**Supplementary materials**

*Journal of Computer-Aided Molecular Design*

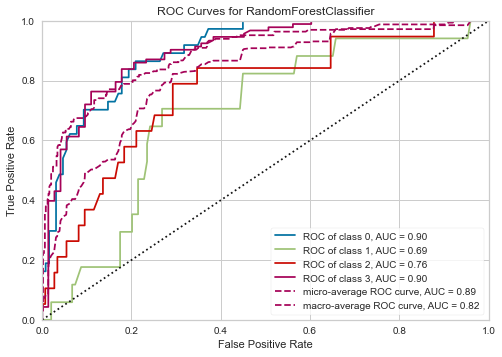
**Advancing the development of Deep Learning and Machine Learning models for oral drugs through diverse descriptor classes: A focus on Pharmacokinetic Parameters (Vdss and PPB)**

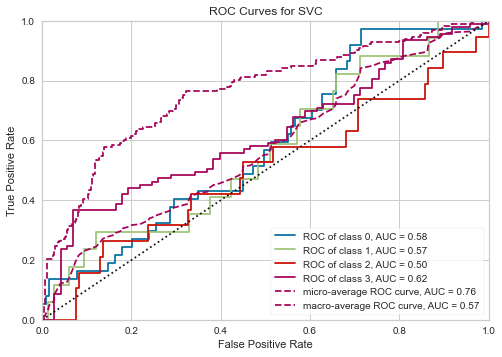
Rakesh Bantu, Samiron Phukan\*, Simon Haydar

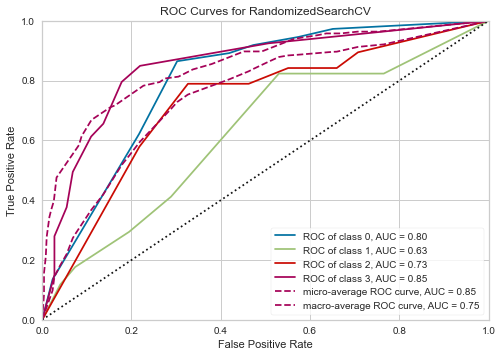
Integrated Drug Discovery,

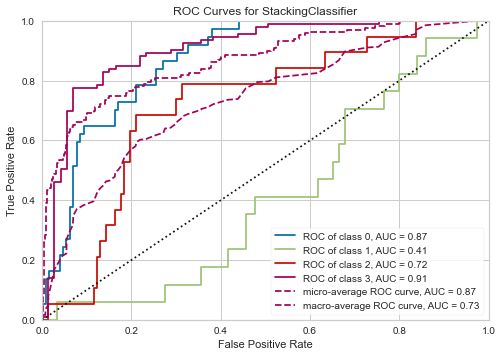
Aragen Lifesciences Ltd, Hyderabad -500076, India

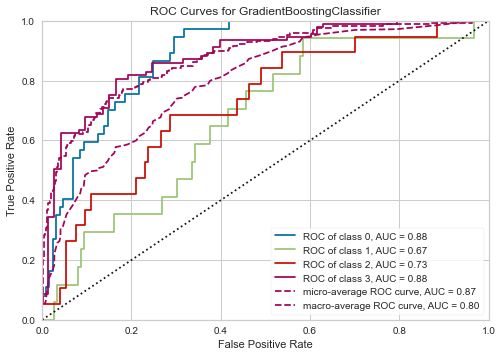
**Figures S(a)I to S(a)X: Receiver operating characteristic (ROC) curves for shared descriptors for PPB prediction (Vdss** ∩ **PPB):**

Fig S(a)I: SVC Fig S(a)II: Random Forest



Fig S(a)III: Decision Tree Classifier Fig S(a)IV: Stacking Classifier



 Fig S(a)V: Gradient Boosting Classifier Fig S(a) VI: Sci-Kit-Learn-Multilayer Perceptron

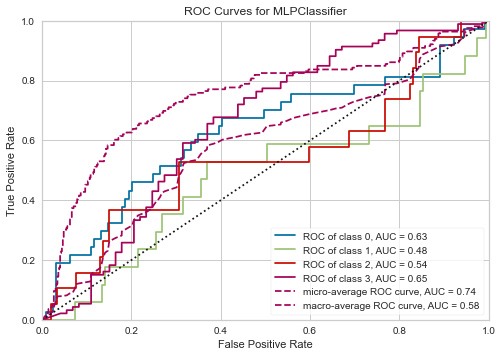
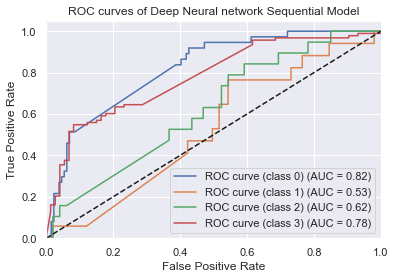
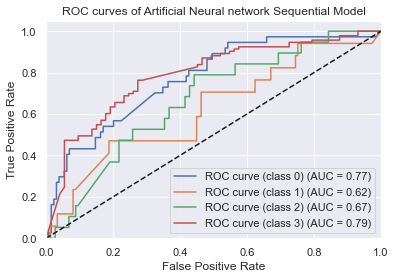
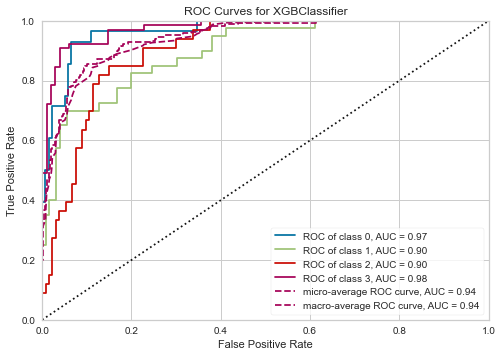
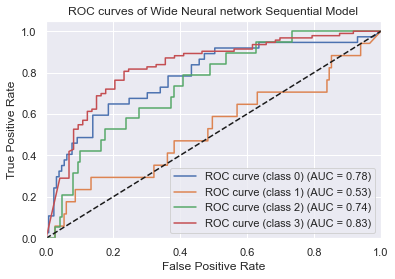


Fig S(a)VII: Artificial neural networks (ANN~~s~~). Fig S(a)VIII: Deep neural networks (DNNs).

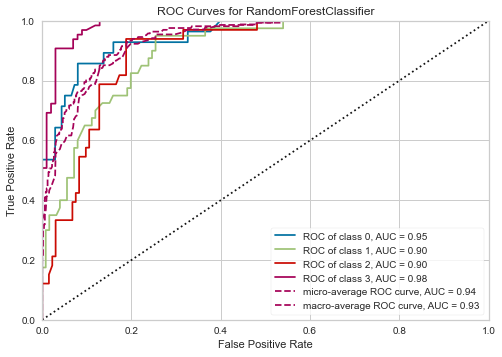


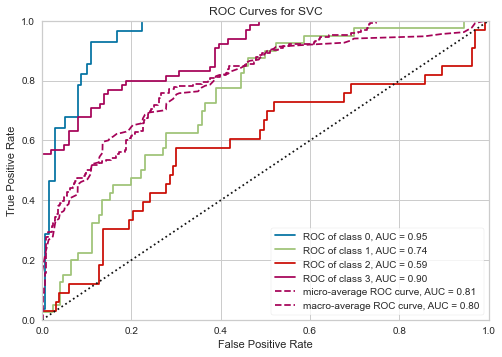


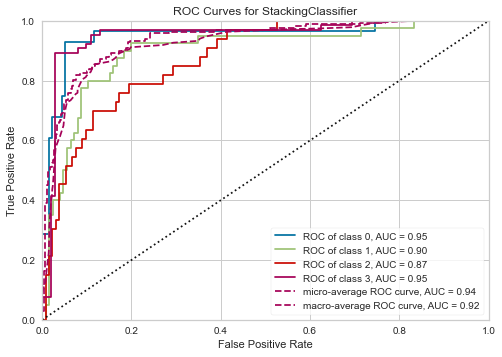
 Fig S(a)IX: Wide Neural networks (WNNs). Fig S(a)X: XG Boost

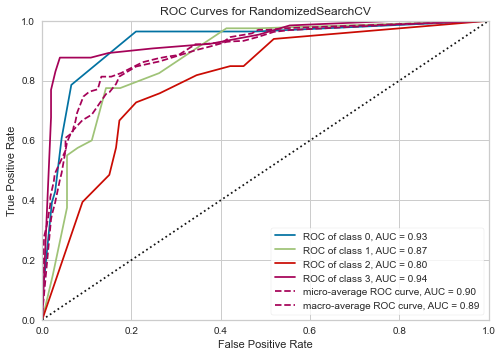


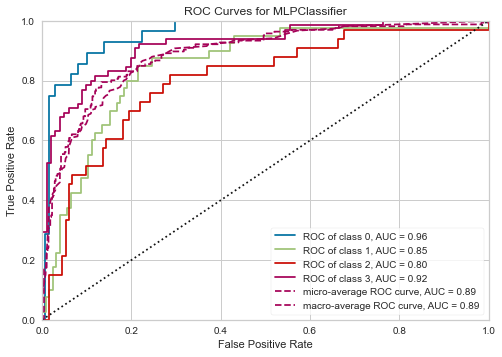
**Figures- S(b)I to S(b)X: Receiver operating characteristic (ROC) curves for Shared Descriptors for Vdss prediction (Vdss** ∩ **PPB):**

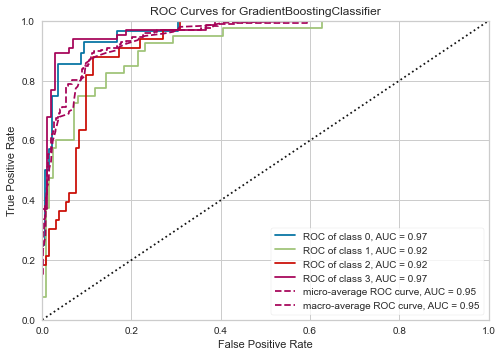
Fig S(b)I: SVC Fig S(b)II: Random Forest

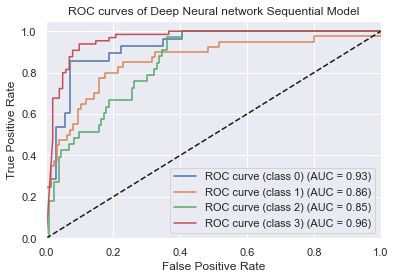


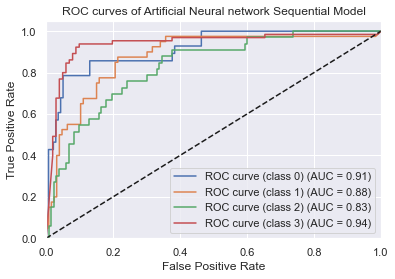
Fig S(b)III: Decision Tree Classifier Fig S(b)IV: Stacking Classifier

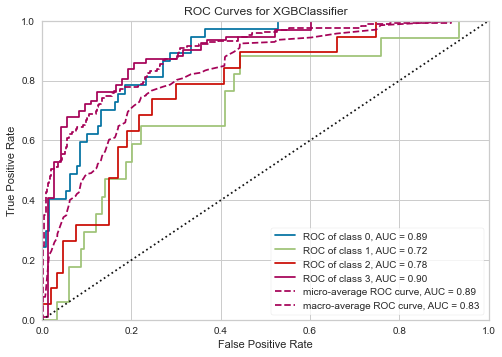


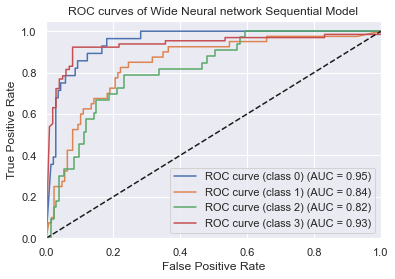
Fig S(b)V: Gradient Boosting Classifier Fig S(b)VI: Sci-Kit-Learn-Multilayer Perceptron



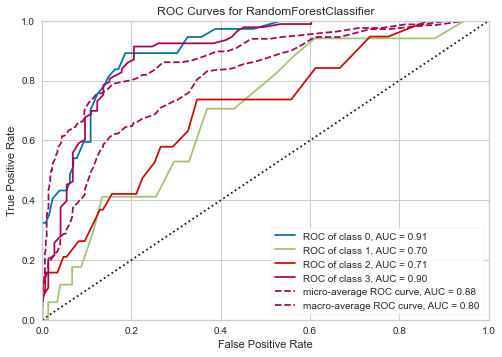
Fig S(b)VII: Artificial neural networks (ANN): Fig S(b)VIII: Deep neural networks (DNN):

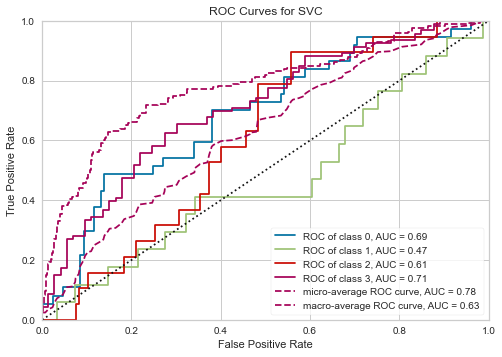


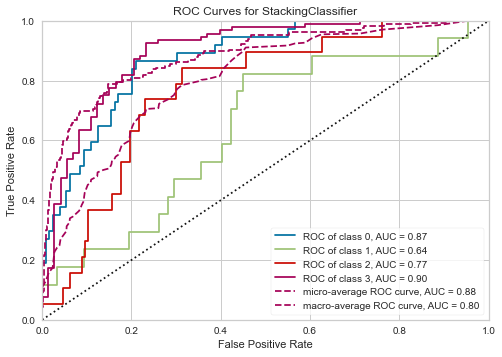
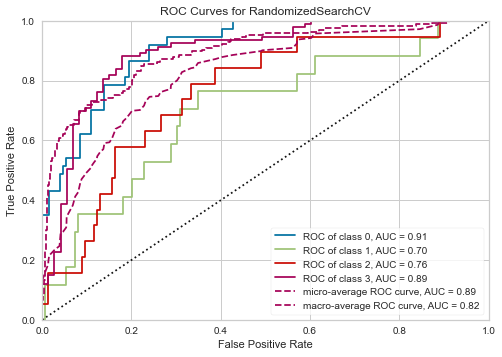
Fig S(b)IX: Wide neural networks (WNN): Fig S(b)X: XG Boost

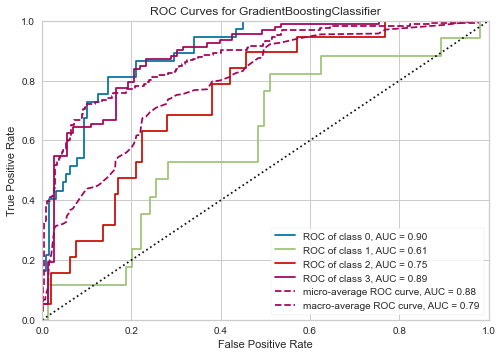
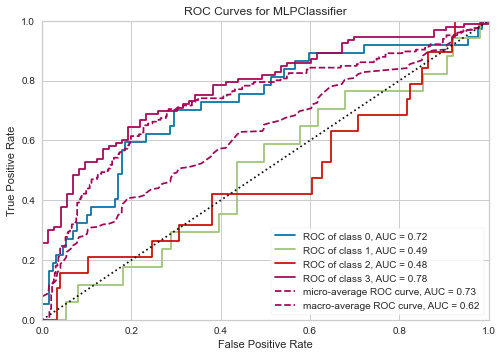


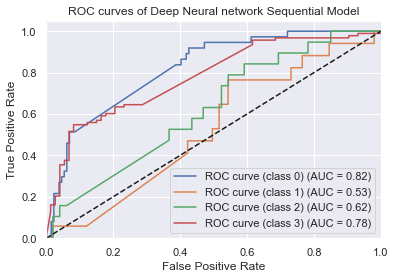
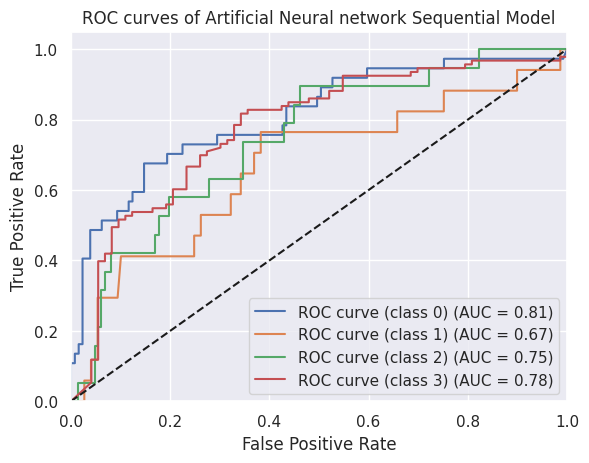
**Figure-S(c)I-S(c)X: Receiver operating characteristic (ROC) curves for PPB Features**

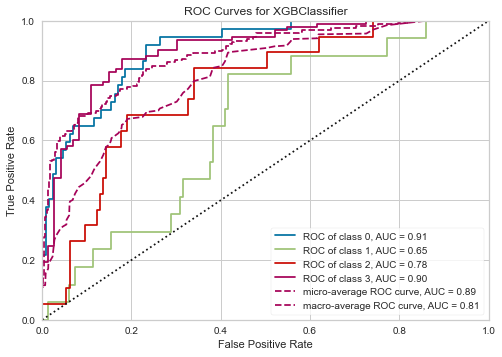
 Fig S(c)I: SVC: Fig S(c)II: Random Forest:

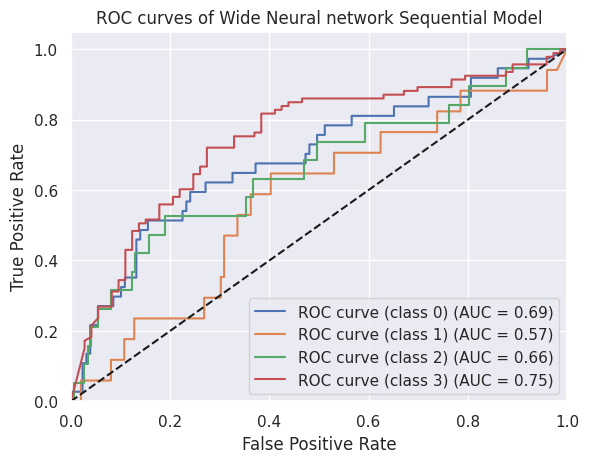


Fig S(c)III: Decision Tree: Fig S(c)IV: Stacking

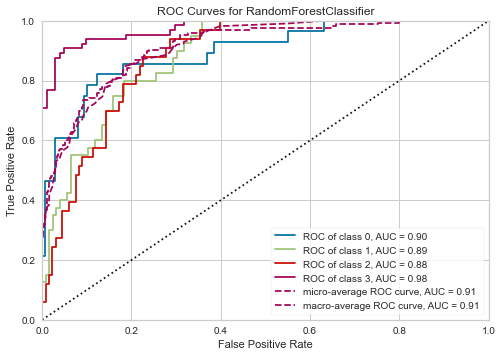
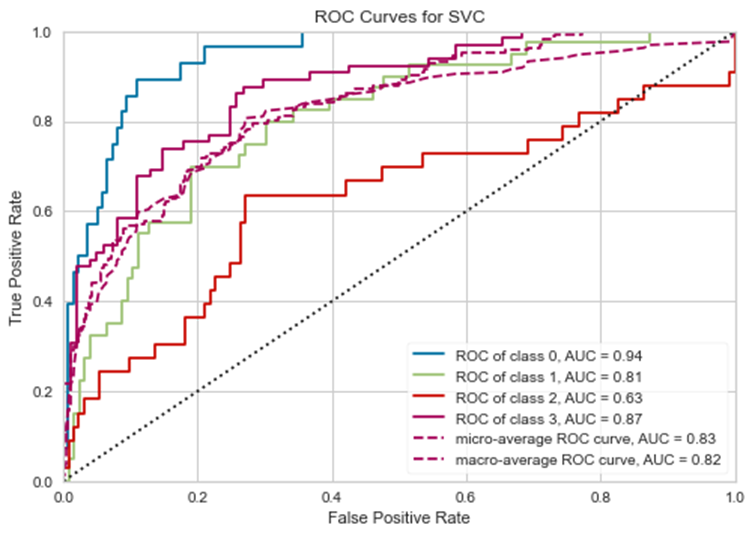
S(c)V: Gradient Boosting Fig S(c)VI: multi-layer perceptron

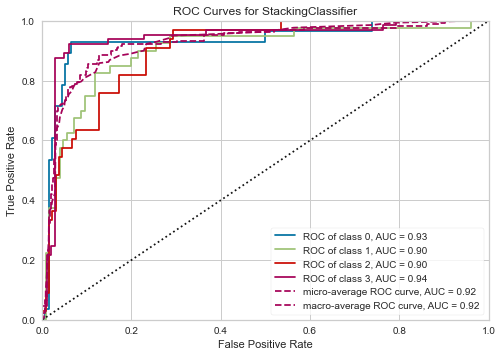
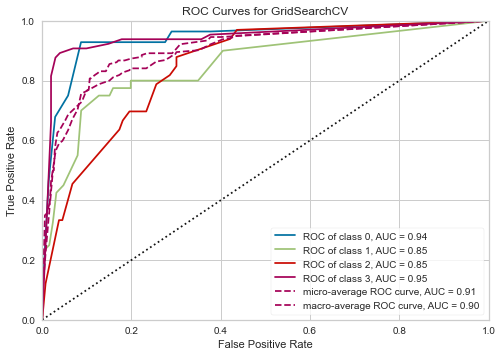
Fig S(c)VII: Artificial neural networks (Ann). Fig S(c)VIII: Deep neural networks (DNN)****

****Fig S(c)IX: Wide neural networks (WNN). Fig S(c)X: XG Boost

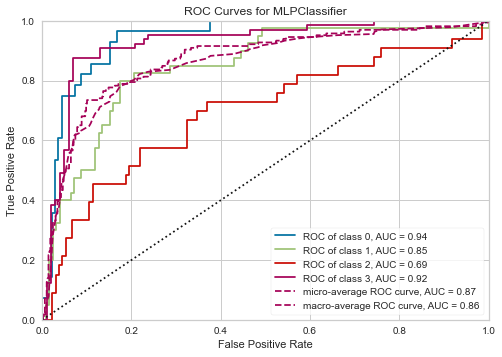
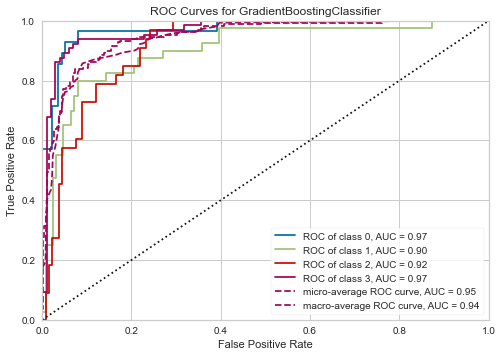
****

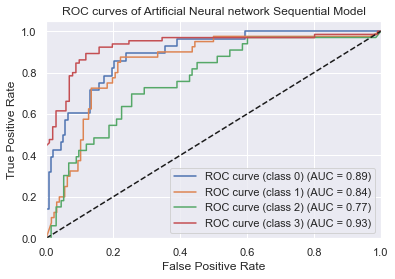
**Figure-S(d)I-S(d)X: Receiver operating characteristic (ROC) curves for Vdss Features:**

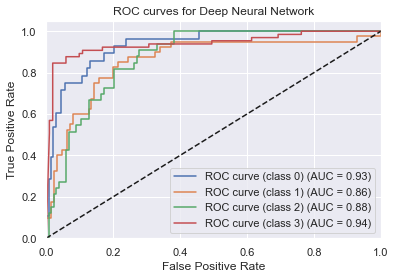
Fig S(d)I: SVC **** Fig S(d)II: Random Forest:

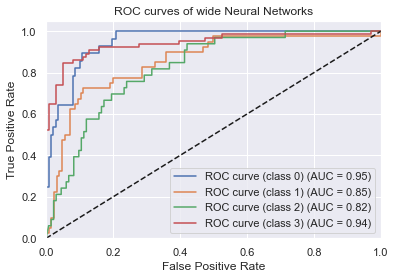
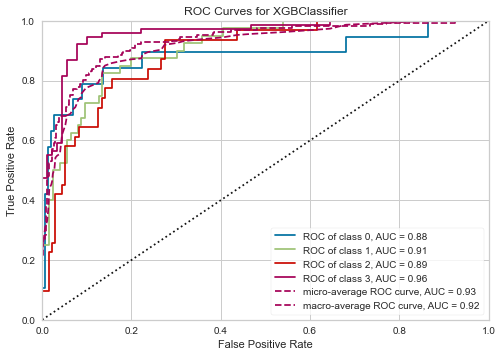
Fig S(d)III: Decision Tree: Fig S(d)IV: Stacking

S(d)V: Gradient Boosting S(d)VI: Multi-Layer Perceptron



Fig S(d)VII: Artificial Neural Network Fig S(d)VIII: Deep Neural Network



Fig S(d)IX: wide Neural Network Fig S(d)X: XG Boost

**Table S3:** The table illustrates the Boruta Selected Descriptors for Vdss and PBB and the shared descriptors for both Vdss and PPB.

|  |  |  |  |
| --- | --- | --- | --- |
| Vdss Descriptors | r\_qp\_PSA | MLFER\_E  PPB Descriptors | MLFER\_E  Shared Descriptors (Vdss ∩ PPB) |
|  | MLFER\_BO | MLFER\_L | piPC5 |
|  | MLFER\_E | piPC1 | piPC6 |
|  | piPC9 | piPC2 | piPC9 |
|  | nHdsCH | piPC3 | MLFER\_BO |
|  | piPC8 | piPC4 | MLFER\_BH |
|  | minHdsCH | piPC5 | CrippenLogP |
|  | AATSC0i | piPC6 | TpiPC |
|  | XLogP | C2SP2 | TopoPSA |
|  | minssNH | SpMin2\_Bhs |  |
|  | ETA\_dEpsilon\_D | piPC9 |  |
|  | minHBd | piPC10 |  |
|  | piPC7 | R\_TpiPCTPC |  |
|  | AATSC0s | SpMax\_D |  |
|  | AATSC2e | SpDiam\_D |  |
|  | TpiPC | MLFER\_BO |  |
|  | ETA\_Epsilon\_4 | MLFER\_BH |
|  | AATS4s | MDEO-12 |
|  | ATSC2e | MDEC-23 |
|  | ATSC0i | ETA\_Beta |
|  | MDEO-11 | SHother |
|  | AATS1i | r\_qp\_PSA |
|  | VE1\_D | nwHBa |
|  | VE2\_D | ETA\_Eta\_F |
|  | maxssNH | ETA\_Eta\_R\_L |
|  | maxsssN | ETA\_Eta\_F\_L |
|  | VE1\_Dze | SpAD\_D |
|  | AATSC1c | ECCEN |
|  | AATS4e | IC5 |
|  | AATS1e | CrippenLogP |
|  | VE3\_D | ZMIC2 |
|  | maxHBd | Kier1 |
|  | TopoPSA | Kier2 |
|  | BIC2 | nAtomP |
|  | VE1\_Dzv | SP-1 |
|  | MLFER\_BH | CrippenMR |
|  | LogP | SpMAD\_D |
|  | piPC6 | TpiPC |
|  | CrippenLogP | TopoPSA |
|  | nBondsD2 | SpMax4\_Bhv |
|  | r\_qp\_QPlogS | SpMin2\_Bhm |
|  | r\_qp\_QPlogPo/w | r\_qp\_volume |
|  | r\_qp\_QPlogBB | EE\_D |
|  | nBondsD | nO |
|  | SHdsCH | SpAbs\_Dze |
|  | BCUTc-1h | r\_qp\_QPpolrz |
|  | r\_qp\_IP (eV) | r\_qp\_QPlogPC16 |
|  | IC2 | r\_qp\_QPlogPo/w |
|  | IC1 | r\_qp\_QPlogS |
|  | VE1\_Dzi | r\_qp\_CIQPlogS |
|  | r\_qp\_QPlogHERG | VE3\_DzZ |
|  | r\_qp\_QPlogKhsa | r\_qp\_QPlogKp |
|  | SHBd | r\_qp\_QPlogKhsa |
|  | MATS2e | r\_qp\_PercentHumanOralAbsorption |
|  | TPSA | r\_qp\_SASA |
|  | SdssC | SpMin2\_Bhv |
|  | GATS2e | r\_qp\_accptHB |
|  | SIC2 | SpMin2\_Bhe |
|  | piPC5 | SpMin2\_Bhi |
|  | nBase | XLogP |
|  | i\_qp\_#amine | WPATH |
|  | SssNH | TPSA |
|  | SdsCH | SpMax4\_Bhi |
|  | i\_qp\_CNS | LogP |
|  | AMR | WTPT-4 |
|  | SssCH2 | SpMin2\_Bhp |
|  | r\_qp\_FISA | AATS7i |