The A Team

Arnab Sarkar

Adora D'Souza

Alykhan Alani

Anas Abidin

Approach

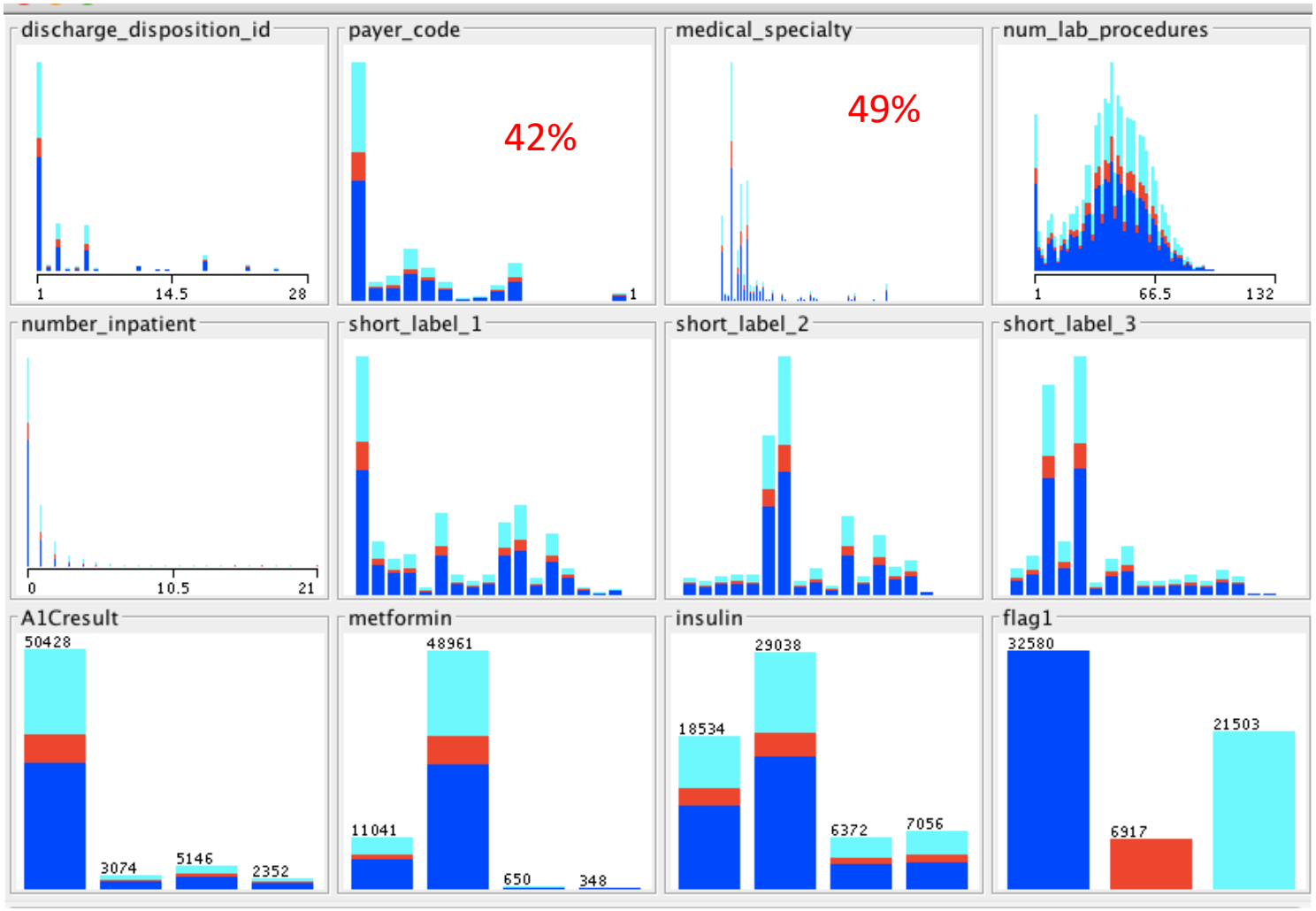
- Instead of using the entire set of feature for prediction, we reduce the dimensionality
  - This was tried using:
    - PCA
    - Mutual Information
    - GMLVQ
- Not all features are equally important or relevant for prediction of readmission

Top features based on predictive value

```
18 long_label_1
17 short_label_1
10 medical_specialty
20 long_label_2
19 short_label_2
22 long_label_3
21 short_label_3
 9 payer_code
16 number_inpatient
43 insulin
26 metformin
 6 discharge_disposition_id
12 num_procedures
25 A1Cresult
 2 gender
```

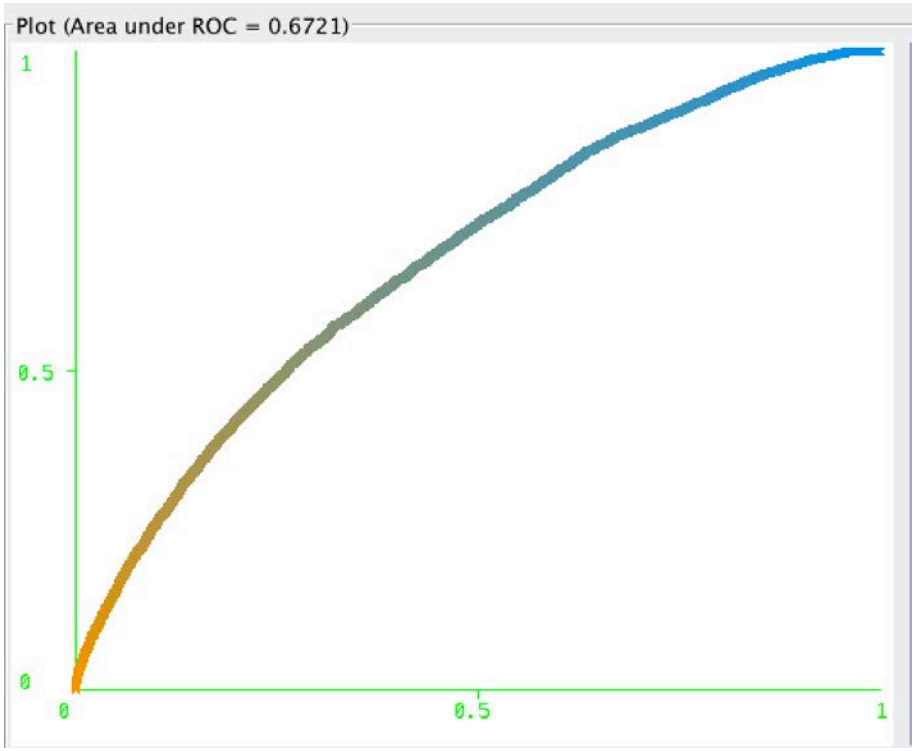

- Lee, E. W. (2012). Selecting the best prediction model for readmission. Journal of Preventive Medicine and Public Health, 45(4), 259-266.
- Futoma, Joseph, Jonathan Morris, and Joseph Lucas. "A comparison of models for predicting early hospital readmissions." Journal of biomedical informatics 56 (2015): 229-238.

Distribution of best features



Findings – 2 class problem

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       11441               62.5191 %
Incorrectly Classified Instances      6859               37.4809 %
Kappa statistic                          0.2426
Mean absolute error                      0.4234
Root mean squared error                  0.4848
Relative absolute error                 85.0789 %
Root relative squared error             97.1946 %
Coverage of cases (0.95 level)          99.3934 %
Mean rel. region size (0.95 level)      97.388  %
Total Number of Instances            18300

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall
                 0.688    0.447    0.638      0.688
                 0.553    0.312    0.607      0.553
Weighted Avg.    0.625    0.384    0.624      0.625

=== Confusion Matrix ===

    a    b   <-- classified as
 6724 3050 |    a = 0
 3809 4717 |    b = 1
```



Findings – 3 class problem

```
Correctly Classified Instances       22238               45.5715 %
Incorrectly Classified Instances     26560               54.4285 %
Kappa statistic                          0.1836
Mean absolute error                      0.405
Root mean squared error                  0.4572
Relative absolute error                 91.1257 %
Root relative squared error             96.9973 %
Coverage of cases (0.95 level)          99.6229 %
Mean rel. region size (0.95 level)      97.6365 %
Total Number of Instances            48798

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall
                 0.528    0.293    0.474      0.528
                 0.440    0.237    0.481      0.440
                 0.400    0.286    0.411      0.400
Weighted Avg.    0.456    0.272    0.455      0.456

=== Confusion Matrix ===

    a    b    c   <-- classified as
 8583 3191 4492 |    a = 0
 4297 7152 4817 |    b = 1
 5242 4521 6503 |    c = 2
```