**Supporting Information**

**Activity Cliffs as Protein-Related Phenomenon: Investigation Using Machine Learning Against Numerous Protein Kinases**

Safa Daoud1, Mutasem Taha2\*

1Department of Pharmaceutical Chemistry and Pharmacognosy, Faculty of Pharmacy, Applied Sciences Private University, Amman, Jordan. Email: s\_daoud@asu.edu.jo

2Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan. Email: mutasem@ju.edu.jo

\*Corresponding Author

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Iteration  **Table S1: Y-Scrambled Accuracy and Cohen’s Kappa values determined for XGBoost model using SMOTE-Enhanced Training Data and Leave-20%-Out Cross-validation** | Accuracy | Cohen's kappa | Iteration | Accuracy | Cohen's kappa | Iteration | Accuracy | Cohen's kappa |
| 1 | 0.602941 | 0.112615 | 42 | 0.573529 | 0.067171 | 83 | 0.573529 | 0.051011 |
| 2 | 0.569853 | 0.028097 | 43 | 0.5625 | 0.036896 | 84 | 0.573529 | 0.042718 |
| 3 | 0.544118 | -0.02779 | 44 | 0.536765 | -0.03528 | 85 | 0.577206 | 0.05715 |
| 4 | 0.5 | -0.11745 | 45 | 0.551471 | -0.01121 | 86 | 0.544118 | -0.02779 |
| 5 | 0.536765 | -0.0308 | 46 | 0.518382 | -0.05122 | 87 | 0.522059 | -0.05439 |
| 6 | 0.551471 | 0.006229 | 47 | 0.588235 | 0.075728 | 88 | 0.540441 | -0.00735 |
| 7 | 0.507353 | -0.08683 | 48 | 0.573529 | 0.025692 | 89 | 0.577206 | 0.05715 |
| 8 | 0.577206 | 0.027605 | 49 | 0.551471 | -0.01567 | 90 | 0.533088 | -0.05035 |
| 9 | 0.636029 | 0.191837 | 50 | 0.522059 | -0.07282 | 91 | 0.511029 | -0.08571 |
| 10 | 0.544118 | -0.02779 | 51 | 0.5625 | 0.011483 | 92 | 0.555147 | 0.016495 |
| 11 | 0.625 | 0.147073 | 52 | 0.540441 | -0.05695 | 93 | 0.573529 | 0.051011 |
| 12 | 0.591912 | 0.089945 | 53 | 0.511029 | -0.09043 | 94 | 0.621324 | 0.159184 |
| 13 | 0.555147 | 0.03721 | 54 | 0.628676 | 0.192949 | 95 | 0.488971 | -0.13469 |
| 14 | 0.569853 | 0.032352 | 55 | 0.518382 | -0.09301 | 96 | 0.580882 | 0.05093 |
| 15 | 0.529412 | -0.03373 | 56 | 0.5625 | 0.020102 | 97 | 0.522059 | -0.07282 |
| 16 | 0.584559 | 0.057178 | 57 | 0.533088 | -0.05963 | 98 | 0.540441 | -0.0293 |
| 17 | 0.496324 | -0.12322 | 58 | 0.5625 | 0.020102 | 99 | 0.514706 | -0.09892 |
| 18 | 0.503676 | -0.10204 | 59 | 0.591912 | 0.097777 | 100 | 0.595588 | 0.103977 |
| 19 | 0.547794 | 0.004523 | 60 | 0.514706 | -0.10379 |  |  |  |
| 20 | 0.591912 | 0.109275 | 61 | 0.584559 | 0.077551 |  |  |  |
| 21 | 0.533088 | -0.02785 | 62 | 0.555147 | 0.003633 |  |  |  |
| 22 | 0.522059 | -0.07282 | 63 | 0.511029 | -0.1246 |  |  |  |
| 23 | 0.540441 | -0.0293 | 64 | 0.591912 | 0.085977 |  |  |  |
| 24 | 0.511029 | -0.09518 | 65 | 0.511029 | -0.10481 |  |  |  |
| 25 | 0.555147 | -0.00513 | 66 | 0.525735 | -0.04404 |  |  |  |
| 26 | 0.547794 | 0.004523 | 67 | 0.511029 | -0.09518 |  |  |  |
| 27 | 0.566176 | 0.038812 | 68 | 0.522059 | -0.07752 |  |  |  |
| 28 | 0.547794 | -0.04472 | 69 | 0.522059 | -0.07282 |  |  |  |
| 29 | 0.518382 | -0.04236 | 70 | 0.555147 | -0.02773 |  |  |  |
| 30 | 0.536765 | -0.03528 | 71 | 0.533088 | -0.02347 |  |  |  |
| 31 | 0.544118 | -0.01885 | 72 | 0.591912 | 0.081975 |  |  |  |
| 32 | 0.540441 | -0.04294 | 73 | 0.529412 | -0.06095 |  |  |  |
| 33 | 0.496324 | -0.14307 | 74 | 0.580882 | 0.059223 |  |  |  |
| 34 | 0.595588 | 0.084231 | 75 | 0.621324 | 0.148139 |  |  |  |
| 35 | 0.518382 | -0.06479 | 76 | 0.610294 | 0.129048 |  |  |  |
| 36 | 0.536765 | -0.0583 | 77 | 0.540441 | -0.00735 |  |  |  |
| 37 | 0.485294 | -0.16041 | 78 | 0.529412 | -0.05631 |  |  |  |
| 38 | 0.503676 | -0.10204 | 79 | 0.566176 | 0.026214 |  |  |  |
| 39 | 0.540441 | -0.03836 | 80 | 0.514706 | -0.07061 |  |  |  |
| 40 | 0.580882 | 0.05093 | 81 | 0.584559 | 0.08152 |  |  |  |
| 41 | 0.507353 | -0.11557 | 82 | 0.573529 | 0.051011 |  |  |  |
| Iteration  **Table S2: Y-Scrambled Accuracy and Cohen’s Kappa values determined for XGBoost model using Training Data (Unaugmented) and Leave-20%-Out Cross-validation** | Accuracy | Cohen's kappa | Iteration | Accuracy | Cohen's kappa | Iteration | Accuracy | Cohen's kappa |
| 1 | 0.5 | -0.13222 | 42 | 0.529412 | -0.01209 | 83 | 0.544118 | -0.06036 |
| 2 | 0.588235 | 0.083734 | 43 | 0.514706 | -0.08932 | 84 | 0.529412 | -0.04716 |
| 3 | 0.5 | -0.1736 | 44 | 0.544118 | -0.00573 | 85 | 0.588235 | 0.050847 |
| 4 | 0.647059 | 0.214629 | 45 | 0.5 | -0.15254 | 86 | 0.470588 | -0.1386 |
| 5 | 0.602941 | 0.108738 | 46 | 0.470588 | -0.17806 | 87 | 0.441176 | -0.26543 |
| 6 | 0.529412 | -0.06562 | 47 | 0.661765 | 0.240777 | 88 | 0.558824 | 9.79E-04 |
| 7 | 0.573529 | 0.05916 | 48 | 0.514706 | -0.12877 | 89 | 0.485294 | -0.17589 |
| 8 | 0.455882 | -0.20038 | 49 | 0.632353 | 0.129098 | 90 | 0.5 | -0.13222 |
| 9 | 0.617647 | 0.134182 | 50 | 0.529412 | -0.06562 | 91 | 0.485294 | -0.1355 |
| 10 | 0.676471 | 0.267385 | 51 | 0.544118 | -0.06036 | 92 | 0.573529 | 0.008048 |
| 11 | 0.455882 | -0.22136 | 52 | 0.602941 | 0.108738 | 93 | 0.588235 | 0.083734 |
| 12 | 0.455882 | -0.26559 | 53 | 0.441176 | -0.26543 | 94 | 0.529412 | -0.02933 |
| 13 | 0.426471 | -0.28738 | 54 | 0.588235 | 0.099338 | 95 | 0.441176 | -0.18207 |
| 14 | 0.514706 | -0.08932 | 55 | 0.470588 | -0.24264 | 96 | 0.544118 | -0.0233 |
| 15 | 0.544118 | -0.00573 | 56 | 0.529412 | -0.04716 | 97 | 0.544118 | -0.00573 |
| 16 | 0.485294 | -0.1355 | 57 | 0.558824 | -0.03553 | 98 | 0.455882 | -0.22136 |
| 17 | 0.588235 | 0.067581 | 58 | 0.676471 | 0.280077 | 99 | 0.588235 | -0.00316 |
| 18 | 0.544118 | -0.00573 | 59 | 0.588235 | 0.083734 | 100 | 0.588235 | 0.099338 |
| 19 | 0.632353 | 0.144869 | 60 | 0.529412 | -0.04716 |  |  |  |
| 20 | 0.5 | -0.11261 | 61 | 0.602941 | 0.092885 |  |  |  |
| 21 | 0.529412 | -0.08475 | 62 | 0.558824 | 0.035005 |  |  |  |
| 22 | 0.514706 | -0.14959 | 63 | 0.588235 | 0.067581 |  |  |  |
| 23 | 0.632353 | 0.174757 | 64 | 0.558824 | 0.018287 |  |  |  |
| 24 | 0.544118 | 0.011257 | 65 | 0.588235 | 0.083734 |  |  |  |
| 25 | 0.691176 | 0.306796 | 66 | 0.573529 | 0.042718 |  |  |  |
| 26 | 0.529412 | -0.04716 | 67 | 0.647059 | 0.24093 |  |  |  |
| 27 | 0.588235 | 0.067581 | 68 | 0.514706 | -0.1087 |  |  |  |
| 28 | 0.544118 | -0.0415 | 69 | 0.676471 | 0.292337 |  |  |  |
| 29 | 0.588235 | 0.050847 | 70 | 0.573529 | 0.025692 |  |  |  |
| 30 | 0.544118 | -0.0415 | 71 | 0.602941 | 0.092885 |  |  |  |
| 31 | 0.573529 | 0.05916 | 72 | 0.529412 | -0.02933 |  |  |  |
| 32 | 0.558824 | -0.01695 | 73 | 0.544118 | -0.00573 |  |  |  |
| 33 | 0.485294 | -0.19718 | 74 | 0.573529 | 0.042718 |  |  |  |
| 34 | 0.441176 | -0.26543 | 75 | 0.661765 | 0.253817 |  |  |  |
| 35 | 0.558824 | 0.035005 | 76 | 0.602941 | 0.138837 |  |  |  |
| 36 | 0.632353 | 0.144869 | 77 | 0.397059 | -0.30769 |  |  |  |
| 37 | 0.632353 | 0.215867 | 78 | 0.588235 | 0.033503 |  |  |  |
| 38 | 0.705882 | 0.322034 | 79 | 0.617647 | 0.102538 |  |  |  |
| 39 | 0.617647 | 0.149182 | 80 | 0.441176 | -0.22233 |  |  |  |
| 40 | 0.485294 | -0.17589 | 81 | 0.411765 | -0.24428 |  |  |  |
| 41 | 0.558824 | 9.79E-04 | 82 | 0.588235 | 0.015512 |  |  |  |
| Iteration  **Table S3: Y-Scrambled Accuracy and Cohen’s Kappa values determined for K\*model using SMOTE-Enhanced Training Data and Leave-20%-Out Cross-validation** | Accuracy | Cohen's kappa | Iteration | Accuracy | Cohen's kappa | Iteration | Accuracy | Cohen's kappa |
| 1 | 0.551471 | 0.023081 | 42 | 0.481618 | -0.14116 | 83 | 0.584559 | 0.089358 |
| 2 | 0.551471 | 0.027205 | 43 | 0.621324 | 0.166389 | 84 | 0.569853 | 0.044898 |
| 3 | 0.525735 | -0.02214 | 44 | 0.507353 | -0.09152 | 85 | 0.551471 | 0.023081 |
| 4 | 0.544118 | 0.011257 | 45 | 0.533088 | -0.04577 | 86 | 0.540441 | 0.009555 |
| 5 | 0.595588 | 0.122889 | 46 | 0.525735 | -0.01787 | 87 | 0.555147 | 0.053165 |
| 6 | 0.540441 | -0.00735 | 47 | 0.540441 | -0.00307 | 88 | 0.547794 | 0.021297 |
| 7 | 0.540441 | -0.01602 | 48 | 0.474265 | -0.12364 | 89 | 0.5625 | 0.036896 |
| 8 | 0.540441 | -0.00735 | 49 | 0.503676 | -0.08331 | 90 | 0.507353 | -0.05512 |
| 9 | 0.518382 | -0.08343 | 50 | 0.580882 | 0.094816 | 91 | 0.610294 | 0.154784 |
| 10 | 0.511029 | -0.07641 | 51 | 0.511029 | -0.06275 | 92 | 0.533088 | -0.01912 |
| 11 | 0.496324 | -0.0901 | 52 | 0.533088 | -0.02347 | 93 | 0.511029 | -0.10481 |
| 12 | 0.566176 | 0.034649 | 53 | 0.606618 | 0.145006 | 94 | 0.496324 | -0.0901 |
| 13 | 0.573529 | 0.067171 | 54 | 0.536765 | -0.01758 | 95 | 0.577206 | 0.084952 |
| 14 | 0.507353 | -0.06399 | 55 | 0.5625 | 0.061035 | 96 | 0.555147 | 0.033137 |
| 15 | 0.507353 | -0.06848 | 56 | 0.558824 | 0.026718 | 97 | 0.529412 | -0.03817 |
| 16 | 0.5625 | 0.049119 | 57 | 0.544118 | 0.002838 | 98 | 0.566176 | 0.066977 |
| 17 | 0.522059 | -0.02791 | 58 | 0.573529 | 0.05916 | 99 | 0.533088 | -0.03673 |
| 18 | 0.569853 | 0.036571 | 59 | 0.525735 | -0.02214 | 100 | 0.540441 | -0.01602 |
| 19 | 0.577206 | 0.07324 | 60 | 0.569853 | 0.049008 |  |  |  |
| 20 | 0.492647 | -0.09575 | 61 | 0.540441 | 0.005383 |  |  |  |
| 21 | 0.580882 | 0.094816 | 62 | 0.518382 | -0.06479 |  |  |  |
| 22 | 0.485294 | -0.12106 | 63 | 0.547794 | 0.004523 |  |  |  |
| 23 | 0.514706 | -0.06149 | 64 | 0.551471 | 0.031294 |  |  |  |
| 24 | 0.533088 | -0.02785 | 65 | 0.558824 | 0.039096 |  |  |  |
| 25 | 0.5 | -0.07987 | 66 | 0.547794 | 0.004523 |  |  |  |
| 26 | 0.613971 | 0.160987 | 67 | 0.544118 | 0.011257 |  |  |  |
| 27 | 0.466912 | -0.14892 | 68 | 0.588235 | 0.106942 |  |  |  |
| 28 | 0.580882 | 0.106089 | 69 | 0.503676 | -0.05639 |  |  |  |
| 29 | 0.544118 | 0.019535 | 70 | 0.488971 | -0.12497 |  |  |  |
| 30 | 0.477941 | -0.14191 | 71 | 0.558824 | 0.055118 |  |  |  |
| 31 | 0.555147 | 0.033137 | 72 | 0.470588 | -0.11526 |  |  |  |
| 32 | 0.470588 | -0.14822 | 73 | 0.540441 | 0.001175 |  |  |  |
| 33 | 0.580882 | 0.079335 | 74 | 0.525735 | -0.01364 |  |  |  |
| 34 | 0.577206 | 0.069269 | 75 | 0.496324 | -0.10405 |  |  |  |
| 35 | 0.577206 | 0.081081 | 76 | 0.511029 | -0.07641 |  |  |  |
| 36 | 0.540441 | 0.029902 | 77 | 0.5 | -0.07987 |  |  |  |
| 37 | 0.514706 | -0.05253 | 78 | 0.5625 | 0.049119 |  |  |  |
| 38 | 0.547794 | 0.012981 | 79 | 0.547794 | 0.021297 |  |  |  |
| 39 | 0.584559 | 0.08152 | 80 | 0.507353 | -0.06399 |  |  |  |
| 40 | 0.529412 | -0.02496 | 81 | 0.573529 | 0.075047 |  |  |  |
| 41 | 0.580882 | 0.094816 | 82 | 0.525735 | -0.01364 |  |  |  |
| Iteration  **Table S4: Y-Scrambled Accuracy and Cohen’s Kappa values determined for K\*model using Training Data (Unaugmented) and Leave-20%-Out Cross-validation** | Accuracy | Cohen's kappa | Iteration | Accuracy | Cohen's kappa | Iteration | Accuracy | Cohen's kappa |
| 1 | 0.632353 | 0.202627 | 42 | 0.441176 | -0.18207 | 83 | 0.529412 | 0.020702 |
| 2 | 0.411765 | -0.28666 | 43 | 0.441176 | -0.22233 | 84 | 0.514706 | -0.05253 |
| 3 | 0.441176 | -0.18207 | 44 | 0.573529 | 0.105263 | 85 | 0.485294 | -0.17589 |
| 4 | 0.485294 | -0.11632 | 45 | 0.485294 | -0.0625 | 86 | 0.470588 | -0.17806 |
| 5 | 0.647059 | 0.186441 | 46 | 0.411765 | -0.24428 | 87 | 0.602941 | 0.076459 |
| 6 | 0.544118 | -0.06036 | 47 | 0.573529 | 0.119643 | 88 | 0.485294 | -0.11632 |
| 7 | 0.602941 | 0.138837 | 48 | 0.485294 | -0.1355 | 89 | 0.588235 | 0.015512 |
| 8 | 0.5 | -0.05764 | 49 | 0.485294 | -0.0625 | 90 | 0.602941 | 0.059426 |
| 9 | 0.470588 | -0.1386 | 50 | 0.588235 | 0.050847 | 91 | 0.529412 | -0.02933 |
| 10 | 0.602941 | 0.166969 | 51 | 0.602941 | 0.153137 | 92 | 0.647059 | 0.277236 |
| 11 | 0.617647 | 0.177674 | 52 | 0.5 | -0.07535 | 93 | 0.661765 | 0.253817 |
| 12 | 0.514706 | -0.07061 | 53 | 0.514706 | -0.03506 | 94 | 0.455882 | -0.10545 |
| 13 | 0.602941 | 0.059426 | 54 | 0.441176 | -0.2435 | 95 | 0.5 | -0.09366 |
| 14 | 0.558824 | 0.035005 | 55 | 0.544118 | -0.0233 | 96 | 0.558824 | 0.018287 |
| 15 | 0.529412 | -0.04716 | 56 | 0.529412 | -0.08475 | 97 | 0.558824 | 0.110724 |
| 16 | 0.573529 | 0.119643 | 57 | 0.514706 | -0.14959 | 98 | 0.544118 | 0.043557 |
| 17 | 0.617647 | 0.252747 | 58 | 0.602941 | 0.124046 | 99 | 0.558824 | 0.051163 |
| 18 | 0.617647 | 0.20432 | 59 | 0.573529 | 0.042718 | 100 | 0.720588 | 0.372816 |
| 19 | 0.470588 | -0.1386 | 60 | 0.602941 | 0.092885 |  |  |  |
| 20 | 0.602941 | 0.059426 | 61 | 0.5 | -0.0405 |  |  |  |
| 21 | 0.514706 | -0.01815 | 62 | 0.441176 | -0.20186 |  |  |  |
| 22 | 0.470588 | -0.10171 | 63 | 0.588235 | 0.015512 |  |  |  |
| 23 | 0.544118 | 0.043557 | 64 | 0.544118 | -0.10021 |  |  |  |
| 24 | 0.632353 | 0.202627 | 65 | 0.544118 | -0.0233 |  |  |  |
| 25 | 0.558824 | 0.035005 | 66 | 0.441176 | -0.2435 |  |  |  |
| 26 | 0.411765 | -0.20461 | 67 | 0.558824 | 0.035005 |  |  |  |
| 27 | 0.5 | 0.007725 | 68 | 0.632353 | 0.160079 |  |  |  |
| 28 | 0.558824 | 0.066789 | 69 | 0.676471 | 0.280077 |  |  |  |
| 29 | 0.647059 | 0.214629 | 70 | 0.5 | -0.13222 |  |  |  |
| 30 | 0.544118 | -0.06036 | 71 | 0.573529 | 0.008048 |  |  |  |
| 31 | 0.617647 | 0.191217 | 72 | 0.441176 | -0.10901 |  |  |  |
| 32 | 0.602941 | 0.108738 | 73 | 0.426471 | -0.22325 |  |  |  |
| 33 | 0.529412 | -0.08475 | 74 | 0.470588 | -0.15799 |  |  |  |
| 34 | 0.5 | -0.11261 | 75 | 0.529412 | -0.02933 |  |  |  |
| 35 | 0.647059 | 0.228004 | 76 | 0.602941 | 0.041754 |  |  |  |
| 36 | 0.411765 | -0.16738 | 77 | 0.632353 | 0.215867 |  |  |  |
| 37 | 0.573529 | 0.075047 | 78 | 0.558824 | 0.081908 |  |  |  |
| 38 | 0.632353 | 0.202627 | 79 | 0.529412 | 0.004575 |  |  |  |
| 39 | 0.470588 | -0.10171 | 80 | 0.544118 | -0.0233 |  |  |  |
| 40 | 0.5 | -0.09366 | 81 | 0.397059 | -0.26497 |  |  |  |
| 41 | 0.411765 | -0.22412 | 82 | 0.544118 | -0.0233 |  |  |  |