**Analysis of variations in input feature set with SVM performance**

In our current solution to domain identification problem, we are employing SVM at two stages, first to classify between single & multi-domain proteins and next in classifying between multi-domain proteins. The below set of experiments compare the performance of SVM at each stage when different training/testing input feature sets are used. The test is done for benchmark\_2, benchmark\_3, ASTRAL SCOP30 and our own created dataset.

For Phase I SVM: Classifying single & multi-domain proteins

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Feature Set** | **Benchmark2** | | | | | **Benchmark3** | | | | | **ASTRAL SCOP30** | | | **Self Created** | | |
|  | Single Correct | Multi Correct | | | Overall | Single Correct | Multi Correct | | | Overall | Single Correct | Multi Correct | Overall | Single Correct | Multi Correct | Overall |
|  |  |
| Contiguous | Non-Contiguous | | Contiguous | Non-Contiguous | |
| **L** | 41/54  (75.93%) | 65/98  (66.33%) | | | 106/152  (66.33%) | 42/55  (76.36%) | 51/77  (66.23% | | | 93/132  (66.23%) |  |  |  |  |  |  |
| 36/66  (54.55%) | 29/32  (90.62%) | | 33/59  (55.93%) | 18/18  (100%) | |  |  |
| **L I** | 43/54  (79.63%) | 62/98  (63.27%) | | | 105/152  (63.27%) | 44/55  (80%) | 49/77  (63.64%) | | | 93/132  (63.64%) |  |  |  |  |  |  |
| 33/66  (50%) | | 29/32  (90.62%) | 31/59  (52.54%) | | 18/18  (100%) |  |  |
| **L I D** | 44/54  (81.48%) | 64/98  (65.31%) | | | 108/152  (65.31%) | 45/55  (81.82%) | 51/77  (66.23%) | | | 96/132  (66.23%) |  |  |  |  |  |  |
| 35/66  (53.03%) | | 29/32  (90.62%) | 33/59  (55.93%) | 18/18  (100%) | |  |  |
| **L I D R** | 39/54  (72.22%) | 78/98  (79.59%) | | | 117/152  (79.59%) | 40/55  (72.73%) | 61/77  (79.22%) | | | 101/132  (79.22%) |  |  |  |  |  |  |
| 48/66  (72.73%) | | 30/32  (93.75%) | 44/59  (74.58%) | 17/18  (94.44%) | |  |  |
| **I** | 47/54  (87.04%) | 86/98  (87.76%) | | | 133/152  (87.76%) | 47/55  (85.45%) | 70/77  (90.91%) | | | 117/132  (90.91%) |  |  |  |  |  |  |
| 62/66  (93.94%) | | 24/32  (75.00%) | 57/59  (96.61%) | 13/18  (72.22%) | |  |  |
| **I D** | 47/54  (87.04%) | 86/98  (87.76%) | | | 133/152  (87.76%) | 47/55  (85.45%) | 70/77  (90.91%) | | | 117/132  (90.91%) |  |  |  |  |  |  |
| 62/66  (93.94%) | | 24/32  (75%) | 57/59  (96.61%) | 13/18  (90.91%) | |  |  |
| **I D R** | 44/54  (81.48%) | 77/98  (78.57%) | | | 121/152  (78.57%) | 45/55  (81.82%) | 61/77  (79.22%) | | | 106/132  (79.22%) |  |  |  |  |  |  |
| 49/66  (74.24%) | | 28/32  (87.50%) | 44/59  (74.58%) | 17/18  (94.44%) | |  |  |
| **D** | 41/54  (75.93%) | 65/98  (66.33%) | | | 106/152  (66.33%) | 42/55  (76.36%) | 52/77  (67.53%) | | | 94/132  (67.53%) |  |  |  |  |  |  |
| 47/66  (71.21%) | | 18/32  (56.25%) | 42/59  (71.19%) | 10/18  (55.56%) | |  |  |
| **D R** | 43/54  (79.63%) | 76/98  (77.55%) | | | 119/152  (77.55%) | 44/55  (80.00%) | 60/77  (77.92%) | | | 104/132  (77.92%) |  |  |  |  |  |  |
| 47/66  (71.21%) | | 29/32  (90.62%) | 43/59  (72.88%) | 17/18  (94.44%) | |  |  |
| **R** | 45/54  (83.33%) | 77/98  (78.57%) | | | 122/152  (78.57%) | 46/55  (83.64%) | 61/77  (79.22%) | | | 107/132  (79.22%) |  |  |  |  |  |  |
| 48/66  (72.73%) | | 29/32  (90.62%) | 44/59  (74.58%) | 17/18  (94.44%) | |  |  |