TCGA\_blca 5 subtypes prediction initial results:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Classes | Luminal\_infiltrated (LI) | Luminal\_papillary (LP) | Luminal (L) | Basal\_squamous (BS) | Neuronal (N) |
| Numbers | 74 | 128 | 26 | 130 | 19 |

Using multi-class SVM classifier with leave-one-out cross validation: we get the following confusion matrix:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | LI | LP | L | BS | N |
| LI | 14 | 28 | 2 | 27 | 3 |
| LP | 11 | 96 | 0 | 21 | 0 |
| L | 8 | 11 | 0 | 6 | 1 |
| BS | 12 | 28 | 0 | 88 | 2 |
| N | 4 | 7 | 1 | 3 | 4 |

Overall classification accuracy=53.55%

Accuracy for LI = 18.92%

Accuracy for LP=75.00%

Accuracy for L=0.0%

Accuracy for BS=67.69%

Accuracy for N=8.16%

Because the unbalanced dataset, it is very hard to predict less representative classes.