|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ML\_Set | overall Recall | overall Precision | overall Senstivity | overall Specificity | overall F-measure | Accuracy | Cohen's kappa |
| NB\_Training | 0.769466 | 0.646712 | 0.769466 | 0.845062 | 0.66898 | 0.678571 | 0.480172 |
| NB\_Test | 0.665598 | 0.575126 | 0.665598 | 0.806433 | 0.595093 | 0.627451 | 0.378448 |
| RF\_Training | 0.972789 | 0.989583 | 0.972789 | 0.981481 | 0.980525 | 0.979592 | 0.960424 |
| RF\_Test | 0.897436 | 0.962963 | 0.897436 | 0.929825 | 0.919786 | 0.921569 | 0.842593 |
| GB\_Training | 0.964212 | 0.992126 | 0.964212 | 0.986111 | 0.977428 | 0.984694 | 0.970352 |
| GB\_Test | 0.850962 | 0.885714 | 0.850962 | 0.894737 | 0.863725 | 0.862745 | 0.728517 |
| GL\_Training | 0.873372 | 0.772942 | 0.873372 | 0.925123 | 0.806424 | 0.836735 | 0.715207 |
| GL\_Test | 0.712073 | 0.622626 | 0.712073 | 0.84152 | 0.643865 | 0.686275 | 0.47013 |

Table 1Evaluation of Mechanistic interpretation model (classification)

|  |  |  |
| --- | --- | --- |
| Model | Mean(Accuracy) | SD (Accuracy) |
| NB | 0.584288 | 0.111398 |
| RF | 0.871384 | 0.074927 |
| GB | 0.875833 | 0.100975 |
| GL | 0.782621 | 0.106943 |

Table 2: 10 Fold CV of Mechanistic interpretation model

|  |  |  |  |
| --- | --- | --- | --- |
| Features | Relative Importance | Scaled Importance | Percentage |
| mtbca2\_pKi | 58.22038 | 1 | 0.112109 |
| ATSC8e | 36.3932 | 0.625094 | 0.070079 |
| mtbca3\_pKi | 35.78751 | 0.61469 | 0.068912 |
| mtbca1\_pKi | 35.07446 | 0.602443 | 0.067539 |
| GATS4c | 22.5575 | 0.38745 | 0.043437 |
| MATS4i | 21.20904 | 0.364289 | 0.04084 |
| BCUTp-1l | 19.00568 | 0.326444 | 0.036597 |
| ATSC5v | 18.00871 | 0.30932 | 0.034678 |
| ATS1v | 16.66892 | 0.286307 | 0.032098 |
| CrippenLogP | 16.00261 | 0.274863 | 0.030815 |
| SpMAD\_D | 15.84103 | 0.272087 | 0.030503 |
| PubchemFP696 | 15.43823 | 0.265169 | 0.029728 |
| n5HeteroRing | 14.94415 | 0.256682 | 0.028776 |
| GATS2m | 14.88371 | 0.255644 | 0.02866 |
| MATS3c | 14.72063 | 0.252843 | 0.028346 |
| AATSC2v | 14.46865 | 0.248515 | 0.027861 |
| ATSC8p | 11.63487 | 0.199842 | 0.022404 |
| ZMIC1 | 10.64654 | 0.182866 | 0.020501 |
| MATS1m | 10.33457 | 0.177508 | 0.0199 |
| nBondsD | 9.832024 | 0.168876 | 0.018933 |
| n6Ring | 9.23626 | 0.158643 | 0.017785 |
| nRotBt | 8.801653 | 0.151178 | 0.016948 |
| nBondsS | 8.683194 | 0.149144 | 0.01672 |
| piPC10 | 7.669606 | 0.131734 | 0.014769 |
| nHBDon\_Lipinski | 7.533685 | 0.129399 | 0.014507 |
| CIC1 | 6.746012 | 0.11587 | 0.01299 |
| PubchemFP643 | 6.69836 | 0.115052 | 0.012898 |
| ATSC2v | 6.375082 | 0.109499 | 0.012276 |
| MATS1c | 6.292087 | 0.108074 | 0.012116 |

Table 3: Top 30 important features for RF Mechanistic interpretation model

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Target | Models and sets | R2 | MAE | MSE | RMSE | MSD | MAPE | Adjusted R2 |
| Mtb β-CAs1 | XGBoost\_Train | 0.85 | 0.34 | 0.18 | 0.43 | 0 | 0.06 | 0.85 |
|  | XGBoost\_Test | 0.7 | 0.44 | 0.31 | 0.56 | 0.22 | 0.08 | 0.7 |
|  | Random\_Forest\_Train | 0.98 | 0.09 | 0.02 | 0.14 | 0.01 | 0.02 | 0.98 |
|  | Random\_Forest\_Test | 0.93 | 0.18 | 0.07 | 0.27 | 0.05 | 0.03 | 0.93 |
|  | GB\_Train | 1 | 0.02 | 0 | 0.04 | 0 | 0 | 1 |
|  | GB\_Test | 0.97 | 0.08 | 0.03 | 0.18 | 0.01 | 0.01 | 0.97 |
|  | GL\_Train | 0.64 | 0.53 | 0.45 | 0.67 | 0.01 | 0.09 | 0.64 |
|  | GL\_Test | 0.58 | 0.54 | 0.46 | 0.68 | 0.12 | 0.09 | 0.58 |
| Mtb β-CA2 | XGBoost\_Train | 1 | 0.02 | 0 | 0.02 | 0 | 0 | 1 |
|  | XGBoost\_Test | 1 | 0.04 | 0 | 0.06 | 0.02 | 0.01 | 1 |
|  | RF\_Train | 0.99 | 0.08 | 0.01 | 0.12 | 0 | 0.01 | 0.99 |
|  | RF\_Test | 0.98 | 0.09 | 0.02 | 0.14 | 0.05 | 0.02 | 0.98 |
|  | GB\_Train | 1 | 0.02 | 0 | 0.05 | 0 | 0 | 1 |
|  | GB\_Test | 1 | 0.02 | 0 | 0.03 | 0.02 | 0 | 1 |
|  | GL\_Train | 0.81 | 0.3 | 0.2 | 0.45 | 0.03 | 0.06 | 0.81 |
|  | GL\_Test | 0.95 | 0.17 | 0.04 | 0.2 | 0.07 | 0.03 | 0.95 |
| Mtb β-CA3 | XGBoost\_Train | 0.88 | 0.26 | 0.13 | 0.36 | 0 | 0.04 | 0.88 |
|  | XGBoost\_Test | 0.42 | 0.47 | 0.42 | 0.65 | 0.19 | 0.08 | 0.42 |
|  | RF\_Train | 0.97 | 0.12 | 0.03 | 0.18 | -0.01 | 0.02 | 0.97 |
|  | RF\_Test | 0.9 | 0.25 | 0.11 | 0.34 | -0.03 | 0.04 | 0.9 |
|  | GB\_Train | 0.99 | 0.02 | 0.01 | 0.08 | 0 | 0 | 0.99 |
|  | GB\_Test | 0.97 | 0.11 | 0.04 | 0.19 | 0.03 | 0.02 | 0.97 |
|  | GL\_Train | 0.61 | 0.5 | 0.39 | 0.62 | 0.05 | 0.08 | 0.61 |
|  | GL\_Test | 0.59 | 0.53 | 0.45 | 0.67 | -0.16 | 0.08 | 0.59 |

Table 4: Evaluation metrices of tested machine learning algorithms for regression QSAR model ( MAE: Mean absolute error, MSE: Mean Squared error, MSD: Mean signed difference, MAPE: Mean absolute percentage error)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Target | Models and sets | R2 | MAE | MSE | RMSE | MSD | MAPE | Adjusted R2 |
| Mtb β-CA1 | XGB\_10CV | 0.75 | 0.45 | 0.31 | 0.56 | 0 | 0.07 | 0.75 |
|  | RF\_10CV | 0.88 | 0.27 | 0.14 | 0.38 | 0 | 0.05 | 0.88 |
|  | GB\_10CV | 0.84 | 0.21 | 0.2 | 0.45 | 0 | 0.03 | 0.84 |
|  | GL\_10CV | 0.59 | 0.57 | 0.51 | 0.71 | 0.04 | 0.1 | 0.59 |
| Mtb β-CA2 | XGB\_10CV | 0.74 | 0.2 | 0.27 | 0.52 | -0.15 | 0.03 | 0.74 |
|  | RF\_10CV | 0.85 | 0.27 | 0.15 | 0.39 | 0.01 | 0.05 | 0.85 |
|  | GB\_10CV | 0.92 | 0.15 | 0.08 | 0.28 | 0.02 | 0.03 | 0.92 |
|  | GL\_10CV | 0.64 | 0.41 | 0.38 | 0.61 | 0 | 0.07 | 0.64 |
| Mtb β-CA3 | XGB\_10CV | 0.59 | 0.46 | 0.45 | 0.67 | -0.07 | 0.07 | 0.59 |
|  | RF\_10CV | 0.69 | 0.34 | 0.3 | 0.55 | -0.07 | 0.05 | 0.69 |
|  | GB\_10CV | 0.79 | 0.22 | 0.2 | 0.45 | -0.03 | 0.03 | 0.79 |
|  | GL\_10CV | 0.52 | 0.57 | 0.48 | 0.69 | 0 | 0.09 | 0.52 |

Table 5: 10-Fold cross validation of Regression QSAR models (MAE: Mean absolute error, MSE: Mean Squared error, MSD: Mean signed difference, MAPE: Mean absolute percentage error)

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Feature | Relative Importance | Correlation |
| GB\_ Mtb β-CA1 | nHBint8 | 355.7089 | 0.652725 |
|  | nAtomP | 231.7431 | 0.635068 |
|  | ATSC8m | 143.3329 | -0.1183 |
|  | GATS4c | 96.25777 | 0.301278 |
|  | MLFER\_E | 51.51185 | 0.606587 |
|  | AATS8v | 43.97047 | 0.510538 |
|  | VE1\_Dzs | 43.18784 | 0.541491 |
|  | MATS1m | 28.07761 | -0.0246 |
|  | GATS2m | 26.39813 | -0.29153 |
|  | PubchemFP16 | 25.26062 | 0.312198 |
| GB\_ Mtb β-CA2 | n5Ring | 121.3352 | 0.756362 |
|  | IC2 | 56.33199 | 0.651682 |
|  | ATSC5v | 39.91257 | -0.4901 |
|  | SpMin7\_Bhe | 31.05248 | 0.481068 |
|  | VE3\_Dze | 19.20901 | -0.23679 |
|  | SpMAD\_Dzp | 16.39583 | 0.601416 |
|  | MATS4i | 7.652126 | 0.610733 |
|  | GATS5p | 5.905156 | 0.391311 |
|  | AATS5e | 5.319247 | -0.04416 |
|  | ATSC5m | 4.069537 | -0.37029 |
| GB\_ Mtb β-CA3 | nHBint6 | 281.2664 | 0.636164 |
|  | SCH-7 | 84.36696 | 0.331864 |
|  | minHssNH | 68.6321 | 0.517751 |
|  | SM1\_Dzs | 37.32567 | -0.19937 |
|  | MATS2e | 31.33493 | 0.239875 |
|  | AATS3v | 24.76011 | 0.376814 |
|  | MATS4c | 22.57057 | 0.113647 |
|  | CIC2 | 21.14907 | 0.003216 |
|  | AATSC1m | 21.07899 | 0.104953 |
|  | MATS2c | 20.6598 | 0.269474 |

Table 6: Top 10 important feature for GB\_ Mtb β-CA1-3 QSAR models

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Hub genes | MCC | DMNC | MNC | Degree | EPC | BottleNeck | EcCentricity | Closeness | Radiality | Betweenness | Stress | ClusteringCoefficient |
| TNF | 684635440.00 | 0.3883 | 72 | 72 | 21.009 | 49 | 0.24704 | 116.9167 | 6.32783 | 3655.757 | 26296 | 0.21831 |
| ALB | 602272276.00 | 0.32767 | 69 | 69 | 19.471 | 17 | 0.24704 | 114.25 | 6.26235 | 5415.203 | 32626 | 0.1867 |
| SRC | 620245947.00 | 0.42351 | 63 | 64 | 20.401 | 27 | 0.24704 | 112.3333 | 6.2564 | 3324.7 | 19820 | 0.24058 |
| STAT3 | 688725056.00 | 0.53622 | 57 | 57 | 20.196 | 1 | 0.19763 | 108.5333 | 6.19687 | 1286.494 | 12340 | 0.32456 |
| HIF1A | 683645992.00 | 0.51609 | 50 | 50 | 18.683 | 8 | 0.19763 | 104.2 | 6.12544 | 1144.957 | 9714 | 0.32571 |
| NFKB1 | 683164740.00 | 0.56265 | 50 | 50 | 19.208 | 1 | 0.24704 | 104.25 | 6.13139 | 774.5789 | 8534 | 0.3551 |
| PTGS2 | 506809868.00 | 0.45514 | 49 | 49 | 17.753 | 12 | 0.24704 | 103.5 | 6.11353 | 1506.332 | 12440 | 0.28912 |
| PPARG | 367115834.00 | 0.48137 | 47 | 47 | 17.48 | 5 | 0.24704 | 103.0833 | 6.12544 | 1506.907 | 11722 | 0.3099 |
| MMP9 | 416855600.00 | 0.55828 | 43 | 43 | 18.423 | 1 | 0.19763 | 99.78333 | 6.0421 | 604.2381 | 5388 | 0.36988 |
| CXCR4 | 14821274.00 | 0.55975 | 35 | 35 | 15.4 | 6 | 0.19763 | 96.11667 | 6.01233 | 428.6465 | 4414 | 0.39664 |

Table 7: Hub genes topological parameters