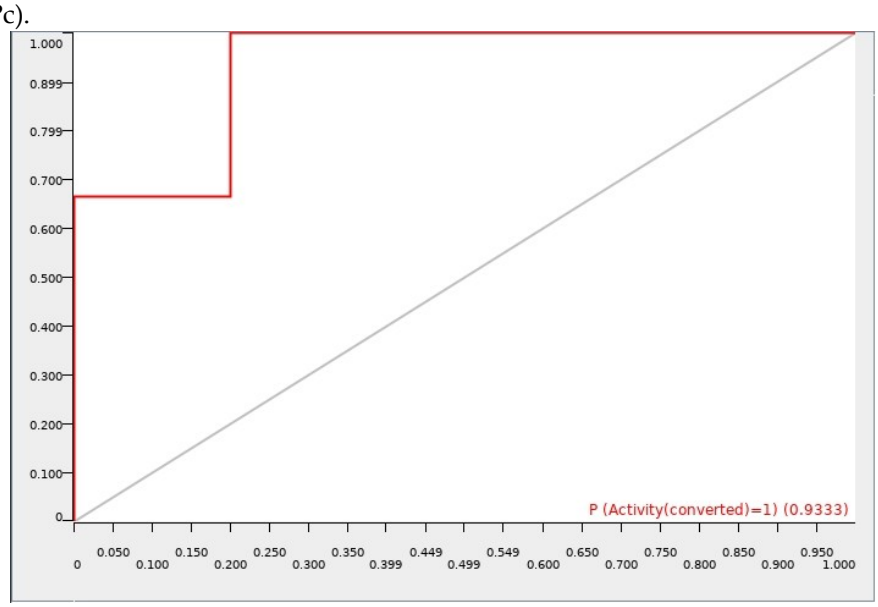
Figure S26. ROC curve obtained for the test set of FabG database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

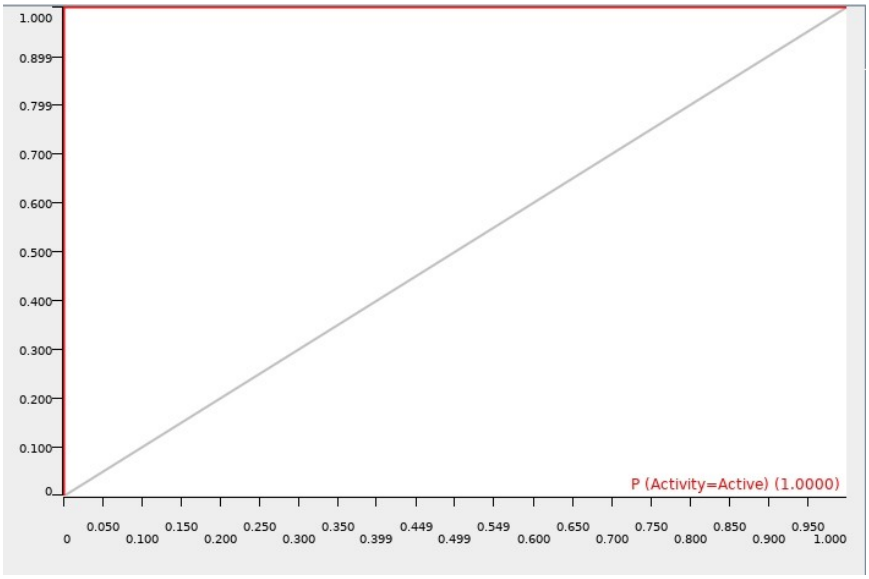
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S27. ROC curve obtained for the cross-validation set of FabG database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

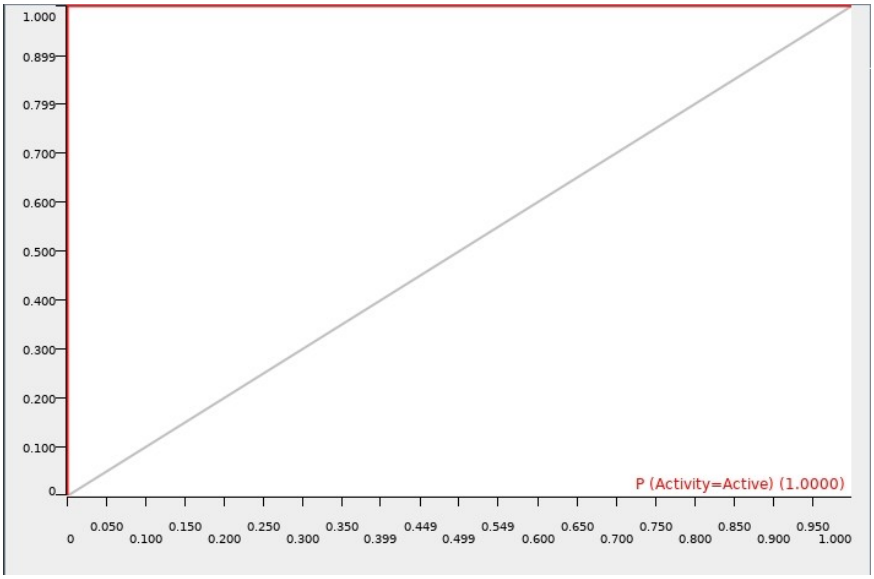
Figure S28. ROC curve obtained for the training set of FabG database for the classificatory k-nearest neighbour (KNN).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

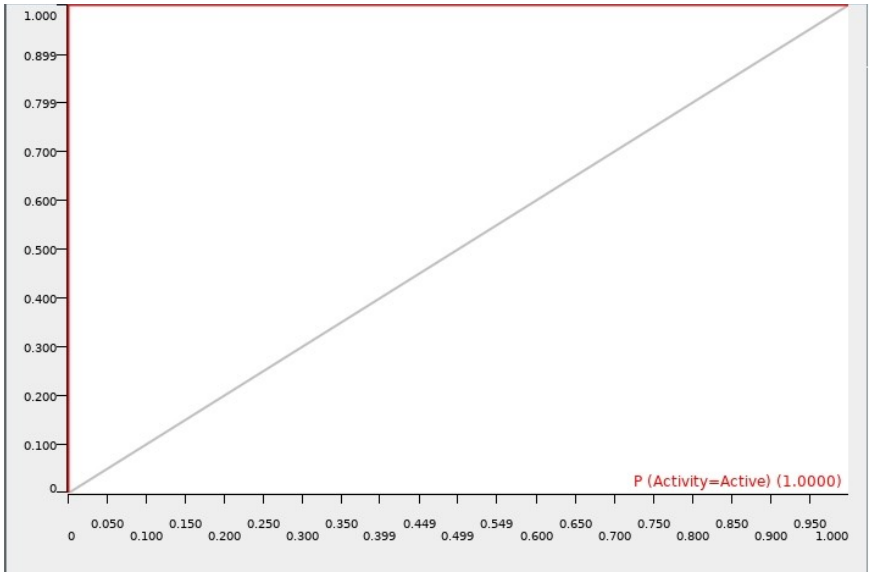
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S29. ROC curve obtained for the test set of FabG database for the classificatory k-nearest neighbour (KNN).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

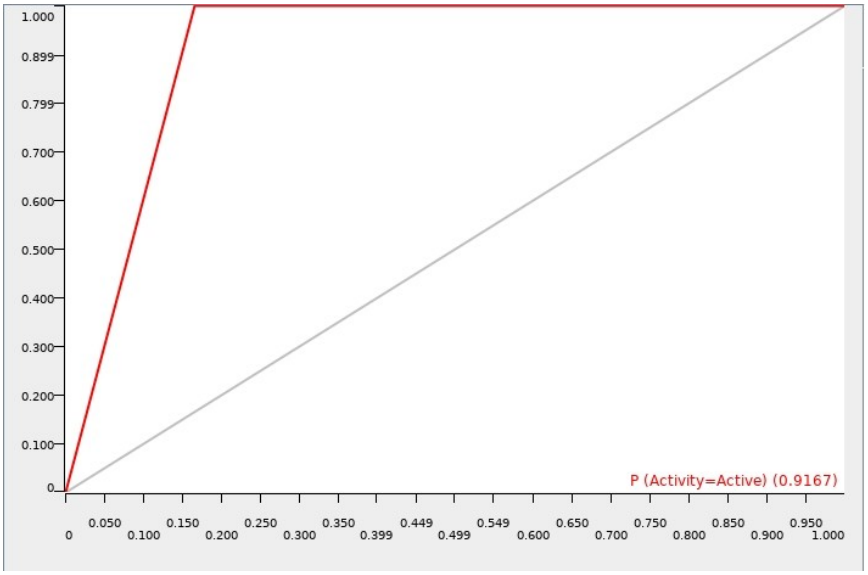
Figure S30. ROC curve obtained for the cross-validation set of FabG database for the classificatory k-nearest neighbour (KNN).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

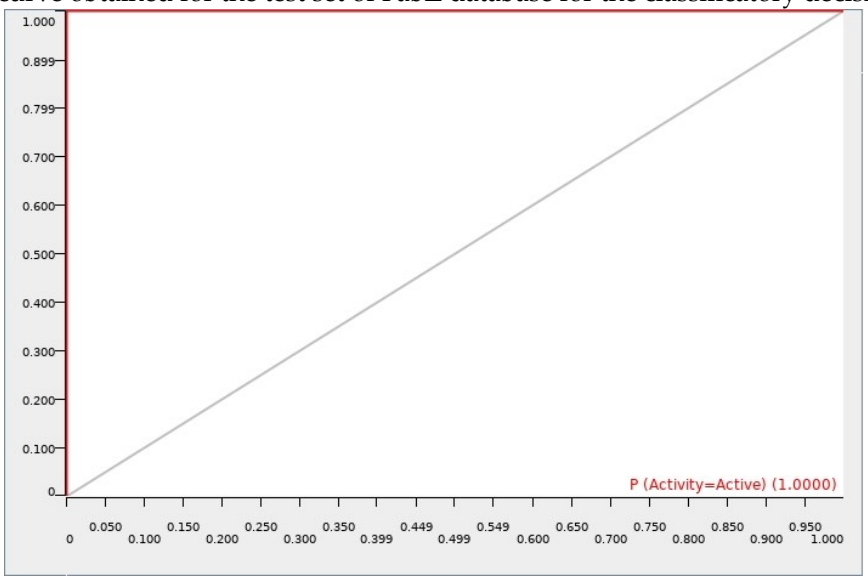
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S31. ROC curve obtained for the training set of FabZ database for the classificatory decision tree (DTc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

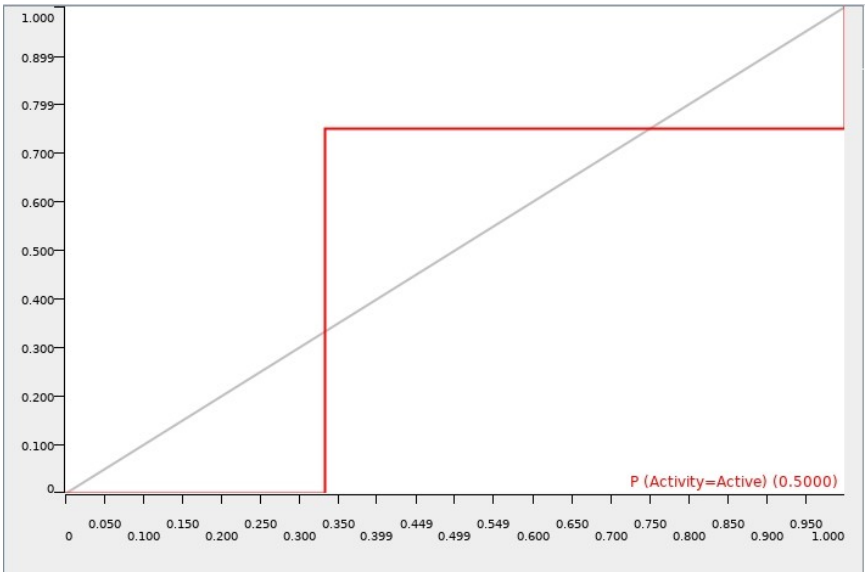
Figure S32. ROC curve obtained for the test set of FabZ database for the classificatory decision tree (DTc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

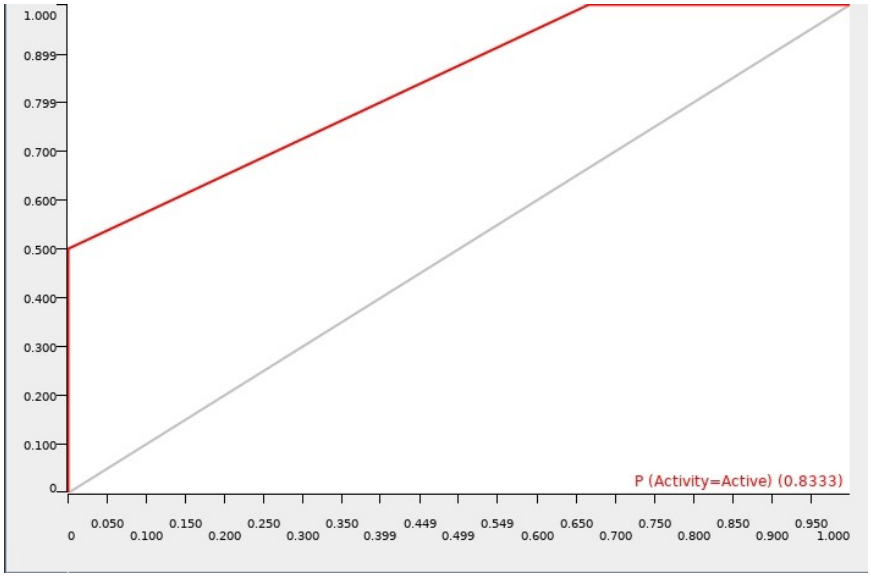
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S33. ROC curve obtained for the cross-validation set of FabZ database for the classificatory decision tree (DTc).



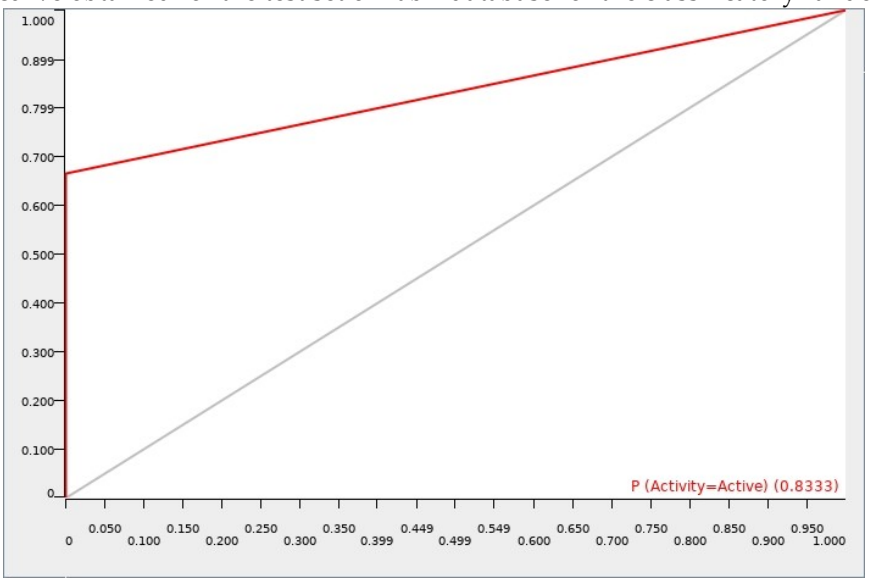
The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

Figure S34. ROC curve obtained for the training set of FabZ database for the classificatory random forest (RFc).



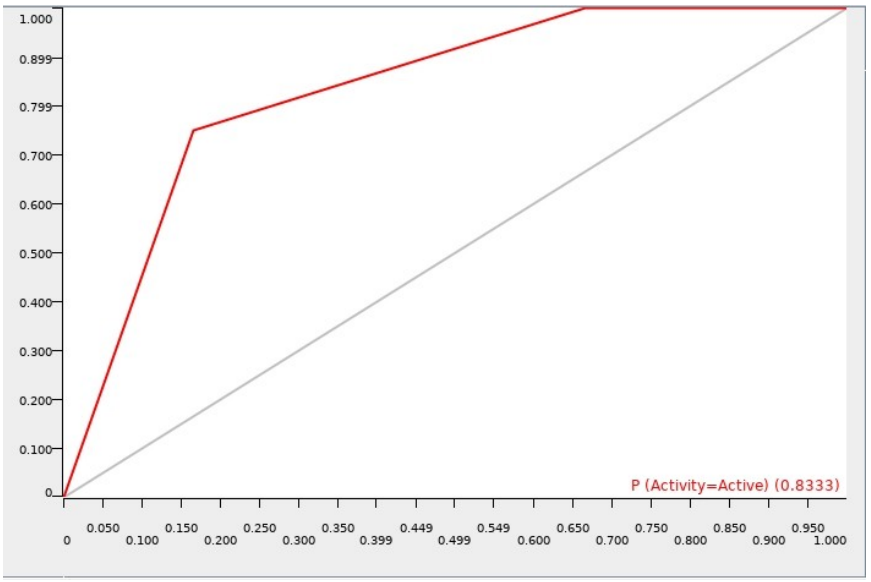
The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

Figure S35. ROC curve obtained for the test set of FabZ database for the classificatory random forest (RFc).



The y axis represents the true positive rates, while the x axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

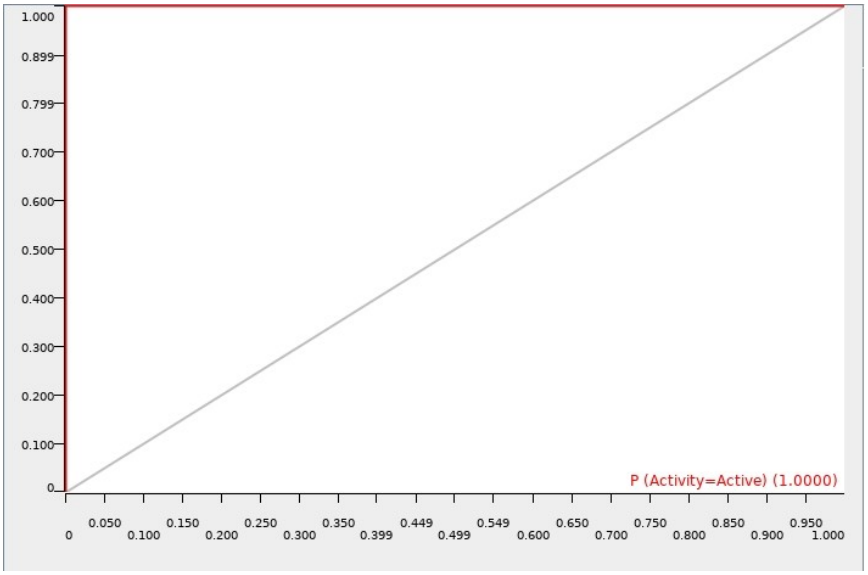
Figure S36. ROC curve obtained for the cross-validation set of FabZ database for the classificatory random forest (RFc).



The y axis represents the true positive rates, while the x axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

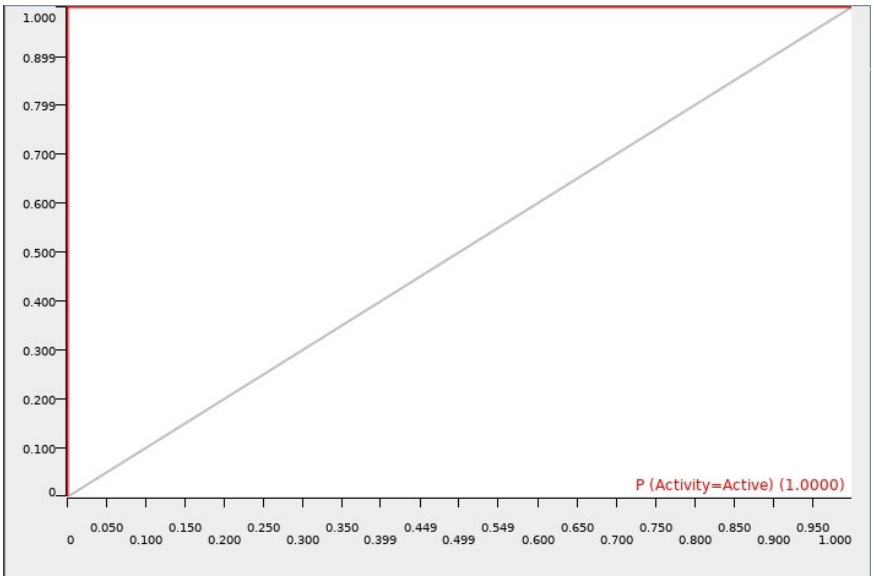
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S37. ROC curve obtained for the training set of FabZ database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

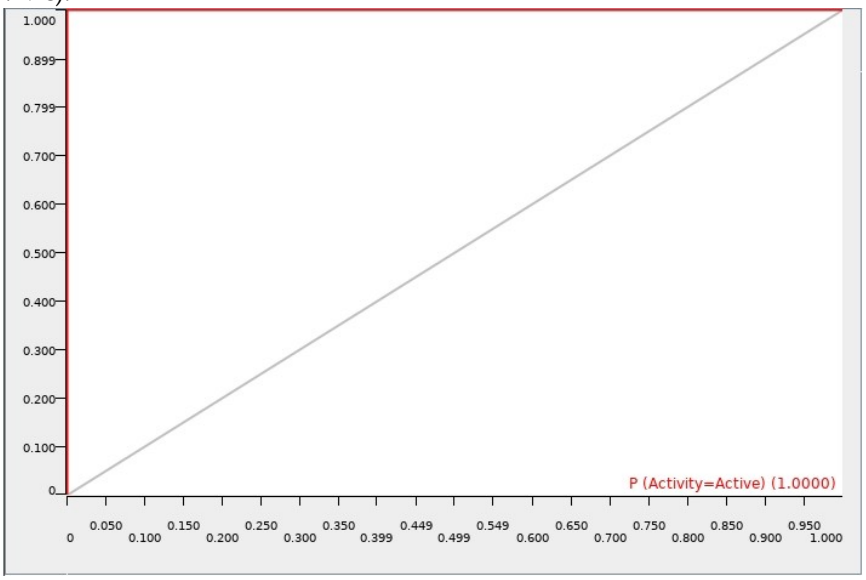
Figure S38. ROC curve obtained for the test set of FabZ database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

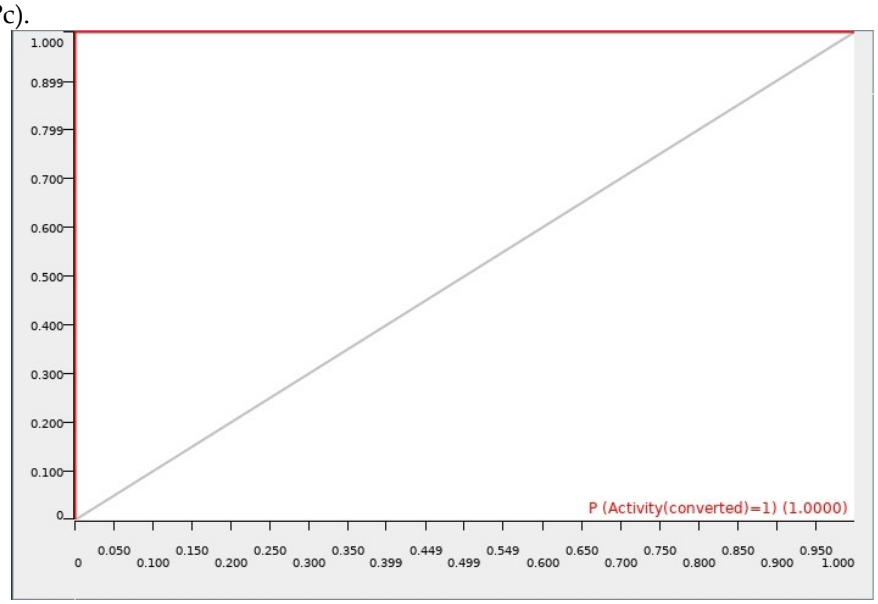
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S39. ROC curve obtained for the cross-validation set of FabZ database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

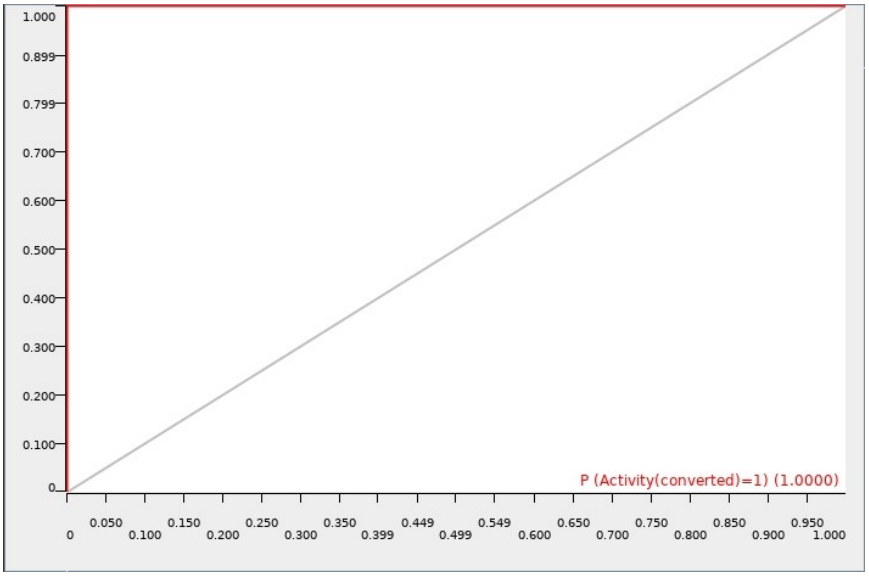
Figure S40. ROC curve obtained for the training set of FabZ database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

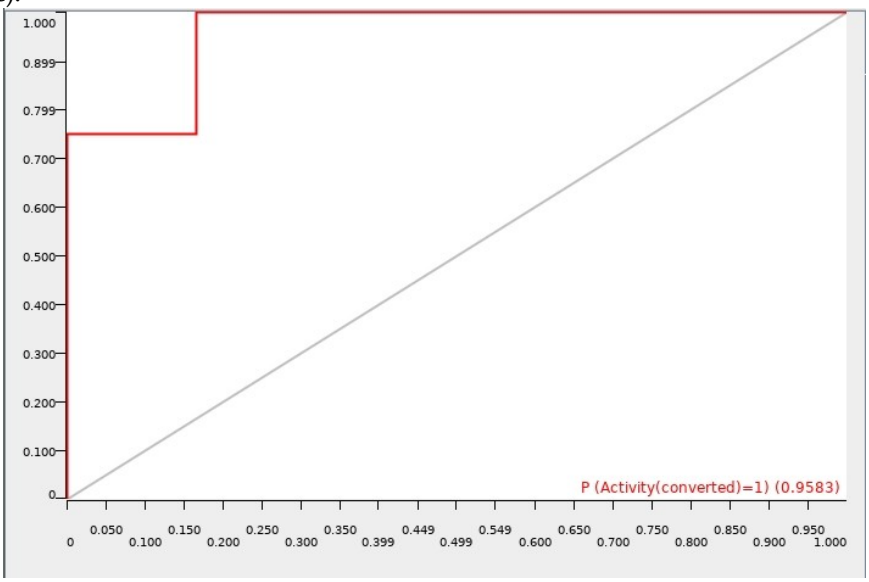
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S41. ROC curve obtained for the test set of FabZ database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

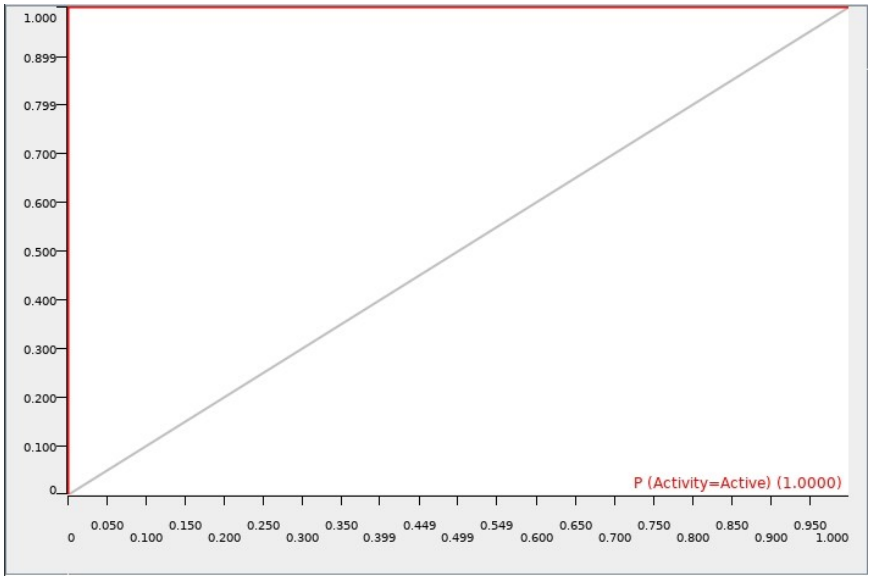
Figure S42. ROC curve obtained for the cross-validation set of FabZ database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

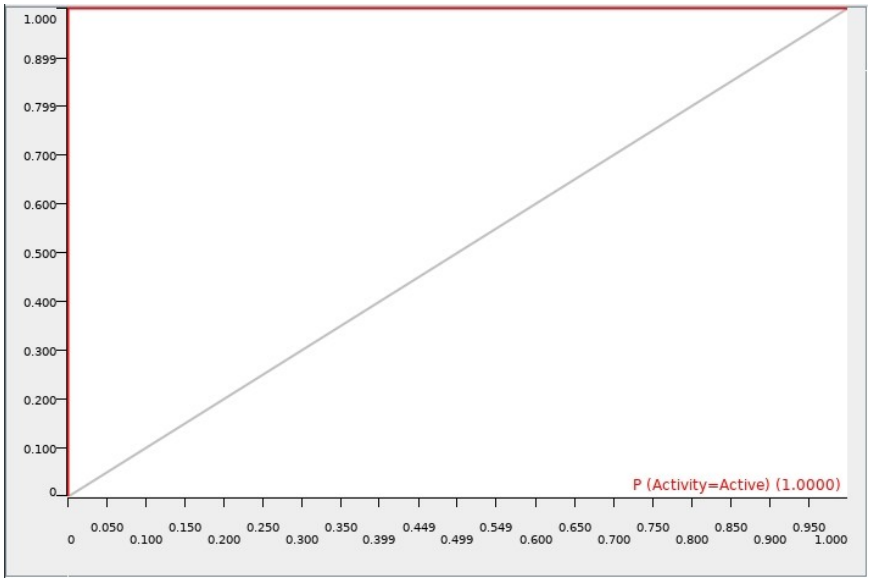
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S43. ROC curve obtained for the training set of FabZ database for the classificatory k-nearest neighbour (KNN).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

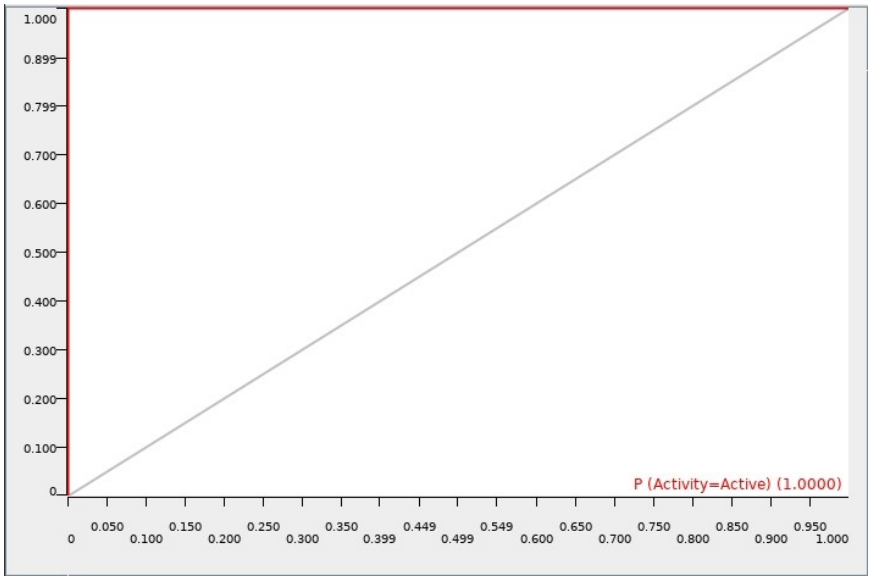
Figure S44. ROC curve obtained for the test set of FabZ database for the classificatory k-nearest neighbour (KNN).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

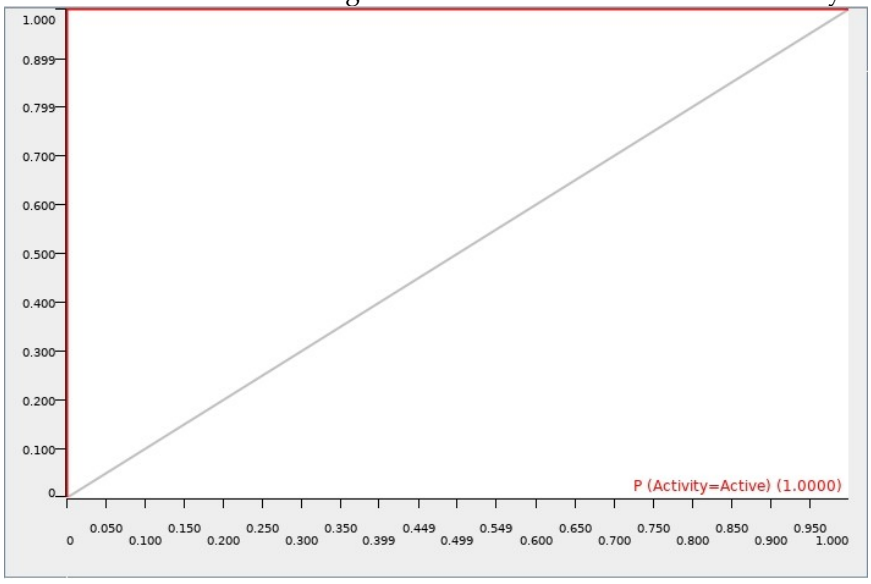
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S45. ROC curve obtained for the cross-validation set of FabZ database for the classificatory k-nearest neighbour (KNN).



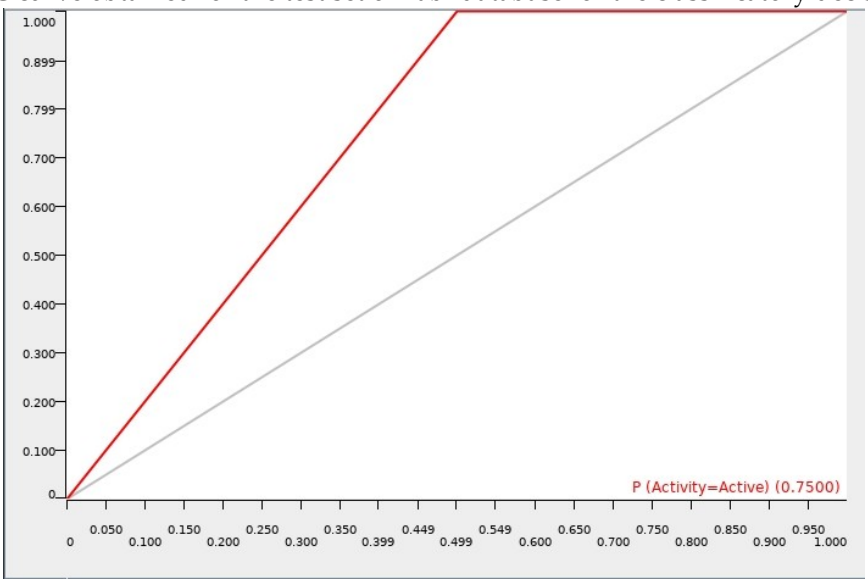
The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

Figure S46. ROC curve obtained for the training set of FabI database for the classificatory decision tree (DTc).



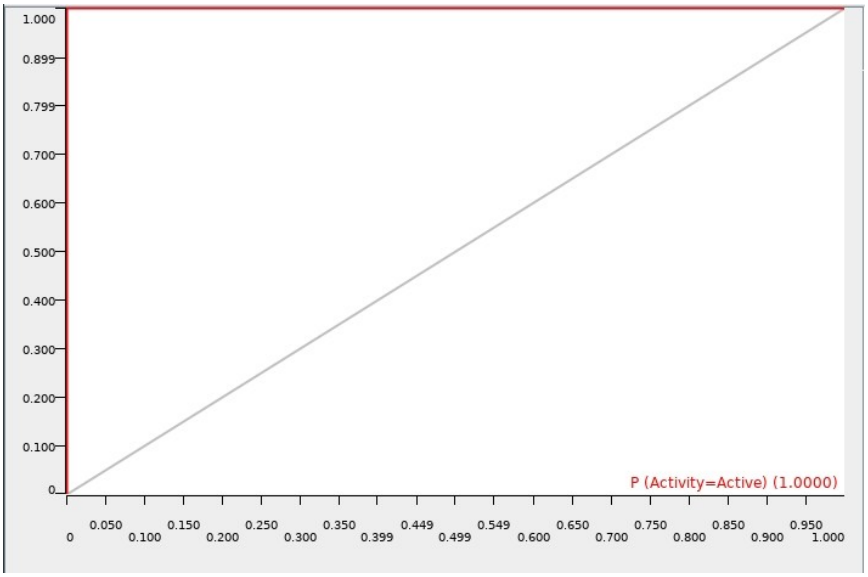
The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

Figure S2.47. ROC curve obtained for the test set of FabI database for the classificatory decision tree (DTc).



The y axis represents the true positive rates, while the x axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

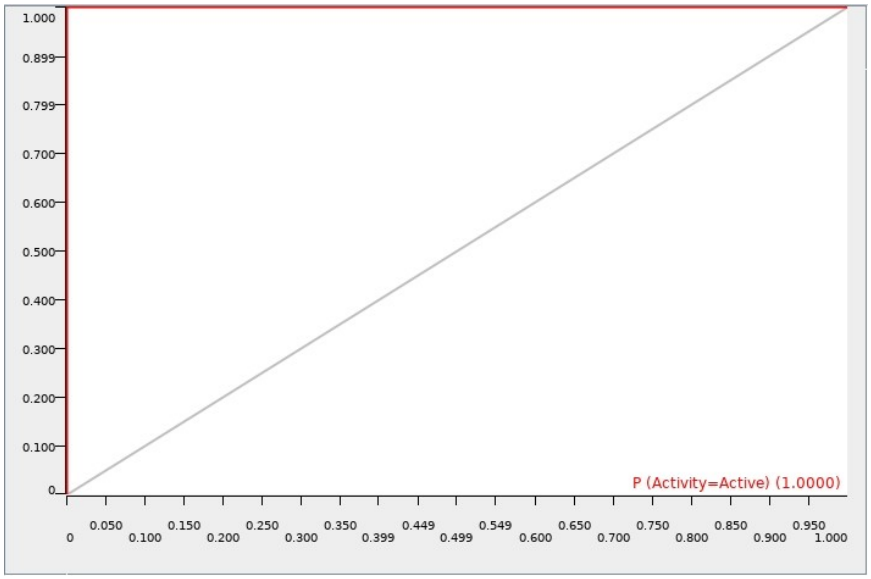
Figure S48. ROC curve obtained for the cross-validation set of FabI database for the classificatory decision tree (DTc).



The y axis represents the true positive rates, while the x axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

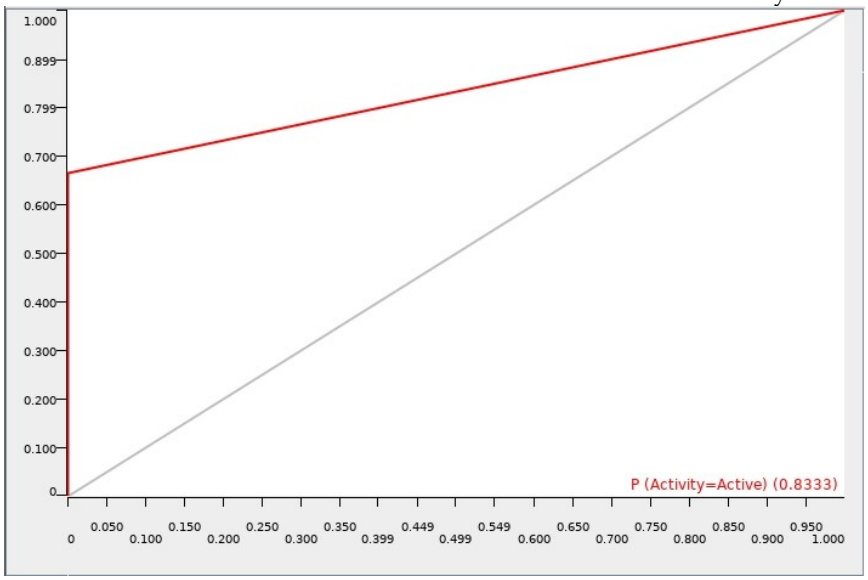
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S49. ROC curve obtained for the training set of FabI database for the classificatory random forest (RFc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

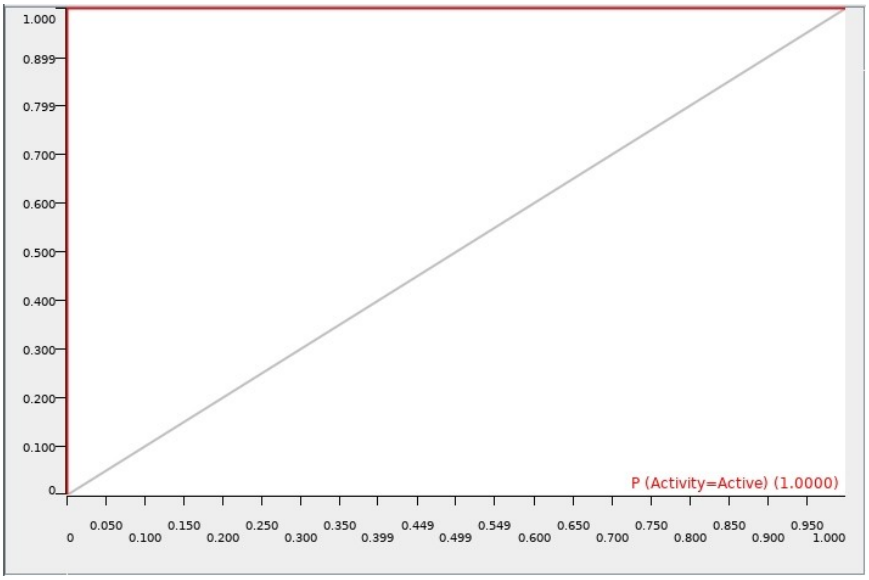
Figure S50. ROC curve obtained for the test set of FabI database for the classificatory random forest (RFc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

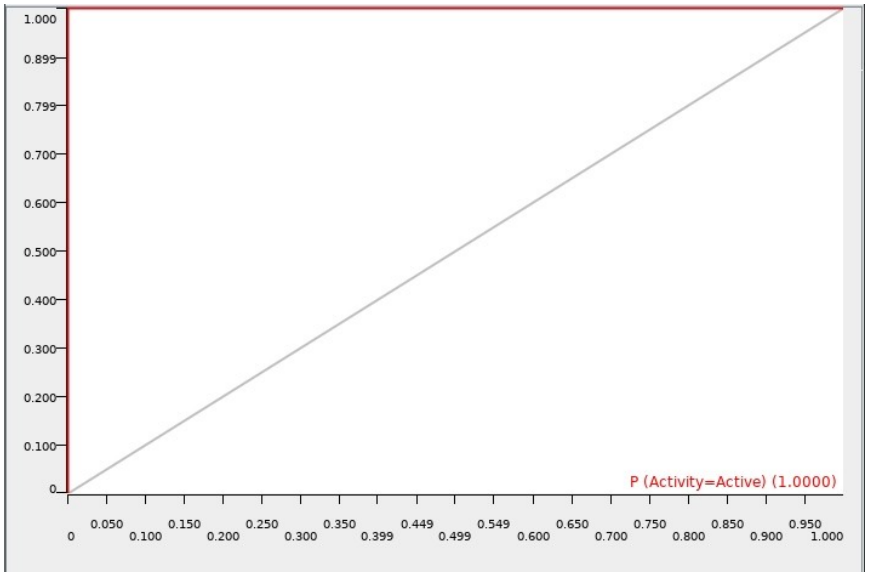
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S51. ROC curve obtained for the cross-validation set of FabI database for the classificatory random forest (RFc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

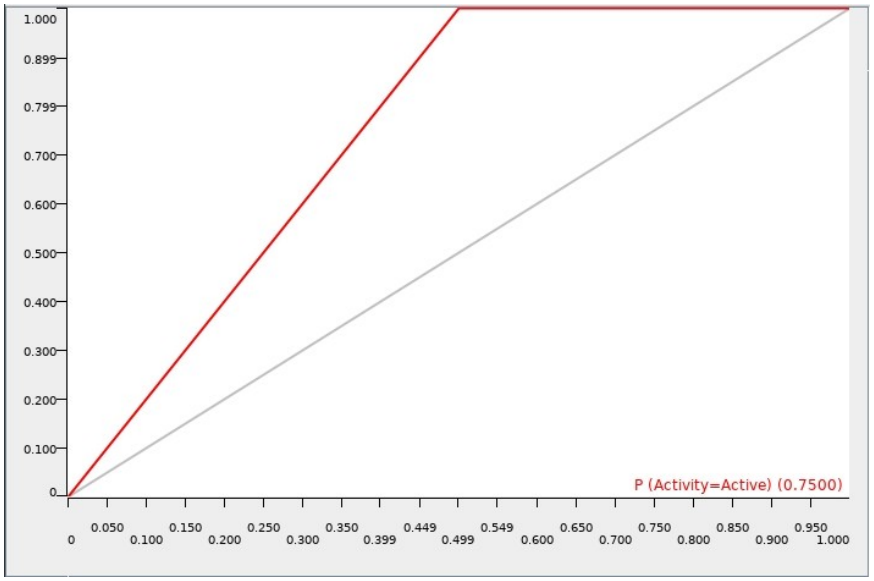
Figure S52. ROC curve obtained for the training set of FabI database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

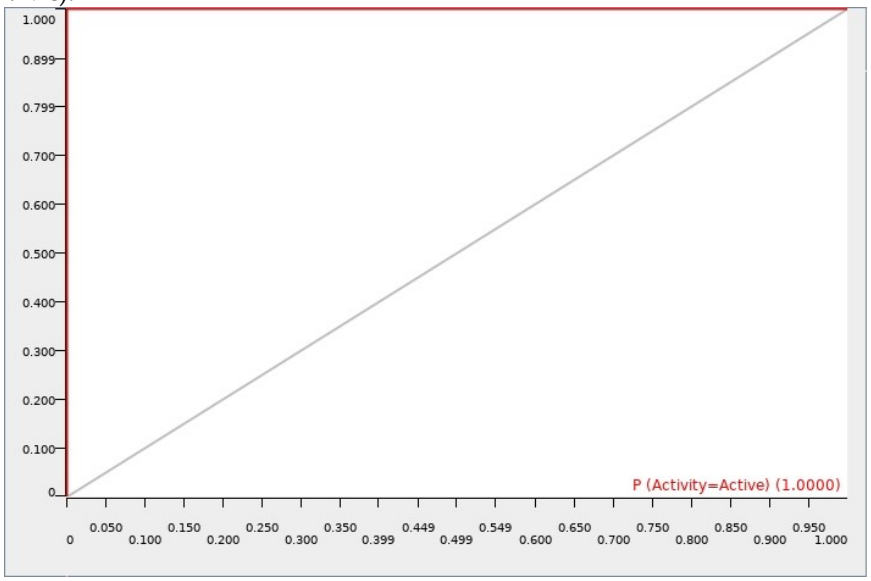
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S53. ROC curve obtained for the test set of FabI database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

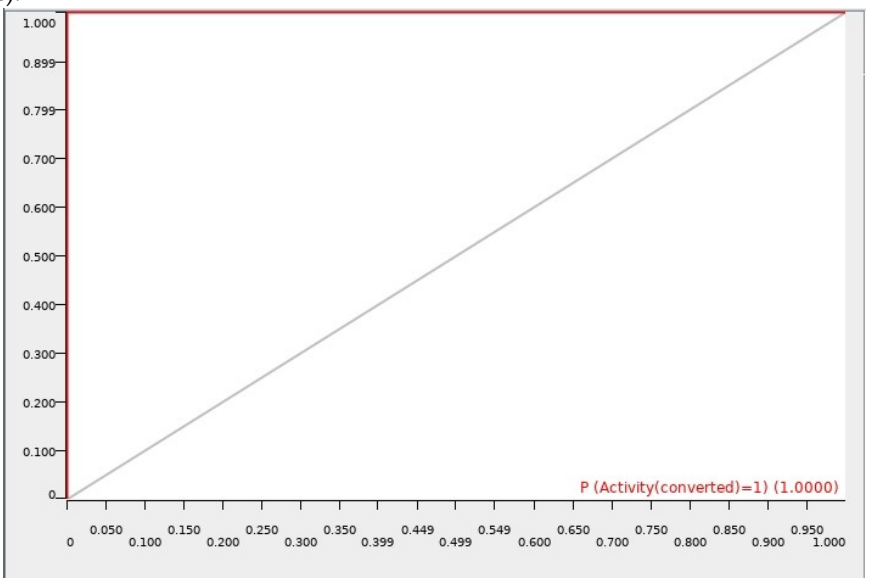
Figure S54. ROC curve obtained for the cross-validation set of FabI database for the classificatory support vector machine (SVMc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

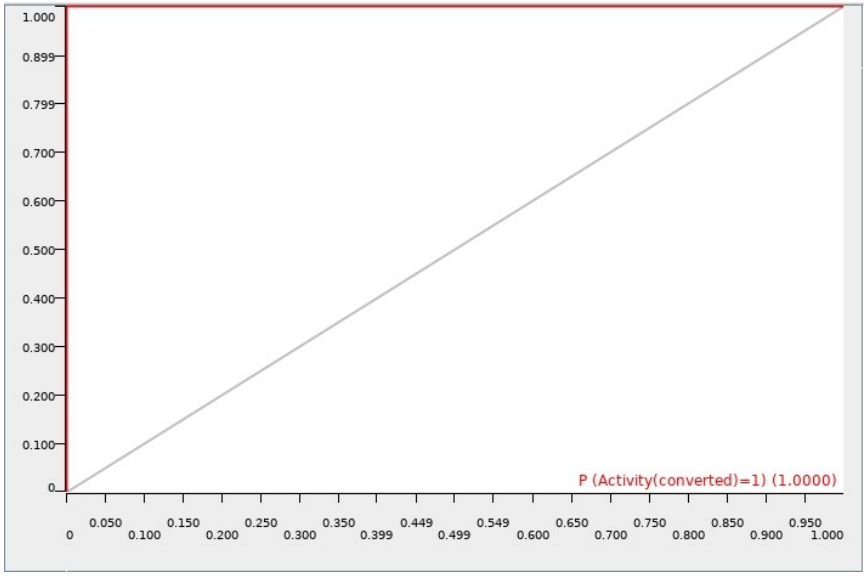
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S55. ROC curve obtained for the training set of FabI database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

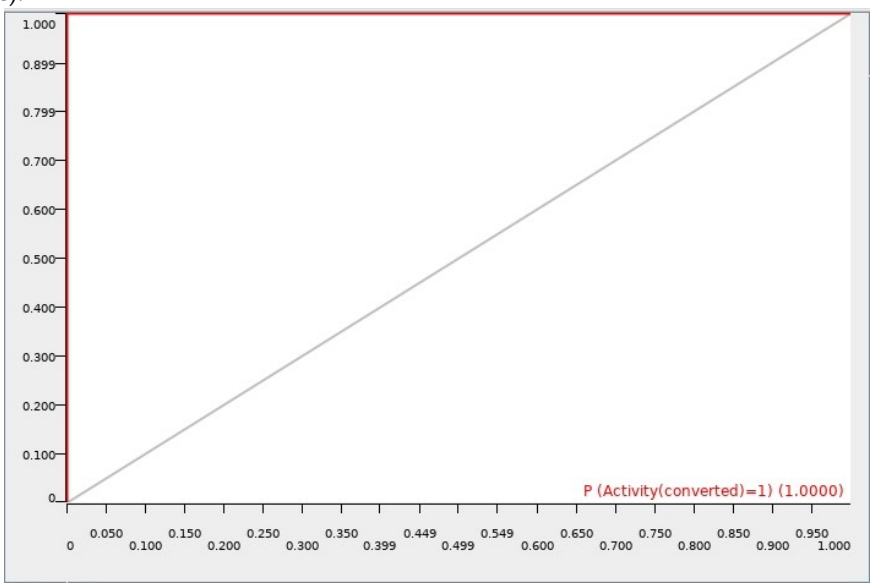
Figure S56. ROC curve obtained for the test set of FabI database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

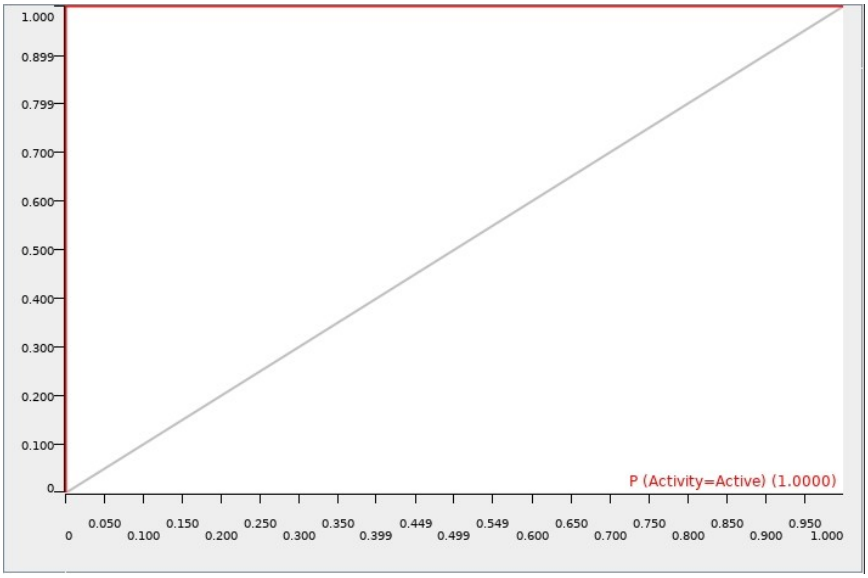
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S57. ROC curve obtained for the cross-validation set of FabI database for the classificatory multilayer perceptron (MLPc).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

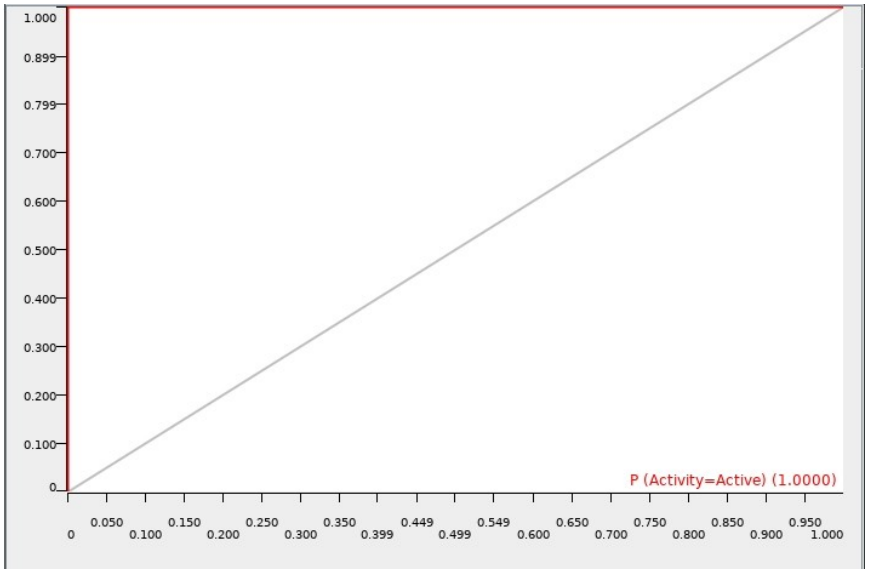
Figure S58. ROC curve obtained for the training set of FabI database for the classificatory k-nearest neighbour (KNN).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

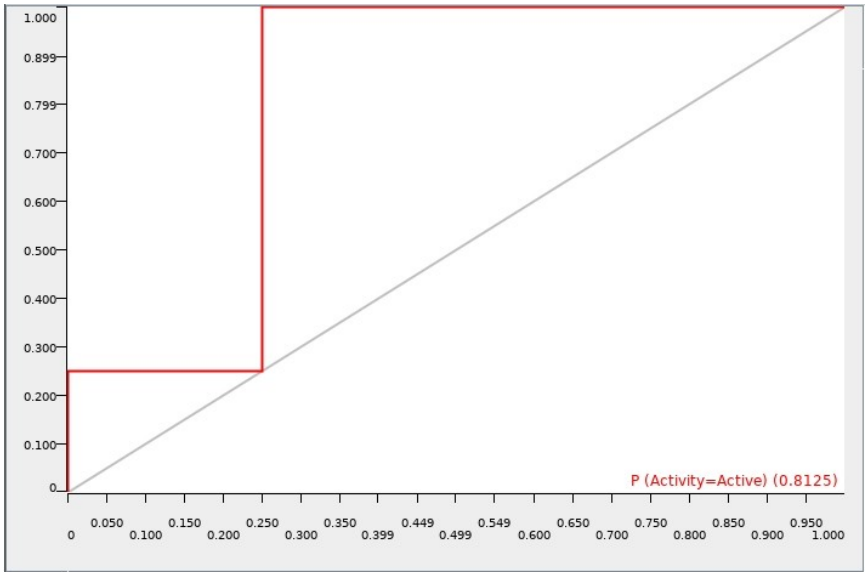
Identification of Pharmacophore Groups with Antimalarial Potential in Flavonoids by QSAR-Based Virtual Screening - ROC curves

Figure S59. ROC curve obtained for the test set of FabI database for the classificatory k-nearest neighbour (KNN).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.

Figure S60. ROC curve obtained for the cross-validation set of FabI database for the classificatory k-nearest neighbour (KNN).



The *y* axis represents the true positive rates, while the *x* axis represents the false positive rates. The red line represents the ROC curve, while the light gray line represents the worst possible performance the model can achieve.